Sources of Variation in US Mortality: A Latent Variable Analysis

Andrew Stokes and Chris Tencza

Population Studies Center
University of Pennsylvania

September 2012

Abstract

Modifiable risk factors exert a strong influence on health and mortality in the United States. Estimation of the effects of modifiable risk factors is challenging as US mortality data do not provide information on the risk factor(s) that gave rise to a particular disease and surveys are rarely large enough to permit detailed analyses. We explore a new indirect method for identifying and estimating mortality effects of leading risk factors, which relies on vital statistics data. Our basic assumption is that most spatial variation in cause-specific mortality rates is a manifestation of a small number of latent variables, variation in which give rise to the observed mortality patterns. We apply factor analysis to extract the major components of mortality variation. We find that the majority of US mortality variation is explained by a small number of factors and that these factors can be meaningfully interpreted in terms of known modifiable risk factors.

Introduction

There is a longstanding interest among demographers in understanding the effects of modifiable risk factors on health and mortality. Prior research on the "actual causes of death" has concluded that about half of annual deaths in the United States are attributable to modifiable risk factors [1] [2]. Recent research has extended the number of modifiable risk factors examined and explored their role in US mortality disparities by gender, race/ethnicity and geography [3] [4]. In addition, the role of modifiable risk factors in poor US performance in international longevity rankings has been examined [5] [6]. The evidence suggests that targeting just 4 of the leading modifiable risk factors could increase overall life expectancy in the United States by more than 4 years and reduce differences in life expectancy across population sub-groups by as much as 20% [4]. Taken together, the accumulated evidence provides strong support for the notion that public health efforts aimed at modifiable risk factors may be the most cost-effective means of improving population health in the US.

Data limitations pose a significant challenge for estimating the impact of risk factors on mortality in a valid and comparable manner. While the National Center for Health Statistics maintains an extensive and high-quality time-series of cause-specific mortality with gender, age, race/ethnicity and geographic detail, information on the risk factor(s) that gave rise to the disease or injury is not available. Thus, studies have traditionally relied on survey data, which has several limitations. First, data are typically only available for a small subset of risk factors. Second, exposure data are often elicited through self-report, introducing systematic errors. Third, surveys are typically not powered for investigating time-trends and/or geographic patterns for small areas such as US counties.

These limitations suggest that new methods for identifying and quantifying modifiable risk factors and their effects on mortality are needed. Indirect attributable risk methods, referred to as such because the effects of a risk factor are investigated through a physical outcome to which the risk is closely related, may represent a promising alternative. These methods have previously been developed to assess smoking-attributable mortality in the United States and internationally [5] [7]. In the previous studies, the death rate from lung-cancer was used as a marker of the cumulative damage caused by smoking and various methods were subsequently applied to translate these rates into estimates of total smoking attributable mortality. A limitation of the previous studies is that they focused only on smoking and were not optimized for estimating the effects of multiple risk factors simultaneously.

In this paper, we devise a new indirect method for identifying and estimating mortality effects of leading risk factors, which relies exclusively on vital statistics data. We assume that a large part of the variation in cause-specific mortality rates over states is a manifestation of a small number of latent variables, variation in which give rise to the observed mortality patterns. We apply factor analysis to the cause-specific mortality data, separately by sex, to extract the

major components of mortality variation. Then we use regression methods to estimate the relative importance of these components to variation in all-cause mortality.[1]

Methods

The premise of our approach is that a few underlying factors determine the large majority of variation in cause-specific mortality across US states. Although these factors are not directly observable, we hypothesize that they are indirectly identifiable through the imprint they leave on patterns of mortality. Latent variables responsible for mortality variation may fall into one of a number of domains that affect health. They may be related to the quality and availability of medical services or to lifestyle (e.g. smoking, alcohol), environmental (e.g. air pollution, pesticides) and/or structural (e.g. poverty) factors [8] [9]. Our methodological strategy is factor analysis, a method for determining the structure embedded in a set of variables and describing that structure in terms of a set of factors which are construed as being representative of latent variables. We argue that this method is appropriate here as individual causes of death are each partial and indirect representations of underlying risk factors. For example, variation in the obesity rate may jointly affect the diabetes and ischemic disease death rates such that there is some common variation among these two causes of death. However, variation in these death rates may also reflect other factors unique to each cause of death, such as cause-of-death certification practices. In this study, factor analysis provides a manner of partitioning the variance among the entire set of causes of death into a component of shared variance (shared across multiple variables) and a component of unique variance (variance specific to a particular variable and random error).

We implement the analysis separately by sex, on state-by-cause matrices. Orthogonal rotation (varimax) is specified and the first 5 factors are retained based on examination of scree-plots. We then implement ordinary least squares regression with the all-cause mortality rate on the left-hand side and scores on each of the retained factors on the right-hand side in order to obtain the association of each latent variable with all-cause mortality.

Mortality data by underlying cause of death, sex, age and US state were obtained for years 2000-2004 from the National Center for Health Statistics and denominators for the corresponding years were derived from the bridged-race population files obtained from the Census Bureau. Causes of death were classified using a list of 56 causes at three different levels of aggregation developed at the University of Washington [10]. The motivation for using this list was two-fold: first, the list was designed to maximize the comparability of cause-specific mortality data over time and across geographic units. Second, the categories were formulated based on considerations of the burden of disease and public health policy. Cause-specific death rates were calculated on the combined 2000-2004 data by sex and age-standardized to the US

[1] Latent variable methods have previously been used to analyze cause-specific mortality patterns and their determinants. In one study, temporal variation in Russian cause-specific mortality was analyzed separately by sex using select causes of death [11]. However, the prior analysis was primarily exploratory and did not seek to estimate the relative contribution of each of the identified factors to overall mortality levels.

2000 census population. Within each sex, if a particular cause contributed fewer than 1000 deaths over the interval 2000-2004, it was eliminated from the analysis. All analyses were conducted using Stata version 11 (Stata Corp, Texas) and R (Gentleman et al.).

Results

Forty-five causes of death for men and 46 for women were entered into the factor analysis. For both males and females, the first factor explained a fourth of the total variance in the data and five factors were responsible for two-thirds of the variance (Table 1). Subsequent factors explained progressively less variance such that ten factors were required to reach 80% of cumulative variance explained and 16 factors to reach 90% of cumulative variance explained. Based on the results in Table 1, we chose to retain five factors for further analysis. Tables 2 and 3 show factor loadings for the five retained factors by sex. The values in these tables represent the correlations between each factor and cause of death. Particularly strong correlations, (ρ greater than 0.50 in absolute value) are highlighted to aid interpretation of the loading patterns. Figures 1 and 2 are US state maps of the factor scores plotted on a continuous scale. In the presentation of the preliminary results, we focus on those for men. Factor 1 appears to be a smoking factor with notably heavy loadings of lung cancer, cerebrovascular disease and COPD. States with high scores on this factor are primarily located in the South. Factor 2 exhibits strong correlations with liver cancer and alcoholic cirrhosis of the liver, suggesting a role for alcohol in this factor. Factor 3 exhibits strong and consistent correlations with accidents and injuries, including motor-vehicle accidents and suicide. States with high scores on this factor are primarily concentrated in the Rocky Mountains region and include Montana, Colorado, Arizona and New Mexico. Factor 4 exhibits strong correlations with HIV/AIDS and homicide. States with high factor scores include Louisiana and Mississippi. Factor 5 is strongly associated with other cancers, diabetes and endocrine disorders, suggesting a potential role for obesity or variation in cause-of-death certification practices. Table 4 shows the OLS model parameters obtained from regressing the age-standardized all-cause mortality rates on the scores for each of the retained factors separately by sex. For males, the highest coefficient observed is that for factor 1, which we have identified as a smoking factor. The next highest coefficient is observed for factor 3, the factor associated with accidents and injuries.

References

1. Mcginnis M, Foege W (1993) Actual causes of death in the United States. JAMA 270. doi:10.1001/jama.291.10.1238.

2. Mokdad AH (2005) Actual Causes of Death in the United States, 2000. JAMA 291: 1238–1245.

3. Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, et al. (2009) The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS Med 6.

4. Danaei G, Rimm EB, Oza S, Kulkarni SC, Murray CJL, et al. (2010) The promise of prevention: the effects of four preventable risk factors on national life expectancy and life expectancy disparities by race and county in the United States. PLoS Med 7: e1000248.

5. Preston SH, Glei DA, Wilmoth JR (2010) Contribution of Smoking to International Differences in Life Expectancy. In: Crimmins E, Preston S, Cohen B, editors. International Differences in Mortality at Older Ages: Dimensions and Sources. Washington, D.C.: National Academy Press.

6. Preston SH, Stokes A (2011) Contribution of obesity to international differences in life expectancy. American journal of public health 101: 2137–2143. doi:10.2105/AJPH.2011.300219.

7. Peto R, Lopez AD, Boreham J, Thun M (2006) Mortality From Smoking in Developed Countries 1950-2000. 2nd Editio. p.

8. McGinnis JM, Williams-Russo P, Knickman JR (2002) The Case For More Active Policy Attention To Health Promotion. Health Affairs 21: 78–93. doi:10.1377/hlthaff.21.2.78.

9. Schroeder SA (2007) We Can Do Better - Improving the Health of the American People: 1221–1228.

10. Naghavi M, Makela S, Foreman K, Brien JO, Pourmalek F, et al. (2010) Algorithms for enhancing public health utility of national causes-of-death data: 1–14.

11. Gavrilova NS, Semyonova VG (2002) Mortality Crisis in Russia : New Health Threats. Population (English Edition): 1–33.

Table 1. Eigenvalues, Discrete and Cumulative Proportion of Variance Explained.						
	Male			Female		
Factor	**Eigenvalues**	**Proportion**	**Cumulative**	**Eigenvalues**	**Proportion**	**Cumulative**
1	11.856347	0.263474	0.263474	11.663528	0.253555	0.253555
2	6.91211	0.153602	0.417077	7.590472	0.16501	0.418565
3	6.061302	0.134696	0.551772	5.405888	0.117519	0.536085
4	2.606227	0.057916	0.609689	3.14187	0.068302	0.604386
5	2.390742	0.053128	0.662816	2.432141	0.052873	0.657259
6	1.838222	0.040849	0.703666	2.134058	0.046393	0.703651
7	1.481899	0.032931	0.736597	1.63323	0.035505	0.739156
8	1.423932	0.031643	0.76824	1.449508	0.031511	0.770667
9	1.176521	0.026145	0.794384	1.257793	0.027343	0.798011
10	1.035385	0.023009	0.817393	1.121126	0.024372	0.822383
11	0.827042	0.018379	0.835772	0.895784	0.019474	0.841856
12	0.78282	0.017396	0.853168	0.798299	0.017354	0.859211
13	0.74583	0.016574	0.869742	0.688927	0.014977	0.874187
14	0.621263	0.013806	0.883548	0.545678	0.011863	0.88605
15	0.570094	0.012669	0.896216	0.488394	0.010617	0.896667
16	0.535856	0.011908	0.908124	0.4839	0.01052	0.907187
17	0.442034	0.009823	0.917947	0.427472	0.009293	0.91648
18	0.362231	0.00805	0.925997	0.395744	0.008603	0.925083
19	0.329068	0.007313	0.933309	0.354817	0.007713	0.932796
20	0.263579	0.005857	0.939167	0.294884	0.006411	0.939207
21	0.24743	0.005498	0.944665	0.283054	0.006153	0.94536
22	0.20855	0.004634	0.9493	0.233435	0.005075	0.950435
23	0.171215	0.003805	0.953104	0.194967	0.004238	0.954673
24	0.150996	0.003355	0.95646	0.165314	0.003594	0.958267
25	0.142421	0.003165	0.959625	0.145916	0.003172	0.961439
26	0.125343	0.002785	0.96241	0.123881	0.002693	0.964132
27	0.081218	0.001805	0.964215	0.108341	0.002355	0.966487
28	0.080323	0.001785	0.966	0.103224	0.002244	0.968731
29	0.049351	0.001097	0.967097	0.080796	0.001756	0.970488
30	0.042876	0.000953	0.968049	0.07233	0.001572	0.97206
31	0.027549	0.000612	0.968662	0.057934	0.001259	0.97332
32	0.02515	0.000559	0.969221	0.047746	0.001038	0.974358
33	0.020588	0.000458	0.969678	0.041847	0.00091	0.975267
34	0.016873	0.000375	0.970053	0.025816	0.000561	0.975829
35	0.009929	0.000221	0.970274	0.018192	0.000395	0.976224
36	0.006312	0.00014	0.970414	0.013323	0.00029	0.976514
37	0.000994	0.000022	0.970436	0.009307	0.000202	0.976716
38	-0.004188	-9.3E-05	0.970343	0.00197	0.000043	0.976759
39	-0.007509	-0.00017	0.970176	0.001547	0.000034	0.976792
40	-0.009685	-0.00022	0.969961	-0.007564	-0.00016	0.976628
41	-0.011568	-0.00026	0.969704	-0.007923	-0.00017	0.976456
42	-0.014162	-0.00032	0.969389	-0.009259	-0.0002	0.976254
43	-0.016365	-0.00036	0.969025	-0.011052	-0.00024	0.976014
44	-0.018586	-0.00041	0.968612	-0.012475	-0.00027	0.975743
45	-0.023671	-0.00053	0.968086	-0.013806	-0.0003	0.975443
				-0.016201	-0.00035	0.975091

Table 2: Factor Loadings for Retained Factors, Male.

Cause of Death	Code	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Tuberculosis	A01	0.276925	0.842001	-0.04704	0.29045	0.000626
HIV/AIDS	A02	-0.13134	0.302538	-0.12195	0.75164	0.001368
Intestinal infections diseases	A04	-0.06676	-0.34343	-0.22586	-0.01551	0.490351
Selected Vaccine Preventable Diseases	A05	-0.0226	0.063317	0.288252	-0.62568	0.145707
Parasitic and Vector Diseases	A07	0.044542	-0.0506	0.309556	0.215519	-0.21285
Meningitis and encephalitis	A081	0.440008	0.350583	0.18266	0.402895	-0.08734
Hepatitis	A082	-0.19633	0.882822	-0.02793	0.092986	-0.04318
Other infectious diseases	A083	0.827646	0.14492	-0.23928	0.199217	0.004374
Respiratory infections	A09	0.505352	0.230635	-0.5256	0.067308	-0.44444
Neonatal conditions	A11	0.539418	-0.21719	-0.02578	0.578262	0.144643
Nutritional Deficiencies	A12	0.748062	-0.06903	0.243252	0.118618	0.393777
Esophagus cancer	B141	0.119752	-0.67088	-0.1058	0.047033	0.322081
Stomach cancer	B142	-0.16875	0.397599	-0.43572	0.489076	-0.15146
Liver cancer	B143	-0.08758	0.735879	-0.20498	0.330718	0.049291
Larynx, trachea, bronchus, and lung cancers	B144	0.75742	-0.24063	0.047624	0.32661	0.347022
Breast cancer	B145	0.122207	-0.30502	-0.36681	0.457042	0.133232
Prostate cancer	B147	0.771084	-0.21209	-0.04371	0.168737	0.02984
Colorectal cancer	B148	0.427304	-0.47947	-0.26022	0.474676	0.29629
Other malignant and benign neoplasms	B149	0.438383	-0.43942	-0.07018	-0.02613	0.631188
Diabetes Mellitus	B15	0.40187	0.076985	0.076286	0.026948	0.765022
Endocrine, nutritional, blood and immune disorders	B16	0.26671	-0.07199	0.110301	0.033514	0.839317
Mental and behavioral disorders, neurological conditions, sense organ diseases	B17	0.361425	-0.08686	0.553023	-0.4424	0.27132
Rheumatic heart disease	B181	-0.11683	-0.06261	0.103776	-0.68608	0.03486
Ischemic heart disease	B182	0.173812	0.101732	-0.42891	0.437549	0.214989

Cerebrovascular disease	B184	0.779189	0.244319	0.072082	-0.27571	0.195391
Other circulatory diseases	B185	0.391463	0.241722	-0.04287	0.400902	0.467933
Respiratory diseases	B19	0.511707	-0.23763	0.298033	-0.17724	0.22282
COPD	B191	0.742393	-0.02247	0.37002	-0.28522	0.099482
Cirrhosis of the liver	B201	0.167499	0.021745	0.250551	0.570641	0.356852
Alcoholic cirrhosis of the liver	B202	-0.193	0.77554	0.107252	-0.34486	-0.12557
Other digestive diseases	B202	0.822866	0.050802	0.190299	-0.20643	0.118235
Skin diseases- Genitourinary diseases- Musculoskeletal diseases	B21	0.616724	0.540568	0.09273	0.072634	0.289923
Congenital anomalies	B22	0.483144	0.286483	0.474689	-0.27959	0.265414
Unintentional injuries	C23	0.712563	-0.37407	0.440977	0.169706	0.092123
Transport injuries	C231	0.607395	0.145383	0.668265	0.124084	0.04546
Exposure to forces of nature	C232	0.218432	-0.02182	0.683053	-0.02451	-0.12821
Falls	C233	-0.01793	-0.18602	0.514517	-0.37861	-0.19139
Exposure to mechanical forces (animate or inanimate)	C234	0.747336	-0.05116	0.3569	-0.13386	0.091498
Accidental drowning and submersion	C235	0.205647	0.271218	0.645959	0.174889	0.073637
Exposure to smoke, fire and flames, contact with heat and hot substances	C236	0.861514	-0.14541	0.128226	0.283289	0.080961
Accidental poisoning by and exposure to noxious substances	C237	-0.24889	0.130114	0.444017	0.300742	0.250877
Self-inflicted injuires	C241	0.261634	0.022755	0.836381	-0.27507	0.067652
Interpersonal violence	C242	0.386393	0.428734	0.023219	0.572326	0.118595
War and civil conflict and legally sanctioned deaths	C243	0.11613	0.607554	0.368573	-0.14443	0.046087
Alcohol poisoning	C25	0.311949	0.088568	0.456873	-0.14824	-0.25048

Table 3: Factor Loadings for Retained Factors, Female.

Cause of Death	Code	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Tuberculosis	A01	0.174905	0.816074	-0.11499	-0.01536	0.064149
HIV/AIDS	A02	-0.25379	0.138168	0.138236	0.793506	-0.02315
STDS, excluding HIV	A03	0.424108	-0.063316	0.034465	0.376308	0.047414
Intestinal infections diseases	A04	-0.14191	-0.183404	0.624766	-0.04457	0.155498
Selected Vaccine Preventable Diseases	A05	0.216673	-0.050489	-0.22798	-0.69289	0.108932
Parasitic and Vector Diseases	A07	0.052581	-0.194165	-0.20591	0.478331	0.08313
Meningitis and encephalitis	A081	0.552731	0.096515	-0.15022	0.114536	0.145483
Hepatitis	A082	-0.02898	0.856091	-0.12421	-0.0577	0.315892
Other infectious diseases	A083	0.679968	0.383186	0.196852	-0.00898	-0.1585
Respiratory infections	A09	0.268976	0.523103	-0.01091	-0.15882	-0.44211
Maternal Conditions	A10	0.096256	0.602982	-0.20073	0.357232	0.015267
Neonatal conditions	A11	0.425911	-0.138076	0.238422	0.673918	-0.09035
Nutritional Deficiencies	A12	0.862827	-0.194451	0.12884	0.045602	0.043425
Esophagus cancer	B141	-0.35609	-0.192685	0.526587	0.13936	-0.19546
Stomach cancer	B142	-0.29877	0.724771	-0.06171	0.250783	-0.37232
Liver cancer	B143	0.137763	0.818669	-0.15206	-0.01768	0.048783
Larynx, trachea, bronchus, and lung cancers	B144	0.186788	-0.196789	0.751362	0.008786	0.26505
Breast cancer	B145	0.167559	-0.070502	0.736897	0.299212	-0.36368
Cervix and Corpus uteri cancer	B146	0.069044	0.188369	0.555151	0.406973	-0.35495
Colorectal cancer	B148	0.163763	-0.212351	0.739173	0.241399	-0.34813
Other malignant and benign neoplasms	B149	0.275156	-0.279316	0.670385	-0.2228	-0.31298
Diabetes Mellitus	B15	0.728685	0.140147	0.347877	0.025456	0.001078
Endocrine, nutritional, blood and immune disorders	B16	0.588494	-0.101902	0.520522	0.116136	0.153004

Mental and behavioral disorders, neurological conditions	B17	0.565237	-0.434226	-0.14531	-0.38761	0.311338
Rheumatic heart disease	B181	-0.09724	-0.038779	-0.08796	-0.64113	0.136975
Ischemic heart disease	B182	-0.17507	0.448269	0.453272	0.44241	-0.17064
Cerebrovascular disease	B184	0.79323	0.164326	0.075284	-0.35455	0.155268
Other circulatory diseases	B185	0.582553	0.249903	0.37683	0.21046	0.184751
Respiratory diseases	B19	0.811927	-0.144398	-0.13769	-0.1198	-0.02855
COPD	B191	0.33804	-0.001474	0.250688	-0.40216	0.56781
Cirrhosis of the liver	B201	0.42633	0.069397	0.060654	0.315304	0.345111
Alcoholic cirrhosis of the liver	B202	-0.21385	0.452537	-0.32137	-0.40972	0.460497
Other digestive diseases	B202	0.79288	-0.000058	0.14425	-0.42796	0.084177
Skin diseases- Genitourinary diseases- Musculoskeletal diseases	B21	0.825325	0.286795	-0.06139	-0.07115	0.122582
Congenital anomalies	B22	0.719545	0.050496	-0.08648	-0.12416	0.229493
Unintentional injuries	C23	0.79335	-0.390747	0.150717	0.140783	0.032573
Transport injuries	C231	0.771915	-0.0688	-0.18592	0.080881	0.417078
Exposure to forces of nature	C232	0.534377	-0.180456	-0.1634	0.0067	0.057554
Falls	C233	0.120199	-0.411255	-0.42676	-0.24685	0.257371
Exposure to mechanical forces (animate or inanimate)	C234	0.722045	0.036125	-0.00026	0.037871	0.221931
Accidental drowning and submersion	C235	0.170101	0.139386	-0.28483	0.02184	0.82229
Exposure to smoke, fire and flames, contact with heat	C236	0.726734	0.013467	0.320055	0.308453	-0.17843
Accidental poisoning by and exposure to noxious substances	C237	0.160022	0.141465	0.127982	0.039945	0.791691
Self-inflicted injuires	C241	0.337411	-0.15046	-0.31774	-0.19647	0.799136
Interpersonal violence	C242	0.700866	0.187497	0.072054	0.513439	0.33426
Alcohol poisoning	C25	0.244785	-0.111301	-0.23224	-0.09243	0.39391

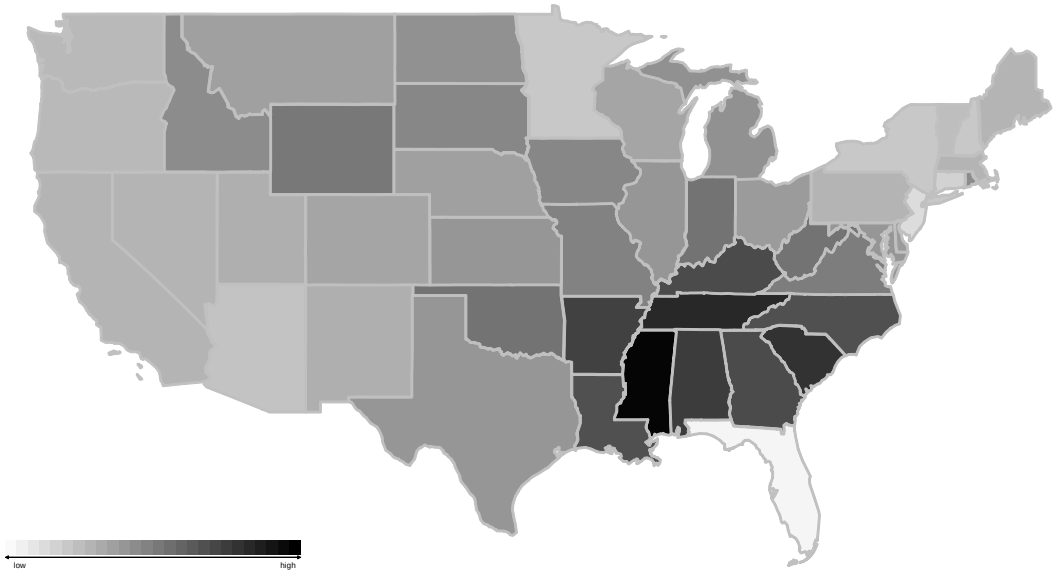
Table 4: OLS Model Parameters

Male				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.007255	3.25E-05	223.0383	< 0.001
factor1	0.000439	3.27E-05	13.42705	< 0.001
factor2	5.77E-05	3.30E-05	1.74528	0.088
factor3	0.000287	3.31E-05	8.657186	< 0.001
factor4	0.000185	3.29E-05	5.628346	< 0.001
factor5	0.0002	3.29E-05	6.081335	< 0.001
				R^2=0.88

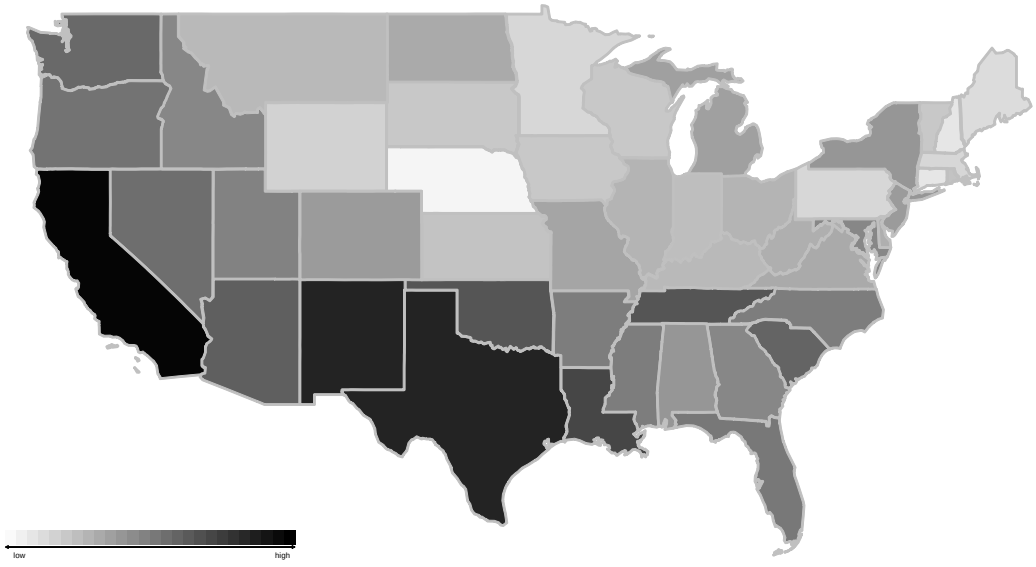
Female				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.007171	2.40E-05	298.2282	< 0.001
factor1	0.000498	2.43E-05	20.50196	< 0.001
factor2	-1.81E-05	2.43E-05	-0.74282	0.461
factor3	0.000294	2.44E-05	12.08947	< 0.001
factor4	0.000127	2.43E-05	5.20407	< 0.001
factor5	1.44E-05	2.44E-05	0.591364	0.557
				R^2=.9297

Figure 1. US State Maps of Male Factor Scores

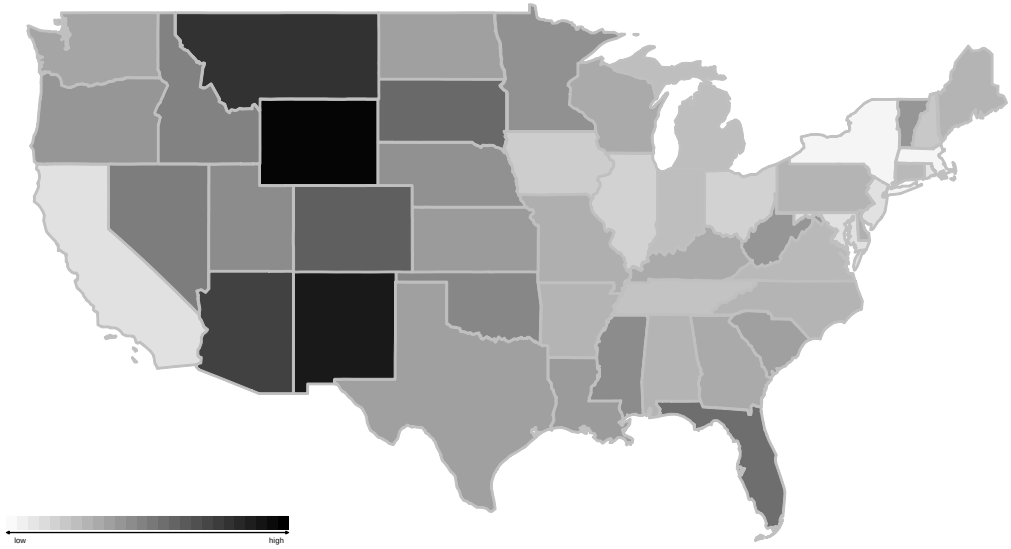
Factor 1:



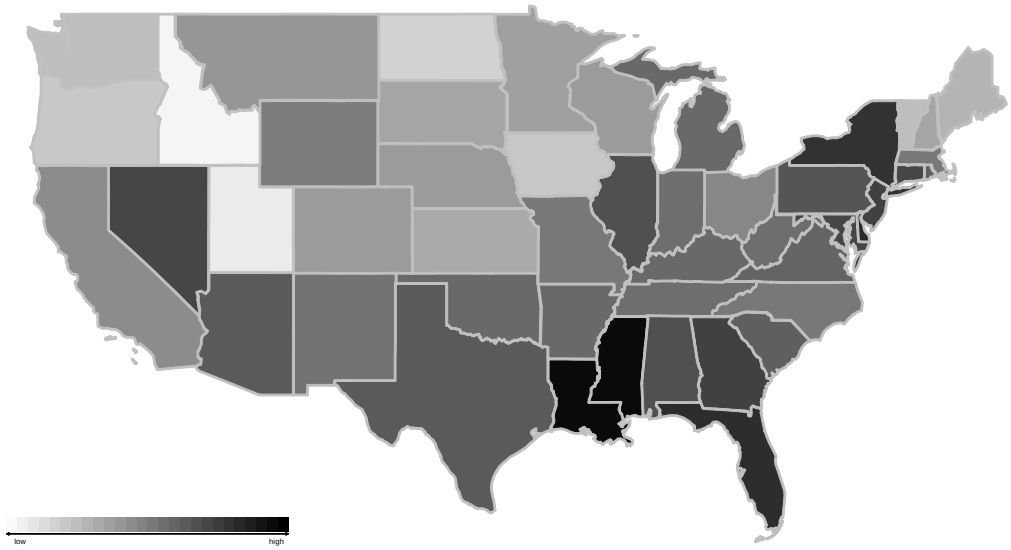
Factor 2:



Factor 3:



Factor 4:



Factor 5:

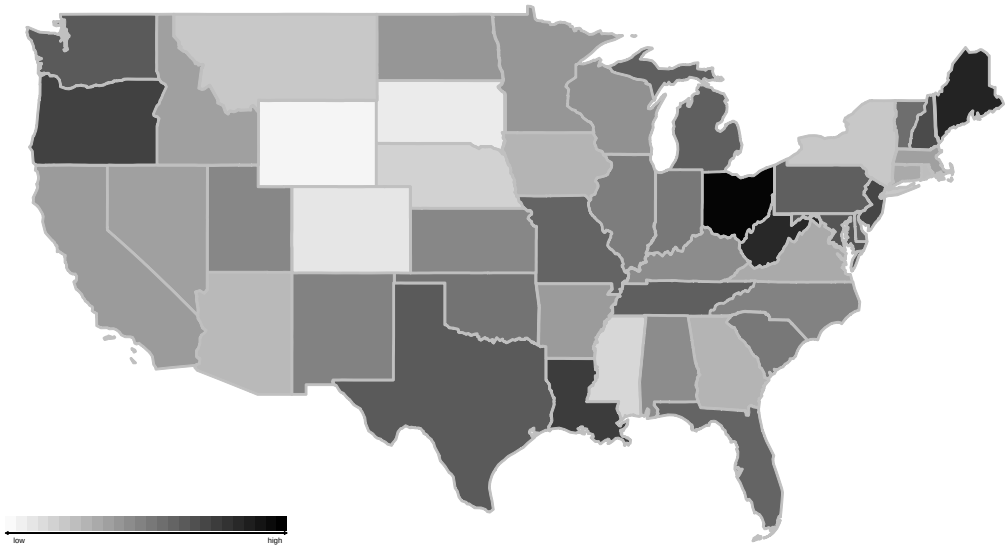
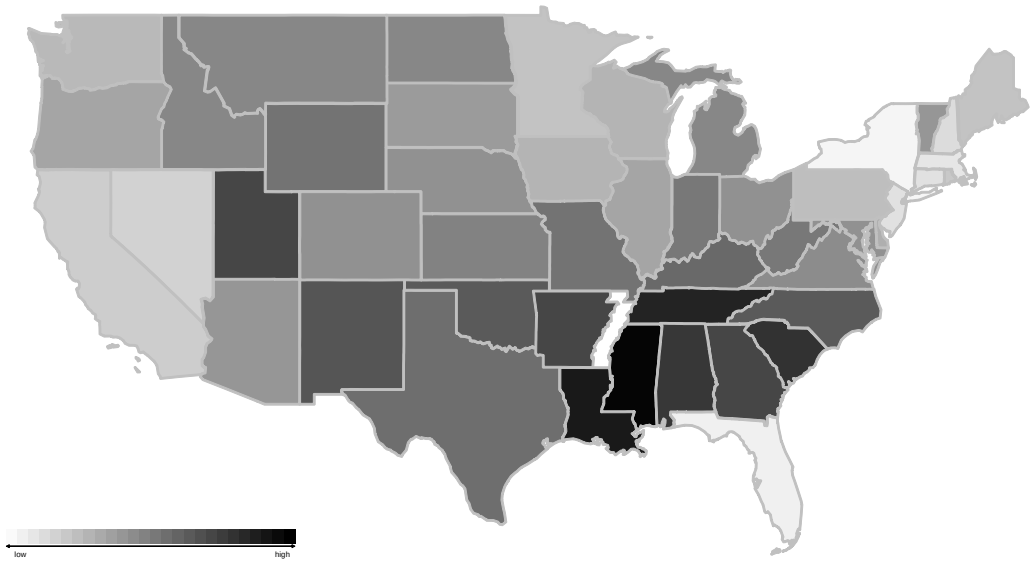
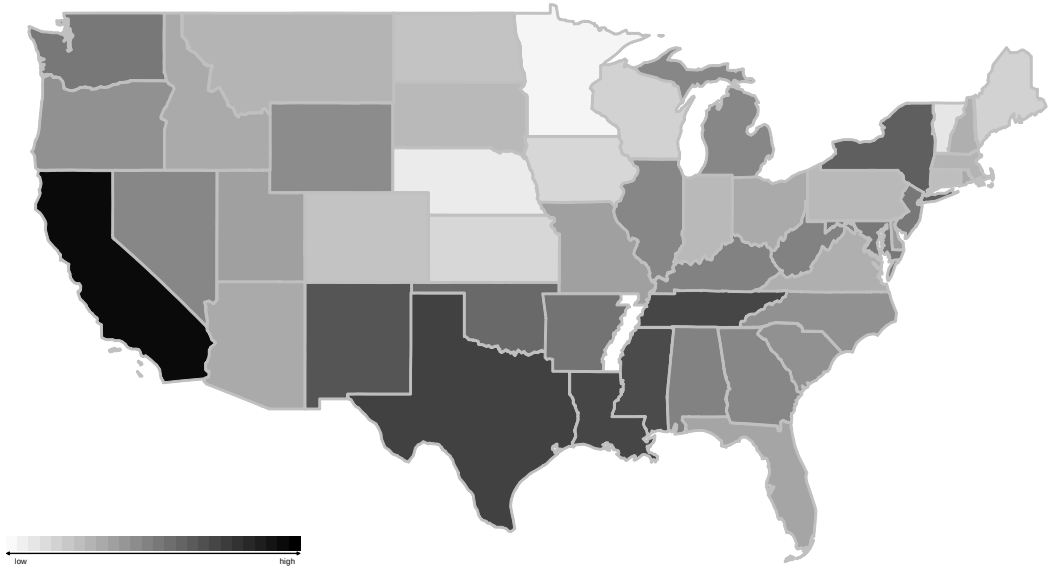


Figure 2. US State Maps of Female Factor Scores

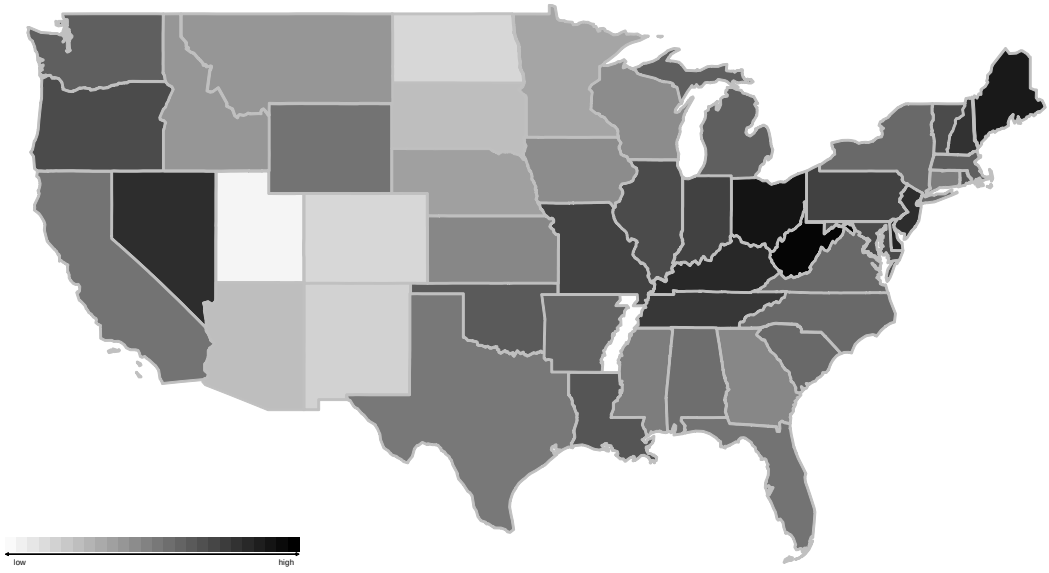
Factor 1:



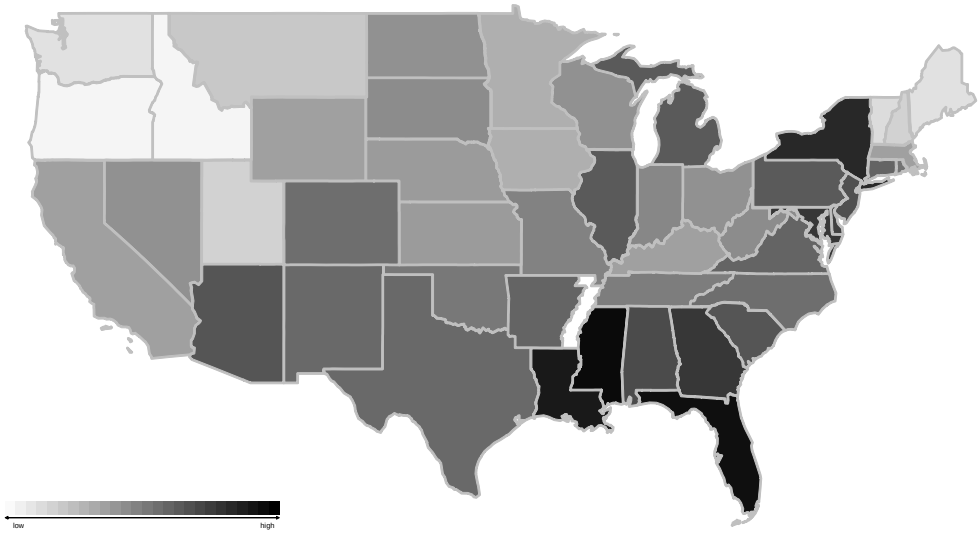
Factor 2:



Factor 3:



Factor 4:



Factor 5: