

ONLINE SUPPLEMENTARY MATERIAL

1. Detailed methods

1.1. Data source

The Medical Expenditure Panel Survey (MEPS) started in 1996 and is a co-sponsored initiative from both the Agency for Healthcare Research and Quality (AHRQ), and the National Centre for Health Statistics¹. There are two major components to MEPS: Household Component and Insurance Component. In this study we used the Household Component of MEPS (MEPS-HC), a nationally representative survey of the non-institutionalized, civilian US population, to estimate health care usage and expenditure. We pooled MEPS-HC data collected over 10 calendar years (2006-2015) to increase the power of statistical analyses and produce estimates that are representative of the US population. The 2015 MEPS-HC is the latest version of MEPS-HC data available at the time of our study.

At the start of each year, the Agency for Healthcare Research and Quality (AHRQ) selects a new panel of sample households for MEPS¹. Each panel consists of five rounds of interviews that takes place over a two-year period with an average recall period of 4-6 months within each assessment. Within one calendar year, there are two overlapping panels of samples. This allows for a continuous and up-to-date estimate at both the individual and household level.

1.2. Explanatory variables

Potential predictors in our models include comorbidities and sociodemographic variables. Comorbidities in our study sample were identified based on ICD-9-CM codes. Sociodemographic factors included age; sex (female and male); race/ethnicity (White, Black, Hispanic and other racial minorities); insurance coverage (uninsured, any private and public only); marital status (married and unmarried); geographic region (Northeast, Midwest, South and West); education level (high school not completed, high school completed, bachelor degree, postgraduate degree and other degree); income level (poor, near poor, low income, middle income and high income); and current smoking status (smoker and non-smoker). Income level in a specific year was based on family household income as a percentage of the federal poverty line (FPL; national level) for that year, accounting for family size and composition, with poor people defined as having an income <100% FPL, near poor as $\geq 100\%$ and <125% FPL, low-income as $\geq 125\%$ and <200% FPL, middle income as $\geq 200\%$ to <400% FPL, and high income as $\geq 400\%$ of FPL¹. For a FPL for a family with specific numbers of adults and children in a specific year, see the website of the United States Census Bureau². For example, the FPL for a family of two adults and two children in 2015 was \$24,036.

1.3. Variable selection in models for healthcare expenditure and work-related outcomes

Observations in MEPS had different probabilities of selection (sampling weights) and were sampled as groups (clustering) with different groups of clusters sampled separately (stratification), and therefore these characteristics were taken into account in our analyses to avoid biases in the point and standard error estimates. Since the number of predictors was large and the stepwise procedure was not supported for the analyses accounting for sampling weights, clustering and stratification, we performed univariable regression to assess the association of each variable with the outcome and selected variables with a p-value smaller than 0.05 to include the final models for health care expenditure, employment status and income (see Tables S1-S3 below). Because we used the zero-inflated negative binomial model (ZINB) for absenteeism which consists of two components (see below for details), and the selection of variables in each component needs to be done simultaneously in the same modelling process, the use of the univariable regression to select variables was deemed inappropriate. Therefore, we used the full model for absenteeism. We also had a face validation of the selected variables with members of our team and other experts in epidemiology, rheumatology and health economics.

1.4. Statistical modelling

We performed statistical analyses using Stata v. 15.1 (Stata Corp., College Station, TX, USA). Because we pooled data across 10 years, we adjusted the person-level weights by dividing them by 10 to obtain annual estimates on the population in the period 2006-2015 following the method described in the AHRQ³. Due to the clustering and stratification panel design of MEPS, we used the stratum of the primary sampling unit and the primary sampling unit variables from the HC-036 Pooled Estimation Linkage File for all variance estimation³. We calculated all standard errors using the Taylor-Series Linearization procedure with replacement as described by the AHRQ.

Given that the observed days of sick leave consisted of excess zeros (> 60% of all observations), we used the ZINB to examine the association between comorbidities and days of sick leave. The ZINB model consisted of two components: the zero-inflation component was a logit model of the probability that sick leave was undetected, resulting in excess zeros in the number of sick leave days, and the count component was a negative binomial model of the true count of sick leave days. We chose the negative binomial distribution in the second component because the positive count of sick leave days in our data were over-dispersed (i.e. the variance is much larger than the mean).

2. Supplementary tables

Supplementary Table S1. Univariable GLM results for total annual healthcare expenditure in patients with rheumatoid arthritis.

Variable	Mean ratio*	95% C.I.	p-value
Hypertension	1.43	(1.23, 1.66)	<0.001‡
Disorders of lipid metabolism	1.26	(1.11, 1.43)	<0.001‡
Diabetes	1.49	(1.30, 1.71)	<0.001‡
Depression	1.52	(1.29, 1.78)	<0.001‡
Disorders of the eye and adnexa	1.48	(1.27, 1.72)	<0.001‡
Asthma	1.51	(1.26, 1.82)	<0.001‡
Chronic obstructive pulmonary disease	1.67	(1.45, 1.93)	<0.001‡
Cancer	1.45	(1.23, 1.71)	<0.001‡
Myocardial infarction	1.70	(1.41, 2.04)	<0.001‡
Osteoporosis	1.37	(1.08, 1.72)	0.008†
Stroke	1.69	(1.40, 2.04)	<0.001‡
Heart failure	2.42	(1.90, 3.07)	<0.001‡
Peptic ulcer	1.24	(0.94, 1.63)	0.128
Sex (Male as reference)			
Female	1.21	(1.05, 1.39)	0.009†
Age	1.01	(1.00, 1.02)	0.001†
Region (West as reference)			
Northeast	1.31	(1.04, 1.65)	0.021†
Midwest	1.27	(1.04, 1.56)	0.021†
South	1.06	(0.90, 1.26)	0.465
Marital status (Unmarried as reference)			
Married	0.99	(0.86, 1.15)	0.937
Race/ethnicity (White as reference)			
Black	0.91	(0.78, 1.06)	0.214

Variable	Mean ratio*	95% C.I.	p-value
Hispanic	0.89	(0.62, 1.28)	0.546
Other	0.84	(0.68, 1.05)	0.122
Income level (Poor/negative as reference)			
Near poor	1.31	(1.03, 1.67)	0.029†
Low income	1.01	(0.83, 1.24)	0.915
Middle income	1.15	(0.95, 1.38)	0.145
High income	1.06	(0.88, 1.29)	0.533
Insurance coverage (Public as reference)			
Any private	0.88	(0.76, 1.02)	0.094
No insurance	0.45	(0.37, 0.55)	<0.001‡
Education (High school not completed as reference)			
High school completed	1.00	(0.85, 1.17)	0.610
Bachelor degree	1.38	(1.06, 1.80)	0.017†
Postgraduate degree	1.18	(0.87, 1.60)	0.293
Other degree	1.29	(1.07, 1.56)	0.006†
Smoking status (Non-smoker as reference)			
Smoker	0.80	(0.68, 0.95)	0.011†

* Values are the exponential form of the beta coefficients for a covariate, which represents the multiplier factor for the outcome when the covariate increases by one; Reference for each comorbidity is absence of condition.

† P < 0.05

‡ P < 0.001

Supplementary Table S2. Univariable logistic regression results for employment status in working-age patients (ages 18-64 years) with rheumatoid arthritis.

Variable*	Odds ratio	95% C.I.	p-value
Hypertension	0.46	(0.37, 0.56)	<0.001‡
Disorders of lipid metabolism	0.58	(0.47, 0.72)	<0.001‡
Diabetes	0.48	(0.37, 0.63)	<0.001‡
Depression	0.41	(0.31, 0.54)	<0.001‡
Disorders of the eye and adnexa	0.71	(0.51, 0.99)	0.048†
Asthma	0.46	(0.34, 0.63)	<0.001‡
Chronic obstructive pulmonary disease	0.26	(0.18, 0.36)	<0.001‡
Cancer	0.46	(0.31, 0.70)	<0.001‡
Myocardial infarction	0.38	(0.24, 0.61)	<0.001‡
Osteoporosis	0.45	(0.26, 0.76)	0.003†
Stroke	0.27	(0.16, 0.46)	<0.001‡
Heart failure	0.04	(0.01, 0.11)	<0.001‡
Peptic ulcer	0.32	(0.13, 0.81)	0.017†
Sex (Male as reference)			
Female	0.69	(0.54, 0.88)	0.003†
Age	0.97	(0.96, 0.98)	<0.001‡
Region (West as reference)			
Northeast	1.33	(0.92, 1.93)	0.133
Midwest	1.20	(0.85, 1.71)	0.296
South	0.88	(0.64, 1.21)	0.424
Marital status (Unmarried as reference)			
Married	1.61	(1.29, 2.02)	<0.001‡
Race/ethnicity (White as reference)			
Black	0.62	(0.49, 0.78)	<0.001‡
Hispanic	0.92	(0.65, 1.29)	0.627

Variable*	Odds ratio	95% C.I.	p-value
Other	0.77	(0.57, 1.03)	0.081
Insurance coverage (Public as reference)			
Any private	12.46	(9.18, 16.92)	<0.001‡
No insurance	4.42	(3.05, 6.40)	<0.001‡
Education (High school not completed as reference)			
High school completed	2.05	(1.55, 2.71)	<0.001‡
Bachelor degree	4.03	(2.49, 6.52)	<0.001‡
Postgraduate degree	6.64	(3.61, 12.22)	<0.001‡
Other degree	2.28	(1.64, 3.19)	<0.001‡
Smoking status (Non-smoker as reference)			
Smoker	0.60	(0.47, 0.77)	<0.001‡

* Reference for each comorbidity is the absence of the condition.

† P < 0.05

‡ P < 0.001

Supplementary Table S3. Univariable GLM results for income in patients with rheumatoid arthritis.

Variable	Mean ratio*	95% C.I.	p-value
Hypertension	0.89	(0.78, 1.02)	0.102
Disorders of lipid metabolism	1.04	(0.90, 1.22)	0.573
Diabetes	0.79	(0.68, 0.92)	0.003†
Depression	0.84	(0.70, 1.00)	0.054
Disorders of the eye and adnexa	1.06	(0.87, 1.28)	0.564
Asthma	0.79	(0.65, 0.96)	0.017†
Chronic obstructive pulmonary disease	0.63	(0.50, 0.78)	<0.001‡
Cancer	0.89	(0.73, 1.08)	0.239
Myocardial infarction	0.92	(0.67, 1.26)	0.610
Osteoporosis	0.66	(0.13, 3.36)	0.611
Stroke	0.94	(0.62, 1.41)	0.759
Heart failure	0.41	(0.27, 0.63)	<0.001‡
Peptic ulcer	0.57	(0.40, 0.80)	0.001†
Sex (Male as reference)			
Female	0.77	(0.68, 0.89)	<0.001‡
Age	1.00	(0.99, 1.01)	0.881
Region (West as reference)			
Northeast	1.06	(0.81, 1.38)	0.657
Midwest	0.80	(0.65, 0.97)	0.028†
South	0.87	(0.74, 1.03)	0.108
Marital status (Unmarried as reference)			
Married	1.16	(1.00, 1.35)	0.051
Race/ethnicity (White as reference)			
Black	0.85	(0.73, 1.00)	0.044†
Hispanic	0.85	(0.68, 1.08)	0.180

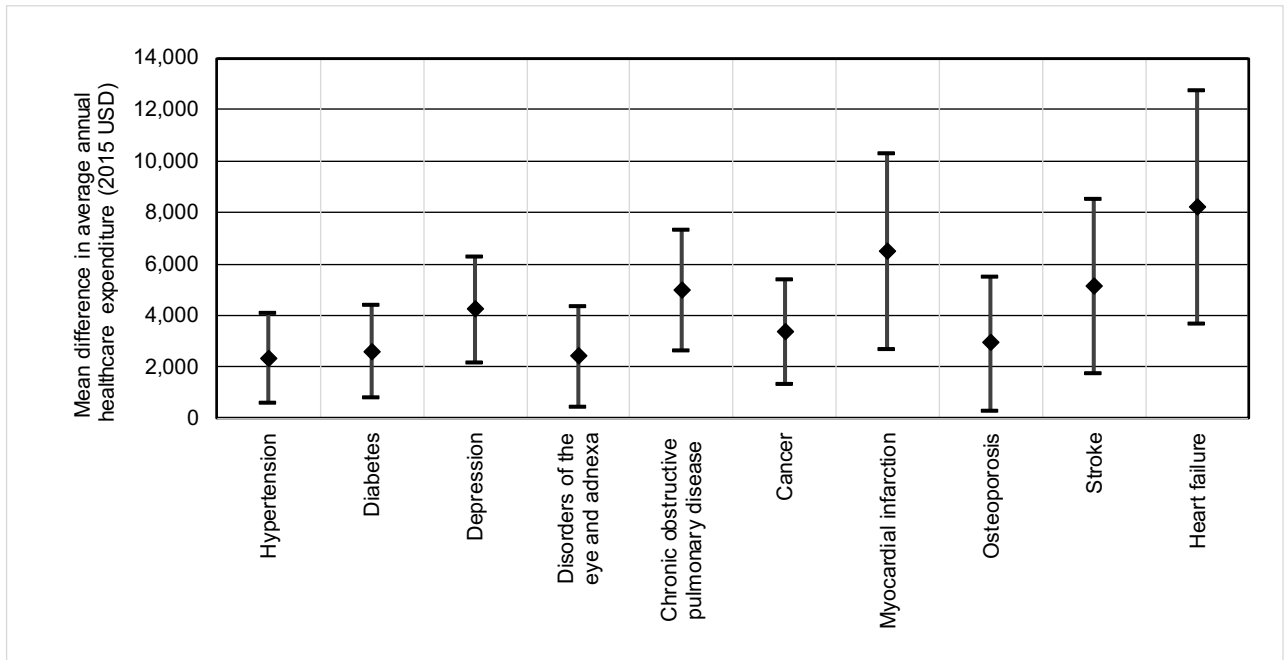
Variable	Mean ratio*	95% C.I.	p-value
Other	1.12	(0.84, 1.49)	0.442
Insurance coverage (Public as reference)			
Any private	2.26	(1.83, 2.79)	<0.001‡
No insurance	1.05	(0.82, 1.36)	0.686
Education (High school not completed as reference)			
High school completed	1.13	(0.82, 1.54)	0.453
Bachelor degree	2.05	(1.47, 2.86)	<0.001‡
Postgraduate degree	2.25	(1.59, 3.19)	<0.001‡
Other degree	1.40	(1.00, 1.96)	0.047†
Smoking status (Non-smoker as reference)			
Smoker	0.81	(0.69, 0.94)	0.008†

* Values are the exponential form of the beta coefficients for a covariate, which represents the multiplier factor for the outcome when the covariate increases by one; Reference for each comorbidity is absence of condition.

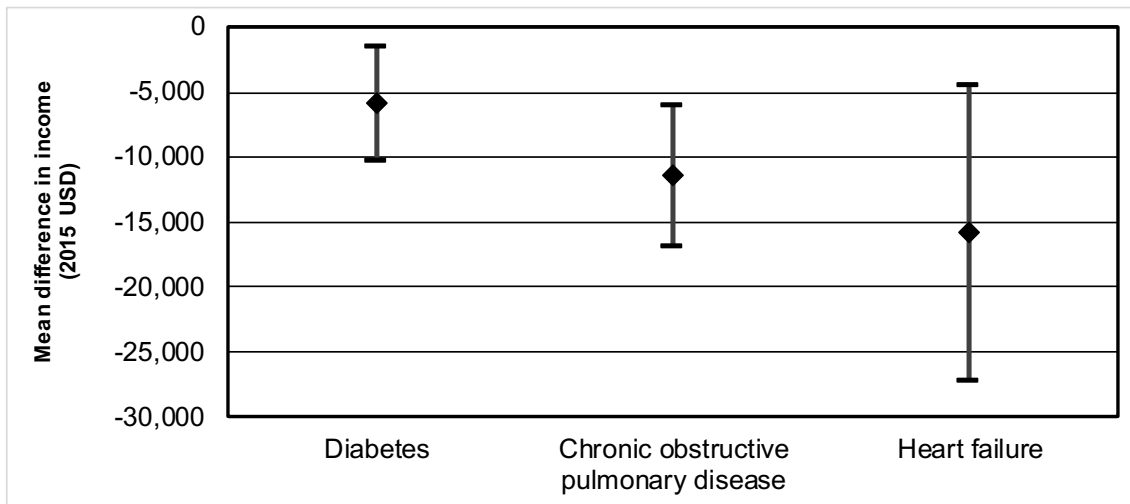
† P < 0.05

‡ P < 0.001

3. Supplementary figures



Supplementary Figure S1. Adjusted mean difference in total annual healthcare expenditure in patients with rheumatoid arthritis by comorbidity. Bars represent 95% confidence intervals.



Supplementary Figure S2. Adjusted mean difference in income in patients with rheumatoid arthritis by comorbidity. Bars represent 95% confidence intervals.

References

1. Medical Expenditure Panel Survey. MEPS HC-181 2015 full year consolidated data file. [Internet. Accessed February 20, 2020] Available from: meps.ahrq.gov/data_stats/download_data/pufs/h181/h181doc.pdf.
2. United States Census Bureau. Poverty thresholds. [Internet. Accessed 9 October 2020] Available from: <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html>.
3. Medical Expenditure Panel Survey. MEPS HC-036: 1996-2015 pooled linkage variance estimation file. [Internet. Accessed February 20, 2020] Available from: meps.ahrq.gov/data_stats/download_data/pufs/h36/h36u15doc.pdf.