American Journal of Epidemiology

Vol. 186, No. 2
DOI: 10.1093/aje/kwx003
Advance Access publication:
June 9, 2017

Practice of Epidemiology

The Identification of a Threshold of Long Work Hours for Predicting Elevated Risks of Adverse Health Outcomes

Sadie H. Conway*, Lisa A. Pompeii, David Gimeno Ruiz de Porras, Jack L. Follis, and Robert E. Roberts

* Correspondence to Dr. Sadie H. Conway, Department of Epidemiology, Human Genetics, and Environmental Sciences at the School of Public Health, University of Texas Health Science Center at Houston, 1200 Pressler Street, RAS W610, Houston, Texas 77030, USA (e-mail: sadie.conway@uth.tmc.edu).

Initially submitted April 13, 2016; accepted for publication August 25, 2016.

Working long hours has been associated with adverse health outcomes. However, a definition of long work hours relative to adverse health risk has not been established. Repeated measures of work hours among approximately 2,000 participants from the Panel Study of Income Dynamics (1986–2011), conducted in the United States, were retrospectively analyzed to derive statistically optimized cutpoints of long work hours that best predicted three health outcomes. Work-hours cutpoints were assessed for model fit, calibration, and discrimination separately for the outcomes of poor self-reported general health, incident cardiovascular disease, and incident cancer. For each outcome, the work-hours threshold that best predicted increased risk was 52 hours per week or more for a minimum of 10 years. Workers exposed at this level had a higher risk of poor self-reported general health (relative risk (RR) = 1.28; 95% confidence interval (CI): 1.06, 1.53), cardiovascular disease (RR = 1.42; 95% CI: 1.24, 1.63), and cancer (RR = 1.62; 95% CI: 1.22, 2.17) compared with those working 35–51 hours per week for the same duration. This study provides the first health risk–based definition of long work hours. Further examination of the predictive power of this cutpoint on other health outcomes and in other study populations is needed.

cancer; cardiovascular disease; long work hours; self-reported general health; work hours; work schedule tolerance

Abbreviations: AUC, area under the receiver operating characteristics curve; CVD, cardiovascular disease; LWH, long work hours; PSID, Panel Study of Income Dynamics; SRGH, self-reported general health; WH/week, work hours per week.

Evidence that exposure to long hours of work increases the risk of deleterious health conditions has been documented in the scientific literature for more than 50 years (1). A growing body of evidence has demonstrated positive associations between long work hours (LWH) and numerous health conditions of varying etiologies, including poor self-reported general health (SRGH) (2), cardiovascular disease (CVD) (3), musculoskeletal disorders (4), work-related injuries (5, 6), depression (7), and sleep disruption (8). However, the number of hours that place a worker at risk has not been established. Prior studies have characterized LWH in terms of overtime (9, 10), extended shift hours (11), or daily (12) or weekly (13) hours of work, which have been categorized inconsistently (3, 14, 15). Much of the LWH literature has defined the exposure as more than 40 work hours per week (WH/week) or more than 8 hours per day. Typically, no justification is provided for the use of these thresholds, with the exception that 40 hours of work per week is a widely-accepted labor standard for full-time work. Inconsistencies in the application of work-hours cutpoints limit comparisons across studies.

A validated threshold for demarcating LWH risk categories has been called for by numerous researchers (16–19). Harrington (19) proposed that increased risk might occur from working more than 48 to 56 hours per week. In response, Folkard argued that the existing evidence was not adequate to establish specific recommendations but conceded that the newly established 48-hour work week set forth in the European Community Working Time Directive was acceptable, given what was known (Simon

Folkard, University of Wales, unpublished data, 1994). Despite this, 2 recent reviews of LWH—one of which focused on cardiovascular outcomes (20) and the other on a variety of health outcomes (14)—used 40 hours of work per week as the boundary between standard and extended work hours, with standard hours serving as the reference group, without providing justification for the 40-hour boundary from a health standpoint. An improved analytical approach that more accurately assesses the risk associated with different durations of work hours is warranted. However, there is currently little evidence on the predictive validity of different LWH cutpoints relative to adverse health risk.

Predictive cutpoints, which estimate the probability of future health risks by maximizing classification accuracy over time, are widely used to designate risk strata (21) and have been applied to such diverse exposures as cancer biomarkers (22), obesity (23), and fall-risk indices (24). The identification of a statistically derived predictive cutpoint requires evaluating a set of possible cutpoints using measures of calibration and discrimination. Unlike diagnostic testing, which characterizes a cutpoint's ability to discriminate patients with a disease from those without, predictive models employ measures of calibration in addition to discrimination in order to evaluate the agreement between predicted probabilities and actual observed risk (25).

To address some of these existing limitations, we conducted a retrospective cohort study among US workers to identify the statistically optimized LWH exposure cutpoints that best predicted 3 adverse health outcomes independent of other factors (21, 26): 1) poor SRGH, which is widely used as a composite measure of overall health and has been shown to be a strong predictor of mortality (27–29); 2) incident CVD, the leading cause of death worldwide (30); and, 3) incident cancer, for which the limited evidence on the relationship with LWH is mixed (31–33), with 2 recent cohort studies reporting conflicting results (31, 32). The goal of this exploratory study was to identify working-hour thresholds with the greatest predictive ability for 3 chronic conditions with varying etiologies, with an emphasis on assessing the utility of our approach.

METHODS

Study overview

We examined the effects of LWH on SRGH, incident CVD, and incident cancer using data from the Panel Study of Income Dynamics (PSID) during 1986–2011. The PSID is an ongoing longitudinal survey of a representative sample of US households with interviews performed annually from study inception in 1968 until 1997 and biennially since 1997 to the present (34, 35). The PSID has captured work-hours data since 1968, self-reported general health data since 1984, and responses to a health survey pertaining to CVD and cancer since 1999.

The work-hours exposure variable was constructed from data collected in each interview year on the number of jobs the respondent worked, including start and stop dates for up to 4 jobs. For each job, data were collected on: 1) the average number of hours worked per week; 2) the average number of hours of overtime worked per week; and 3) the number of weeks not worked due to illness, vacation, strike, unemployment, or layoff. These data were used to calculate the average number of hours worked per week in the previous year on all jobs by dividing the total annual hours spent working per given year by 50 weeks, according to the definition by the US Bureau of Labor Statistics of an annual full-time worker (36). From this calculation, we determined the mean of the average number of hours worked per week for each participant across their period of time in the study. Work-hours data for participants reporting incident cases were censored in the year in which the incident case was reported as well as for all subsequent years because they were no longer considered to be at risk.

This study protocol was approved by the University of Texas Health Science Center Committee for the Protection of Human Subjects.

Construction of study cohorts

Participants were included in the study if they reported being at least 18 years of age at study baseline (1986) and reported at least 10 total years of nonzero work-hours data over the 25-year study period, either consecutively or nonconsecutively (n = 8,251). Individuals who dropped out of the study prior to the start of the disease outcome reporting period in 1999 and for whom outcome data was never collected (n = 1,696), individuals who dropped out of the study after 1999 but before reporting any outcome data (n = 1,342), and individuals who did not provide adequate data for determining outcome prevalence or incidence for any outcome (n = 1,764) were excluded from the analytical cohort. This remaining sample (n = 3,449) was used to develop 3 outcome-specific cohorts for SRGH, CVD, and cancer. Due to health differences between part-time and full-time workers, we further restricted the outcome-specific study cohorts to full-time workers (i.e., those averaging ≥35 work hours per week over their study observation period) (37, 38). We also excluded those who reported developing an outcome but who did not report work hours for a total of at least 10 years prior to reporting the incident case.

SRGH cohort. Starting in 1986, SRGH was captured in each interview, with participants asked to categorize their health, in general, as "excellent, very good, good, fair, or poor." We dichotomized SRGH as good (excellent, very good, good) or poor (fair or poor). An incident case of poor SRGH was defined as the first report of poor SRGH following at least 1 previous report of good SRGH and no previous reports of poor SRGH. Prevalent cases were defined as those whose first reported SRGH status was poor SRGH, including those reporting poor SRGH at baseline.

The final SRGH cohort was comprised of 2,206 workers, with 766 reports of incident poor SRGH (35%). From the starting sample of 3,449, we excluded 174 participants who were missing SRGH data over the study period, 170 prevalent cases, 309 participants who did not have at least 10 years of work-hours exposure, and 590 who worked part-time.

CVD cohort. The observation period for CVD began in 1999, with data collected biennially thereafter. Participants were asked whether a doctor had ever diagnosed them as having had a heart attack, high blood pressure, hypertension,

coronary heart disease, angina, congestive heart failure, or stroke. A dichotomous CVD index variable was developed for each interview year, with affirmative answers to any CVD-related question defined as a positive CVD outcome in that year. Incident CVD was defined as first reports of any of the CVD-related outcomes following negative responses to all CVD questions in the previous year. Prevalent cases were defined as those who indicated having previously diagnosed CVD at the start of the observation period in 1999.

The final CVD cohort included 1,698 participants, with 777 cases of incident CVD (45.8%). From the starting sample of 3,449, we excluded 305 participants who were missing CVD data over the study period, 657 prevalent cases, 45 participants who worked less than 10 years over the study period, and 744 who worked part-time.

Cancer cohort. Cancer outcomes were treated similarly to CVD, with the observation period beginning in 1999 when participants were asked whether a doctor had ever told them that they had cancer or a malignant tumor, excluding skin cancer. Cases were defined by an affirmative response to this question in any year. Incident cancer was defined as affirmative reports of cancer following the previous year's report of not ever having been diagnosed with cancer. A prevalent case was defined as an affirmative response to ever having cancer at the start of the observation period in 1999.

The final cancer cohort was comprised of 2,196 participants, with 263 cases of incident cancer (12.0%). From the starting sample of 3,449, we excluded 191 participants who were missing cancer data, 82 prevalent cases, 172 participants who worked less than the 10-year requirement, and 808 who worked part-time.

Statistical analysis

A series of univariate tests of model fit, calibration, and discrimination were used to evaluate a comprehensive set of potential work-hours cutpoints from 36 WH/week to 65 WH/week. Small cell sizes precluded the analysis of cutpoints above 65 WH/week. At each cutpoint, participants were defined as exposed to LWH if the mean of their hours worked per week across the study duration was equal to or greater than the cutpoint being tested for the specific outcome being examined.

Model fit was assessed for each work-hours cutpoint using the Bayesian information criterion, which does not require that models being compared also be nested. Model calibration was evaluated using Somers *D* statistics, which provided a measure of the performance of each LWH cutpoint as a predictor of the study outcomes (39), and relative risks for each univariate cutpoint model were calculated using Poisson regression, which indicated the probability of the outcome given the exposure of each LWH cutpoint (25). To account for unequal exposure durations, an offset variable was constructed by dividing the number of waves of work-hours data that were reported by the number of possible study waves. One-sided 95% confidence intervals were calculated for all *D* statistics and relative risks; 1-sided analyses were performed because the study question focused on the potential effects of a positive association between LWH and adverse health outcomes. Model discrimination was assessed by calculating the Youden index (*J*) for each LWH exposure cutpoint model, which was maximized at the work-hours threshold with optimized differentiating ability in terms of the outcome, giving equal weight to sensitivity and specificity. For the purpose of this study, the statistically optimized cutpoint was defined as the cutpoint that maximized the Youden index, which is a commonly used measure of the overall effectiveness of a cutpoint dichotomized from a continuous exposure (26, 40, 41). Additional measures of discrimination included sensitivity, defined as the percentage of the total number of participants with the outcome who were classified as exposed to LWH at the cutpoint being tested; specificity, defined as the percentage of the total number of participants without the outcome who were classified as unexposed; the area under the receiver operating characteristics curve (AUC), which indicated how well each cutpoint separated the cohorts into those with and without the outcome; positive and negative predictive values, which indicated the probability of disease among participants who were exposed or unexposed (42); and likelihood ratios, both positive and negative, which represented the probability of exposure among participants with the outcome (i.e., positive likelihood ratio) and the probability of exposure among participants without the outcome (i.e., negative likelihood ratio) (43). Likelihood ratios provide an alternative to the *J* statistic for identifying statistically optimized cutpoints (i.e., the cutpoints that maximize the positive likelihood ratio and/or minimize the negative likelihood ratio), and agreement between the likelihood ratios and the *J* statistic was noted (42, 44, 45). Additionally, we analyzed the exposure as a continuous variable to determine the statistically optimized cutpoint had it not been restricted to a natural number. A post-hoc exploratory analysis of the relative risks and *D* statistic confidence intervals and beta coefficients was performed to assess statistical differences among the identified measures of calibration and discrimination (46, 47).

We conducted univariate analyses to describe those who did and did not work at or above the LWH threshold. We conducted post-hoc descriptive analyses to examine the degree to which the study cohorts overlapped in terms of the study outcomes. We also examined whether differences existed between respondents and nonrespondents with respect to hours worked per week. All analyses were performed using Stata/SE, version 13.1 (StataCorp LP, College Station, Texas).

RESULTS

When combined, the 3 cohorts represented a total of 2,306 participants, of whom 62.0% were male and 72.6% were white and non-Hispanic (results not shown). At study baseline, participants had a mean age of 32.8 (standard error, 0.2) years and the majority (52.6%) had attained at least some level of postsecondary education. The combined population reported working an average of 44.5 (standard error, 0.2) hours per week over the study duration. Almost half of the study population (48.0%, $n = 1{,}107$) reported at least 1 of the outcomes of interest, with 29.9% ($n = 690$) of respondents reporting only 1 outcome, 16.1% ($n = 373$) reporting 2 outcomes, and 1.9% ($n = 44$) reporting all 3 outcomes.

The model fit and calibration information on LWH exposure and the outcome of poor SRGH status is shown in Table 1.

The cutpoint of 52 WH/week demonstrated the best model fit, yielded the only statistically significant *D* coefficient, and produced the only relative risk that reached statistical significance (relative risk = 1.28; 95% confidence interval: 1.06, 1.53). All measures of discrimination also were maximized at 52 WH/week, although 51 WH/week performed equally well in terms of the AUC.

For the outcome of CVD (Table 2), the work-hours cutpoints of 52 through 56 WH/week demonstrated equivalent model fit. The *D* statistic was maximized at 52 WH/week, with significant relationships seen for exposure cutpoints from 49 WH/week to 62 WH/week. An increased risk of incident CVD was demonstrated for cutpoints from 50 WH/week to 63 WH/week. Although several cutpoints maximized the various measures of discrimination, 52 WH/week was the best-performing cutpoint in terms of discrimination, as it was the only cutpoint that maximized the *J* statistic and was one of 3 cutpoints that maximized the AUC and minimized the negative likelihood ratio.

Table 1. Measures[a] of Model Fit, Calibration, and Model Discrimination for Long Work Hours and Self-Reported General Health (n = 2,206; Cases[b] = 766), Panel Study of Income Dynamics, United States, 1986–2011

Cutpoint (Weekly Work Hours)	Proportion Above Cutpoint, %	Measure of Model Fit and Calibration				Model Discrimination							
		Bayesian Information Criterion	RR[c]	95% CI[c]	Somers *D* Statistic	Youden Index (*J*)	Sensitivity, %	Specificity, %	AUC (c)	LR (+)	LR (–)	% PPV	% NPV
36	95.5	5.47E+07	0.87	0.64, 1.19	−0.0062	0.000	95.0	4.3	0.50	0.99	1.15	34.6	62.0
37	88.4	5.46E+07	0.77	0.63, 0.94	−0.0277	0.000	87.5	11.1	0.50	0.98	1.13	34.3	62.5
38	81.5	5.47E+07	0.88	0.74, 1.05	−0.0138	0.003	81.7	18.6	0.50	1.00	0.98	34.8	65.7
39	74.0	5.47E+07	0.90	0.77, 1.06	−0.0134	0.000	73.1	25.5	0.50	0.98	1.05	34.3	64.1
40	65.2	5.47E+07	0.91	0.78, 1.05	−0.0141	0.000	63.8	34.1	0.50	0.97	1.06	34.0	63.9
41	57.1	5.46E+07	0.85	0.74, 0.99	−0.0333	0.000	54.7	41.6	0.50	0.94	1.09	33.2	63.3
42	50.3	5.46E+07	0.87	0.75, 1.01	−0.0263	0.000	47.8	48.4	0.50	0.93	1.08	33.0	63.6
43	44.8	5.47E+07	0.95	0.82, 1.10	0.0006	0.000	43.9	54.8	0.50	0.97	1.03	34.0	64.7
44	40.0	5.47E+07	0.94	0.81, 1.08	−0.0054	0.000	39.3	59.6	0.50	0.97	1.02	34.1	64.9
45	35.5	5.47E+07	0.92	0.79, 1.07	−0.0103	0.000	34.3	63.9	0.50	0.95	1.03	33.6	64.7
46	31.3	5.47E+07	0.90	0.78, 1.05	−0.0148	0.000	30.3	68.2	0.50	0.95	1.02	33.6	64.8
47	27.7	5.47E+07	0.94	0.81, 1.09	−0.0052	0.006	28.1	72.5	0.50	1.02	0.99	35.2	65.5
48	24.0	5.47E+07	1.00	0.86, 1.17	0.0097	0.022	25.5	76.8	0.51	1.10	0.97	36.8	66.0
49	21.0	5.47E+07	1.01	0.86, 1.19	0.0093	0.026	22.7	79.9	0.51	1.13	0.97	37.5	66.0
50	17.7	5.47E+07	1.05	0.88, 1.25	0.0135	0.029	19.6	83.3	0.51	1.17	0.97	38.4	66.1
51	15.0	5.46E+07	1.12	0.94, 1.34	0.0223	0.032	17.1	86.1	0.52	1.23	0.96	39.6	66.2
52	12.9	5.45E+07	1.28	1.06, 1.53	0.0400[d]	0.040	15.5	88.5	0.52	1.35	0.95	41.8	66.3
53	10.9	5.46E+07	1.16	0.95, 1.42	0.0212	0.023	12.4	89.9	0.51	1.22	0.97	39.4	65.9
54	9.5	5.46E+07	1.21	0.98, 1.51	0.0251	0.021	10.8	91.3	0.51	1.24	0.98	39.7	65.8
55	7.8	5.46E+07	1.20	0.95, 1.52	0.0197	0.017	8.9	92.9	0.51	1.24	0.98	39.8	65.7
56	6.8	5.46E+07	1.22	0.97, 1.54	0.0200	0.021	8.1	94.0	0.51	1.34	0.98	41.6	65.8
57	5.8	5.46E+07	1.21	0.95, 1.56	0.0170	0.017	6.9	94.8	0.51	1.33	0.98	41.4	65.7
58	4.6	5.46E+07	1.22	0.92, 1.61	0.0135	0.013	5.5	95.8	0.51	1.32	0.99	41.2	65.6
59	3.9	5.47E+07	1.15	0.84, 1.57	0.0081	0.008	4.4	96.4	0.50	1.23	0.99	39.5	65.5
60	3.4	5.47E+07	1.02	0.71, 1.47	0.0019	0.004	3.7	96.7	0.50	1.12	1.00	37.3	65.4
61	3.0	5.47E+07	1.05	0.71, 1.54	0.0024	0.003	3.1	97.2	0.50	1.10	1.00	36.9	65.4
62	2.4	5.47E+07	1.13	0.75, 1.71	0.0049	0.000	2.3	97.6	0.50	1.00	1.00	34.6	65.3
63	2.0	5.47E+07	0.98	0.60, 1.60	−0.0003	0.000	1.7	97.8	0.50	0.79	1.00	29.5	65.2
64	1.8	5.47E+07	0.89	0.52, 1.51	−0.0023	0.000	1.4	98.0	0.50	0.71	1.01	27.5	65.2
65	1.5	5.47E+07	1.02	0.60, 1.74	0.0009	0.000	1.3	98.3	0.50	0.78	1.00	29.4	65.2

Abbreviations: AUC, area under the receiver operating characteristics curve; CI, confidence interval; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; RR, relative risk.

[a] All calculations take into account clustering and probability weighting.

[b] Incident cases of poor SRGH were defined as the first report of poor SRGH after 1986 following at least 1 previous report of good SRGH.

[c] Relative risk and corresponding 95% confidence interval from univariate Poisson regression.

[d] $P < 0.05$.

Table 2. Measures[a] of Model Fit, Calibration, and Model Discrimination for Long Work Hours and Incident Cardiovascular Disease ($n = 1,698$; Cases[b] = 777), Panel Study of Income Dynamics, United States, 1986–2011

Cutpoint (Weekly Work Hours)	Proportion Above Cutpoint, %	Measure of Model Fit and Calibration				Model Discrimination							
		Bayesian Information Criterion	RR[c]	95% CI[c]	Somers *D* Statistic	Youden Index (*J*)	Sensitivity, %	Specificity, %	AUC (c)	LR (+)	LR (−)	% PPV	% NPV
36	94.6	5.66e+07	1.30	0.94, 1.81	0.0225	0.015	95.5	6.0	0.51	1.02	0.75	46.1	61.1
37	87.4	5.66e+07	1.13	0.92, 1.39	0.0278	0.028	89.7	13.1	0.51	1.03	0.78	46.6	60.2
38	79.6	5.66e+07	1.14	0.96, 1.34	0.0439	0.032	82.4	20.8	0.52	1.04	0.85	46.7	58.4
39	71.9	5.66e+07	1.14	0.99, 1.31	0.0538	0.046	75.7	28.9	0.52	1.06	0.84	47.3	58.5
40	63.4	5.66e+07	1.14	1.01, 1.30	0.0623	0.056	68.7	36.9	0.53	1.09	0.85	47.9	58.3
41	55.6	5.66e+07	1.06	0.94, 1.19	0.0395	0.025	58.3	44.2	0.51	1.04	0.94	46.8	55.7
42	48.5	5.66e+07	1.07	0.96, 1.20	0.0413	0.027	52.3	50.5	0.51	1.06	0.95	43.7	52.2
43	42.7	5.66e+07	1.09	0.97, 1.22	0.0494	0.028	46.6	56.2	0.51	1.06	0.95	47.3	55.5
44	37.9	5.66e+07	1.04	0.93, 1.17	0.0312	0.024	41.6	60.8	0.51	1.06	0.96	47.2	55.2
45	34.2	5.66e+07	1.02	0.90, 1.14	0.0191	0.009	36.6	64.4	0.50	1.03	0.99	46.4	54.6
46	30.1	5.66e+07	1.01	0.89, 1.14	0.0158	0.012	31.9	69.3	0.51	1.04	0.98	46.7	54.7
47	26.3	5.66e+07	1.03	0.91, 1.16	0.0196	0.023	29.1	73.2	0.51	1.08	0.97	47.8	55.0
48	22.8	5.66e+07	1.09	0.97, 1.24	0.0387	0.033	26.0	77.3	0.52	1.15	0.96	49.1	55.3
49	19.6	5.66e+07	1.12	0.98, 1.27	0.0416[d]	0.036	23.0	80.6	0.52	1.19	0.96	50.0	55.4
50	16.2	5.66e+07	1.21[d]	1.06, 1.38	0.0594	0.045	20.2	84.3	0.52	1.28	0.95	52.0	55.6
51	13.4	5.65e+07	1.28[e]	1.12, 1.47	0.0672[e]	0.051	17.8	87.3	0.53	1.40	0.94	54.1	55.7
52	11.1	5.64e+07	1.42[e]	1.24, 1.63	0.0827[e]	0.058	15.7	90.1	0.53	1.59	0.94	57.3	55.9
53	9.7	5.64e+07	1.41[e]	1.21, 1.63	0.0699[e]	0.048	13.3	91.5	0.52	1.57	0.95	56.9	55.6
54	8.3	5.64e+07	1.45[e]	1.25, 1.69	0.0695[e]	0.051	12.1	93.1	0.53	1.74	0.94	59.5	55.6
55	6.7	5.64e+07	1.50[e]	1.28, 1.77	0.0629[e]	0.050	10.3	94.7	0.52	1.94	0.95	62.0	55.6
56	5.6	5.64e+07	1.49[e]	1.26, 1.76	0.0563[e]	0.046	9.1	95.4	0.52	2.00	0.95	62.8	55.5
57	4.7	5.66e+07	1.37[e]	1.12, 1.66	0.0344[d]	0.029	6.7	96.2	0.51	1.76	0.97	59.8	55.0
58	3.6	5.66e+07	1.35[d]	1.09, 1.67	0.0272[d]	0.022	5.5	96.6	0.51	1.64	0.98	58.1	54.8
59	3.1	5.66e+07	1.44[e]	1.15, 1.80	0.0285[d]	0.023	4.8	97.5	0.51	1.91	0.98	61.7	54.8
60	2.6	5.66e+07	1.37[d]	1.07, 1.76	0.0222[d]	0.018	4.1	97.7	0.51	1.81	0.98	60.4	54.7
61	2.4	5.66e+07	1.40[d]	1.09, 1.80	0.0221[d]	0.018	3.9	97.9	0.51	1.87	0.98	61.2	54.7
62	2.1	5.66e+07	1.42[d]	1.08, 1.87	0.0174[d]	0.013	3.0	98.4	0.51	1.82	0.99	60.5	54.6
63	1.8	5.66e+07	1.38[d]	1.02, 1.87	0.0137	0.011	2.6	98.5	0.51	1.69	0.99	58.8	54.5
64	1.7	5.66e+07	1.29	0.91, 1.82	0.0089	0.006	2.1	98.6	0.50	1.46	0.99	55.2	54.4
65	1.5	5.66e+07	1.38	0.97, 1.96	0.0101	0.007	1.9	98.8	0.50	1.62	0.99	57.7	54.4

Abbreviations: AUC, area under the receiver operating characteristics curve; CI, confidence interval; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; RR, relative risk.

[a] All calculations take into account clustering and probability weighting.

[b] Incident cases of CVD were defined as the first report of any of the physician diagnosed CVD-related outcomes (e.g., heart attack, high blood pressure, hypertension, coronary heart disease, angina, congestive heart failure, or stroke) after 1999 following at least 1 negative response to having ever been diagnosed with any of the CVD-related outcomes.

[c] Relative risk and corresponding 95% confidence interval from univariate Poisson regression.

[d] $P < 0.05$.

[e] $P < 0.01$.

For the outcome of cancer (Table 3), the best model fit was observed for the work-hours cutpoints of 52 WH/week, 54 WH/week, and 56 WH/week. Although significant *D* statistics were generated at several cutpoints, 52 WH/week maximized the *D* statistic. The relative risk calculations demonstrated significant increases in the risk of incident cancer for cutpoints from 50 WH/week to 57 WH/week. The cutpoint of 52 WH/week maximized the *J* statistic, AUC, and negative predictive value while minimizing the negative likelihood ratio. Although other cutpoints yielded equivalent AUC values, no other cutpoint performed as well across all measures of model fit, calibration, and discrimination as 52 WH/week.

Table 3. Measures[a] of Model Fit, Calibration, and Model Discrimination for Long Work Hours and Incident Cancer ($n = 2,196$; Cases[b] = 263), Panel Study of Income Dynamics, United States, 1986–2011

Cutpoint (Weekly Work Hours)	Proportion Above Cutpoint, %	Measure of Model Fit and Calibration				Model Discrimination							
		Bayesian Information Criterion	RR[c]	95% CI[c]	Somers *D* Statistic	Youden Index (*J*)	Sensitivity, %	Specificity, %	AUC (c)	LR (+)	LR (−)	% PPV	% NPV
36	94.6	3.62E+07	0.69	0.45, 1.04	−0.0192	0.000	91.6	5.0	0.50	0.96	1.67	11.6	81.5
37	87.4	3.62E+07	0.88	0.64, 1.22	−0.0055	0.000	85.9	12.4	0.50	0.98	1.13	11.8	86.6
38	79.6	3.62E+07	0.86	0.65, 1.13	−0.0145	0.000	78.3	20.2	0.50	0.98	1.07	11.8	87.3
39	71.9	3.62E+07	0.82	0.64, 1.05	−0.0316	0.000	70.3	27.9	0.50	0.98	1.06	11.7	87.4
40	63.4	3.62E+07	0.97	0.77, 1.23	0.0054	0.009	64.3	36.7	0.50	1.01	0.97	12.1	88.3
41	55.6	3.62E+07	1.06	0.85, 1.34	0.0294	0.021	57.4	44.6	0.51	1.04	0.95	12.4	88.5
42	48.5	3.62E+07	1.00	0.80, 1.25	0.0123	0.002	48.7	51.5	0.50	1.00	1.00	12.0	88.1
43	42.7	3.62E+07	0.99	0.79, 1.24	0.0093	0.000	42.6	57.3	0.50	1.00	1.00	12.0	88.0
44	37.9	3.62E+07	1.04	0.83, 1.31	0.0229	0.005	38.4	62.1	0.50	1.01	0.99	12.1	88.1
45	34.2	3.62E+07	1.06	0.85, 1.34	0.0279	0.000	34.2	65.8	0.50	1.00	1.00	12.0	88.0
46	30.1	3.62E+07	1.15	0.91, 1.45	0.0462	0.017	31.6	70.1	0.51	1.06	0.98	12.6	88.3
47	26.3	3.62E+07	1.22	0.96, 1.55	0.0564	0.030	28.9	74.1	0.51	1.11	0.96	13.2	88.4
48	22.8	3.62E+07	1.27	1.00, 1.63	0.0611[d]	0.035	25.9	77.7	0.52	1.16	0.95	13.6	88.5
49	19.6	3.62E+07	1.16	0.89, 1.51	0.0355	0.023	21.7	80.7	0.51	1.12	0.97	13.2	88.3
50	16.2	3.62E+07	1.32[d]	1.01, 1.74	0.0560[d]	0.036	19.4	84.2	0.52	1.23	0.96	14.3	88.5
51	13.4	3.61E+07	1.45[d]	1.10, 1.92	0.0657[d]	0.051	17.9	87.2	0.53	1.40	0.94	16.0	88.6
52	11.1	3.60E+07	1.62[e]	1.22, 2.17	0.0742[d]	0.055	16.0	89.5	0.53	1.53	0.94	17.2	88.7
53	9.7	3.61E+07	1.57[d]	1.14, 2.16	0.0612[d]	0.045	13.7	90.8	0.52	1.49	0.95	16.9	88.6
54	8.3	3.60E+07	1.71[e]	1.23, 2.38	0.0650[d]	0.052	12.9	92.3	0.53	1.68	0.94	18.6	88.6
55	6.7	3.61E+07	1.70[d]	1.18, 2.46	0.0543[d]	0.036	9.9	93.7	0.52	1.58	0.96	17.7	88.4
56	5.6	3.60E+07	1.86[e]	1.27, 2.70	0.0564[d]	0.041	9.1	94.9	0.52	1.80	0.96	19.7	88.5
57	4.7	3.61E+07	1.67[d]	1.12, 2.48	0.0383[d]	0.033	7.6	95.7	0.52	1.75	0.97	19.2	88.4
58	3.6	3.62E+07	1.40	0.89, 2.22	0.0182	0.015	4.9	96.5	0.51	1.43	0.98	16.3	88.2
59	3.1	3.62E+07	1.54	0.96, 2.46	0.0212	0.017	4.6	97.1	0.51	1.57	0.98	17.6	88.2
60	2.6	3.62E+07	1.46	0.86, 2.46	0.0157	0.013	3.8	97.5	0.51	1.53	0.99	17.2	88.2
61	2.4	3.62E+07	1.62	0.97, 2.73	0.0193	0.016	3.8	97.8	0.51	1.75	0.98	19.2	88.2
62	2.1	3.62E+07	1.65	0.97, 2.83	0.0162	0.016	3.4	98.1	0.51	1.84	0.98	20.0	88.2
63	1.8	3.62E+07	1.45	0.78, 2.69	0.0097	0.010	2.7	98.3	0.51	1.61	0.99	17.9	88.1
64	1.7	3.62E+07	1.49	0.80, 2.75	0.0102	0.011	2.7	98.4	0.51	1.66	0.99	18.4	88.1
65	1.5	3.62E+07	1.57	0.81, 3.05	0.0099	0.009	2.3	98.7	0.50	1.70	0.99	18.8	88.1

Abbreviations: AUC, area under the receiver operating characteristics curve; CI, confidence interval; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; RR, relative risk.

[a] All calculations take into account clustering and probability weighting.

[b] Incident cases of cancer were defined as the first report of physician diagnosed cancer or malignant tumor (excluding skin cancer) after 1999 following at least 1 previous report of not ever having been diagnosed with cancer or a malignant tumor.

[c] Relative risk and corresponding 95% confidence interval from univariate Poisson regression.

[d] $P < 0.05$.

[e] $P < 0.01$.

The analyses of the LWH exposure as a continuous variable (results not shown) identified the optimal LWH cutpoints of 51.8 WH/week, 52.2 WH/week, and 52.1 WH/week for poor SRGH, incident CVD, and incident cancer, respectively. In the post-hoc analysis, we observed no significant differences between the relative risk or *D* beta coefficients for the cutpoints of 50 WH/week to 54 WH/week for each outcome; however, significant differences between 50 WH/week and 52 WH/week at the 0.10 level for the outcome of CVD ($P = 0.08$) were observed. While the comparison of the relative risk and *D* statistic confidence intervals indicated an overlap across cutpoints, there was a clear pattern of increasing magnitude up to 52 WH/week, with a decline thereafter.

Table 4. Demographic Characteristics[a] at Study Baseline According to Work-Hours Exposure Category (35–51 Hours per Week vs. ≥52 Hours per Week) for Each Outcome-Specific Cohort, Panel Study of Income Dynamics, United States, 1986

Characteristic	Average Work Hours per Week According to Outcome Cohort					
	CVD Cohort (n = 1,698)		Cancer Cohort (n = 2,196)		SRGH Cohort (n = 2,206)	
	35–51 Hours (n = 1,485), %	≥52 Hours (n = 213), %	35–51 Hours (n = 1,952), %	≥52 Hours (n = 244), %	35–51 Hours (n = 1,921), %	≥52 Hours (n = 285), %
Sex[b,c]						
Male	61.4	89.2	62.0	89.5	61.5	87.3
Female	38.6	10.8	38.0	10.5	38.5	12.7
Age, years[d]	32.5 (0.3)[e]	34.3 (0.8)[e]	33.7 (0.2)	34.5 (0.7)	33.9 (0.2)	35.0 (0.7)
Educational level (highest completed)[b]						
Did not complete high school	6.5	5.5	6.9[c]	5.9[c]	7.0	6.5
High-school diploma	61.3	51.0	61.7[c]	49.2[c]	61.2	52.2
College degree	32.2	43.4	31.4[c]	44.9[c]	31.8	41.3
Racial/ethnic category[b]						
White, non-Hispanic	88.0	95.1	86.9	92.9	87.1	92.0
Black, non-Hispanic	7.2	2.4	8.6	3.0	8.2	4.3
Hispanic	3.3	1.5	2.9	2.8	3.0	2.6
Other	1.5	1.0	1.6	1.3	1.7	1.1
Marital status[b]						
Married/cohabiting	76.9[e]	85.5[e]	75.6[c]	87.0[c]	75.8[f]	85.5[f]
Not married or cohabiting	23.1[e]	14.5[e]	24.4[c]	13.0[c]	24.2[f]	14.5[f]
Number of children in the household[d]	1.0 (0.04)	1.2 (0.09)	1.1 (0.04)	1.2 (0.08)	1.0 (0.03)[e]	1.2 (0.08)[e]
Employment status[b,c]						
Self-employed	13.3	32.5	12.5	37.0	12.0	34.3
Employed by others	86.7	67.5	87.5	63.0	88.0	65.7
Industry[b]						
Services	68.2	69.9	68.1	69.9	68.8	67.7
Nonservices	31.8	30.1	31.9	30.2	31.2	32.3
Occupation[b]						
Manual	29.3	31.6	29.2	32.8	28.5	33.5
Nonmanual	70.7	68.4	70.8	67.2	71.5	66.5
Pay status[b,c]						
Salaried	46.8	55.4	46.5	52.3	47.2	51.4
Hourly	39.3	13.6	39.8	11.8	39.7	14.9
Other arrangement	13.8	31.0	13.7	35.9	13.1	33.7
Poor SRGH status[b]						
Yes	27.5	25.7	32.8	32.6	32.9[e]	40.6[e]
No	72.5	74.3	67.2	67.4	67.1[e]	59.4[e]
Incident CVD[b]						
Yes	42.6[c]	59.8[c]	54.9[e]	63.5[e]	55.0[f]	65.1[f]
No	57.4[c]	40.2[c]	45.1[e]	36.5[e]	45.0[f]	34.9[f]

Table continues

Table 4. Continued

Characteristic	Average Work Hours per Week According to Outcome Cohort					
	CVD Cohort ($n = 1,698$)		Cancer Cohort ($n = 2,196$)		SRGH Cohort ($n = 2,206$)	
	35–51 Hours ($n = 1,485$), %	≥52 Hours ($n = 213$), %	35–51 Hours ($n = 1,952$), %	≥52 Hours ($n = 244$), %	35–51 Hours ($n = 1,921$), %	≥52 Hours ($n = 285$), %
Incident cancer[b]						
Yes	12.2[f]	21.3[f]	14.5[f]	22.3[f]	14.1[e]	21.4[e]
No	87.8[f]	78.7[f]	85.5[f]	77.7[f]	85.9[e]	78.6[e]

Abbreviations: CVD, cardiovascular disease; SRGH, self-reported general health.
[a] Proportions adjusted for probability weighting; statistical tests take into account clustering and probability weighting.
[b] χ^2 test.
[c] $P < 0.001$.
[d] Adjusted Wald test; values expressed as mean (standard error).
[e] $P < 0.05$.
[f] $P < 0.01$.

Descriptive data on the study sample are presented in Table 4 by average WH/week, with the ≥52 WH/week category representing the exposure of LWH. Across all 3 outcome-specific cohorts, those who worked 52 WH/week or more were more likely to be male, married, self-employed, and paid through nontraditional pay structures (defined as not paid hourly or salaried) than were those not reporting LWH.

Among the 1,107 (48%) of study participants that reported at least 1 study outcome, 15.9% ($n = 176$) worked an average of 52 hours or more per week (results not shown). At baseline, the participants who would go on to develop any of the 3 health outcomes reported working an average of 3 hours more per week than participants who did not report poor SRGH, incident CVD, or incident cancer (42.8 versus 40.9 hours, respectively; $P < 0.005$). These differences in work hours remained present over the study duration; those who developed an outcome reported working approximately 1.5 hours more per week than those who did not (45.1 versus 43.7 hours, respectively; $P < 0.005$).

In our examination of overlap across outcome-specific cohorts, we observed that approximately 71% ($n = 1,627$ of 2,306) of participants were included in all 3 outcome-specific cohorts, with 68% ($n = 1,107$) reporting at least 1 outcome. Of those, 29.2% ($n = 373$) reported 2 outcomes, and 3.9% ($n = 44$) reported all 3. Among those reporting more than 1 outcome of interest ($n = 417$), concurrent cases of CVD and poor SRGH (79.9%; 333 of 417) were most common.

In the examination of LWH exposure patterns of nonrespondents who did not provide outcome data, there were no differences in the mean WH/week relative to respondents (43.6 vs. 43.6 mean WH/week, respectively ($P = 0.925$)).

DISCUSSION

To our knowledge, there have been no published evaluations of the predictive ability of specific LWH cutpoints for different adverse health outcomes. In this study, we observed that the dichotomized exposure cutpoint of 52 work hours per week or greater, on average, over a minimum of 10 years was the optimized cutpoint for predicting increased risk of poor SRGH, incident CVD, and incident cancer. The 52 WH/week cutpoint was more accurate in terms of calibration than discrimination when compared with lower or higher cutpoints, showing good model calibration but limited model discrimination.

Our findings that adverse health outcomes were associated with working 52 hours per week or more are consistent with evidence reported from previous cohort studies. For SRGH, a previous study showed a relationship between increasing overtime hours and poor subjective health, but the level of work hours was analyzed as a continuous measure and, thus, a specific range of work hours beyond which the risk of poor SRGH increased was not identified (48). Prior studies have demonstrated increased risk of CVD at work-hours durations of 45 hours per week (49), 50 hours per week (50), 55 hours per week (51), 11 hours per day (52), 11–12 hours per day (53), and 3–4 hours of overtime per day (53). To our knowledge, only 3 previous studies have examined the relationship between LWH and cancer, with contradictory results (31–33). Our findings are consistent with those of a recent study reporting a significant dose-response relationship between LWH and cancer incidence, most notably among women, which analyzed a longitudinal US cohort and employed similar exposure and outcome measures to those used here (31). Two additional studies examining LWH and cancer incidence reported mixed results, with some evidence of an increase in incident breast cancer (32, 33). Because all 3 of these previous studies used different LWH categories, the estimates are not directly comparable, which could be remedied by a more consistent approach to defining work hours in terms of health risk–based thresholds.

Each of the models used to test the cutpoints against the outcomes of interest was univariate, as is standard practice for the initial identification of predictive thresholds (54–57). Employing

this approach highlighted the dichotomized work-hours cutpoint that was the strongest predictive indicator of the outcome across all members of each cohort, regardless of demographic, socioeconomic, or occupational characteristics. However, in each of these analyses, the small values of *J* and AUC suggested that the univariate cutpoint models had poor distinguishing ability, which would likely be improved through multivariate modeling (21, 40). Although no statistical differences were found for the relative risk and *D* beta coefficients for the cutpoints of 50–54 WH/week, none were expected, as the assignment of exposure status (and those participants who constituted the exposed portion of the sample) varied little from one cutpoint to the next larger cutpoint (46, 47). However, the suggestion of statistical difference at a significance level of 0.1 and the trends demonstrated by the confidence intervals increasing to and decreasing from 52 WH/week supported our findings that 52 WH/week is an appropriate starting point for defining LWH in more sophisticated models that consider possible confounding effects as well as the influence of mediators and moderators of the work hours–health relationship.

The identification of the same LWH cutpoint (52 WH/week) as the best predictor for each of the outcomes was unexpected. Although a sizeable proportion of participants were included in all 3 cohorts, the overlap in the cohorts was primarily among those who did not develop the outcomes. Given the high rates of CVD incidence in the US adult population (58) and the broad criteria used when evaluating SRGH (28), it was not surprising that some participants developed incident cases of more than 1 outcome.

PSID provides limited information about why data are missing with respect to loss to follow-up, refusal to answer, or death. Our analysis of nonrespondents suggests that selection bias is unlikely, because participation was not differentially related to exposure (59). It is possible that condition severity precluded their further participation for health reasons (60). In the event that cases went unreported by nonrespondents, our results would be biased towards the null.

The measures of association resulting from these analyses are unadjusted, and we recognize that confounding may be present. However, the goal of this study was to take this important initial step of identifying working-hour thresholds with the greatest predictive ability for the outcomes of interest. These estimates provide researchers with a starting point (approximately 52 hours worked per week) for assessing LWH in future multivariate models.

The longitudinal nature of this study in a large, representative sample of US households allowed us to identify statistically optimized cutpoints for LWH using repeated measures of working hours. The PSID cohort is not specifically sampled to produce a representative sample of US workers; therefore, we applied PSID sampling weights to improve the generalizability of the study findings to US workers. An additional strength is that we had ample data to examine exposures over a minimum of 10 years, which we used as a criterion for inclusion. We recognize that requiring a minimum of 10 years of work-hours data precluded our ability to draw conclusions for those reporting fewer years of data. We chose to be conservative in our inclusion criteria to ensure that participant work-hours averages reflected long-term patterns and to allow adequate time for the onset of the outcomes following exposure.

In 2006, a multidisciplinary team of experts argued that flawed methodological approaches prevented the accurate classification of health risks among those who work extended hours (17), yet this approach to evaluating the effect of LWH on workers' health has not changed. Further research using other representative data sets of US workers is necessary to assess the predictive value of the exposure cutpoint of 52 WH/week as a risk factor of detrimental health conditions as well as to examine time-varying patterns of working long hours on health outcomes.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas (Sadie H. Conway, Lisa A. Pompeii); Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, San Antonio, Texas (David Gimeno Ruiz de Porras); Department of Mathematics, Computer Science, and Cooperative Engineering, University of St. Thomas, Houston, Texas (Jack L. Follis); and Division of Health Promotion and Behavioral Sciences, School of Public Health, University of Texas Health Science Center at Houston, San Antonio, Texas (Robert E. Roberts).

Funding for this project was received from the National Institute for Occupational Safety and Health (grant 5T42OH008421).

Conflict of interest: none declared.

REFERENCES

1. Buell P, Breslow L. Mortality from coronary heart disease in California men who work long hours. *J Chronic Dis*. 1960; 11(6):615–626.
2. Ettner SL, Grzywacz JG. Workers' perceptions of how jobs affect health: a social ecological perspective. *J Occup Health Psychol*. 2001;6(2):101–113.
3. Virtanen M, Heikkilä K, Jokela M, et al. Long working hours and coronary heart disease: a systematic review and meta-analysis. *Am J Epidemiol*. 2012;176(7):586–596.
4. Lipscomb JA, Trinkoff AM, Geiger-Brown J, et al. Work-schedule characteristics and reported musculoskeletal disorders of registered nurses. *Scand J Work Environ Health*. 2002; 28(6):394–401.
5. Simpson CL, Severson RK. Risk of injury in African American hospital workers. *J Occup Environ Med*. 2000;42(10): 1035–1040.
6. Dembe AE, Erickson JB, Delbos RG, et al. The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. *Occup Environ Med*. 2005;62(9):588–597.
7. Hilton MF, Whiteford HA, Sheridan JS, et al. The prevalence of psychological distress in employees and associated occupational risk factors. *J Occup Environ Med*. 2008;50(7): 746–757.

8. Kivistö M, Härmä M, Sallinen M, et al. Work-related factors, sleep debt and insomnia in IT professionals. *Occup Med (Lond)*. 2008;58(2):138–140.
9. Theorell T, Rahe RH. Behavior and life satisfactions characteristics of Swedish subjects with myocardial infarction. *J Chronic Dis*. 1972;25(3):139–147.
10. Kawakami N, Akachi K, Shimizu H, et al. Job strain, social support in the workplace, and haemoglobin A1c in Japanese men. *Occup Environ Med*. 2000;57(12):805–809.
11. Proctor SP, White RF, Robins TG, et al. Effect of overtime work on cognitive function in automotive workers. *Scand J Work Environ Health*. 1996;22(2):124–132.
12. Jex SM, Bliese PD. Efficacy beliefs as a moderator of the impact of work-related stressors: a multilevel study. *J Appl Psychol*. 1999;84(3):349–361.
13. Kroenke CH, Spiegelman D, Manson J, et al. Work characteristics and incidence of type 2 diabetes in women. *Am J Epidemiol*. 2007;165(2):175–183.
14. Bannai A, Tamakoshi A. The association between long working hours and health: a systematic review of epidemiological evidence. *Scand J Work Environ Health*. 2014;40(1):5–18.
15. van der Hulst M. Long workhours and health. *Scand J Work Environ Health*. 2003;29(3):171–188.
16. Spurgeon A, Harrington JM, Cooper CL. Health and safety problems associated with long working hours: a review of the current position. *Occup Environ Med*. 1997;54(6):367–375.
17. Caruso CC, Bushnell T, Eggerth D, et al. Long working hours, safety, and health: toward a National Research Agenda. *Am J Ind Med*. 2006;49(11):930–942.
18. Sparks K, Cooper C, Fried Y, et al. The effects of hours of work on health: a meta-analytic review. *J Occup Organ Psychol*. 1997;70(4):391–408.
19. Harrington JM. Shift work and health—a critical review of the literature on working hours. *Ann Acad Med Singapore*. 1994; 23(5):699–705.
20. Kivimäki M, Jokela M, Nyberg ST, et al. Long working hours and risk of coronary heart disease and stroke: a systematic review and meta-analysis of published and unpublished data for 603,838 individuals. *Lancet*. 2015;386(10005):1739–1746.
21. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem*. 2008;54(1): 17–23.
22. Simon R. Advances in clinical trial designs for predictive biomarker discovery and validation. *Curr Breast Cancer Rep*. 2009;1(4):216–221.
23. Okosun IS, Rotimi CN, Forrester TE, et al. Predictive value of abdominal obesity cut-off points for hypertension in blacks from west African and Caribbean island nations. *Int J Obes Relat Metab Disord*. 2000;24(2):180–186.
24. Olsson Möller U, Kristensson J, Midlöv P, et al. Predictive validity and cut-off scores in four diagnostic tests for falls—a study in frail older people at home. *Phys Occup Ther Geriatr*. 2012;30(3):189–201.
25. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007; 115(7):928–935.
26. Schisterman EF, Perkins NJ, Liu A, et al. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005; 16(1):73–81.
27. Kuhn R, Rahman O, Menken J. Survey measures of health: how well do self-reported and observed indicators measure health and predict mortality. In: National Research Council Committee on Population; Cohen B, Menken J, eds. *Aging in Sub-Saharan Africa: Recommendation for Furthering Research*. Washington, DC: National Academies Press; 2006: 314–344.
28. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med*. 2009;69(3):307–316.
29. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav*. 1997;38(1):21–37.
30. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2016 update: a report from the American Heart Association. *Circulation*. 2016;133(4): e38–e360.
31. Dembe AE, Yao X. Chronic disease risks from exposure to long-hour work schedules over a 32-year period. *J Occup Environ Med*. 2016;58(9):861–867.
32. Heikkila K, Nyberg ST, Madsen IE, et al. Long working hours and cancer risk: a multi-cohort study. *Br J Cancer*. 2016; 114(7):813–818.
33. Nielsen NR, Stahlberg C, Strandberg-Larsen K, et al. Are work-related stressors associated with diagnosis of more advanced stages of incident breast cancers? *Cancer Causes Control*. 2008;19(3):297–303.
34. University of Michigan Institute for Social Research. *PSID: Data on Families Since 1968*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan; 2014. http://psidonline.isr.umich.edu/Guide/Brochures/PSID.pdf. Accessed April 11, 2014.
35. McGonagle KA, Schoeni RF, Sastry N, et al. The Panel Study of Income Dynamics: overview, recent innovations, and potential for life course research. *Longit Life Course Stud*. 2012;3(2):188–209.
36. United States Bureau of Labor Statistics. *Appendix C. How to Compute Your Firm's Incidence Rate for Safety Management.* Washington, DC: United States Bureau of Labor Statistics; 2006. http://www.bls.gov/iif/oshwc/osh/os/osh06_appc.pdf. Accessed November 21, 2014.
37. Conway SH, Pompeii LA, Roberts RE, et al. Dose-response relation between work hours and cardiovascular disease risk: findings from the Panel Study of Income Dynamics. *J Occup Environ Med*. 2016;58(3):221–226.
38. Ross CE, Mirowsky J. Does employment affect health? *J Health Soc Behav*. 1995;36(3):230–243.
39. Newson R. Confidence intervals for rank statistics: Somers' D and extensions. *Stata J*. 2006;6(3):309–334.
40. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950; 3(1):32–35.
41. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006;163(7): 670–675.
42. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. 2nd ed. Philadelphia, PA: W.B. Saunders Company; 1985.
43. McGee S. Simplifying likelihood ratios. *J Gen Intern Med*. 2002;17(8):647–650.
44. Radack KL, Rouan G, Hedges J. The likelihood ratio. An improved measure for reporting and evaluating diagnostic test results. *Arch Pathol Lab Med*. 1986;110(8):689–693.
45. Higgins PB, Gower BA, Hunter GR, et al. Defining health-related obesity in prepubertal children. *Obes Res*. 2001;9(4): 233–240.
46. Clogg CC, Petkova E, Haritou A. Statistical methods for comparing regression coefficients between models. *Am J Sociol*. 1995;100(5):1261–1293.

47. Paternoster R, Brame R, Mazerolle P, et al. Using the correct statistical test for the equality of regression coefficients. *Criminology*. 1998;36(4):859–866.
48. Taris TW, Ybema JF, Beckers DG, et al. Investigating the associations among overtime work, health behaviors, and health: a longitudinal study among full-time employees. *Int J Behav Med*. 2011;18(4):352–360.
49. Holtermann A, Mortensen OS, Burr H, et al. Long work hours and physical fitness: 30-year risk of ischaemic heart disease and all-cause mortality among middle-aged Caucasian men. *Heart*. 2010;96(20):1638–1644.
50. Tarumi K, Hagihara A, Morimoto K. A prospective observation of onsets of health defects associated with working hours. *Ind Health.* 2003;41(2):101–108.
51. O'Reilly D, Rosato M. Worked to death? A census-based longitudinal study of the relationship between the numbers of hours spent working and mortality risk. *Int J Epidemiol*. 2013; 42(6):1820–1830.
52. Kivimäki M, Batty GD, Hamer M, et al. Using additional information on working hours to predict coronary heart disease: a cohort study. *Ann Intern Med*. 2011;154(7): 457–463.
53. Virtanen M, Ferrie JE, Singh-Manoux A, et al. Overtime work and incident coronary heart disease: the Whitehall II prospective cohort study. *Eur Heart J*. 2010;31(14): 1737–1744.
54. Faraggi D, Simon R. A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. *Stat Med.* 1996;15(20):2203–2213.
55. de Edwardes MD. Univariate random cut-points theory for the analysis of ordered categorical data. *J Am Stat Assoc*. 1997; 92(439):1114–1123.
56. Williams BA, Mandreker JN, Mandrekar SJ, et al. *Finding Optimal Cutpoints for Continuous Covariates with Binary and Time-to-Event Outcomes*. Rochester, MN: Department of Health Sciences Research, Mayo Clinic; 2006. (Technical Report Series #79).
57. Vanniyasingam T, Rodseth RN, Lurati Buse GA, et al. Predicting the occurrence of major adverse cardiac events within 30 days of a vascular surgery: an empirical comparison of the minimum p value method and ROC curve approach using individual patient data meta-analysis. *Springerplus*. 2016;5:304.
58. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation*. 2015;131(4):e29–e322.
59. Austin MA, Criqui MH, Barrett-Connor E, et al. The effect of response bias on the odds ratio. *Am J Epidemiol*. 1981;114(1): 137–143.
60. Montgomery MP, Kamel F, Hoppin JA, et al. Characteristics of non-participation and potential for selection bias in a prospective cohort study. *Am J Ind Med*. 2010;53(5):486–496.