Interval Regression

Clara E. “Libby” Dismuke-Greer, PhD

March 10,2021

HERC
Health Economics Resource Center

VA
U.S. Department of Veterans Affairs
Veterans Health Administration
Health Services Research & Development Service

Outline

- “Stings in the tails”
- Types of interval data
- Bias using OLS with interval data
- STATA intreg for simple interval regression with normal distributions
- STATA eintreg for sample selection or endogeneity
- Other intreg extensions
- R Package and SAS
- Example using earnings for individuals with SCI
- Example using wait times for primary care in 10 OECD countries

Poll Question 1

What experience do you have with interval data?

A. No experience with interval data or interval models estimation.

B. Experience with interval data but not interval models estimation.

C. Experience with interval data and interval models estimation.

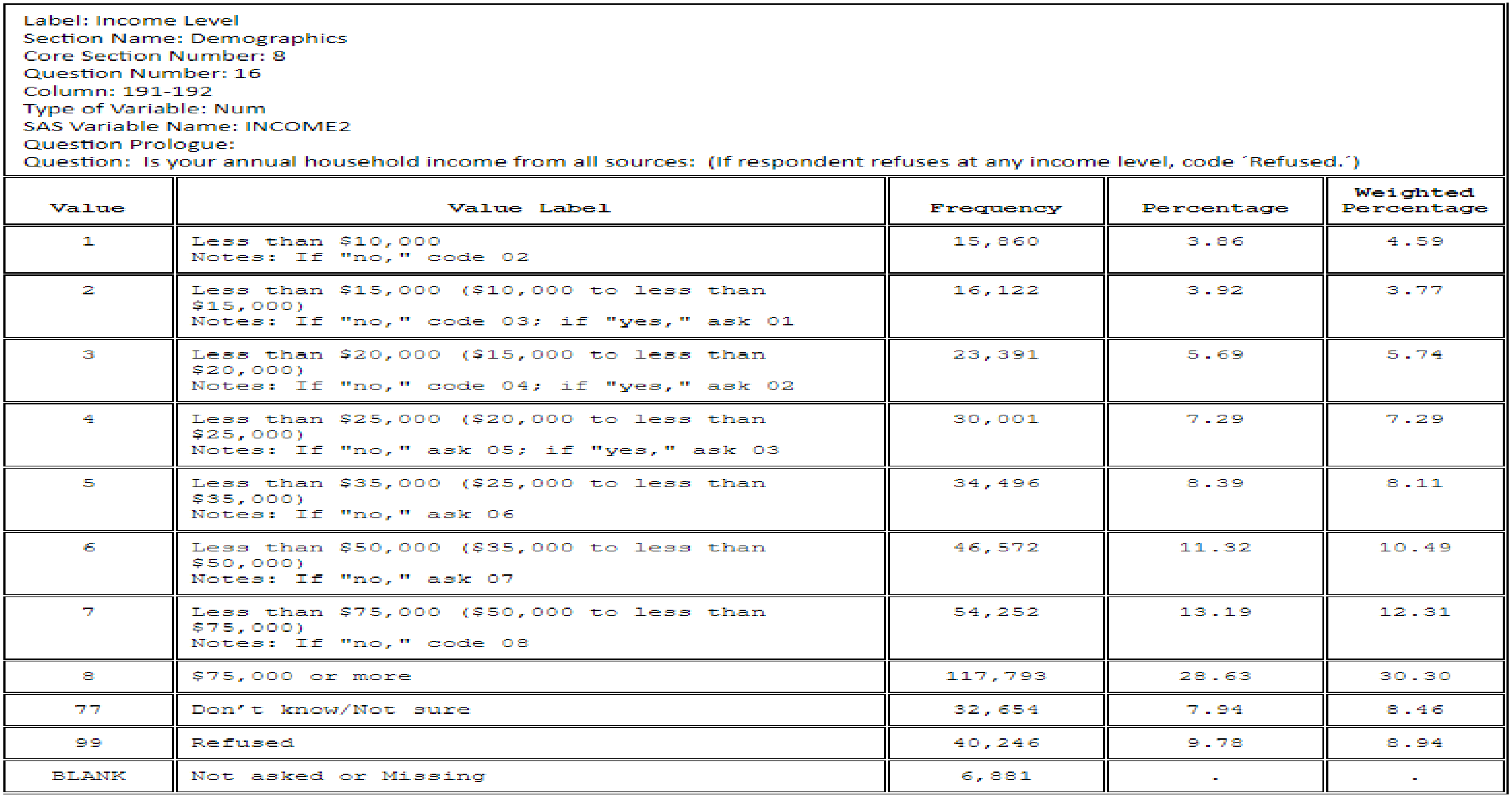
“Stings in the Tails” (1)

- Interval data occurs when only the lower and upper interval bounds of a variable are observed and the true value which lies between the bounds is unknown. (2)
- Instead of measuring the dependent variable on a continuous scale, the scale is divided into η_κ intervals where κ indicates in which of the η_κ intervals an observation falls. (2)
- This leads to information loss since the distribution shape within the intervals is unknown. (2)
- In the survey statistics field, especially among government surveys, requesting data such as income in intervals has been shown to reduce item non-response because it offers a higher level of privacy protection. (2)

Types of Interval Data

- Income reported in health surveys such as
- Behavioral Risk Factor Surveillance System (BRFSS),
- National Health Interview Survey (NHIS)
- and Medicare Current Beneficiary Survey (MCBS) are reported in intervals.
- Wait times (3)

BRFSS 2019



Label: Income Level
Section Name: Demographics
Core Section Number: 8
Question Number: 16
Column: 191-192
Type of Variable: Num
SAS Variable Name: INCOME2
Question Prologue:
Question: Is your annual household income from all sources: (If respondent refuses at any income level, code ´Refused.´)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Less than $10,000 Notes: If "no," code 02	15,860	3.86	4.59
2	Less than $15,000 ($10,000 to less than $15,000) Notes: If "no," code 03; if "yes," ask 01	16,122	3.92	3.77
3	Less than $20,000 ($15,000 to less than $20,000) Notes: If "no," code 04; if "yes," ask 02	23,391	5.69	5.74
4	Less than $25,000 ($20,000 to less than $25,000) Notes: If "no," ask 05; if "yes," ask 03	30,001	7.29	7.29
5	Less than $35,000 ($25,000 to less than $35,000) Notes: If "no," ask 06	34,496	8.39	8.11
6	Less than $50,000 ($35,000 to less than $50,000) Notes: If "no," ask 07	46,572	11.32	10.49
7	Less than $75,000 ($50,000 to less than $75,000) Notes: If "no," code 08	54,252	13.19	12.31
8	$75,000 or more	117,793	28.63	30.30
77	Don’t know/Not sure	32,654	7.94	8.46
99	Refused	40,246	9.78	8.94
BLANK	Not asked or Missing	6,881	.	.

OLS Bias

- While OLS regression on the midpoints of the intervals is easily applied, it comes with the disadvantage of giving biased estimation results. (2)
- This approach disregards the uncertainty stemming from the unknown true distribution of the data within the intervals and therefore leads to biased parameter estimates. (2)
- Its performance relies on the number of intervals and estimation results are only comparable to more advanced methods when the number of intervals is very large. (2)

Poll Question 2

Which statistical package do you use? Check all that apply

A. Stata

B. SAS

C. R

D. SPSS

E. Other

STATA Intreg

- intreg fits a linear model with an outcome measured as point data, interval data, left-censored data, or right-censored data.
- As such, it is a generalization of the model fit by tobit.
- Regression on x1 and x2 of an interval-measured dependent variable with lower endpoint y lower and upper endpoint y upper intreg y_lower y_upper x1 x2.
- Coefficients are interpreted directly.

STATA Intreg

- With robust standard errors: intreg y_lower y_upper x1 x2, vce(robust)
- Model heteroskedasticity in the conditional variance as a function of x3 : intreg y_lower y_upper x1 x2, het(x3)
- Adjust for complex survey design using svyset data: svy: intreg y_lower y_upper x1 x

Eintreg for sample selection or endogeneity

- eintreg fits an interval regression model that accommodates any combination of endogenous covariates, nonrandom treatment assignment, and endogenous sample selection.
- Continuous, binary, and ordinal endogenous covariates are allowed.
- Treatment assignment may be endogenous or exogenous.

Intreg Extensions

- Xteintreg - fits a random-effects interval regression model that accommodates endogenous covariates, treatment, and sample selection in the same way as eintreg and also accounts for correlation of observations within panels or within groups.

Other intreg extensions

- bayes:intreg –Bayseian interval regression
- fmm:intreg- Finite mixtures of interval regression models
- meintreg- Multilevel mixed-effects interval regression
- stintreg- parametric models for interval-sensored survival-time data
- xtintreg- Random-effects interval-data regression models

R package and SAS

- R package kdeAlgo() Estimates statistical indicators and its standard errors from interval censored data.
- SAS uses proc lifereg
- Ex. Proce lifereg data=intreg+data; class type; model (lgpa ugpa)=…/d=normal; run;

Example of intreg and extintreg

- Employment status, hours working, and gainful earnings after spinal cord injury: relationship with pain, prescription medications for pain, and nonprescription opioid use (4)
- Participants included 4670 adults with traumatic SCI of at least one-year duration who were enrolled in a study of health and longevity.
- Earnings were measured with 16 categories ranging from a low of<$10,000 to a high of >$175,000.

Methods

- We used standard interval regression (intreg) to estimate the association of pain and pain medications with conditional earnings (conditional on being employed).
- We used extended interval regression (extintreg) to estimate the association of pain and pain medications with unconditional earnings (full sample).

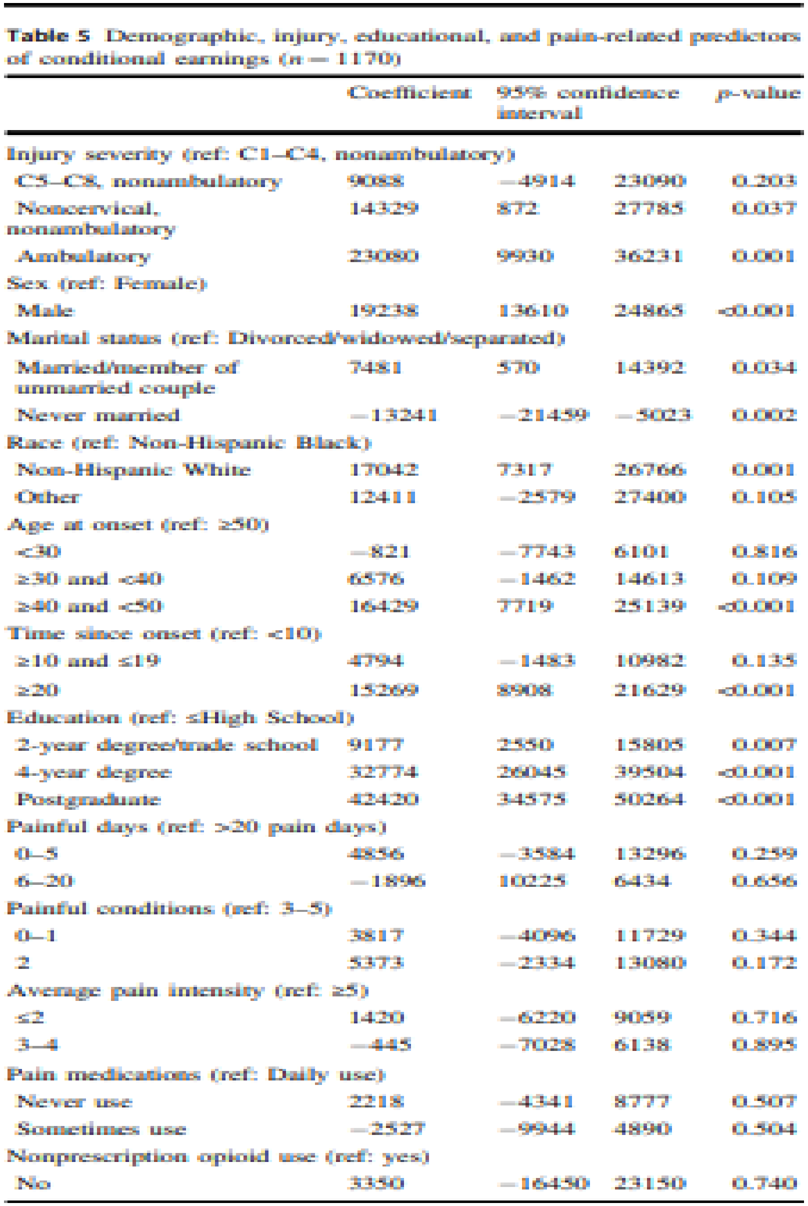


Table 5 Demographic, injury, educational, and pain-related predictors of conditional earnings ($n = 1170$)

	Coefficient	95% confidence interval		p-value
Injury severity (ref: C1–C4, nonambulatory)				
C5–C8, nonambulatory	9088	−4914	23090	0.203
Noncervical, nonambulatory	14329	872	27785	0.037
Ambulatory	23080	9930	36231	0.001
Sex (ref: Female)				
Male	19238	13610	24865	<0.001
Marital status (ref: Divorced/widowed/separated)				
Married/member of unmarried couple	7481	570	14392	0.034
Never married	−13241	−21459	−5023	0.002
Race (ref: Non-Hispanic Black)				
Non-Hispanic White	17042	7317	26766	0.001
Other	12411	−2579	27400	0.105
Age at onset (ref: ≥50)				
<30	−821	−7743	6101	0.816
≥30 and <40	6576	−1462	14613	0.109
≥40 and <50	16429	7719	25139	<0.001
Time since onset (ref: <10)				
≥10 and ≤19	4794	−1483	10982	0.135
≥20	15269	8908	21629	<0.001
Education (ref: ≤High School)				
2-year degree/trade school	9177	2550	15805	0.007
4-year degree	32774	26045	39504	<0.001
Postgraduate	42420	34575	50264	<0.001
Painful days (ref: >20 pain days)				
0–5	4856	−3584	13296	0.259
6–20	−1896	10225	6434	0.656
Painful conditions (ref: 3–5)				
0–1	3817	−4096	11729	0.344
2	5373	−2334	13080	0.172
Average pain intensity (ref: ≥5)				
≤2	1420	−6220	9059	0.716
3–4	−445	−7028	6138	0.895
Pain medications (ref: Daily use)				
Never use	2218	−4341	8777	0.507
Sometimes use	−2527	−9944	4890	0.504
Nonprescription opioid use (ref: yes)				
No	3350	−16450	23150	0.740

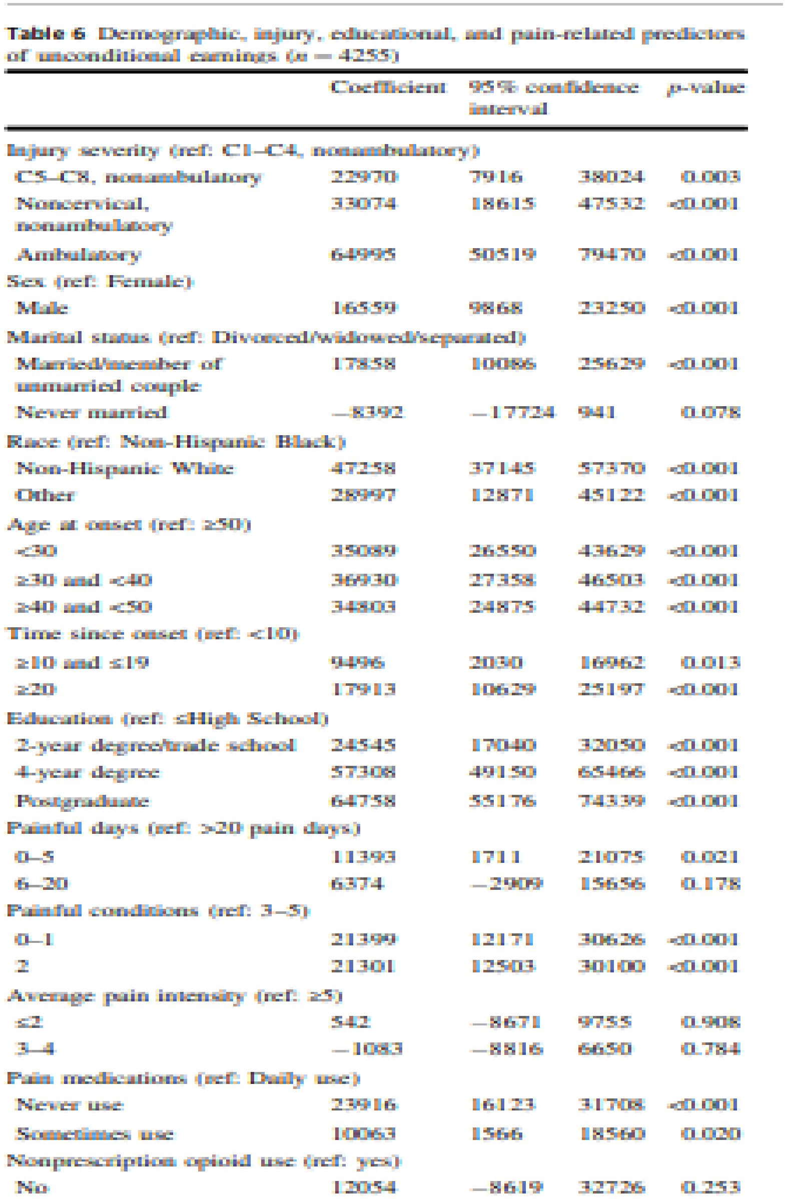


Table 6 Demographic, injury, educational, and pain-related predictors of unconditional earnings ($n = 4255$)

	Coefficient	95% confidence interval		p-value
Injury severity (ref: C1–C4, nonambulatory)				
C5–C8, nonambulatory	22970	7916	38024	0.003
Noncervical, nonambulatory	33074	18615	47532	<0.001
Ambulatory	64995	50519	79470	<0.001
Sex (ref: Female)				
Male	16559	9868	23250	<0.001
Marital status (ref: Divorced/widowed/separated)				
Married/member of unmarried couple	17858	10086	25629	<0.001
Never married	−8392	−17724	941	0.078
Race (ref: Non-Hispanic Black)				
Non-Hispanic White	47258	37145	57370	<0.001
Other	28997	12871	45122	<0.001
Age at onset (ref: ≥50)				
<30	35089	26550	43629	<0.001
≥30 and <40	36930	27358	46503	<0.001
≥40 and <50	34803	24875	44732	<0.001
Time since onset (ref: <10)				
≥10 and ≤19	9496	2030	16962	0.013
≥20	17913	10629	25197	<0.001
Education (ref: ≤High School)				
2-year degree/trade school	24545	17040	32050	<0.001
4-year degree	57308	49150	65466	<0.001
Postgraduate	64758	55176	74339	<0.001
Painful days (ref: >20 pain days)				
0–5	11393	1711	21075	0.021
6–20	6374	−2909	15656	0.178
Painful conditions (ref: 3–5)				
0–1	21399	12171	30626	<0.001
2	21301	12503	30100	<0.001
Average pain intensity (ref: ≥5)				
≤2	542	−8671	9755	0.908
3–4	−1083	−8816	6650	0.784
Pain medications (ref: Daily use)				
Never use	23916	16123	31708	<0.001
Sometimes use	10063	1566	18560	0.020
Nonprescription opioid use (ref: yes)				
No	12054	−8619	32726	0.253

Comparing intreg and extintreg

- Injury severity C5-C8 becomes significant in unconditional earnings as does all age categories, and time since onset 20-19 years.
- Painful days 0-5 and all painful conditions becomes significant as well.
- Finally, pain medications becomes significant with never use being associated with the highest earnings.

Wait Time Example

- Socioeconomic inequalities in waiting times for primary care across ten OECD countries. (3)
- Waiting time measured by time reported to see an MD or RN from Commonwealth Fund survey.
- Interval regression used since responses are in intervals. Eg. Same day, Next Day, 2-5 days, 6-7 days, 8-14 days, more than 2 weeks, never, with a separate model for each country.

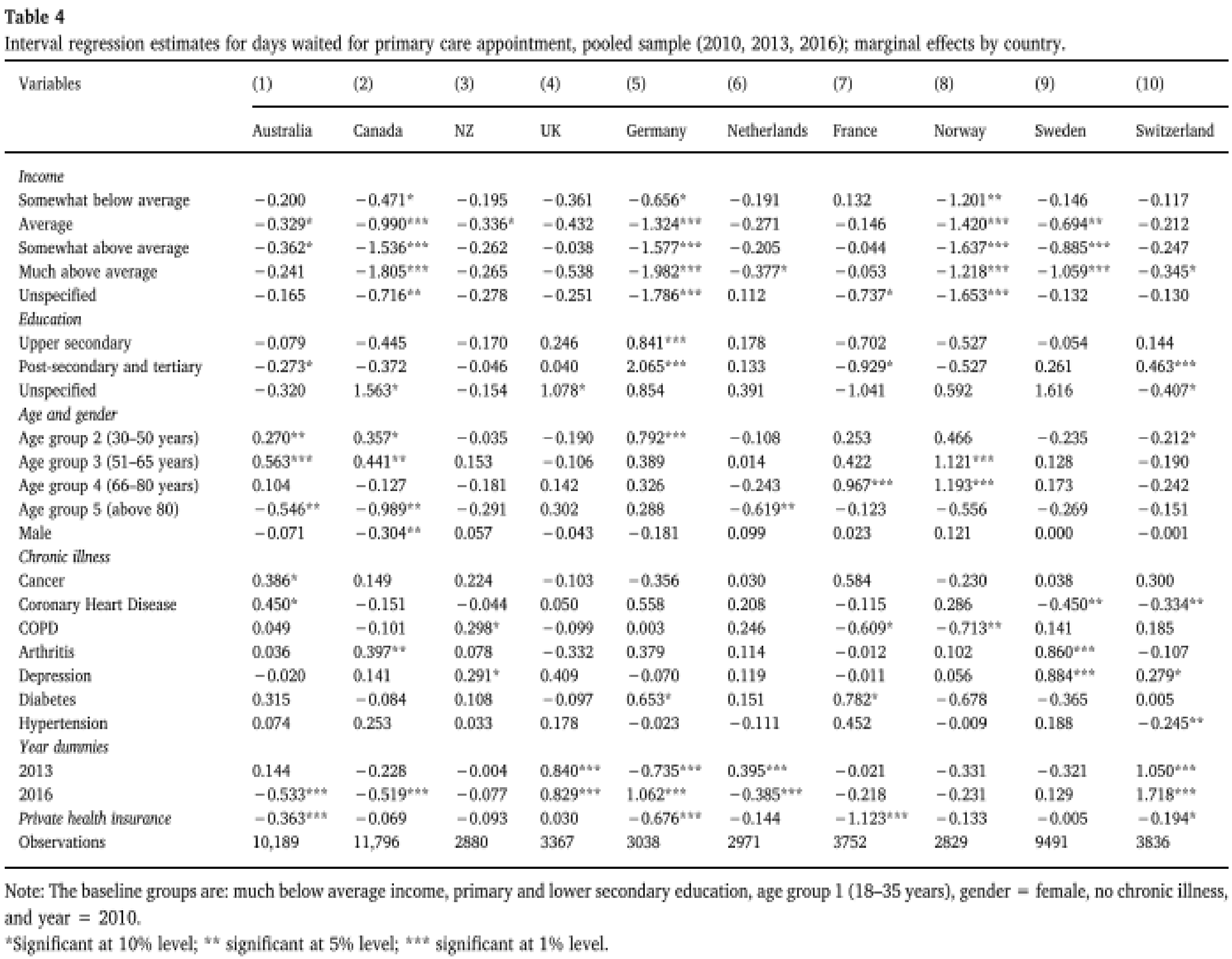


Table 4

Interval regression estimates for days waited for primary care appointment, pooled sample (2010, 2013, 2016); marginal effects by country.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Australia	Canada	NZ	UK	Germany	Netherlands	France	Norway	Sweden	Switzerland
Income										
Somewhat below average	−0.200	−0.471*	−0.195	−0.361	−0.656*	−0.191	0.132	−1.201**	−0.146	−0.117
Average	−0.329*	−0.990***	−0.336*	−0.432	−1.324***	−0.271	−0.146	−1.420***	−0.694**	−0.212
Somewhat above average	−0.362*	−1.536***	−0.262	−0.038	−1.577***	−0.205	−0.044	−1.637***	−0.885***	−0.247
Much above average	−0.241	−1.805***	−0.265	−0.538	−1.982***	−0.377*	−0.053	−1.218***	−1.059***	−0.345*
Unspecified	−0.165	−0.716**	−0.278	−0.251	−1.786***	0.112	−0.737*	−1.653***	−0.132	−0.130
Education										
Upper secondary	−0.079	−0.445	−0.170	0.246	0.841***	0.178	−0.702	−0.527	−0.054	0.144
Post-secondary and tertiary	−0.273*	−0.372	−0.046	0.040	2.065***	0.133	−0.929*	−0.527	0.261	0.463***
Unspecified	−0.320	1.563*	−0.154	1.078*	0.854	0.391	−1.041	0.592	1.616	−0.407*
Age and gender										
Age group 2 (30–50 years)	0.270**	0.357*	−0.035	−0.190	0.792***	−0.108	0.253	0.466	−0.235	−0.212*
Age group 3 (51–65 years)	0.563***	0.441**	0.153	−0.106	0.389	0.014	0.422	1.121***	0.128	−0.190
Age group 4 (66–80 years)	0.104	−0.127	−0.181	0.142	0.326	−0.243	0.967***	1.193***	0.173	−0.242
Age group 5 (above 80)	−0.546**	−0.989**	−0.291	0.302	0.288	−0.619**	−0.123	−0.556	−0.269	−0.151
Male	−0.071	−0.304**	0.057	−0.043	−0.181	0.099	0.023	0.121	0.000	−0.001
Chronic illness										
Cancer	0.386*	0.149	0.224	−0.103	−0.356	0.030	0.584	−0.230	0.038	0.300
Coronary Heart Disease	0.450*	−0.151	−0.044	0.050	0.558	0.208	−0.115	0.286	−0.450**	−0.334**
COPD	0.049	−0.101	0.298*	−0.099	0.003	0.246	−0.609*	−0.713**	0.141	0.185
Arthritis	0.036	0.397**	0.078	−0.332	0.379	0.114	−0.012	0.102	0.860***	−0.107
Depression	−0.020	0.141	0.291*	0.409	−0.070	0.119	−0.011	0.056	0.884***	0.279*
Diabetes	0.315	−0.084	0.108	−0.097	0.653*	0.151	0.782*	−0.678	−0.365	0.005
Hypertension	0.074	0.253	0.033	0.178	−0.023	−0.111	0.452	−0.009	0.188	−0.245**
Year dummies										
2013	0.144	−0.228	−0.004	0.840***	−0.735***	0.395***	−0.021	−0.331	−0.321	1.050***
2016	−0.533***	−0.519***	−0.077	0.829***	1.062***	−0.385***	−0.218	−0.231	0.129	1.718***
Private health insurance	−0.363***	−0.069	−0.093	0.030	−0.676***	−0.144	−1.123***	−0.133	−0.005	−0.194*
Observations	10,189	11,796	2880	3367	3038	2971	3752	2829	9491	3836

Note: The baseline groups are: much below average income, primary and lower secondary education, age group 1 (18–35 years), gender = female, no chronic illness, and year = 2010.

*Significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Main Points

- When data is reported in intervals, OLS is biased.
- Intreg and many extensions are available in STATA
- R package smicd
- SAS package Proc Lifereg
- Coefficients are interpreted directly

References

- 1. Conroy RM. Stings in the tails: Detecting and dealing with censored data. The Stata Journal (2005) 5, Number 3, pp. 395–404.
- 2. Walter P. The R Package smicd: Statistical Methods for Interval Censored Data. https://cran.r-project.org/web/packages/smicd/vignettes/vignetteSmicd.pdf.

References

- 3. Martin S, Siciliani L, Smith P. Socioeconomic inequalities in waiting times for primary care across ten OECD countries. Social Science & Medicine 263(2020).
- 4. Krause J, Dismuke-Greer CE, Reed KS, Li C. Employment status, hours working, and gainful earnings after spinal cord injury: relationship with pain, prescription medications for pain, and nonprescription opioid use. Spinal Cord (2020) 58:275-293.

Questions?

For more information contact:

Clara.Dismuke@va.gov

or

Visit the HERC website at
www.herc.research.va.gov

Or

Email us at HERC@va.gov

Or

Call us at (650) 617-2630