Combining the Individual Health Survey with the Aggregate Mortality Data to Estimate the Disability-Free Life Expectancy

Kosuke Imai

Department of Politics
Princeton University

November 28, 2007

- Imai, Kosuke, and Samir Soneji. (2007). "On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and Its Extension." *Journal of the American Statistical Association*. Vol. 102, No. 480 (December), pp. 1199-1211.

- Paper, computer code, and replication data available at `http://imai.princeton.edu/`

Ecological Inference in the Social Sciences

- Ecological fallacy (Robinson, 1950): Aggregate correlation between race and literacy rate is -0.733, while individual correlation is -0.339.

	Black	White	
Literate	?	?	Y_i
Illiterate	?	?	$1 - Y_i$
	X_i	$1 - X_i$	

  where i represents a region, and X_i and Y_i are observed proportions.

- Applications in political science:
  - Voting rights, racial voting, redistricting
  - Study of voting behavior in Nazi Germany, Iraq, etc.
  - Split ticketing

Combining Aggregate and Individual Data

- Aggregate mortality data:
  - death counts from vital statistics
  - population for each age cohort from census

- Individual health survey data:
  - A battery of questions about health and disability
  - National Health Interview Survey, the Medicare Current Beneficiary Survey, and the National Health and Nutrition Examination Survey

- Goal: Use both aggregate and individual data to estimate the disability-free life expectancy (DFLE)

What is Disability-Free Life Expectancy (DFLE)?

- Life Expectancy $\equiv$ Expected number of years of life
- DFLE $\equiv$ Expected number of years of life spent in good health
- Research questions:
  - People live longer now, but are additional years of life spent in good health?
  - Is life expectancy increasing faster than disability rates are decreasing?
  - How does DFLE differ across gender and racial groups?
- Mortality-morbidity debate:
  - Decline in mortality rates only reflects a decline in the fatality rate of chronic diseases (Gruenberg, 1977; Kramer, 1980).
  - If the onset of the chronic condition can be postponed, morbidity will be compressed into a shorter period of time (Fries, 1980).
  - Decline in mortality rates leads to the increased prevalence of milder chronic diseases (Manton, 1982).

Sullivan's Method

- Sullivan (1971) proposed a method to estimate DFLE from period life table and consecutive cross-section disability survey
  - Applicable when longitudinal data are NOT available
  - DIFFICULT to follow a cohort from birth to death

- Academic researchers use it to estimate:
  - DFLE by socioeconomic status (Molla et al. 2004)
  - DFLE by educational levels (Bronnum-Hansen et al. 2004)
  - DFLE by occupational groups (Bronnum-Hansen et al. 2000)
  - DFLE over time (Graham et al. 2004; Crimmins et al 1989)
  - DFLE by chronic disease status (Manuel & Schultz, 2004)
  - DALY across different regions in the world (Murray & Lopez, 1996)
  - DALY for 191 member states of WHO (Musgrove et al. 2000)

The Problems with Sullivan's Method

- Sullivan (1971) provided no formal justification
- Conflicting results over required assumptions:
  - Transition probability from healthy to disabled must be 'large' (Newman 1988)
  - Recovery probability must be negligible (Palloni et al 2005)
  - Mortality risk of the disabled and healthy must be homogeneous (Palloni et al 2005)
- Controversies:
  - Underestimates DFLE due to bias in the estimation of disability prevalence (Rogers et al. 1990)
  - Bias due to non-stationary population (Mathers 1991)
  - Does the method allow for transitions from disabled to healthy? (Barendregt et al. 1994,95; Van De Water et al. 1995)

Overview of the Paper

1. Identify the key assumptions of Sullivan's method
2. Derive a large-sample variance that takes into account for all the estimation uncertainty
3. Show that these assumptions are unlikely to hold in real data
4. Propose an extension to Sullivan's method which relaxes these assumptions
5. Applications to the 1907 and 1912 US Birth Cohorts

Theoretical Definition of Life Expectancy and DFLE

- Life expectancy at age $x \in [0, \infty)$ for a cohort born at time y

$$e(x,y) = \frac{1}{l(x,y)} \int_x^\infty l(t,y)\, dt$$

  where $l(x,y)$ is the survival function

- DFLE at age x for this cohort

$$e^{DF}(x,y) = \frac{1}{l(x,y)} \int_x^\infty [1 - \pi(t,y)]\ l(t,y)\, dt$$

  where $\pi(x,y)$ is the proportion disabled at age x for this cohort

Period Life Table

- Theoretical definitions given in a continuous-time framework
- Data come in a discrete form: Period (rather than cohort) life table

- Age intervals $[x, x + n_x)$: the last interval is $[\omega, \infty)$
- ${}_{n_x}P_x$: mid-year population (Census)
- ${}_{n_x}D_x$: total number of deaths (Vital statistics)
- ${}_{n_x}a_x$: average person-years lived in the interval among those dying in the interval (Complete life tables)

- Stationarity Assumptions of Period Life Table:
  1. The age-specific hazard rate is constant over time: $\mu(x,y) = \mu(x)$
  2. The birth rate is constant over time
  3. The net migration rates are 0 at all ages

Estimating Life Expectancy from Period Life Table

- Life expectancy for a hypothetical cohort

$$e_x = \frac{1}{l_x} \sum_{i \in \mathcal{A}_x} {}_{n_i}L_i$$

  - ${}_{n_x}L_x$: total number of person-years in the interval $[x, x + n_x)$

$$ {}_{n_x}L_x = n_x l_{x+n_x} + l_x \, {}_{n_x}q_x \, {}_{n_x}a_x $$

  - $\mathcal{A}_x = \{i \in \mathcal{A} : x \leq i\}$ where $\mathcal{A}$ is a set of starting ages for all age intervals
  - ${}_{n_x}q_x$: conditional probability of death in the interval
  - l_x: proportion of survivors at age x

- It can be shown that $e_x = e(x)$ under the stationarity assumptions

1999 US Period Life Table

Age	l_x	q_x	a_x	L_x	e_x
20	0.986	0.001	0.506	0.986	55.851
25	0.982	0.001	0.500	0.981	51.101
30	0.977	0.001	0.495	0.976	46.338
35	0.971	0.001	0.500	0.970	41.597
40	0.963	0.002	0.500	0.962	36.916
45	0.952	0.003	0.500	0.950	32.323
50	0.935	0.004	0.500	0.933	27.853
55	0.911	0.007	0.499	0.908	23.516
60	0.875	0.011	0.501	0.870	19.391
65	0.820	0.016	0.500	0.813	15.499
70	0.743	0.025	0.500	0.733	11.839
75	0.638	0.038	0.500	0.626	8.356
80	0.505	0.059	0.500	0.490	4.873
85+	0.345	1.000	0.951	0.328	0.951

Sullivan's Method

- Sullivan's Estimator:

$$\hat{e}_x^{DF} \equiv \frac{1}{l_x} \sum_{i \in \mathcal{A}_x} (1 - {}_{n_i}\hat{\pi}_i)\, {}_{n_i}L_i$$

  where ${}_{n_i}\hat{\pi}_x$ is the sample fraction of the disabled survey respondents within the age interval $[i, i + n_i)$.

- The standard variance estimator of $\hat{e}_x^{DF}$ assumes that the mortality rate is known (rather than estimated)

Statistical Foundation of Sullivan's Method

1. Consistency and Unbiasedness of Sullivan's Method
   - Consistent (but not unbiased) if the mortality rate is estimated
   - Required assumptions: 3 stationary assumptions of period life table, and the stationarity of age-specific disability prevalence $\pi(x, y) = \pi(x)$ for all y
   - Resolves controversies
   - Corrects Sullivan's original formula for ${}_{n_x}\hat{\pi}_x$
2. Derivation of consistent and unbiased variance estimators
3. Derivation of large-sample variance when the mortality rate is estimated rather than assumed to be known
4. A consistent estimator of this large-sample variance

Age	Estimated Disability Prevalence	Estimated DFLE	$_nm_x$ Known 95% C.I. Lower	$_nm_x$ Known 95% C.I. Upper	$_nm_x$ Estimated 95% C.I. Lower	$_nm_x$ Estimated 95% C.I. Upper
20	0.01	54.47	54.40	54.53	54.37	54.56
25	0.00	49.73	49.67	49.79	49.64	49.82
30	0.00	44.98	44.91	45.04	44.89	45.07
35	0.01	40.25	40.19	40.32	40.16	40.34
40	0.01	35.59	35.53	35.66	35.51	35.68
45	0.01	31.02	30.96	31.09	30.94	31.10
50	0.01	26.58	26.52	26.64	26.50	26.66
55	0.02	22.26	22.19	22.32	22.18	22.34
60	0.01	18.17	18.11	18.24	18.10	18.25
65	0.03	14.31	14.24	14.37	14.24	14.38
70	0.05	10.70	10.64	10.77	10.64	10.77
75	0.08	7.30	7.24	7.37	7.24	7.37
80	0.15	4.05	3.99	4.11	3.99	4.11
85	0.20	0.77	0.75	0.78	0.74	0.79

Extension of Sullivan's Method

- Avoid the tenuous assumption of stationary mortality and disability
- Popular approach: Multi-state life table method
- Models transition probabilities
- Requires a large-scale longitudinal disability survey
- Theoretical assumptions:
  1. Transition probabilities follow continuous-time first-order Markov process
  2. Functional form assumption about the average person-years spent in a state
  3. Additional assumptions are also required in practice

- Cohort DFLE: Extend Sullivan's method to a cohort life table and consecutive cross-section surveys
- Weaker assumptions, less stringent data requirements

Bounding the Cohort DFLE

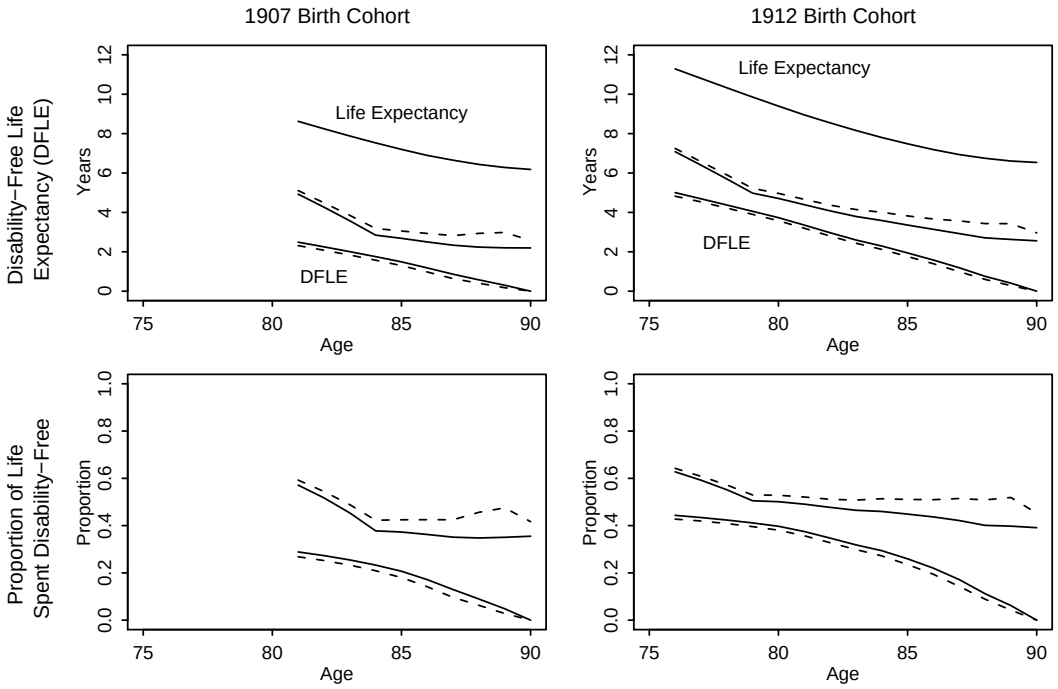
- Disability surveys may not cover all age groups
- Cohort DFLE is not identified
- Tighter bounds with monotonicity assumptions:
  1. Disability prevalence of a given birth cohort in the last age interval of interest is greater than or equal to that of the preceding interval
  2. Disability prevalence of a given birth cohort in the first age interval of interest is less than or equal to that of the next interval
- Can bound the cohort DFLE
- Consistent estimation of bounds is possible
- Beran (1988)'s bootstrap method to construct asymptotically exact and balanced CI

Data

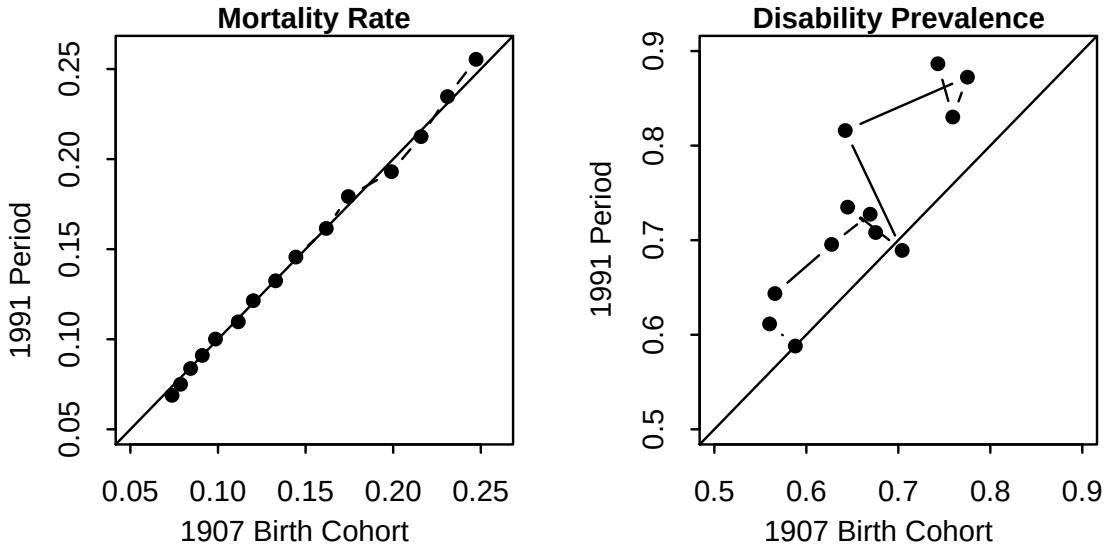
- Self-Reported Disability: at least one Activity of Daily Living deficiency (bathing, dressing, getting in or out of bed, using the toilet, and eating).
- Data
  - Mortality: 1988 to 2003 Death Counts (US Vital Statistics)
  - Population: 1988 to 2003 Census Estimates (US Census Bureau)
  - Disability: 1991 to 2003 US Medicare Current Beneficiary Survey

	Mortality Data		Disability Data	
Birth Cohort	From	To	From	To
1907	81	96	84	96
1912	76	91	79	91

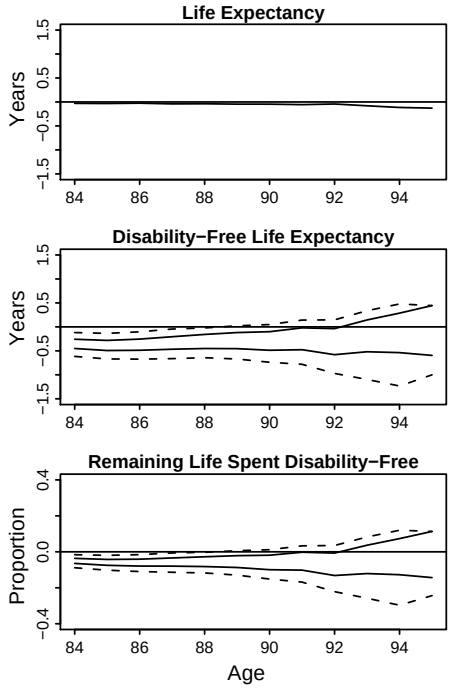
DFLE and Proportion of Life Spent Disability-Free



Stationary Mortality and Non-stationary Morbidity



Cohort (1907) Versus Period (1991) DFLE



Concluding Remarks

- We establish a statistical foundation of Sullivan's method:
  - Under stationarity assumptions, Sullivan's method is valid
  - We derive a large-sample variance accounting for additional uncertainty about the mortality rate
- We extend Sullivan's method to cohort life tables:
  - No stationarity assumptions required
  - Avoid strong assumptions about transition probabilities that are necessary for multi-state methods
  - Only repeated cross-section disability data are required
- Empirical analyses of two birth cohorts:
  - DFLE may not have been increasing as fast as life expectancy
  - Nearly stationary mortality but non-stationary disability
  - Significant cohort and period differences in DFLE, especially at older ages