"Wealth, Health and Child Development: Evidence from Administrative Data on Swedish Lottery Players"

Online Appendix

David Cesarini
Erik Lindqvist
Robert Östling
Björn Wallace

October, 2015

Table of Contents

I. Appendix Figures

Panel A. Adult Sample

Panel B. Pre-lottery Children

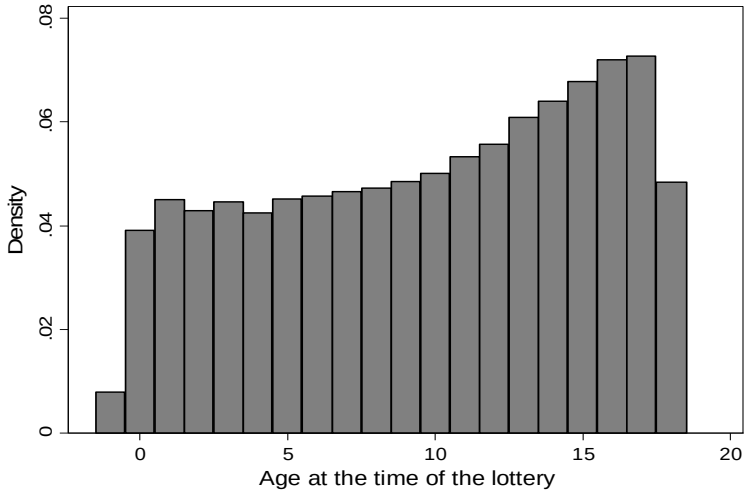


FIGURE AI
Age Distribution of Lottery Players and Pre-lottery Children

This figure shows the age distribution (measured at the time of the lottery event) for the adult lottery sample (Panel A) and pre-lottery children (Panel B). Because the sample of pre-lottery children includes some children that were conceived but not born at the time of the lottery, some children are recorded as being of age -1.

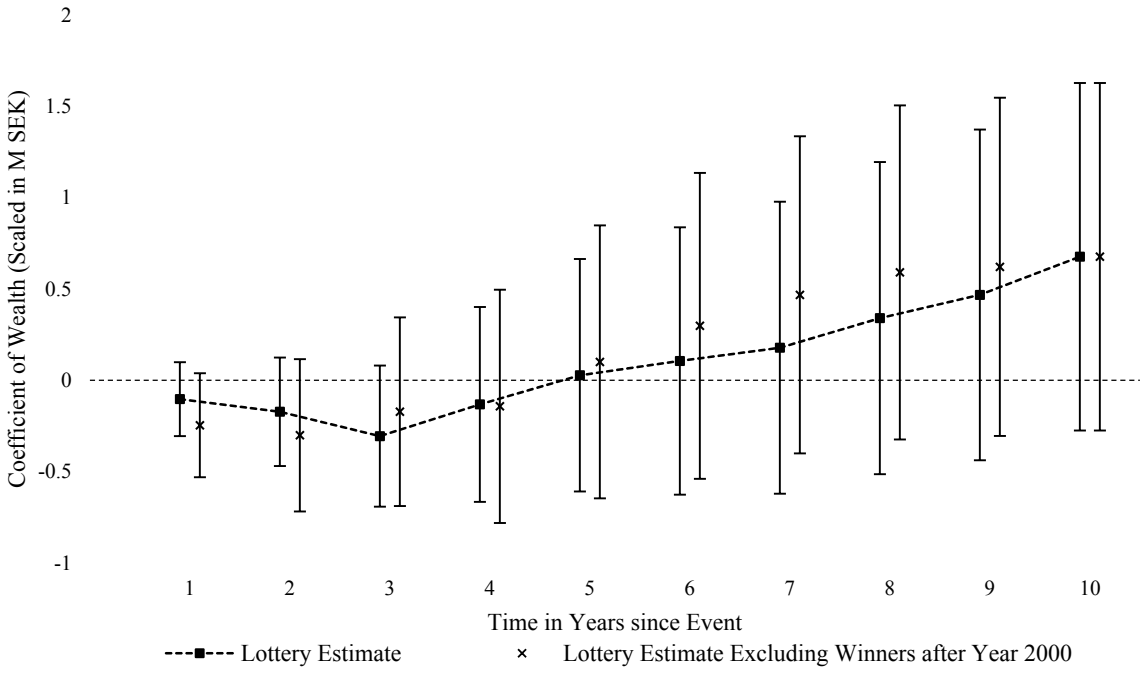


FIGURE AII
Wealth and Mortality (Fixed Sample Composition)

This figure contrasts our baseline lottery estimates of the effect of wealth on mortality to estimates from a sample restricted to players who win in 2000 or earlier. This restriction ensures that the sample composition is held constant in all analyses. Standard errors are clustered by individual, and the error bars give 95% confidence intervals of the coefficient.

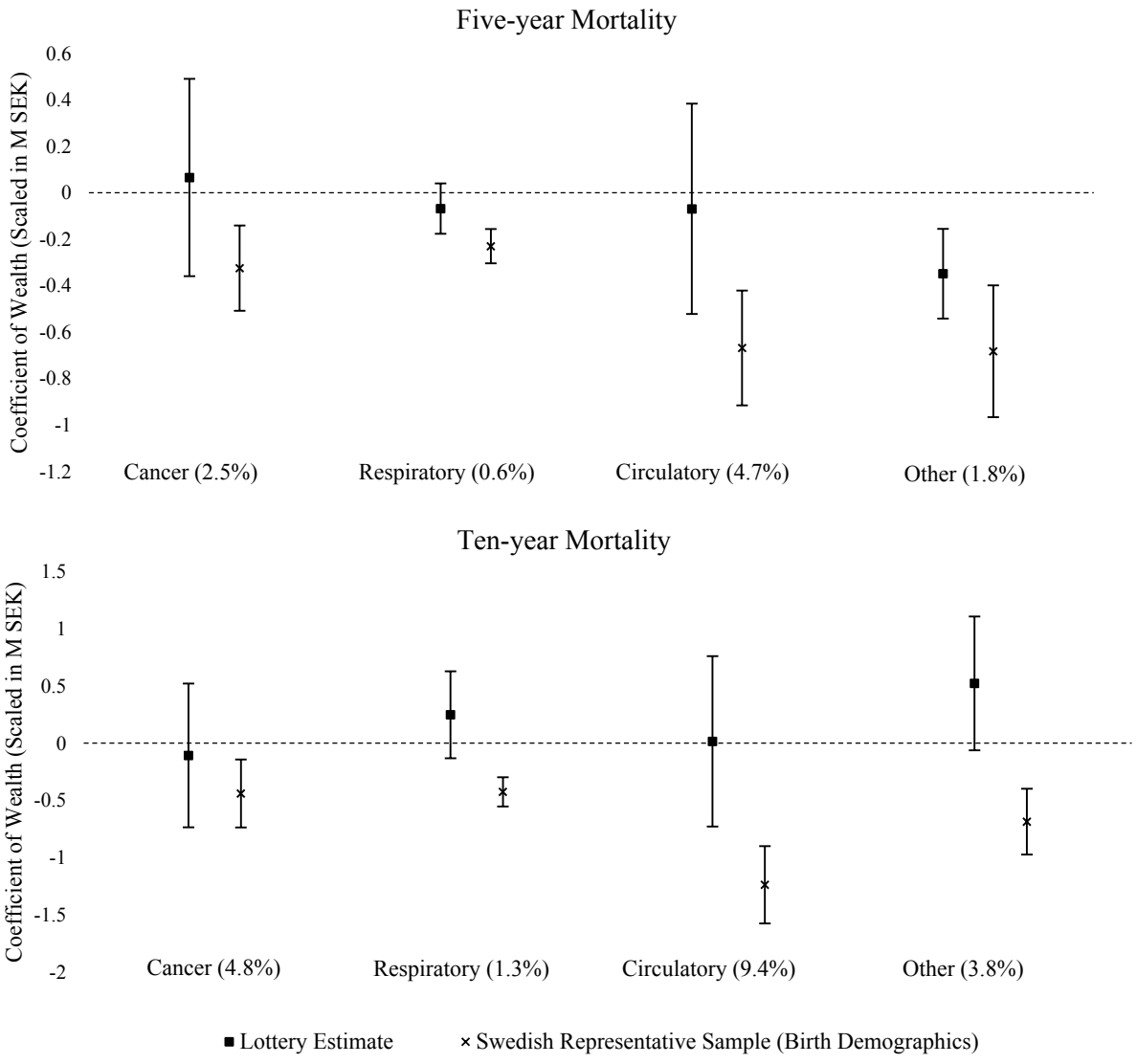


FIGURE AIII
Wealth and Mortality from Common Causes

This figure contrasts our lottery-based estimates of the effect of wealth on mortality due to common causes to gradients estimated in the Swedish representative sample. The representative sample has been reweighted to match the sex and age distribution of our sample of lottery winners. Gradients are estimated with controls for birth demographics. Standard errors are clustered by individual, and the error bars give 95% confidence intervals of the coefficient. The fraction of individuals in the lottery sample that die of each cause is shown within parentheses.

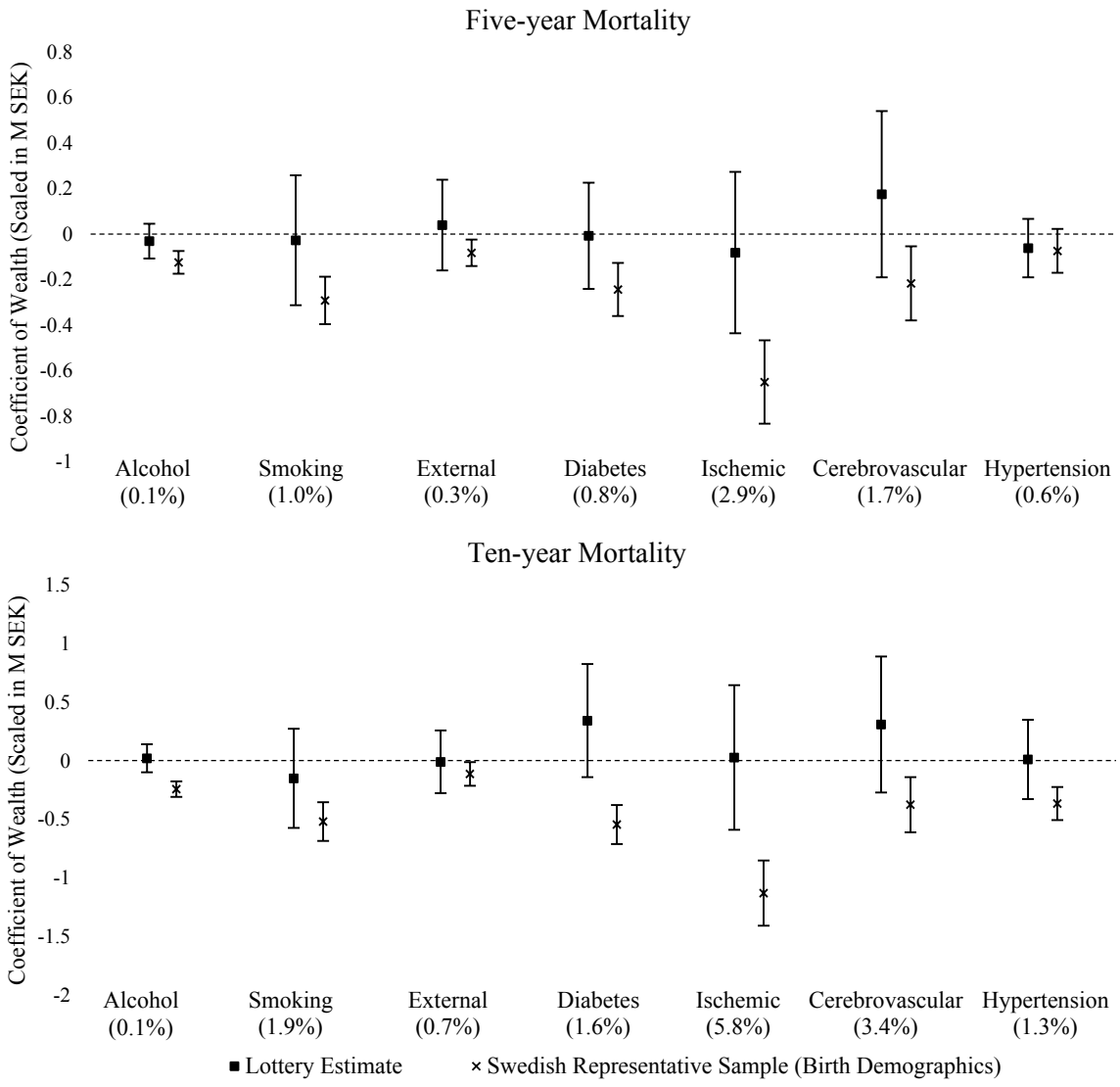


FIGURE AIV

Wealth and Mortality from Hypothesis-based Causes

This figure contrasts our lottery-based estimates of the effect of wealth on mortality due to hypothesis-based causes to gradients estimated in the Swedish representative sample. The representative sample has been reweighted to match the sex and age distribution of our sample of lottery winners. Gradients are estimated with controls for birth demographics. Standard errors are clustered by individual, and the error bars give 95% confidence intervals of the coefficient. The fraction of individuals in the lottery sample that die of each cause is shown within parentheses.

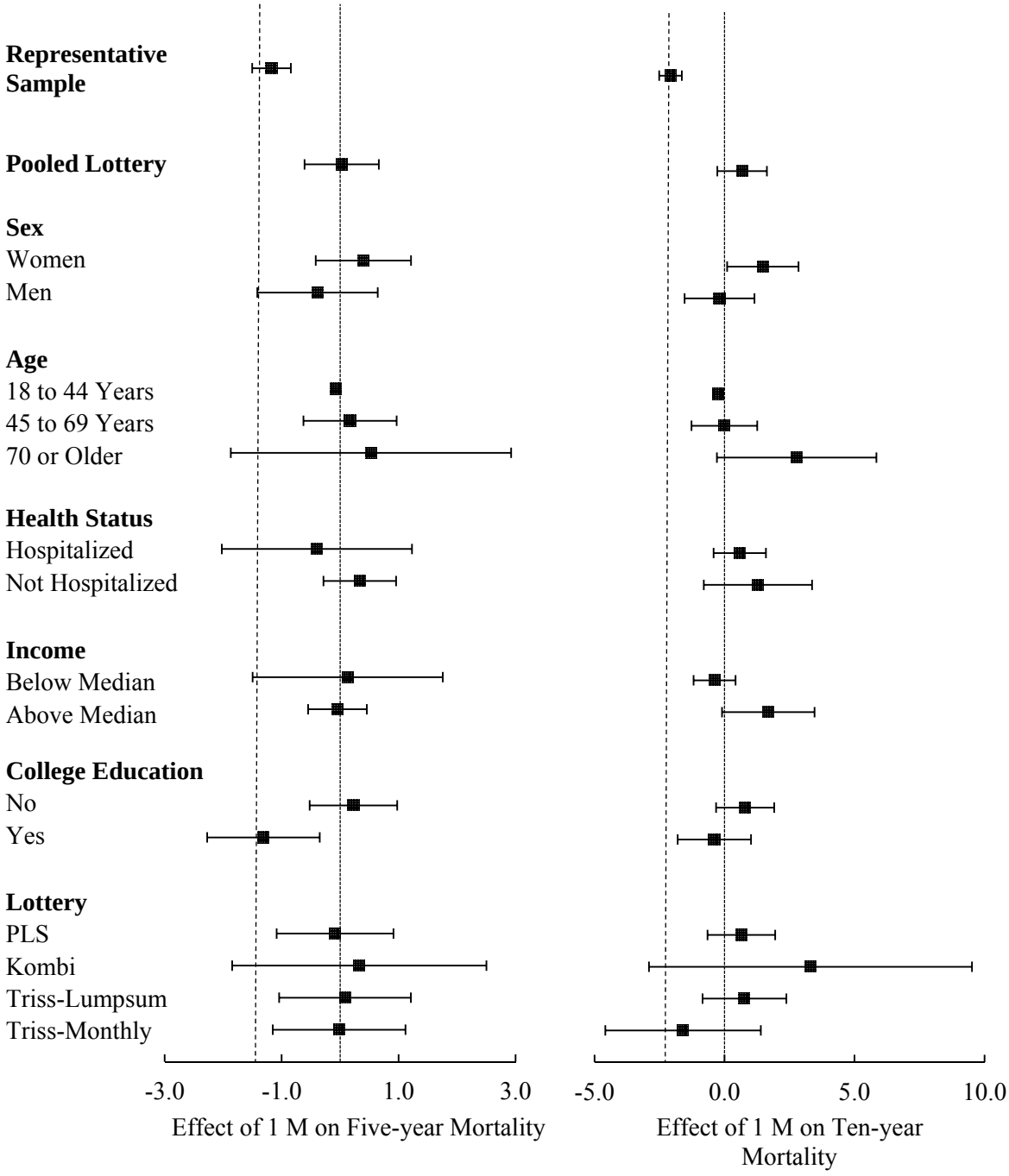


FIGURE AV
Heterogenous Effects of Wealth on Mortality

This figure shows the estimated effect of wealth on five- and 10-year mortality in different subsamples. As a benchmark, the uppermost estimate and the dashed line shows the Swedish mortality gradient. The estimates are also reported in Tables VIII and AX-AXII and are obtained from regressions in which prize amount and all control variables are interacted with indicator variables for different subsamples. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase in mortality per 1 million SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual, and the error bars give 95% confidence intervals of the coefficient.

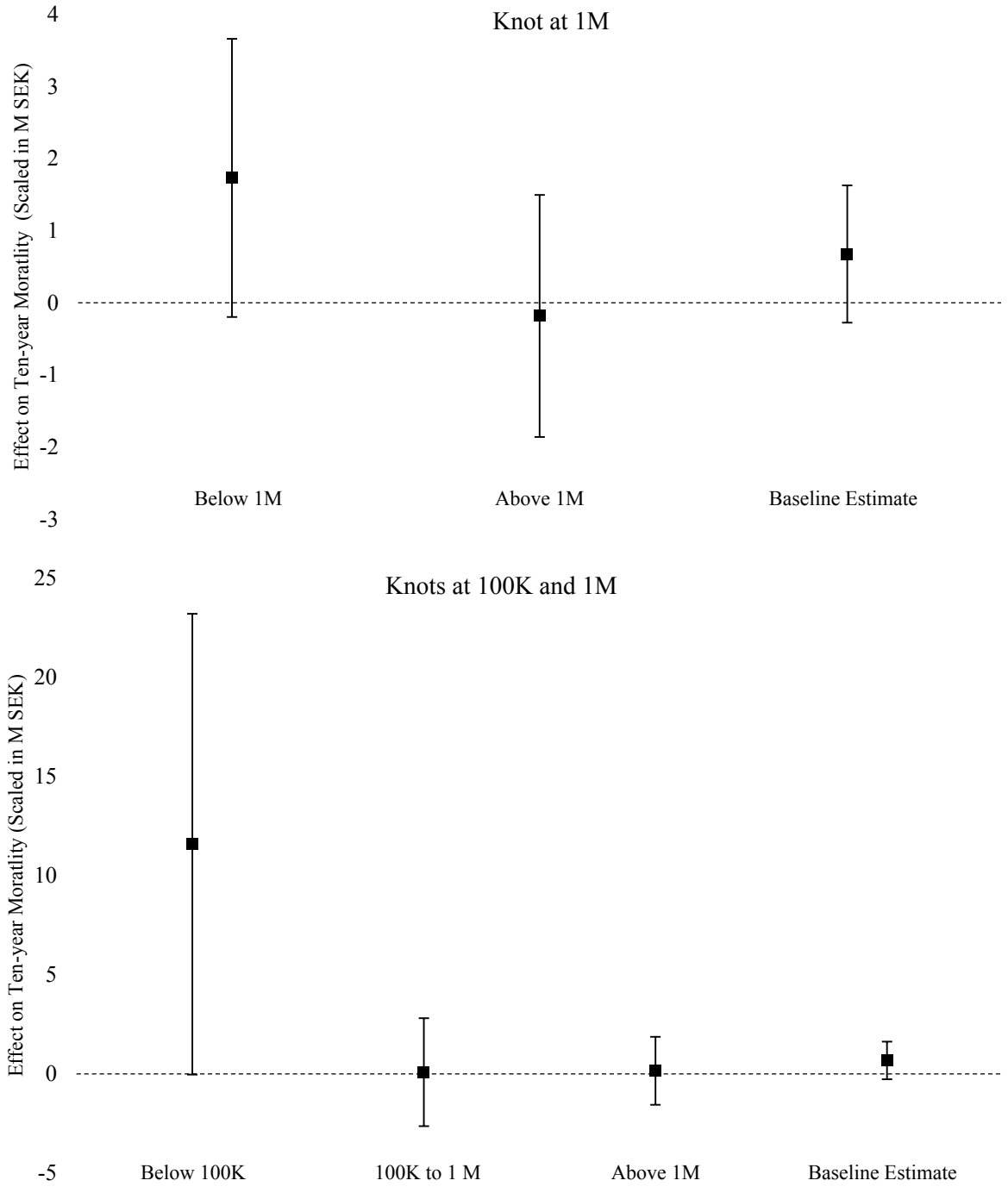


FIGURE AVI
Non-linear Effect of Wealth on Ten-year Mortality

This figure illustrates the estimates from spline regressions with a knot at 1 million (top panel) and knots at 100K and 1M SEK (bottom panel). These estimates are also reported in Table AXIII. The outcome variable is an indicator variable indicating whether the lottery player was dead 10 years after the lottery event. Wealth is scaled so that a coefficient of 1 denotes a 1 percentage unit increase in morality per 1M SEK won. All regressions include the full set of baseline controls. Standard errors are clustered by individual, and the error bars give 95% confidence intervals of the coefficient.

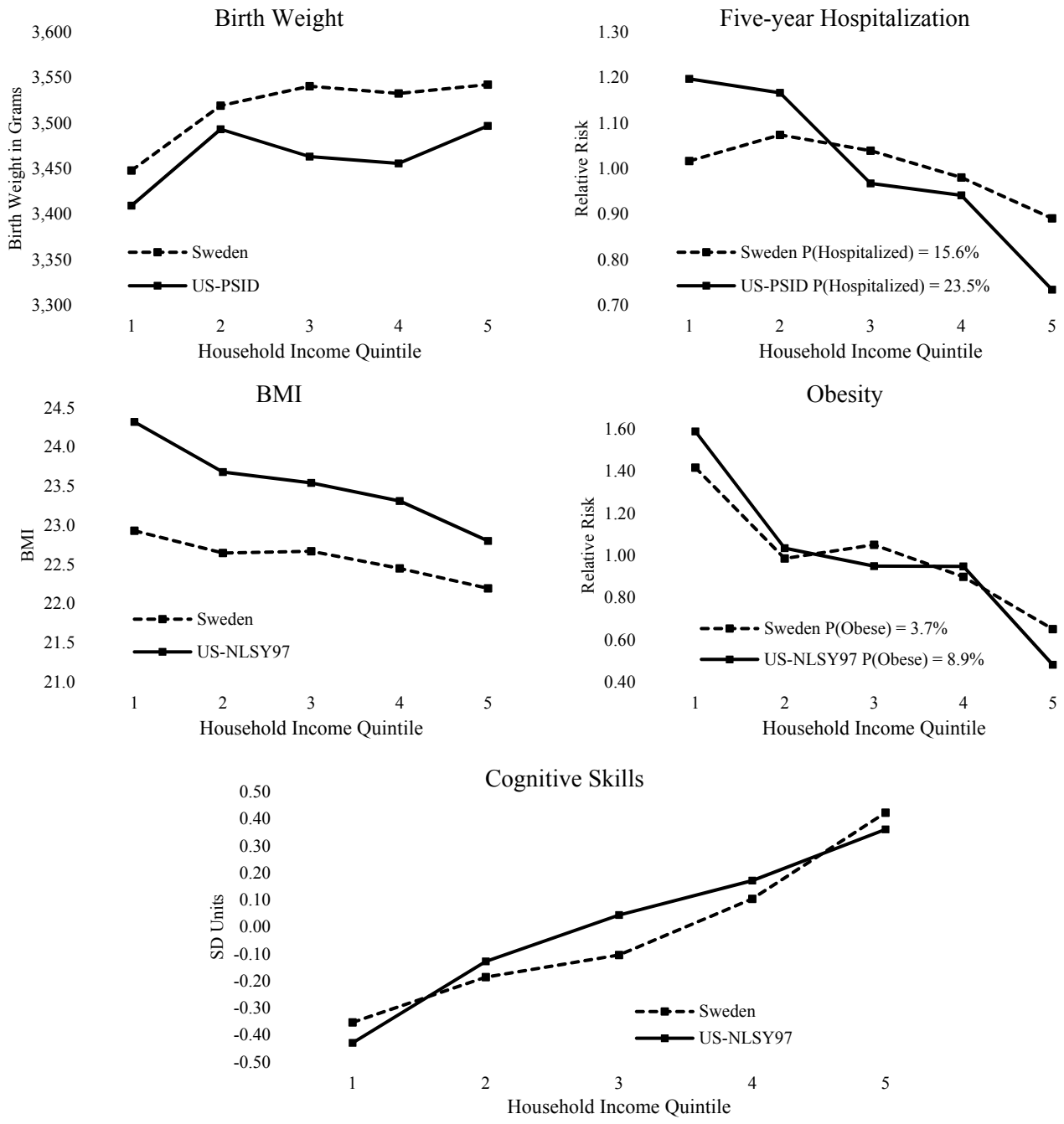


FIGURE AVII
Comparison to US Gradients

This figure compares the relationship between household income quintile and child outcomes in representative samples from Sweden and the US. For five-year hospitalizations and obesity, the figures shows the risk relative to the sample average. The US gradients are estimated for non-Hispanic Whites using the National Longitudinal Survey of Youth 1997 Cohort (NLSY97) and the Panel Study of Income Dynamics (PSID). The Swedish gradients have been constructed to maximize comparability with the US gradients, see OA Section IX.E. for further details. To avoid cluttering the figure, we do not report mean income by quintile in each analysis. For cognitive skills, the average income in the first quintile is 22K USD in the US and 23K USD in the Sweden, whereas the fifth quintile is 180K USD in the US and 121K USD in Sweden.

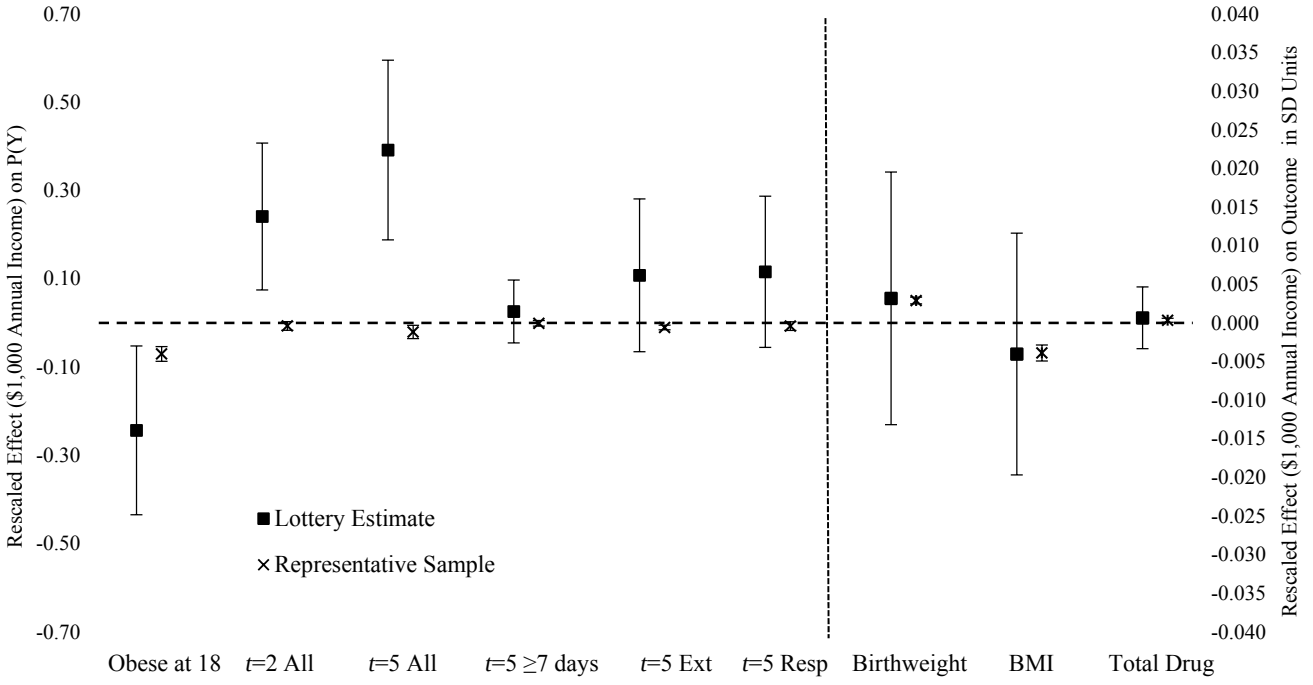


FIGURE AVIII
Household Income and Selected Child Health Outcomes

This figure compares our lottery-based estimates of the causal impact of wealth on selected health outcomes to the household-income gradients estimated in a large representative sample of children. For the binary variables to the left of the dashed line, we report estimates from a linear probability model in which income scaled so a coefficient of 0.1 corresponds to a 0.1 percentage point increase per $1,000 in permanent annual income. The coefficients to the right of the dashed line are expressed in units of standard deviations of the dependent variable. Gradients include controls for birth demographics and the lottery estimates additionally include demographic controls for the winning parent. Standard errors are clustered using an iterative algorithm that assigns siblings and half-siblings to the same cluster. Ext.: External Causes. Resp.: Respiratory Causes. BMI: Body Mass Index. Total Drug: sum of DDDs prescribed 2006-2010.

II Appendix Tables

TABLE AI

PRE-DETERMINED CHARACTERISTICS BELOW/ABOVE PRIZE AMOUNT

	Pooled Lottery Sample			PLS		Kombi		Triss-Lumpsum		Triss-Monthly	
	Average	Difference	*p*	Difference	*p*	Difference	*p*	Difference	*p*	Difference	*p*
Prize in M SEK	0.015	-0.125		-0.073		-1.220		-0.668		-2.264	
Birth Year	1935.9	0.136	[0.162]	0.093	[0.356]	-0.051	[0.910]	1.237	[0.019]	1.292	[0.260]
Female	51.0%	-0.004	[0.230]	-0.003	[0.361]	0.001	[0.947]	-0.015	[0.353]	-0.072	[0.060]
Nordic Born	97.2%	0.000	[0.668]	0.000	[0.982]	0.002	[0.731]	0.011	[0.189]	0.004	[0.848]
# Children	1.6	0.001	[0.947]	0.006	[0.473]	-0.021	[0.728]	-0.115	[0.009]	-0.145	[0.133]
College-Graduate	18.0%	0.002	[0.321]	0.002	[0.353]	0.010	[0.553]	0.005	[0.691]	-0.022	[0.495]
Married	57.8%	0.004	[0.198]	0.004	[0.157]	0.058	[0.012]	-0.021	[0.204]	-0.083	[0.031]
Retired	38.9%	-0.005	[0.069]	-0.004	[0.155]	0.001	[0.949]	-0.039	[0.004]	0.014	[0.619]
Labor Income/1000	127.5	0.655	[0.425]	0.494	[0.558]	-3.218	[0.633]	7.577	[0.111]	-2.098	[0.860]
Hospitalized	31.9%	0.002	[0.380]	0.003	[0.349]	-0.021	[0.349]	-0.002	[0.911]	0.051	[0.110]
Hosp. $\geq$ 7 days	15.8%	-0.002	[0.470]	-0.001	[0.513]	-0.020	[0.241]	0.002	[0.793]	0.011	[0.568]
Hosp. Cancer	3.9%	0.000	[0.809]	0.000	[0.917]	-0.006	[0.571]	0.005	[0.364]	0.012	[0.422]
Hosp. Respiratory	3.5%	0.000	[0.659]	0.000	[0.704]	-0.017	[0.129]	0.004	[0.484]	0.010	[0.353]
Hosp. Circulatory	10.2%	0.000	[0.800]	0.000	[0.955]	-0.002	[0.924]	0.002	[0.773]	0.046	[0.008]
N		439,234		387,813		46,486		4,250		685	

Notes. This table compares pre-determined characteristics for lottery players who win a prize below or above the median prize in each cell. The first column shows the unweighted average in the pooled lottery sample. For the pooled sample, and each of the lotteries separately, the column "Difference" shows the difference within each cell between the winners of prizes below or at the median compared to winners above the median. We test whether this difference is significantly different from zero by regressing the outcome variable on cell fixed effects and a binary indicator variable for whether the prize is above or below the median in the cell. The *p*-value for the indicator variable is reported in the table. Standard errors are clustered at the level of the individual.

TABLE AII

TESTING FOR ENDOGENOUS ATTRITION

	1 if Domiciled Abroad at…			1 if Missing from…		
	$t = 2$	$t = 5$	$t = 10$	Ninth Grade Register	Conscription Register	National Tests Register
Effect (M SEK)	0.034	0.162	0.355	-0.440	1.627	0.990
SE	(0.054)	(0.135)	(0.243)	(0.688)	(1.819)	(1.440)
p (analytical)	[0.525]	[0.228]	[0.144]	[0.522]	[0.371]	[0.492]
Proportion	0.43%	0.62%	0.85%	2.99%	26.57%	6.97%
N	415,215	378,099	296,904	76,436	49,617	28,628

Notes. This table reports the estimated effects of wealth on the likelihood of being domiciled outside Sweden on December 31 at $t = 2$, 5, and 10 (left panel) conditional on being alive, as well as the likelihood that pre-lottery children are missing in the registers used to generate the child development variables (right panel). Wealth is scaled so that a regression coefficient of 1.00 implies that 1M SEK increases the likelihood of being domiciled abroad or missing from registers by one percentage point. All regressions include the full set of baseline controls. Standard errors are clustered by individual (left panel) or using an iterative algorithm that assigns siblings and half-siblings to the same cluster (right panel).

TABLE AIII

THE EFFECT OF WEALTH ON FERTILITY

	Winners below Age 50		Winners below Age 50 with Pre-Lottery Children	
	All	Women	All	Women
Effect (M SEK)	2.415	-0.276	1.556	-0.506
SE	(1.333)	(1.731)	(1.514)	(1.469)
p (analytical)	[0.070]	[0.873]	[0.304]	[0.731]
p (resampling)	[0.052]			
Mean	0.44	0.40	0.16	0.13
N	125,045	59,491	57,649	29,670

Notes. This table reports the estimated effects of wealth on the number of children born after a lottery win (up until 2010). The left panel reports results for all winners aged below 50, whereas the right panel is restricted to the sub-sample of winners with pre-lottery children. Wealth is scaled so that a coefficient of 1.00 implies that 1M SEK is estimated to increase post-lottery fertility by 0.01 children. All regressions include the full set of baseline controls. Standard errors are clustered by individual. Resampling-based p values described in OA Section VII are reported whenever the analytical p-value is below 10%.

TABLE AIV

SIMILARITY OF ADULT LOTTERY WINNERS TO THE GENERAL POPULATION BY LOTTERY

	Pooled Lottery Sample		Individual Lottery Samples			
	Unweighted	Prize-Weighted	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Birth Year	1935.9	1943.3	1934.9	1942.2	1950.7	1954.5
Female	51.0%	49.2%	52.1%	41.9%	50.4%	47.0%
Nordic Born	97.2%	95.8%	97.1%	98.3%	94.1%	93.7%
# Children	1.63	1.69	1.59	1.92	1.78	1.68
College-Graduate	18.0%	18.3%	18.1%	17.0%	18.1%	20.9%
Married	57.8%	55.5%	58.1%	55.7%	50.5%	51.1%
Retired	38.9%	27.8%	38.9%	41.3%	21.0%	15.5%
Labor Income/1000	127.5	158.3	124.4	148.3	168.4	202.5
Hospitalized	31.9%	27.4%	31.8%	33.2%	26.3%	22.5%
Hospitalized ≥ 7 days	15.8%	12.0%	16.2%	13.2%	9.2%	6.4%
Hospitalized for Cancer	3.9%	3.5%	3.9%	4.2%	3.4%	3.8%
Hospitalized for Respiratory	3.5%	3.3%	3.4%	4.2%	3.2%	2.2%
Hospitalized for Circulatory	10.2%	7.5%	9.9%	13.7%	7.2%	5.7%
N	439,234	439,234	387,813	46,486	4,250	685

	Unweighted Random Population Samples		Random Population Samples: Sex- and Age Reweighted to Distribution of Above Lottery			
	1990	2000	1990	2000	2000	2000
Birth Year	1942.5	1951.5	1934.9	1942.2	1950.7	1954.5
Female	51.1%	51.6%	52.1%	41.9%	50.4%	47.0%
Nordic Born	94.3%	91.1%	94.9%	92.4%	91.3%	90.8%
# Children	1.48	1.51	1.77	1.90	1.66	1.65
College-Graduate	13.2%	23.1%	15.0%	22.7%	25.7%	28.9%
Married	50.7%	44.9%	54.5%	56.1%	49.6%	48.5%
Retired	21.7%	20.8%	34.7%	37.5%	19.4%	14.4%
Labor Income/1000	141.1	146.2	108.5	137.5	162.2	186.9
Hospitalized	28.7%	26.7%	34.6%	33.4%	26.3%	24.5%
Hospitalized ≥ 7 days	13.8%	10.3%	18.8%	14.9%	10.1%	8.6%
Hospitalized for Cancer	2.6%	2.8%	4.9%	5.5%	3.5%	3.4%
Hospitalized for Respiratory	3.0%	3.1%	4.5%	4.6%	3.1%	2.9%
Hospitalized for Circulatory	6.0%	7.2%	11.6%	13.9%	7.5%	6.8%

Notes. This table compares the baseline characteristics of lottery players to those of the general population. The first column in the top panel reports unweighted summary statistics for the adult estimation sample and the second column summary statistics weighting each observation by prize amount. The next four columns provide unweighted descriptive statistics by lottery. Each lottery sample is compared to representative samples of adults aged 18 and above. For PLS, we use a sample from 1990 and reweight it so that its age and sex distribution exactly matches that of the PLS sample. For the remaining three lotteries, we proceed analogously using a sample from 2000. Baseline characteristics are measured the year before the lottery event.

TABLE AV

REPRESENTATIVENESS OF PARENTS IN POOLED LOTTERY SAMPLE

Panel A: Parents of Pre-Lottery Children

	Pooled Lottery Sample		Sex- and Age-Reweighted Representative Sample	
	Unweighted	Prize-Weighted	Weighted to Match (1)	Weighted to Match (2)
	(1)	(2)	(3)	(4)
Birth Year	1951.5	1958.9	1951.5	1958.9
Female	48.4%	45.3%	48.4%	45.3%
Nordic Born	95.7%	92.4%	92.6%	89.9%
# Children	2.18	2.15	1.73	1.65
College-Graduate	31.6%	27.6%	26.1%	31.6%
Married	69.7%	58.3%	56.7%	49.5%
Retired	0.4%	0.1%	0.4%	0.1%
Labor Income/1000	228.9	250.2	201.2	225.5
Hospitalized	19.6%	18.8%	22.5%	19.1%
Hospitalized ≥ 7 days	5.5%	4.8%	7.7%	5.6%
Hospitalized for Cancer	1.5%	1.0%	2.0%	1.6%
Hospitalized for Respiratory	1.7%	2.2%	2.1%	1.8%
Hospitalized for Circulatory	1.8%	1.7%	2.4%	2.2%
N	68,584	68,584	71,669	69,914

Panel B: Parents of Post-Lottery Children

	Pooled Lottery Sample		Sex- and Age-Reweighted Representative Sample	
	Unweighted	Prize-Weighted	Weighted to Match (5)	Weighted to Match (6)
	(5)	(6)	(7)	(8)
Birth Year	1968.4	1969.9	1968.4	1969.9
Female	45.8%	45.8%	45.8%	45.8%
Nordic Born	96.7%	89.6%	93.6%	90.6%
# Children	0.30	0.47	0.53	0.68
College-Graduate	16.7%	23.7%	11.5%	22.6%
Married	12.2%	14.9%	16.5%	20.4%
Retired	0.0%	0.0%	0.0%	0.0%
Labor Income/1000	125.0	165.6	106.1	142.2
Hospitalized	16.0%	14.4%	18.8%	17.4%
Hospitalized ≥ 7 days	3.0%	3.5%	4.4%	3.7%
Hospitalized for Cancer	0.4%	0.5%	0.6%	0.6%
Hospitalized for Respiratory	2.4%	1.6%	2.5%	2.2%
Hospitalized for Circulatory	0.5%	0.2%	0.5%	0.6%
N	34,187	34,187	61,213	56,603

Notes. This table compares the baseline characteristics of lottery players with pre- and post-lottery children to those of the general population. Columns (1) and (5) reports unweighted summary statistics for the pooled lottery sample and columns (2) and (6) summary statistics with observations weighted by prize amount. Columns (3) and (4) shows descriptive statistics for a representative sample of adults aged 18 and above who have been reweighted to match the age and sex distribution of (1) and (2), respectively, and columns (7) and (8) the representative sample re-weighted to match columns (5) and (6). PLS winners have been matched to a representative sample from 1990 and the other lotteries to one from 2000. The number of observations irefers to the total number of lottery prizes in columns (1)-(2) and (5)-(6) and to individuals in columns (3)-(4) and (7)-(8). Baseline characteristics are measured the year before the lottery event.

TABLE AVI

REPRESENTATIVENESS OF PLAYERS' CHILDREN

Panel A: Pre-Lottery Children

	Pooled Lottery Sample	Individual Lottery Samples				Weighted Representative Sample
		PLS	Kombi	Triss-Lumpsum	Triss-Monthly	
Demographics						
Birth Year	1983.5	1982.5	1993.0	1992.8	1995.6	1983.5
Female	49.4%	49.1%	53.4%	47.7%	51.9%	49.4%
Child Health						
BMI	22.4	22.4	23.1	23.1	23.5	22.5
Obese (BMI > 30)	2.9%	2.8%	4.7%	6.9%	5.6%	3.5%
Hospitalized within 2 years	10.9%	11.0%	9.5%	10.8%	6.5%	11.5%
Hospitalized within 5 years	18.0%	18.1%	16.1%	18.3%	15.2%	19.2%
Hospitalized within 5 Years ($\geq$ 7 days)	2.8%	2.8%	2.4%	2.8%	0.6%	3.0%
Hospitalized within 5 Years (Respiratory)	4.2%	4.2%	3.0%	4.8%	4.8%	4.4%
Hospitalized within 5 Years (External)	5.2%	5.2%	5.3%	4.7%	6.1%	5.7%
Total Drug Consumption (DDDs)	468.3	478.7	332.5	375.2	234.9	516.2
Child Development						
GPA	0.271	0.283	0.069	-0.216	0.098	0.001
Cognitive Skills	0.169	0.174	0.023	-0.127	0.228	-0.009
Noncognitive Skills	0.143	0.146	0.028	-0.091	0.112	-0.016
Test Scores Swedish	0.163	0.175	0.081	-0.086	-0.028	-0.018
Test Scores English	0.105	0.114	0.034	-0.037	0.001	-0.019
Test Scores Mathematics	0.204	0.228	-0.012	-0.118	0.127	-0.007

Panel B: Post-Lottery Children

	Pooled Lottery Sample	Individual Lottery Samples				Weighted Representative Sample
		PLS	Kombi	Triss-Lumpsum	Triss-Monthly	
Demographics						
Birth Year	2001.5	2001.4	2007.3	2004.9	2007.2	2001.5
Female	48.8%	48.8%	47.8%	51.4%	45.2%	48.8%
Infant Health						
Birth Weight (grams)	3546.8	3546.9	3522.0	3567.1	3472.1	3535.1
Low Birth Weight (< 2500 grams)	4.1%	4.1%	5.0%	3.1%	6.3%	4.1%
Premature Birth (< 37 weeks)	6.3%	6.3%	8.0%	6.6%	3.1%	5.8%

Notes. This table compares demographic characteristics and outcome variables for the pre- and post-lottery children to the children of a representative sample which has been reweighted to match the birth year and sex distributions of the pre-lottery children (Panel A) and post-lottery children (Panel B). Note that the variables are sometimes (always, in the case of post-lottery children) measured after the lottery event and hence the measured outcomes may be endogenous to the outcome of the lottery. BMI: Body Mass Index. DDD: Defined Daily Dose.

TABLE AVII

AGGREGATION OF ICD AND ATC CLASSIFICATIONS INTO COMMON AND HYPOTHESIS-BASED SOURCES OF ILL HEALTH

	Hospitalization and Mortality			Drug Prescription
	ICD-8	ICD-9	ICD-10	ATC
Common Causes				
Circulatory	390-458	390-459	I00-I99	C
Cancer	140-239	140-239	C00-D48	L01
Respiratory	460-519	460-519	J00-J99	R
Other	All other except 630-679	All other except 630-679	All other except O	All other except G03
Hypothesis-Based Causes				
Diabetes	250	250	E10-E14	A10
Ischemic Heart Disease	410-414	410-414	J	
Hypertension	401-404	401-405	I10-I15	
Heart Disease				C01,C02,C03,C04,C07,C08,C09,C10
Cerebrovascular	430-438	430-438	I60-I69	B01
Tobacco	140-151, 155, 157, 160-162, 180, 188, 189.0, 189.1, 197.8, 205.0	140-151, 155, 157, 160-162, 180, 188, 189.0, 189.1, 205.0	C0-C16, C22, C25, C30-C34, C53, C64-C65, C67, C92.0, C92.4, C92.5	
Alcohol	291, 303, 571.0, 980	291, 303, 305.0, 357.5, 425.5, 535.3, 571.0-571.3, 980	E24.4, F10, G31.2, G62.1, G72.1, I42.6, K29.2, K70, K85.2, K86.0, T51, X45, X65, Y15	
External Causes	E800-999	E800-999	S00-T98	
Mental Health				N05, N06A

Notes. This table describes the aggregation of the World Health Organization's International Classification of Diseases (ICD) diagnoses and ATC: Anatomical Therapeutic Chemical (ATC) classification codes into common and hypotheses-based causes.

TABLE AVIII

WEALTH AND MORTALITY RISK

Panel A: Wealth and Two-year Mortality

	Lottery Sample	Swedish 2000 Sample	Swedish 2000 Sample	HRS Sample	Lottery Sample Probit	Lottery Sample Before 2000
Effect/Gradient (M SEK)	-0.173	-0.646	-0.439	-0.377	-0.229	-0.302
SE	(0.152)	(0.113)	(0.113)	(0.062)	(0.292)	(0.213)
p (analytical)	[0.255]	[<0.001]	[<0.001]	[<0.001]	[0.433]	[0.156]
Controls	Baseline	Birth	Baseline	Birth	Baseline	Baseline
N	431,064	39,019	39,019	19,641	431,064	367,863
Proportion Dead	3.68%	3.87%	3.87%	3.90%	3.68%	3.54%

Panel B: Wealth and Five-year Mortality

	Lottery Sample	Swedish 2000 Sample	Swedish 2000 Sample	HRS Sample	Lottery Sample Probit	Lottery Sample Before 2000
Effect/Gradient (M SEK)	0.027	-1.572	-1.135	-1.051	0.131	0.100
SE	(0.325)	(0.171)	(0.171)	(0.092)	(0.393)	(0.381)
p (analytical)	[0.934]	[<0.001]	[<0.001]	[<0.001]	[0.739]	[0.793]
Controls	Baseline	Birth	Baseline	Birth	Baseline	Baseline
N	418,002	39,019	39,019	19,641	418,002	367,863
Proportion Dead	9.55%	10.07%	10.07%	10.53%	9.55%	9.21%

Panel C: Wealth and 10-year Mortality

	Lottery Sample	Swedish 2000 Sample	Swedish 2000 Sample	HRS Sample	Lottery Sample Probit	Lottery Sample Before 2000
Effect/Gradient (M SEK)	0.676	-2.787	-2.062	-1.822	0.397	0.676
SE	(0.485)	(0.225)	(0.225)	(0.133)	(0.461)	(0.485)
p (analytical)	[0.164]	[<0.001]	[<0.001]	[<0.001]	[0.389]	[0.164]
Controls	Baseline	Birth	Baseline	Birth	Baseline	Baseline
N	367,863	39,019	39,019	19,641	367,863	367,863
Proportion Dead	19.29%	20.92%	20.92%	22.61%	19.29%	19.29%

Notes. This table reports the baseline OLS estimates of the effect of wealth on mortality along with the gradient between wealth and mortality in Swedish and US population samples reweighted to match the age and sex distribution of the lottery winners. The last two columns report the effect of lottery wealth estimated using a Probit model and the OLS estimates when the sample is restricted to those who won in 2000 or earlier. The table reports average marginal effects scaled so that a coefficient of 1.00 denotes a 1 percentage point increase in mortality per 1M SEK. Standard errors are clustered by individual.

TABLE AIX

WEALTH AND CAUSE-SPECIFIC MORTALITY RISK

	Common Causes				Hypothesis-Based Causes						
	Cancer	Circ.	Resp.	Other	Alcohol	Cerebro.	Diabetes	External	Hyper.	Ischemic	Smoking
Panel A: Five years after win											
OLS Effect (M SEK)	0.066	-0.069	-0.068	0.099	-0.031	0.175	-0.008	0.039	-0.062	-0.081	-0.027
SE	(0.217)	(0.231)	(0.055)	(0.148)	(0.039)	(0.186)	(0.119)	(0.102)	(0.065)	(0.181)	(0.146)
p	[0.762]	[0.765]	[0.220]	[0.504]	[0.426]	[0.347]	[0.949]	[0.699]	[0.347]	[0.653]	[0.852]
Proportion	2.5%	4.7%	0.6%	1.8%	0.1%	1.7%	0.8%	0.3%	0.6%	2.9%	1.0%
Gradient (M SEK)	-0.273	-0.462	-0.178	-0.222	-0.090	-0.106	-0.176	-0.062	-0.029	-0.527	-0.270
SE	(0.095)	(0.129)	(0.036)	(0.100)	(0.022)	(0.083)	(0.060)	(0.030)	(0.050)	(0.096)	(0.054)
Proportion	2.5%	4.5%	0.6%	2.4%	0.2%	1.9%	0.9%	0.4%	0.7%	2.8%	1.0%
Heterogeneity *p*	[0.153]	[0.138]	[0.096]	[0.071]	[0.193]	[0.168]	[0.207]	[0.340]	[0.693]	[0.029]	[0.118]
Panel B: 10 years after win											
OLS Effect (M SEK)	-0.107	0.015	0.247	0.521	0.020	0.307	0.340	-0.011	0.009	0.026	-0.152
SE	(0.320)	(0.379)	(0.194)	(0.298)	(0.061)	(0.296)	(0.246)	(0.136)	(0.172)	(0.314)	(0.216)
p	[0.737]	[0.968]	[0.202]	[0.080]	[0.745]	[0.299]	[0.166]	[0.936]	[0.956]	[0.935]	[0.482]
Proportion	4.8%	9.4%	1.3%	3.8%	0.1%	0.1%	1.6%	0.7%	1.3%	5.8%	1.9%
Gradient	-0.327	-0.899	-0.324	-0.511	-0.182	-0.202	-0.416	-0.095	-0.325	-0.911	-0.469
SE	(0.154)	(0.175)	(0.066)	(0.148)	(0.030)	(0.122)	(0.085)	(0.054)	(0.073)	(0.143)	(0.084)
Proportion	5.2%	9.1%	1.3%	5.4%	0.4%	3.6%	2.0%	0.8%	1.8%	5.7%	2.2%
Heterogeneity *p*	[0.535]	[0.028]	[0.005]	[0.002]	[0.003]	[0.112]	[0.004]	[0.564]	[0.074]	[0.007]	[0.170]

Notes. This table reports OLS estimates of the effect of wealth on cause-specific mortality, as well as the gradient for a representative sample reweighted to match the age and sex distribution of the lottery sample. Wealth is scaled so that a coefficient of 1.00 denotes a 1 percentage point increase in the likelihood of dying per 1M SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. The heterogeneity *p*-value is from a two-sided *t*-test of the null hypothesis that the gradient and causal parameter are equal. Circ.: circulatory disease. Resp.: respiratory disease. Cerebro.: cerebrovascular disease. Hyper.: hypertension.

TABLE AX

THE EFFECT OF WEALTH ON MORTALITY RISK BY AGE

Panel A: Wealth and Two-year Mortality (N = 431,064)

	Age 18-44	Age 45-69	Age 70+
Effect (M SEK)	-0.050	-0.134	-0.379
SE	(0.030)	(0.168)	(0.628)
p (analytical)	[0.093]	[0.423]	[0.547]
p (resampling)	[0.274]		
Heterogeneity p		[0.773]	
Proportion Dead	0.14%	1.33%	10.12%

Panel B: Wealth and Five-year Mortality (N = 418,002)

	Age 18-44	Age 45-69	Age 70+
Effect (M SEK)	-0.072	0.170	0.528
SE	(0.037)	(0.407)	(1.223)
p (analytical)	[0.055]	[0.676]	[0.666]
p (resampling)	[0.545]		
Heterogeneity p		[0.745]	
Proportion Dead	0.37%	3.70%	25.90%

Panel C: Wealth and 10-year Mortality (N = 367,863)

	Age 18-44	Age 45-69	Age 70+
Effect (M SEK)	-0.246	-0.006	2.775
SE	(0.101)	(0.645)	(1.564)
p (analytical)	[0.015]	[0.992]	[0.076]
p (resampling)	[0.428]		[0.092]
Heterogeneity p		[0.147]	
Proportion Dead	0.93%	9.13%	51.16%

Notes. This table reports estimates of the effect of wealth on mortality from regressions in which lottery prize and all control variables are interacted with indicator variables for the different age groups. Wealth is scaled so that a coefficient of 1.00 denotes a 1 percentage point increase in mortality per 1M SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. Resampling-based p values described in OA Section VII are reported whenever the analytical p-value is below 10%. The heterogeneity p-value is from an F-test of equal effects in the different age groups.

TABLE AXI

THE EFFECT OF WEALTH ON MORTALITY RISK BY SEX, COLLEGE, HEALTH, AND INCOME

Panel A: Wealth and Two-year Mortality (N = 431,064)

	Sex		College-Graduate		Hospitalized		Income > Median	
	Female	Male	Yes	No	No	Yes	Below	Above
Effect (M SEK)	-0.183	-0.245	0.035	-0.239	-0.161	0.068	-0.350	-0.066
SE	(0.174)	(0.256)	(0.263)	(0.177)	(0.101)	(0.496)	(0.341)	(0.092)
p (analytical)	[0.293]	[0.338]	[0.894]	[0.177]	[0.111]	[0.891]	[0.306]	[0.472]
Heterogeneity p	[0.841]		[0.387]		[0.651]		[0.422]	

Panel B: Wealth and Five-year Mortality (N = 418,002)

	Sex		College-Graduate		Hospitalized		Income > Median	
	Female	Male	Yes	No	No	Yes	Below	Above
Effect (M SEK)	0.398	-0.386	-1.311	0.228	0.337	-0.397	0.130	-0.046
SE	(0.416)	(0.525)	(0.491)	(0.382)	(0.317)	(0.830)	(0.672)	(0.256)
p (analytical)	[0.339]	[0.463]	[0.008]	[0.550]	[0.288]	[0.632]	[0.847]	[0.859]
p (resampling)			[0.008]					
Heterogeneity p	[0.242]		[0.013]		[0.409]		[0.807]	

Panel C: Wealth and 10-year Mortality (N = 367,863)

	Sex		College-Graduate		Hospitalized		Income > Median	
	Female	Male	Yes	No	No	Yes	Below	Above
Effect (M SEK)	1.473	-0.193	-0.392	0.792	0.587	1.284	1.684	-0.381
SE	(0.701)	(0.685)	(0.720)	(0.571)	(0.513)	(1.063)	(0.907)	(0.412)
p (analytical)	[0.036]	[0.778]	[0.586]	[0.166]	[0.253]	[0.227]	[0.063]	[0.356]
p (resampling)	[0.010]						[0.052]	
Heterogeneity p	[0.089]		[0.198]		[0.555]		[0.038]	

Notes. This table reports estimates of the effect of wealth on mortality from regressions in which prize amount and all control variables are interacted with indicator variables for different subsamples. Median income is measured in the year before the lottery event. Wealth is scaled so that a coefficient of 1.00 denotes a 1 percentage point increase in morality per 1M SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. Resampling-based p-values described in OA Section VII are reported whenever the analytical p-value is below 10%. The heterogeneity p-value is from an F-test of equal effects in the different subsamples. Hospitalized: whether a winner was hospitalized or not during the five years preceding the lottery event.

TABLE AXII

THE EFFECT OF WEALTH ON MORTALITY RISK BY LOTTERY

Panel A: Wealth and Two-year Mortality

	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Effect (M SEK)	-0.387	0.638	-0.306	-0.141
SE	(0.328)	(0.702)	(0.184)	(0.223)
p (analytical)	[0.237]	[0.363]	[0.098]	[0.527]
p (resampling)			[0.178]	
N	387,813	39,012	3,675	564

Panel B: Wealth and Five-year Mortality

	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Effect (M SEK)	-0.087	0.327	0.084	-0.016
SE	(0.510)	(1.109)	(0.575)	(0.579)
p (analytical)	[0.865]	[0.768]	[0.883]	[0.977]
N	387,813	26,932	2,874	383

Panel C: Wealth and 10-year Mortality

	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Effect (M SEK)	0.647	3.305	0.766	-1.594
SE	(0.663)	(3.170)	(0.823)	(1.525)
p (analytical)	[0.329]	[0.297]	[0.352]	[0.298]
N	356,949	8,954	1,823	137

Notes. This table reports estimates of the effect of wealth on mortality in the different lottery samples. Wealth is scaled so that a coefficient of 1.00 denotes a 1 percentage point increase in morality per 1M SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. Resampling-based p described in OA Section VII are reported whenever the analytical p-value is below 10%.

TABLE AXIII

NON-LINEAR EFFECTS OF WEALTH ON MORTALITY RISK

Panel A: Wealth and Two-year Mortality

	Spline Regressions					Excluding Prizes…		
	Knot at 1M		Knots at 100K and 1M					
	< 1M	> 1M	< 100K	100K to 1M	> 1M	> 4M	> 2M	< 10K
Effect (M SEK)	-0.197	-0.158	3.201	-0.687	-0.098	-0.097	-0.235	-0.142
SE	(0.403)	(0.179)	(3.213)	(0.584)	(0.183)	(0.256)	(0.298)	(0.158)
p (analytical)	[0.625]	[0.376]	[0.319]	[0.239]	[0.595]	[0.705]	[0.431]	[0.371]
Heterogeneity *p*	[0.938]			[0.551]				

Panel B: Wealth and Five-year Mortality

	Spline Regressions					Excluding Prizes…		
	Knot at 1M		Knots at 100K and 1M					
	< 1M	> 1M	< 100K	100K to 1M	> 1M	> 4M	> 2M	< 10K
Effect (M SEK)	0.007	0.041	6.582	-0.994	0.200	-0.086	-0.259	0.242
SE	(0.666)	(0.520)	(4.762)	(0.949)	(0.528)	(0.408)	(0.470)	(0.352)
p (analytical)	[0.991]	[0.938]	[0.167]	[0.295]	[0.705]	[0.834]	[0.582]	[0.492]
Heterogeneity *p*	[0.973]			[0.370]				

Panel C: Wealth and 10-year Mortality

	Spline Regressions					Excluding Prizes…		
	Knot at 1M		Knots at 100K and 1M					
	< 1M	> 1M	< 100K	100K to 1M	> 1M	> 4M	> 2M	< 10K
Effect (M SEK)	1.732	-0.183	11.578	0.081	0.152	1.028	0.951	0.766
SE	(0.984)	(0.856)	(5.926)	(1.389)	(0.875)	(0.602)	(0.669)	(0.535)
p (analytical)	[0.078]	[0.830]	[0.051]	[0.954]	[0.862]	[0.088]	[0.155]	[0.152]
Heterogeneity *p*	[0.220]			[0.116]				

Notes. This table reports non-linear estimates of the effect of wealth on mortality risk: spline regressions with a knot at 1M SEK, and with knots at 100K and 1M SEK, as well as estimates when prizes above 4M and 2M SEK are excluded. The last column shows the estimated effect when prizes below 10K SEK are excluded (but non-winning controls in Kombi are included). Wealth is scaled so that a coefficient of 1.00 denotes a 1 percentage point increase in mortality per 1M SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. The heterogeneity *p*-value is from an *F*-test of equal effects at different prize levels.

TABLE AXIV

THE EFFECT OF WEALTH ON DRUG UTILIZATION: EXTENSIVE MARGIN

	All	Common Causes				Hypothesis-Based Causes			
	Total	Cancer	Circ.	Resp.	Other	Cerebo.	Diabetes	Heart	Mental Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Lottery Sample (N = 279,784). Dependent Variable: 1 if Prescribed Non-zero Amount 2006-2010									
Effect (M SEK)	-0.566	0.163	-0.858	-0.200	-0.618	1.009	-0.113	-0.719	-0.082
SE	(0.438)	(0.145)	(0.708)	(0.799)	(0.472)	(0.655)	(0.383)	(0.700)	(0.740)
p (analytical)	[0.197]	[0.261]	[0.226]	[0.802]	[0.190]	[0.123]	[0.769]	[0.304]	[0.912]
p (resampling)	[0.158]	[0.256]	[0.224]	[0.851]	[0.112]	[0.106]	[0.859]	[0.324]	[0.897]
Proportion Drug Consumers	93.53%	1.03%	56.11%	46.53%	91.59%	32.30%	8.15%	54.18%	33.73%
Panel B: Representative Sample (N = 36,454). Dependent Variable: 1 if Prescribed Non-Zero Amount 2006-2010									
Gradient (M SEK)	0.243	0.047	-0.829	-0.236	0.249	-0.922	-1.048	-0.924	-0.532
SE	(0.150)	(0.095)	(0.315)	(0.364)	(0.168)	(0.334)	(0.218)	(0.318)	(0.348)
p (analytical)	[0.105]	[0.622]	[0.009]	[0.517]	[0.138]	[0.006]	[<0.001]	[0.004]	[0.126]
Proportion Drug Consumers	94.03%	1.03%	57.50%	48.69%	92.34%	33.22%	9.08%	55.50%	37.33%
Heterogeneity p	[0.081]	[0.502]	[0.970]	[0.968]	[0.083]	[0.009]	[0.034]	[0.790]	[0.582]

Notes. This table reports the effect of wealth on extensive margin drug prescription, as well as the gradient between wealth and drug prescription in a representative sample reweighted to match the birth year and sex distribution of the lottery sample. Wealth is scaled so that a regression coefficient of 1.00 denotes a 1 percentage point increase in drug usage per 1M SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. The table also reports resampling-based p-values described in OA Section VII. The heterogeneity p-value is from a two-sided t-test of the null hypothesis that the gradient and causal parameter are equal. Circ.: circulatory disease. Resp.: respiratory disease. Cerebro.: cerebrovascular disease.

TABLE AXV

WEALTH AND DRUG UTILIZATION: INTENSIVE MARGIN

	All	Common Causes				Hypothesis-Based Causes			
	Total	Cancer	Circ.	Resp.	Other	Cerebo.	Diabetes	Heart	Mental Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Lottery Sample (N = 279,784). Dependent Variable: Sum of DDDs Prescribed 2006-2010									
Effect (M SEK)	-97.56	n.a.	-22.61	13.68	-88.63	6.50	4.46	-22.61	-32.50
SE	(84.62)	n.a.	(36.59)	(25.18)	(59.59)	(8.68)	(11.07)	(36.59)	(10.33)
p (analytical)	[0.249]	n.a.	[0.537]	[0.587]	[0.137]	[0.454]	[0.687]	[0.537]	[0.002]
p (resampling)	[0.268]	n.a.	[0.523]	[0.276]	[0.198]	[0.446]	[0.595]	[0.535]	[0.018]
Total Consumption 2006-2010 (DDDs)	4,375	n.a.	1,558	201	2,616	350	138	1,558	312
SD	6,168	n.a.	2,490	750	4,858	638	644	2,490	945
Panel B: Representative Sample (N = 36,454). Dependent Variable: Sum of DDDs Prescribed 2006-2010									
Gradient (M SEK)	-194.15	n.a.	-99.56	-19.00	-75.59	-11.50	-16.37	-99.56	-33.06
SE	(44.112)	n.a.	(17.195)	(5.336)	(36.610)	(4.678)	(4.942)	(17.195)	(6.216)
p (analytical)	[<0.001]	n.a.	[<0.001]	[<0.001]	[0.039]	[0.014]	[<0.001]	[<0.001]	[<0.001]
Total Consumption 2006-2010 (DDDs)	4,659	n.a.	1,611	237	2,811	362	155	1,611	393
SD	6,500	n.a.	2,534	832	5,149	649	675	2,534	1,130
Heterogeneity p	[0.311]	n.a.	[0.057]	[0.204]	[0.852]	[0.068]	[0.086]	[0.057]	[0.963]

Notes. This table shows the effect of wealth on drug prescription at the intensive margin, as well as the gradient between wealth and drug prescription for a representative sample reweighted to match the birth year and sex distribution of the lottery sample. Wealth is scaled in million SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. The table also reports resampling-based p-values described in OA Section VII. The heterogeneity p-value is from a two-sided t-test of the null hypothesis that the gradient and causal parameter are equal. No estimate for cancer drugs is reported, because data on dosage are unavailable. Circ.: circulatory disease. Resp.: respiratory disease. Cerebro.: cerebrovascular disease.

TABLE AXVI

THE EFFECT OF WEALTH ON MENTAL HEALTH DRUGS BY SUB-CATEGORY

	Mental Health (2+6)	Psycholeptics (N05, 3+4+5)	Antipsychotics (N05A)	Anxiolytics (N05B)	Hypnotics & Sedatives (N05C)	Antidepressants (N06A)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect (M SEK)	-32.50	-23.68	-1.38	-7.45	-14.86	-8.82
SE	(10.33)	(6.31)	(1.41)	(2.54)	(4.73)	(7.27)
p (analytical)	[0.002]	[<0.001]	[0.329]	[0.003]	[0.002]	[0.226]
p (resampling)	[0.018]	[0.012]	[0.635]	[0.040]	[0.012]	[0.292]
Total Consumption 2006-2010 (DDDs)	312	167	12	27	128	144
SD of Total Consumption 2006-2010	945	636	161	225	484	522

Notes. This table shows the effect of wealth on different sub-categories of mental health drugs. Wealth is scaled in million SEK and the outcome variable is the sum of DDDs prescribed 2006-2010. All regressions include the full set of baseline controls. Standard errors are clustered by individual. Resampling-based *p*-values are described in OA section VII. The number of observations in all regressions is 279,784.

TABLE AXVII

THE EFFECT OF WEALTH ON DRUG UTILIZATION: ROBUSTNESS

	All	Common Causes				Hypothesis-Based Causes			
	Total	Cancer	Circ.	Resp.	Other	Cerebo.	Diabetes	Heart	Mental Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Count Model									
Incidence rate ratio (IRR)	0.978	n.a.	0.989	1.057	0.965	1.033	n.a.	0.989	0.886
SE	(0.022)	n.a.	(0.026)	(0.098)	(0.027)	(0.029)	n.a.	(0.026)	(0.038)
p (clustered)	[0.329]	n.a.	[0.681]	[0.550]	[0.195]	[0.250]	n.a.	[0.681]	[0.005]
p (overdispersion)	[0.332]	n.a.	[0.678]	[0.257]	[0.243]	[0.297]	n.a.	[0.678]	[0.031]
Total Consumption 2006-2010 (DDDs)	4,375	n.a.	1,558	201	2,616	350	138	1,558	312
SD	6,168	n.a.	2,490	750	4,858	638	644	2,490	945
Winsorized (99 pct)									
Effect (M SEK)	-133.66	n.a.	-31.64	-6.19	-99.34	4.84	0.97	-31.64	-18.97
SE	(69.83)	n.a.	(30.88)	(7.37)	(48.17)	(8.12)	(7.59)	(30.88)	(9.20)
p (analytical)	[0.056]	n.a.	[0.305]	[0.401]	[0.039]	[0.551]	[0.899]	[0.305]	[0.039]
p (resampling)	[0.074]	n.a.	[0.336]	[0.492]	[0.094]	[0.539]	[0.813]	[0.292]	[0.100]
Total Consumption 2006-2010 (DDDs)	4239	n.a.	1513	175	2489	341	117	1513	283
SD	5453	n.a.	2267	527	4062	605	485	2267	719

Notes. The upper panel reports the estimated incidence rate ratios from GLM regressions with Poisson density and a logarithmic link function. Wealth is scaled in million SEK. For example, an incidence rate ratio of 0.886 for mental health drugs implies that the expected count is (100-88.6)% lower for those who win 1M SEK compared to winning nothing, i.e., approximately 35.5 fewer DDDs. There is clear evidence of overdispersion, but GLM models with gamma or negative binomial density do not converge. We therefore report p-values based on standard errors that take overdispersion into account, but these standard errors do not account for within-individual correlations. No estimates are reported for cancer drugs, because data on dosage are unavailable, and for diabetes drugs, because the estimator did not converge. The lower panel shows the estimated effect on the total number of DDDs winsorized at the 99th percentile with standard errors clustered by individual. The lower panel also reports resampling-based *p*-values described in OA Section VII. All regressions include the full set of baseline controls and the number of observations in all regressions is 279,784. Circ.: circulatory disease. Resp.: respiratory disease. Cerebro.: cerebrovascular disease.

TABLE AXVIII

THE EFFECT OF WEALTH ON CAUSE-SPECIFIC HOSPITALIZATIONS

	Hospitalization…		Common Causes				Hypothesis-Based Causes						
	Any	≥ 7 days	Cancer	Circ.	Resp.	Other	Alcohol	Cerebro.	Diabetes	External	Hyper.	Ischemic	Smoking
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Panel A: Five years													
OLS Effect (M SEK)	0.393	0.437	0.085	0.465	-0.192	-0.134	0.201	0.047	0.220	0.410	0.298	0.404	-0.028
SE	(0.617)	(0.474)	(0.290)	(0.445)	(0.317)	(0.587)	(0.147)	(0.219)	(0.244)	(0.360)	(0.334)	(0.295)	(0.129)
p (analytical)	[0.524]	[0.356]	[0.769]	[0.295]	[0.545]	[0.819]	[0.171]	[0.832]	[0.367]	[0.254]	[0.372]	[0.171]	[0.830]
Proportion	38.3%	19.2%	5.5%	15.1%	5.1%	28.4%	0.3%	3.6%	2.7%	8.1%	4.7%	5.8%	0.8%
Gradient	-0.764	-1.006	-0.187	-0.508	-0.447	-0.974	-0.203	-0.209	-0.500	-0.294	-0.310	-0.489	-0.107
SE	(0.316)	(0.250)	(0.161)	(0.256)	(0.154)	(0.294)	(0.053)	(0.137)	(0.126)	(0.190)	(0.191)	(0.165)	(0.068)
Proportion	37.7%	16.9%	5.6%	16.3%	5.0%	27.3%	1.0%	3.2%	3.3%	8.7%	6.8%	6.0%	0.9%
Heterogeneity *p*	[0.095]	[0.007]	[0.413]	[0.058]	[0.469]	[0.200]	[0.010]	[0.324]	[0.009]	[0.084]	[0.114]	[0.008]	[0.587]
Panel B: 10 years													
OLS Effect (M SEK)	-0.027	0.106	-0.087	-0.209	-0.045	0.070	0.255	-0.178	0.019	0.233	-0.065	0.148	0.171
SE	(0.770)	(0.654)	(0.406)	(0.616)	(0.433)	(0.787)	(0.259)	(0.295)	(0.302)	(0.563)	(0.454)	(0.443)	(0.207)
p (analytical)	[0.972]	[0.871]	[0.830]	[0.734]	[0.917]	[0.929]	[0.325]	[0.546]	[0.951]	[0.679]	[0.886]	[0.739]	[0.409]
Proportion	51.2%	25.9%	8.4%	21.3%	7.5%	40.4%	0.5%	5.0%	3.5%	12.7%	7.8%	8.3%	1.1%
Gradient (M SEK)	-0.288	-0.768	0.495	-0.557	-0.451	-0.586	-0.289	0.014	-0.665	-0.077	-0.557	-0.679	-0.056
SE	(0.334)	(0.306)	(0.241)	(0.313)	(0.214)	(0.343)	(0.066)	(0.189)	(0.177)	(0.258)	(0.283)	(0.220)	(0.103)
Proportion	50.9%	23.4%	9.1%	25.3%	8.2%	37.1%	1.4%	5.0%	5.2%	13.9%	15.2%	8.9%	1.5%
Heterogeneity *p*	[0.756]	[0.226]	[0.217]	[0.615]	[0.401]	[0.445]	[0.042]	[0.583]	[0.050]	[0.617]	[0.359]	[0.095]	[0.327]

Notes. This table reports OLS and Probit estimates of the effect of wealth on cause-specific hospitalizations, as well as the gradient for a representative sample reweighted to match the age and sex distribution of the lottery sample. Wealth is scaled so that a regression coefficient of 1.00 denotes a 1 percentage point increase in hospitalization per 1M SEK. All regressions include the full set of baseline controls. Standard errors are clustered by individual. The heterogeneity *p*-value is from a two-sided *t*-test of the null hypothesis that the gradient and causal parameter are equal. Circ.: circulatory disease. Resp.: respiratory disease. Cerebro.: cerebrovascular disease. Hyper.: hypertension.

TABLE AXIX

THE EFFECT OF WEALTH ON HEALTH CARE UTILIZATION BY AGE

Panel A: Wealth and Five-year Hospitalization
(N = 378,099)

	Age 18-44	Age 45-69	Age 70+
Effect (M SEK)	0.927	-0.540	1.024
SE	(1.032)	(0.905)	(1.478)
p (analytical)	[0.369]	[0.551]	[0.488]
Heterogeneity p		[0.480]	

Panel B: Wealth and Five-year Health Index
(N = 418,002)

	Age 18-44	Age 45-69	Age 70+
Effect (M SEK)	-0.006	0.228	1.090
SE	(0.056)	(0.385)	(0.875)
p (analytical)	[0.911]	[0.553]	[0.213]
Heterogeneity p		[0.385]	

Panel C: Wealth and Drug Consumption
(All Drugs, Total DDDs 2006-2010, N = 279,784)

	Age 18-44	Age 45-69	Age 70+
Effect (M SEK)	109.861	-155.074	-277.256
SE	(118.548)	(125.163)	(312.292)
p (analytical)	[0.354]	[0.215]	[0.375]
Heterogeneity p		[0.221]	

Panel D: Wealth and Mental Health Drug Consumption
(Total DDDs 2006-2010, N = 279,784)

	Age 18-44	Age 45-69	Age 70+
Effect (M SEK)	-38.952	-28.294	13.336
SE	(16.269)	(14.081)	(37.214)
p (analytical)	[0.017]	[0.045]	[0.720]
p (resampling)	[0.114]	[0.122]	
Heterogeneity p		[0.434]	

Notes. This table reports the estimated effect of wealth on adult health care utilization from regressions in which lottery winnings and all control variables are interacted with indicator variables for the different age groups. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include the full set of baseline controls. Higher values for the health index denotes worse health. Standard errors are clustered by individual. Resampling-based p-values described in OA Section VII are reported whenever the analytical p-value is below 10%. The heterogeneity p-value is from an F-test of equal effects in the different age groups.

TABLE AXX

THE EFFECT OF WEALTH ON HEALTH CARE UTILIZATION BY SEX, COLLEGE, HEALTH, AND INCOME

Panel A: Wealth and Five-year Hospitalization (N = 378,099)

	Sex		College-Graduate		Hospitalized		Income > Median	
	Female	Male	Yes	No	No	Yes	Below	Above
Effect (M SEK)	1.296	-0.450	0.820	0.311	0.236	1.521	1.246	-0.026
SE	(0.886)	(0.861)	(1.534)	(0.677)	(0.708)	(1.271)	(1.013)	(0.775)
p (analytical)	[0.144]	[0.601]	[0.593]	[0.645]	[0.739]	[0.231]	[0.219]	[0.973]
Heterogeneity p	[0.158]		[0.762]		[0.377]		[0.319]	

Panel B: Wealth and Five-year Health Index (N = 418.002)

	Sex		College-Graduate		Hospitalized		Income > Median	
	Female	Male	Yes	No	No	Yes	Below	Above
Effect (M SEK)	0.551	0.030	-1.484	0.544	0.329	0.553	0.513	0.062
SE	(0.353)	(0.421)	(0.380)	(0.314)	(0.272)	(0.640)	(0.520)	(0.255)
p (analytical)	[0.119]	[0.942]	[0.000]	[0.083]	[0.227]	[0.387]	[0.324]	[0.809]
p (resampling)			[0.006]	[0.066]				
Heterogeneity p	[0.344]		[0.000]		[0.747]		[0.436]	

Panel C: Wealth and Drug Consumption (All Drugs, Total DDDs 2006-2010, N = 279,784)

	Sex		College-Graduate		Hospitalized		Income > Median	
	Female	Male	Yes	No	No	Yes	Below	Above
Effect (M SEK)	-123.317	-59.481	-148.630	-92.525	-191.059	144.329	-160.447	-45.604
SE	(124.030)	(120.507)	(204.777)	(92.636)	(76.750)	(272.679)	(174.726)	(93.819)
p (analytical)	[0.320]	[0.622]	[0.468]	[0.318]	[0.013]	[0.597]	[0.358]	[0.627]
p (resampling)					[0.026]			
Heterogeneity p	[0.712]		[0.803]		[0.236]		[0.563]	

Panel D: Wealth and Mental Health Drug Consumption (Total DDDs 2006-2010, N = 369,905)

	Sex		College-Graduate		Hospitalized		Income > Median	
	Female	Male	Yes	No	No	Yes	Below	Above
Effect (M SEK)	-35.526	-30.661	-64.055	-28.419	-31.705	-51.529	-61.844	-16.051
SE	(15.840)	(13.906)	(19.886)	(12.180)	(10.000)	(30.559)	(20.750)	(11.459)
p (analytical)	[0.025]	[0.027]	[0.001]	[0.020]	[0.002]	[0.092]	[0.003]	[0.161]
p (resampling)	[0.080]	[0.092]	[0.036]	[0.080]	[0.016]	[0.214]	[0.038]	
Heterogeneity p	[0.817]		[0.126]		[0.538]		[0.053]	

Notes. This table reports the estimated effect of wealth on adult health care utilization from regressions in which lottery winnings and all control variables are interacted with indicator variables for different subsamples. Income is measured the year prior to the lottery event. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include the full set of baseline controls. Higher values for the health index denotes worse health. Standard errors are clustered by individual. Resampling-based p-values described in OA Section VII are reported whenever the analytical p-value is below 10%. The heterogeneity p-value is from an F-test of equal effects in the different subsamples. Hospitalized: whether a winner was hospitalized or not during the five years preceding the lottery event.

TABLE AXXI

THE EFFECT OF WEALTH ON HEALTH CARE UTILIZATION BY LOTTERY

Panel A: Wealth and Five-year Hospitalization

	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Effect (M SEK)	0.774	0.367	-0.764	1.598
SE	(1.003)	(2.359)	(0.940)	(1.659)
p (analytical)	[0.440]	[0.876]	[0.416]	[0.336]
N	350,069	24,921	2,741	368

Panel B: Wealth and Five-year Health Index

	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Effect (M SEK)	0.238	0.075	0.518	-0.252
SE	(0.398)	(0.903)	(0.487)	(0.510)
p (analytical)	[0.551]	[0.934]	[0.288]	[0.621]
N	387,813	26,932	2,874	383

Panel C: Wealth and Drug Consumption (All Drugs, Total DDDs 2006-2010)

	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Effect (M SEK)	32.793	-76.223	-148.021	-188.129
SE	(169.530)	(367.399)	(111.723)	(169.401)
p (analytical)	[0.847]	[0.836]	[0.185]	[0.268]
N	254,056	22,914	2,459	355

Panel D: Wealth and Mental Health Drug Consumption (Total DDDs 2006-2010)

	PLS	Kombi	Triss-Lumpsum	Triss-Monthly
Effect (M SEK)	-20.323	-36.386	-38.669	-51.225
SE	(20.136)	(34.170)	(12.457)	(31.966)
p (analytical)	[0.313]	[0.287]	[0.002]	[0.110]
p (resampling)			[0.062]	
N	254,056	22,914	2,459	355

Notes. This table reports the estimated effect of wealth on adult health care utilization in the different lottery samples as well as in the full sample including shared prizes in Triss-Lumpsum. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include the full set of baseline controls. Higher values for the health index denotes worse health. Standard errors are clustered by individual. Resampling-based p-values described in OA Section VII are reported whenever the analytical p-value is below 10%.

TABLE AXXII

NON-LINEAR EFFECTS OF WEALTH ON HEALTH CARE UTILIZATION

Panel A: Wealth and Five-year Hospitalization

	Spline Regressions					Excluding Prizes…		
	Knot at 1M		Knots at 100K and 1M					
	< 1M	> 1M	< 100K	100K to 1M	>1M	> 4M	> 2M	< 10K
Effect (M SEK)	1.653	-0.468	3.895	1.313	-0.414	1.116	1.387	0.624
SE	(1.336)	(0.985)	(8.997)	(1.902)	(1.008)	(0.817)	(0.955)	(0.666)
p (analytical)	[0.216]	[0.635]	[0.665]	[0.490]	[0.681]	[0.172]	[0.147]	[0.349]
Heterogeneity *p*	[0.278]			[0.538]				

Panel B: Wealth and Five-year Health Index

	Spline Regressions					Excluding Prizes…		
	Knot at 1M		Knots at 100K and 1M					
	< 1M	> 1M	< 100K	100K to 1M	>1M	> 4M	> 2M	< 10K
Effect (M SEK)	0.443	0.196	9.676	-0.964	0.420	0.269	0.149	0.379
SE	(0.528)	(0.442)	(3.683)	(0.755)	(0.449)	(0.327)	(0.374)	(0.291)
p (analytical)	[0.401]	[0.657]	[0.009]	[0.202]	[0.350]	[0.412]	[0.690]	[0.193]
Heterogeneity *p*	[0.760]			[0.038]				

Panel C: Wealth and Drug Consumption (All Drugs, Total DDDs 2006-2010)

	Spline Regressions					Excluding Prizes…		
	Knot at 1M		Knots at 100K and 1M					
	< 1M	> 1M	< 100K	100K to 1M	>1M	> 4M	> 2M	< 10K
Effect (M SEK)	96.124	-214.834	679.460	11.556	-202.989	41.643	80.426	-165.400
SE	(222.089)	(105.380)	(1366.495)	(297.775)	(107.432)	(138.625)	(168.703)	(86.281)
p (analytical)	[0.665]	[0.041]	[0.619]	[0.969]	[0.059]	[0.764]	[0.634]	[0.055]
Heterogeneity *p*	[0.266]			[0.494]				

Panel D: Wealth and Mental Health Drug Consumption (Total DDDs 2006-2010)

	Spline Regressions					Excluding Prizes…		
	Knot at 1M		Knots at 100K and 1M					
	< 1M	> 1M	< 100K	100K to 1M	>1M	> 4M	> 2M	< 10K
Effect (M SEK)	-22.260	-38.697	-85.524	-13.089	-39.981	-30.400	-11.601	-33.464
SE	(25.819)	(16.165)	(212.352)	(38.395)	(16.684)	(14.946)	(19.256)	(10.875)
p (analytical)	[0.389]	[0.017]	[0.687]	[0.733]	[0.017]	[0.042]	[0.547]	[0.002]
Heterogeneity *p*	[0.650]			[0.855]				

Notes. This table reports non-linear estimates of the effect of wealth on health care utilization: spline regressions with a knot at 1M SEK, and with knots at 100K and 1M SEK, as well as estimates when prizes above 4M and 2M SEK are excluded. The last column shows the estimated effect when prizes below 10K SEK are excluded (but non-winning controls in Kombi are included). Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include the full set of baseline controls. Higher values for the health index denotes worse health. Standard errors are clustered by individual. Standard errors are clustered by individual. The heterogeneity *p*-value is from an *F*-test of equal effects at different prize amount ranges.

TABLE AXXIII

OVERVIEW OF OUTCOME VARIABLES USED IN INTERGENERATIONAL ANALYSES

	Outcome Measured (Year and/or Age)	Age at Lottery Win	Included Cohorts	Additional Restrictions
Infant Health				
Birth Weight (grams)	At Birth		1987-2010	
Low Birth Weight (< 2500 g.)	At Birth		1987-2010	
Premature (gestation < 37 w.)	At Birth		1987-2010	
Hospitalized within…				
2 years, all causes	1986-2010	Age -1 to 18	1968-2009	
5 years, all causes	1986-2010	Age -1 to 18	1968-2005	
2 years, all causes ≥ 7 days	1986-2010	Age -1 to 18	1968-2009	
5 years, all causes ≥ 7 days	1986-2010	Age -1 to 18	1968-2005	
2 years, respiratory disease	1986-2010	Age -1 to 16	1970-2009	
5 years, respiratory disease	1986-2010	Age -1 to 13	1973-2005	
2 years, external causes	1986-2010	Age -1 to 16	1970-2009	
5 years, external causes	1986-2010	Age -1 to 13	1973-2005	
Body Mass				
BMI	1987-2010, Age ~18	Men aged -1 to 17	1968-1992	
Overweight (BMI > 25)	1987-2010, Age ~18	Men aged -1 to 17	1968-1992	
Obese (BMI > 30)	1987-2010, Age ~18	Men aged -1 to 17	1968-1992	
Drug Consumption				
Total	2006-2010	Age -1 to 18	1968-2007	Child alive in 2010
Mental Health	2006-2010, age 15+	Age -1 to 18	1968-1993	Child alive in 2010
Allergy & Asthma	2006-2010, age 0-18	Age -1 to 15	1990-2007	Child alive in 2010
ADHD	2006-2010, age 6-18	Age -1 to 15	1990-2002	Child alive in 2010
Child Development				
Cognitive Skills	1987-2010	Men aged -1 to 17	1969-1992	
Noncognitive Skills	1987-2010	Men aged -1 to 17	1969-1992	
GPA	1988-2009	Age -1 to 14	1972-1996	
National Test Swedish	2003-2009	Age -1 to 14	1988-1996	
National Test English	2003-2009	Age -1 to 14	1988-1996	
National Test Math	2003-2009	Age -1 to 14	1986-1996	
Parental Investments				
Parental Leave (in days)	1993-2010, Age 0-3		1993-2007	Excluded if 0-3 year younger sibling
Smoke during Pregnancy	1987-2010		1987-2010	
Net Wealth, after 5 years	1999-2007	Age -1 to 18	1976-2003	
School Quality	1988-2009	Age -1 to 14	1972-1996	
Mother/Father's Mental Health	2006-2010			Mother/father alive in 2010; won before 2006; and at least one child ≤ 18 years of age in 2010

Notes. This table summarizes the outcome variables used in the intergenerational analyses and corresponds to Tables 3 to 5 in the pre-analysis plan. The first two columns list the year and age at which outcomes are measured. The fourth column lists the included cohorts for each outcome variable, and the last column lists any additional sample restrictions. Note that for a couple of outcome variables, the numbers above differs by one year compared to those in the pre-analysis plan. This is due to data limitations that became apparent only when we started analyzing the data. Another discrepancy compared to the pre-analysis plan is that school quality can only be measured in the year of graduation at age 15 or 16. BMI: Body Mass Index. ADHD: Attention Deficit Hyperactivity Disorder. Age -1: children that were conceived but not born at the time of the lottery win.

TABLE AXXIV

THE EFFECT OF WEALTH ON CHILD OUTCOMES: LOTTERY WINNINGS PER CHILD

Panel A: Child Health

	Infant Health	Hospitalizations					Drug Cons.	BMI	
	Birth Weight (grams)	All Causes		≥ 7 days	Respi-ratory	External	Total	BMI	Obese
		t = 2	*t = 5*	*t = 5*	*t = 5*	*t = 5*			
Effect (M SEK)	-43.624	2.119	3.443	0.227	0.949	1.017	8.895	-0.050	-2.692
SE	(40.887)	(0.747)	(0.913)	(0.320)	(0.776)	(0.768)	(27.656)	(0.366)	(1.124)
p (analytical)	[0.286]	[0.005]	[<0.001]	[0.478]	[0.222]	[0.185]	[0.748]	[0.890]	[0.017]
p (resampling)		[<0.001]	[<0.001]						[0.266]
Proportion/mean	3546.8	10.9%	18.0%	2.8%	4.8%	4.9%	468.3	22.4	2.9%
SD	588.4						1540.5	3.2	
N	54,575	114,160	111,064	111,064	75,382	75,382	112,223	32,646	32,646

Panel B: Child Development

	Cognitive Skills	Noncognitive Skills	GPA	Swedish	English	Math
Effect (M SEK)	-0.191	-0.034	-0.030	-0.054	-0.105	-0.032
SE	(0.076)	(0.124)	(0.045)	(0.069)	(0.085)	(0.093)
p (analytical)	[0.012]	[0.781]	[0.515]	[0.432]	[0.218]	[0.734]
p (resampling)	[0.056]					
Mean	0.169	0.143	0.271	0.163	0.105	0.204
SD	0.976	0.976	0.938	0.996	0.990	1.013
N	36,435	31,550	74,187	25,079	25,286	23,990

Notes. This table reports the effect of wealth on child health (Panel A) and development (Panel B) when the treatment variable is prize amount per child. In our infant health analyses, prize amount is divided by the number of pre- and post-lottery children of the winner (at the time infant health variables are measured), whereas prize amount is divided by the number of pre-lottery children in all other analyses. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include controls for birth characteristics of the child as well as demographic characteristics for the winning parent. Standard errors are clustered using an iterative algorithm that assigns siblings and half-siblings to the same cluster. Resampling-based *p*-values described in OA Section VII are reported whenever the analytical *p*-value is below 10%. BMI: Body Mass Index.

TABLE AXXV

NON-LINEAR EFFECTS OF WEALTH ON CHILD OUTCOMES

	Child Health									Child Development					
	Birth Weight (grams)	Hospitalization: All Causes $t = 2$	Hospitalization: All Causes $t = 5$	Hospitalization: $\geq$ 7 days $t = 5$	Hospitalization: Resp. $t = 5$	Hospitalization: External $t = 5$	Total Drug Cons.	BMI	Obese	Cognitive Skills	Non-cognitive Skills	GPA	Swedish	English	Math
								Excluding...							
Prizes > 4M															
Effect (M SEK)	-12.542	2.468	3.681	0.856	-0.040	1.239	55.167	-0.189	-1.484	-0.057	-0.022	-0.039	-0.070	-0.110	-0.065
SE	(40.163)	(1.308)	(1.620)	(0.747)	(0.980)	(1.419)	(64.579)	(0.255)	(0.695)	(0.071)	(0.082)	(0.042)	(0.058)	(0.065)	(0.079)
p (analytical)	[0.755]	[0.059]	[0.023]	[0.252]	[0.968]	[0.383]	[0.393]	[0.459]	[0.033]	[0.419]	[0.790]	[0.350]	[0.227]	[0.092]	[0.408]
p (resampling)		[0.024]	[0.010]						[0.653]					[0.092]	
Proportion/mean	3550.0	10.9%	18.0%	2.8%	4.8%	4.9%	468.6	22.4	2.9%	0.169	0.143	0.271	0.163	0.106	0.204
N	54,549	114,066	110,991	110,991	75,325	75,325	112,137	32,641	32,641	36,429	31,544	74,156	25,055	25,261	23,967
Prizes > 2M															
Effect (M SEK)	-65.460	1.143	3.067	0.571	-1.049	-0.790	-10.949	0.167	-1.287	-0.098	-0.042	-0.036	0.039	0.004	-0.085
SE	(68.900)	(1.864)	(2.333)	(0.961)	(1.326)	(1.336)	(104.811)	(0.284)	(0.695)	(0.095)	(0.095)	(0.065)	(0.119)	(0.133)	(0.141)
p (analytical)	[0.342]	[0.540]	[0.189]	[0.553]	[0.429]	[0.554]	[0.917]	[0.555]	[0.064]	[0.303]	[0.663]	[0.575]	[0.741]	[0.975]	[0.547]
p (resampling)									[0.719]						
Proportion/mean	3550.0	10.9%	18.0%	2.8%	4.8%	4.9%	468.6	22.4	2.9%	0.169	0.143	0.272	0.164	0.106	0.205
N	54,519	113,881	110,871	110,871	75,236	75,236	111,981	32,627	32,627	36,410	31,529	74,092	25,006	25,210	23,919
PLS prizes < 10K															
Effect (M SEK)	-18.230	2.447	3.931	0.373	1.154	1.306	18.706	-0.182	-2.292	-0.110	0.002	-0.028	-0.040	-0.092	-0.041
SE	(29.944)	(0.763)	(0.943)	(0.341)	(0.798)	(0.819)	(27.511)	(0.266)	(1.082)	(0.056)	(0.083)	(0.028)	(0.042)	(0.046)	(0.054)
p (analytical)	[0.543]	[0.001]	[<0.001]	[0.274]	[0.148]	[0.111]	[0.497]	[0.494]	[0.034]	[0.047]	[0.977]	[0.315]	[0.341]	[0.046]	[0.448]
p (resampling)		[<0.001]	[<0.001]						[0.002]	[0.048]				[0.088]	
Proportion/mean	3,550.0	10.2%	17.3%	2.6%	4.4%	4.8%	413.6	22.5	3.4%	0.134	0.090	0.190	0.119	0.068	0.084
N	5,892	18,507	15,476	15,476	9,867	9,867	17,059	3,656	3,656	4,187	3,557	9,145	3,908	3,928	3,791

Notes. This table reports the effect of wealth on child outcomes when prizes above 4M and 2M SEK are excluded, as well as when prizes below 10K SEK are excluded (but non-winning controls in Kombi are included). Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include controls for the children's birth characteristics and demographic characteristics for the winning parent. Standard errors are clustered using an iterative algorithm that assigns siblings and half-siblings to the same cluster. Resampling-based p-values described in OA Section VII are reported whenever the analytical p-value is below 10%. Resp.: respiratory disease. BMI: Body Mass Index.

TABLE AXXVI

THE EFFECT OF WEALTH ON CHILD OUTCOMES BY LOTTERY

	Child Health									Child Development					
	Birth Weight (grams)	Hospitalization: All Causes $t=2$	Hospitalization: All Causes $t=5$	Hospitalization: $\geq$ 7 days $t=5$	Hospitalization: Resp. $t=5$	Hospitalization: External $t=5$	Total Drug Cons.	BMI	Obese	Cognitive Skills	Non-cognitive Skills	GPA	Swedish	English	Math
PLS															
Effect (M SEK)	-13.595	1.495	2.967	-1.370	-0.301	-0.265	4.045	-0.225	-1.707	-0.118	-0.115	-0.013	-0.022	-0.051	0.036
SE	(66.859)	(2.049)	(2.255)	(0.522)	(1.403)	(1.586)	(117.924)	(0.269)	(0.438)	(0.088)	(0.084)	(0.054)	(0.090)	(0.127)	(0.127)
p (analytical)	[0.839]	[0.466]	[0.188]	[0.009]	[0.830]	[0.867]	[0.973]	[0.404]	[<0.001]	[0.178]	[0.174]	[0.806]	[0.812]	[0.691]	[0.778]
p (resampling)				[0.144]					[0.288]						
Kombi															
Effect (M SEK)	10.627	2.366	5.996	2.393	0.823	-1.000	-28.155	-0.373	-5.835	-0.168	0.288	0.171	-0.069	0.076	0.179
SE	(79.549)	(2.606)	(5.077)	(2.279)	(3.565)	(3.418)	(84.059)	(0.703)	(2.474)	(0.385)	(0.426)	(0.187)	(0.234)	(0.223)	(0.270)
p (analytical)	[0.894]	[0.364]	[0.238]	[0.294]	[0.818]	[0.770]	[0.738]	[0.596]	[0.019]	[0.663]	[0.500]	[0.360]	[0.768]	[0.732]	[0.508]
p (resampling)									[0.553]						
Triss-Lumpsum															
Effect (M SEK)	-8.583	2.880	3.584	0.487	1.665	1.356	16.058	-0.052	-2.918	-0.116	-0.006	-0.042	-0.054	-0.137	-0.078
SE	(35.319)	(1.080)	(1.242)	(0.476)	(1.073)	(1.125)	(29.661)	(0.414)	(1.969)	(0.047)	(0.114)	(0.037)	(0.052)	(0.057)	(0.071)
p (analytical)	[0.808]	[0.008]	[0.004]	[0.306]	[0.121]	[0.229]	[0.588]	[0.900]	[0.140]	[0.015]	[0.955]	[0.249]	[0.301]	[0.016]	[0.270]
p (resampling)		[0.004]	[0.012]							[0.060]				[0.064]	
Triss-Monthly															
Effect (M SEK)	-467.072	-0.225	1.610	0.446	-1.318	-0.006	2.196					-0.100	-0.068	-0.133	-0.070
SE	(145.655)	(1.195)	(2.550)	(0.441)	(1.911)	(1.991)	(20.153)					(0.087)	(0.176)	[0.144]	(0.131)
p (analytical)	[0.004]	[0.851]	[0.529]	[0.315]	[0.493]	[0.998]	[0.913]					[0.255]	[0.701]	[0.361]	[0.595]
p (resampling)	[0.010]														

Notes. This table reports the effect of wealth on child outcomes for the different lottery samples. Estimating the effect on BMI, cognitive and noncognitive skills for Triss-Monthly is not possible, because the effective sample size is too small to yield meaningful estimates. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include controls for the children's birth characteristics and demographic characteristics for the winning parent. Standard errors are clustered using an iterative algorithm that assigns siblings and half-siblings to the same cluster. Resampling-based p-values described in OA Section VII are reported whenever the analytical p-value is below 10%. Resp.: respiratory disease. BMI: Body Mass Index.

TABLE AXXVII

THE EFFECT OF WEALTH CHILD OUTCOMES BY INCOME AND SEX OF WINNING PARENT

	Child Health									Child Development					
	Birth Weight (grams)	Hospitalization: All Causes $t = 2$	Hospitalization: All Causes $t = 5$	Hospitalization: $\geq$ 7 days $t = 5$	Hospitalization: Resp. $t = 5$	Hospitalization: External $t = 5$	Total Drug Cons.	BMI	Obese	Cognitive Skills	Noncog. Skills	GPA	Swedish	English	Math
Household Income															
Bottom Quartile	-158.109	2.851	3.625	0.955	4.899	-0.499	3.451	-0.472	-7.050	-0.215	-0.482	-0.059	-0.057	-0.137	-0.125
SE	(50.244)	(1.399)	(1.863)	(0.963)	(1.538)	(1.128)	(83.919)	(0.483)	(1.856)	(0.106)	(0.107)	(0.047)	(0.047)	(0.055)	(0.068)
p (analytical)	[0.002]	[0.042]	[0.052]	[0.322]	[0.001]	[0.658]	[0.967]	[0.329]	[0.000]	[0.043]	[0.000]	[0.209]	[0.229]	[0.013]	[0.067]
p (resampling)	[0.102]	[0.050]	[0.058]		[0.008]				[0.004]	[0.026]	[0.002]			[0.224]	[0.282]
Quartiles 2 to 4	16.808	1.978	3.748	0.084	-0.328	1.359	10.427	-0.147	-1.566	-0.104	0.036	-0.030	-0.020	-0.094	0.020
SE	(30.470)	(0.894)	(1.034)	(0.320)	(0.575)	(0.904)	(27.633)	(0.234)	(1.134)	(0.058)	(0.079)	(0.033)	(0.053)	(0.056)	(0.067)
p (analytical)	[0.581]	[0.027]	[0.000]	[0.793]	[0.569]	[0.133]	[0.706]	[0.529]	[0.167]	[0.075]	[0.651]	[0.357]	[0.704]	[0.091]	[0.768]
p (resampling)		[0.026]	[0.006]							[0.276]				[0.138]	
Heterogeneity p	[0.003]	[0.597]	[0.953]	[0.390]	[0.001]	[0.196]	[0.937]	[0.545]	[0.012]	[0.360]	[0.000]	[0.621]	[0.604]	[0.588]	[0.130]
N	54,575	114,160	111,064	111,064	75,382	75,382	112,223	32,646	32,646	36,435	31,550	74,187	25,079	25,286	23,990
Winning Parent															
Mother	16.265	3.748	4.241	1.448	0.301	0.852	6.399	-0.604	-2.277	-0.051	-0.043	0.003	-0.031	-0.175	-0.056
SE	(42.769)	(1.201)	(1.569)	(0.681)	(0.828)	(1.396)	(36.458)	(0.275)	(1.017)	(0.092)	(0.091)	(0.042)	(0.066)	(0.064)	(0.089)
p (analytical)	[0.704]	[0.002]	[0.007]	[0.034]	[0.716]	[0.542]	[0.861]	[0.028]	[0.025]	[0.582]	[0.637]	[0.948]	[0.640]	[0.006]	[0.530]
p (resampling)		[0.000]	[0.008]	[0.050]				[0.074]	[0.484]					[0.022]	
Father	-76.720	1.377	3.465	-0.601	1.885	1.698	20.220	0.175	-1.451	-0.182	-0.006	-0.060	-0.109	-0.101	-0.046
SE	(35.804)	(1.046)	(1.155)	(0.211)	(1.102)	(0.981)	(44.582)	(0.288)	(0.970)	(0.063)	(0.102)	(0.044)	(0.047)	(0.069)	(0.087)
p (analytical)	[0.032]	[0.188]	[0.003]	[0.004]	[0.087]	[0.083]	[0.650]	[0.543]	[0.135]	[0.004]	[0.956]	[0.166]	[0.022]	[0.145]	[0.600]
p (resampling)	[0.024]		[0.004]	[0.252]	[0.040]	[0.096]				[0.046]			[0.048]		
Heterogeneity p	[0.096]	[0.136]	[0.690]	[0.004]	[0.251]	[0.620]	[0.810]	[0.050]	[0.557]	[0.237]	[0.785]	[0.298]	[0.338]	[0.431]	[0.936]
N	54,575	114,160	111,064	111,064	75,382	75,382	112,223	32,646	32,646	36,435	31,550	74,187	25,079	25,286	23,990

Notes. This table separately reports the effect of wealth on child outcomes for children with biological parents that have a total disposable income above or below the 25th percentile in the year prior to the lottery event as well as by the sex of the winning parent. The estimates are obtained from regressions in which prize amount and all control variables are interacted with indicator variables for different subsamples. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include controls for the children's birth characteristics and demographic characteristics for the winning parent. Standard errors are clustered using an iterative algorithm that assigns siblings and half-siblings to the same cluster. Resampling-based p-values described in OA Section VII are reported whenever the analytical p-value is below 10%. The heterogeneity p-value is from an F-test of equal effects in the two subsamples. Resp.: respiratory disease. BMI: Body Mass Index.

TABLE AXXVIII

THE EFFECT OF WEALTH ON CHILD OUTCOMES BY AGE AND SEX

	Child Health								Child Development					
	Hospitalization					Total Drug Cons.	BMI	Obese	Cognitive skills	Non-cognitive skills	GPA	Swedish	English	Math
	All Causes		≥ 7 days	Resp.	External									
	t = 2	*t* = 5	*t* = 5	*t* = 5	*t* = 5									
Child Age														
Below 9	2.272	4.833	-0.019	1.551	0.634	42.708	-0.327	-4.187	-0.052	0.188	-0.044	-0.085	-0.122	-0.051
SE	(1.123)	(1.365)	(0.434)	(1.022)	(0.995)	(48.752)	(0.347)	(2.646)	(0.112)	(0.105)	(0.047)	(0.038)	(0.051)	(0.069)
p (analytical)	[0.043]	[<0.001]	[0.965]	[0.129]	[0.524]	[0.381]	[0.346]	[0.114]	[0.643]	[0.072]	[0.347]	[0.025]	[0.017]	[0.460]
p (resampling)	[0.042]	[<0.001]								[0.178]		[0.134]	[0.034]	
Above 9	2.472	2.324	0.347	-0.331	1.683	-29.792	-0.040	-1.397	-0.142	-0.091	0.012	0.094	0.023	0.103
SE	(1.150)	(1.560)	(0.507)	(0.522)	(1.722)	(34.601)	(0.321)	(0.850)	(0.054)	(0.066)	(0.032)	(0.062)	(0.081)	(0.081)
p (analytical)	[0.032]	[0.136]	[0.493]	[0.527]	[0.328]	[0.389]	[0.901]	[0.100]	[0.008]	[0.171]	[0.710]	[0.130]	[0.773]	[0.202]
p (resampling)	[0.018]								[0.044]					
Heterogeneity *p*	[0.902]	[0.258]	[0.582]	[0.103]	[0.626]	[0.223]	[0.557]	[0.323]	[0.460]	[0.008]	[0.330]	[0.014]	[0.130]	[0.149]
N	114,160	111,064	111,064	75,382	75,382	112,223	32,646	32,646	36,435	31,550	74,187	25,079	25,286	23,990
Child's Sex														
Female	2.900	3.493	0.679	0.533	1.330	23.596					-0.058	0.048	-0.092	-0.011
SE	(1.215)	(1.445)	(0.577)	(1.184)	(1.037)	(49.923)					(0.043)	(0.074)	(0.067)	(0.092)
p (analytical)	[0.017]	[0.016]	[0.239]	[0.652]	[0.200]	[0.636]					[0.179]	[0.518]	[0.166]	[0.902]
p (resampling)	[0.012]	[0.012]												
Male	1.490	3.302	-0.091	1.561	0.610	-5.724					0.003	-0.051	-0.050	-0.011
SE	(1.082)	(1.432)	(0.405)	(1.256)	(1.238)	(27.239)					(0.033)	(0.047)	(0.061)	(0.070)
p (analytical)	[0.168]	[0.021]	[0.823]	[0.214]	[0.622]	[0.834]					[0.924]	[0.275]	[0.408]	[0.877]
p (resampling)		[0.020]												
Heterogeneity *p*	[0.410]	[0.930]	[0.263]	[0.570]	[0.644]	[0.603]					[0.243]	[0.255]	[0.646]	[0.997]
N	114,160	111,064	111,064	75,382	75,382	112,223					74,187	25,079	25,286	23,990

Notes. This table separately reports heterogeneous effects of wealth on child outcomes depending on the age of the child when the parent won and the sex of the child. The estimates are obtained from regressions in which lottery winnings and all control variables are interacted with indicator variables for different subsamples. The effect on BMI, cognitive and noncognitive skills cannot be estimated separately for female and male children, because conscription data is only available for men. Wealth is scaled so that a marginal effect of 1.00 denotes a 1 percentage point increase per 1M SEK for binary outcomes, whereas it is scaled in million SEK for other outcomes. All regressions include controls for the children's birth characteristics and demographic characteristics of the winning parent. Standard errors are clustered using an iterative algorithm that assigns siblings and half-siblings to the same cluster. Resampling-based *p*-values described in OA Section VII are reported whenever the analytical *p*-value is below 10%. The heterogeneity *p*-value is from an *F*-test of equal effects in the two subsamples. Resp.: respiratory disease. BMI: Body Mass Index.

TABLE AXXIX

PROXIES FOR PARENTAL INVESTMENT

	Pre-Lottery Children				Post-Lottery Children		
	Wealth	School Quality	Mental Health Drugs		Maternal Leave	Paternal Leave	Smoked Pregnancy
			Mother	Father			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Effect (M SEK)	0.006	-0.009	-35.526	-30.661	-18.466	5.411	-0.953
SE	(0.007)	(0.011)	(15.833)	(13.913)	(15.880)	(8.361)	(1.040)
p (analytical)	[0.399]	[0.394]	[0.025]	[0.028]	[0.245]	[0.518]	[0.359]
p (resampling)	[0.342]	[0.470]	[0.092]	[0.126]	[0.178]	[0.250]	[0.262]
Mean/Proportion	0.08	0.02	389.61	224.02	385.99	79.79	6.93%
SD	0.18	0.27	1002.49	868.67	177.82	119.80	
N	36,982	74,459	147,969	131,815	9,868	11,750	51,753
Unit of analysis	Child	Child	Winner	Winner	Child	Child	Child

Notes. This table shows the estimated effect of wealth on different proxies for parental investment. Column (1) reports the effect on child wealth five years after the lottery event. Child wealth is measured in million SEK and winsorized at the 99th percentile. School quality refers to the average GPA in the school in which the child finished ninth grade. Mental health drug consumption is the total number of DDDs consumed by the mother/father for 2006-2010. Maternal/paternal leave is measured in number of days of paid parental leave. Smoking during pregnancy is an indicator variable indicating whether the mother reported having smoked during pregnancy. Wealth is scaled in million SEK prices in columns (2)-(6). The coefficient for mother's smoking in column (7) is scaled so that a regression coefficien of 1.00 denotes a 1 percentage point increase per 1M SEK. All regressions include controls for the children's birth characteristics and demographic characteristics for the winning parent. Standard errors are clustered using an iterative algorithm that assigns siblings and half-siblings to the same cluster in all regressions in which children is the unit of analysis, and by individual for parental mental health. Resampling-based *p* -values are described in OA Section VII.

III. The PLS Sample

In this section we describe the Swedish prize-linked savings (PLS) data that were used to construct the PLS sample. We begin with a brief history of Swedish PLS accounts.[1]

III.A. Background on the PLS Accounts

The history of Swedish PLS accounts can be traced back to 1949 when subsidized savings accounts for the youth ("Ungdomens Lönsparande") were introduced. The holders of these accounts were randomly assigned money in regular prize draws. During the 1960s, a similar program for adults ("Allmänna Lönsparandet") was established, and in 1972, the two were merged ("Nya Lönsparandet"). All of these programs were subsidized by the Swedish government, but administered by the banking sector. In 1984, the government decided to stop the subsidies and the last prize draw was held in September 1985. Because PLS accounts were popular, banks expressed an interest in continuing to offer the product, and as a consequence, two new types of unsubsidized PLS accounts were introduced to replace the old program.

PLS FIGURE I
PLS lottery draw
Photograph from one of the VK lottery draws conducted in the early 1990s (date unknown). The man in the dark suit is Per Ekström (head of the VK unit at the Swedish Bank Föreningssparbanken), who kindly provided us with the picture.

[1] The sources that we use are listed in the Bibliography under Additional Sources, PLS (Section X.A), at the end of this Appendix.

The first of these was operated by the commercial and state-owned banks and was called "vinnarkontot" (VK, "winners' account"). The second was operated by the savings and loans banks and was called "miljonkontot" (MK, "million account"). As part of the transition to the new system, existing PLS accounts were converted to VK and MK accounts. Although the PLS accounts were fully privatized in 1985, the government continued to strictly regulate and monitor them due to their lottery component, and consequently the relevant authorities carefully supervised draws. PLS Figure I above depicts one of the lottery draws.

PLS TABLE I

EVOLUTION OF VK AND MK ACCOUNTS

	Vinnarkontot (VK)		Miljonkontot (MK)	
	Accounts	Balance	Accounts	Balance
1986	1,898,211	11,460	2,013,000	10,900
1989	2,460,574	13,718	2,403,000	13,405
1994	2,183,369	7,205	1,873,000	6,565
1998	1,317,715	3,527		
2003			1,138,000	3,029

Notes. This table shows the change over time in the number of VK and MK accounts and their aggregate balances. It is based on documents provided by Per Ekström and Johan Hansing. No information is available for MK in 1998 and VK in 2003. Balance is given in millions of SEK at current prices.

Like the subsidized PLS accounts that preceded them, VK and MK accounts were popular. In 1983, there were more than 1.5 million PLS accounts in Sweden, and by 1989, this number had increased to well over 4.8 million, split almost equally between VK and MK. PLS Table I outlines the evolution of the number of accounts by type and their aggregate balances from 1986 to 2003. As can be seen, after a substantial increase in the 1980s, the number of accounts gradually declined during the 1990s. Several reasons can explain this development. First, although PLS accounts were not subsidized after 1985, the old accounts were automatically enrolled in the new system operated by the banks. Many account holders likely did not immediately respond to these new incentives by closing their accounts. Second, following the reform, PLS accounts increased in popularity. The reasons for this development are unclear, but it could be related to the fact that the government allowed the banks to substantially increase the lottery share of total interest paid. Third, the tax reform in l99l probably contributed to the decline in the popularity of the accounts. Prior to the reform, both labor and capital income were taxed according to a progressive tax scheme, whereas prizes from PLS accounts were taxed at a flat rate of 40%, initially 30%, of the gross prize amount. After the reform, the government introduced a capital gains tax of 30% and similarly reduced the tax rate on PLS

prizes to 30%, making PLS accounts relatively less attractive to many savers. Fourth, and perhaps most importantly, nominal interest rates dropped dramatically in the 1990s, lowering the total nominal value of the prize pool. In addition, Sweden suffered a major recession in the early 1990s and some individuals may have opted to close their accounts due to liquidity needs.

In response to the declining popularity of the PLS accounts, all participating banks except one (Nordbanken/Nordea) stopped offering VK accounts in 1999, and in 2003, the VK accounts were finally abolished. Likewise, the MK accounts were discontinued in 2004. Unfortunately, to our knowledge, the MK data have not been preserved, and consequently our study is based uniquely on the VK accounts. In what follows, we use the terms 'PLS' or 'PLS Sample' to refer to the VK accounts for which we have data.

III.B. Lottery Rules

During the entire period of study, PLS account holders could win prizes in three different categories: birthday prizes, fixed prizes, and odds prizes. Birthday prizes, which we ignore throughout, were small prizes (100 or 250 SEK) awarded to account holders born on randomly chosen days. Fixed prizes were prizes whose size was independent of the account balance of the winning account. The fixed prizes varied in size, from 1,000 to 2 million SEK. Odds prizes were prizes that paid winners a multiple (1, 10 or 100) of their account balance, with the maximum prize being capped at 1 million SEK (up until May 1996). Prize plans determined the distribution of odds and fixed prizes for each draw. The prize plans were frequently revised, but adjustments were relatively modest. In addition, an "extra prize" was occasionally awarded to ensure that the prize pool of each draw was fully exhausted.

As we explain in Section III.C, we have detailed account-level data obtained from digitized microfiche cards for the period December 1986 to December 1994 (the 'fiche period'). In addition, we have information about all prizes awarded between October 1979 and September 2003 (the 'full period'). We briefly describe the data available before the fiche period in this Appendix, although we do not use this data in our analyses. PLS Table II summarizes the most important changes to prizes during the period 1979-2003, whereas PLS Table III provides additional details for the fiche period.

PLS TABLE II

SUMMARY OF PRIZE PLANS 1979-2003

	Frequency	Largest Fixed Prize	Odds prizes: Type	Odds prizes: Capped?	Tickets Observed
1979-1982: Sep	Annual	250,000	None		No
1983-1985: Mar	Bi-annual	250,001	None		No
1983-1985: Sep	Bi-annual	400,000	None		No
1985-1986: May and Nov	Bi-annual	1,000,000	1/10/100	1,000,000	No
Dec 1986 to Oct 1992	Monthly	1,000,000	10/100	1,000,000	Yes
Nov 1992 to Dec 1994	Monthly	2,000,000	1/10/100	1,000,000	Yes
Jan 1995 to Apr 1996	Monthly	2,000,000	1/10/100	1,000,000	No
May 1996 to Sep 1996	Monthly	1,000,000	10	100,000	No
Dec 1996 to Dec 1998	Quarterly	1,000,000	10	100,000	No
Mar 1999 and Jun 1999	Data missing				
Sep 1999 to Jun 2000	Quarterly	1,000,000	10	100,000	No
Sep 2000 to Sep 2003	Quarterly	10,000	10	100,000	No

Notes. This table summarizes the key information in the prize plans for the full period.

PLS TABLE III

DETAILED PRIZE PLANS FOR THE FICHE PERIOD (DEC 1986 TO DEC 1994)

	Fixed prizes						Odds prizes (%)		
	2M	1M	0.5M	0.1M	10,000	1,000	100	10	1
Dec 86 to Feb 90		3	2	20	200	2,000	80	20	
Mar 90 to Oct 92		3		20	*400*	3,000	50	50	
Nov 92 to Feb 94	1	1		20	200	3,000	33	33	33
Mar 94 to Dec 94	1	1		10	100	1,500	22*	43*	35*

Notes. This table shows the allocation of prizes during the fiche period according to the prize plans. The exact rules determining the distribution of odds prizes between March 1994 and December 1994 are not clear, and we therefore report the actual distribution of prizes awarded.

Prior to 1985, the eligibility for participation in prize draws was subject to restrictions on account balances. For 1979 to 1982, draws were annual and the increase in the balance during the preceding year determined the number of tickets associated with an account. In 1983, a second yearly draw was added and the number of tickets for this draw was unrelated to the balance increase and instead determined by the lowest balance in the two preceding years. After 1985, the number of lottery tickets in each draw was simply equal to the account balance at the time of the draw divided by 100 and rounded down to the nearest integer. However, accounts had to have a minimum balance of 800 SEK to be eligible (1,000 SEK for the draw in May 1985). In 1986, the number of draws was increased and from December 1986 to September

1996 draws were monthly, and thereafter quarterly.

III.C. Digitizing the Fiches and Data Processing

We compiled the PLS data from two primary sources: microfiche volumes and prize lists. We describe each in turn.

Microfiche volumes. We have 99 volumes of microfiche cards. Each volume corresponds to a specific lottery draw and contains around 150 microfiche cards with information about all accounts that were eligible to participate in the draw. For each account, the microfiche cards contain the account number, the personal identification number (PIN) of the owner, and the balance. An example of the microfiche cards can be seen in PLS Figure II. Each microfiche card has about 250 images, and each image, in turn, contains approximately 50 rows ordered by account number, with each row representing a unique account. PLS Figure III shows an example image from the August 1991 draw. The first column (*LÖPNUMMER*) lists the number of lottery tickets associated with each account. The second column (*KONTONUMMER*) shows the account numbers. The third column (*SALDO*) lists the account balances, the fourth column (*ANTAL ANDELAR*) the number of tickets, and the fifth column (*PERSONNUMMER*) shows the account holders' PINs. PINs are redacted from the example image for confidentiality reasons.

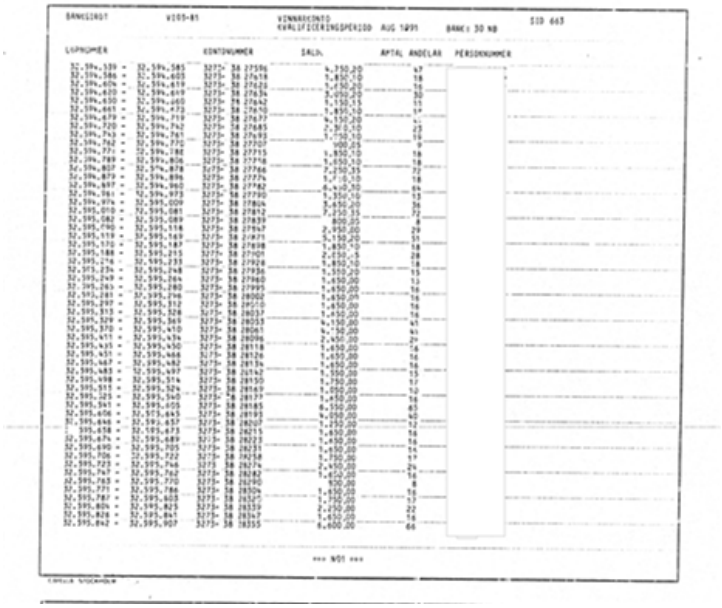
Of the 99 volumes, 97 correspond to the monthly draws conducted during the fiche period and one to a practice draw in November 1986. We use data from the practice draw to link individuals to accounts and to impute account balances for later draws. The final volume belongs to the transition period 1985-1986 with bi-annual draws and does not include account balances, and is consequently not used in this paper.

Prize lists. We have the prize lists from all but three draws for the period October 1979 to September 2003. The prize list data is almost complete, although accounts from some banks are missing from some draws (see PLS Table V). An example of a prize list can be found in PLS Figure IV. Each prize list contains the winning account numbers, as well as the amount and type of prize won. All prize lists were entered twice separately by different research assistants and then double-checked for discrepancies. The lists do not contain neither PINs nor account balances. However, for odds prizes, they can in most instances be used to infer account balances because prize amounts were multiples thereof.

PLS Figure II

A sample microfiche card

Each microfiche card contains around 250 microfiche images (see PLS Figure III for an example image).



PLS Figure III

A sample microfiche image from August 1991

PINs redacted for privacy reasons.

UTLOTTADE FASTA VINSTER OCH ODDSVINSTER FÖR HANDELSBANKEN

Kontonummer	Vinstbelopp kr
6001- 101 640 021	1 000
6001- 154 254 851	1 000
6101- 241 990 742	1 000
6102- 147 460 581	1 000
6102- 309 467 241	1 000
6105- 125 574 681	1 000
6105- 137 741 081	10 780
6105- 199 690 251	20 744
6105- 355 899 531	12 735
6106- 121 809 862	1 000
6106- 128 471 492	1 000
6107- 058 749 926	1 000
6107- 058 749 926	1 000
6107- 134 955 102	1 000
6107- 168 908 972	1 000
6107- 382 679 482	10 000
6109- 159 474 671	1 000
6109- 365 267 031	1 000
6109- 398 695 571	1 000
6110- 309 536 901	1 000
6110- 358 608 481	1 000
6111- 390 843 911	10 000

Kontonummer	Vinstbelopp kr
6142- 718 863 178	1 000
6144- 097 430 358	1 000
6144- 303 256 648	1 000
6144- 528 994 158	10 000
6144- 536 094 578	1 000
6145- 546 654 398	1 000
6145- 711 765 588	1 000
6145- 791 226 468	1 000
6151- 159 333 202	1 000
6151- 303 957 611	1 000
6151- 323 089 542	1 000
6151- 382 597 532	10 000
6152- 165 867 051	1 000
6153- 157 194 191	1 000
6153- 257 954 821	1 000
6153- 372 298 931	1 000
6155- 208 010 831	1 000
6155- 273 755 331	1 000
6156- 161 884 962	1 000
6158- 332 879 852	1 000
6159- 295 460 822	1 000
6161- 166 657 662	1 000

Kontonummer	Vinstbelopp kr
6196- 125 188 102	1 000
6196- 177 444 002	1 000
6196- 276 411 102	1 000
6197- 049 888 366	1 000
6201- 154 056 952	1 000
6201- 227 523 911	1 000
6201- 335 927 742	1 122
6202- 117 920 991	1 000
6202- 135 700 892	1 000
6203- 185 600 131	1 000
6203- 278 922 392	10 000
6203- 355 706 261	1 000
6203- 381 309 452	1 000
6204- 044 245 858	1 000
6204- 089 880 048	100 000
6204- 110 561 791	1 000
6204- 242 041 701	1 000
6204- 253 136 431	5 240
6204- 377 428 361	9 341
6204- 391 386 891	1 000
6205- 393 591 662	1 000
6206- 314 389 822	1 000

PLS FIGURE IV

A sample prize list for one of the banks in our data

Each prize list contains the account number of the winning account and the amount won.

Digitizing and processing the microfiche data. The microfiche cards were first scanned and then converted to text using the optical character recognition (OCR) software *ABBYY FlexiCapture*. Below, we describe the steps that we followed to process the OCR data from all volumes except the one covering the transition period. The goal of the processing was to (i) provide a reliable mapping of account numbers to PINs and to (ii) accurately infer the account balance of each account in each draw. The steps described below were taken separately for each of the 98 volumes.

Column 1. LÖPNUMMER, lottery ticket number. The first column of the microfiche images contains information about the lottery ticket numbers associated with each account. More precisely, the set of tickets associated with an account is given by the integer of an interval delimited by two values separated by a hyphen. We processed the OCR output for column 1 as follows:

1. We removed all non-numeric characters except for hyphens and dots.
2. We converted the interval into two numeric variables *Start* and *End* using the hyphen to separate the two numbers.
3. We created a variable called *andelar2suspect* to flag suspicious observations, initially setting its value to 2 for all observations.
4. For rows that had both a *Start* and *End* value, we sorted the data according to the order in which the OCR software read them. If correctly read, lottery ticket numbers should be sorted in increasing order.

5. Given the structure of the data, the value of *Start* for a row *k* should be greater than the value of *End* for row *k-1*, and likewise, the value of *End* for a row *k* should be smaller than the value of *Start* for row *k+1*. If both of these conditions were satisfied, we changed the value of *andelar2suspect* to 1.
6. If all data were correctly read and sorted around a row *k*, the value of *Start* for row *k* should be exactly equal to the value of *End+1* for row *k-1*, and the value of *End* for row *k* should be exactly equal to *Start+1* on row *k+1*. If both of these conditions were satisfied, we changed the value of *andelar2suspect* to 0.

The variable *andelar2suspect* will thus take the value 2 if we have reason to suspect an OCR reading error, 1 if we find some indications of an error, and 0 if we find no indications of an error.

7. For each row, we created a variable *andelar2* denoting the number of tickets held in the prize draw, by subtracting *End* from *Start* and adding 1.

Column 2. KONTONUMMER, account number. The second column of the microfiche images contains the account numbers. We structure them in a way that enables us to use them for quality control of the OCR output. The first four digits of any account number are determined by the bank where the account was opened, and the last digit is a checksum digit that must satisfy certain logical conditions. Finally, observations on a given microfiche image are always ordered by account number in ascending order. We processed the OCR output for column 2 as follows:

1. We removed all non-numeric characters except for hyphens.
2. We verified that the account number consisted of four digits and a hyphen followed by a string of digits. If not, it was set to missing.
3. We created a variable called *miss* to flag suspicious observations, initially setting its value to 0 for all observations. Next, we set this variable to l if an account number was missing following step 2, and equal to 2 if the first four digits were either less than 1101 or greater than 9002 (invalid codes). Finally, we set *miss* to 3 if the checksum digit was incorrect and equal to 4 if the account appeared twice in the data (accounts should only appear once in any given draw).
4. Because the data on the microfiche images should be sorted by account number, we used the following iterative procedure to flag suspicious observations:

   (a) For all accounts with *miss* equal to 0, we sorted the data by row number, separately for each image.

   (b) If an account number was smaller than the account number in the preceding row,

or larger than that in the following row, we set *miss* equal to 5.

(c) We excluded observations with *miss* equal to 5 and then re-sorted the data by row number. For the accounts that did not appear in the right order, we set *miss* to 6.

(d) We continued this process iteratively until no account numbers were incorrectly sorted on any given image.

5. Finally, we converted the account number to a numeric variable and stored it as *kontonummer*.

Note that our sorting algorithm uses information from one image at a time. The reason for this is that we encountered some instances where account numbers were not sorted in ascending order across images.

Column 3. SALDO, account balance. The third column of the microfiche images contains the account balance, from which the number of tickets can be inferred since accounts were allocated 1 ticket per 100 SEK in the balance. Consequently, we used this column to generate a second measure of the number of tickets held. We processed the OCR output for column 3 as follows:

1. We removed all non-numeric characters except for dots and commas.
2. For each row, we verified that the account balance was of the format "XXX,XX" or "XXX.XXX,XX". If not, the variable was set to missing.
3. Next, we converted the balance into a number by removing dots and commas. The number was then divided by 10,000 and rounded down to the nearest integer. We stored the output as *andelar3*.

To better see the logic behind the transformation in step 3, recall that each account was awarded 1 ticket per 100 SEK (subject to satisfying the minimum balance requirement of 800 SEK). In the microfiche images, account balances are reported down to two decimal points. Now consider an account with a balance of 2350.85 SEK. In step 3, this amount would first be transformed to the number 235085, thereafter divided by 10,000 before finally being rounded down to the nearest integer. The resulting number 23 corresponds to the number of lottery tickets in the draw.

Column 4. ANTAL ANDELAR, number of tickets. The fourth column of the microfiche images contains the number of tickets associated with each account. From the OCR output, we simply removed all non-numeric characters (typically symbols like ",","." and "-") and converted the result into a numeric variable called *andelar1*.

Column 5. PERSONNUMMER, PIN. The fifth and final column of the microfiche images contains the PIN of each account holder in the format "YYMMDD-XXXZ", where the last digit

Z is a checksum digit. We processed the OCR output for column 5 as follows:

1. We removed all non-numeric characters except for hyphens.
2. For each row, we verified that the PIN contained 8+4 characters separated by a hyphen, and that the date format as well as the checksum digit was correct. All incorrect PINs were set to missing.
3. We converted the output from step 2 to a numeric variable and added 9.9×10^{11} to the resulting number. We call this variable *personnummer*. The reason for implementing the 9.9×10^{11} transformation is the fact that the PINs of individuals born in the first decade of the 20th century begin with zeros which would otherwise be lost in processing.

Manually typed microfiche data. As an additional quality control measure, we manually entered the account information for four subsets of accounts that are of particular interest. The resulting data, based on visual inspection of high-resolution scans of the original microfiche cards, are of high quality, because the human eye is able to read the scans in almost every instance where the OCR processing is problematic.

Large winners. We defined winners of a prize of 100,000 SEK (at current prices) or more as large winners. In total, there were 4,866 accounts associated with large winners. Of these 3,705 belonged to the fiche period, 640 to the pre-fiche period and 521 to the post-fiche period. We processed these accounts as follows:

1. For the fiche period, we manually entered the PIN and account balance for the winning month. We were able to identify all but 18 of these winners from the scans. Further, reviewing all the 3,705 accounts, we discovered that three were mistakenly associated with different PINs in different months in the processed OCR data. These errors were corrected upon visual inspection.
2. For the pre-fiche period, we used the November 1986 volume (the earliest month for which we have account-level data) to manually record the PINs of winning accounts. This allowed us to determine the PIN in 559 out of 640 cases.
3. For the post-fiche period, we used the December 1994 volume (the latest month for which we have account-level data) to manually record the PINs of winning accounts. This allowed us to determine the PIN in 504 out of 521 cases.

Winners with missing PINs. We looked up all accounts with missing PINs that won a prize (including prizes below 100,000 SEK) during the fiche period as well as accounts with missing PINs found in in the digitized microfiche data that won outside the fiche period. In total, there were 708 such accounts and we were able to identify both the PIN and account balance in the month of a win for 508 of these.

Suspicious values of andelarX variables. Our data contain some 300 observations for which the values of the three *andelarX* variables coincided and were less than 8. Because the implied underlying balances should have made the accounts ineligible to participate in the lotteries, we suspected that these observations reflected OCR processing errors. To verify this suspicion, we took a random sample of 25% of these accounts and found that all of them did indeed have 8 or more lottery tickets. We corrected the balances for the randomly sampled accounts and set the variable to missing for the remaining 75%.

Large deposits. We also checked a few accounts for which we observed very large deposits. Large deposits were not allowed for most of the period of study, but appear to have taken place anyway in a small number of cases. These data contain 758 observations with manually typed number of lottery tickets, but fewer accounts since each account in the data appears more than once.

Whenever we manually entered a PIN, we verified that it was correctly formatted and corrected a small number of errors. In addition, we also double-checked a few cases in which the number of lottery tickets was typed incorrectly (detected based on discrepancies between OCR readings and manual entries). The resulting data file contains 6440 manually typed observations.

Merging fiche volumes. We merged the data from the 98 volumes into one large panel file comprising more than 180 million observations.

Digitizing and processing the prize list data. The prize lists were manually double-typed. After entry, we verified the checksum digits for all account numbers, as well as the number of prizes and sum of prizes for each draw and corrected a small number of data-entry errors. The resulting data file contains information on 555,202 unique prizes (in one case, the account number was missing, so we have 555,20l complete observations). Some accounts won multiple times (the maximum in the data is 30) and occasionally so within a single draw (maximum is 4).

Imputation and quality control. Below, we outline how we constructed the variables for the number of tickets held from the digitized and processed data. In addition, we provide evidence that our preferred measure – *imputeandelar* – is precise. The measurement errors for the number of tickets held is evaluated by computing the correlation between our variables and the values implied by actual account balances, which are available for the 9,611 observations that were either odds prize winners or entered manually (recall that for odds prize winners, the balance for the winning month can almost always be inferred from the size of the prize, which is observed on the prize lists).

1. First, we constructed a variable called *andelarsafe.* If all of our three measures – *andelar1, andelar2*, and *andelar3* – had the same value greater than 7, we set *andelarsafe* to this value, otherwise we set it to missing. We also set *andelarsafe* to missing if *andelar2suspect* was equal to 2. The resulting variable is defined for 3,575 out of the 9,611 observations with actual account balances, and regressing the actual number of lottery tickets held on *andelarsafe* gives an R^2 of 0.9990.
2. Next, we constructed a variable called *andelarsafe2* along the same principles. The only difference compared to *andelarsafe* is that we required that two or more, instead of all three, of the *andelarX* variables had to have the same value. The resulting variable is defined for 5,801 out of the 9,611 observations, and the R^2 is 0.9982.
3. Our third measure is called *andelarsafe3*. We set it equal to *andelarsafe2* if *andelarsafe2* was non-missing. If *andelarsafe2* was missing, we used the following algorithm to determine whether each of the variables *andelarX* takes on a "reasonable" value in the OCR output:
   (a) For each account in a particular draw, we defined the variables *pre_andelarsafe* and *post_andelarsafe*. We set the variable *pre_andelarsafe* to the non-missing value of *andelarsafe* observed most recently prior to the draw. If *andelarsafe* was always missing before the draw we set *pre_andelarsafe* to missing. We defined the variable *post_andelarsafe* analogously and classified the value of *andelarX* as reasonable if it was between *pre_andelarsafe* and *post_andelarsafe*.
   (b) If *andelarX* was lower than *pre_andelarsafe* and *post_andelarsafe*, we classified the value as reasonable only if the deposit that would be required to explain the difference between *andelarX* and *post_andelarsafe* was at most 1,000 SEK per month. This cut-off is not based on actual VK rules, but rather it was arrived at through a process of trial and error.
   (c) If *andelarX* was greater than *pre_andelarsafe* and *post_andelarsafe*, we classified the value as reasonable only if the deposit that would be required to explain the difference between *andelarX* and *pre_andelarsafe* was at most 1,000 SEK per month.

   After classifying each observation of *andelar1*, *andelar2*, and *andelar3* as either reasonable or unreasonable, we proceeded as follows:
   (d) If *andelarsafe3* was missing and *andelar3* was reasonable, we set *andelarsafe3* equal to *andelar3.*

(e) If *andelarsafe3* was missing after step (a) and *andelar2* was reasonable, we set *andelarsafe3* equal to *andelar2.*

(f) If *andelarsafe3* was missing after step (b) and *andelar1* was reasonable, we set *andelarsafe3* equal to *andelar1.*

The resulting variable is defined for 6,540 out of the 9,611 observations, and the R^2 is 0.9950.

As a final step, we imputed missing values. First, we calculated a linear inter- and extrapolation based on *andelarsafe3* with the restriction that the values were between 8 and 5,000, setting implied values outside this interval equal to the nearest boundary. We then used the estimated values to fill in empty observations so that we covered all months between the first and last time an account was observed in the data (either on microfiche cards or prize lists). In doing so, we primarily filled in months during the period December 1986 to December 1994, but we also imputed values outside of this period for accounts that won before or after the fiche period. The resulting variable is defined for 9,105 out of the 9,611 observations, and the R^2 is 0.9887.

Having verified that the imputation seemed to work well, we again imputed missing values, this time using both *andelarsafe3* and the 9,611 observations that were either entered manually or involved odds prize winners. We call the resulting variable *imputeandelar* and given that it has such a high precision we use it as our primary measure for number of tickets. In the paper we consequently refer to *imputeandelar* as 'number of tickets' or 'account balance'. Note, however, that while it is a precise measure, there is some uncertainty regarding when an account was opened and closed. Therefore, we do not use imputed values from outside the fiche period in our analyses.

Linking accounts to individuals. In this section, we describe how we linked accounts to the PINs of the account holders.

1. First, we merged the manually typed microfiche data with the microfiche panel data using manually typed PINs whenever available.
2. Next, we dropped PINs that could not be linked to an actual person (living or deceased). There were approximately 6,000 such observations in the data, presumably as a result of OCR reading errors.
3. For each account, we calculated the modal value of the PINs linked to the account, provided that (i) at least five successful PINs were matched to the account and (ii) the fraction of matched PINs that were equal to the modal PIN was at least 80%. Next, we dropped non-modal PINs that are likely to reflect OCR reading errors. If for an account

the PIN was not equal to the mode in a particular month, but was successfully linked to the modal PIN both in the observation prior to and after the month in question, we set the non-modal PIN to missing. Likewise, if two non-modal but adjacent PINs were preceded and followed by the modal PIN, we set the non-modal PINs to missing. Next, we identified all cases in which the first successfully matched PIN was non-modal and set it to missing provided that the modal PIN was observed at least once within three months. We then proceeded analogously for the last successfully matched PINs.

4. Next, we deleted all accounts that were not associated with a unique PIN after completion of steps 1 through 3. In this step we eliminated roughly l% to 2% of the winning accounts.
5. Finally, we replaced all remaining missing observations by the unique PINs associated with each account.

To assess how reliably we mapped accounts to individuals, we repeated the above procedure, but without using the information about manually typed PINs in step 1. Comparing the owners thus identified to the actual owners for the 4,854 accounts with manually typed PINs we found that these were identical in all but three cases.

Unfortunately, even so, occasional errors may remain. For example, it is possible that some accounts have switched owners and that we only observe the PIN of one of the owners once. If so, the algorithm will incorrectly assign these accounts to one of the PINs for the whole period. However, this is likely to be rare. In the full panel data file, which consists of some 184 million observations, we find only about 30,000 instances when we set the first or last appearing non-modal PIN to missing (as described in step 3 above).

An indication that some misclassification remains in our panel data is that 0.1% of individuals are associated with more than five accounts. Most notably, one individual is identified as the owner of 36 accounts, which clearly is unreasonable given that VK rules only allowed one account per bank and that there were less than 36 banks. Restricting attention to accounts successfully matched to a PIN at least 10 times, the maximum number of accounts observed for an individual is 24. Most likely this individual has a PIN that corresponds to rare but systematic mistakes the OCR software makes (e.g., “11” appear to be overrepresented among pre-processed PINs).

III.D. Additional Quality Control

PLS Table IV demonstrates that during the fiche period our linking algorithm was

successful in identifying the PIN of the owner for 98% to 99% of the prize-winning accounts. Column 1 shows the number of prizes that should have been awarded according to the prize plans and column 2 the number of prizes entered manually from the prize lists, whereas column 3 lists the number of prizes belonging to accounts that could reliably be linked to PINs. PLS Table V reports the same information for all fixed prizes outside the fiche period. For this period, we restrict attention to fixed prizes because the prize amounts for odds prizes are functions of the account balances that we typically do not observe. Nonetheless, surprisingly many winning accounts could be linked to individuals in the post-fiche period. For example, 95% of the prizes awarded in 2002 could be linked to PINs using information from the fiche period. However, for the pre-fiche period fewer winning accounts could be linked to individuals, which is one reason for why we do not include pre-fiche period winners in our analyses.

PLS TABLE IV

NUMBER OF PRIZES SUCCESSFULLY MATCHED TO PINS (FICHE PERIOD)

	# Prizes According to Prize Plan	# Prizes Manually Entered from Prize Lists	# Linked to PIN
	(1)	(2)	(3)
Dec 1986 to Dec 1987	29,240	29,240	28,858
1988	27,051	27,051	26,712
1989	27,022	27,022	26,727
1990	39,008	39,008	38,614
1991	41,351	41,351	40,914
1992	41,177	41,177	40,646
1993	40,028	40,028	39,489
1994	24,205	24,205	23,816

Notes. Column 1 shows the number of yearly prizes according to the prize plans and column 2 the corresponding number of prizes in the manually entered prize list data. Column 3 shows the number of prizes that could be reliably linked to a PIN.

III.E. Constructing the Final Estimation Sample

We impose two restrictions on the PLS data when constructing the final estimation sample. First, we drop all prizes from the pre-fiche period (1979-1986). The reason for this is that we are only able to identify the PINs of winners who kept their accounts until the beginning of the fiche period (December 1986), and prize amount may interact with unobserved characteristics

in determining the likelihood that a winning account is closed down. Such interactions would imply a relationship between predetermined characteristics and prize amount in the sample of accounts that win before the fiche period and whose owners we are able to identify.

PLS TABLE V

NUMBER OF FIXED PRIZES SUCCESSFULLY MATCHED TO PINS (OUTSIDE FICHE PERIOD)

	# Fixed Prizes According to Prize Plan (1)	# Fixed Prizes Manually Entered from Prize Lists (2)	# Linked to PIN (3)	Comment (4)
Old PLS Accounts				
1979	N/A	7,958	4,402	
1980	N/A	8,964	5,529	
1981	N/A	11,645	7,758	
1982	N/A	14,364	10,341	
1983	N/A	12,025	9,298	
1984	N/A	22,001	17,879	
1985	N/A	26,655	23,013	
Transition Period PLS				
1985	10	10	9	
1986	10	10	10	
Post Fiche Period PLS				
1995	18,224	18,225	17,775	One 1,000 SEK odds prize coded as fixed prize
1996	44,372	42,879	41,348	Prize lists missing for a few banks in some draws and one 1,000 SEK odds prize codes as a fixed prize
1997	36,299	35,758	33,917	Prize list missing for one bank
1998	23,307	23,193	21,652	Prize list missing for one bank in March 1988
1999	N/A	3,845	3,712	Prize lists from March, June and December missing
2000	N/A	14,050	13,473	
2001	N/A	14,257	13,610	
2002	N/A	12,944	12,291	
2003	N/A	9,238	8,710	

Notes. Column 1 shows the number of prizes according to the prize plan (not available for all years) and column 2 shows the corresponding number of prizes in the manually entered prize list data. Column 3 shows the number of prizes that could be reliably linked to a PIN. Column 4 reports which data that is known to be missing.

Second, we restrict the sample to prize-winning accounts. We impose this restriction since

our full panel from the fiche period, which contains winning and non-winning accounts, provides some indications that the owners of missing accounts have systematically different pre-win characteristics from the owners of non-missing accounts. These differences arise because our algorithm reliably identifies around 98.7% of the PINs of the owners of prize-winning accounts compared to approximately 86.8% of all accounts (see PLS Table VI).

PLS TABLE VI

NUMBER OF WINNING AND NON-WINNING ACCOUNTS FOR SPECIFIC DRAWS

	# Winning accounts		Identified winning accounts (%)	# Non-winning accounts		Identified non-winning accounts (%)
	Actual	Data		Actual	Data	
Jan 1987	2,238	2,201	98.4	1,895,973	1,630,797	86.0
Jan 1988	2,247	2,211	98.4	2,293,943	2,008,961	87.6
Jan 1989	2,246	2,218	98.8	2,415,611	2,140,537	88.6
Jan 1990	2,240	2,216	98.9	2,458,334	2,166,890	88.1
Dec 1990	3,444	3,405	98.9	2,475,837	2,146,249	86.7
Jan 1992	3,432	3,385	98.6	2,427,801	2,082,075	85.8
Jan 1993	3,356	3,316	98.8	2,356,883	1,993,532	84.6

Notes. This table shows the actual number of winning and non-winning accounts as well as the number of accounts for which we managed to both successfully identify PINs and impute the number of lottery tickets. The information on the actual number of accounts is based on documents provided by Johan Hansing.

An account may be missing in the final data for two major reasons. Both are relevant for understanding why our population of prize-winning accounts is essentially attrition free. First, the algorithm that we use is systematically biased in the direction of assigning a non-winning account a later start date than its true start date. For example, consider a non-winning account opened in February 1993 but not identified by the OCR software until June 1994 (because of idiosyncratic differences in the quality of microfiche cards over time). This account will first appear in our data in June 1994. Because the information about prize-winning accounts is entered manually from the prize lists, we have the account numbers of all winning accounts in every draw. The algorithm that we use to construct the final panel takes advantage of the fact that if an account won a prize at time *t*, it must have existed at time *t*. For example, consider another account opened in February 1993 that was not identified by the OCR software until June 1994, but that won a prize in March 1993. Our algorithm will then identify this account as opened in March 1993.

The upshot is that an identification strategy based on matching winning accounts to non-winning accounts with identical balances will compare virtually all of the winning accounts to a

non-random subset of non-winning accounts. In particular, the non-winning accounts will be disproportionately composed of accounts that have been in existence for longer. Comparing winning and non-winning accounts with identical balances does therefore not guarantee conditional independence of the amount won. In the example in the previous paragraph, the winning account that was opened in February 1993 will not be matched with the non-winning account that was opened in February 1993. Instead, it will be matched with non-winning accounts that were on average opened at an earlier date.

Second, our algorithm may fail to reliably match an account number to a PIN. The matching rate is higher for prize-winning accounts because on average, a winning account has participated in a larger number of draws. Therefore, a winning account appears on more microfiche cards, increasing the probability that the OCR software is able to identify the unique mapping from account number to PIN. In addition, prize-winning account numbers are entered manually from prize lists and are therefore observed without error.

After dropping accounts that never won as well as accounts that won in the pre-fiche period (prior to December 1986), the resulting data set contains 445,656 prizes won by 346,385 unique accounts. Of all winning accounts during and after the fiche period, we were able to match 336,561 to individuals following the procedure outlined above. These accounts were owned by 330,369 unique individuals, with 324,464 individuals being associated with one winning account, 5,646 with two, 233 with three, 24 with four, and 2 with five. In total, after restricting the sample to accounts that can be reliably matched to individuals, 431,496 month by prize-winning account combinations remain in our data set. Out of these, 204 involve an individual winning on two different accounts in the same draw, all of which are excluded from our analysis. In addition, we drop odds prizes won in the post-fiche period and odds prizes won by accounts for which the number of lottery tickets could not be reliably imputed. Finally, we drop individuals with PINs that did not pass the quality control outlined in Section IX.A. The resulting final data set consists of 420,671 month by prize-winning account combinations. In our adult analyzes, we also omit 6,348 observations because the winner was below 18 years of age at the time of the draw. We call the data restricted to winners in the fiche and post-fiche periods ‘Prize Data’.

III.F. Constructing the PLS Cells (Adult Analyses)

In this section, we motivate our method for defining cells in our PLS sample. To simplify the exposition, we illustrate each step with data from two hypothetical PLS draws, denoted 1

and 2. The hypothetical data (shown in PLS Table VII) have exactly the same structure as the actual PLS data, but we have purposely selected values of the key variables that allow us to illustrate subtleties that arise in defining the cells.

PLS TABLE VII

HYPOTHETICAL PRIZE DATA USED TO ILLUSTRATE PLS MATCHING

AccountID	DrawID	Prize	Prize Type	PersonID	Balance	Year
101	1	8000	10-Odds	201	8	1987
101	1	80000	100-Odds	201	8	1987
203	1	1000	Fixed	202	12	1987
347	1	57000	10-Odds	303	57	1987
1019	1	2800	1-Odds	543	28	1987
1321	1	2600	1-Odds	817	26	1987
1922	1	58000	10-Odds	890	58	1987
2081	1	100000	Fixed	902	24	1987
2321	1	1000	Fixed	1261	56	1987
2689	1	5700	Fixed	1344	57	1987
2689	1	10000	Fixed	1344	57	1987
2776	1	1000	Fixed	1526	108	1987
2776	1	1000	Fixed	1526	108	1987
…						
203	2	1000	Fixed	202	8	1988
889	2	1000	Fixed	900	28	1988
892	2	26000	10-Odds	1079	26	1988
1579	2	100000	Fixed	1100	134	1988
1992	2	1000	Fixed	1390	26	1988

In the data, each row is a prize won in a particular draw. For each prize, the data contains the identifier of the winning account (*Account ID*) and the identifier of the account's owner (*PersonID*). There is a unique mapping from the *Account ID* to *PersonID*. However, since one person can hold multiple accounts, the converse is not true. Conceptually, we can think of our matching procedure as consisting of three steps. We describe each in turn.

Step 1. Generate cells for odds prize winners. We begin with the Prize Data and generate the variable *Pnum* which is defined as the total number of prizes won by the account holder in any particular draw. Next, we drop all individuals for which *Pnum* exceeds 1, leaving us with a data set where each row has a unique combination of *DrawID* and *PersonID*. The resulting hypothetical data set is shown in PLS Table VIII.

For all accounts in the data set, we discretize the imputed balance variable in increments of (i) 1 for account balances between 8 and 10 (ii) 2 for account balances between 10 and 200 (iii)

5 for balances between 200 and 400 tickets, and (iv) 50 for account balances exceeding 400. The resulting variable *Disc* takes the value 1 if the account balance is in the interval [8,9), 2 if it is in [9,10), 3 if it is in [10,12) and so on up to 230 in the highest interval [4950,5000).

PLS TABLE VIII

Dropping Winners of Multiple Prizes

AccountID	DrawID	Prize	Prize Type	PersonID	Balance	Year	Pnum
~~101~~	~~1~~	~~8000~~	~~10-Odds~~	~~201~~	~~8~~	~~1987~~	~~2~~
~~101~~	~~1~~	~~80000~~	~~100-Odds~~	~~201~~	~~8~~	~~1987~~	~~2~~
203	1	1000	Fixed	202	12	1987	1
347	1	57000	10-Odds	303	57	1987	1
1019	1	2800	1-Odds	543	28	1987	1
1321	1	2600	1-Odds	817	26	1987	1
1922	1	58000	10-Odds	890	58	1987	1
2081	1	100000	Fixed	902	24	1987	1
2321	1	1000	Fixed	1261	56	1987	1
~~2689~~	~~1~~	~~5700~~	~~Fixed~~	~~1344~~	~~57~~	~~1987~~	~~2~~
~~2689~~	~~1~~	~~10000~~	~~Fixed~~	~~1344~~	~~57~~	~~1987~~	~~2~~
~~2776~~	~~1~~	~~1000~~	~~Fixed~~	~~1526~~	~~108~~	~~1987~~	~~2~~
~~2776~~	~~1~~	~~1000~~	~~Fixed~~	~~1526~~	~~108~~	~~1987~~	~~2~~
…							
203	2	1000	Fixed	202	8	1988	1
889	2	1000	Fixed	900	28	1988	1
892	2	26000	10-Odds	1079	26	1988	1
1579	2	100000	Fixed	1100	134	1988	1
1992	2	1000	Fixed	1390	26	1988	1

We then drop from the sample the following:

1. All fixed prizes whose value of the *Disc* variable does not exactly match that of at least one odds prize winning account in the month of the draw, i.e. accounts 203 and 2081 in draw 1 and accounts 203, 889 and 1579 in draw 2.
2. Odds prizes with a unique *Disc* value in the winning draw, i.e. accounts 1019, 1321 and 1922.

Next, we generate the variable *Cell* and assign to it a unique value for each combination of *Disc* and *DrawID*. To reduce the number of cell fixed effects to be estimated, we further restrict the sample to cells with at least five observations. We call the resulting dataset ‘Odds Prize Data’.

In the hypothetical data, account 347 wins an odds prize in draw 1 and is matched to account 2321. We assign the owners of these accounts to the first cell (*Cell* = 1). Similarly, in

draw 2, account 892 wins an odds prize and is matched to account 1992, and we assign the owners to the second cell (*Cell* = 2). PLS Table IX shows the construction of the odds prize cells for our hypothetical data. Finally, in the construction of cells for the odds prizes, we omit a few cells for which there were no suitable controls as well as cells where the total sum of prizes was below 100,000 SEK. The latter restriction is implemented in order to reduce the total number of cells.

PLS TABLE IX

CONSTRUCTING ODDS PRIZE CELLS

AccountID	DrawID	Prize	Prize Type	PersonID	Balance	Year	Pnum	Disc	Cell
~~203~~	~~1~~	~~1000~~	~~Fixed~~	~~202~~	~~12~~	~~1987~~	~~1~~	~~4~~	
347	1	57000	10-Odds	303	57	1987	1	26	1
~~1019~~	~~1~~	~~2800~~	~~1-Odds~~	~~543~~	~~28~~	~~1987~~	~~1~~	~~12~~	
~~1321~~	~~1~~	~~2600~~	~~1-Odds~~	~~817~~	~~26~~	~~1987~~	~~1~~	~~11~~	
~~1922~~	~~1~~	~~58000~~	~~10-Odds~~	~~890~~	~~58~~	~~1987~~	~~1~~	~~27~~	
~~2081~~	~~1~~	~~100000~~	~~Fixed~~	~~902~~	~~24~~	~~1987~~	~~1~~	~~10~~	
2321	1	1000	Fixed	1261	56	1987	1	26	1
…									
~~203~~	~~2~~	~~800~~	~~Fixed~~	~~202~~	~~8~~	~~1988~~	~~1~~	~~1~~	
~~889~~	~~2~~	~~1000~~	~~Fixed~~	~~900~~	~~28~~	~~1988~~	~~1~~	~~12~~	
892	2	26000	10-Odds	1079	26	1988	1	11	2
~~1579~~	~~2~~	~~100000~~	~~Fixed~~	~~1100~~	~~134~~	~~1988~~	~~1~~	~~65~~	
1992	2	1000	Fixed	1390	26	1988	1	11	2

Step 2. Generate Cells for Fixed Effects. We begin with the Prize Data and drop all odds prizes. We then drop all rows in the data with *DrawID* by *PersonID* combinations that appear in Odds Prize Data. Since prize amount is independent of the account balance conditional on the number of fixed prizes in a given draw, this does not introduce any bias in the estimated effect of lottery wealth on outcomes. However, by dropping these observations, we are restricting the sample to fixed prizes that were not matched to an odds prize in step 1. As shown in PLS Table X this implies that we drop account 2321 from draw 1 and 1992 from draw 2.

PLS TABLE X

DROPPING ODDS PRIZES

AccountID	DrawID	Prize	Prize Type	PersonID	Balance	Year
~~101~~	~~1~~	~~8000~~	~~10-Odds~~	~~201~~	~~8~~	~~1987~~
~~101~~	~~1~~	~~80000~~	~~100-Odds~~	~~201~~	~~8~~	~~1987~~
203	1	1000	Fixed	202	12	1987
~~347~~	~~1~~	~~57000~~	~~10-Odds~~	~~303~~	~~57~~	~~1987~~
~~1019~~	~~1~~	~~2800~~	~~1-Odds~~	~~543~~	~~28~~	~~1987~~
~~1321~~	~~1~~	~~2600~~	~~1-Odds~~	~~817~~	~~26~~	~~1987~~
~~1922~~	~~1~~	~~58000~~	~~10-Odds~~	~~890~~	~~58~~	~~1987~~
2081	1	100000	Fixed	902	24	1987
~~2321~~	~~1~~	~~1000~~	~~Fixed~~	~~1261~~	~~56~~	~~1987~~
2689	1	5700	Fixed	1344	57	1987
2689	1	10000	Fixed	1344	57	1987
2776	1	1000	Fixed	1526	108	1987
2776	1	1000	Fixed	1526	108	1987
…						
203	2	1000	Fixed	202	8	1988
889	2	1000	Fixed	900	28	1988
~~892~~	~~2~~	~~26000~~	~~10-Odds~~	~~1079~~	~~26~~	~~1988~~
1579	2	100000	Fixed	1100	134	1988
~~1992~~	~~2~~	~~1000~~	~~Fixed~~	~~1390~~	~~26~~	~~1988~~

In the remaining sample of fixed prizes, we define *Pnum*, the total number of fixed prizes won by the account holder in the draw, and *Sum_FP*, the total value of these prizes. Next, we drop observations so that we only keep one account per individual per draw, implying that winners who won multiple fixed prizes in a given draw only appear once. PLS Table XI shows that this implies that in our hypothetical data we drop the second observation for *PersonID* 1344 from the first draw as well as the equivalent observation for *PersonID* 1526. Finally, we generate a variable called *Cell* and assign it a unique value for each unique combination of *DrawID* and *Pnum.* The values of the *Cell* variable are selected so that there is no overlap with the cells defined in step 1. We call the resulting dataset ‘Fixed Prize Data’.

PLS TABLE XI

FIXED PRIZE CELLS

AccountID	DrawID	Prize	Prize Type	PersonID	Balance	Year	Pnum	PSum	Cell
203	1	1000	Fixed	202	12	1987	1	1000	10001
2081	1	100000	Fixed	902	24	1987	1	100000	10001
2689	1	1000	Fixed	1344	57	1987	2	11000	10003
~~2689~~	~~1~~	~~10000~~	~~Fixed~~	~~1344~~	~~57~~	~~1987~~	~~2~~	~~11000~~	
2776	1	1000	Fixed	1526	108	1987	2	2000	10003
~~2776~~	~~1~~	~~1000~~	~~Fixed~~	~~1526~~	~~108~~	~~1987~~	~~2~~	~~2000~~	
…									
203	2	1000	Fixed	202	8	1988	1	1000	10002
889	2	1000	Fixed	900	28	1988	1	1000	10002
1579	2	100000	Fixed	1100	134	1988	1	100000	10002

Step 3. Combining the Samples. In the final step we combine the Fixed Prize Data with the Odds Prize Data. We then defined a unique identifier across rows which we call *UniqueID* as well as a variable *Win* that is equal to *Prize* if the observation is from the Odds Prize Data and *PSum* otherwise. The *UniqueID* is convenient because it allows us to manage account holders who won on more than one occasion. This gives us the final estimation sample (which can easily be converted to panel format and merged to register data using *PersonID* and *Year*). The final hypothetical data set is shown in PLS Table XII.

PLS TABLE XII

THE FINAL SAMPLE

AccountD	DrawID	Prize	Prize Type	PersonID	Balance	Year	Pnum	PSum	Cell	Win	Unique ID
347	1	57000	10-Odds	303	57	1987	1		1	57000	A303
2321	1	1000	Fixed	1261	56	1987	1		1	1000	A1261
892	2	26000	10-Odds	1079	26	1988	1		2	26000	B1079
1992	2	1000	Fixed	1390	26	1988	1		2	1000	B1390
203	1	1000	Fixed	202	12	1987	1	1000	10001	1000	A202
2081	1	100000	Fixed	902	24	1987	1	100000	10001	100000	A902
2689	1	1000	Fixed	1344	57	1987	2	11000	10003	11000	A1344
2776	1	1000	Fixed	1526	108	1987	2	2000	10003	2000	A1526
203	2	1000	Fixed	202	8	1988	1	1000	10002	1000	B202
889	2	1000	Fixed	900	28	1988	1	1000	10002	1000	B900
1579	2	100000	Fixed	1100	134	1988	1	100000	10002	100000	B1100

III.G. Constructing the PLS Cells (Child Analyses)

The construction of the cells for the child analyses follows the same procedure as for the adult analyses, with three exceptions. First, we drop all odds prize winners from the estimation sample. Second, because we drop all odds prize winners, there is no need to use fixed prize winners as matched controls. The fixed prize cells are thus based on the full set of fixed prize winners. Third, we construct these fixed prize cells using the same procedure as for the adult sample (see step 2 above), except that we also condition on the winning parent's number of pre-lottery children at the time of the win. In other words, if we let *Children* denote the number of pre-lottery children, then all children of winners with the same unique combination of *Children*, *DrawID*, and *Pnum* will form one cell.

III.H. Estimation

As PLS Table XII shows, the data in our final estimation sample are cross-sectional in nature. In our empirical analyses we normalize the time of the lottery to 0 and estimate regression equations of the form

$$Outcome_i = \beta_0 + \beta_1 \cdot Win_{i,0} + X_i \cdot \beta_2 + Z_{i,-1} \cdot \gamma + \varepsilon_i,$$

where *i* indexes individuals by their unique identifiers *UniqueID*, X_i is the vector of cell fixed effects and $Z_{i,-1}$ is a vector of controls. The key assumption needed for β_1 to have a causal interpretation is that *Win* is uncorrelated with ε_i conditional on the cell fixed effects. Because we condition on the cell fixed effects, all of our identifying variation is coming from within-cell differences in the total prize amount won.

It is worth noting two things about the controls. First, we only include them in order to absorb more of the variance of the residual and to improve the precision of our estimates. The coefficients in the vector γ are thus not of particular interest to us. Second, the covariates are always measured the year prior to the year in which the draw took place. In other words, if for example an individual wins the lottery in 1986, the covariates in the $Z_{i,-1}$ vector are measured in 1985.

One unusual feature of our data is that individuals who win in more than one draw appear multiple times. As a consequence, the error terms ε_i are correlated within individuals as well as across cells. In addition, some outcome variables and baseline controls are time varying, and the covariates will therefore not typically have the same values across observations that share a *PersonID.* For example, suppose we match the final estimation sample to the *Cause of Death Register* and find that the individual who appears twice in the final hypothetical sample (PLS

Table XII, *PersonID* = 203) died in 1989. Suppose next that we are interested in one-year mortality (1 if deceased). The outcome variable will take on value 1 for the 1988 observation (*UniqueID* = B202) and 0 for the 1987 observation (*UniqueID* = A202). Moreover, in the hypothetical mortality regressions the controls for observation B202 are measured in 1987 and the controls for observation A202 are measured in 1986. Therefore, "duplicating" such observations is necessary if one wishes to retain them in the data, since simply controlling for two cell fixed effects is not feasible if an individual appears in two cells.

In principle, we could have avoided all duplication by restricting the sample only to individuals who never won a prize in the past. This would however have come at the cost of lower precision since it would imply using less of the variation in lottery prizes. Therefore, we do not limit our main estimation sample to first-time winners. Importantly, duplicating observations does not lead to a violation of exogeneity. To see why, note that, conditional on the account balance, the amount won is completely uncorrelated with whether or not an individual also won an additional prize in a previous draw. In other words, in a given cell, winners of large amounts are as likely as winners of small amounts to have won a prize in the same, or in a previous, draw.

III.I. Identification

This section provides a more formal argument in support of our identification strategy for fixed prizes. The critical identifying assumption that we make in studying the PLS sample is that conditional on the number of fixed prizes won, the total sum of fixed prizes won is independent of an individual's account balance (and therefore, also plausibly independent of any other predetermined characteristics).

To see why this assumption should hold under the rules of the PLS lottery, consider an account holder with a balance of t tickets who participates in a PLS draw with a total number of T lottery tickets. The lottery this account holder participates in is an experiment in which t objects are randomly sampled, without replacement, from a population with a total of T objects: n_1 prizes of fixed size p_1, n_2 prizes of fixed size p_2, … , n_{m-1} prizes of fixed size p_{m-1} and finally, n_m "prizes" of size 0 (we think of non-winning tickets as having won a prize of 0).

Let Y_1 denote the number of prizes of size p_1 won by the account holder, Y_2 the number of prizes of size p_2, and so on up to Y_m. The random vector $Y = (Y_1,\ldots,Y_m)$ then follows a multivariate hypergeometric distribution with probability density function

$$P(Y_1 = y_1, \ldots, Y_m = y_m) = \frac{\prod_{i=1}^{m} \binom{n_i}{y_i}}{\binom{T}{t}},$$

where $\sum_{i=1}^{m} y_i = t$ and $y_i \leq n_i$ for all i. Let the random variable X denote the number of (non-zero) prizes won. Conditioning on the account with t tickets having won exactly $X = x$ (non-zero) prizes is equivalent to setting $Y_m = t - x$ (i.e. the account "won" $t - x$ zero prizes) and the conditional probability density function is given by

$$P(Y_1 = y_1, \ldots, Y_{m-1} = y_{m-1} \mid Y_m = t - x) = \frac{P(Y_1 = y_1, \ldots, Y_{m-1} = y_{m-1}, Y_m = t - x)}{P(Y_m = t - x)},$$

where $P(Y_m = t - x)$ is the marginal probability density function for the zero prizes, which follows a hypergeometric distribution. We therefore have

$$P(Y_1 = y_1, \ldots, Y_{m-1} = y_{m-1} \mid Y_m = t - x) = \frac{\frac{\prod_{i=1}^{m-1} \binom{n_i}{y_i} \binom{n_m}{t-x}}{\binom{T}{t}}}{\frac{\binom{n_m}{t-x}\binom{T-n_m}{t-(t-x)}}{\binom{T}{t}}} = \frac{\prod_{i=1}^{m-1} \binom{n_i}{y_i}}{\binom{T-n_m}{x}}.$$

Note that this conditional distribution is independent of the number of tickets held by the account holder. The sum of fixed prizes won is therefore a random variable $Tot = Y' \cdot p$ and is hence also conditionally independent of t.

For example, if an account holder wins one prize, the probability that this prize is of size p_1 is

$$P(Y_1 = 1, Y_2 = 0, Y_3 = 0, \ldots, Y_{m-1} = 0 \mid Y_m = t - 1) = \frac{\binom{n_i}{1}}{\binom{T-n_m}{1}} = \frac{n_i}{T-n_m},$$

i.e., simply the number of prizes of size p_1 divided by the total number of (non-zero) prizes.

However, the argument fails to hold for odds prize winners because the prize amounts p are *not* independent of t.

IV. The Triss Scratch-Off Sample

This section describes the Triss scratch-off lottery data that were used to construct the Triss-Lumpsum and Triss-Monthly samples. We begin with a brief background on Triss.

IV.A. Background on the Triss Data

Triss is a popular scratch-off lottery run by the Swedish state-owned gaming operator Svenska Spel. Lottery tickets can be bought in convenience stores throughout the country and participants can win a large number of different prizes. In our paper we focus on two prize categories, TV-Triss and Klöver. The primary reason for doing so is that we are able to identify virtually all winners of these prizes from publicly available sources.

In particular, winners of TV-Triss and Klöver are typically invited to participate in prize draws on a popular morning TV show ("TV4 Morgon"). However, there are more winners than slots on the show, so not all winners are allowed to take part. At the show, TV-Triss winners draw a new scratch-off lottery ticket from a stack of 90 or 100. All tickets look identical, so there is no strategic element involved in choosing. The ticket drawn determines a lump sum prize amount, ranging from 50,000 to 5 million SEK. Klöver winners, who are paid in installments, instead draw a ticket determining a monthly prize ranging from 10,000 to 50,000 SEK, before drawing a second ticket determining a duration of 10 to 50 years for the monthly installments. This two-stage process ensures that the size and duration of installments are independent. In both lotteries, prizes are net of taxes, and for Klöver the monthly installments are indexed by inflation. If a Klöver winner dies before receiving his or her last payment, the money owed is transferred to the deceased's estate. Henceforth, and in the paper, we call the TV-Triss sample 'Triss-Lumpsum' and the Klöver sample 'Triss-Monthly'.

AV.B. Constructing the Final Estimation Sample

Using information from the Swedish Gambling Authority (SGA) we identified the changes to Triss prize plans that were registered between 1994 and 2011. For Triss-Lumpsum the prize plan only changed once, on July 19, 2000, whereas there were five changes for Triss-Monthly: August 14, 2001; November 15, 2006; March 20, 2007; December 15, 2007; and April 27, 2009. All of these changes were minor. But, there is some uncertainty regarding the completeness of the information from SGA.

Winners of Triss-Lumpsum or Triss-Monthly prizes who do not wish to participate in the TV show have several options. First, they can participate in the draw but not have their

participation broadcasted. Second, they can choose to send a delegate – typically a family member – to the TV show. Finally, winners can instead of participating in the draw accept a lump sum payment of 50,000 (Triss-Lumpsum) or 500,000 SEK (Triss-Monthly). Since the expected values from participating are higher than the fixed lump sum payment, this option is rarely exercised. For example, the 50,000 SEK offered to Triss-Lumpsum winners is equal to the worst possible outcome in the Triss-Lumpsum draw.

Typically, the televised draw takes two to three minutes, during which the host makes small talk with the participant (either the winner or a delegate). In the TV show, the winner's name and hometown is always disclosed, and while the topic of the conversation varies, it often focuses on family, interests, and plans for spending the prize money.

The original data. Svenska Spel supplied us with a spreadsheet containing information on 5,057 Triss-Lumpsum prizes awarded between March 1994 and October 2011 and 824 Triss-Monthly prizes awarded between October 1997 and October 2011 (this lottery only began in 1997). The spreadsheet contains information on televised as well as untelevised lottery draws, and each entry represents a unique draw. For each entry the spreadsheet lists a "participant," typically the owner of the winning ticket, but if a delegate was sent its name is usually listed instead.

With a few rare exceptions, the spreadsheet contains the participants' name, address, phone number, and age. For each draw, it also lists the date, type of lottery, as well as the prize won. Finally, the spreadsheet contains a column with supplementary information about participants entered by Svenska Spel. For instance, if the participant is a delegate, there is often a note outlining the delegate's relationship to the actual winner. Likewise, there is sometimes information about shared ownership, and from 2000 and on, there is typically some information on how winners plan to spend their prize money.

Step A. Prize data: cleaning and descriptives. Because opting out is not attractive, this rarely happens. In total, 10 winners opted out of the lottery. Of these, only two were Triss-Lumpsum winners who preferred the 50,000 SEK lump sum to the lottery, and the remaining eight were Triss-Monthly winners preferring the 500,000 SEK lump sum. These observations are excluded from the lottery data, leaving 5,871 prizes. In an additional 125 cases, the information on prize amount was either missing or inconsistent with the prize plan. Using televised draws from the archives of the National Library of Sweden, a research assistant was able to recover the correct prize amount for 111 out of these. We drop the 14 observations for which we were unable to verify the prize amount, leaving us with a sample of 5,857 prizes. Out of these, 5,045 were Triss-Lumpsum and 812 Triss-Monthly. The prizes were awarded on 5,276

different dates, implying that some 90% (5,276/5,857) of winners participated in the TV show.

The distribution of nominal Triss-Lumpsum prizes is shown in Triss Table I. As can be seen both the median and modal value is 100,000 SEK. The average nominal prize is 277,037 SEK, and the average real prize is 307,612 SEK (in 2010 prices), giving a total real prize amount of 1.55 billion SEK.

TRISS TABLE I

DISTRIBUTION OF NOMINAL LUMPSUM PRIZES

Amount (SEK)	Number of prizes	Share (%)
5,000,000	57	1.13
4,000,000	23	0.46
3,000,000	46	0.91
2,000,000	67	1.33
1,000,000	78	1.55
800,000	4	0.08
500,000	233	4.62
300,000	409	8.11
200,000	837	16.59
100,000	1, 926	38.18
50,000	1,365	27.06
TOTAL	5,045	100.00

Triss Table II shows the distribution of monthly installments for Triss-Monthly. The modal installment is 10,000, the median 15,000, and the average 16,385 SEK. However, note that the 50,000 SEK installment was only introduced in 2007.

TRISS Table II

DISTRIBUTION OF MONTHLY INSTALLMENTS

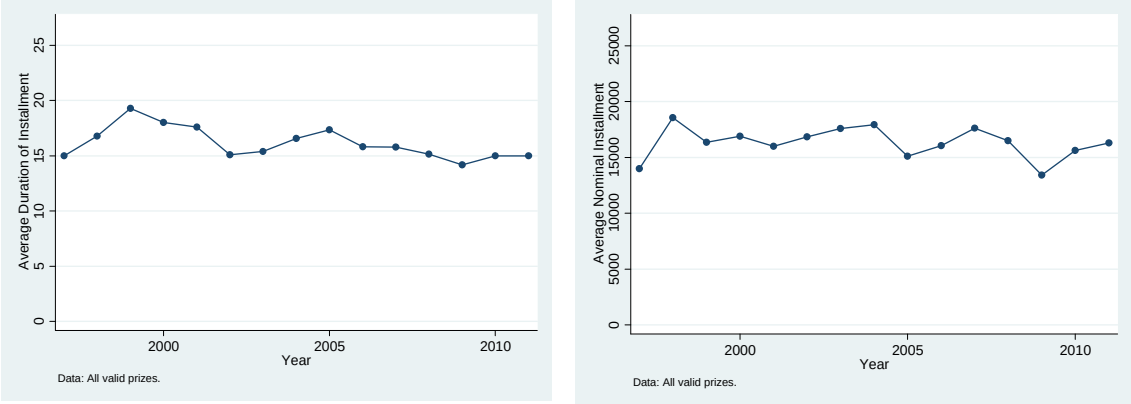
Amount (SEK)	Number of prizes	Share (%)
50,000	24	2.96
25,000	117	14.41
20,000	126	15.52
15,000	242	29.80
10,000	303	37.32
TOTAL	812	100.00

Triss Table III shows the distribution of the duration of Triss-Monthly installments. The modal duration is 10, the median 15, and the average 16.09 years. However, note that the 50 years duration was only introduced in 2010.

TRISS TABLE III

DISTRIBUTION OF DURATION OF INSTALLMENTS

Duration (years)	Number of prizes	Share (%)
50	1	0.12
25	137	16.87
20	198	24.38
15	174	21.43
10	302	37.19
TOTAL	812	100.00

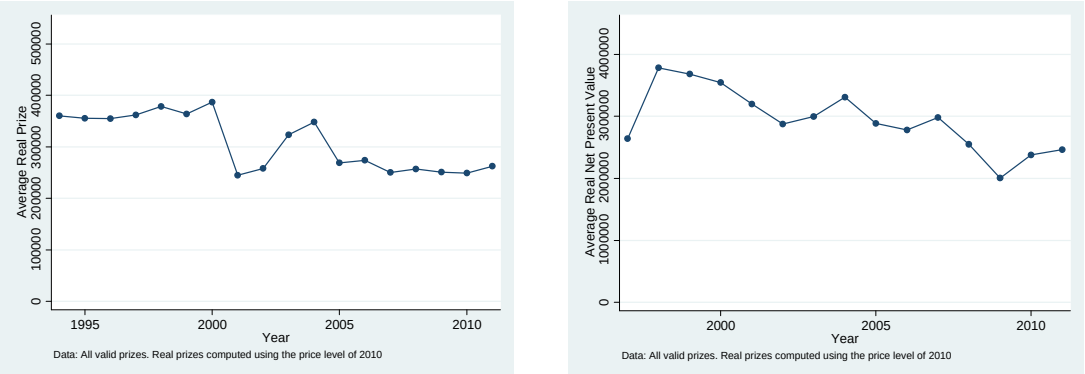


TRISS FIGURE I

The evolution of average duration (left) and nominal installments (right) over time

The time trends of the average duration and nominal installments for Triss-Monthly are shown in Triss Figure I. Despite the relatively frequent changes to the prize plan, there is little movement in the average duration and size of the installments. Assuming a discount rate of 2%, we can calculate the net present value (NPV) of Triss-Monthly installments in 2010 prices. The NPVs range from 1,059,930 to 12,061,635, with an average of 2,915,408 SEK, giving a total real prize amount of 2.37 billion SEK. Thus the prizes in the two Triss lotteries sum up to more than 3.9 billion SEK.

Triss Figure II shows the evolution of real prize amounts in Triss-Lumpsum and Triss-Monthly over time. Most notably, the average prize amount for Triss-Lumpsum falls in 2000 in response to the new prize plan. There is also a downward trend for Triss-Monthly, reflecting the fact that Svenska Spel did not adjust the prize plans for inflation.



TRISS FIGURE II
The evolution of real prizes in Triss-Lumpsum (left) and Triss-Monthly (right) over time

Step B: Coding of shared prizes. A feature of Triss is that friends or relatives may share ownership of tickets. Conceptually, we are interested in establishing who owned a winning ticket at the time that it was purchased. For example, consider a group of friends who pool resources to participate in the lottery. In this case, each participant clearly owns a share of the ticket and is *ex ante* entitled to a share of any prize won. On the other hand, we do not consider the mere intention of sharing a prize with friends or family as shared ownership. For example, consider a grandmother who announces she will share some of the prize money with her grandchildren. Her decision to share is endogenous to winning the lottery and we therefore do not consider the grandchildren as owners of the original ticket. The key question is whether a contract, implicit or explicit, to share the prize can be said to have existed before it was revealed that the ticket entitled its owner to participate in the TV show. For this reason, statements that part of the prize money may be given to friends or relatives are not coded as a 'shared prize', because such statements just reveal an intention of how to spend the money, not an *ex ante* transfer of ownership.

Fortunately, the information supplied in the supplementary information column is often sufficiently rich to clearly establish the *ex ante* ownership of winning tickets. In each case for which the column contains an entry suggesting that the prize was shared, we recorded all the available information about the co-owners and sought to identify them (see step D below). We also recorded the number of co-owners. In doing so, we only sought to identify individuals whose information suggested shared *ex ante* ownership. Thus we did not code as shared prizes cases in which the winner stated an intention to "give away" the prize money. Note that we do not code statements of shared ownership between spouses as shared prizes.

Although the information available on sharing is typically unambiguous, some cases require judgment calls. For instance, it is not always clear whether ownership was shared or whether the owner just intended to give away part of the prize. This is especially true after 2000, when the supplementary information column typically contains information about how winners intend to spend their prize money. For each case where we find indications that the prize may be shared, we code the degree of certainty that the *ex ante* ownership really was shared, the number of people sharing the prize (when possible), and the type of relationship between the person appearing in the original data and their potential co-winners.

In general, we assume that each co-winner holds an equal stake in the prize, unless it is explicitly stated otherwise.

Step C. Identifying participants' PINs. The spreadsheet supplied to us by Svenska Spel typically contains basic demographic information about the participants. It does however not contain their PINs, which are required to match individuals to register data.

Step C1. First, we identified the PINs of the recorded participants. In order to do this, for each draw, we identified the names and PINs of all individuals domiciled at the address provided by the participant to Svenska Spel using publicly available registers. If a participant had the same name as an individual at the address, and if the age implied by that individual's PIN was consistent with the information in the spreadsheet, then we assigned the PIN to the participant. If the address search did not produce a match, we instead looked up the phone number provided by the participant, again using publicly available registers. Often, we were able to recover the address of the person to whom the phone number was registered. Using the address, identifying the PIN of the registered owner is straightforward. If the name and age implied by the PIN match the information in the spreadsheet, then we assigned the PIN to the participant.

The PINs of some participants who could not be identified via their address, but had (very) rare names, were assigned based on a successful name and age match to an individual domiciled in the same region. In doing so, we erred on the side of caution. Further, to avoid spurious matches, we deleted participants who in the publicly available registers were not listed as alive at the time of the lottery win, or who did not have a recorded gender or year of birth.

Using age, address and phone number, we were able to match PINs to 4,359 participants associated with 4,369 draws (10 persons were matched to two draws), corresponding to 74% (4,369/5,881) of the prizes in our data.

Step C2. If a person with a common name moved and changed phone numbers before 2008, identifying him or her using the above approach is typically not possible. The reason for this is

that easily accessible public records only go back a few years in time. We therefore solicited the help of Statistics Sweden (SCB) to identify additional participants. SCB in turn conducted a three-stage search. In the first stage, they sought to match participants to PINs based on age, family and given names, postal code, and the first four letters of the street address at the end of the year prior to the win. Only unique matches were kept. In the second stage, SCB specifically looked at remaining participants that had been tentatively identified in step C1, but for whom we had not declared a match. In the third and final stage, the first stage search was repeated for the remaining participants, but extended to the year after the win. Out of the participants we were unable to identify in step C1, SCB identified 1,400. In addition to the PINs, they also provided us with a code for the sources used in the identification process, which allows us to construct measures for the accuracy of the PINs.

TRISS TABLE IV

IDENTIFICATION OF PARTICIPANTS (TYPE 1)

Step	Code	Search (within step)	Coders	# Prizes	Explanation
C1	0	1	CLÖW	4,369	Match on age, family name, at least one given name, street address, and postal code. In some cases, match on phone number. Only unique matches kept.
C2	1	1,2	SCB	862	Match on age, family name, postal code, postal address one given name, and the first four letters of the street address.
C2	2	1,2	SCB	141	Match on age, family name, postal code, postal address, and one given name.
C2	3	2,3	SCB	340	Same as code 1, except spelling mistakes in names, addresses, nicknames instead of given names in the original data, or errors in the postal code.
C2	5	2,3	SCB	20	Same as code 1, except failure to match on address.
C2	6	2,3	SCB	15	Only one person with the same name in the postal address and in a reasonable age range.
C2	7	2,3	SCB	19	Only one person in the country with the same name, or the only name that fits within a reasonable age span.
C2	8	2,3	SCB	3	Same as code 1, but failure to match on stated age.
C3	1	1	CLÖW	20	Almost completely certain (name, age, region, telephone/street address matches).
C3	2	1	CLÖW	9	Quite certain (name, age)
C3	3	1	CLÖW	10	Uncertain

Notes. CLÖW = authors; SCB = Statistics Sweden.

Step C3. In the third step, we made a final effort to manually identify the remaining participants using their names, age, phone numbers, addresses and SCBs databases. This way,

we were able to identify an additional 39 participants with varying degrees of certainty.

In total, we identified the participants PINs for 4,369 + 1,400 + 39 = 5,808 out of the 5,881 prizes, or 98.8%, representing 5,789 unique participants. Triss Table IV summarizes the procedure.

Step D. Identifying winners not on the participant list. There were 125 winners who sent a delegate to participate in the TV show. Of these, 32 appear in the original data from Svenska Spel, whereas the name of the delegate is given in 93 cases. Poor health is a common reason for declining to participate, and the delegate is usually a younger relative. Furthermore, in some instances, the spreadsheet from Svenska Spel notes that the winner shares ownership of the lottery ticket. In these cases, additional information about either family relationships or names and addresses of co-winners often allow us to identify the original owners and co-winners.

We sought to identify all individuals who are not listed as participants in the spreadsheet, but who nevertheless owned a share in a winning ticket. That is, either winners who sent a delegate to the show or co-winners who shared the prize but did not appear on the show. We call this type of winners 'Type 2 winners'. All Type 2 winners are linked to a specific prize, and hence to a participant.

Using the spreadsheet, we were able to identify a subset of Type 2 winners. For example, the supplementary information column may record that a winner "shares the lottery ticket with a friend", followed by the friend's name and address. Like for the participants, we excluded individuals who were not listed as alive in publicly available registers at the time of the lottery win, or who did not have a recorded gender or year of birth. With the help of SCB, we were able to identify a total of 372 Type 2 winners associated with 253 prizes. In 56 cases, this identification was straightforward, since the spreadsheet contained their PINs. An additional 57 were identified using other information such as name and address. Finally, augmenting the available demographic information on participants and co-winners with any family relationships recorded in the spreadsheet SCB was able to identify 259 Type 2 winners.

Step E. Matching prizes and participants. As explained above, out of 5,881 prizes, 5,857 were valid (step A) and 5,808 could be matched to a participant (step B). Matching valid prizes to identified participants gives us a sample of 5,786 participants. Because 19 out of these appeared twice on the TV show there are 5,767 unique participants.

Matching the set of identified Type 2 winners to the set of valid prizes, dropping one duplicate and one who was also a participant, leaves 368 unique Type 2 winners, none of which won more than once.

Step F. Sample restrictions and final samples. To construct the final samples, we start with the matched sample in step E, and then impose the following restrictions:

1. We excluded the 93 participants who were delegates without a stake in the prize.
2. We excluded 5 winners who won twice in the same cell (defined in Section IV.C) and therefore lack a valid control group.
3. We excluded the 39 participants we identified in step C3 (see Triss Table IV) and 3 Type 2 winners whose PINs were uncertain.
4. We deleted 35 winners (20 participants and 15 Type 2 winners) for whom the coding of sharing and delegates was uncertain.
5. We deleted 1 winner who died before the draw.
6. We deleted 37 winners (all participants) who shared their prize but for whom information regarding the number of co-winners is uncertain.

Imposing these restrictions leaves us with a basic sample of 5,645 prizes and 5,925 different winners who won a total of 5,938 times. Thus, we are able to match at least one winner to 96.0% (5,645/5,881) of all the prizes in the original sample from Svenska Spel.

Depending on the context, we consider two additional restrictions that leave us with four different final samples (see Triss Table V). First, we exclude all 2011 winners since we typically lack data on outcome variables for this year. Second, we also exclude winners of shared prizes since the coding of these is uncertain.

TRISS TABLE V

FINAL SAMPLE

	Including shared prizes		Excluding shared prizes	
	#Prizes	#Winners	#Prizes	#Winners
Including 2011	5,645	5,925	5,242	5,231
Excluding 2011	5,323	5,596	4,954	4,944

IV.C. Constructing the Triss Cells

The cells used in the adult analyses are defined by the unique combination of year and prize plan, implying that we construct separate cells for the Triss-Lumpsum and Triss-Monthly lotteries. For the child analyses, we also condition on the winning parent's number of pre-lottery children, as explained above for the PLS sample.

IV.D. Quality Control and Robustness Checks

Finally, we examine the coding of shared prizes and the quality of the Svenska Spel data.

Sharing. As discussed in Section IV.B step B, we define a prize as 'shared' if ownership of a winning lottery ticket was shared prior to qualifying for the TV show, that is before buying (or scratching) the ticket. In contrast, cases where the owner of the ticket only pledges to "share" part of the winnings with friends or relatives are not classified as shared ownership. Unfortunately, given the available information, it is not always straightforward to disentangle these two different types of situations from each other. We therefore classify shared prizes into three different categories depending on our confidence in the underlying information, although in practice we only use two of these. 'Shared_1' implies we are confident in a prize being shared (303 cases), whereas 'Share_3' denotes a lower level of confidence (126). Finally, 'Share_2' refers to cases where we believe that the participant made the decision to share the prize before appearing on the TV show (1). The number of prizes coded as Share_3 increases sharply between 2002 and 2003 as a response to the shift in the type of information provided by Svenska Spel. Whether we underestimate the number of shared prizes before this shift, or overestimate it after the shift, is an open question.

Triss Table VI gives the average prize shares for participants and Type 2 winners. As can be seen in column 1, the average share of participants is close to 96% for both Triss-Lumpsum and Triss-Monthly. The explanation for these high averages is that the vast majority of participants are not coded as sharing prizes. In column 2, the average share for Type 2 winners is reported. While some Type 2 winners also do not share the prize, most do, and the average share is 36.0% for Triss-Lumpsum and 38.3% for Triss-Monthly.

TRISS TABLE VI

SHARE BY WINNER TYPE

	Participants and non-sharing Type-2 winners (%)	Type-2 winners part of a shared prize (%)
Average Share	95.7	36.4
Average Share: Triss-Lumpsum	95.8	36.0
Average Share: Triss-Monthly	95.9	38.3

The fact that the average shares do not vary systematically across lotteries provides some reassurance regarding the quality of our coding of shared prizes. As discussed in Section IV.B step A, the average value of a Triss-Monthly prize is approximately 10 times that of a Triss-

Lumpsum prize. Owners who have not pledged to share their prize prior to qualifying for the TV show thus face a much higher cost of sharing if they won a Triss-Monthly prize.

As discussed in section IV.B step B, the supplementary information column often allows us to infer with whom the participant shares the prize. Despite the fact that we do not actually code prizes shared amongst spouses as shared, most of the sharing takes place within the family. The most popular form of sharing is with children (26%), followed by friends (20%), mothers (15%), sisters (13%), and cohabitants (6%).

TRISS TABLE VII

SHARE BY PRIZE AND PARTICIPANT CHARACTERISTICS

	(1)	(2)	(3)	(4)
Constant	0.967***	0.963***	0.959***	0.944***
	(0.0107)	(0.0107)	(0.00244)	(0.0104)
Age	0.000202	0.000314**		
	(0.000155)	(0.000154)		
Female	-0.0252***	-0.0250***		
	(0.00434)	(0.00432)		
Nordic	-0.00899	-0.0127		
	(0.00829)	(0.00845)		
Married at t = -1	0.00256	0.000766		
	(0.00458)	(0.00462)		
Wage earnings at t = 1 (1000 SEK)	5.15e-07	1.82e-05		
	(1.42e-05)	(1.49e-05)		
Amount won (million SEK)			-0.0046	0.0048*
			(0.0036)	(0.0029)
Year*Prize plan FE	No	Yes	Yes	Yes
Sample	All	All	Lumpsum	Monthly
N	5,614	5,614	4,851	786
R^2	0.007	0.026	0.014	0.049
p-value joint significance	0.000	0.000	-	-

Notes. Sample restricted to participants and non-sharing Type 2 winners (see text for definitions). Amount won is the nominal lump sum prize and the nominal NPV of monthly installments, assuming a 2% discount rate on top of inflation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Triss Table VII shows the share of prizes held by participants and non-sharing Type 2 winners as a function of basic socioeconomic characteristics (columns 1 and 2) and prize amount (columns 3 and 4). We do not include Type 2 winners who have a stake in a shared prize, since we are interested in what factors predict sharing. The result that stands out from columns 1 and 2 is that women are less likely to share. However, this result may simply reflect

that men are more likely to appear on the TV show in case of a shared prize. Columns 3 and 4 show the relationship between prize amount and the participant's share. The coefficient for Triss-Monthly is marginally statistically significant (*p*-value 0.099), but both coefficients are small in economic terms. The coefficient estimate for Triss-Lumpsum implies that a 1 million SEK increase in prize amount predicts a 0.46 percentage points (4,600 SEK) lower share for the participant. In contrast, the coefficient estimate for Triss-Monthly instead predicts a 0.48 percentage points (4,800 SEK) higher share.

We also evaluate the coding of shared prizes by matching the Triss data to the Swedish Wealth Registry. Between 1999 and 2007, the Swedish Wealth Registry collected high-quality data on wealth for the entire Swedish population. We first estimate a regression of the following form

$$Wealth_{i0} = \beta_0 + \beta_1 Prize_{i0} + \beta_2 Split_{i0} * Prize_{i0} + \beta_3 Split_{i0} + \beta_4 Wealth_{i-1} + u_{i0},$$

where $Wealth_{i0}$ is the winner's individual net wealth the year after winning the lottery, $Prize_{i0}$ is the prize amount (before sharing across co-winners), and $Split_{i0}$ is a dummy equal to 1 if the lottery prize was shared. In addition, we control for year by prize-plan fixed effects. If our coding of shared prizes is accurate, we should be able to reject that $\beta_2 = 0$.

In our second regression, we consider the specification

$$Wealth_{i0} = \beta_0 + \beta_1 Share_{i0} * Prize_{i0} + \beta_2 Split_{i0} * Share_{i0} * Prize_{i0} + \beta_3 Split_{i0} + \beta_4 Wealth_{i0} + u_{i0},$$

where compared to above, the only difference is that we multiply the prize amount won by $Share_{i0}$, person *i*'s share of the prize. Here instead, if our coding of shared prizes is accurate, we should not be able to reject that $\beta_2 = 0$. In other words, taking the share into account, whether or not the prize was split (and thus the share < 1) should not matter for the effect on wealth. Since we are focusing on the immediate impact on wealth, the monthly installments in Triss-Monthly are not well suited for evaluating our coding of shared prizes in this context. Triss Table VIII therefore only reports the results for Triss-Lumpsum. Studying column 1, we can see that the coefficient on prize amount for non-shared prizes is 0.491, implying that net wealth increases by 0.49 SEK for every 1 SEK won.[2] As expected, the increase in wealth is much smaller for shared prizes, with a 0.20 SEK increase per 1 SEK won by co-winners. This result confirms one important aspect of shared prizes – that multiple parties indeed share the prize money.

[2] The gradient between wealth and prize amount won is not equal to 1 for many reasons: winners may give away part of their prize (e.g., to spouses or children), increase consumption, buy cars or other consumer durables, renovate their houses, hide their wealth in offshore accounts or in cash, or buy property that is not taxed to its full market value.

However, the fact that a prize is shared *ex post* is not, in itself, proof that ownership *ex ante* was shared. Column 2 shows the results from the second specification where we interact the amount won with the share. Here we instead express lottery winnings in terms of the amount won by a particular person. As explained above, in the absence of coding error in shared prizes or heterogeneity between sharing and non-sharing winners, the fact that a prize is split should not interact with the effect on net financial wealth. The point estimate for the interaction effect between split prizes and prize amount won adjusted for share in column 2 implies that for shared prizes net financial wealth increases by 0.55 SEK for every 1 SEK won. This number is close to our estimate for non-sharing winners (0.49) and the difference is not statistically significant, suggesting that our coding of shared prizes is accurate.

TRISS TABLE VIII

EVALUATING CODING OF SHARED PRIZES USING THE WEALTH REGISTER

	(1)	(2)
Amount won (Lumpsum)	0.491***	
	(0.0413)	
Split*Amount won (Lumpsum)	-0.293***	
	(0.0653)	
Share*Amount won (Lumpsum)		0.491***
		(0.0413)
Split*Share*Amount won (Lumpsum)		0.0634
		(0.115)
Split	24,261	20,703
	(19,547)	(18,011)
Lagged net financial wealth	1.051***	1.052***
	(0.0146)	(0.0145)
Constant	8,755	8,477
	(9,140)	(9,130)
N	2,110	2,110
R^2	0.881	0.882

Notes. The dependent variable is (nominal) net financial wealth. Sample is restricted to lumpsum winners between year 2000 and 2007. Both regressions include fixed effect for prize plan by year. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Manual verification of data quality. The National Library of Sweden have tape recordings of most televised Triss draws. We accessed a subset of these for two purposes. First, as described in Section IV.B step A, we recovered the prize amount in 111 out of the 125 cases for which this information was missing or incompatible with the prize plans. Second, we used

another independent sample of tape recordings to assess the reliability of the information in the data that Svenska Spel provided us with.

In doing so, we started by randomly selecting 500 prize draws from our original list and ordered the tape recordings of the TV show for the relevant days. This way we were able to retrieve 456 televised draws, or 91% of the random sample. Research assistants then reviewed the recordings, following a protocol, and documented any information disclosed about the participant in the show (name, hometown, whether the participant was a delegate, and prize sharing).

Triss Table IX shows the correspondence between the spreadsheet data from Svenska Spel and our coding based on the tape recordings. Of course, an inconsistency between the spreadsheet and the televised draw does not necessarily imply that the spreadsheet is incorrect. The name of the participant was inconsistent in six cases (1.3%), which generally meant that the spelling differed in a non-trivial way. For instance, in one case the order of the given and family names was reversed. In four cases (0.9%), the information on hometown was inconsistent, possibly because these winners had moved in between reporting their win (which entitled them to participate in the show) to Svenska Spel and the airing of the show. Finally, in one instance, the duration of a Triss-Monthly prize was inconsistent with the spreadsheet (15 years in the spreadsheet, 10 years in the TV show). Overall, our coding of televised draws suggests the quality of the data from Svenska Spel is very high.

TRISS TABLE IX

QUALITY OF BASIC INFORMATION ORIGINAL FILE

	Perfectly consistent	Minor inconsistency	Inconsistent
Name	443	7	6
Postal address	452	-	4
Prize	455	-	1

In addition to using the televised draws to assess the basic data quality, we also use them to evaluate our coding of delegates and shared prizes based on the supplementary information column in the spreadsheet. By cross-matching our coding of shared prizes from the spreadsheet and the televised draws, we can get a sense of the accuracy of our coding and the fraction of shared prizes in our data. Even so, since a risk of over- or underreporting of shared prizes exists both in the spreadsheet and in the TV data, we are not able to determine the exact fraction of shared prizes. For instance, some winners could make statements suggesting that ownership is

shared even when it is not, whereas others may abstain from reporting shared ownership.

Of the 456 sampled draws for which we have both spreadsheet and TV data, the spreadsheet suggested that prizes were shared in 41 cases (9.0%, compared to 7.7% for the full sample). Out of these, the TV data indicated that prizes were shared in 27 (66%). In addition, the TV data indicated that the prize was shared in a further 13 (2.9%) cases that were not recorded in the spreadsheet. Notably, either using only the spreadsheet information or the TV data, we find that the estimated frequency of shared prizes is almost identical (9.0% and 8.8%, respectively).

Finally, we also investigated the quality of the information regarding the number of co-winners and their identities. For both of these domains we have spreadsheet and TV data in 19 cases (out of 27), and for both this information was consistent except in one case.

Overall, the relatively high level of overlap between the information in the spreadsheet and the data from the televised draws suggests that although we cannot perfectly identify the sharing of prizes, the error resulting from measurement error in the share won is likely to be small.

V. The Kombi Sample

V.A. Background on Kombi

Kombi is a monthly subscription lottery run by A-lotterierna, a limited company owned by the Swedish Social Democratic Party and its youth movement, since 1981. In 1998 lottery tickets cost 175 SEK, but in 2000 the price was increased to 200 SEK. Subscribers are billed monthly, usually by direct debit, and ticket owners automatically participate in regular prize draws where they can win cash prizes or merchandize. The number of monthly tickets is capped, with the cap being 180,000 in 1998 and 225,000 in 2010. The overwhelming majority of Kombi subscribers buy one or two lottery tickets per month.

V.B. Constructing the Cells and the Final Estimation Sample

A-lotterierna provided SCB with two data sets on our behalf. The first – 'Winners' – contains information about all large prizes, 1 million SEK net of taxes and above, awarded 1998 to 2010 as well as some prizes awarded in 2011. For each prize, the file contains information about the PIN of the winner, the prize amount, and the date on which the prize was awarded. In total there are 499 large prizes won by 496 unique individuals. The second data set – 'Kombi Panel' – is a panel that tracks all the subscription ticket owners who supplied their PIN over time (approximately 98%).

We drop 31 prizes won in 2011 (most outcome variables only extend to 2010) and an additional 6 prizes because the winner was missing from Kombi Panel. The PINs of all the remaining winners passed the quality control outlined in Section IX.A, leaving us with 462 prizes won by 459 individuals.

Our empirical strategy is to compare each winner of a large prize to "matched controls" who did not win a large prize, but owned an identical number of tickets at the time of the draw. In the adult analyses, we matched each winner of a large prize to (up to) 100 controls. If there were more than 100 possible matched controls, we choose the 100 that were most similar in terms of age and gender, randomly breaking ties if the 100^{th} control was of the same age and gender as the winner. One winner could be matched to 8 controls, another to 51, and one to 65. The remaining winners were all matched to 100 controls. Thus, the final sample comprises 462+8+51+65+459*100 = 46,486 observations.

Note that our matching procedure does not prevent an individual from being drawn as a matched control in more than one draw, nor does it prevent the winner of a large prize from

being drawn as a matched control in a month in which he or she did not win. The reason for this is that imposing either restriction would give rise to a selected sample wherein the large winners are compared to a possibly non-random sample of non-winners with the same number of tickets. Conceptually, we treat an individual drawn multiple times in the matched data as multiple observations, therefore there are 40,366 unique individuals in our final sample of 46,486 observations.

For the intergenerational analyses, we proceeded analogously and matched winners to (up to) 100 controls with the same number of lottery tickets and pre-lottery children. If there were more than 100 possible matched controls, we used those with the children who were the most similar in terms of age and gender. In total, there are 58 winners in the Kombi sample that have pre- or post-lottery children. Out of these, one could not be matched to any control, one to 9 controls, another to 18, and one to 45 controls. The remaining were matched to 100 controls. Thus, the final sample comprises 58 + 0 + 9 + 18 + 45 + 54*100 = 5,530 observations at the parental level. The 58 winners have 89 pre-lottery children and 10 post-lottery children, so in total the sample consists of 99 children of winners matched to 9,456 children of controls.

VI. Variable Definitions

VI.A. Mortality Variables (Adult Analyses)

We use data from the *National Cause of Death Register* to define mortality variables as follows.

- *Deceased* – Equal to 1 in year *t* if the individual has a recorded date of death that predates the month of the lottery draw in year *t*.
- *Deceased Cancer* – Equal to *1* in year *t* if the individual has a recorded date of death that predates the month of the lottery draw in year *t* and the primary cause of death is listed as cancer. See Table AVII for a description of the ICD codes used to classify cancer.
- The variables *Deceased Circulatory*, *Deceased Respiratory* and *Deceased Other* are defined analogously. Deaths due to hypothesis-based causes are also defined analogously, with one exception: we only require that one of the diagnosis codes on the death certificate matches the diagnosis code of the disease.

VI.B. Drug Prescription Variables (Adult Analyses)

We define a set of variables based on information in the *Prescribed Drug Register* from 2006 to 2010.

- *Any consumption* – Equal to 1 in case an individual was prescribed a nonzero amount of any drug listed in the register between 2006 and 2010.
- *Any consumption of drugs in category X* – Equal to 1 in case an individual was prescribed a nonzero amount of any drug in category X between 2006 and 2010. Table AVII shows what ATC codes belong to each category.
- *Total consumption*. The total amount of Defined Daily Doses (DDD) prescribed in all categories between 2006 and 2010.
- *Total consumption of drugs in category*. The total amount of Defined Daily Doses (DDD) prescribed in category X between 2006 and 2010.

VI.C. Hospitalization Variables (Adult Analyses)

We use data from the *National Inpatient Register* to define the following variables.

- *Hospitalized (Any) within t years after the win* – Equal to *1* in year *t* if the individual was hospitalized at least once between winning the lottery and year *t*.

- *Hospitalized (*>=7 days*) within t years after the win* – Equal to *1* in year *t* if the individual was hospitalized at least once and spent a total of 7 nights or more overnight in the hospital between winning the lottery and year *t*. The number of overnight stays is inferred from the entry and discharge dates.
- *Health Index.* We construct a health index using the 2000 representative sample, excluding individuals that are also in our lottery sample, and run a Probit regression in which the dependent variable is a binary variable equal to 1 if the individual was deceased in the year 2005. In this regression, we include sex interacted with age bins of five years, the Charlson co-morbidity index (defined below) as well as indicator variables for hospitalization due to any of the common or hypothesis-based causes listed in Table AVII. For each cause of hospitalization we include one indicator variable defined using only the main hospitalization cause and one variable defined based on all listed causes. In addition we also include indicator variables indicating whether the person was hospitalized at all, more than7 days during a year or more than 28 days during one year between 1995 and 1999. The coefficients from this regression are then used to predict five-year mortality for the lottery players in our sample. The health index is defined as the predicted mortality multiplied by 100 and the index is set to 100 if the individual has deceased.

The variables *Hospitalized for X within t years after the win,* where X is either one of the common or one of the hypothesis-based causes, are defined analogously to the hospitalization variables above. See Appendix Table AVII for a description of the ICD codes used to classify the categories.

VI.D. Net Wealth and Other Variables

- *Net Wealth (Sweden)* – Total assets at market value in thousands of SEK measured in 2010 prices and winsorized at the 1st and 99th percentiles. The variable is obtained from Statistics Sweden's *Wealth Register* and is only available for the years 1999 to 2007. The register contains annual information about the year-end value of assets and debt for all Swedes for the period 1999-2007 (Statistiska Centralbyrån 2006; Waldenström 2008). It is based on information that Swedish banks, financial institutions and government agencies were legally required to supply SCB with. In 2007 the Swedish wealth tax was abolished and as a consequence some of the requirements on financial institutions to report to SCB were relaxed and the register discontinued.

The register includes information about the value of various types of real estate owned (major categories: owner-occupied dwellings, cooperative flats, land and vacation homes), financial assets (major categories: bank account balances, indirectly and directly held stocks and bonds), and debt (major categories: mortgages and student debt). The net wealth variable in the register is constructed by summing the market value of all assets owned and then subtracting all debt. For households whose wealth exceeds the threshold for wealth taxation, SCB also has additional data on other assets, including non-listed stocks, cars, boats and jewelry. Before the tax was abolished, this threshold was set to 1.5 million SEK for single households and 3 million SEK for cohabiting married couples, cohabitants with children and parents living with children under 18.

The variables in the register compare favorably with other data sources in terms of quality. Nevertheless, the data have some limitations (Statistiska Centralbyrån 2006) that are relevant for interpreting our results, e.g. those in Figure I.

First, and most importantly, the value of real estate wealth is harder to assess reliably than the value of financial wealth because real estate is less standardized and because SCB does not always have accurate information about the size and condition of properties. SCB therefore assigns an imputed value to each property using available information about the unit along with information about house prices in the vicinity. Therefore, if some winners spend part of their lottery wealth on home improvements – as suggested by for instance Kaplan (1978) – then the increased resale value of the home will not show up as increased real estate wealth unless the unit is sold immediately, whereas the financial cost of the improvement will cause financial wealth to decrease.

Second, the net wealth variable will typically not include cash, cars, or other durables, merchandise, assets transferred to other family members, or money that has been concealed from the tax authority. The purchase of a car (or some other consumer durable) worth 100,000 SEK will thus typically reduce measured wealth by 100,000, even though actual net wealth has only declined by 100,000 minus the resale value of the car.

For these reasons, we caution against interpreting the $s = 0$ coefficient estimates depicted in Figure I as evidence that 40% of lottery wealth is consumed in the year-of-win.

- *Domiciled Abroad* – 1 if an individual is registered as living outside Sweden on December 31 of the year in question. Obtained from *Total Population Register*.

VI.E. Proxy Charlson Co-Morbidity Index (Adult Analyses)

We used the STATA program charlson.ado (Stagg 2006) to construct the proxy Charlson index. This program assigns individuals a co-morbidity index score based on their hospitalization records. The program takes as its input ICD9 CM or ICD10 codes. In our data, however, we observed ICD8 codes for the period up to and including 1986, ICD9 codes 1987-1996, and ICD10 codes for the period 1997-2010. We therefore proceeded as follows to transform the data into a format the Stata program can read:

Step A. We converted the ICD8 codes to ICD9 codes using the cross-walk provided by the Swedish National Board of Welfare (Socialstyrelsen 2013a). If the conversion table assigns an ICD8 to a unique ICD9 code ("target"), we convert the ICD8 code to that unique target. If an ICD8 code is mapped to multiple ICD9 targets, we assign it to the ICD9 code that is most frequently observed in our sample within the set of target ICD9 codes.

The data released to us are a mix of three- and four-digit ICD codes, so to determine the frequency of a diagnosis code, we proceeded as follows: First, we built a frequency table of target ICD9 codes. If a 4-digit ICD9 target code is in the cross-walk but is *not* observed empirically, we checked whether its 3-digit parent category is in our sample. In this case, we add the parent category to the list of candidate target codes. We discard all other target codes that we do not observe directly or indirectly (via the parent category).

To illustrate the way we treat multiple mappings, consider the ICD8 code "011." The corresponding ICD9 codes from the crosswalk are 010.8, 011.0, 011.1, 011.2, 011.3, 011.4, 011.5, 011.6, 011.7, 011.8, 011.9, 012.2, and 771.2. Because we observe none of these codes in our sample, we replace these candidate ICD9 codes with their parent categories 010, 011, 012, and 771. Out of these, only 011 and 771 are present in our sample, the latter being the most frequently observed candidate ICD9 code. Hence our algorithm maps ICD8 code "011" to ICD9 code "771."

Using this method, we are able to map 66% of ICD8 codes.

Step B. We used the Stata program to convert the hospitalization code to a Charlson index. In computing the Charlson index, we used all diagnoses codes associated with the longest hospitalization of the individual in any given year. Because the mapping from *ICD8* to *ICD9* is imperfect and because the Stata program was designed to process hospitalization codes

provided in the *ICD9 CM* format, we refer to the resulting variable as our proxy Charlson Index. Nevertheless, this variable is strongly predictive of mortality in our data, including in the pre-1987 data.

VI.F. Baseline Covariates (Adult Analyses)

Our set of baseline controls for the adult analyses is composed of a set of birth, demographic and health variables.

Birth characteristics

- *Age* – Year minus year of birth, where information about the latter is obtained from the *Total Population Register.*
- *Female* – 1 if an individual is female.
- *Nordic Born* – 1 if there is no information in the *Total Population Register* that an individual was born outside the Nordic countries.

Demographic characteristics

- *College-Educated* – 1 in year *t* if the individual has completed at least three years of college education. From 1990 and onward, this variable is obtained from *Longitudinal integration database for health insurance and labour market studies*. For the pre-1990 period we use the information on educational attainment in the *Population and Housing Census 1970*. Individuals with a college degree in 1990 but not 1970 are classified as having a college degree from the year in which they turned 30 years of age.
- *Labor Income Last Year / 1000* – Labor earnings in thousand SEK measured in 2010 prices. The variable is obtained from Statistics Sweden's *LISA* for 1990-2010 and from their *Income and Taxation Register* for earlier years and is winsorized at the 1st and 99th percentiles.
- *Married* – 1 if an individual is married. The variable is obtained from Statistics Sweden's *LISA for 1990-2010 and from the Total Population Register for earlier years.*
- *Retired* – 1 if an individual received a non-zero income from old-age pension. The variable is obtained from Statistics Sweden's *LISA* for 1990-2010 and from their *Income and Taxation Register* for earlier years.

Health characteristics

- *Charlson* – Comorbidity measured by a proxy for the Charlson index using data from the Swedish *Inpatient Register*. Additional details on variable construction are provided in a separate section above.

- *Hospitalized (Any)* during the 5 years before the win
- *Hospitalized (≥7 days)* during the 5 years before the win
- *Hospitalized for Cancer* during the 5 years before the win
- *Hospitalized for Respiratory* during the 5 years before the win
- *Hospitalized for Circulatory* during the 5 years before the win

VI.G. Baseline Covariates (Child Analyses)

The baseline covariates used in the child analyses include the birth and demographic characteristics (except for *Retired*) of the winning parent, and the birth characteristics of the child.

VI.H. Infant Health (Post-Lottery Children)

Our list of pre-specified outcomes contains three measures of infant health from the *Medical Birth Register:*

- *Birth weight* (in grams)
- *Low birth weight* – Equal to *1* in for birth weight below 2,500 grams
- *Preterm births* – Equal to *1* in for gestation length below 37 weeks

VI.I. Hospitalization Variables (Pre-lottery Children)

We use data from the *National Inpatient Register* to define the following variables

- *Hospitalized (Any)* within *t* years after the win
- *Hospitalized (>=7 days)* within *t* years after the win
- *Hospitalized for external causes* within *t* years after the win
- *Hospitalized for respiratory causes* within *t* years after the win

See Appendix Table AVII for a description of the ICD codes used to classify the categories *External Causes* and *Respiratory Causes*. For hospitalizations due to external causes and respiratory disease, we restrict the sample to children aged 18 or below during the entire time window for which the variable is defined. For example, when studying respiratory disease in children over the two-year horizon, we restrict the sample to children who were at most 16 years of age in the year of the lottery event. For the omnibus category, we make no such restrictions.

VI.J. Drug Prescription Variables (Pre-Lottery Children)

Drug prescription data are available from 2006 to 2010 in the *Prescribed Drug Register*. Our list of pre-specified outcomes includes the following outcomes derived from this register:

- *Asthma & Allergy* (ATC codes R01, R03 and R06)
- *Mental Health* (ATC codes N05 and N06A).
- *ADHD* (ATC codes N06BA)
- *Total* (net of above three categories and ATC code G03)

Each variable is computed as the sum of daily doses prescribed over the 2006-2010 period. In case data for the entire period is unavailable, or the child is outside of the relevant age range for some years, we take the average and multiply by 5. We exclude children for whom we have less than 3 years of data. For Mental Health, we exclude years in which the child had not yet turned 15. For ADHD, we include children aged 6-18. In both cases, these cutoffs were chosen based on the exploratory analyses which showed strong age-related patterns of consumption. For asthma and allergy, we only keep observations from years in which the child was 18 years of age or below.

VI.K. BMI Variables (Pre-lottery Children)

We consider three different BMI measures from the *Conscription Register*:

- BMI
- *Overweight* – Equal to 1 if BMI > 25
- *Obese* – Equal to 1 if BMI > 30

Data from the C*onscription Register* are available until 2010. Conscription is only compulsory for men and takes place around the age of 18, so our estimation sample is de facto restricted to sons born in the window 1968-1992.

VI.L. Skill and School Achievement Variables (Pre-lottery Children)

We use data from the *Conscription Register* to obtain two measures of skills:

- *Cognitive skills*. We normalize the conscript's score on "inskrivningsprovet" – test of cognitive ability, similar to the AFQT. We normalize the score by draft cohort so the resulting variable has mean zero and standard deviation one.

- *Non-cognitive skills*. We normalize the conscript's score on the "ability to cope with war stress" as measured in a psychologist interview at the military enlistment. We undertake the same type of normalization as for cognitive skills.

We also use data on grades from the *Ninth Grade Register*.

- *Grade point average (GPA) in ninth grade*. We normalize the GPA by graduation year. More precisely, for each individual, we compute the percentile ranking for that person's GPA graduating in the same year, and convolute this percentile ranking with the inverse of the standard normal distribution. We perform the normalization separately by year because we know the Swedish school system had grade inflation during our period of study, especially in the later years (Vlachos 2010). We set the GPA variable to missing for the students who cannot be identified in the register or who went to a school that did not grade students.

Finally, we use information from national tests in core subjects. Students in ninth grade take standardized national tests in Swedish, English, and Mathematics.

- *National tests in Swedish*. We normalize the grades separately for each year, using the same normalization as for cognitive and non-cognitive skills.
- *National tests in English*. Defined analogously to National Tests in Swedish.
- *National tests in Mathematics*. Defined analogously to National Tests in Swedish.

VI.M. Parental Behavior Variables (Pre- or Post-lottery Children)

- *Child net wealth 5 years after the win* (pre-lottery children). We obtain information on net wealth for the years 1999-2007 from the *Swedish Wealth Registry*.
- *Maternal smoking* (post-lottery children). We use data on maternal smoking during pregnancy from the *Medical Birth Register*. We construct a dummy variable equal to 1 if the mother reported smoking a non-zero amount of cigarettes per day when registered at the maternity clinic.
- *Parental mental health* (pre-lottery children). This variable is measured for mothers and fathers separately, and is defined as the sum of DDD prescribed over the 2006-2010 period in the ATC categories N05 ("psycholeptics") and N06A ("antidepressants"). We restrict the sample to winners who satisfy three conditions: (1) they were alive in 2010, (2) they had at least one pre-lottery child aged 18 or below in 2010, and (3) they won their prize in 2005 or earlier.

- *Parental leave* (post-lottery children). Beginning in 1993, the registers contain information about the total number of days of parental leave claimed by each individual. For each post-lottery child, we compute the total number of parental leave days claimed by the winning parent in the years that the child was 0-3 years old. A limitation of the register data is that we only observe the total amount of parental leave claimed by a parent, not the child for which the benefits were claimed. To avoid misclassification or double counting, we therefore restrict our analyses to post-lottery children without siblings 0-3 years younger in age.
- *School quality* (pre-lottery children). For most children, we define school quality as the average GPA, normalized by year as explained above, in the child's graduating school in the year of graduation from primary school. In our data, 2% of graduating children graduate from a school in which we observe fewer than 20 children in the graduation year. For such children, we define the school-quality variable as a 10-year average GPA. A small number of children also graduate from schools that are not required to assign grades, because they rely on special pedagogical strategies. For such children, we instead impute average grades using information about national test scores in Mathematics, English, and Swedish (whenever available) and conscription test scores (cognitive and non-cognitive skills, whenever available).

VII. Inference

The accuracy of the analytical standard errors that we report throughout the paper relies on an asymptotic approximation that may give rise to misleading inferences in finite samples. In this section, we describe Monte Carlo experiments conducted to (i) quantify the magnitude of finite-sample bias in our various estimation samples and outcome variables and (ii) generate the permutation-based *p*-values we report in the main text.

VII.A. Description of Monte Carlo Experiments

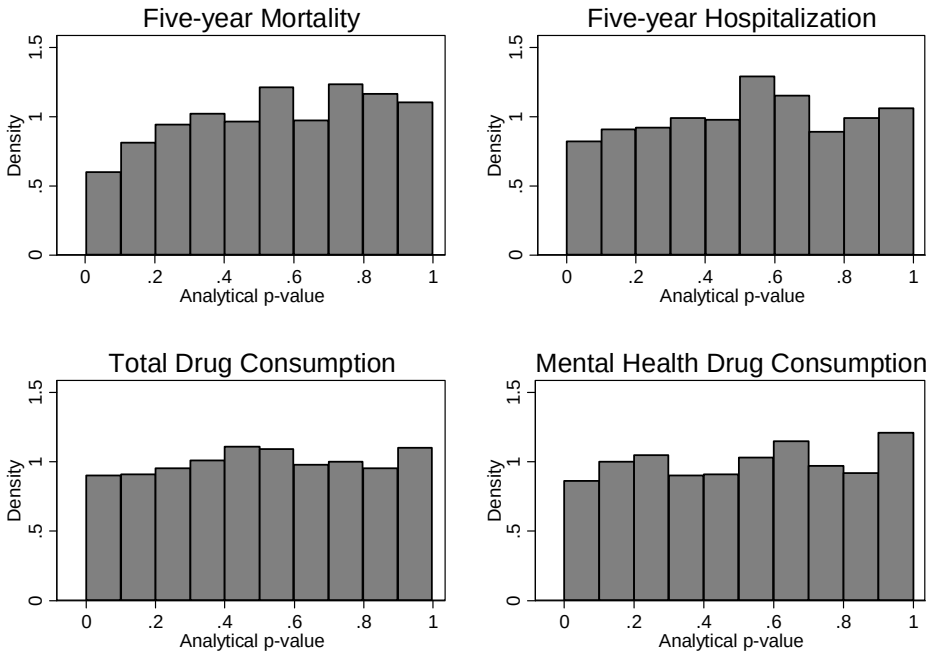
We simulated 1,000 data sets by randomly permuting the prize columns in each cell of the pooled adult sample. In the intergenerational analyzes, we proceeded analogously so that siblings always received the same permuted prize. In our simulations, the prize amount is by construction (conditionally) independent of the post-lottery outcome. It is hence straightforward to simulate an approximate finite-sample distribution of our test statistics under the null hypothesis that the effect of wealth is zero.

For each outcome and estimation sample, we simulate the finite-sample distribution of the coefficient on wealth, its analytical standard error, and the conventional *p*-value (i.e., the *p*-value derived from the analytical standard error). Because we know the rules that were used to assign prizes, the data-generating process we use is likely to be a very good approximation of the process that actually generated the data. We refer to the simulated finite-sample distribution as "approximate" to emphasize that the process we use does not incorporate any effects that a lottery draw may have on the composition of subsequent lottery populations.

Our permutation-based *p*-values are then defined as $p = 2 \times \min(1 - \tau, \tau)$, where τ is the percentile ranking of the actual regression coefficient in the simulated distribution of coefficients.

Testing for Bias in Analytical Standard Errors. Because the studied prizes are conditionally randomly assigned, the simulated distribution of *p*-values should be uniform if the analytical standard errors are unbiased. The simulations thus allow us to evaluate if the analytical standard errors are biased under the null that the effect of wealth is zero.

In the pooled adult sample, our simulations show little evidence of bias. To illustrate the point graphically, Inference Figure I shows the simulated *p*–value distribution for four typical outcomes: five-year hospitalization, five-year mortality, total drug consumption (DDDs), and mental health. We find a slight tendency for under-rejection of five-year mortality, but otherwise, we find little evidence of bias.

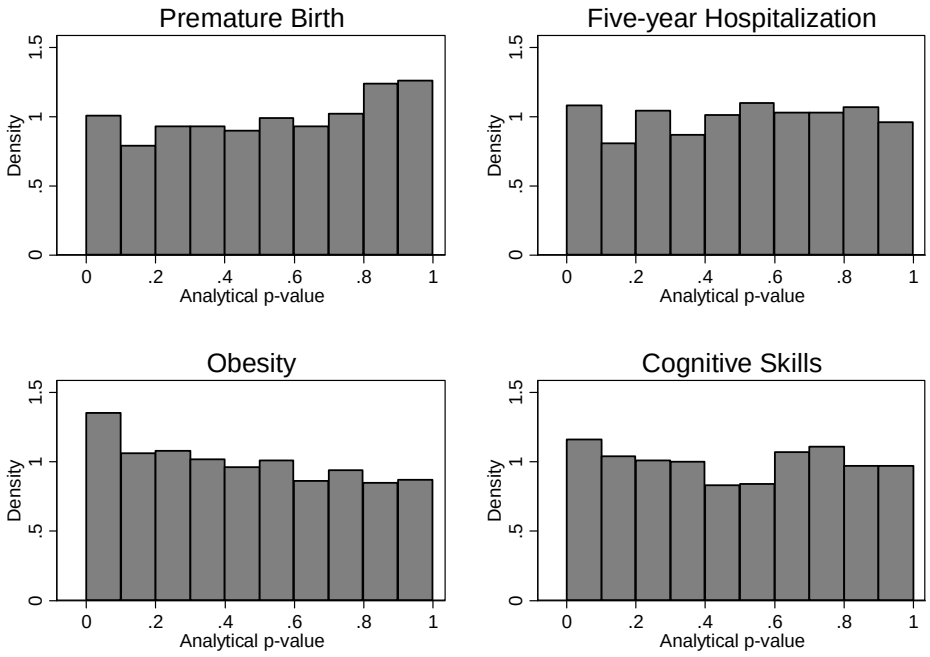


INFERENCE FIGURE I

Standard Errors

Simulated distribution of conventional *p*-values when the effect of wealth is zero.

An analogous analysis of the outcomes considered in the intergenerational analyses show slightly stronger indications of small-sample bias. The results for premature birth, five-year hospitalizations, obesity, and cognitive skills can be seen in Inference Figure II. We find a slight tendency for over-rejection for obesity and cognitive skills.



INFERENCE FIGURE II
Standard Errors
Simulated distribution of conventional *p*-values when the effect of wealth is zero.

Consistent with the analyses reported above, readers can verify that in the pooled adult analyses, the conventional and permutation-based *p*-values are consistently very similar, whereas they differ slightly more in the intergenerational analyses. In additional analyses conducted in subsamples of the simulated pooled adult or intergenerational samples, we found that analytical standard errors can sometimes be quite misleading, though only for rare and skewed outcomes. Whenever the analytical standard errors appear at odds with the permutation-based *p*-values, we rely on the permutation-based *p*-values, which are more robust to distributional assumptions.

VIII. Coverage and Quality of Main Health Data Sets

In this section, we discuss the coverage and quality of the three key population-wide administrative registers maintained by the Swedish National Board of Health and Welfare (NBHW) that we use in our analyses. By international standards the data quality is high, reflecting a tradition of vital population statistics dating back to 1749.

VIII.A. Coverage and Quality in the Swedish Cause of Death Register (CDR)

In principle coverage in CDR is complete, but PINs are missing for approximately 0.25% of the deaths, albeit rarely so after 1997. Until the beginning of July 1991, information on causes of death was obtained by SCB from the local parish registers of the Church of Sweden who were responsible for population registration (Socialstyrelsen 2010). Thereafter, the issuing doctor or institution was instead made responsible for submitting certificates of cause of death directly to SCB, and later on to the NBHW. Based on the submitted information, the contributing, including underlying, causes of death in CDR were coded by SCB and the NBHW according to the most recent applicable version of WHO's International Classification of Diseases (ICD).

Prior to July 1991 information about causes of death is almost complete, with the number of deaths with missing information being 0.005% in 1985 (Socialstyrelsen 2010). However, missing information has increased somewhat in recent years, reaching 1.3% in 2012. Likewise, the number of deaths with inadequately specified underlying cause of death has increased from 0.8% in 1975 to 2.7% in 2012 (Socialstyrelsen 2013b). There are also a number of additional factors that may affect the quality and comparability over time of the data in CDR. For instance, there is a marked decline in autopsies conducted from in about 37% of deaths in 1987 to in about 12% in 2010 (Socialstyrelsen 2013b). Furthermore, changes in coding from one version of ICD to another as well as changes to national coding practices may affect comparability over time. There have been a number of attempts to evaluate the quality in the diagnoses in CDR, for details see Socialstyrelsen (2007) and in particular Johansson, Björkenstam and Westerling (2009).

VIII.B. Coverage and Quality in the Swedish Inpatient Register (IPR)

The IPR is a part of the *Swedish National Patient Register*. Although it is not suitable for studying all types of diagnoses or medical care, the overall quality is arguably high. PINs are

available for 97.1% of all discharges between 1964 and 2008, and coverage is almost complete for individuals aged one and above resident in Sweden (Ludvigsson et al 2011). A primary diagnosis is listed for 99% of overall discharges. Over the years there have been a number of attempts to evaluate the quality of coding in the IPR. Most notably, NBHW has conducted two randomized studies using actual patient journals to verify coding (Socialstyrelsen 2009). Reviewing the journals it was found that 13% of the diagnoses at the 3-digit ICD level were misclassified in 1986 and 12% in 1990. Overall, transcription errors were rare and misclassifications were split fairly evenly between coding errors, the diagnosis in the journal having the incorrect ICD code in IPR, and misdiagnosis based on re-evaluation of the medical information available in the journal. A recent review of the literature found that somewhere between 85% and 95% of diagnoses in the IPR are valid (Ludvigsson et al 2011).

While the quality of IPR is arguably high, there are a number of changes in coding practices and time trends that could affect comparability over time. In part these changes are due to underlying changes in epidemiological trends and medical technology as well as the organization of care (Socialstyrelsen 2013c). The perhaps most notable example hereof is that the number discharges associated with psychiatric care is down 17% from 1987 to 2011. Naturally, changes in recommended coding practices and versions of ICD introduce discontinuities that may affect both comparability as well as the number of discharges associated with particular diagnoses. For instance, when the NBHW broadened the definition of acute (and subsequent) myocardial infarction(s) in their revised diagnostic guidelines in 2001 this lead to a significant increase in the number of diagnoses as well as a decrease in the diagnoses of other related heart diseases (Socialstyrelsen 2012a). Also, in recent years, the number of secondary diagnoses in IPR has increased somewhat, from 0.7 in the beginning of the millennium to 1.9 in 2011.

VIII.C. Coverage and quality in Swedish Prescribed Drug Register (PRD)

The PRD was created in 1999 and contains information on all prescribed drugs dispensed by pharmacies in Sweden (Socialstyrelsen 2013d). For drugs prescribed after June 30, 2005, the PIN of the individual who was prescribed the drug is typically recorded. The register contains information about the names and ATC-codes of drugs, as well as the number of packages dispensed and number of DDDs per package. The ATC classifications and DDD values may change over time and PRD is retroactively updated with respect to any such changes (Socialstyrelsen 2013d). In addition, DDD is not always available. For instance, cytotoxic drugs

and vaccines are not assigned DDDs. Coverage with respect to drugs dispensed is complete, but the PIN is missing for up to 0.6% of the observations (after June 30, 2005). There are a number of explanations for missing PINs, including clerical error, immigration status and protected identity. Most notably, if an individual chooses not to have a prescription dispensed within the framework of the public subsidy for drugs and healthcare, and if the prescription is not for a drug that is classified as narcotic, it is not compulsory to register the PIN in the prescription (Socialstyrelsen 2013f).

Like most similar registers, PRD only covers dispensed prescriptions, and consequently there is no information available on prescriptions that were not collected (Socialstyrelsen 2013e). Although this could potentially be a concern, most prescriptions are in fact collected within a month of them being issued. While PRD covers all prescribed drugs dispensed at pharmacies, it does not cover drugs that are dispensed at hospitals and other health care units. For the second half of 2005 the prescriptions in PRD made up 84% of total DDDs sold in Sweden and 77% of expenditures (Wettermark et al 2007). By 2011 the share of expenditures had dropped to 71.7% (Socialstyrelsen 2012b). Presumably this change is primarily due to an increase in costs for drugs dispensed at hospitals and other health care units. Whereas the share of DDDs for hospitals and other health care units in 2011 was almost the same as in 2005, 3%, the share of costs had increased from about 15% to about 20%.

VIII.D. Coverage and Quality in the Swedish Medical Birth Register (MBR)

The MBR was created in 1973 and covers pregnancies resulting in births in Sweden (Socialstyrelsen 2002). The register contains information about pregnancies, mothers and a number of birth outcomes. Coverage is almost complete, and in a typical year about 99% of all births are registered. A number of attempts have been made to evaluate the quality of the information in the register, and while some variables should be used with care, key variables are typically of high quality (Cnattingius et al. 1990; Socialstyrelsen 2002). The most ambitious evaluation of the register was made in 2002 and found that well of 99% of the children born between 1973 and 1998 had correct PINs. Likewise, or the same time period, well over 99% of children had a registered birth weight, and studying the recorded birth weight in more than 500 actual medical journals only one transcription error was identified. More recently, 99.2% of all live births in 2011 were registered in MBR (Socialstyrelsen 2013f). Of course, the measure that we use for preterm birth is potentially more problematic than that for birth weight, since gestational age is difficult to measure (Socialstyrelsen 2002; Behrman and Butler 2007).

However, both the MBR data on gestational age and our measure of preterm birth are well established and widely used.

IX. Other Analyses

IX.A. Quality Control of PINs

In our empirical analyses, we discard a small fraction of individuals whose personal identification numbers fail quality control. A personal identification number fails quality control if it satisfies at least one of the following criteria:

- The number was flagged by Statistics Sweden as a duplicate (perhaps due to a clerical error at birth) and hence could not be reliably linked to a single individual.
- The dates of death in the *National Cause of Death Register* and the *Total Population Register* are mutually inconsistent.
- Year of birth or sex is missing for the individual in the *Total Population Register*.
- The individual was already deceased before the month of winning. Because individual bank accounts are not immediately closed after an individual is deceased, such patterns in the data need not reflect an error. However, we elected not to include these individuals because we do not observe any health outcomes for them after they win the lottery and because we suspect some of these observations may be data errors.
- The individual was not yet born or above age 110 at the time of winning.

IX.B. U.S. Mortality Gradients

Data Sources. We use version L of the processed HRS data files provided by RAND (Chien et al. 2013).

Sample Restrictions. The results are based on the AHEAD cohort that was interviewed in 1993. In the files provided by RAND, the AHEAD cohort is known as Wave 2. We only include respondents who were alive at the time of the interview (R2IWSTAT = 1).

Wealth. Wealth is calculated from the RAND variable H2ATOTA (total wealth excluding secondary residence), which measures net wealth. We convert wealth to units of million SEK at 2010 prices as follows:

1. We use the variables R2IWENDY and R2IWENDM (year/month of interview end date) to assign a reference date for reported wealth.
2. We use the monthly non-seasonally adjusted CPI series CPIAUCNS on FRED (2015) to convert wealth to 2010m1 USD.
3. We use the monthly SEK-to-USD exchange rate EXSDUS on FRED (2013) to convert 2010m1 USD to 2010m1 SEK.
4. We divide the resulting number by 1,000,000 to convert wealth to million SEK.

5. We winsorize wealth below the 1st and above the 99th percentile, replacing values below/above the threshold with the percentile values.

Adjusting for Household Size. Wealth as reported in HRS is at the household level. To obtain wealth per respondent, we divide by 2 whenever the observation is part of a couple household (H2CPL = 1).

Death Indicator. We compute indicators whether a respondent died within 1,2,...10 years after the interview end month/year. More specifically, we set ISDEADx = 1 whenever the year and month of death are not missing (RADFLAG = 1) and the month/year of death occurred before the (interview end month/year + x), where x in {1,2,...10} years.

Age/Gender Cells and Swedish Lottery Weights. We reweight the HRS sample based on the total amount of prize money for each age/gender cell (where age is measured in years at the interview end date, obtained from the R2AGEY_E RAND variable). For ages below 35, we collapse observations into two [age = 35]/male and [age = 35]/female cells in the lottery and HRS sample. The weight for these cells is the total prize money won by men/women with age <= 35.

IX.C. Wealth and Adult Health in Sweden

We estimate the Swedish gradients for adults using a representative sample of 50,000 individuals taken in year 2000. Wealth refers to net wealth according to the *Swedish Wealth Register* measured in year 2000 (see Section VI.D for details). Wealth is measured in million SEK and deflated to 2010 prices. For mortality, the mortality index and hospitalizations, the representative sample has been re-weighted according to the age and gender distribution of the full sample of lottery winners, and these health outcomes are measured in 2001 to 2010. Since drug prescriptions are only observed 2006-2010, we weight the representative according to the birth year and gender among the sample of lottery players that is used in drug prescription analyses, i.e. lottery players that won before 2006 and were alive in 2010. We use the same controls when estimating the gradients as in the lottery analyses and the controls are measured in 1999.

IX.D. Parental Income and Child Outcomes in Sweden

Some of the child outcomes are available for limited subsamples, and we therefore use all available representative samples to be able to estimate reliable gradients, i.e. all children to parents in seven Swedish representative samples of 50,000 people each (the representative

samples are taken in 1980, 1985, 1990, 1995, 2000, 2005, and 2010). Parental income is measured as the total disposable income of both parents during the first 10 years of the child's life expressed in million SEK (deflated to 2010 prices). Since age variability among children is low (all are below 18), we have not reweighted the representative sample to match the sample of lottery players' children. Parental income is available from 1978 to 2010, which implies that we are including children born between 1978 and 2000.

Since the different child outcome variables are only available during certain years and at certain ages (as described in Table AXXIII), the sample varies for each outcome variable. Infant health is measured at birth, whereas parental income in the ten years following birth. Child development variables are measured at the ages shown in Table AXXIII. Hospitalization data is available for all ages, but we chose to focus on children aged 10, which is the average age in our sample of pre-lottery children. This implies that five-year hospitalizations refer to hospitalizations during age 11-15 and two-year hospitalizations to hospitalization during age 11-12. Drug prescription is measured in 2006-2010 and we use the same age constraints as listen in Table AXXIII.

We include controls for birth demographics of the child, but no parental controls.

IX.E. Comparison to US Household-Income Gradients

Figure AVII shows how the mean of some selected child outcomes vary by income quintile in Sweden and the United States. In this section, we provide additional details about this comparative analysis. We focus on a selected set of key outcomes measured similarly in Swedish and US samples, and describe sample selection, the definition of the outcome variables, and other steps we took to maximize comparability. The US gradients are estimated using two different data sources: the *National Longitudinal Survey of Youth 1997 Cohort* (NLSY97) and the *Panel Study of Income Dynamics* (PSID). Below, we describe each cohort analysis in turn.

NLSY97. The NLSY97 is a nationally representative and longitudinal survey of youths born between 1980 and 1984 (Bureau of Labor Statistics 2015a). These youths were first surveyed in 1997 and have subsequently been surveyed on an annual basis. Of the 8,494 youth surveyed, 6,748 belong to what is conventionally referred to as the cross-sectional sample. This sample was designed to be representative of the US population. All our analyses are based on members of the cross-sectional sample, and we restrict the sample to Non-Hispanic Whites. Imposing these two restrictions leaves 4,807 individuals: 2,317 women and 2,490 men. Of

these, 872 are born in 1980, 1010 in 1981, 986 in 1982, 980 in 1983 and 959 in 1984.

The NLSY97 only contains information about a respondent's household income from 1997 and onward. In the first survey-wave of 1997, respondents 12 to 17 years old and it is therefore not feasible to measure household-income in the first ten years of a child's life (as is done in our main analyses). Our measure of household income is instead the average household income in the youth's household in all survey years up to and including the year the youth turns 19. We use the annual inflation data from the Consumer Price Index for All Urban Consumers (FRED 2015) to convert incomes to year-2010 $1,000. We were able to construct a household-income variable for the households of 4,552 out of the 4,807 non-Hispanic youths in the cross-sectional sample.

We consider three child outcomes available in NLSY97: BMI, Obesity and Cognitive Skills.

BMI and Obesity. We use data from the annual surveys conducted between 1997 and 2004 to construct these variables. In each wave, respondents were asked to report their height and body weight, allowing us to construct BMI. To maximize comparability with the Swedish analyses, we use the BMI variable in the year closest to the year in which the respondent turns 18 and we use 30 as the cutoff to define the obesity indicator. A BMI variable is available for 4,797 out of 4,807 respondents. For over the 99% of the respondents with a non-missing BMI variable, it was obtained in the year the respondent turned 17, 18 or 19.

Cognitive Skills. NLSY97 respondents were asked to take a computerized version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB). Overall, the test measures skills in 12 separate domains, but the raw score attained in each domain are not directly comparable across individuals because the item difficulty is adjusted dynamically as a function of the test-taker's performance. Conveniently, the NLSY supplies an aggregate percentile-ranking of each NLSY97 test-taker that is based on his or her overall performance in four domains: mathematics, arithmetic reasoning, word knowledge and paragraph comprehension. The percentile-ranking is conducted separately in respondents grouped into three-month age groups Bureau of Labor Statistics (Bureau of Labor Statistics 2015b). This variable is available for 3945 of the 4,807 members of our sample. The implied response rate (82%) similar to the overall response rate in the cross-sectional NLSY97 sample (81%).

Our measure is of Cognitive Skills is constructed by convoluting the percentile ranking with the inverse of the standard normal distribution, so that the estimated gradients can be interpreted as in population standard-deviation units, and demeaning the transformed variable.

Swedish Gradients. The main difference between the NLSY97 analyses and the Swedish

income gradients is that the NLSY97 sample only allows us to measure household-income fairly late during development. To construct a comparable Swedish income measure, we use all 35,898 children born between 1980 and 1984 to members of our representative samples. We calculate the average of the child's household-income, taking the average from the year the child turns 16 until the year they turn 19. We then split the sample into quintiles by household-income, and compare Swedish and US sample means for each of the three outcomes.

PSID. The Panel Study of Income Dynamics (PSID) is an American dynastic longitudinal household survey (Becketti et al. 1988). It was launched in 1968 with a nationally representative sample of 2,930 families and a low income oversample of 1,872 families. Individuals join the survey if they are born to or adopted by a current sample member.

To begin our analyses, we construct an identifier, hereafter "ID", by combining information about the PSID family from which they descend (INTNUM-68) and their unique sequence number within that family (PERNUM). The ID variable uniquely identifies each PSID member. In each year Y during which the survey is administered, household-level variables are assigned an identifier of the form (INTNUM-Y).

To track individuals across time, the PSID provides a cross-year index that maps the ID to a family-identifier (in the form INTNUM-Y). This allows us to link each member of a household, including children residing in the household, to variables such as household income. The main survey of the PSID was administered annually between 1968 and 1997, and biennially thereafter. The survey asks household members a detailed set of questions about various types of income, including income in years during which they were not interviewed (thus making it possible to generate longitudinal data series of annual income, despite the biennial surveying). From these responses, the PSID derives a household-income variable. We use this variable as our measure of household income in our analyses.

Our data on child outcomes are from two waves of the PSID Child Development Supplement (hereafter, CDS), an add-on survey intended to provide researchers with high-quality information about children's development (Hofferth et al. 1997). Wave 1 of the survey was administered in 1997 and all children aged 0 to 12 at the time were eligible to participate. Wave 1 contains 3,563 children (from the 88% of families who agreed to participate). In 2002, the PSID sought to re-interview the Wave 1 respondents, obtaining data from 2,907 children.

In total, 3,563 children participated in at least one of the waves, but as in the NLSY97 analyses we restrict the sample non-Hispanic White children. This restriction leaves us with 1,628 observations. From this sample, we drop children for whom we do not have information about the household-income in the first ten years of the child's life. This second restriction

leaves us with a sample of 1,024 non-Hispanic White children. In our analyses, we also use data on child's birth year, sex and race. These variables are all available in the CDS supplement.

We estimate gradients with respect to two variables: five-year "all-cause" hospitalization risk and birth weight.

Birth Weight. Wave 1 of the CDS asks each child's Primary Caregiver (PCG) to report the child's birth weight in pounds and ounces. For comparability with the Swedish analyses, we convert the reported birth weight to grams. We have 1,016 non-missing observations. Of these, 23 are based on birth weight rounded to the nearest pound because the PCG only supplied a birth weight rounded to the nearest pound.

Hospitalization. Wave I of the CDS also asks the PCG if the child has had an overnight hospitalization even since birth. If the answer is affirmative, the PCG is also asked to report the number of hospitalizations, and date and cause of the most recent stay. The date of the most recent hospitalization allows us to determine if the child has been hospitalized overnight in the past five years. In Wave 2, the same information is gathered about the time period since the previous survey data. For comparability with the Swedish gradients, we define our hospitalization variable over a five-year window. The procedure we use to generate our binary *Hospitalization* variable varies by child's birth year.

For children born 1991 or earlier, we use the Wave 1 (1997) data to define the *Hospitalization* variable. For these children, we set the variable to 1 if the child's PGS indicated that the child had an overnight hospitalization since 1992 and 0 if no hospitalization occurred. If there are missing data that prevent us from inferring if a hospitalization occurred in the past five years, we set the variable to missing. Because children born 1991 were aged 5-12 at the time of the survey 1997, the resulting variable is a 5-year "all-cause" hospitalization. The children were aged 0 to 7 at the beginning of the five-year window over which the variable is defined.

For children born in 1992 or later, we use Wave 2 responses to determine the value of the *Hospitalization* variable. We set the variable to 1 if the child is reported to have had an in-patient hospitalization event between 1997 and 2002 and 0 if the PGS indicates no such hospitalization event took place. The resulting variable is defined over the five year window 1997-2002 and hence comparable to our 5-year "all-cause" hospitalization. In 1997, the children for whom we use Wave 2 data to define this variable were aged 0 to 5.

Our final *Hospitalization* variable is available for 1,000 respondents; 804 born 1991 or earlier and 196 born 1992 or later. To investigate how average birth weights and hospitalization risk vary by income quintile, we used the Consumer Price Index for All Urban Consumers (FRED 2015) to convert the ten household-incomes to year-2010 $1,000. We then used the

1997 child-level demographic weights provided by the CDS supplement to estimate the quintiles of our 10-year household income variable. The quintile-specific means and standard deviations are also reweighted using these weights.

Swedish Gradients. To calculate quintile-specific averages that are comparable to the PSID estimates, we include all 69,612 children aged 5 to 12 in 2000 to members of our representative samples. We computed the gross household income in the first ten years of each child's life and calculated the quintile of this distribution for each birth cohort. We then split the sample into quintiles by household-income, and compare Swedish and US birth weight means and hospitalization probabilities across quintiles.

X. Bibliography

X.A. Additional Sources, PLS

Per Ekström, Documents and note from interview conducted in Stockholm on November l7, 20l0.

Johan Hansing, Documents and notes from interview conducted in Stockholm on September 9, 20l0.

Kungl. Maj:t, Proposition 1972:22. *Kungl. Maj:ts proposition med förslag om skattefrihet vid premiering av sparande, m.m*.

Regeringen, Proposition 1976/77:47. *Om sparstimulerande åtgärder*.

Regeringen, Proposition 1978/79:45. *Om villkoren för lönsparande på vinstsparkonto*.

Regeringen, Proposition 1980/81:59. *Om villkoren för lönsparande på vinstsparkonto*.

Regeringen, Proposition 1980/81:l67. *Om villkoren for lönsparande på vinstsparkonto*.

Regeringen, Proposition 1984/85:51. *Om bosparande for ungdomar och om andra sparfrågor*.

Regeringen, Proposition 1990/91:100. *Förslag till statsbudget för budgetåret 1991/92*.

Finansdepartementet, Regeringsbeslut 1985-07-26, 2821/85.

Finansdepartementet, Regeringsbeslut 1986-08-28, 3366/86.

Finansdepartementet, Regeringsbeslut 1989-10-05, 3176/89.

Finansdepartementet, Regeringsbeslut 1991-04-18, 3499/90.

Finansdepartementet, Regeringsbeslut 1993-12-16, 4849/93.

Finansdepartementet, Regeringsbeslut 1994-09-15, 4849/93.

X.B. References

Becketti, Sean, William Gould, Lee Lillard, and Finis Welch. "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation," *Journal of Labor Economics*, 6 (1988), 472-492.

Behrman, Richard E., and Adrienne Stith Butler, ed., *Preterm Birth: Causes, Consequences, and Prevention*, (Washington, DC: National Academies Press, 2007).

Bureau of Labor Statistics, U.S. Department of Labor, "National Longitudinal Survey of Youth 1997," available at https://www.nlsinfo.org/content/cohorts/nlsy97 (accessed 30 September, 2015a).

Bureau of Labor Statistics, U.S. Department of Labor, "Appendix 10: CAT-ASVAB Scores," available at https://www.nlsinfo.org/content/cohorts/nlsy97/other-documentation/codebook-supplement/appendix-10-cat-asvab-scores (accessed 30 September, 2015a).

Chien, Sandy, Nancy Campbell, Orla Hayden, Michael Hurd, Regan Main, Josh Mallett, Craig Martin, Erik Meijer, Angela Miu, Michael Moldoff, Susann Rohwedder, and Patricia St. Clair, *RAND HRS Data Documentation, Version M*, (Santa Monica, CA: Version M, RAND Center for the Study of Aging, 2013).

Cnattingius, Sven, Anders Ericson, Jan Gunnarskog, and Bengt Källén, "A Quality Study of a Medical Birth Registry," *Scandinavian Journal of Social Medicine*, 18 (1990), 143-148.

Federal Reserve Bank of St. Louis, FRED, "Sweden/U.S. Foreign Exchange Rate [EXSDUS], 2013", available at http://research.stlouisfed.org/fred2/series/EXSDUS (accessed October 21, 2013).

Federal Reserve Bank of St. Louis, FRED, "U.S. Bureau of Labor Statistics, Consumer Price

Index for All Urban Consumers: All Items [CPIAUCNS]", available at http://research.stlouisfed.org/fred2/series/CPIAUCNS/ (accessed September 6, 2015).

Hofferth, Sandra, Pamela E. Davis-Kean, Jean Davis, and Jonathan Finkelstein, *The Child Development Supplement to the Panel Study of Income Dynamics: 1997 User Guide,* (Ann Arbor, MI: Survey Research Center, Institute for Social Research, 1997).

Johansson, Lars. A., Charlotte Björkenstam, and Ragnar Westerling, "Unexplained Differences between Hospital and Mortality Data Indicated Mistakes in Death Certification: An Investigation of 1,094 Deaths in Sweden During 1995," *Journal of Clinical Epidemiology*, 62 (2009), 1202-1209.

Kaplan, H. Roy, *Lottery winners: How they won and how winning changed their lives*, (New York, NY: Harper & Row, 1978).

Ludvigsson, Jonas F., Eva Andersson, Anders Ekbom, Maria Feychting, Jeong-Lim Kim, Christina Reuterwall, Mona Heurgren, and Petra Otterblad Olausson, "External Review and Validation of the Swedish National Inpatient Register," *BMC Public Health*, 11 (2011), 450.

Socialstyrelsen, *Utvärdering av det svenska Medicinska födelseregistret*, (Stockholm, SE: Socialstyrelsen, 2002).

Socialstyrelsen, *Dödsorsaker 2004 – Sveriges officiella statistik*, (Stockholm, SE: Socialstyrelsen, 2007).

Socialstyrelsen, *Kvalitet och innehåll i patientregistret –Utskrivningar från slutenvården 1964–2007 och besök i specialiserad öppenvård (exklusive primärvårdsbesök) 1997–2007*, (Stockholm, SE: Socialstyrelsen, 2009).

Socialstyrelsen, *Dödsorsaksstatistik – Historik, produktionsmetoder och tillförlitlighet*, (Stockholm, SE: Socialstyrelsen, 2010).

Socialstyrelsen, *Hjärtinfarkter 1987–2011 – Sveriges officiella statistik*, (Stockholm, SE:

Socialstyrelsen, 2012a).

Socialstyrelsen, *Förstudie kring individdata för rekvisitionsläkemedel*, Stockholm, SE: Socialstyrelsen, 2012b).

Socialstyrelsen, ”Konverteringstabell mellan ICD-8/ICD-9/ICD-8”, 2013a,. available at http://www.socialstyrelsen.se/klassificeringochkoder/diagnoskoder/konverteringstabeller. (accessed October 21, 2013).

Socialstyrelsen, *Dödsorsaker 2012 – Sveriges officiella statistik*, (Stockholm, SE: Socialstyrelsen, 2013b).

Socialstyrelsen, *Sjukdomar i sluten vård 1987–2011 – Sveriges officiella statistik*, (Stockholm, SE: Socialstyrelsen, 2013c).

Socialstyrelsen, *Läkemedel, statistik för år 2012 – Sveriges officiella statistik*, (Stockholm, SE: Socialstyrelsen, 2013d).

Socialstyrelsen, ”Läkemedelsregistret – Bortfall och kvalitet”, 2013e, available at http://www.socialstyrelsen.se/register/halsodataregister/lakemedelsregistret/bortfallochkvalitet, (accessed 19 November, 2013e).

Socialstyrelsen, *Graviditeter, förlossningar och nyfödda barn – Sveriges officiella statistik*, (Stockholm, SE: Socialstyrelsen, 2013f).

Stagg, Vicki, “CHARLSON: Stata module to calculate Charlson index of comorbidity,” Boston College, Department of Economics, 2006, available at https://ideas.repec.org/c/boc/bocode/s456719.html (accessed 30 September, 2015).

Statistiska centralbyrån, *Förmögenhetsstatistik 2004 – Sveriges officiella statistik*, (Stockholm, SE: Statistiska centralbyrån, 2006).

Vlachos, Jonas, “Betygets värde,” Konkurrensverket, Uppdragsforskningsrapport 2010:6, 2010.

Waldenström, Daniel, "Hur ska framtidens förmögenhetsstatistik se ut?," *Ekonomisk Debatt*, 35 (2008), 72-77.

Wettermark, Björn, Niklas Hammar, C. Michael Fored, Andrejs Leimanis, Petra Otterblad Olausson, Ulf Bergman, Ingemar Persson, Anders Sundström, Barbro Westerholm, and Måns Rosén, "The New Swedish Prescribed Drug Register – Opportunities for Pharmacoepidemiological Research and Experience from the First Six Months," *Pharmacoepidemiology and Drug Safety*, 16 (2007), 726-735.