

Propensity score analysis exploring the impact of smoking and drinking on the prognosis of patients with oral cancer

Xiaodan Bao BS^{1,2} | Fengqiong Liu PhD^{1,2} | Qing Chen BS^{1,2} | Lin Chen BS^{1,2} | Jing Lin BS^{1,2} | Fa Chen PhD^{1,2} | Jing Wang MD³ | Yu Qiu MD⁴ | Bin Shi PhD⁴ | Lizhen Pan BS⁴ | Lisong Lin PhD⁴ | Baochang He PhD^{1,2} 

¹Department of Epidemiology and Health Statistics, Fujian Provincial Key Laboratory of Environment Factors and Cancer, School of Public Health, Fujian Medical University, Fujian, China

²Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Fujian Key Laboratory of Tumor Microbiology, Fujian Medical University, Fujian, China

³Laboratory Center, The Major Subject of Environment and Health of Fujian Key Universities, School of Public Health, Fujian Medical University, Fujian, China

⁴Department of Oral and Maxillofacial Surgery, The First Affiliated Hospital of Fujian Medical University, Fujian, China

Correspondence

Baochang He, Department of Epidemiology and Health Statistic, School of Public Health, Fujian Medical University, 1 Xueyuan Road, Fuzhou 350108, China.
Email: prof_hbc@163.com

Funding information

Startup Fund for Scientific Research of Fujian Medical University, Grant/Award Number: 2017XQ1011; High-level Talents research Start-up Project of Fujian Medical University, Grant/Award Number: XRCZX2018001; Joint Funds for the Innovation of Science and Technology of Fujian province, Grant/Award Number: 2017Y9103; Scientific Research Talents Training Project of Health and Family Planning Health Commission in Fujian Province, Grant/Award Numbers: 2018-1-71, 2017-ZQN-57; Program for New Century Excellent Talents in Fujian Province University, Grant/Award Number: 2018B029; Fujian Natural Science Foundation Program, Grant/Award Number: 2019J01314

Abstract

Background: To explore the effect of smoking and drinking on survival of patients with oral cancer by comparing the characteristics and survival of nonsmoking and nondrinking (NSND) patients in contrast to smoking and/or drinking (SD) patients.

Methods: This prospective study including 1165 patients with oral cancer was conducted in Fujian, China from January 2005 to January 2019. The patients were categorized to two groups, the NSND group and SD group. We compared overall survival and disease-specific survival between the two groups using the Kaplan-Meier method and Cox proportional hazards regression before and after propensity score matching (PSM) to explore the effect of smoking and drinking on the prognosis of patients with oral cancer.

Results: NSND patients accounted for 55.45% (646 patients) of all the patients with oral cancer. SD patients with oral cancer tended to be older and mainly are male (98.46%) and with more advanced disease status. There are trends toward both higher risk of all-cause death ($HR = 1.678$; 95% CI: 1.086-2.594) and oral cancer specific death ($HR = 1.632$; 95% CI: 1.044-2.552) in SD patients with oral cancer before PSM. After PSM, the association is still significant, with adjusted HR of 1.897 (95% CI: 1.138-3.165) for all-cause death and adjusted HR of 1.764 (95% CI: 1.043-2.983) for oral cancer-specific

Xiaodan Bao and Fengqiong Liu contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Head & Neck* published by Wiley Periodicals, Inc.

death. Additionally, PSM can improve the HR value and result in a stronger association.

Conclusions: Social and clinical characteristics of NSND patients differed from SD patients with oral cancer. SD patients with oral cancer have higher all-cause mortality and oral cancer-specific mortality than NSND patients.

KEY WORDS

all-cause mortality, nondrinking, nonsmoking, oral cancer, oral cancer-specific mortality

1 | INTRODUCTION

Oral cancer is malignant tumor seriously endangering quality of life and leading to significant mortality.¹⁻³ According to recent data from the GLOBOCAN 2018, the age-standardized mortality of oral cancer was 2.8 of 100000 for male and 1.2 of 100000 for female.⁴

It has been well established that tobacco smoking and alcohol consumption are major risk factors for the development of oral cancer. In addition, the combination of tobacco smoking and alcohol consumption has a synergistic effect on oral cancer.⁵⁻⁹ In recent years, nonsmoking and nondrinking (NSND) patients have been identified as a distinct and growing subset of oral cancer. Nevertheless, the survival outcomes of this specific group patients have been largely neglected.

Results from the current few studies have been controversial, with some studies pointing out both smoking and drinking were linked to a higher risk of death in oral cancer,¹⁰⁻¹² while other studies showed that the prognosis of NSND oral squamous cell carcinoma (OSCC) patients is no different from that of smoking and/or drinking (SD) patients,¹³ or even worse.¹⁴ In fact, data has showed that NSND patients with oral cancer are younger and more likely to be female and with earlier disease stages.^{13,14} In addition, previous studies have reported that smokers have a higher prevalence of genetic mutations than nonsmokers.^{15,16} Differences in the distribution of these confounders may lead to the controversial results observed in prognosis of NSND patients. Propensity score matching (PSM) could equalize the initially differences between two groups and increase the reliability of results. However, thus far, no prospective study about the prognosis of NSND using PSM has been reported.

The objective of this study is to compare the social and clinical characteristics, as well as the prognosis of

NSND and SD patients with oral cancer from southeast China by using PSM.

2 | PATIENTS (OR SUBJECTS) AND METHODS

2.1 | Study population

A total of 1165 patients with oral cancer were consecutively recruited from The First Affiliated Hospital of Fujian Medical University from January 2005 to August 2018. All patients were histologically confirmed primary oral cancer cases. As described previously,¹⁷ all patients were Chinese Han population and resided in Fujian Province for more than 10 years. Patients who with recurrent oral cancer or distant metastasized cancer were excluded. This prospective study was approved by the Institutional Review Board (IRB) of Fujian Medical University (Fuzhou, China) and performed in line with the ethical standards described in the Declaration of Helsinki.

2.2 | Data collection

Tumor characteristics (TNM stage, pathological grading, pathological type, etc.) were retrieved from patients' medical records. Demographic characteristics (occupation, education level, origin, etc.) and lifestyle habits (including smoking and drinking status) data were collected via interview-based structured and self-reported questionnaires distributed at diagnosis. Details of the treatment have been provided in previous reports.¹⁸ Mortality data were sourced from telephone interview with 6-month intervals or hospital patient attendance records. We also verify the death of patients through national death surveillance system. The cause of death was recorded in the Resident Medical Certificate of Death Form which is

issued by physician. The follow-up was consistent until January 2019.

2.3 | Definition of variables

This study adopted a uniform InterCHANGE lifestyle questionnaire, which was compiled by the university of Utah and the institute of oncology, Chinese academy of medical sciences, united with famous epidemiologists in China and the actual situation in China. Subjects were asked to report the total number of years smoking and the average cigarettes consumption level per day. As for alcohol drinking, information about average number of drinks per week and total number of years drinking was collected.

All patients were staged using the AJCC staging system, version 7, 2010. The AJCC staging system before 2010 was slightly differ from systems used after 2010, so we updated the disease staging information before 2010 using clinical pathological reports according to the new AJCC staging system. All-cause mortality was defined as the death from any cause. And the second outcome of interest was death from oral cancer. The occupational classification was determined by the professional classification of the People's Republic of China. Farmers refer to agricultural, forestry, animal husbandry, and fishery production personnel; workers refer to production and transportation operators. Office workers refer to staff of the government or the principal of a group or a unit, professional and technical personnel, clerical and related personnel. Illiteracy refers to those who have never been to school and cannot read. The classification of urban or rural areas was determined according to the administrative division of China. The origin is derived from household registry information that includes the birth place and residence of the participant.

2.4 | Statistical analysis

The cumulative smoking exposure was calculated as pack-years of smoking which was the product of the number of packs consumed per day and number of years of smoking. Sensitivity analysis was conducted with regard to smoking and alcohol consumption to define appropriate threshold for smoking and drinking, and to define the SD and NSND group.

Chi-square test was used to compare demographic characteristics between NSND group and SD group. The survival curves were estimated by the Kaplan-Meier method and compared by the log-rank test. Univariate and multivariate Cox regression models were used to

assess the effects of smoking and drinking. A propensity score (PS) applying the method of nearest neighbor matching within a specified caliper distance (calipers of width equal to 0.02) without replacement was used to assess the effects of smoking and drinking, in order to minimize the potential confounding bias which could influence the results. The NSND group was 1:1 matched on age, gender, occupation, education level, residence, BMI, TNM stage, and pathological type with SD group. All statistical analyses were performed with the R software version 3.6.1. A *P* value of $<.05$ was taken to be significant.

3 | RESULTS

Sensitivity analysis was conducted using different threshold of smoking, the effect of smoking was more significant when the cutoff point for smoking was 10 pack years or 20 pack years compared with 100 cigarettes in life time especially for oral cancer specific mortality (Table 1). While the effect of drinking was more significant when the cutoff point for drinking was at least seven drinks per week continuously for at least 1 year compared with at least one drink a week continuously for at least 6 months (Table 1). We also analyzed the prognosis of patients according to different levels smoking duration and smoking frequency (Table S1), and found that longer duration and higher consumption frequency is positively related to worse prognosis. In addition to smoking, we also explored the relationship between different alcohol drinking duration and frequency and prognosis of oral cancer, and found that patients who drank more than seven times a week had significantly poorer overall survival (Table S1).

Based on the results of sensitivity analysis, smokers (ex and current smoker) were defined as those who had smoked at least 10 pack years cigarettes in their lives. Patients were considered to be drinkers (ex and current drinker) if they consumed alcohol at least seven drinks per week continuously for at least 1 year. Patients were categorized as NSND or SD. A total of 646 patients (55.45%) were identified as NSND from 1165 patients with oral cancer, while SD group patients represent approximately 44.55% (519) of the oral cancer population. A peak at 60 to 70 years was observed in the NSND group, while a peak at 50 to 60 years was seen in SD group (Figure 1). Moreover, we observed extremely difference in sex distribution between NSND and SD groups (Figure 1). The SD group was predominantly male (98.46%), while the NSND group was dominated by female patients (65.63%). NSND group has higher proportion of patients with $BMI \geq 24 \text{ kg/m}^2$ than SD group. In

TABLE 1 Survival analysis for SD vs NSND patients with oral cancer based on different threshold of smoking and drinking classification in full cohort

	Number of patients (%)	All-cause mortality HR (95%CI)	Oral cancer specific mortality HR (95%CI)
Threshold of smoking			
100 cigarettes ^a			
No	620 (53.22)	1.000	1.000
Yes	545 (46.78)	1.586 (1.001, 2.512)	1.564 (0.975, 2.507)
10 pack years ^b			
No	680 (58.37)	1.000	1.000
Yes	485 (41.63)	1.842 (1.195, 2.840)	1.782 (1.143, 2.777)
20 pack years ^c			
No	731 (62.75)	1.000	1.000
Yes	434 (37.25)	1.743 (1.166, 2.604)	1.666 (1.104, 2.513)
Threshold of drinking			
>1 times/week ^d			
No	756 (64.89)	1.000	1.000
Yes	408 (35.01)	1.029 (0.724, 1.462)	1.048 (0.729, 1.508)
>7 times/week ^e			
No	937 (80.43)	1.000	1.000
Yes	228 (19.57)	1.463 (1.064, 2.011)	1.317 (0.905, 1.915)

Note: All adjusted for age, gender, occupation, education level, residence, BMI, TNM stage, pathological grading and pathological type, pathological grading and adjuvant therapy.

^aSmokers (ex and current smoker) were defined as those who had smoked at least 100 cigarettes in their lives.

^bSmokers (ex and current smoker) were defined as those who had smoked at least 10 pack years cigarettes in their lives.

^cSmokers (ex and current smoker) were defined as those who had smoked at least 20 pack years cigarettes in their lives.

^dDrinkers (ex and current smoker) were defined as at least one drink a week continuously for at least 6 months.

^eDrinkers (ex and current smoker) were defined as at least seven drink per week continuously for at least 1 year.

addition, NSND patients were more likely to present at earlier disease stage, with 44.16% in I and II stages, 55.83% at T1 and T2 classifications; however, no difference was observed in N classification. Details of demographics and clinical characteristics of all patients were listed in Table 2.

Before PSM, log-rank test showed that patients with SD had a significantly higher all-cause mortality ($P = .002$, Figure 2) and oral cancer-specific mortality ($P = .010$, Figure 2) than those with NSND. The variables that were not equally distributed between the groups and are potential confounders (age, gender, occupation, education level, residence, BMI, TNM stage, pathological type, pathological grading, and adjuvant therapy) were included and the enter method were used to develop the cox regression model models to adjust the potential confounding effect of those factors.

There are three cox regression model used: (a) In the crude model, no confounding factors were adjusted; (b) in the second cox model, only gender and age were

adjusted; and (c) in the full model, confounding factors such as age, gender, occupation, education level, residence, BMI, TNM stage, pathological type, pathological grading, and adjuvant therapy were adjusted. After fully adjustment of potential confounders, cox regression analysis showed that SD patients still had an increased risk of all-cause mortality (HR: 1.678, 95% CI: 1.086-2.594) and oral cancer specific mortality (HR: 1.632, 95% CI: 1.044-2.552) compared with the NSND group (Table 4).

PSM was performed in term of age, gender, occupation, education level, residence, BMI, TNM stage, and pathological type, since the distribution of these factors was unbalanced between SD group and NSND group. Out of 1165 patients, 185 NSND and 185 SD patients were selected, details of demographics and clinical characteristics of all patients were listed in Table 3. After PSM, higher all-cause mortality was still observed in patients with SD compared to those with NSND in both log rank test ($P = .003$, Figure 2) and cox regression analysis (adjusted HR: 1.897, 95% CI: 1.138-3.165). Similarly,

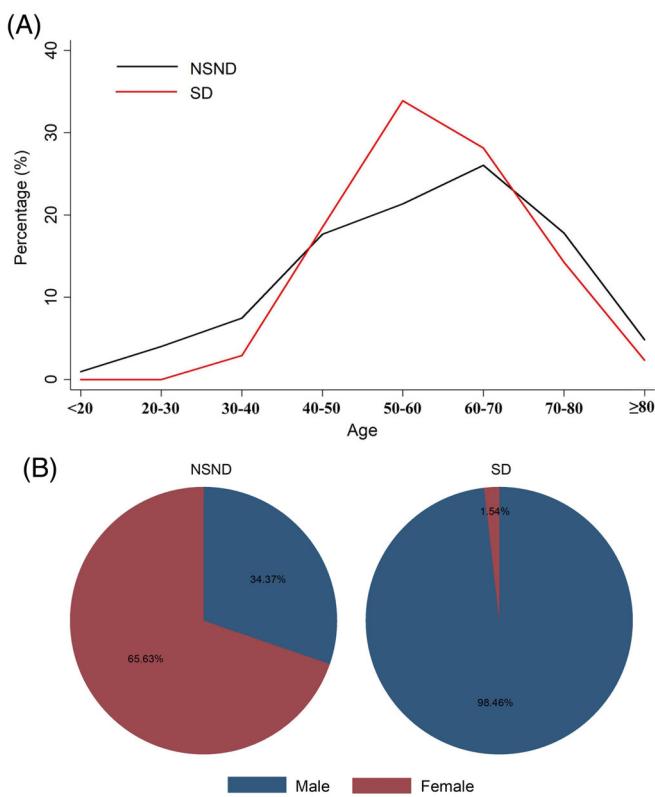


FIGURE 1 A, Age distribution of smoking and/or drinking (SD) patients compared to nonsmoker nondrinkers (NSND) patients. B, Sex distribution of SD patients compared to NSND patients [Color figure can be viewed at wileyonlinelibrary.com]

difference of oral cancer specific mortality was observed between NSND and SD patient in log rank test ($P = .013$, Figure 2) and cox regression (adjusted HR: 1.764, 95% CI: 1.043-2.983). Additionally, PSM can improve the HR value of both all-cause mortality (1.897 vs 1.678) and oral cancer specific mortality (1.764 vs 1.632) and enhance significance of association compared with the results of unmatching (Table 4).

4 | DISCUSSION

Oral cancer is traditionally a disease dominated by male smokers and drinkers. Our results showed that NSND patients accounted for 55.45% in patients with oral cancer, which was higher than the 20%- to 30% reported in the literatures.^{13,14,19} Compared to other studies in which 30% to 50% of NSND patients with oral cancer were female,^{20,21} our study population had a higher proportion of female (65.63%) in NSND patients. In general, the rates of smoking or drinking are lower for women in the southeastern China. Therefore, these findings indicate that there may be other important risk factors for oral cancer in the southeastern China besides smoking and

drinking, and more attention should be paid to the increasing NSND patient population.

Our study also showed that the average age of NSND patients was about 10 years older compared with SD patients, similar patterns were also repeated reported in numerous studies.^{13,14,21} This phenomenon suggested that tobacco and alcohol exposure may lead to development of oral cancer 10 years earlier than the non-exposure population. Although NSND patients have older age than SD patients, the prognosis of NSND is better than SD patients in terms of all-cause mortality, which indicated the protective effect of NSND on the population.

In addition to gender and age, data showed that NSND patients were less advanced than SD patients with oral cancer in TNM stage and pT classification. And patients in the NSND group had better nutritional status (BMI $\geq 24 \text{ kg/m}^2$ accounted for 27.40% in NSND group and 19.08% in SD group, respectively), which may partly explain the better prognosis observed in NSND group, since a large number of studies confirmed that patients with poor nutritional status may be more intolerant of the side effects of therapy and have adverse survival outcomes.^{22,23}

Largely owing to the differed distribution of important social and clinical relevant variables between the SD group and the NSND group, the prognosis regarding NSND has been controversial.¹⁰⁻¹⁴ Hence, we applied PSM, which is a matching method that could equalize the initially differences thus substantially decrease the bias in order to make a more reasonable comparison between groups in our study. This method was first proposed by Rosenbaum and Rubin in 1983²⁴ and is becoming more and more widely used.²⁵⁻²⁸ In our study, PSM method was adopted to make the SD group and NSND group match in many social and clinical characteristics including age, gender, occupation, education level, residence, BMI, TNM stage, and pathological type, and we found that the survival outcome of SD group was worse than that of NSND group in both before and after matching analysis. However, PSM can improve the HR value and enhance significance of association compared with results of unmatching, which validated the role of smoking and drinking in prognosis of oral cancer, and at the same time verified the potential value of PSM method.

Several limitations should be noticed when interpreting the results of this study. First, misclassification may occur because drinking and smoking are self-reported. Second, we only evaluated the prognostic value of the smoking and drinking status of patients with oral cancer at pretreatment. Hence, future studies are needed to collect information of posttreatment smoking and drinking status. Finally, this study is a single-center study

TABLE 2 Demographics of the patients in the SD group vs the NSND group before propensity score matching (N = 1165)

Variables	NSND (646) No. of patients (%)N	SD (519) No. of patients (%)N (%)	P
Sex			<.001
Male	222 (34.37)	511 (98.46)	
Female	424 (65.63)	8 (1.54)	
Age (years)			.001
<50	194 (30.03)	111 (21.39)	
≥50	452 (69.97)	408 (78.61)	
Occupation			<.001
Farmer	77 (12.11)	120 (23.21)	
Worker	353 (55.50)	230 (44.49)	
Office worker and other	206 (32.39)	167 (32.30)	
Education level			<.001
Illiteracy	101 (15.64)	26 (5.01)	
Primary-middle school	371 (57.43)	378 (72.83)	
High school and above	174 (26.93)	115 (22.16)	
Origin			.031
Urban area	349 (54.02)	313 (60.31)	
Rural area	297 (45.98)	206 (39.69)	
BMI (kg/m^2)			.002
18.5-23.9	394 (60.99)	340 (65.51)	
<18.5	75 (11.61)	80 (15.41)	
≥24	177 (27.40)	99 (19.08)	
TNM Stage			.001
I	112 (19.55)	54 (11.46)	
II	141 (24.61)	108 (22.93)	
III	106 (18.50)	84 (17.83)	
IV	214 (37.35)	225 (47.78)	
pT classification			<.001
T ₁	128 (22.61)	68 (14.38)	
T ₂	188 (33.22)	140 (29.60)	
T ₃	81 (14.31)	76 (16.07)	
T ₄	169 (29.86)	189 (39.95)	
pN classification			.529
N ₀	403 (70.70)	313 (66.74)	
N ₁	77 (13.51)	67 (14.29)	
N ₂	86 (15.09)	85 (18.12)	
N ₃	4 (0.70)	4 (0.85)	
Pathological grading			.887
Well	103 (21.73)	86 (20.67)	
Moderate	123 (25.95)	113 (27.16)	
Poor	248 (52.32)	217 (52.17)	
Pathological type			<.001
Squamous cell carcinoma	504 (78.63)	448 (86.65)	
Adenocarcinoma	137 (21.37)	69 (13.35)	

(Continues)

TABLE 2 (Continued)

Variables	NSND (646) No. of patients (%)N	SD (519) No. of patients (%)N (%)	P
Family history of cancer			.749
No	541 (83.75)	431 (83.04)	
Yes	105 (16.25)	88 (16.96)	
Adjuvant therapy			.323
No	324 (53.47)	248 (49.80)	
RT	101 (16.67)	82 (16.47)	
CT	64 (10.56)	70 (14.06)	
CRT	117 (19.30)	98 (19.67)	
Hypertension			.153
No	473 (73.22)	399 (76.88)	
Yes	173 (26.78)	120 (23.12)	
Diabetes			.126
No	578 (89.47)	478 (92.10)	
Yes	68 (10.53)	41 (7.90)	

Abbreviations: BMI, body mass index; CT, chemotherapy; CRT, chemoradiotherapy; NSND, nonsmoking and nondrinking; RT, radiotherapy; SD, smoking and/or drinking.

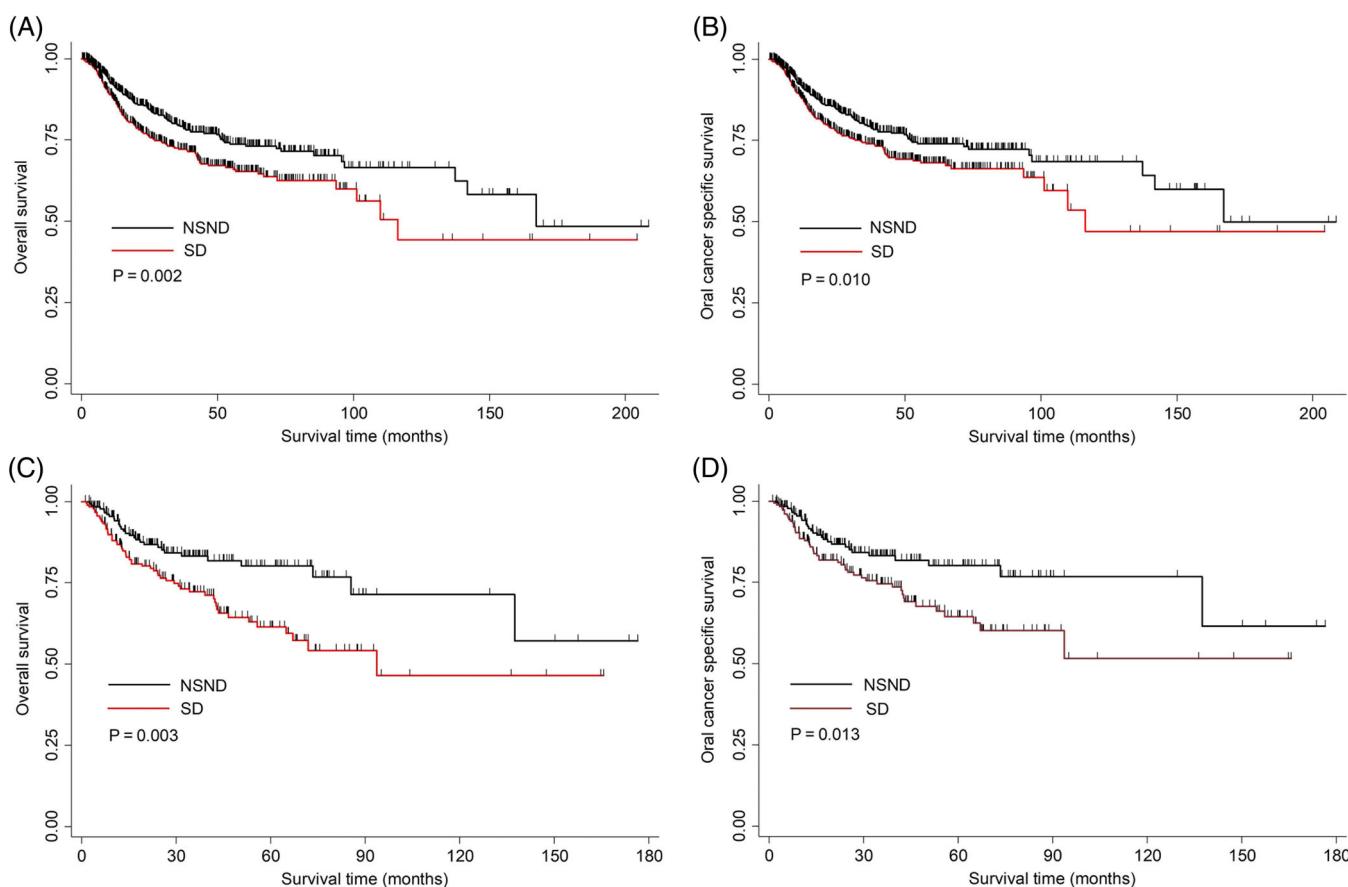


FIGURE 2 A, Overall survival rates of smoking and/or drinking (SD) patients and nonsmoker nondrinkers (NSND) patients before propensity score matching. B, Oral cancer specific survival rates of SD patients and NSND patients before propensity score matching. C, Overall survival rates of SD patients and NSND patients after propensity score matching. D, Oral cancer-specific survival rates of SD patients and NSND patients after propensity score matching [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Demographics in the SD group vs the NSND group after propensity score matching (N = 370)

Variables	NSND (185) No. of patients (%)	SD (185) No. of patients (%)	P
Sex			1.000
Male	179 (96.76)	179 (96.76)	
Female	6 (3.24)	6 (3.24)	
Age (years)			.051
<50	62 (33.51)	45 (24.32)	
≥50	123 (66.49)	140 (75.68)	
Occupation			.990
Farmer	32 (17.30)	31 (16.76)	
Worker	102 (55.14)	103 (55.68)	
Office worker and other	51 (27.57)	51 (27.57)	
Education level			.568
Illiteracy	11 (5.95)	12 (6.49)	
Primary-middle school	96 (51.89)	105 (56.76)	
High school and above	78 (42.16)	68 (36.76)	
Origin			.298
Urban area	91 (49.19)	101 (54.59)	
Rural area	94 (50.81)	84 (45.41)	
BMI (kg/m^2)			.569
18.5–23.9	116 (51.56)	106 (20.50)	
<18.5	16 (7.11)	18 (3.48)	
≥24	53	61 (11.80)	
TNM Stage			.136
I	27 (14.59)	41 (22.16)	
II	44 (23.79)	30 (16.22)	
III	35 (18.92)	34 (18.38)	
IV	79 (42.70)	80 (43.24)	
pT classification			.055
T ₁	30 (16.48)	44 (24.04)	
T ₂	58 (31.87)	37 (20.22)	
T ₃	29 (15.93)	31 (16.94)	
T ₄	65 (35.72)	71 (38.80)	
pN classification			.629
N ₀	122 (66.67)	129 (70.10)	
N ₁	25 (13.66)	28 (15.22)	
N ₂	34 (18.58)	25 (13.59)	
N ₃	2 (1.09)	2 (1.09)	
Pathological grading			.258
Well	34 (24.11)	25 (16.89)	
Moderate	39 (27.66)	40 (27.03)	
Poor	68 (48.23)	83 (56.08)	
Pathological type			.719
Squamous cell carcinoma	154 (84.62)	159 (85.95)	
Adenocarcinoma	28 (15.38)	26 (14.05)	

(Continues)

TABLE 3 (Continued)

Variables	NSND (185) No. of patients (%)	SD (185) No. of patients (%)N (%)	P
Family history of cancer			1.000
No	148 (80.00)	148 (80.00)	
Yes	37 (20.00)	37 (20.00)	
Adjuvant therapy			.623
No	99 (56.25)	90 (50.28)	
RT	23 (13.07)	28 (15.64)	
CT	18 (10.23)	24 (13.41)	
CRT	36 (20.45)	37 (20.67)	
Hypertension			1.000
No	136 (73.51)	136 (73.51)	
Yes	49 (26.49)	49 (26.49)	
Diabetes			.459
No	171 (92.43)	167 (90.27)	
Yes	14 (7.57)	18 (9.73)	

Abbreviations: BMI, body mass index; CT, chemotherapy; CRT, chemoradiotherapy; NSND, nonsmoking and nondrinking; RT, radiotherapy; SD, smoking and/or drinking.

TABLE 4 Survival analysis for SD vs NSND patients with oral cancer

	Number of censored (%)	Number of death (%)	Log rank P	Crude model	Age- and sex-adjusted model ^a	Fully adjusted model ^b
Before propensity score matching						
All-cause mortality			0.002			
NSND	519 (57.80)	127 (47.57)		1.000	1.000	1.000
SD	379 (42.20)	140 (52.43)		1.474 (1.159, 1.875)	1.722 (1.206, 2.459)	1.678 (1.086, 2.594)
Oral cancer specific mortality			0.010			
NSND	522 (57.17)	124 (49.21)		1.000	1.000	1.000
SD	391 (42.83)	128 (50.79)		1.380 (1.078, 1.768)	1.623 (1.127, 2.337)	1.632 (1.044, 2.552)
After propensity score matching						
All-cause mortality			0.003			
NSND	155 (54.58)	30 (34.88)		1.000	1.000	1.000
SD	129 (45.42)	56 (65.12)		1.933 (1.240, 3.013)	2.010 (1.288, 3.137)	1.897 (1.138, 3.165)
Oral cancer specific mortality			0.013			
NSND	156 (53.61)	29 (36.71)		1.000	1.000	1.000
SD	135 (46.39)	50 (63.29)		1.778 (1.125, 2.811)	1.844 (1.165, 2.918)	1.764 (1.043, 2.983)

Abbreviation: NSND: nonsmoking and nondrinking; SD: smoking and/or drinking.

^aAdjusted for age, gender.

^bAdjusted for age, gender, occupation, education level, residence, BMI, TNM stage, pathological type, pathological grading and adjuvant therapy.

that may limit generalization to other populations to some extent. Therefore, future multicenter research will be necessary.

In conclusion, our study showed that nearly a half of patients with oral cancer are NSND patients in China. The overall survival and disease specific survival of

NSND patients are better than SD patients after PS matching for critical social and clinical characteristics. In other words, smoking and drinking provide any additional prognostic information beyond the traditional prognostic factors in patients with oral cancer.

ACKNOWLEDGMENTS

This work was supported by grants from Fujian Natural Science Foundation Program (No. 2019J01314); Program for New Century Excellent Talents in Fujian Province University (No. 2018B029); The Scientific Research Talents Training Project of Health and Family Planning Health Commission in Fujian Province (No. 2017-ZQN-57, No. 2018-1-71), Joint Funds for the Innovation of Science and Technology of Fujian province (No. 2017Y9103), the High-level Talents research Start-up Project of Fujian Medical University (No. XRCZX2018001), Startup Fund for Scientific Research of Fujian Medical University (No. 2017XQ1011).

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

AUTHOR CONTRIBUTIONS

B.H. constructed the study design. X.B. and F.L. contributed to data interpretation, and manuscript drafting. J.W. and F.C. contributed to statistical analysis. L.C., C.Q., J.L., and L.P. participated in the clinical investigation, contributed to the epidemiological data collection. Y.Q., L.L., and B.S. revised the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All procedures performed in this study involving human participants were in accordance with the ethical standards of the Institutional Review Board (IRB) of Fujian Medical University (2011053). Written informed consent was obtained from all individual participants included in the study.

ORCID

Baochang He  <https://orcid.org/0000-0002-5357-2171>

REFERENCES

- van Dijk BA, Brands MT, Geurts SM, Merkx MA, Roodenburg JL. Trends in oral cavity cancer incidence, mortality, survival and treatment in The Netherlands. *Int J Cancer*. 2016;139:574-583.
- Schwam ZG, Judson BL. Improved prognosis for patients with oral cavity squamous cell carcinoma: analysis of the National Cancer Database 1998-2006. *Oral Oncol*. 2016;52:45-51.
- Chen F, Lin L, Yan L, Qiu Y, Cai L, He B. Preoperative neutrophil-to-lymphocyte ratio predicts the prognosis of Oral squamous cell carcinoma: a large-sample prospective study. *J Oral Maxillofac Surg*. 2017;75:1275-1282.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
- Blot WJ, McLaughlin JK, Winn DM, et al. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res*. 1988; 48:3282-3287.
- Chen F, He BC, Yan LJ, et al. Tea consumption and its interactions with tobacco smoking and alcohol drinking on oral cancer in Southeast China. *Eur J Clin Nutr*. 2017;71:481-485.
- Castellsague X, Quintana MJ, Martinez MC, et al. The role of type of tobacco and type of alcoholic beverage in oral carcinogenesis. *Int J Cancer*. 2004;108:741-749.
- Andre K, Schraub S, Mercier M, Bontemps P. Role of alcohol and tobacco in the aetiology of head and neck cancer: a case-control study in the Doubs region of France. *Eur J Cancer B Oral Oncol*. 1995;31B:301-309.
- Hashibe M, Brennan P, Chuang SC, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the international head and neck cancer epidemiology consortium. *Cancer Epidemiol Biomarkers Prev*. 2009;18:541-550.
- Beynon RA, Lang S, Schimansky S, et al. Tobacco smoking and alcohol drinking at diagnosis of head and neck cancer and all-cause mortality: results from head and neck 5000, a prospective observational cohort of people with head and neck cancer. *Int J Cancer*. 2018;143:1114-1127.
- Abrahao R, Anantharaman D, Gaborieau V, et al. The influence of smoking, age and stage at diagnosis on the survival after larynx, hypopharynx and oral cavity cancers in Europe: the ARCAge study. *Int J Cancer*. 2018;143:32-44.
- Sharp L, McDevitt J, Carsin AE, Brown C, Comber H. Smoking at diagnosis is an independent prognostic factor for cancer-specific survival in head and neck cancer: findings from a large, population-based study. *Cancer Epidemiol Biomarkers Prev*. 2014;23:2579-2590.
- DeAngelis A, Breik O, Koo K, et al. Non-smoking, non-drinking elderly females, a 5year follow-up of a clinically distinct cohort of oral squamous cell carcinoma patients. *Oral Oncol*. 2018;86:113-120.
- Koo K, Barrowman R, McCullough M, Iseli T, Wiesenfeld D. Non-smoking non-drinking elderly females: a clinically distinct subgroup of oral squamous cell carcinoma patients. *Int J Oral Maxillofac Surg*. 2013;42:929-933.
- Li R, Faden DL, Fakhry C, et al. Clinical, genomic, and metagenomic characterization of oral tongue squamous cell carcinoma in patients who do not smoke. *Head Neck*. 2015;37:1642-1649.
- Pickering CR, Zhang J, Neskey DM, et al. Squamous cell carcinoma of the oral tongue in young non-smokers is genetically similar to tumors in older smokers. *Clin Cancer Res*. 2014;20: 3842-3848.
- Chen F, Lin L, Liu F, et al. Three prognostic indexes as predictors of response to adjuvant chemoradiotherapy in patients with oral squamous cell carcinoma after radical surgery: a large-scale prospective study. *Head Neck*. 2019;41:301-308.
- Liu F, Chen F, Huang J, et al. Prospective study on factors affecting the prognosis of oral cancer in a Chinese population. *Oncotarget*. 2017;8:4352-4359.

19. Harris SL, Kimple RJ, Hayes DN, Couch ME, Rosenman JG. Never-smokers, never-drinkers: unique clinical subgroup of young patients with head and neck squamous cell cancers. *Head Neck*. 2010;32:499-503.
20. La Vecchia C, Tavani A, Franceschi S, Levi F, Corrao G, Negri E. Epidemiology and prevention of oral cancer. *Oral Oncol*. 1997;33:302-312.
21. Kruse AL, Bredell M, Gratz KW. Oral squamous cell carcinoma in non-smoking and non-drinking patients. *Head Neck Oncol*. 2010;2:24.
22. Righini CA, Timi N, Junet P, Bertolo A, Reyt E, Atallah I. Assessment of nutritional status at the time of diagnosis in patients treated for head and neck cancer. *Eur Ann Otorhinolaryngol Head Neck Dis*. 2013;130:8-14.
23. Kono T, Sakamoto K, Shinden S, Ogawa K. Pre-therapeutic nutritional assessment for predicting severe adverse events in patients with head and neck cancer treated by radiotherapy. *Clin Nutr*. 2017;36:1681-1685.
24. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41-55.
25. Chen H, Laba JM, Boldt RG, et al. Stereotactic ablative radiation therapy versus surgery in early lung cancer: a meta-analysis of propensity score studies. *Int J Radiat Oncol Biol Phys*. 2018;101:186-194.
26. Farquhar DR, Tanner AM, Masood MM, et al. Oral tongue carcinoma among young patients: an analysis of risk factors and survival. *Oral Oncol*. 2018;84:7-11.
27. Miro O, Gil V, Martin-Sanchez FJ, et al. Morphine use in the ED and outcomes of patients with acute heart failure: a propensity score-matching analysis based on the EAHFE registry. *Chest*. 2017;152:821-832.
28. Warschkow R, Guller U, Koberle D, et al. Perioperative blood transfusions do not impact overall and disease-free survival after curative rectal cancer resection: a propensity score analysis. *Ann Surg*. 2014;259:131-138.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Bao X, Liu F, Chen Q, et al. Propensity score analysis exploring the impact of smoking and drinking on the prognosis of patients with oral cancer. *Head & Neck*. 2020; 1–11. <https://doi.org/10.1002/hed.26099>

Supplement table1 Adjusted hazard ratios, and 95% confidence intervals for tobacco smoking and alcohol drinking and oral cancer in full cohort

	All-cause mortality	Oral cancer specific mortality
	HR (95%CI)	HR (95%CI)
Tobacco smoking duration		
Never	1.000	1.000
0-45 years	1.427(0.885,2.302)	1.382(0.845,2.261)
>45 years	2.280(1.263,4.116)	2.358(1.288,4.317)
Tobacco smoking frequency		
Never	1.000	1.000
0-10cigarettes/day	0.909(0.343,2.413)	0.977(0.366,2.603)
>10cigarettes/day	1.654(1.037,2.636)	1.617(1.002,2.612)
Alcohol drinking duration		
Never	1.000	1.000
0-35 years	0.956(0.630,1.450)	0.976(0.635,1.499)
>35 years	1.136(0.732,1.764)	1.145(0.726,1.804)
Alcohol drinking frequency		
Never	1.000	1.000
0-7 times/week	0.937(0.611,1.438)	0.848(0.518,1.390)
>7 times/week	1.400(1.021,1.979)	1.231(0.822,1.844)

Abbreviation: All adjusted for age, gender, occupation, education level, residence, BMI, TNM stage, pathological grading and pathological type, pathological grading and adjuvant therapy



Single-centre comparison of robotic and open pancreateoduodenectomy: a propensity score-matched study

Lapo Bencini¹ · Federica Tofani¹ · Claudia Paolini¹ · Carla Vaccaro¹ · Paolo Checcacci¹ · Mario Annecchiarico¹ · Luca Moraldi¹ · Marco Farsi² · Simone Polvani³ · Andrea Coratti¹

Received: 18 September 2019 / Accepted: 24 December 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Background Panreatoduodenectomy for pancreatic head and periampullary cancers is still associated with high perioperative morbidity and mortality. The aim of this study was to compare the short-term outcomes of robot-assisted pancreateoduodenectomy (RAPD) and open pancreateoduodenectomy (OPD) performed in a high-volume centre.

Methods A single-centre, prospective database was used to retrospectively compare the early outcomes of RAPD procedures to standard OPD procedures completed between January 2014 and December 2018. Of the 121 included patients, 78 underwent RAPD and 43 underwent OPD. After propensity score matching (PSM), 35 RAPD patients were matched with 35 OPD patients with similar preoperative characteristics.

Results There were no statistically significant differences in most of the baseline demographics and perioperative outcomes in the two groups after PSM optimization with the exception of the operative time (530 min (RAPD) versus 335 min (OPD) post-match, $p < 0.000$). No differences were found between the two groups in terms of complications (including pancreatic leaks, 11.4% in both OPD and RAPD), perioperative mortality, reoperations or readmissions. Earlier refeeding was obtained in the RAPD group vs. the OPD group (3 vs. 4 days, $p = 0.002$). Although the differences in the length of the hospital stay and blood transfusions were not statistically significant, both parameters showed a positive trend in favour of RAPD. The number of harvested lymph nodes was similar and oncologically adequate.

Conclusions RAPD is a safe and oncologically adequate technique to treat malignancies arising from the pancreatic head and periampullary region. Several perioperative parameters resulted in trends favouring RAPD over OPD, at the price of longer operating time. Data should be reinforced with a larger sample to guarantee statistical significance.

Keywords Robotic pancreateoduodenectomy · Pancreatic surgery · Robotic surgery

Pancreatic cancer (PC) is the fourth leading cause of death from cancer for females and the fifth leading cause for males in Western countries [1]. Its peak incidence occurs in the seventh and eighth decades of life, without significant sex differences [2]. Unfortunately, the incidence and mortality of PC in the United States have remained approximately stable over the past two decades [3], thus reflecting poor

medical improvements. In a recent worldwide review of incidence and mortality in 185 countries, there were more than 400,000 annual newly diagnosed pancreatic cancer patients, and equal number was expected to die from the disease [4].

Despite the advances in modern chemoradiotherapy, the best and only chance of cure for patients with PC is an oncologically adequate surgical resection aimed at completely removing gross and microscopic disease (R0). An early curative intent surgery is the best predictor of outcome [5]. An integrated, multidisciplinary plan is mandatory to combine the benefits of surgery, chemotherapy and radiotherapy to obtain the best results [6, 7].

Unfortunately, less than 20% of patients suffering from PC are eligible for curative surgery, which has a 5-year survival rate of approximately 20% [7, 8]. Moreover, pancreatic surgery is still associated with a perioperative morbidity of

✉ Lapo Bencini
lapbenc@tin.it

¹ Surgical Oncology and Robotics, Careggi University Hospital, Florence, Italy

² General Surgery, Le Scotte University Hospital, Siena, Italy

³ Gastroenterology Research Unit, University of Florence, Florence, Italy

45% and a mortality rate of less than 5% [9, 10]. Despite the high morbidity and mortality associated with the challenging pancreateoduodenectomy (PD), the centralization of major pancreatic resections to high-volume centres is expected to reduce perioperative complications and to improve survival [11]. However, PD for pancreatic head and periampullary cancers is one of the most challenging surgical procedures and requires the highest level of surgical expertise.

Since the first PD performed by Codivilla in 1898, many technical modifications have been undertaken in the procedure with the aim of improving patient outcomes [12]; however, almost none of these have demonstrated drastic differences over the others. For example, no difference in oncological outcomes, morbidity, mortality or survival has been noted between pylorus-preserving PD and the classic Whipple procedure or between pancreateojejunostomy and pancreaticogastrostomy [13, 14].

Interestingly, pancreatic surgery belonged exclusively to the field of open surgery until 1994, when the first laparoscopic PD (LPD) was performed by Gagner and Pomp [15]. Despite the theoretical benefits of a faster recovery and the potential oncological advantages related to a shorter interval to the receipt of adjuvant chemotherapy, only a few LPDs have been reported in the literature since then [16, 17]. Three randomized trials compared minimally invasive pancreateoduodenectomy with the open procedure. The first trial reported a shorter length of hospital stay after LPD in high-volume centres (>40 LPD/OPD per year, >150 LPDs performed), although it was underpowered to compare major morbidity [18]. A Spanish trial reported similar results [19]. In contrast, the third trial from the Netherlands (LEOPARD-2) was interrupted due to inferior results in the laparoscopic arm [20].

The intrinsic limitations of laparoscopy, such as non-articulated instruments, a lack of depth perception and its use in confined spaces, are the main obstacles to safely approaching the challenge of pancreatic surgery. Conversely, in the last decade, robotic surgery has been introduced to improve the feasibility of minimally invasive challenging procedures, such as pancreatic surgery, with encouraging results.

The first robotic-assisted PD (RAPD) was published in 2003 by Julianotti and colleagues from a large community hospital [21]. Although the indication for robotic surgery in pancreatic disease is still controversial due to the lack of large oncological datasets, the preliminary experiences are encouraging [22].

Robotic surgery is expected to minimize the trauma created in the exposure and handling of tissues and may offer the opportunity to combine the advantages of both minimally invasive and open surgery. The main advantages of the robotic system are an optimal and ergonomic surgeon position, deeper and more stable high-definition 3D vision,

endo-wrist arm technology (articulation of the instruments with 7 degrees of freedom), motion scaling and tremor filtration. Despite these theoretical advantages, RAPD remains a very challenging operation; its use is reserved to specialized centres with great experience in both pancreatic surgery and robotics.

The aim of this study was to compare the short-term outcomes RAPD and OPD performed in a high-volume centre.

Materials and methods

Patients and setting

In 2014, a structured robotics programme was started at Careggi Main Regional and University Hospital, Florence, Italy. A prospective database of all the robotic procedures was created to evaluate the results and the oncologic outcomes. According to the study's purposes, all consecutive patients who underwent elective RAPDs between 1 January 2014 and 31 December 2018 were collected and retrospectively compared with the open pancreateoduodenectomies (OPDs) performed during the same period.

The indication for surgery, for both open and robotic procedures, included malignant and borderline pancreatic nodules after a discussion at multidisciplinary oncology rounds (MOR). Briefly, the routine preoperative diagnosis of pancreatic disease was performed by conventional ultrasound, multidetector computed tomography (MDCT), magnetic resonance imaging (MRI), endoscopic ultrasonography, and Ca19-9 measurement. Fine-needle biopsy or cytology was obtained selectively.

The same experienced surgeon performed (or supervised using the second console) all the robotic procedures using both the da Vinci SI® and XI® Surgical System (Intuitive Surgical, Sunnyvale, CA, USA), while the open procedures were performed by the same surgeon and other experienced surgeons (more than 50 PD completed before 2014). The decision to perform a robotic operation or a laparotomy was balanced according to the preoperative imaging and patient characteristics. The only rigid exclusion criterion for RAPD was tumours with suspected vascular involvement at the preoperative imaging. Procedure-specific informed consent was obtained from all patients, including detailed consent to the use of the robotic assistance when needed.

Preoperative variables included baseline characteristics, such as age, sex, body mass index (BMI), comorbidities (Charlson comorbidity index) [23], past abdominal surgical history, MDCT/MRI-scan information (vascular/organ involvement), American Society of Anaesthesiologists (ASA) classification [24], and Eastern Cooperative Oncology Group (ECOG) performance status. Moreover, we divided patients into subgroups according to the tumour's

location (head/uncinate, common biliary duct—CBD, duodenum or others).

Operative data were extracted by the operating theatre software, while the perioperative outcomes were collected from the clinical digital files. The International Study Group on Pancreatic Fistula definition (ISGPS) was used to classify Postoperative Pancreatic Fistula (POPF) [25]. The ISGPS definition was also used to classify post-pancreatectomy haemorrhage (PPH) [26]. Additionally, the incidence of delayed emptying syndrome (DGE) was collected according to the ISGPS definition [27]. The general complications (minor or major) were recorded based on the Clavien–Dindo Classification [28].

Patients were followed until discharge or 30 days postoperatively (whichever occurred later) during outpatient visits or by phone calls in selected cases. The primary objective was to assess the postoperative outcomes (limited to in-hospital or 30-day events). The final pathological report was obtained in all cases. The Union for International Cancer Control (UICC) TNM classification was used for pathological staging of the tumours [29].

Surgical technique

The OPD were performed according to a standardized Whipple-Kausch technique or by a pylorus-preserving method. Interestingly, the same steps were also followed during RAPD, with very few technical differences. In the RAPD, both the da Vinci SI® and XI® Surgical System (Intuitive Surgical, Sunnyvale, CA, USA) procedures were used, depending on the availability.

Briefly, the operating room was arranged with the patient in the supine, mild reverse Trendelenburg position. The robotic cart and trocar ports were placed as shown in Fig. 1.

The procedure was a fully robotic technique. In the presence of peritoneal adhesions, an operative laparoscopy was performed before docking the robotic system to allow adequate robotic port placement or to verify the absence of metastases or peritoneal carcinomatosis.

The robotic procedure started with the mobilization of the right colon; the lesser sac was opened and explored to exclude any tumour infiltration. The surgical steps reflected the standardized counter clockwise technique, including Kocher's manoeuvre, superior mesenteric vein (SMV) dissection, duodenum or stomach transection, cholecystectomy and common bile duct section, Treitz's section, and pancreatic division with dissection of the uncinate process, with the fourth robotic arm facilitating the retraction.

After the confirmation of negative resection margins, the reconstructive phase of the operation started, according to the classic clockwise approach (pancreatic anastomosis, biliary reconstruction and gastrointestinal anastomosis). The pancreatic texture was intraoperatively defined by the

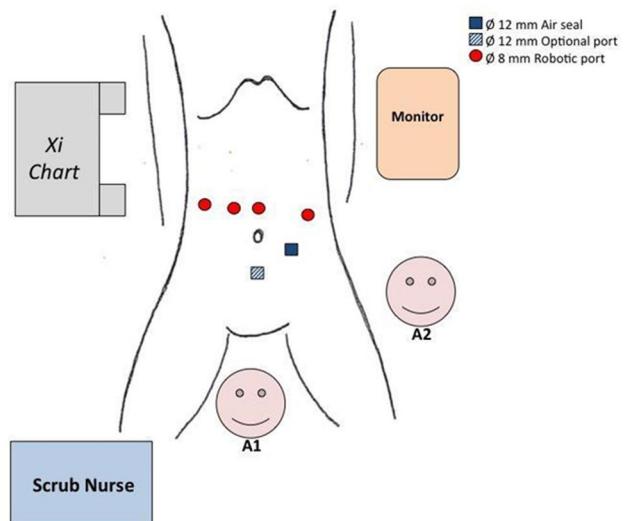


Fig. 1 Cart and trocar positioning and theatre arrangement during RAPD

operating surgeon as either firm or soft. Because of the impossibility to palpate the gland during RAPD (despite the partial replacement of the tactile feedback by visual feedback), pancreatic texture was assessed or confirmed on the specimen. The pancreatic duct diameter was acquired intraoperatively by probing the ductal orifice with serial-sized dilators or estimated by preoperative radiological investigations. One or two drainage tubes were also placed in the upper abdominal quadrants at the end of the procedure.

Statistics

Statistical and descriptive analyses were performed using the Statistical Package for Social Science (SPSS) version 20 (SPSS Inc., Chicago, Illinois, USA).

All data were collected prospectively and reviewed retrospectively. Propensity score matching (PSM) was performed with the R package MatchIt procedure with the nearest-neighbour 1-to-1 method [30] to minimize the effects of preoperative selection biases between the OPD and RAPD groups.

Covariates were important parameters for the overall score of patients prior to the surgery (e.g. sex, age, BMI, smoking and alcohol consumption, CHA, ASA, ECOG). Cases with missing data in the matching variables were omitted from the analysis. After matching, 35 patients who received RAPD and 35 matched patients who received OPD were used for the statistical analysis.

Group and subgroup comparisons based on continuous data were performed using the non-parametric Mann–Whitney *U* test, while discrete variables were compared using

the Chi-square or Fisher's exact test. The statistical level of significance was defined as a p value <0.05 .

Results

All 121 consecutive patients who underwent PD in our surgical department from January 2014 to December 2018 were prospectively collected. The entire cohort included 77 patients who underwent a planned OPD, 44 RAPD and 6 who started as RAPD but needed an early conversion to OPD (13.6%) due to intense adhesions. We decided to merge the converted RAPD with the group of OPD according to the study's purposes. The final analysis involved 83 OPDs to match 38 RAPDs.

An attempt to limit the obvious selection biases among robotic and open procedures was planned through a propensity score analysis to balance for possible preoperative confounders and to achieve balanced exposure groups at baseline. Overall, 13 preoperative variables were included in the model, reducing the initial cohort. Thirty-five of 38 RAPD were matched (1:1) to 35 of the 83 OPDs. Each variable was weighted as 1 in the final calculation. After matching, all variables were more homogeneous (Fig. 2).

The demographic and baseline characteristics, including age and BMI, were well matched among groups in the pre- and post-match comparisons. Comorbidities and performance indexes, including the Charlson, ECOG and ASA scores, were also similar. Other specific oncological parameters, such as CA19-9 level, jaundice, weight loss before

surgery and the need for neoadjuvant chemotherapies, were not significantly different. Details of the matching for demographic and baseline characteristics are given in Table 1.

When considering the tumour characteristics, the only parameter that was significantly different in the pre-match and post-match comparison was the anatomic distribution of the primary lesion, which was mostly in the pancreatic head for the OPD cases (80 and 51.4%, $p=0.006$, respectively). Tumour size, the number of harvested nodes and tumour staging were similar before and after matching. Nevertheless, the histopathologic final definition was significantly different in both the pre- and post-match comparisons (65.7 and 45.7%, $p=0.04$, respectively) (Table 2).

When considering the perioperative results, the robotic group was reported to have a significantly longer operative time than the open group (median 560 min versus 367 min, $p < 0.000$) in the pre-match comparison, while the difference continued to be significant even after PSM (median 530 min versus 335 min, $p < 0.000$). Moreover, additional organ resections and vascular debridement were performed exclusively in the open group. Anastomotic (pancreatic) technique differences were also incomparable between the two groups because robotic operations included only duct-to-mucosa reconstruction, while in the open group, a larger spectrum of anastomosis was evaluated according to the surgeons' experiences. The consistency of the pancreatic gland was described as similar in the two groups pre- and post-matching (Table 2).

Interestingly, the occurrence of overall intraoperative complications was a quite rare event in the open group

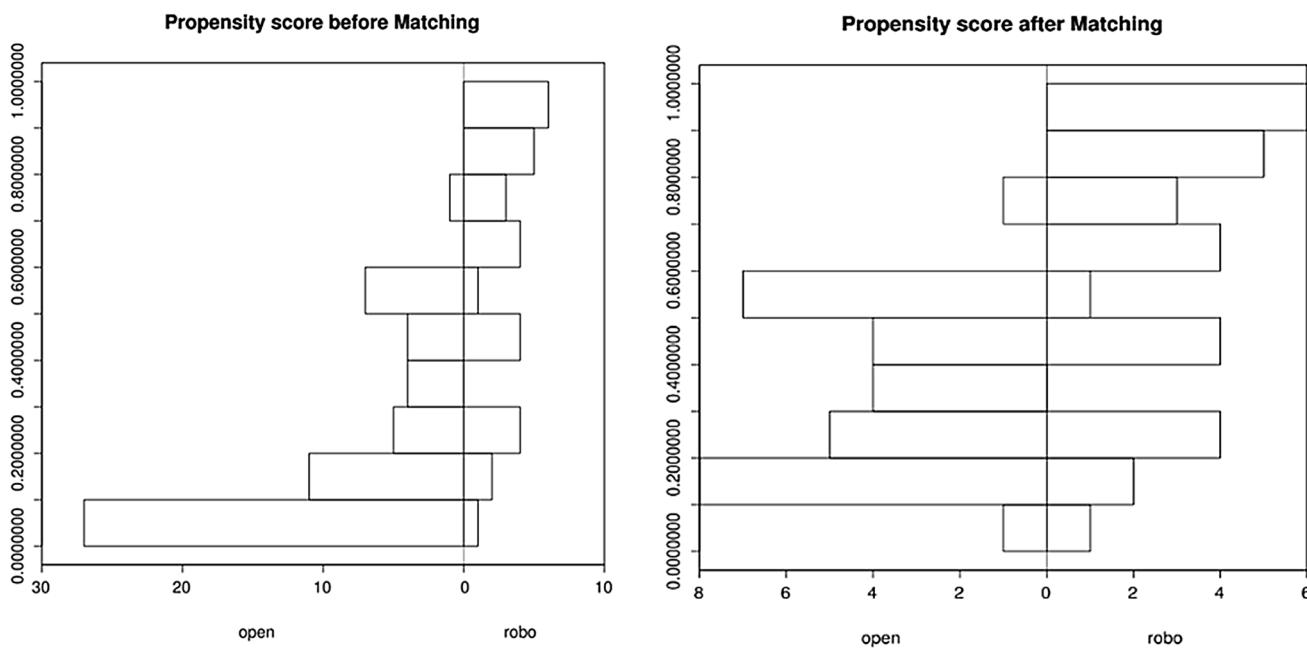


Fig. 2 Distribution of propensity score (pre- and post-matching)

Table 1 Baseline and demographic characteristics of the pre-match and post-match cohort of patients who underwent PD

Baseline characteristics	OPD pre N=83	RAPD pre N=38	p value	OPD post N=35	RAPD post N=35	p value
Age (years)	74 (56–91)	60 (42–73)	0.06	69 (50–88)	70.5 (42–85)	0.95
Sex, N (%)						
Female	39 (47)	16 (42.1)	0.69	19 (56.3)	16 (45.7)	0.51
Male	44 (53)	22 (57.9)		16 (45.7)	19 (56.3)	
BMI (kg/m^2)	24 (14–38)	26 (18–32)	0.07	24 (18–38)	26 (18–32)	0.19
CHARLSON score, N (%)						
2	33 (40.2)	17 (44.7)	0.92	18 (51.5)	16 (45.8)	0.86
3	13 (15.9)	8 (21.1)		7 (20)	7 (20)	
4	12 (14.6)	6 (15.8)		4 (11.4)	5 (14.3)	
5	15 (18.3)	6 (15.8)		6 (17.1)	6 (17.1)	
6	6 (7.3)	1 (2.6)		0 (0)	1 (2.8)	
7	2 (2.4)	0 (0)		0 (0)	0 (0)	
9	1 (1.2)	0 (0)				
ASA score, N (%)						
1	6 (7.2)	3 (7.9)	0.09	3 (8.6)	3 (8.6)	0.05
2	47 (56.6)	29 (76.3)		24 (68.6)	27 (77.1)	
3	28 (33.7)	5 (13.2)		8 (22.8)	5 (14.3)	
4	2 (2.4)	1 (2.6)		0 (0)	0 (0)	
ECOG performance score, N (%)						
0	69 (83.1)	37 (97.4)	0.19	34 (97.1)	34 (97.1)	1.000
1	8 (9.6)	0 (0)		0 (0)	0 (0)	
2	1 (1.2)	0 (0)		0 (0)	0 (0)	
3	4 (4.8)	1 (2.6)		1 (2.9)	1 (2.9)	
4	1 (1.2)	0 (0)		0 (0)	0 (0)	
MOR, N (%)	69 (83.1)	32 (84.2)	1.000	30 (85.7)	30 (85.7)	1.000
Pre-op diabetes, N (%)	22 (26.5)	7 (18.4)	0.36	6 (17.1)	7 (20%)	0.75
Previous abdominal surgery, N (%)	47 (56.6)	17 (44.7)	0.24	21 (60)	16 (45.7)	0.23
Pre-op stenting, N (%)	47 (56.6)	17 (44.7)	0.24	17 (48.6)	17 (48.6)	1.000
Pre-op jaundice, N (%)	55 (66.3)	24 (63.1)	0.83	21 (60%)	22 (62.8)	0.80
Weight loss (> 5 kg/6 m), N (%)	44 (53)	14 (36.8)	0.11	16 (45.7)	12 (34.3)	0.32
CA19.9 (U/mL)	102 (2–6642)	85 (2–1617)	0.3	70 (2–2617)	143 (2–1617)	0.80
Pre-op biopsy, N (%)	29 (34.9)	16 (42.1)	0.54	15 (42.8)	15 (42.8)	1.000
Neoadjuvant treatment, N (%)	7 (8.4)	1 (2)	0.43	0 (0)	0 (0)	1.000

OPD open pancreateoduodenectomy, RAPD robot-assisted pancreateoduodenectomy, BMI body mass index, ASA American Society of Anaesthesiologists, ECOG Eastern Cooperative Oncology Group, MOR multidisciplinary oncology rounds, DM diabetes mellitus

without a statistical significance in both the pre- and post-match comparisons, with a consistent trend to better results in the robotic group (8.4% vs. 0%, $p=0.09$ and 5.7% vs. 0%, $p=0.15$, respectively). Intraoperative results are summarized in Table 3. The overall morbidity and mortality rates were acceptable and highly comparable between open and robotic procedures for both the pre-match comparison and the post-match adjustment (Table 4).

The occurrence of POPF was 14.4% versus 13.1% between OPD and RAPD, respectively, in the pre-match comparison ($p=0.84$) and 11.4% vs. 11.4% in the post-match comparison ($p=1.00$); the POPF was equally distributed in

grades B and C according to the brand-new classification of the International Study Group on Pancreatic Fistula definition (ISGPS). Other pancreatic-specific complications, such as post-pancreatectomy haemorrhage (PPH), the reoperation rate, readmissions and the need for blood transfusions, were also similar (Table 4).

The bowel canalization, including both gases and faeces, occurred with a similar median time in the early postoperative period. Nevertheless, the median day of resumption of solid foods (creams, yogurt or chopped meat) was earlier for the robotic group in the pre- and post-match test (3 vs. 4 days, $p=0.002$). Additionally, the overall length of

Table 2 Tumour characteristics of the pre-match and post-match cohort of patients who underwent PD

Tumour characteristics	OPD pre N=83	RAPD pre N=28	p value	OPD post N=35	RAPD post N=35	p value
Localization, N (%)						
Pancreas	58 (69.9)	20 (52.7)	0.000*	28 (80)	18 (51.4)	0.006*
Duodenum	9 (10.8)	17 (44.7)		4 (11.4)	16 (45.7)	
CBD	11 (13.3)	1 (2.6)		3 (8.7)	1 (2.8)	
Others	5 (6)	0 (0)		0 (0)	0 (0)	
Histology, N (%)						
Ductal adenocarcinoma	46 (55.4)	18 (47.4)	0.01*	23 (65.7)	16 (45.7)	0.04*
Duodenum adenocarcinoma	6 (7.2)	13 (34.2)		3 (8.6)	13 (37.1)	
CCK	11 (13.3)	1 (2.6)		3 (8.6)	1 (2.8)	
NET	6 (7.2)	2 (5.3)		4 (11.4)	2 (5.7)	
IPMN/MCN	5 (6)	2 (5.3)		2 (5.3)	1 (3.3%)	
Pancreatitis	1 (1.2)	—		—	—	
Other	8 (9.6)	2 (5.3)		—	2 (5.7)	
AJCC stage						
Not a cancer (confirmed)	14 (16.9)	6 (15.8)				
0	1 (1.2)	0 (0%)	0.33	3 (8.6)	5 (14.3)	0.27
Ia	7 (8.4)	2 (5.3)		6 (17.1)	2 (5.7)	
Ib	4 (4.8)	7 (18.4)		2 (5.7)	7 (20)	
IIa	18 (21.7)	10 (26.3)		6 (17.1)	8 (22.9)	
IIb	31 (37.3)	11 (28.9)		15 (42.9)	11 (41.4)	
III	6 (7.2)	2 (5.3)		3 (8.6)	2 (5.7)	
IV	2 (2.4)	0 (0%)		0 (0)	0 (0)	
Tumour size (mm)	35 (2–51)	30 (18–40)	0.2	37 (2–51)	30 (18–40)	0.92
Wirsung duct diameter (mm)	3 (2–10)	3 (2–12)	0.7	3.5 (2–7)	3 (2–12)	0.79
Lymph nodes harvested (N)	23 (2–67)	22 (7–60)	0.88	23 (3–62)	22 (7–60)	0.85
Lymph node ratio (N+/N)	0 (0–0.7)	0 (0–0.2)	0.16	0.3 (0–0.4)	0 (0–0.2)	0.18

Bold and asterisks values with statistic significance ($p < 0.05$)

OPD open pancreateoduodenectomy, RAPD robot-assisted pancreateoduodenectomy, AJCC American Joint Committee on Cancer

hospital stay was similar for robotic operations and open procedures (8 vs. 10, $p=0.09$ in the pre-match and 8 vs. 10, $p=0.13$ in the post-match).

Discussion

The role of robotics in pancreatic surgery has been discussed in several papers [12, 31, 32], although its wide application is far from reality. Nevertheless, a minimally invasive approach to pancreatic malignancies could have a crucial impact in the oncologic perspective, reducing the interval between surgery and adjuvant therapies. Moreover, improved results in hand-sewn anastomoses, lymphadenectomy and the management of major bleeding are facilitated by the magnified 3D intraoperative view and the articulate endoscopic instruments. Furthermore, the availability of a second console is crucial in enhancing dedicated training in pancreatic surgery, including more and more challenging

steps, and facilitating both senior and younger surgeons. Unfortunately, large robotic series are still lacking, and the supposed benefits of the technique have never been demonstrated. Retrospective/prospective series with accurate data analyses are considered precursors to quality controlled large observational studies if the general principles of oncologic surgery are maintained.

Interestingly, the capabilities of the da Vinci Surgical System® can overcome the technical limitations of laparoscopy to reproduce complex open procedures pointing towards the failed spread of laparoscopic PD. Therefore, in our centre, laparoscopic PD has never been performed, and any effort is directed towards robotic applications. Nevertheless, major laparoscopic pancreatic resection continues to be adopted at a slower pace and only in a few highly selective centres. Indeed, whereas less than 500 laparoscopic PD have been reported in the literature in almost 20 years [16, 17], more than 350 robot-assisted PD have been published in the last few years [33, 34].

Table 3 Intraoperative results (pre-match and post-match cohort)

Intraoperative results	OPD pre N=83	RAPD pre N=38	p value	OPD post N=35	RAPD post N=35	p value
Operative time (min)	367 (270–520)	560 (465–670)	0.000*	335 (220–565)	530 (405–660)	0.000*
Type of resection, N (%)						
Pylorus-preserving PD	58 (69.8)	17 (44.8)	0.008*	24 (69)	16 (46)	0.053
Whipple PD	25 (30.2)	21 (55.2)		11 (31)	19 (54)	
Additional organs resected, N (%)	11 (13.3)	0 (0)	0.01*	2 (5.7)	0 (0)	0.15
Vascular resection, N (%)	20 (24.1)	0 (0)	0.001*	8 (22.8)	0 (0)	0.003*
Type of anastomosis, N (%)						
Pancreatojejunostomy	11 (13.3)	0 (0)	0.02*	7 (20)	0 (0)	0.007*
Duct-to-mucosa	61 (73.5)	36 (94.7)		23 (65.7)	33 (94.3)	
Pancreatogastrostomy	8 (9.6)	2 (5.3)		5 (14.3)	2 (5.7)	
None	3 (3.6)	0 (0)		0 (0)	0 (0)	
Consistency of the pancreas, N (%)						
Hard/firm pancreas	37 (44.6)	16 (42.1)	0.96	17 (48.6)	16 (45.7)	0.94
Soft pancreas	37 (44.6)	18 (47.4)		18 (51.4)	15 (42.8)	
Unknown	9 (10.8)	4 (10.5)		0 (0)	0 (0)	
Intraoperative complications, N (%)	7 (8.4)	0 (0)	0.09	2 (5.7)	0 (0)	0.15

Bold and asterisks values with statistic significance ($p < 0.05$)

OPD open pancreateoduodenectomy, RAPD robot-assisted pancreateoduodenectomy, POPF postoperative pancreatic fistula, PPH post-pancreatectomy haemorrhage, DGS delayed gastric syndrome

Nonrandomized studies and meta-analyses comparing minimally invasive techniques with open pancreatic resections show comparable complications (including the incidence of POPF), reoperations, mortality, and numbers of harvested lymph nodes as a surrogate of good oncologic outcome [35–37]. Wound infections, hospital stay length, blood loss, transfusion rate and R1 resections were significantly lower in patients who underwent minimally invasive resections [35].

Our series represents one of the few and large experiences with RAPD. Although the present study suffers from being inherently retrospective, highly selective and subject to early learning curves with a high cost burden, these early short-term results are within optimal parameters. Patients with a very advanced pancreatic neoplasm with proven or suspected vascular involvement or who received neoadjuvant chemotherapy for locally advanced stages were excluded a priori from a minimally invasive approach, thus representing an obvious bias. Conversely, periampullary small tumours in younger patients were considered an excellent target for RAPD. Another limitation of the present study was the inclusion of converted RAPD in the open group, according to the limits of the pre-determined aim. This choice was based on the fact that all the conversions were decided in the early step of the interventions due to severe adhesions or the presence of advanced disease. Of note, all converted patients were excluded by the PSM process.

The more interesting findings of the present study were that, in both the pre- and post-matched comparisons, early feeding and several perioperative minor outcomes tended to favour RAPD compared to OPD. After the propensity score calculation, the subgroups of patients remained substantially similar in the relevant aspects, including the incidence of POPF, morbidity, mortality and hospital stay length (Table 1). Interestingly, in our study, the entire postoperative course of each patient was followed by the same team (surgeons, anaesthesiologists, nurses, dieticians and physiotherapists) until the time of discharge, reducing any bias in postoperative management.

Overall, no significant differences were found between the two groups in terms of complications and perioperative mortality, with a similar percentage of POPF demonstrated in the two groups (11.4% after PSA adjustment), which was consistent with the current literature [34]; “biochemical leak” (defined in the previous nomenclature as grade A by ISGPs [25] was the most common event with minimal impact on the perioperative course. Although no significant difference between the two groups was observed in terms of B/C POPF, the RAPD group presented with a higher risk of fistulas due to the different types of histology and location of the neoplasms (more adenocarcinomas of the pancreatic head versus periampullary location, Table 2). According to the work of Callery and colleagues [38], the pancreatic texture in the malignancies of the pancreatic gland is more consistent than the periampullary lesions (due to the

Table 4 Outcomes and postoperative complications (pre-match and post-match cohort)

	Postoperative complications (30 days/in-hospital)	OPD pre N=83	RAPD pre N=38	p value	OPD post N=35	RAPD post N=35	p value
Overall morbidity, N (%)	43 (51.8)	20 (52.6)	1.000	15 (42.8)	17 (48.5)	0.63	
Mortality, N (%)	4 (4.8)	1 (2.6)	1.000	0 (0)	1 (2.9)	1.000	
Clavien–Dindo classification			0.30				0.61
0	24 (28.9)	17 (44.7)		14 (40)	17 (48.6)		
I	16 (16.9)	6 (15.8)		8 (22.8)	6 (17.1)		
II	21 (25.3)	4 (10.5)		7 (20)	4 (11.4)		
IIIa	7 (8.4)	2 (5.3)		2 (5.7)	2 (5.7)		
IIIb	7 (8.4)	5 (13.2)		3 (8.6)	3 (8.6)		
IVa	1 (1.2)	2 (5.3)		0 (0)	2 (5.7)		
IVb	5 (6)	1 (2.6)		1 (2.9)	0 (0)		
V	4 (4.8)	1 (2.6)		0 (0)	1 (2.9)		
POPF N (%)	12 (14.4)	5 (13.1)	0.84	4 (11.4)	4 (11.4)	1.00	
Grading of POPF, N (%)							
B	5 (6)	3 (7.9)	0.78	2 (5.7)	2 (5.7)	1.00	
C	7 (8.4)	2 (5.2)		2 (5.7)	2 (5.7)		
PPH, N (%)	8 (9.6)	4 (10.5)	1.000	4 (11.4)	3 (8.6)	0.69	
Grading of PPH, N (%)							
A	0 (0)	1 (2.6)	0.25	0 (0)	1 (2.8)	0.03*	
B	1 (2.5)	2 (5.2)		0 (0)	2 (6.7)		
C	7 (87.5)	1 (2.6)		4 (11.4)	0 (0)		
DGS, N (%)	12 (14.4)	5 (13.1)	0.84	5 (14.2)	4 (11.4)	0.72	
Blood transfusion, N (%)	32 (38.6)	7 (18.4)	0.03*	8 (22.8)	6 (17.1)	0.55	
Refeeding (days)	3.5 (2–8)	3 (2–7)	0.002*	4 (2–19)	3 (1–5)	0.002*	
Gas canalization (days)	3 (1–9)	4 (2–7)	0.5	3 (1–7)	3 (2–7)	0.36	
Faeces canalization (days)	4 (1–10)	5 (3–7)	0.46	4 (3–8)	5 (2–9)	0.29	
Reoperation, N (%)	14 (16.9)	8 (21.1)	0.61	4 (11.4)	6 (17.1)	0.49	
Readmission, N (%)	7 (8.4)	3 (7.9)	1.000	4 (11.4)	3 (8.6)	0.69	
Hospital stay (days)	10 (6–110)	8 (6–68)	0.09	10 (7–110)	8 (6–40)	0.13	

Bold and asterisks values with statistic significance ($p < 0.05$)

OPD open pancreateoduodenectomy, RAPD robot-assisted pancreateoduodenectomy, POPF postoperative pancreatic fistula, PPH post-pancreatectomy haemorrhage, DGS delayed gastric syndrome

fibrosis associated with ductal cancer). The equal percentage of POPFs in both groups could suggest a relatively better outcome with the robotics. Moreover, this result could be sustained by the reduced blood loss and the technical advantages offered by robotic surgery, which allows the anastomosis of a very small pancreatic duct.

A nearly constant result is a longer operative time in the robotic series, in line with the majority of published studies (335 versus 530 min for OPD and RAPD, respectively, after PSM, $p > 0.001$). This aspect, together with the increased costs of supplies and operative room employment, is recognized drawbacks of robotic technology, although these aspects are expected to be reduced as the surgical teams gain experience.

According to recent studies, patients undergoing PD for pancreatic adenocarcinoma require perioperative transfusions in 20% of cases [10]. In our series, the blood

transfusion rate was lower for RAPD (38.6 (OPD) and 18.4% (RAPD), $p=0.03$ in the unmatched cohort); while not statistically significant after PSM (22.8 (OPD) and 17.1% (RAPD), $p=0.55$), the trend favoured RAPD.

Refeeding in the RAPD group started significantly faster (one day) than in the OPD group before and after PSM, although the rate of DGE in both groups was similar (and low) (Table 4). Moreover, the length of the hospital stay was similar in the two groups, with a median value that was higher in the OPD group (not statistically significant).

The number of harvested lymph nodes was comparable and oncologically adequate according to the AJCC guidelines [29], with a median number for the post-match results of 23 for the OPD group and 23 for the RAPD group. The few patients with < 12 lymph nodes harvested were not affected by confirmed malignancies without the need for an extended lymphadenectomy.

A recent similar retrospective multicentre PSM cohort study comparing minimally invasive (including both robotic, laparoscopic and hybrid procedures) and open PD found no differences in the 30-day major morbidity, mortality and length of stay between the two groups. However, minimally invasive PD was associated with a doubled rate of POPF and longer operative times [32]. In addition to the approach itself, other factors could have influenced this outcome, such as an insufficient balance between the groups at baseline (residual confounding), underreporting in some registry data and the presence of many operating surgeons from different countries. Furthermore, the 2005 ISGPF definition of POPF can be interpreted in many ways, and the POPF registration after discharge may not be accurate in registry studies.

A large systematic review and meta-analysis suggested advantages of robotic pancreatectomy over the open approach in several aspects. First, a significantly lower overall complication rate and a reduced reoperation rate were observed. No significant differences were observed in POPF or mortality between robotic surgery and open surgery. However, the long-term survival rate and lymph node retrieval were not evaluated due to incomplete data [39].

Interestingly, a lymph node ratio (LNR) >0.4 has been associated with a risk of death comparable to that due to the presence of a metastatic disease [40]. In this case series, the LNR of patients who underwent OPD was 0.3, while the LNR of patients treated with RAPD was 0 in the absence of statistical significance ($p=0.18$). The excellent and comparable lymph node clearance with both techniques reflects the consistent standardization achieved with a dedicated oncological-oriented surgical team.

Our data confirmed that robotic pancreatectomy can be performed safely and according to current oncological standards (adequate resection margins and appropriate lymphadenectomy), with comparable results to those obtained for open pancreatectomy, except for the duration of surgery. Moreover, many perioperative parameters (refeeding, hospital stay length, blood transfusions) seemed to favour RAPD over OPD.

Conclusions

The advent of robotic technology might circumvent the technical limitations of laparoscopic pancreatic surgery. Many of the advocated advantages regarding the robotic platform are more evident when dealing with challenging pancreatoduodenectomy, while the distal pancreatectomies are easily approached by pure laparoscopy. However, a minimally invasive approach to pancreatic malignancies could have a crucial impact in the oncologic perspective, reducing the interval between surgery and adjuvant therapies. Furthermore, the availability of a second console is crucial in

enhancing dedicated training in pancreatic surgery, including more and more challenging steps, and facilitating both senior and younger surgeons.

Our data confirmed that robotic pancreatectomy can be performed safely and according to current oncological standards in selected patients, and several perioperative parameters (refeeding, hospital stay length, blood transfusions) seem to favour RAPD over OPD, although at the price of longer operating times.

Although large, randomized controlled trials are very difficult to design (if not impractical and/or impossible) due to the need for local expertise, wide caseloads and problems of informed consent, further studies are of crucial importance to assess the definitive superiority of one technique over another.

Compliance with ethical standards

Disclosures Drs Lapo Bencini, Federica Tofani, Claudia Paolini, Carla Vaccaro, Paolo Checcacci, Mario Annecchiarico, Luca Moraldi, Marco Farsi, Simone Polvani and Andrea Coratti have no conflicts of interest or financial ties to disclose.

References

- Ward E, Sherman RL, Henley SJ, Siegel D, Feuer E, Firth AU, Kohler BA, Scott S, Ma J, Anderson RN, Benard V, Cronin K (2019) Annual Report to the Nation on the Status of Cancer, 1999–2015, featuring cancer in men and women ages 20–49. *J Natl Cancer Inst.* <https://doi.org/10.1093/jnci/djz106>
- Rawla P, Sunkara T, Gaduputtic V (2019) Epidemiology of pancreatic cancer: global trends etiology and risk factors. *World J Oncol.* <https://doi.org/10.14740/wjon1166>
- Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruh J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds) (2019) SEER Cancer Statistics Review, 1975–2016, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2016/, based on November 2018 SEER data submission, posted to the SEER web site, April 2019
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* <https://doi.org/10.3322/caac.21492>
- Vreeland TJ, Katz MHG (2018) Timing of pancreatic resection and patient outcomes: is there a difference? *Surg Clin North Am.* <https://doi.org/10.1016/j.suc.2017.09.006>
- Adamska A, Domenichini A, Falsca M (2017) Pancreatic ductal adenocarcinoma: current and evolving therapies. *Int J Mol Sci.* <https://doi.org/10.3390/ijms18071338>
- Tempero MA, Malafa MP, Al-Hawary M, Asbun H, Bain A, Behrman SW, Benson AB, Binder E, Cardin DB, Cha C, Chiorean EG, Chung V, Czito B, Dillhoff M, Dotan E, Ferrone CR, Hardacre J, Hawkins WG, Herman J, Ko AH, Komanduri S, Koong A, LoConte N, Lowy AM, Moravek C, Nakakura EK, O'Reilly EM, Obando J, Reddy S, Scaife C, Thayer S, Weekes CD, Wolff RA, Wolpin BM, Burns J, Darlow S (2017) Pancreatic adenocarcinoma, Version 2.2017, NCCN Clinical Practice Guidelines

- in oncology. *J Natl Compr Cancer Netw.* <https://doi.org/10.6004/jnccn.2017.013>.
8. Cooperman AM, Bruckner H, Snady H, Hammerman H, Fader A, Feld M, Golier F, Rush T, Siegal J, Kasmin F, Cohen S, Wayne MG, Iskandar ME, Steele JG (2018) Cancer of the pancreas—actual 5, 10, and 20+ year survival: the lucky and fortunate few. *Surg Clin North Am.* <https://doi.org/10.1016/j.suc.2017.09.007>
 9. Cameron JL, He J (2015) Two thousand consecutive pancreaticoduodenectomies. *J Am Coll Surg.* <https://doi.org/10.1016/j.jamcollsurg.2014.12.031>
 10. Dosch AR, Grigorian A, Delaplain PT, Bartholomew TS, Won EJ, Gabriel V, Wolf RF, Juric Z, Imagawa DK (2019) Perioperative blood transfusion is associated with an increased risk for post-surgical infection following pancreaticoduodenectomy. *HPB (Oxford).* <https://doi.org/10.1016/j.hpb.2019.03.374>
 11. Van Putten M, Nelen SD, Lemmens VEPP, Stoot JHMB, Hartgrink HH, Gisbertz SS, Spillenaar Bilgen EJ, Heisterkamp J, Verhoeven RHA, Nieuwenhuijzen GAP (2018) Overall survival before and after centralization of gastric cancer surgery in the Netherlands. *Br J Surg.* <https://doi.org/10.1002/bjs.10931>
 12. Griffin JF, Poruk KE, Wolfgang CL (2015) Pancreatic cancer surgery: past, present, and future. *Chin J Cancer Res.* <https://doi.org/10.3978/j.issn.1000-9604.2015.06.07>
 13. Hüttner FJ, Fitzmaurice C, Schwarzer G, Seiler CM, Antes G, Büchler MW, Diener MK (2016) Pylorus-preserving pancreaticoduodenectomy (pp Whipple) versus pancreaticoduodenectomy (classic Whipple) for surgical treatment of periampullary and pancreatic carcinoma. *Cochrane Database Syst Rev.* <https://doi.org/10.1002/14651858.CD006053.pub6>
 14. Cheng Y, Briarava M, Lai M, Wang X, Tu B, Cheng N, Gong J, Yuan Y, Pilati P, Mocellin S (2017) Pancreaticojejunostomy versus pancreaticogastrostomy reconstruction for the prevention of postoperative pancreatic fistula following pancreaticoduodenectomy. *Cochrane Database Syst Rev.* <https://doi.org/10.1002/14651858.CD012257.pub2>
 15. Gagner M, Pomp A (1994) Laparoscopic pylorus-preserving pancreaticoduodenectomy. *Surg Endosc* 8(5):408–410
 16. Boggi U, Amorese G, Vistoli F, Caniglia F, De Lio N, Perrone V, Barbarello L, Belluomini M, Signori S, Mosca F (2015) Laparoscopic pancreaticoduodenectomy: a systematic literature review. *Surg Endosc.* <https://doi.org/10.1007/s00464-014-3670-z>
 17. Croome KP, Farnell MB, Que FG, Reid-Lombardo KM, Truty MJ, Nagorney DM, Kendrick ML (2014) Total laparoscopic pancreaticoduodenectomy for pancreatic ductal adenocarcinoma: oncologic advantages over open approaches? *Ann Surg.* <https://doi.org/10.1097/SLA.0000000000000937>
 18. Palanivelu C, Senthilnathan P, Sabnis SC, Babu NS, Srivatsan Gurumurthy S, Anand Vijai N, Nalankilli VP, Praveen Raj P, Parthasarathy R, Rajapandian S (2017) Randomized clinical trial of laparoscopic versus open pancreaticoduodenectomy for periampullary tumours. *Br J Surg.* <https://doi.org/10.1002/bjs.10662>
 19. Poves I, Burdío F, Morató O IM, Radosevic A, Ilzarbe L, Visa L, Grande L (2018) Comparison of perioperative outcomes between laparoscopic and open approach for pancreaticoduodenectomy: the PADULAP randomized controlled trial. *Ann Surg.* <https://doi.org/10.1097/SLA.0000000000002893>
 20. Van Hilst J, de Rooij T, Bosscha K, Brinkman DJ, van Dieren S, Dijkgraaf MG, Gerhards MF, de Hingh IH, Karsten TM, Lips DJ, Luyer MD, Busch OR, Festen S, Besselink MG, Dutch Pancreatic Cancer Group (2019) Laparoscopic versus open pancreaticoduodenectomy for pancreatic or periampullary tumours (LEOPARD-2): a multicentre, patient-blinded, randomised controlled phase 2/3 trial. *Lancet Gastroenterol Hepatol.* [https://doi.org/10.1016/S2468-1253\(19\)30004-4](https://doi.org/10.1016/S2468-1253(19)30004-4)
 21. Giulianotti PC, Coratti A, Angelini M, Sbrana F, Ceconi S, Balestracci T, Caravaglios G (2003) Robotics in general surgery: personal experience in a large community hospital. *Arch Surg* 138:777–784
 22. Cai J, Ramanathan R, Zenati MS, Abbas A, Hogg ME, Zeh HJ, Zureikat AH (2019) Robotic pancreaticoduodenectomy is associated with decreased clinically relevant pancreatic fistulas: a propensity-matched analysis. *J Gastrointest Surg.* <https://doi.org/10.1007/s11605-019-04274-1>
 23. Charlson ME, Ales KL, MacKenzie PP (1987) A new method of classifying prognostic comorbidity in longitudinal studies—development and validation. *J Chron Dis* 40(5):373–383
 24. Ament R (1979) Origin of the ASA classification. *Anesthesiology* 51(2):179
 25. Bassi C, Marchegiani G, Dervenis C, Sarr M, Abu Hilal M, Adham M, Allen P, Andersson R, Asbun HJ, Besselink MG, Conlon K, Del Chiaro M, Falconi M, Fernandez-Cruz L, Fernandez-Del Castillo C, Fingerhut A, Friess H, Gouma DJ, Hackert T, Izicki J, Lillemoe KD, Neoptolemos JP, Olah A, Schulick R, Shrikhande SV, Takada T, Takaori K, Traverso W, Vollmer CR, Wolfgang CL, Yeo CJ, Salvia R, Büchler M, International Study Group on Pancreatic Surgery (ISGPS) (2017) The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 years after. *Surgery.* <https://doi.org/10.1016/j.surg.2016.11.014>
 26. Wente MN, Veit JA, Bassi C, Dervenis C, Fingerhut A, Gouma DJ, Izicki JR, Neoptolemos JP, Padbury RT, Sarr MG, Yeo CJ, Büchler MW (2007) Postpancreatectomy hemorrhage (PPH)—an International Study Group of Pancreatic Surgery (ISGPS) definition. *Surgery* 142(1):20–25
 27. Wente MN, Bassi C, Dervenis C, Fingerhut A, Gouma DJ, Izicki JR, Neoptolemos JP, Padbury RT, Sarr MG, Traverso LW, Yeo CJ, Büchler MW (2007) Delayed gastric emptying (DGE) after pancreatic surgery: a suggested definition by the International Study Group of Pancreatic Surgery (ISGPS). *Surgery* 142(5):761–768
 28. Dindo D, Demartines N, Clavien P-A (2004) Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 240:205–213
 29. Edge SB, Compton CC (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol.* <https://doi.org/10.1245/s10434-010-0985-4>
 30. Ho EO, Imai K, King G, Stuart EA (2011) Match it: nonparametric preprocessing for parametric causal inference. *J Stat Softw.* <https://doi.org/10.18637/jss.v042.i08>
 31. Coratti A, Annecchiarico M (2014) Robot-assisted pancreatic surgery. *Br J Surg.* <https://doi.org/10.1002/bjs.9416>
 32. Klompmaker S, van Hilst J, Wellner UF, Busch OR, Coratti A, D'Hondt M, Dokmak S, Festen S, Kerem M, Khatkov I, Lips DJ, Lombardo C, Luyer M, Manzoni A, Molenaar IQ, Rosso E, Saint-Marc O, Vansteenkiste F, Wittel UA, Bonsing B, Groot Koerkamp B, Abu Hilal M, Fuks D, Poves I, Keck T, Boggi U, Besselink MG, European Consortium on Minimally Invasive Pancreatic Surgery (E-MIPS) (2018) Outcomes after minimally-invasive versus open pancreaticoduodenectomy: a pan-European propensity score matched study. *Ann Surg.* <https://doi.org/10.1097/SLA.00000000000002850>
 33. Boggi U, Napoli N, Costa F, Kauffmann EF, Menonna F, Iacopi S, Vistoli F, Amorese G (2016) Robotic-assisted pancreatic resections. *World J Surg.* <https://doi.org/10.1007/s00268-016-3565-3>
 34. Zureikat AH, Postlewait LM, Liu Y, Gillespie TW, Weber SM, Abbott DE, Ahmad SA, Maithel SK, Hogg ME, Zenati M, Cho CS, Salem A, Xia B, Steve J, Nguyen TK, Keshava HB, Chalikonda S, Walsh RM, Talamonti MS, Stocker SJ, Bentrem DJ, Lumpkin S, Kim HJ, Zeh HJ 3rd, Kooby DA (2016) A multi-institutional comparison of perioperative outcomes of robotic and open pancreaticoduodenectomy. *Ann Surg.* <https://doi.org/10.1097/SLA.0000000000001869>

35. Lei P, Wei B, Guo W, Wei H (2014) Minimally invasive surgical approach compared with open pancreaticoduodenectomy: a systematic review and meta-analysis on the feasibility and safety. *Surg Laparosc Endosc Percutan Technol.* <https://doi.org/10.1097/SLE.0000000000000054>
36. Kendrick ML, van Hilst J, Boggi U, de Rooij T, Walsh RM, Zeh HJ, Hughes SJ, Nakamura Y, Vollmer CM, Kooby DA, Asbun HJ (2017) Minimally Invasive Pancreatic Resection Organizing Committee minimally invasive pancreatectomy. *HPB (Oxford)*. <https://doi.org/10.1016/j.hpb.2017.01.023>
37. Nickel F, Haney CM, Kowalewski KF, Probst P, Limen EF, Kalkum E, Diener MK, Strobel O, Müller-Stich BP, Hackert T (2019) Laparoscopic versus open pancreaticoduodenectomy: a systematic review and meta-analysis of randomized controlled trials. *Ann Surg.* <https://doi.org/10.1097/SLA.0000000000003309>
38. Callery MP, Pratt WB, Kent TS, Chaikof EL, Vollmer CM Jr (2013) A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreatectomy. *J Am Coll Surg.* <https://doi.org/10.1016/j.jamcollsurg.2012.09.002>
39. Zhao W, Liu C, Li S, Geng D, Feng Y (2018) Sun (2018) Safety and efficacy for robot-assisted versus open pancreaticoduodenectomy and distal pancreatectomy: a systematic review and meta-analysis. *Surg Oncol.* 27(3):468–478. <https://doi.org/10.1016/j.suronc.2018.06.001>
40. Yamada S, Fujii T, Hirakawa A, Kanda M, Sugimoto H, Kodera Y (2016) Lymph node ratio as parameter of regional lymph node involvement in pancreatic cancer. *Langenbecks Arch Surg.* <https://doi.org/10.1007/s00423-016-1412-52016>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Claims data analyses unable to properly characterize the value of neurologists in epilepsy care

Chloe E. Hill, MD, MS, Chun Chieh Lin, PhD, MBA, James F. Burke, MD, MS, Kevin A. Kerber, MD, MS, Lesli E. Skolarus, MD, MS, Gregory J. Esper, MD, MBA, Brandon Magliocco, MS, and Brian C. Callaghan, MD, MS

Neurology® 2019;92:e973-e987. doi:10.1212/WNL.0000000000007004

Correspondence

Dr. Hill

chloehi@med.umich.edu

Abstract

Objective

To determine the association of a neurologist visit with health care use and cost outcomes for patients with incident epilepsy.

Methods

Using health care claims data for individuals insured by United Healthcare from 2001 to 2016, we identified patients with incident epilepsy. The population was defined by an epilepsy/convulsion diagnosis code (ICD codes 345.xx/780.3x, G40.xx/R56.xx), an antiepileptic prescription filled within the succeeding 2 years, and neither criterion met in the 2 preceding years. Cases were defined as patients who had a neurologist encounter for epilepsy within 1 year after an incident diagnosis; a control cohort was constructed with propensity score matching. Primary outcomes were emergency room (ER) visits and hospitalizations for epilepsy. Secondary outcomes included measures of cost (epilepsy related, not epilepsy related, and antiepileptic drugs) and care escalation (including EEG evaluation and epilepsy surgery).

Results

After participant identification and propensity score matching, there were 3,400 cases and 3,400 controls. Epilepsy-related ER visits were more likely for cases than controls (year 1: 5.9% vs 2.3%, $p < 0.001$), as were hospitalizations (year 1: 2.1% vs 0.7%, $p < 0.001$). Total medical costs for epilepsy care, nonepilepsy care, and antiepileptic drugs were greater for cases ($p \leq 0.001$). EEG evaluation and epilepsy surgery occurred more commonly for cases ($p \leq 0.001$).

Conclusions

Patients with epilepsy who visited a neurologist had greater subsequent health care use, medical costs, and care escalation than controls. This comparison using administrative claims is plausibly confounded by case disease severity, as suggested by higher nonepilepsy care costs. Linking patient-centered outcomes to claims data may provide the clinical resolution to assess care value within a heterogeneous population.

From the Health Services Research Program (C.E.H., C.C.L., J.F.B., K.A.K., L.E.S., B.C.C.), Department of Neurology, University of Michigan, Ann Arbor; Department of Neurology (G.J.E.), Emory University, Atlanta, GA; and American Academy of Neurology (B.M.), Minneapolis, MN.

Go to *Neurology.org/N* for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

ER = emergency room; ICD = International Classification of Diseases; VNS = vagal nerve stimulation.

As payment becomes increasingly tied to high-value neurologic care, it is imperative that we use rigorous measurement of quality and cost.¹ This is particularly relevant to the care of patients with epilepsy, a disease with 1% prevalence in the United States, complex medical needs, and high economic burden.^{2,3} Total annual direct health care costs per person with epilepsy are estimated at \$10,000 to \$48,000,⁴ with indirect costs projected to be substantially higher.⁵

Epilepsy is defined as at least 1 unprovoked seizure with an elevated risk of seizure recurrence.⁶ Therefore, epilepsy encompasses a heterogeneous population of patients with a broad range of etiologies and severities.² Practice guidelines advise that many patients presenting with seizures can initially be managed in the primary care setting.⁷ However, early expert care may be warranted for patients with persistent seizures.⁸ Prior study suggests that patients with epilepsy may benefit from specialized care,⁹ but the full effect of neurologists has not been well characterized.

With the current implementation of the Medicine Access and CHIP Reauthorization Act and heightened interest in alternative payment models,¹⁰ careful consideration of clinical outcomes and cost is necessary to determine how to maximize the value of neurologist care. We, the American Academy of Neurology Health Services Research Subcommittee, chose to investigate value using administrative claims data, which has previously been used to assess the role of neurologists in other diseases such as stroke and headache.^{11–13} This study, funded by the American Academy of Neurology, aims to determine the association of a neurologist visit on health care use and cost outcomes for patients with incident epilepsy using a large, private insurance claims dataset.

Methods

Data source and study population

We performed a retrospective analysis using data from the OptumInsight Clininformatics Data Mart (Optum.com, Eden Prairie, MN), a database of inpatient medical, outpatient medical, pharmacy, and laboratory administrative claims for individuals insured by United Healthcare, from 2001 to 2016. This database includes deidentified information for ≈12 to 14 million annual covered lives, for a total of ≈73 million unique lives over the study period.

We identified adult patients with incident epilepsy defined by (1) an ICD diagnosis code for epilepsy or convulsion (345.xx/780.3x or G40.xx/R56.xx), (2) a prescription filled for an antiepileptic medication at the time of diagnosis or in the subsequent 2 years, and (3) neither an epilepsy-related

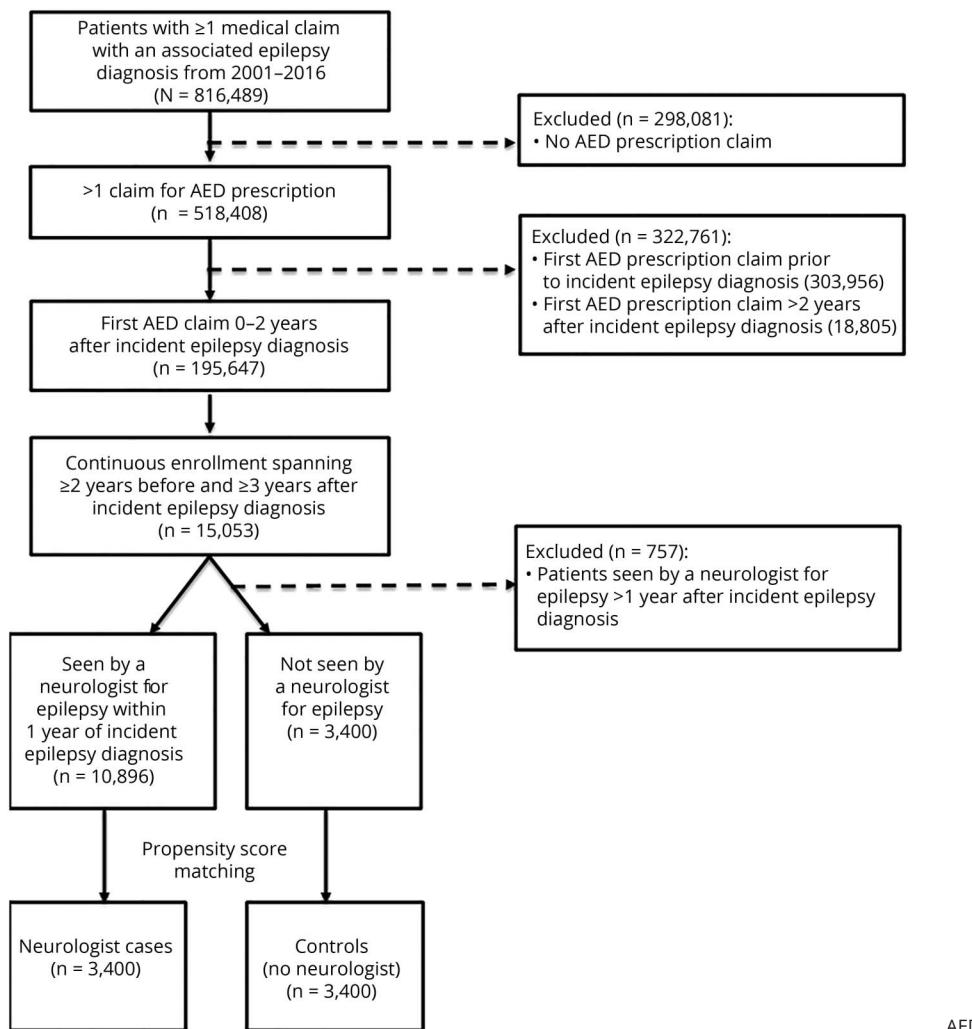
diagnosis code nor an antiepileptic prescription in the preceding 2 years. This definition of epilepsy, which includes both diagnostic code and antiepileptic medication, is based on validated criteria for administrative claims data (positive predictive value 84%).^{14,15} Patients were excluded if they were not enrolled in insurance coverage for 5 continuous years, with at least 2 years of prediagnosis data and at least 3 years of follow-up data. In addition, to have a sufficient follow-up period to assess neurologist influence on longer-term outcomes, patients were excluded if their visit to a neurologist occurred >1 year after their incident diagnosis (figure 1).

Exposure

The exposure of interest was a neurologist encounter for epilepsy (primary or secondary ICD diagnosis code). A neurologist was identified by provider category code or the National Uniform Claim Committee taxonomy code in the Optumlab data. The case patients were compared to the population of patients with incident epilepsy who did not see a neurologist for epilepsy.

To adjust for pretreatment observable differences between cases and controls, a subsample was created through propensity score modeling.¹⁶ A propensity score was calculated from multilevel logistic regression to estimate the probability of not being seen by a neurologist, conditional on matching variables that included the following: age, sex, 17 Charlson Comorbidity Index categories, preexisting psychiatric comorbid conditions (anxiety: ICD codes 293.84, 300.0x; depression: 296.2x, 296.30, 296.31, 296.32, 296.33, 296.34, 296.35, 296.36, 296.82, 311.00; bipolar disorder: 296.0–296.16, 296.4–296.81, 296.83–296.89, 301.13; alcohol dependence/abuse: 303.9x, 305.0x), preexisting neurologic comorbid conditions (head injury: 850.xx–854.xx; malignant brain tumor: 191.xx, 192.1, 198.3; benign brain tumor: 225.0, 225.2, 237.5, 237.6, 239.6; tuberous sclerosis: 759.5; migraine headache: 346.xx; nonmigrainous headache: 307.81, 339.xx, 784.0; neuropathy: 356.xx, 357.xx [excluding 357.0 and 357.81]; chronic pain: 053.10, 053.11, 053.12, 053.13, 333.94, 337.2x, 338.0–338.2x, 338.3, 338.4, 350.1, 353.6, 722.8x, 723.1, 423.2, 423.3, 423.4, 724.03, 724.1, 724.2, 724.3, 724.4, 724.5, 729.1, 729.2, 780.96; of note, cerebrovascular disease was captured as a Charlson comorbidity), total medical expenditures in the 2 years before diagnosis, medical service use in the 2 years before incident diagnosis (hospitalization, skilled nursing facility admission, emergency room [ER] visit, neurologist visit), psychiatric treatment in the 2 years before incident diagnosis (psychiatrist visit claims, antidepressant medication by Optum pharmacy claim category), benefit design (exclusive provider organization, health maintenance organization, preferred provider organization, indemnity, point of service, other), days of available claims data before incident diagnosis (eligibility time), setting of incident

Figure 1 Study flow diagram



AED = antiepileptic drug.

epilepsy diagnosis (hospital, skilled nursing facility, ER, clinic), year of incident diagnosis, and a random region-level intercept defined by hospital service area (Dartmouth Atlas of Health Care, dartmouthatlas.org) to account for geographic variation in neurologist availability. A 1:1 nearest-neighbor match was performed without replacement. Baseline characteristics before propensity score matching are summarized in table 1.

Lag time variables

For the cases, the index date was defined as the first neurology encounter after the incident epilepsy diagnosis. Each control patient was then assigned an index date that corresponded to his/her matched case. The length of time between the incident epilepsy diagnosis and the index date (the lag time, capped at 1 year by exclusion criterion) and the health care use within that period varied from patient to patient. Therefore, all analyses were adjusted for relevant lag time variables, including the following: length of lag time, diagnosis of status epilepticus or intractable epilepsy (by ICD diagnosis code), ER visits (epilepsy related, trauma/injury related),

hospitalization for epilepsy, medical costs (epilepsy related, not epilepsy related, antiepileptic medication), antiepileptic prescription, care for comorbid conditions (bone health evaluation, antidepressant prescription, psychiatric care), and occurrence of epilepsy-specific care (EEG, presurgical evaluation, intracranial electrode implantation, epilepsy surgery, vagal nerve stimulation [VNS]). For women of childbearing age, defined as 18 to 44 years, valproic acid prescription was also included (table 2).

Outcomes

Our primary outcomes were ER visits and hospitalizations for epilepsy (primary ICD diagnosis code).

Cost outcomes included epilepsy-related expenditures (anti-epileptic medications, head CT, brain MRI, EEG, epilepsy surgical procedures, epilepsy outpatient encounters, epilepsy hospital encounters), non–epilepsy-related expenditures, and antiepileptic drug costs. Antiepileptic drug use was assessed by medication continuation after initial prescription, days of medication supply (measured by summation of all

Table 1 Baseline characteristics for neurologist and no neurologist cohorts before propensity score matching

Variable	Baseline characteristics		Likelihood of being seen by a neurologist		Adjusted OR (95% CI)	<i>p</i> Value
	Neurologist cases (n = 10,896)	No neurologist controls (n = 3,400)	<i>p</i> Value			
Age, mean (SD), y	55.5 (SD 19.0)	61.0 (SD 19.0)	<0.001 ^a	0.98 (0.98–0.99)	<0.001	
Sex, n (%)			0.44 ^b			
Male	4,793 (44.0)	1,487 (43.7)	—	Referent	—	
Female	6,098 (56.0)	1,913 (56.3)	—	1.02 (0.94–1.12)	0.96	
Unknown	5 (0.05)	0 (0)	—	—	—	
Charlson comorbid conditions, n (%)						
Myocardial infarction	743 (6.8)	203 (6.0)	0.08 ^b	1.13 (0.94–1.37)	0.19	
Congestive heart failure	1,294 (11.9)	408 (12.0)	0.85 ^b	0.97 (0.84–1.12)	0.68	
Peripheral vascular disease	1,702 (15.6)	456 (13.4)	0.002 ^b	1.22 (1.07–1.40)	0.003	
Cerebrovascular disease	3,513 (32.2)	837 (24.6)	<0.001 ^b	1.61 (1.43–1.81)	<0.001	
COPD	3,255 (29.9)	907 (26.7)	<0.001 ^b	1.05 (0.95–1.16)	0.32	
Dementia	692 (6.4)	242 (7.1)	0.11 ^b	0.96 (0.80–1.14)	0.64	
Paralysis	591 (5.4)	154 (4.5)	0.04 ^b	0.86 (0.70–1.07)	0.17	
Diabetes mellitus	1,415 (13.0)	470 (13.8)	0.21 ^b	0.92 (0.81–1.04)	0.17	
Diabetes mellitus with complications	1,047 (9.6)	315 (9.3)	0.55 ^b	1.00 (0.85–1.17)	0.96	
Renal disease	1,054 (9.7)	312 (9.2)	0.39 ^b	1.17 (0.99–1.37)	0.06	
Mild liver disease	160 (1.5)	49 (1.4)	0.91 ^b	0.95 (0.67–1.35)	0.79	
Moderate to severe liver disease	65 (0.6)	28 (0.8)	0.15 ^b	0.64 (0.39–1.03)	0.07	
Peptic ulcer disease	386 (3.5)	127 (3.7)	0.60 ^b	0.89 (0.71–1.11)	0.30	
Rheumatologic disease	503 (4.6)	137 (4.0)	0.15 ^b	1.01 (0.82–1.25)	0.91	
HIV/AIDS	59 (0.5)	16 (0.5)	0.62 ^b	1.04 (0.57–1.88)	0.91	
Cancer	1,276 (11.7)	341 (10.0)	0.007 ^b	1.19 (1.03–1.39)	0.02	
Metastatic solid tumor	207 (1.9)	56 (1.7)	0.34 ^b	0.78 (0.56–1.09)	0.15	
Psychiatric comorbid conditions, n (%)						
Anxiety	2,150 (19.7)	613 (18.0)	0.03 ^b	0.98 (0.87–1.10)	0.74	
Depression	897 (8.2)	269 (7.9)	0.55 ^b	1.02 (0.86–1.21)	0.82	
Bipolar disorder	225 (2.1)	110 (3.2)	<0.001 ^b	0.52 (0.40–0.67)	<0.001	
Alcohol dependency/abuse	370 (3.4)	130 (3.8)	0.24 ^b	0.84 (0.67–1.05)	0.13	
Neurologic comorbid conditions, n (%)						
Head injury	560 (5.1)	168 (4.9)	0.65 ^b	0.87 (0.71–1.06)	0.16	
Brain tumor, malignant	133 (1.2)	27 (0.8)	0.04 ^b	1.24 (0.75–2.04)	0.41	
Brain tumor, benign	252 (2.3)	58 (1.7)	0.03 ^b	1.08 (0.77–1.51)	0.66	
Tuberous sclerosis	3 (0.03)	0 (0)	0.33 ^b	—	—	
Migraine	1,138 (10.4)	272 (8.0)	<0.001 ^b	0.78 (0.67–0.91)	0.002	
Nonmigrainous headache	3,556 (32.6)	809 (23.8)	<0.001 ^b	1.13 (1.01–1.26)	0.03	
Neuropathy	731 (6.7)	205 (6.0)	0.16 ^b	0.94 (0.78–1.13)	0.50	

Continued

Table 1 Baseline characteristics for neurologist and no neurologist cohorts before propensity score matching (continued)

Variable	Baseline characteristics		Likelihood of being seen by a neurologist		Adjusted OR (95% CI)	p Value
	Neurologist cases (n = 10,896)	No neurologist controls (n = 3,400)	p Value			
Chronic pain	5,153 (47.3)	1,338 (39.4)	<0.001 ^b	1.12 (1.02–1.23)	0.02	
Medical service use in preceding 2 y, n (%)						
Hospitalization	2,489 (22.8)	744 (21.9)	0.24 ^b	0.90 (0.80–1.02)	0.11	
Skilled nursing facility	423 (3.9)	197 (5.8)	<0.001 ^b	0.75 (0.60–0.92)	0.007	
ER visits, mean (SD)	1.54 (2.99)	1.45 (2.79)	0.09 ^a	0.98 (0.96–1.00)	0.01	
Neurologist visits, mean (SD)	0.31 (0.46)	0.16 (0.36)	<0.001 ^a	2.37 (2.10–2.67)	<0.001	
Psychiatrist visit	1,179 (10.8)	319 (9.4)	0.02 ^b	1.02 (0.87–1.20)	0.80	
Antidepressant medication	1,682 (15.4)	630 (18.5)	<0.001 ^b	0.63 (0.55–0.72)	<0.001	
Total expenditures in preceding 2 y, mean (SD), \$	41,751 (102,253)	38,743 (106,390)	0.15 ^a	1.04 (1.02–1.07)	<0.001	
Eligibility time, mean (SD), d	1,678 (834)	1,744 (881)	<0.001 ^a	1 (1–1)	0.11	
Benefit design, n (%)						
Exclusive provider organization	911 (8.4)	165 (4.9)	—	1.00 (0.83–1.20)	0.006	
Health maintenance organization	2,971 (27.3)	1,748 (51.4)	—	0.41 (0.36–0.46)	<0.001	
Indemnity	503 (4.6)	119 (3.5)	—	0.96 (0.76–1.22)	0.06	
Other	1,610 (14.8)	357 (10.5)	—	1.00 (0.85–1.18)	<0.001	
Point of service	4,274 (39.2)	815 (24.0)	—	Referent	—	
Preferred provider organization	627 (5.8)	196 (5.8)	—	0.68 (0.56–0.82)	0.03	
Setting of incident diagnosis, n (%)						
Hospital	997 (9.2)	657 (19.3)	—	0.30 (0.27–0.34)	<0.001	
Skilled nursing facility	42 (0.4)	73 (2.2)	—	0.14 (0.09–0.22)	<0.001	
ER	5,763 (52.9)	1,536 (45.2)	—	Referent	—	
Clinic	4,094 (37.6)	1,134 (33.4)	—	0.75 (0.68–0.83)	<0.001	
Year of incident diagnosis, n (%)						
2003	570 (5.2)	142 (4.9)	—	Referent	—	
2004	785 (7.2)	194 (5.7)	—	1.05 (0.79–1.40)	<0.001	
2005	1,005 (9.2)	494 (14.5)	—	0.61 (0.47–0.80)	<0.001	
2006	864 (7.9)	268 (7.9)	—	0.69 (0.51–0.93)	0.11	
2007	947 (8.7)	275 (8.1)	—	0.70 (0.51–0.94)	0.15	
2008	1,127 (10.3)	355 (10.4)	—	0.68 (0.51–0.92)	0.06	
2009	1,254 (11.5)	343 (10.1)	—	0.85 (0.63–1.14)	0.14	
2010	1,350 (12.4)	387 (11.4)	—	0.78 (0.58–1.05)	0.89	
2011	1,143 (10.5)	338 (9.9)	—	0.79 (0.58–1.06)	0.75	
2012	1,209 (11.1)	378 (11.1)	—	0.75 (0.55–1.01)	0.67	
2013	642 (5.9)	226 (6.7)	—	0.69 (0.50–0.95)	0.20	

Abbreviations: CI = confidence interval; COPD = chronic obstructive pulmonary disease; ER = emergency room; OR = odds ratio..

^a By t test.

^b By χ^2 test.

Table 2 Lag time variables

Variable	Neurologist cases (n = 3,400)	No neurologist controls (n = 3,400)	p Value
Length of lag time, mean (SD), d	15.9 (52.4)	15.9 (52.4)	1.00 ^a
Status epilepticus, n (%)	26 (0.8)	17 (0.5)	0.17 ^b
Intractable epilepsy, n (%)	16 (0.5)	19 (0.6)	0.61 ^b
ER visit, n (%)			
ER visit for epilepsy	96 (2.8)	35 (1.0)	<0.001 ^b
ER visit for trauma/injury	26 (0.8)	21 (0.6)	0.46 ^b
Hospitalization for epilepsy	7 (0.2)	9 (0.3)	0.62 ^b
Medical costs (SD), \$			
Epilepsy-related medical costs	1,296 (6,407)	897 (6,027)	0.008 ^a
Non-epilepsy-related medical costs	4,912 (45,392)	4,867 (28,200)	0.96 ^a
Antiepileptic medication costs	32 (308)	28 (284)	0.58 ^a
Antiepileptic medication			
AED prescription, n (%)	338 (9.9)	339 (10.0)	0.97 ^b
Antiepileptic pill-days (SD)	8.4 (43.2)	8.1 (36.4)	0.76 ^a
Generation 2 prescription, n (%)	160 (4.7)	178 (5.2)	0.32 ^a
Care for comorbid conditions, n (%)			
Bone health evaluation	25 (0.7)	14 (0.4)	0.08 ^b
Antidepressant prescription	73 (2.2)	92 (2.7)	0.13 ^b
Psychiatric care	56 (1.7)	72 (2.1)	0.15 ^b
Care escalation, n (%)			
EEG	99 (2.9)	68 (2.0)	0.02 ^b
Presurgical evaluation	0 (0)	0 (0)	—
Intracranial electrode implantation	0 (0)	0 (0)	—
Epilepsy surgery	4 (0.1)	1 (0.03)	0.18 ^b
VNS	0 (0)	0 (0)	—
Among women 18–44 y of age, n (%)			
Valproic acid prescription	4 (1.0)	3 (0.7)	0.69 ^b

Abbreviations: AED = antiepileptic drug; ER = emergency room; VNS = vagal nerve stimulation.

^a By t test.^b By χ^2 test.

antiepileptic daily doses prescribed over the year), and prescription of generation 2 antiepileptic drugs (included generation 1 drugs were carbamazepine, clonazepam, ethosuximide, ethotoin, mephobarbital, methsuximide, phenobarbital, phenytoin, primidone, and valproic acid; included generation 2 drugs were brivaracetam, clobazam, eslicarbazepine, ezogabine, felbamate, gabapentin, lacosamide, lamictal, levetiracetam, oxcarbazepine, perampanel, pregabalin, rufinamide, tiagabine, topiramate, vigabatrin, and zonisamide).

Quality of care was examined through measures of care escalation, including use of EEG monitoring, epilepsy

presurgical evaluation (brain PET, brain SPECT, fMRI, Wada), epilepsy surgery, VNS, and antiepileptic therapy augmentation for breakthrough seizures (defined as an increase in antiepileptic drug dose or new antiepileptic prescription ≤ 30 days after an ER visit). We also measured attention to psychiatric disorders (antidepressant medication prescription and psychiatrist visits), bone health evaluation (vitamin D monitoring and dual-energy x-ray absorptiometry scan), and specific consideration of women of childbearing age (valproic acid prescription). Presence of intractable epilepsy (345.01, 345.11, 345.41, 345.51, 345.61, 345.71, 345.81, 345.91, G40.01x, G40.11x, G40.21x, G40.31x, G40.A1x,

G40.B1x, G40.41x, G40.803, G40.804, G40.813, G40.814, G40.823, G40.824, G40.91x) and status epilepticus (345.3x, G40.001, G40.011, G40.101, G40.111, G40.201, G40.211, G40.301, G40.311, G40.A01, G40.A11, G40.B01, G40.B11, G40.401, G40.411, G40.501, G40.801, G40.803, G40.811, G40.813, G40.821, G40.823, G40.901, G40.911) was noted, as well as presentations to the ER for trauma or injury (800.xx–829.xx, 830.xx–839.xx, 840.xx–848.xx, 850.xx–854.xx, 860.xx–869.xx, 870.xx–897.xx, 910.xx–924.xx, 940.xx–949.xx, 959.xx, E810.x–E825.x, S00–S99, T07–T14, T20–T32).

Sensitivity analyses

We evaluated variations on the definition of epilepsy and the window of continuous insurance coverage to investigate the effect on outcomes. These secondary analyses were (1) the requirement of 2 epilepsy diagnosis codes (increased diagnostic specificity), (2) only 1 year of prediagnosis data and only 2 years of follow-up data (increased generalizability), and (3) no restriction on data before the incident epilepsy diagnosis code (prevalent epilepsy).

Statistical analysis

Descriptive statistics were calculated to summarize demographics and clinical characteristics of the cases and controls. Covariate balance was examined before and after propensity score matching with χ^2 tests for categorical variables and *t* tests for continuous variables. Separate regression models (logistic regression for binary outcomes and linear regression for continuous outcomes) were built to estimate the association between neurologist visit and each individual outcome over time. Specifically, each outcome was estimated per year (e.g., total expenditures in years 1–5 as measured from index date) for each patient. Multilevel regression models were fit for each outcome after adjustment for patient age, lag time variables, and baseline covariates that were still unbalanced after matching (chronic obstructive pulmonary disease, preexisting anxiety diagnosis, preexisting chronic pain, eligibility time, and setting of incident epilepsy diagnosis) and including a random participant-level intercept. The influence of neurologist visits over time on each outcome was estimated with average marginal effects, with the random intercept set at its mean. Two-sided *p* values were reported and were considered significant at *p* < 0.05. Statistical analyses were performed with SAS 9.4 (SAS institute, Cary, NC) and STATA 14 (StataCorp, College Station, TX).

All results presented are adjusted for unbalanced baseline covariates, patient age, and lag time variables and are reported as predicted probabilities, which may be interpreted as adjusted outcomes over time comparing patients who did and did not have a neurologist visit.

Standard protocol approvals, registrations, and patient consents

This study used deidentified data and was determined to be exempt from review by the University of Michigan Institutional Review Board.

Data availability

The full dataset, OptumInsight Clininformatics Data Mart, is available through Optum (Optum.com).

Results

Demographic and clinical characteristics

We identified a total of 10,896 patients who saw a neurologist at or after their incident epilepsy diagnosis and 3,400 patients who did not see a neurologist for epilepsy. After propensity score matching, 3,400 cases (neurologist) and 3,400 controls (no neurologist) remained. Baseline characteristics of the study groups are presented in table 3. At the time of diagnosis, some small differences persisted between cases and controls: cases were more likely to have chronic obstructive pulmonary disease (30.0% vs 26.7%, *p* = 0.002), a diagnosis of anxiety (20.6% vs 18.0%, *p* = 0.008), and a diagnosis of chronic pain (43.7% vs 39.4%, *p* < 0.001). Compared to controls, cases were more frequently diagnosed in clinic (35.5% vs 33.4%, *p* = 0.007) than at an ER or inpatient facility. Lastly, cases were eligible for insurance coverage for a shorter length of time before incident diagnosis (mean 1,680 days vs 1,744, *p* = 0.002). Medical use and total medical expenditures in the 2 years preceding epilepsy diagnosis were similar between the 2 cohorts.

Cost: Acute care use

ER visits for epilepsy were more likely for cases than controls (predicted probabilities year 1: 5.9% vs 2.3%, *p* < 0.001), and the magnitude of difference decreased over time (year 5: 1.0% vs 0.4%, *p* < 0.001) (figure 2A). A similar pattern was observed with the number of ER visits for epilepsy (predicted visits year 1: 0.14 vs 0.07, *p* < 0.001; year 5: 0.04 vs 0.02, *p* < 0.03). Hospitalization for epilepsy was also higher for cases than controls (predicted probabilities year 1: 2.1% vs 0.7%, *p* < 0.001); this difference also decreased over time (year 5: 0.4% vs 0.2%, *p* < 0.001) (figure 2B).

Cost: Total medical expenditures

Epilepsy-related total medical costs were higher for cases than controls (predicted costs year 1: \$5,464 vs \$2,364, *p* < 0.001), and costs declined for both cohorts through the study period (year 5: \$2,111 vs \$1,051, *p* < 0.001) (figure 2C). Total cost of nonepilepsy care was similarly greater for cases than controls (year 1: \$38,082 vs \$32,135; year 5: \$28,861 vs \$26,638, *p* = 0.001) (figure 2D). Antiepileptic drug costs remained substantially higher for cases throughout the study period (*p* < 0.001) (figure 2E).

Quality of care: Measures of standard evaluation and care escalation

Evaluation by EEG was much more common for cases than controls, most notably in the first year (predicted probabilities year 1: 78.6% vs 7.2%, *p* < 0.001) (table 4). Completion of epilepsy surgery was more common for cases than controls (year 1: 0.2% vs 0.05%, *p* = 0.001), although surgery was rare in both cohorts, occurring in total for only 47 cases (1.4%) and 16 controls (0.5%). There were no

Table 3 Baseline characteristics for neurologist and no neurologist cohorts after propensity score matching

Variable	Neurologist cases (n = 3,400)	No neurologist controls (n = 3,400)	p Value
Age, mean (SD), y	60.2 (18.2)	61.0 (18.5)	0.07 ^a
Sex, n (%)			0.59 ^b
Male	1,476 (43.4)	1,487 (43.7)	—
Female	1,923 (56.6)	1,913 (56.3)	—
Unknown	1 (0.03)	0 (0)	—
Charlson comorbid conditions, n (%)			
Myocardial infarction	220 (6.5)	203 (6.0)	0.39 ^b
Congestive heart failure	440 (12.9)	408 (12.0)	0.24 ^b
Peripheral vascular disease	506 (14.9)	456 (13.4)	0.08 ^b
Cerebrovascular disease	888 (26.1)	837 (24.6)	0.16 ^b
COPD	1,020 (30.0)	907 (26.7)	0.002 ^b
Dementia	269 (7.9)	242 (7.1)	0.21 ^b
Paralysis	154 (4.5)	154 (4.5)	1.00 ^b
Diabetes mellitus	497 (14.6)	470 (13.8)	0.35 ^b
Diabetes mellitus with complications	346 (10.2)	315 (9.3)	0.20 ^b
Renal disease	313 (9.2)	312 (9.2)	0.97 ^b
Mild liver disease	51 (1.5)	49 (1.4)	0.84 ^b
Moderate-severe liver disease	29 (0.9)	28 (0.8)	0.89 ^b
Peptic ulcer disease	155 (4.6)	127 (3.7)	0.09 ^b
Rheumatologic disease	144 (4.2)	137 (4.0)	0.67 ^b
HIV/AIDS	17 (0.5)	16 (0.5)	0.86 ^b
Cancer	363 (10.7)	341 (10.0)	0.38 ^b
Metastatic solid tumor	63 (1.9)	56 (1.7)	0.52 ^b
Psychiatric comorbid conditions, n (%)			
Anxiety	700 (20.6)	613 (18.0)	0.008 ^b
Depression	312 (9.2)	269 (7.9)	0.06 ^b
Bipolar disorder	124 (3.7)	110 (3.2)	0.35 ^b
Alcohol dependency/abuse	145 (4.3)	130 (3.8)	0.36 ^b
Neurologic comorbid conditions, n (%)			
Head injury	167 (4.9)	168 (4.9)	0.96 ^b
Brain tumor, malignant	27 (0.8)	27 (0.8)	1.00 ^b
Brain tumor, benign	59 (1.7)	58 (1.7)	0.93 ^b
Tuberous sclerosis	0 (0)	0 (0)	1.00 ^b
Migraine	291 (8.6)	272 (8.0)	0.40 ^b
Nonmigrainous headache	878 (25.8)	809 (23.8)	0.05 ^b
Neuropathy	214 (6.3)	205 (6.0)	0.65 ^b
Chronic pain	1,484 (43.7)	1,338 (39.4)	<0.001 ^b

Continued

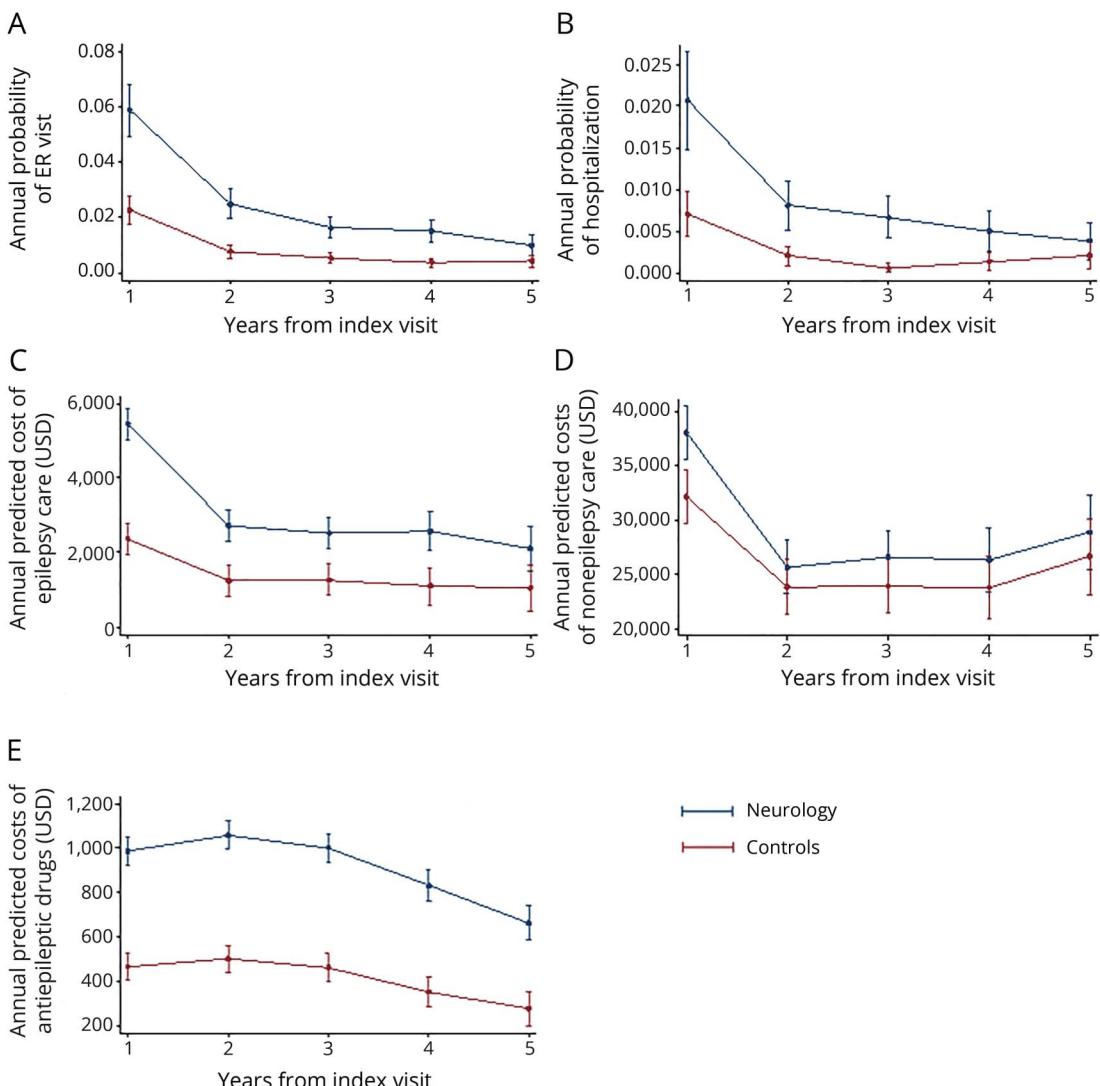
Table 3 Baseline characteristics for neurologist and no neurologist cohorts after propensity score matching (continued)

Variable	Neurologist cases (n = 3,400)	No neurologist controls (n = 3,400)	p Value
Medical service use in preceding 2 y			
Hospitalization, n (%)	765 (22.5)	744 (21.9)	0.54 ^b
Skilled nursing facility admission, n (%)	181 (5.3)	197 (5.8)	0.40 ^b
ER visits, mean (SD), n	1.60 (3.49)	1.45 (2.79)	0.06 ^a
Neurologist visits, mean (SD), n	0.16 (0.37)	0.16 (0.36)	0.45 ^a
Psychiatrist visit	351 (10.3)	319 (9.4)	0.19 ^b
Antidepressant medication	666 (19.6)	630 (18.5)	0.27 ^b
Expenditures in preceding 2 y, mean (SD), \$	38,562 (109,561)	38,743 (106,390)	0.95 ^a
Eligibility time, mean (SD), d	1,680 (830)	1,744 (881)	0.002 ^a
Benefit design, n (%)			0.23 ^b
Exclusive provider organization	158 (4.7)	165 (4.9)	—
Health maintenance organization	1,671 (49.2)	1,748 (51.4)	—
Indemnity	1,47 (4.3)	119 (3.5)	—
Other	395 (11.6)	357 (10.5)	—
Point of service	830 (24.4)	815 (24.0)	—
Preferred provider organization	199 (5.9)	196 (5.8)	—
Setting of incident diagnosis, n (%)			0.007 ^b
Hospital	626 (18.4)	657 (19.3)	—
Skilled nursing facility	41 (1.2)	73 (2.2)	—
ER	1,526 (44.9)	1,536 (45.2)	—
Clinic	1,207 (35.5)	1,134 (33.4)	—
Year of incident diagnosis, n (%)			0.57 ^b
2003	172 (5.1)	142 (4.2)	—
2004	221 (6.5)	194 (5.7)	—
2005	462 (13.6)	494 (14.5)	—
2006	268 (7.9)	268 (7.9)	—
2007	257 (7.6)	275 (8.1)	—
2008	345 (10.2)	355 (10.4)	—
2009	348 (10.2)	343 (10.1)	—
2010	415 (12.2)	387 (11.4)	—
2011	343 (10.1)	338 (9.9)	—
2012	359 (10.6)	378 (11.1)	—
2013	210 (6.2)	226 (6.7)	—

Abbreviations: COPD = chronic obstructive pulmonary disease; ER = emergency room.

^a By t test.^b By χ^2 test.

Figure 2 Health care use for epilepsy and costs of care



Predicted (A) epilepsy-related emergency room (ER) visits, (B) epilepsy-related hospitalizations, (C) epilepsy medical costs, (D) nonepilepsy medical costs, and (E) antiepileptic drugs costs each year for cases compared to controls with 95% confidence intervals. USD = US dollars.

differences between cohorts in probability of presurgical evaluation, intracranial electrode implantation, use of VNS, or augmentation of antiepileptic medication after an ER visit for seizure.

Quality of care: Attention to epilepsy comorbid conditions and women of childbearing age

Cases had a higher rate of bone health evaluation, as measured by vitamin D testing and dual-energy x-ray absorptiometry scan (predicted probabilities year 1: 6.1% vs 4.0%; year 5: 10.3% vs 8.3%, $p < 0.001$). Antidepressant prescription did not differ between cohorts; however, cases more commonly had psychiatrist visits, particularly in the first year (year 1: 4.7% vs 2.6%, $p < 0.001$). Valproic acid prescription for women of childbearing age did not differ between cohorts.

Quality of care: Antiepileptic drug use patterns

Cases were more likely to be continued on antiepileptic medication after initial prescription compared to controls (predicted probabilities year 1: 94.3% vs 84.4%, $p < 0.001$); this proportion declined over time for both groups (year 5: 65.1% vs 47.0%, $p < 0.001$). Among patients who were prescribed antiepileptic drugs, cases had higher total days of antiepileptic medication supply than controls ($p < 0.001$). Prescription of generation 2 antiepileptic medications was more likely for cases than controls ($p < 0.001$).

Measures of disease severity and trauma/injury

Status epilepticus and intractable epilepsy were more likely for cases than controls ($p < 0.001$). Presentations to the ER for trauma or injury were similar between cohorts.

Table 4 Outcomes for neurologist and no neurologist cohorts (predicted probabilities)

Outcome	Neurologist cases	No neurologist controls	p Value
Evaluation by EEG, n (%)			<0.001 ^a
Year 1	78.6	7.2	—
Year 2	8.9	0.9	—
Year 3	6.6	1.1	—
Year 4	5.9	0.5	—
Year 5	4.9	0.2	—
Presurgical evaluation, n (%)			0.11 ^a
Year 1	0.05	0.02	—
Year 2	0.03	<0.01	—
Year 3	0.04	<0.01	—
Year 4	0.03	0.01	—
Year 5	—	0.02	—
Epilepsy surgery, n (%)			0.001 ^a
Year 1	0.2	0.05	—
Year 2	0.01	—	—
Year 3	0.01	—	—
Year 4	0.01	0.01	—
Year 5	0.01	—	—
Intracranial electrode implantation, n (%)			—
Year 1	0.1	—	—
Year 2	—	—	—
Year 3	0.1	—	—
Year 4	0.2	—	—
Year 5	—	—	—
VNS, n (%)			0.42 ^a
Year 1	<0.01	<0.01	—
Year 2	<0.01	<0.01	—
Year 3	<0.01	<0.01	—
Year 4	<0.01	<0.01	—
Year 5	<0.01	<0.01	—
AED augmentation after ER visit, n (%)			0.78 ^a
Year 1	43.5	45.0	
Year 2	34.9	13.4	
Year 3	31.9	10.2	
Year 4	35.2	16.6	

Table 4 Outcomes for neurologist and no neurologist cohorts (predicted probabilities) (continued)

Outcome	Neurologist cases	No neurologist controls	p Value
Year 5			24.3 10.6
Bone health evaluation, n (%)			<0.001 ^a
Year 1	6.1	4.0	
Year 2	7.7	4.9	
Year 3	7.9	5.7	
Year 4	10.2	6.8	
Year 5	10.3	8.3	
Antidepressant prescription, n (%)			0.69 ^a
Year 1	11.3	11.6	
Year 2	12.8	13.4	
Year 3	15.8	15.6	
Year 4	20.7	21.2	
Year 5	24.0	24.2	
Psychiatrist visit, n (%)			<0.001 ^a
Year 1	4.7	2.6	
Year 2	2.6	1.9	
Year 3	2.6	1.6	
Year 4	2.3	1.6	
Year 5	2.0	2.0	
Valproic acid for women 18–44 y, n (%)			0.54 ^a
Year 1	0.6	0.5	
Year 2	0.3	0.3	
Year 3	0.3	0.1	
Year 4	0.5	0.1	
Year 5	0.3	0.04	
Any AED prescription, n (%)			<0.001 ^a
Year 1	94.3	84.4	—
Year 2	82.9	75.9	—
Year 3	71.9	56.5	—
Year 4	68.6	49.7	—
Year 5	65.1	47.0	—
Total AED pill-d			<0.001 ^a
Year 1	164.7	114.5	—
Year 2	174.0	131.9	—

Continued

Table 4 Outcomes for neurologist and no neurologist cohorts (predicted probabilities) (continued)

Outcome	Neurologist cases	No neurologist controls	p Value
Year 3	170.8	128.8	—
Year 4	166.3	122.0	—
Year 5	161.6	118.1	—
Generation 2 AED prescription, n (%)			<0.001 ^a
Year 1	88.8	60.2	—
Year 2	70.7	50.0	—
Year 3	54.5	32.3	—
Year 4	53.0	26.3	—
Year 5	52.8	21.3	—
Status epilepticus, n (%)			<0.001 ^a
Year 1	1.1	0.2	—
Year 2	0.2	0.03	—
Year 3	0.2	0.03	—
Year 4	0.2	0.04	—
Year 5	0.2	0.02	—
Intractable epilepsy, n (%)			<0.001 ^a
Year 1	4.1	0.09	—
Year 2	0.8	0.02	—
Year 3	0.6	0.02	—
Year 4	0.5	0.02	—
Year 5	0.4	0.03	—
ER evaluation for trauma/injury, n (%)			0.50 ^a
Year 1	7.6	7.2	—
Year 2	6.6	7.1	—
Year 3	7.1	5.8	—
Year 4	6.5	6.4	—
Year 5	6.8	5.5	—

Abbreviations: AED = antiepileptic drug; ER = emergency room; VNS = vagal nerve stimulation.

^a By χ^2 test.

Sensitivity analyses

Alterations in the definition of our population, both those more restrictive and those less restrictive (including epilepsy prevalence), led to no significant changes in our primary outcomes. The requirement of 2 epilepsy diagnoses resulted in higher rates of care use overall and narrowed the magnitude of difference between patient cohorts. Total epilepsy-related

costs remained greater for cases; however, total non–epilepsy-related costs were no longer significantly greater for cases than controls (figure 3).

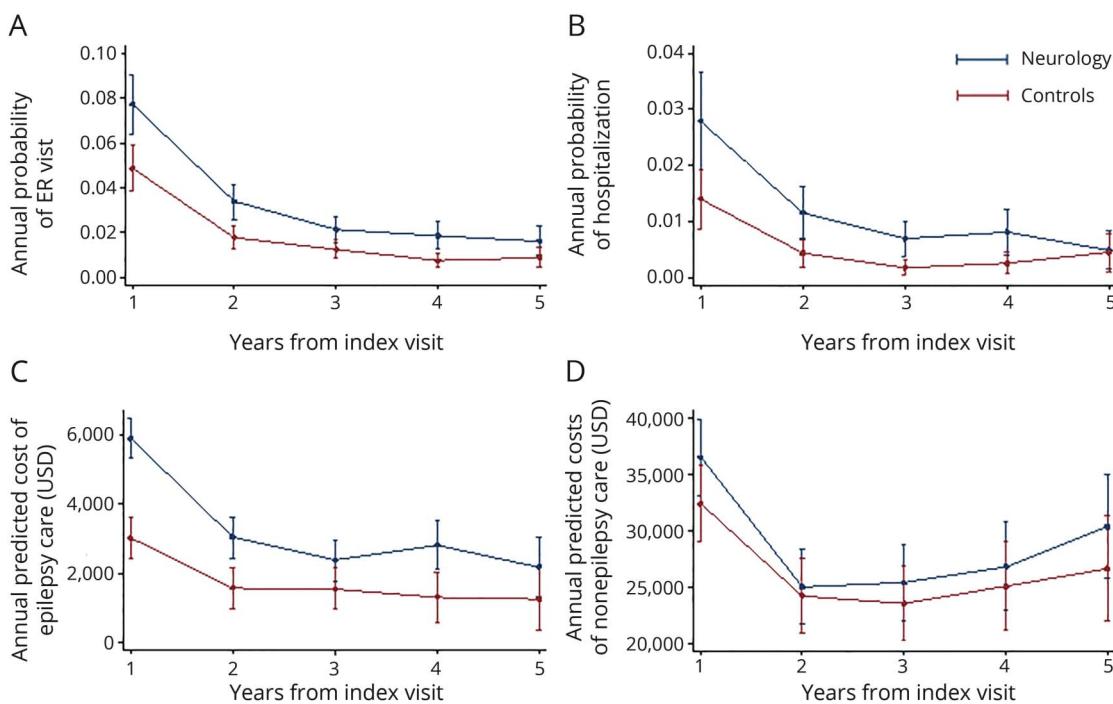
Discussion

In a large cohort of privately insured patients, we found that patients with incident epilepsy who visited a neurologist had higher subsequent epilepsy-related ER visit and hospitalization rates compared to patients who did not see a neurologist. This difference was greatest in the first year and decreased over time. A similar pattern was observed for epilepsy and nonepilepsy medical costs, with higher costs for patients who visited a neurologist, most notably in the first year.

There are multiple ways to interpret the finding of higher epilepsy-related health care use and cost among cases than controls. The first is that neurologist involvement directly leads to increased ER visits and hospitalizations with their associated costs. A prior investigation of patients with acute stroke found that neurologist management was associated with more extensive testing (however, these patients also had improved outcomes)¹⁷; it is a possibility that neurologist management of patients with incident epilepsy increases costs without improving all clinical outcomes. A second explanation is that patients referred to a neurologist have more severe disease or differ from nonreferred patients in other ways that increase use and cannot be measured. While we adjusted for a comprehensive set of variables through application of a 43-factor propensity score, adjustment for remaining dissimilarities, and consideration of lag time events, there are undoubtedly unmeasured differences. Residual confounding is most strongly suggested by our observation that non–epilepsy medical costs, an outcome that should be largely outside of neurologist influence, were also higher for cases than controls. It is conceivable that neurologist involvement encourages patients to seek additional health care directly through referral to other specialists and testing and/or indirectly through resultant proximity to care facilities and providers. However, it is more likely that patients with complicated illness are directed to neurologists for care.

Additional findings further support the notion of confounding by disease severity. The most compelling argument is that when we applied a more stringent disease definition (2 epilepsy diagnostic codes) in sensitivity analysis, the difference in epilepsy care use and cost between cohorts contracted, and nonepilepsy care costs became comparable. Several other factors also suggest that the case cohort is a more medically complex population (although these may have alternative explanations). Considerable comorbidity burden has been described for patients with incident epilepsy, most notably concomitant neurologic disease and psychiatric disorders,^{2,18} and we found more frequent psychiatrist visits for cases than controls. Characterization of antiepileptic use revealed both a greater proportion of the cases continued on antiepileptic

Figure 3 Health care use for epilepsy and medical costs: sensitivity analysis



For the population defined with the requirement of 2 epilepsy diagnosis codes, predicted (A) epilepsy-related emergency room (ER) visits, (B) epilepsy-related hospitalizations, (C) epilepsy medical costs, and (D) nonepilepsy medical costs each year for cases compared to controls with 95% confidence intervals. USD = US dollars.

drugs after initial diagnosis and higher total days of drug supply each year, suggestive of recurrent or refractory seizures. Both status epilepticus and intractable epilepsy were more common within the case cohort. Lastly, nearly every measurable indicator of comorbidity was higher among cases, which cautions that this cohort is likely sicker in unmeasured ways as well. Both of these factors are potential confounders by disease severity that are difficult to address with claims data.

The results of our study suggest that even rigorously adjusted claims data are likely inadequate to assess the value of care in epilepsy. The importance and difficulty of proper risk adjustment have been highlighted in recent investigations of claims data for stroke patients and patients with headache.^{12,19} Several epilepsy-specific considerations make measurement of care quality particularly elusive for this patient population. Classic measures of disease severity such as epilepsy etiology, seizure frequency, and seizure type²⁰ are not readily abstracted from insurance claims. Expenditures are known to be concentrated in patients with drug-resistant epilepsy,^{5,21} yet claims-based identification of this subpopulation is challenging.²² Epilepsy is a chronic condition, typically managed incrementally over a patient's lifetime, and therefore difficult to capture in a several-year study window. Even selection of clinical outcomes is not straightforward. While patients, providers, and payers generally prefer to avoid hospitalizations, not all hospitalizations for epilepsy are adverse

outcomes. Frequently, patients are electively admitted to the hospital to clarify their diagnosis or to receive a therapeutic intervention such as epilepsy surgery. Some important patient outcomes for epilepsy management such as seizure freedom and antiepileptic drug tolerability cannot be assessed through claims data. Moreover, certain features of the illness are nearly impossible to disentangle for interpretation. Does a diagnosis of intractable epilepsy indicate severe underlying disease, insufficient management, or differences in coding behaviors? Does therapy escalation indicate physician responsiveness, severe underlying disease, or both?

Nonetheless, closer examination of how epilepsy-related expenditures differed between cases and controls suggests some potential benefits of neurologist care and additional opportunities to measure quality and value. Recent studies have proposed that antiepileptic medications are a major driver of epilepsy-related spending, accounting for 8% to 77% of direct costs.²¹ We observed that antiepileptic medication costs were substantially higher for cases than controls; however, there may be unmeasured benefit from such spending. Cases were more commonly prescribed generation 2 antiepileptic drugs. While newer medications are pricier than older drugs and may not be more effective for seizures,²³ it is generally accepted that newer medications are better tolerated and associated with decreased morbidity.²⁴ Antiepileptic medication typically should not be discontinued until several years of seizure freedom have elapsed,²⁵ and recent studies

have shown that retention rates for certain generation 2 antiepileptic drugs are greater than those for certain older medications.^{26,27} In our study, continued treatment with an antiepileptic drug was more common for cases than controls. While beyond the scope of this study, identification of drug-resistant patients and quantification of indirect costs should be included in future investigations. We also found that cases more frequently underwent diagnostic evaluation and surgery; therefore, neurologists appear to be important facilitators of targeted therapies such as epilepsy surgery. However, the benefits of surgery may not be realized for several years.²⁸ Other holistic care aspects, including cases receiving bone health evaluation more frequently than controls and a greater proportion of cases being seen by psychiatrists, could be further studied to determine whether fracture risk or treatment for mood disorders is more favorable for cases than controls. Additional morbidity associated with long-term use of certain antiepileptic drugs such as metabolic syndrome, neuropathy, cerebellar atrophy, gingival hyperplasia, weight gain, hepatotoxicity, and anemia could be studied with a longer observation period. Furthermore, consideration of use and cost outcomes over a longer follow-up period is particularly important because the tendency for patients with epilepsy to be eligible for disability and therefore receive insurance through Medicare or Medicaid is high. A reduction in long-term medical costs and even disability payments for this population could have dramatic effects on future federal and state expenditures.

Measurement of patient-centered metrics is a critical component of comprehensive assessment of the value of neurologist care. A randomized controlled trial would be optimal for comparing neurologist care to no neurologist care and could allow collection of rich, multidimensional clinical data. However, the ethical standing of randomizing a patient with complex, highly morbid, treatable disease to nonexpert care is questionable. Another potentially more feasible way to answer this research question is through the linking of claims data to patient registry data, which ideally would include patient-centered metrics such as quality of life, seizure freedom, mood, and return to work.

The typical challenges associated with administrative claims data apply to this study. We analyzed data collected primarily for billing purposes and therefore lacked fine clinical resolution. In addition, there are likely undescribed differences in how neurologists code encounters compared to non-neurologists. Retrospectively identifying patients with epilepsy is difficult; patients may have seizures from conditions other than epilepsy (e.g., hypoglycemia or alcohol withdrawal) or may have convulsive episodes due to alternative pathophysiology (e.g., psychogenic nonepileptic spells or syncope). Claims data have limited clinical details to allow the determination of correct alternative diagnosis. In this way, we may have failed to measure additional benefit of neurologist care. Selection bias due to referral is inherent in any valuation of specialist care; however, an extensive collection of potential

confounding factors was considered through application of propensity score matching. In fact, our risk adjustment strategy was substantially more robust and comprehensive than the current proposals to measure physician value, which would also use administrative data. In addition, control patients were not required to have seen a nonneurologist provider for epilepsy care after incident diagnosis, while case patients were required to see a neurologist for epilepsy care by inclusion criterion. However, a substantial proportion of controls received epilepsy-related care in the 5 years after the index date: 83.5% of controls in year 1 (compared to 94.3% of cases) and 24.7% of controls in year 5 (compared to 31% of cases). As noted, using hospitalization as an outcome is complicated because not all hospitalizations for epilepsy are adverse outcomes; some hospitalizations may represent an appropriate care escalation. Several important measures relating to women of childbearing age such as folic acid prescription, contraception use, and pregnancy complications could not be evaluated because of limitations in claims data. Lastly, the generalizability of our findings may be limited by the use of data from a single, private insurer and by the small percentage of patients who met study criteria.

ER visits, hospitalizations, and medical costs were greater for patients with incident epilepsy who visited a neurologist compared to those who did not. However, our results suggest that this comparison may be confounded by other factors such as disease severity and that claims data have limited ability to characterize the value of epilepsy care provided by physicians. This study has important ramifications for neurologists and other specialists in an era of quality payment programs and alternative payment model development. New payment models should be cautious with penalties until better risk adjustment strategies are available or linkage of patient registry data becomes possible.

Author contributions

Dr. Hill and Dr. Lin contributed to study design, data analysis and interpretation, and manuscript drafting and revision. Dr. Burke and Dr. Kerber contributed to study design and conceptualization, data analysis and interpretation, and manuscript revision. Dr. Skolarus, Dr. Esper, and Mr. Magliocco contributed to study conceptualization and manuscript revision. Dr. Callaghan contributed to study design and conceptualization, data analysis and interpretation, and manuscript revision.

Study funding

Dr. Hill is supported by the Mirowski Family Fund and the Family of Jonathan Rothberg. Dr. Burke is supported by NIH K08 NS082597 and R01 MD008879. Dr. Kerber is supported by NIH/National Center for Research Resources K23 RR024009, Agency for Healthcare Research and Quality (AHRQ) R18 HS017690, NIH/National Institute on Deafness and Other Communication Disorders R01 DC012760, and AHRQ R18 HS022258. Dr. Skolarus is supported by NIH R01 MD008879, U01 AG032947, and R01 MD011516.

Dr. Callaghan is supported by NIH K23 NS079417 and Veterans Affairs Clinical Science Research and Development Merit CX001504. The study was proposed and executed by the American Academy of Neurology Health Services Research Subcommittee. The study was funded by the American Academy of Neurology.

Disclosure

C. Hill, C. Lin, J. Burke, and K. Kerber report no disclosures relevant to the manuscript. L. Skolarus performed consulting for Bracket Global. G. Esper is a consultant for NeuroOne, Inc. He performs medicolegal consultations. B. Magliocco reports no disclosures relevant to the manuscript. B. Callaghan receives research support from Impeto Medical Inc. He performs medical consultations for Advance Medical, consults for a Patient-Centered Outcomes Research Institutegrant, consults for the immune tolerance network, and performs medical legal consultations. Go to Neurology.org/N for full disclosures.

Publication history

Received by *Neurology* July 22, 2018. Accepted in final form October 25, 2018.

References

1. Gorelick PB. Adaptation of neurological practice and policy to a changing US health-care landscape. *Lancet Neurol* 2016;15:444–450.
2. England MJ, Liverman CT, Schultz AM, Strawbridge LM; Institute of Medicine Committee on Public Health Dimensions of the Epilepsies. *Epilepsy Across the Spectrum: Promoting Health and Understanding*. Washington, DC: National Academies Press; 2012.
3. Gooch CL, Pracht E, Borenstein AR. The burden of neurological disease in the United States: a summary report and call to action. *Ann Neurol* 2017;81:479–484.
4. Begley CE, Durgin TL. The direct cost of epilepsy in the United States: a systematic review of estimates. *Epilepsia* 2015;56:1376–1387.
5. Begley CE, Famulari M, Annegers JF, et al. The cost of epilepsy in the United States: an estimate from population-based clinical and survey data. *Epilepsia* 2000;41:342–351.
6. Fisher RS, Acevedo C, Arzimanoglou A, et al. ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* 2014;55:475–482.
7. Labiner DM, Bagic AI, Herman ST, Fountain NB, Walczak TS, Gumnit RJ. Essential services, personnel, and facilities in specialized epilepsy centers: revised 2010 guidelines. *Epilepsia* 2010;51:2322–2333.
8. Desforges JF, Scheuer ML, Pedley TA. The evaluation and treatment of seizures. *N Engl J Med* 1990;323:1468–1474.
9. Ney JP, Johnson B, Knabel T, Craft K, Kaufman J. Neurologist ambulatory care, health care utilization, and costs in a large commercial dataset. *Neurology* 2016;86:367–374.
10. Clough JD, McClellan M. Implementing MACRA. *JAMA* 2016;315:2397.
11. Mitchell JB, Ballard DJ, Whisnant JP, Ammering CJ, Samsa GP, Matchar DB. What role do neurologists play in determining the costs and outcomes of stroke patients? *Stroke* 1996;27:1937–1943.
12. Callaghan BC, Burke JF, Kerber KA, et al. The association of neurologists with headache health care utilization and costs. *Neurology* 2018;90:e525–e533.
13. Ney JP, Sico JJ, Klein BC, Magliocco B, Callaghan BC, Esper GJ. Association of neurologist care with headache expenditures: a population-based, longitudinal analysis. *Cephalgia* 2018;38:1876–1884.
14. Holden EW, Grossman E, Nguyen HT, et al. Developing a computer algorithm to identify epilepsy cases in managed care organizations. *Dis Manag* 2005;8:1–14.
15. Moura LM, Price M, Cole AJ, Hoch DB, Hsu J. Accuracy of claims-based algorithms for epilepsy research: revealing the unseen performance of claims-based studies. *Epilepsia* 2017;58:683–691.
16. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757–763.
17. Goldstein LB, Matchar DB, Hoff-Lindquist J, Samsa GP, Horner RD. VA Stroke Study: neurologist care is associated with increased testing but improved outcomes. *Neurology* 2003;61:792–796.
18. Martin RC, Faught E, Richman J, et al. Psychiatric and neurologic risk factors for incident cases of new-onset epilepsy in older adults: data from U.S. Medicare beneficiaries. *Epilepsia* 2014;55:1120–1127.
19. Schwartz J, Wang Y, Qin L, et al. Incorporating stroke severity into hospital measures of 30-day mortality after ischemic stroke hospitalization. *Stroke* 2017;48:3101–3107.
20. Scheffer IE, Berkovic S, Capovilla G, et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* 2017;58:512–521.
21. Strzelczyk A, Reese JP, Dodel R, Hamer HM. Cost of epilepsy: a systematic review. *Pharmacoeconomics* 2008;26:463–476.
22. Burneo JG, Shariff SZ, Liu K, Leonard S, Saposnik G, Garg AX. Disparities in surgery among patients with intractable epilepsy in a universal health system. *Neurology* 2016;86:72–78.
23. Beghi E, Atzeni L, Garattini L. Economic analysis of newer antiepileptic drugs. *CNS Drugs* 2008;22:861–875.
24. French JA, Gazzola DM. New generation antiepileptic drugs: what do they offer in terms of improved tolerability and safety? *Ther Adv Drug Saf* 2011;2:141–158.
25. Camfield P, Camfield C. When is it safe to discontinue AED treatment? *Epilepsia* 2008;49:25–28.
26. Arif H, Buchsbaum R, Piero J, et al. Comparative effectiveness of 10 antiepileptic drugs in older adults with epilepsy. *Arch Neurol* 2010;67:408–415.
27. Zeng QY, Fan TT, Zhu P, et al. Comparative long-term effectiveness of a monotherapy with five antiepileptic drugs for focal epilepsy in adult patients: a prospective cohort study. *PLoS One* 2015;10:e0131566.
28. Picot MC, Jaussent A, Neveu D, et al. Cost-effectiveness analysis of epilepsy surgery in a controlled cohort of adult patients with intractable partial epilepsy: a 5-year follow-up study. *Epilepsia* 2016;57:1669–1679.

 Open Access Full Text Article

ORIGINAL RESEARCH

Hand grip strength and chronic obstructive pulmonary disease in Korea: an analysis in KNHANES VI

Su Hwan Lee
 Soo Jung Kim
 Yeji Han
 Yon Ju Ryu
 Jin Hwa Lee
 Jung Hyun Chang

Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Ewha Medical Research Institute, Ewha Womans University School of Medicine, Seoul, Korea

Background: Muscle mass is known to be associated with mortality in elderly adults. Because hand grip strength (HGS) is known as a simple assessment tool for muscular strength, many researchers have studied the association between HGS and disease. However, empirical evidence for the relationship between chronic obstructive pulmonary disease (COPD) and HGS is still controversial. The aim of this study was to evaluate the association between COPD and HGS, using Korean population data.

Methods: This was a population-based cross-sectional study. Data were obtained from the sixth Korean National Health and Nutrition Examination Survey, which was conducted from 2013 to 2015. To reduce the effects of HGS-related factors and potential confounding factors, propensity score matching was used to match subjects with and without COPD.

Results: Among 14,930 subjects, 832 were enrolled in each group (non-COPD and COPD) after propensity score matching. COPD subjects did not have lower HGS than non-COPD subjects (non-COPD vs COPD, male, 38.0 ± 7.0 vs 38.9 ± 7.0 kg, $P=0.044$, female, 23.8 ± 4.6 vs 24.2 ± 4.9 kg, $P=0.342$). Lung function was classified by Global Initiative for Chronic Obstructive Lung Disease stages and was not significantly associated with HGS. For male COPD subjects, there was a significant correlation between HGS and the EuroQol Five-Dimension Questionnaire (EQ5D) utility score index, which is an indicator of quality of life that adjusts for age and body mass index ($r=0.201$, $P<0.001$). The correlation was absent for female subjects ($r=0.098$, $P=0.170$).

Conclusion: COPD subjects did not have lower HGS than non-COPD subjects. HGS did not associate with lung function. However, the HGS of male COPD subjects was positively associated with EQ5D utility score index, an indicator of quality of life. HGS may be helpful as an additional method to the evaluation of quality of life in male COPD patients.

Keywords: chronic obstructive pulmonary disease, hand strength, quality of life

Introduction

Chronic obstructive pulmonary disease (COPD) plays a major role in global morbidity and is expected to be a major cause of death worldwide in 2030.¹ Progression of COPD often involves worsening dyspnea, decreased quality of life, hospitalization, need for medical resources and increased risk of mortality.² Therefore, it is important to evaluate COPD status for COPD management. The use of forced expiratory volume in 1 second (FEV₁) has been the severity assessment of choice for COPD; however, reliance on this method has decreased as several other assessment methods have been introduced.^{1,3,4}

Cardiorespiratory fitness and muscular strength are associated with mortality and morbidity in several diseases.^{5,6} As an indicator of muscle strength, hand grip strength (HGS), which is a simple measure of upper limb muscle function, has been linked to

Correspondence: Jung Hyun Chang
 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Ewha Medical Research Institute, Ewha Womans University School of Medicine, 1071 Anyangcheon-ro, Yangcheon-gu, Seoul 07985, Korea
 Tel +82 2 2650 5686
 Fax +82 2 2650 2559
 Email hs1017@ewha.ac.kr

mortality from cardiovascular disease in a large longitudinal population study.⁷ Various comorbidities such as hypertension, coronary artery occlusive disease, stroke and COPD have been associated with low HGS.⁷ Although physical activity is known to be a predictor of mortality in COPD,^{8,9} the association of COPD with HGS is controversial. Some studies have reported that HGS in COPD is unrelated to hospitalization.⁷ In two studies, muscle strength showed no correlation with FEV₁ in COPD patients, and muscle strength in these patients was comparable to that of healthy subjects.^{10,11} In contrast, other studies have shown that HGS in COPD is associated with mortality and that subjects with moderate to severe COPD had lower HGS than did healthy subjects.^{12,13}

The aim of this study was to evaluate the association between COPD and HGS, after adjusting for comorbidities known to be associated with HGS, using population-based data from the Korean National Health and Nutrition Examination Survey (KNHANES VI), which was conducted from 2013 to 2015.

Materials and methods

Study design and populations

This used a population-based cross-sectional design. The data were obtained from KNHANES VI, which are nationwide cross-sectional surveys that evaluate the health and nutrition status of Korean populations from 2013 to 2015. Subjects analyzed in this study were aged >40 years who completed the questionnaire and performed spirometry between 2014 and 2015. KNHANES VI data contained detailed information on demographics; smoking status; physician-diagnosed comorbidities such as hypertension, stroke, ischemic heart disease and diabetes mellitus (DM); activity limitations; lung function and the EuroQol Five-Dimension Questionnaire (EQ5D) for health-related quality of life.

Definition

Smoking status was defined as current smoker, ex-smoker or never smoker (smoked <100 cigarettes during the lifetime). The several comorbidities included only those diagnosed by physicians. Anemia was defined as <13 g/dL hemoglobin for men and <12 g/dL for women.

Spirometry

Pulmonary function tests were conducted by trained medical technicians according to the manual of the American Thoracic Society/European Respiratory Society (ATS/ERS) Task Force, using dry rolling seal spiroimeters (Model 2130; Sensor Medics, Yorba Linda, CA, USA).¹⁴ COPD was

defined when FEV₁ divided by forced vital capacity (FVC) was <0.7, in accordance with Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines.¹ The severity of lung function was classified according to the percentage of predicted FEV₁. Subjects with FEV₁ ≥80% were classified as GOLD stage 1, those with 50% ≤ FEV₁ <80% as GOLD stage 2, those with 30% ≤ FEV₁ <50% as GOLD stage 3 and those with FEV₁ <30% as GOLD stage 4. Subjects with an FEV₁/FVC ≥0.7 were identified as non-COPD.

HGS

HGS was measured three times in each hand using a digital grip strength dynamometer (TKK 5401; Takei Scientific Instruments Co., Ltd., Tokyo, Japan). Trained medical technicians instructed the seated subjects to hold the dynamometer with the second finger nodes of the working hand at 90° to the handle and to squeeze the handle as firmly as they could. After subjects slowly stood up, HGS was measured during expiration. A 60-second rest period was given after each HGS measure. The HGS used in the analysis was the highest of the six measured values.¹⁵

The EQ5D

EQ5D was developed by the EuroQol Group to evaluate multidimensional health-related quality of life.¹⁶ The EQ5D includes a descriptive section and a valuation section; however, KNHANES VI has only been conducted with the descriptive section. The descriptive section includes mobility, self-care, usual activities, pain/discomfort and anxiety/depression, and each is assessed according to three functional levels: no problems, some problems or extreme problems. The functional level was converted into an EQ5D utility score index using a specific Korean valuation set developed by the time trade-off protocol at the Korean Centers for Disease Control and Prevention.¹⁷ We analyzed the EQ5D utility score index and descriptive sections for COPD subjects.¹⁸

Statistical analysis

All continuous values are described as mean ± standard deviation, and categorical values are reported as absolute numbers and percentages. The Student's *t*-test and one-way analysis of variance were used for analyzing continuous values. Categorical values were analyzed using the χ^2 test or Fisher's exact test. To reduce the effect of HGS-related factors and potential confounding factors, propensity score matching was used to match COPD and non-COPD subjects. The factors of age, sex, body mass index (BMI), cardiovascular disease (including hypertension, dyslipidemia and ischemic

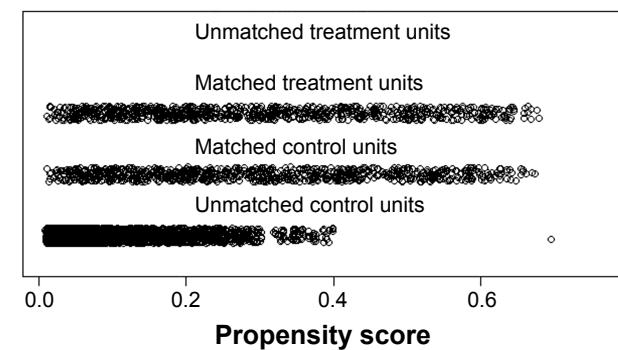


Figure 1 Distribution of propensity score.

heart disease), stroke, DM and depression were considered during the process of propensity score matching (Figure 1). The Kruskal–Wallis test was used to investigate associations among HGS, EQ5D and lung function in COPD subjects. To assess the association between HGS and EQ5D, a partial correlation analysis was performed to adjust for age and BMI. In all comparisons, a *P*-value of <0.05 was considered statistically significant. SPSS version 20 (IBM Corporation, Armonk, NY, USA) was used for the statistical analysis; propensity score matching was performed using the R program version 3.3.3 for Windows.

Ethics statement

The KNHANES VI was conducted by the Korea Center for Disease Control and Prevention (KCDC). All survey protocols were approved by the KCDC Institutional Review Board (approval numbers 2013-07CON-03-4CP, 2013-12EXP-03-5C and 2015-01-02-6C), and participants provided informed consent before participating in the study, which was conducted in accordance with the ethical principles of the Declaration of Helsinki. All data of KNHANES VI are coded, publicly available and freely available.

Results

Baseline characteristics of the subjects

Among the 14,930 subjects, the 5,857 who completed the questionnaire and performed a spirometry adequately were included in this study. Among these, 4,984 (85.1%) were classified as non-COPD and 873 (14.9%) had COPD. After propensity score matching, 832 subjects were enrolled in each group. Data for all subjects in both groups were analyzed, except those for whom results for HGS or EQ5D were missing (Figure 2).

Table 1 shows a comparison of the non-COPD and COPD groups. Before propensity score matching, the baseline characteristics of the two groups were very different. The 873 (14.9%) COPD subjects were older and had a higher

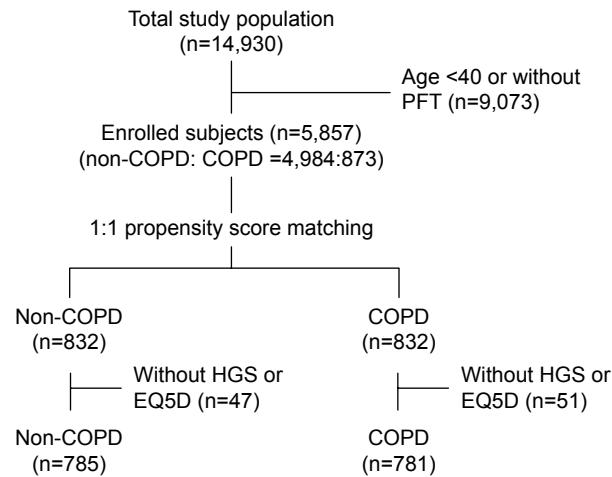


Figure 2 Flow diagram for the study.

Abbreviations: COPD, chronic obstructive pulmonary disease; EQ5D, EuroQol Five-Dimension Questionnaire; HGS, hand grip strength; PFT, pulmonary function test.

proportion of males than did the non-COPD group. COPD subjects also had higher proportions of various comorbidities than did the non-COPD group. Among the subjects with HGS data, male COPD subjects showed lower HGS than did male non-COPD subjects (non-COPD vs COPD 41.7 ± 7.2 vs 38.9 ± 6.9 kg, $P < 0.001$). A similar pattern of results was observed for female subjects (non-COPD vs COPD 25.7 ± 4.6 vs 24.2 ± 4.9 kg, $P < 0.001$). After propensity score matching, no significant statistical differences remained, except for lung function and smoking status.

Comparison of HGS in non-COPD and COPD groups after propensity score matching

There was no difference in HGS between non-COPD and COPD groups (non-COPD vs COPD, 34.7 ± 8.9 vs 35.2 ± 8.8 kg, $P = 0.266$). KNHANES VI data were analyzed by dividing into male and female because each sex had been shown to have a significant difference of baseline HGS in previous studies.^{7,15} For males, there was a statistically significant difference in HGS between non-COPD and COPD subjects, with COPD subjects scoring slightly higher (non-COPD vs COPD, 38.0 ± 7.0 vs 38.9 ± 7.0 kg, $P = 0.044$). However, there was no such difference among females (non-COPD vs COPD 23.8 ± 4.6 vs 24.2 ± 4.9 kg, $P = 0.342$) (Table 2).

Figure 3 shows the associations among HGS, EQ5D and lung function in subjects with COPD. Lung function classified as GOLD stages was not significantly associated with either HGS or with EQ5D utility score in either sex. However, there was a significant correlation between HGS and EQ5D utility score in male subjects after adjusting for

Table 1 Subject demographics before and after propensity score matching

Variables	Overall series			Propensity score matching		
	Non-COPD n=4,984	COPD n=873	P-value	Non-COPD n=832	COPD n=832	P-value
Age, years	56.3±10.3	65.3±9.1	<0.001	64.9±9.3	65.4±9.1	0.308
Male sex	1,929 (38.7)	644 (73.8)	<0.001	620 (74.5)	606 (72.8)	0.436
Height, cm	160.8±8.7	164.0±8.58	<0.001	163.2±8.3	163.9±8.62	0.136
BMI, kg/m ²	24.2±3.1	24.0±2.8	0.053	23.9±2.8	24.0±2.8	0.424
HTN	1,312 (26.6)	399 (39.2)	<0.001	330 (39.7)	338 (40.6)	0.689
Dyslipidemia	956 (19.4)	188 (21.7)	0.256	178 (21.4)	188 (22.6)	0.554
Stroke	92 (1.9)	32 (3.7)	<0.002	29 (3.5)	32 (3.8)	0.696
IHD	125 (2.6)	42 (5)	<0.001	34 (4.1)	42 (5)	0.348
DM	464 (9.4)	133 (15.4)	<0.001	137 (16.5)	132 (15.9)	0.739
Arthritis	792 (16.8)	145 (17.4)	0.643	154 (18.5)	145 (17.4)	0.566
Depression	282 (5.7)	28 (3.2)	0.008	35 (4.2)	28 (3.4)	0.369
Self-reported functional limitation	379 (7.7)	81 (9.4)	0.208	79 (9.5)	81 (9.7)	0.833
Ever smoker, n (%)	1,756 (35.6)	581 (67.2)	<0.001	494 (59.4)	551 (66.2)	0.004
FEV ₁ , % of predicted value	95.1±11.9	79.6±15.5	<0.001	96.0±12.6	79.7±15.6	<0.001
GOLD stage I		444 (50.9)			425 (51.1)	
GOLD stage II		394 (45.1)			374 (45.0)	
GOLD stage III		35 (4.0)			33 (4.0)	

Note: Data represented as mean ± SD or n (%).

Abbreviations: BMI, body mass index; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus; FEV₁, forced expiratory volume in 1 second; GOLD, Global Initiative for Chronic Obstructive Lung Disease; HTN, hypertension; IHD, ischemic heart disease; SD, standard deviation.

age and BMI ($r=0.201, P<0.001$), while female subjects did not show the correlation ($r=0.098, P=0.170$).

Relationship between HGS and EQ5D in male subjects with COPD

Since HGS and the EQ5D utility score index were correlated after adjusting for age and BMI in male COPD subjects,

further analyses of the relationship between HGS and EQ5D were performed. Figure 4 shows the statistical significance of analysis EQ5D and male subjects with COPD groups classified as HGS interquartile range. The highest HGS group had higher EQ5D utility scores than did the lowest HGS group ($P<0.001$). Analyses of groups according to HGS interquartile range and the descriptive section of the EQ5D

Table 2 Comparison of clinical characteristics between non-COPD and COPD subjects

Variables	Male (n=1,179)			Female (n=387)		
	Non-COPD n=597	COPD n=582	P-value	Non-COPD n=188	COPD n=199	P-value
Age, years	64.5±9.3	65.0±9.1	0.325	66.0±9.3	65.6±9.0	0.653
Age category, years						
40–49	50 (8.4)	42 (7.2)	0.554	12 (6.4)	10 (5.0)	0.818
50–59	124 (20.8)	110 (18.9)		30 (16)	38 (19.1)	
60–69	248 (41.5)	240 (41.2)		78 (41.5)	79 (39.7)	
≥70	175 (29.3)	190 (32.6)		68 (36.2)	72 (36.2)	
Smoking status						
Never	136 (22.8)	76 (13.1)	<0.001	176 (93.6)	183 (92)	0.704
Ex-smoker	328 (54.9)	298 (51.2)		8 (4.3)	9 (4.5)	
Current	133 (22.3)	208 (35.7)		4 (2.1)	7 (3.5)	
Anemia	55 (9.6)	34 (6.1)	0.039	17 (9.6)	17 (9.2)	0.905
HGS, kg	38.0±7.0	38.9±7.0	0.044	23.8±4.6	24.2±4.9	0.342
FEV ₁ , % of predicted value	95.3±12.5	79.5±14.7	<0.001	98.3±12.9	80.0±17.6	<0.001
GOLD stage						
GOLD I		300 (51.5)			102 (51.3)	
GOLD II		261 (44.8)			86 (43.2)	
GOLD III, IV		21 (3.6)			11 (5.5)	

Note: Data represented as mean ± SD or n (%).

Abbreviations: COPD, chronic obstructive pulmonary disease; SD, standard deviation; HGS, hand grip strength; FEV₁, forced expiratory volume in 1 second; GOLD, Global Initiative for Chronic Obstructive Lung Disease.

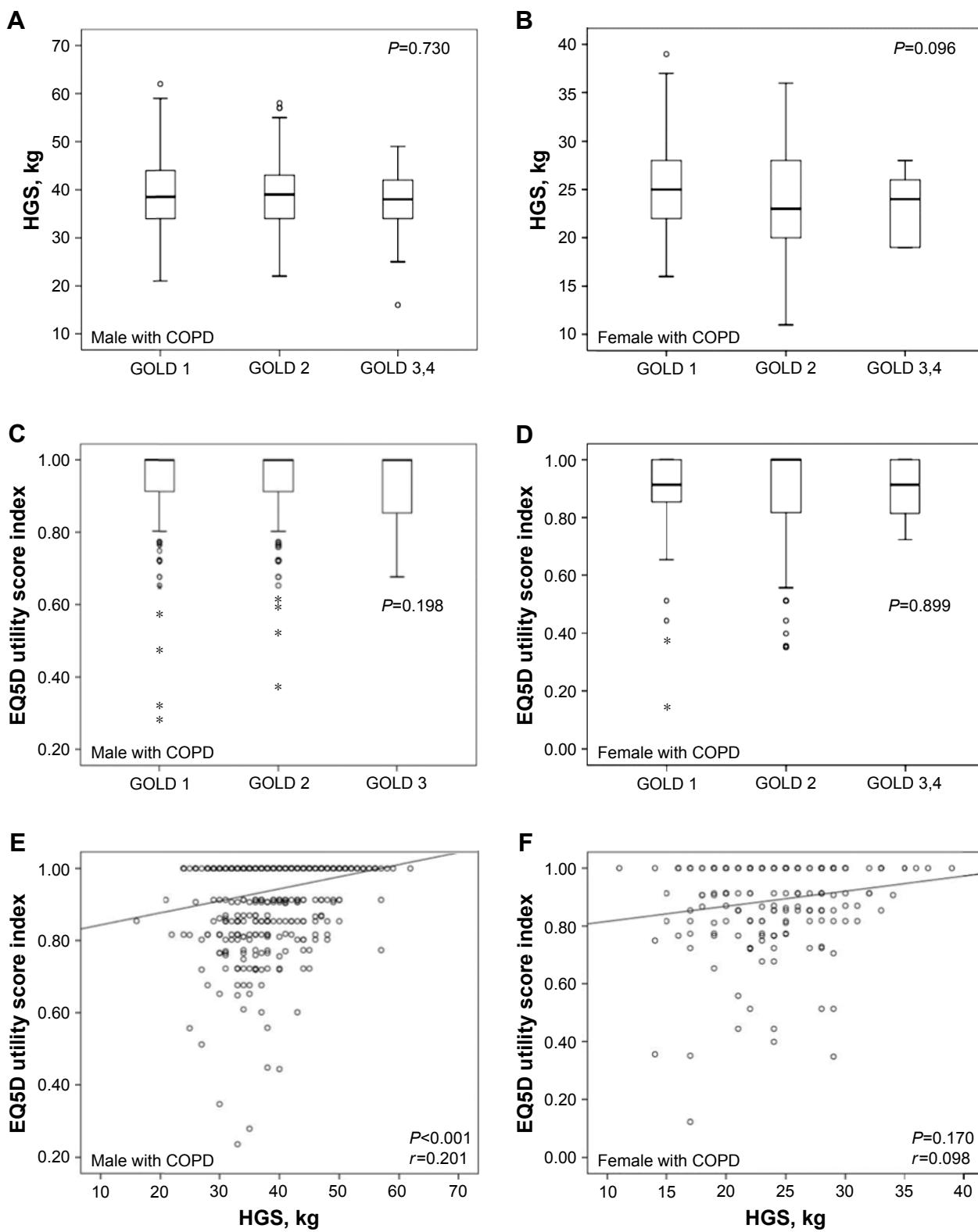


Figure 3 Associations among HGS, EQ5D score and lung function in COPD subjects.

Notes: (A) Association between GOLD stage and HGS in male subjects with COPD. (B) Association between GOLD stage and HGS in female subjects with COPD. (C) Association between GOLD stage and EQ5D utility score in male subjects with COPD. (D) Association between GOLD stage and EQ5D utility score in female subjects with COPD. (E) Correlation between HGS and EQ5D utility score in male subjects with COPD after adjusting for age and BMI. (F) Correlation between HGS and EQ5D utility score in female subjects with COPD after adjusting for age and BMI. *Indicates a variable that was away from the median.

Abbreviations: COPD, chronic obstructive pulmonary disease; HGS, hand grip strength; EQ5D, EuroQol Five-Dimension Questionnaire; GOLD, Global Initiative for Chronic Obstructive Lung Disease; BMI, body mass index.

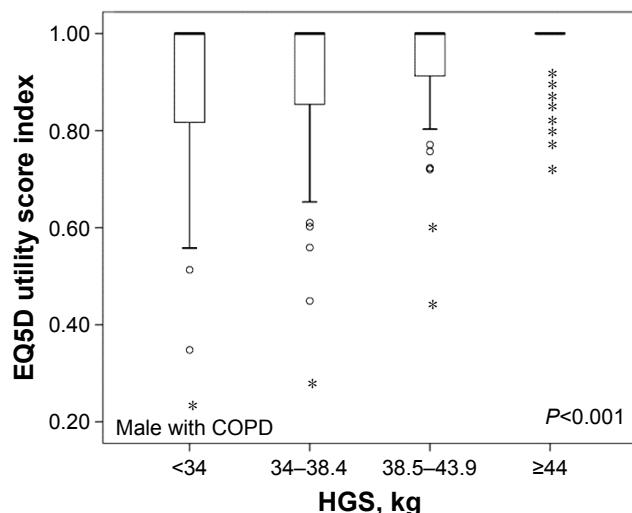


Figure 4 Classification of male subjects according to HGS interquartile ranges and the association with EQ5D utility scores.

Note: *Indicates a variable that was away from the median.

Abbreviations: COPD, chronic obstructive pulmonary disease; HGS, hand grip strength; EQ5D, EuroQol Five-Dimension Questionnaire.

are shown in Table 3. Lower HGS group had problems with mobility, performing usual activities, pain/discomfort and anxiety/depression as well as a lower EQ5D utility score. Only the descriptive section of self-care showed no differences according to HGS.

Relationship between HGS and smoking in male subjects

The association between smoking and HGS was analyzed because previous studies reported that smoking affects HGS.^{19,20} There were significant difference among never smoker, ex-smoker and current smoker in male subjects who were not considered aged (37.6 ± 6.9 vs 38.2 ± 6.9 vs 39.6 ± 7.2 kg, $P=0.002$). However, no significant correlations were observed in groups of male subjects classified as under 65 years or over 65 years (never smoker vs ex-smoker vs current smoker, 41.5 ± 7.1 vs 42.4 ± 6.4 vs 42.3 ± 6.7 kg, $P=0.566$ in age <65 years; 35.3 ± 5.7 vs 36.0 ± 6.0 vs 35.3 ± 5.4 kg, $P=0.388$ in age ≥ 65 years). In female subjects, the relationship could not be analyzed due to small sample size in female smoker.

In a subgroup of male COPD subjects, there was no statistical significance among never smoker, ex-smoker and current smoker (38.2 ± 7.1 vs 38.5 ± 6.7 vs 39.7 ± 7.3 kg, $P=0.110$). In addition, no association was found in EQ5D utility score and smoking status ($P=0.768$).

Discussion

This cross-sectional population-based study showed that COPD subjects did not have less HGS than propensity-matched

Table 3 Classification of male subjects according to HGS interquartile ranges and the association with EQ5D utility scores

Variables	HGS, kg				P-value
	<34	34-38.4	38.5-43.9	≥44	
EQ5D descriptive section, n (%)					
Mobility					
No problem	86 (67.2)	122 (74.8)	128 (89.5)	132 (89.2)	<0.001
Some problem	40 (31.3)	40 (24.5)	14 (9.8)	16 (10.8)	
Extreme problem	2 (1.6)	1 (0.6)	1 (0.7)	0 (0)	
Self-care					
No problem	120 (93.8)	154 (94.5)	139 (97.2)	146 (98.6)	0.113
Some problem	8 (6.3)	9 (5.5)	4 (2.8)	2 (1.4)	
Extreme problem	0 (0)	0 (0)	0 (0)	0 (0)	
Usual activities					
No problem	104 (81.3)	143 (87.7)	137 (95.8)	142 (95.9)	<0.001
Some problem	23 (18.0)	20 (12.3)	6 (4.2)	6 (4.1)	
Extreme problem	1 (0.8)	0 (0)	0 (0)	0 (0)	
Pain/discomfort					
No problem	91 (71.1)	110 (73.6)	119 (83.2)	128 (86.5)	0.003
Some problem	34 (26.6)	36 (22.1)	23 (16.1)	20 (13.5)	
Extreme problem	3 (2.3)	7 (4.3)	1 (0.7)	0 (0)	
Anxiety/depression					
No problem	112 (87.5)	136 (83.4)	135 (94.4)	141 (95.3)	0.002
Some problem	15 (11.7)	26 (16.0)	8 (5.6)	7 (4.7)	
Extreme problem	1 (0.8)	1 (0.6)	0 (0)	0 (0)	
EQ5D utility score (mean ± SD)	0.91±0.1	0.92±0.1	0.96±0.1	0.97±0.1	<0.001

Abbreviations: HGS, hand grip strength; EQ5D, EuroQol Five-Dimension Questionnaire; SD, standard deviation.

non-COPD subjects. In male subjects, those with COPD actually had significantly greater HGS, while there was no difference among females. HGS was also associated with EQ5D in male COPD subjects only.

Some studies have reported that muscle mass was associated with mortality in elderly adults and with functional outcome in critically ill patients.^{21,22} HGS is known as a simple assessment tool for nutritional status, systemic muscle mass and overall muscular strength because of its correlation with several muscular strength measurements such as knee and elbow extension.^{23–25} Therefore, many researchers have studied the association between HGS and mortality and have reported associations between the two.^{26–28} A recent systemic review of HGS has also reported that HGS is associated with increased risk of cardiovascular disease mortality in diverse populations.²⁹ One longitudinal study showed that HGS was a predictor of all causes of mortality.²⁶ However, the relationships between COPD and HGS in terms of lung function, severity of COPD and mortality are still controversial.^{10–13}

In this study, although the HGS of COPD subjects in both sexes were significantly lower than those of non-COPD subjects before propensity score matching, these results were suspected because they did not exclude factors that affect HGS such as cardiovascular disease, age and BMI. After propensity score matching to reduce the effects of several factors, HGS was higher in COPD subjects than in non-COPD subjects, and this finding was statistically significant in males. Cardiovascular disease is known to be associated with HGS, and the observed high prevalence of cardiovascular disease in subjects with COPD before propensity score matching in our study seems to offer at least a partial explanation of the conflicting results.^{7,13} Strandkvist et al¹³ reported that no difference in HGS between COPD and non-COPD was found and the subjects with cardiovascular disease had significantly lower HGS compared to subjects without cardiovascular disease.

A recent study reported that COPD subjects with GOLD stage 3 or 4 had lower HGS than did non-COPD subjects,¹³ while other studies have shown no correlation between HGS and COPD.^{10,11} In our COPD subjects, there was no significant correlation between HGS and lung function divided by GOLD stage. However, this result is similarly controversial to previous studies because the number of subjects of GOLD stages 3 and 4 in our study is small.^{10–13} Therefore, to confirm the correlation between HGS and lung function, a large population study that considers the factors affecting HGS is needed.

This study showed the association between HGS and COPD in Korea and specifically showed that HGS is correlated with the EQ5D index of quality of life in males with COPD. A similar association between muscle strength and quality of life has been reported in other studies.^{9,30} In our male subjects with COPD, subjects with lower HGS also had significantly lower EQ5D scores. Although the correlation coefficient of HGS and EQ5D utility score calculated from 0 point to 1 point was not high, it seemed clear when comparing the descriptive sections of EQ5D expressed in three functional levels with HGS. In groups classified according to HGS, HGS was negatively associated with problems in all EQ5D sections except self-care. We supposed that the description sections except for anxiety/depression section of the EQ5D involve activities requiring muscle movements; therefore, the association between HGS and EQ5D is to be expected. Furthermore, the EQ5D is indicative of overall quality of life, not dyspnea specifically, and the fact that HGS is not associated with lung function may have affected EQ5D results. In terms of smoking affecting HGS,^{19,20} female COPD might have affected the analysis of HGS due to small sample numbers. However, HGS results according to smoking habits in male COPD subjects showed no association with smoking and HGS. A previous study in Korea reported mean HGS values of 40.2 kg in males and 24.2 kg in females,³¹ whereas the corresponding mean values in COPD subjects in our study were 35.77 and 21.76 kg. The relatively greater decrement in HGS of males compared with females in our study might explain the result that the association of EQ5D with HGS was only observed in male subjects with COPD.

This study has the strength of being based on nationwide large-scale data on the relationship between COPD and HGS, which has not been widely studied in Korea. In addition, this study was able to study the association of COPD with HGS more accurately by using the propensity score matching method to adjust for variables other than COPD that are known to be associated with HGS, such as cardiovascular disease and stroke.

This study had some limitations. First, we used the fixed FEV₁/FVC criteria of pre-bronchodilator spirometry data. Because subjects did not use bronchodilator, it is possible that asthma subjects were mixed with the COPD subjects. And these cases could represent over-diagnoses of COPD, thereby resulting in misclassification or dilution of the COPD group.^{32,33} Second, the number of samples in the severe COPD group enrolled was low and the evaluation of HGS was limited to mild to moderate COPD group because

the proportion of COPD patients with GOLD stage 3 or 4 among total subjects was relatively small. Third, there is a limit to the HGS evaluation in female COPD subjects because the number of female subjects was small. Finally, the cross-sectional study design presents its own inherent limitations.

Conclusion

After propensity score matching, COPD subjects in this population-based study did not have lower HGS than non-COPD subjects. For males, the difference of HGS was statistically significant in subjects with and without COPD and in the unexpected direction; however, there was no difference in HGS in females. In addition, no association was found between HGS and lung function, whether classified as GOLD stage. However, the HGS of male COPD subjects was positively associated with EQ5D, an indicator of quality of life. HGS may be helpful as an additional method to the evaluation of quality of life in male COPD patients.

Disclosure

The authors report no conflicts of interest in this work.

References

- Vestbo J, Hurd SS, Agusti AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med.* 2013;187(4):347–365.
- Decramer M, Rennard S, Troosters T, et al. COPD as a lung disease with systemic consequences – clinical impact, mechanisms, and potential for early intervention. *COPD.* 2008;5(4):235–256.
- Karlah M, Fleig Mayer A, Maurici R, Pizzichini MM, Jones PW, Pizzichini E. The COPD Assessment Test: what do we know so far?: a systematic review and meta-analysis about clinical outcomes prediction and classification of patients into GOLD stages. *Chest.* 2016;149(2):413–425.
- Gupta N, Pinto LM, Morogan A, Bourbeau J. The COPD assessment test: a systematic review. *Eur Respir J.* 2014;44(4):873–884.
- Ortega FB, Silventoinen K, Tynelius P, Rasmussen F. Muscular strength in male adolescents and premature death: cohort study of one million participants. *BMJ.* 2012;345:e7279.
- Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA.* 2009;301(19):2024–2035.
- Leong DP, Teo KK, Rangarajan S, et al; Prospective Urban Rural Epidemiology (PURE) Study Investigators. Prognostic value of grip strength: findings from the prospective urban rural epidemiology (PURE) study. *Lancet.* 2015;386(9990):266–273.
- Waschki B, Kirsten A, Holz O, et al. Physical activity is the strongest predictor of all-cause mortality in patients with COPD: a prospective cohort study. *Chest.* 2011;140(2):331–342.
- Troosters T, van der Molen T, Polkey M, et al. Improving physical activity in COPD: towards a new paradigm. *Respir Res.* 2013;14:115.
- Marino DM, Marrara KT, Ike D, De Oliveira AD Jr, Jamami M, Di Lorenzo VA. Study of peripheral muscle strength and severity indexes in individuals with chronic obstructive pulmonary disease. *Physiother Res Int.* 2010;15(3):135–143.
- Gosselink R, Troosters T, Decramer M. Distribution of muscle weakness in patients with stable chronic obstructive pulmonary disease. *J Cardiopulm Rehabil.* 2000;20(6):353–360.
- Burton C, Ter Riet G, Puhan MA, et al. Handgrip weakness and mortality risk in COPD: a multicentre analysis. *Thorax.* 2016;71(1):86–87.
- Strandkvist VJ, Backman H, Roding J, Stridsman C, Lindberg A. Hand grip strength is associated with forced expiratory volume in 1 second among subjects with COPD: report from a population-based cohort study. *Int J Chron Obstruct Pulmon Dis.* 2016;11:2527–2534.
- Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J.* 2005;26(2):319–338.
- Roberts HC, Denison HJ, Martin HJ, et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. *Age Ageing.* 2011;40(4):423–429.
- Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33(5):337–343.
- Lee YK, Nam HS, Chuang LH, et al. South Korean time trade-off values for EQ-5D health states: modeling with observed values for 101 health states. *Value Health.* 2009;12(8):1187–1193.
- Nolan CM, Longworth L, Lord J, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax.* 2016;71(6):493–500.
- Al-Obaidi S, Al-Sayegh N, Nadar M. Smoking impact on grip strength and fatigue resistance: implications for exercise and hand therapy practice. *J Phys Act Health.* 2014;11(5):1025–1031.
- Saito T, Miyatake N, Sakano N, et al. Relationship between cigarette smoking and muscle strength in Japanese men. *J Prev Med Public Health.* 2012;45(6):381–386.
- Puthucheary ZA, Hart N. Skeletal muscle mass and mortality – but what about functional outcome? *Crit Care.* 2014;18(1):110.
- Srikanthan P, Karlamangla AS. Muscle mass index as a predictor of longevity in older adults. *Am J Med.* 2014;127(6):547–553.
- Wind AE, Takken T, Helders PJ, Engelbert RH. Is grip strength a predictor for total muscle strength in healthy children, adolescents, and young adults? *Eur J Pediatr.* 2010;169(3):281–287.
- Springstrol KA, Gal NJ, Ford AL, Whiting SJ, Dahl WJ. Evaluation of handgrip strength and nutritional risk of congregate nutrition program participants in Florida. *J Nutr Gerontol Geriatr.* 2016;35(3):193–208.
- Granic A, Jagger C, Davies K, et al. Effect of dietary patterns on muscle strength and physical performance in the very old: findings from the Newcastle 85+ study. *PLoS One.* 2016;11(3):e0149699.
- Sasaki H, Kasagi F, Yamada M, Fujita S. Grip strength predicts cause-specific mortality in middle-aged and elderly persons. *Am J Med.* 2007;120(4):337–342.
- Bohannon RW. Hand-grip dynamometry predicts future outcomes in aging adults. *J Geriatr Phys Ther.* 2008;31(1):3–10.
- Bohannon RW. Muscle strength: clinical and prognostic value of hand-grip dynamometry. *Curr Opin Clin Nutr Metab Care.* 2015;18(5):465–470.
- Chainani V, Shaharyar S, Dave K, et al. Objective measures of the frailty syndrome (hand grip strength and gait speed) and cardiovascular mortality: a systematic review. *Int J Cardiol.* 2016;215:487–493.
- Calik-Kutukcu E, Savci S, Saglam M, et al. A comparison of muscle strength and endurance, exercise capacity, fatigue perception and quality of life in patients with chronic obstructive pulmonary disease and healthy subjects: a cross-sectional study. *BMC Pulm Med.* 2014;14:6.
- Yoo JI, Choi H, Ha YC. Mean hand grip strength and cut-off value for sarcopenia in Korean adults using KNHANES VI. *J Korean Med Sci.* 2017;32(5):868–872.
- Roberts SD, Farber MO, Knox KS, et al. FEV1/FVC ratio of 70% misclassifies patients with obstruction at the extremes of age. *Chest.* 2006;130(1):200–206.
- Hwang YI, Kim CH, Kang HR, et al. Comparison of the prevalence of chronic obstructive pulmonary disease diagnosed by lower limit of normal and fixed ratio criteria. *J Korean Med Sci.* 2009;24(4):621–626.

International Journal of COPD**Publish your work in this journal**

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols.

Submit your manuscript here: <http://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>

Dovepress

This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

RESEARCH ARTICLE

Open Access



CrossMark

Propensity score analysis of lung cancer risk in a population with high prevalence of non-smoking related lung cancer

Kuei-Feng Lin^{1†}, Hsiu-Fu Wu^{1,2†}, Wei-Chun Huang^{2,3}, Pei-Ling Tang³, Ming-Ting Wu^{1,2} and Fu-Zong Wu^{1,2,4*}

Abstract

Background: Lung cancer has been the leading cause of cancer-related mortality worldwide among both men and women in recent years. There is an increase in the incidence of nonsmoking-related lung cancer in recent years. The purpose of the present study was to investigate multiple potential risk factors for nonsmoking-related lung cancer among Asian Ethnic Groups.

Methods: We used a propensity score-matched cohort analysis for this study. We retrospectively review the medical record of 1975 asymptomatic healthy subjects (40 ~ 80 years old) who voluntarily underwent low-dose chest CT from August 2013 to October 2014. Clinical information and nodule characteristics were recorded.

Results: A propensity score-matched cohort analysis was applied to adjust for potential bias and to create two comparable groups according to family history of lung cancer. For our primary analysis, we matched 392 pairs of subjects with family history of lung cancer and subjects without history. Logistic regression showed that female gender and a family history of lung cancer were the two most important predictor of lung cancer in the endemic area with high prevalence of nonsmoking-related lung cancer (OR = 11.199, 95% CI = 1.444–86.862; OR = 2.831, 95% CI = 1.000136–8.015). In addition, the number of nodules was higher in subjects with family history of lung cancer in comparison with subjects without family history of lung cancer (OR = 1.309, 95% CI = 1.066–1.607).

Conclusions: In conclusion, risk-based prediction model based on the family history of lung cancer and female gender can potentially improve efficiency of lung cancer screening programs in Taiwan.

Keywords: Non-smoker lung cancer, Propensity score matching, Lung adenocarcinoma spectrum, Risk factor

Background

Lung cancer has been the leading cause of cancer-related mortality worldwide among both men and women in recent years [1–3]. The landmark National Lung Screening Trial (NLST) evaluated the benefits of low-dose computed tomography (LDCT) for screening of heavy smokers (≥ 30 pack-years) and found that annual screening by LDCT yielded a relative reduction of lung cancer mortality of 20% among those screened when compared to chest radiography [3]. Smoking is the major risk factor for lung cancer,

but an increase in the incidence of nonsmoking-related lung cancer in recent years has been addressed [4–8]. There has been an increase in the prevalence of non-smoking associated lung cancers in Asian countries such as China, Taiwan, Korea, and Japan over the past few years [9, 10]. Previous studies suggested that a potential association among nonsmokers who had lung adenocarcinoma with associated risk factors such as age, gender, body mass index (BMI), history of lung cancer, and personal cancer history [8, 11]. A major concern has remained regarding that selection bias that occurs as a result of self-referral or physician referral in the setting of these studies designs, which is ordinarily considered a threat to both internal and external validity of the studies [8, 11]. Propensity score matching method is increasingly being used currently and a useful statistical technology in observational studies to ensure that

* Correspondence: cmvwu1029@gmail.com

†Equal contributors

¹Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan

²Faculty of Medicine, School of Medicine, Institute of Clinical Medicine, National Yang Ming University, Taipei, Taiwan

Full list of author information is available at the end of the article

propensity score is balanced across treatment and control groups as an alternative to conventional covariate adjustment in logistic regression models [12]. Using propensity score matching analysis, clinical/demographic characteristics of subjects between the groups with family history of lung cancer (+) versus without family history of lung cancer (-) could be balanced out, thus mimicking randomized controlled trial design. The purpose of the present study was to investigate potential risk factors for nonsmoking-related lung cancer among Asian population based on propensity score matching analysis which could reduce selection bias and potential baseline differences between the two groups.

Methods

Study population and cohort

A flow diagram describing the subject recruitment and exclusions is shown in Fig. 1. We retrospectively analyzed 1975 (1083 males and 892 females) asymptomatic healthy subjects (age range 40 to be 80-year-old) who voluntarily underwent self-paid LDCT exam at the health check-up center of Kaohsiung Veterans General Hospital from August 2013 to October 2014. Clinical information included gender, age, BMI, family history of lung cancer, and family history of other cancers in first and second-degree relatives was collected. Moreover, nodular characteristics were recorded according to ACR Lung-RADS classification shown in Table 1 [13, 14]. Categories 1 (negative) and 2 (benign appearance) correspond to negative screening results, and categories 3 (probably benign) and 4 (suspicious) correspond to positive screening results. Category 4 is divided into 4A, 4B, and 4X, based on the level of suspicion of malignancy

according to the nodule size and characteristics summarized in Table 1. Increases in the probability of malignancy are expressed by assigning either subcategory, 4A (5%–15%), 4B (>15%), or 4X (additional finding such as spiculation or enlarged lymph nodes). The average follow-up time of subjects with suspicious nodules was 1.6 ± 0.5 years after the baseline LDCT.

Among 1975 screened subjects, 72.8% (1438/1975) of the screened subjects were never-smokers, 16.5% (326/1975) were current smokers, and 10.7% (211/1975) were former smokers. Only 7.5% (149/1975) of the study subjects would have been eligible for screening based on the NLST enrollment criteria. Among 1975 screened subjects, there were 27 subjects diagnosed with non-smoking related lung cancer (two lung cancer subjects with smoking were excluded). Definition of non-smoking related lung cancer was defined as the lung adenocarcinoma spectrum such as adenocarcinoma in situ, minimally invasive adenocarcinoma, and invasive adenocarcinoma diagnosed by surgical or biopsy proof. Histopathologic diagnosis of atypical adenomatous hyperplasia was excluded from this study.

Covariate and propensity score matching

All the subjects were divided to two groups: the group with family history of lung cancer (398 subjects) and the group without family history of lung cancer (1577 subjects). However, 87 patients were excluded because of missing data on BMI profiles. We used a 1:1 propensity score-matched pair method combined with covariate adjustment to analyze patients with and without family history of lung cancer shown in Fig. 1. The unbalanced conditions at baseline between the two groups were

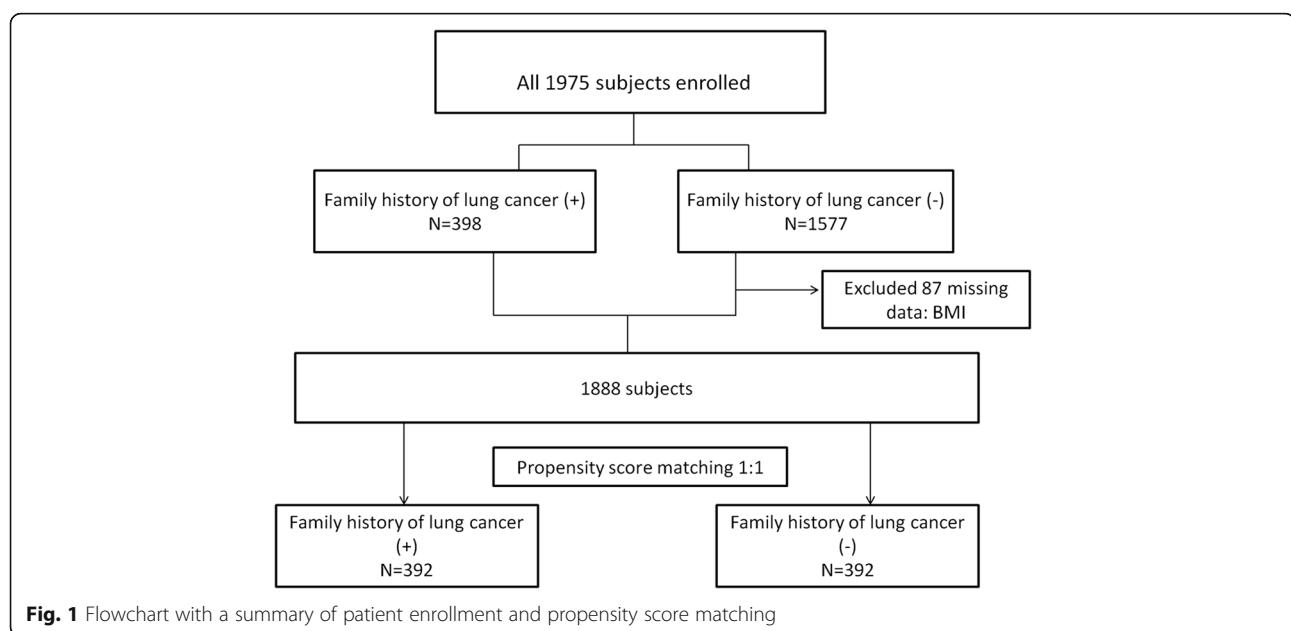


Table 1 Summary of Lung-RADS Classification^a

Lung-RADS	Baseline screening	Subsequent screening
1	No nodules; nodules with calcification	No nodules; nodules with calcification
2	Solid/part solid: < 6 mm GGN: < 20 mm	Solid/part solid: < 6 mm GGN: < 20 mm or unchanged/slowly growing Category 3–4 nodules unchanged at ≥3 mo
3	Solid: ≥ 6 to <8 mm Part solid: ≥ 6 mm with solid component <6 mm GGN: ≥ 20 mm	Solid: New ≥ 4 to <6 mm Part solid: New <6 mm GGN: New ≥ 20 mm
4A	Solid: ≥ 8 mm to <15 mm Part solid: ≥ 8 mm with solid component ≥ 6 and <8 mm	Solid: Growing < 8 mm or new ≥ 6 and < 8 mm Part solid: ≥ 6 mm with new or growing solid component < 4 mm
4B	Solid: ≥ 15 mm Part solid: Solid component ≥ 8 mm	Solid: New or growing and ≥ 8 mm Part solid: ≥ 6 mm with new or growing solid component ≥ 4 mm
4X	Category 3 or 4 nodules with additional features; imaging findings that increase suspicion of malignancy	Category 3 or 4 nodules with additional features; imaging findings that increase suspicion of malignancy

GGN ground-glass nodule

^a Size is the average diameter rounded to the nearest whole number. Growth is a size increase >1.5 mm

Lung-RADS: The ACR Lung Imaging Reporting and Data System

controlled by using PS matching with covariate adjustment. The 1:1 PS matching yielded matched pairs of 392 subjects with family history of lung cancer and 392 patients without family history of lung cancer, resulting in no differences in age, gender, BMI, and the proportion of other cancers of family history.

LDCT imaging acquisition and interpretation

All scans were performed with a 16-slice multi-detector CT (Somatom Sensation 16, Siemens Healthcare, Erlangen, Germany) and a 64-slice multi-detector CT (Aquilion 64; Toshiba Medical Systems) from the lung apex to the base without contrast enhancement. The LDCT examination protocols met the CMS (Centers for Medicare & Medicaid Services) requirement of the volume CT dose index (CTDI-vol) ≤ 3.0 milligray (mGy) for standard-size patients based on recommendations of the ACR and Society of Thoracic Imaging for different vendors setting [15]. Scans were obtained with the subjects in supine position at end inspiration. The data were reconstructed with filtered back projection, a slice thickness of 2 mm, and an increment of 2 mm, using a smooth convolution kernel (Siemens B30f, Toshiba FC02). All studies were evaluated on lung and mediastinal windows on a picture-archiving and communication system and reported by two experienced thoracic radiologists with 8 and 12 years of experience, respectively.

Statistical analysis

Statistical analysis was performed using SPSS® v17.0 for Windows (SPSS, Inc., Chicago, IL) and the SAS® software package (SAS Institute, Inc., Cary, NC). To minimize the effect of potential confounders on selection bias, propensity

scores were generated by using the multiple logistic regressions to estimate the probability that subjects have family history of lung cancer or not. The covariates entered into the propensity score were age, gender, and BMI. Propensity score matching (1:1 match) was performed to adjust for differences in baseline clinical characteristics, yielding a total of 784 subjects: 392 subjects with family history of lung cancer and 392 subjects without family history of lung cancer (SAS Institute, Inc., Cary, NC).

Baseline characteristics were performed as mean ± standard deviation (SD).

Comparisons between the two groups were performed by using the independent T-test for continuous data and chi-square test for categorical data before and after PS matching. The Fisher exact chi-square test was used to analyze when the smallest expected value is less than 5. Multiple logistic regression models were developed, and odds ratios (ORs) were used to evaluate risk factors associated with lung cancer. Data analysis was performed using SPSS® v17.0 for Windows (SPSS, Inc., Chicago, IL).

Results

We retrospectively review the medical record of 1975 asymptomatic healthy subjects (40 ~ 80 years old) who voluntarily underwent low-dose chest CT (1083 males, 892 females) from August 2013 to October 2014. We identified 398 patients with family history of lung cancer while the other 1577 patients without family history of lung cancer shown in Fig. 1. The baseline characteristics in the pre-match and post-match cohorts are presented in Table 2.

Table 2 Patient characteristics before and after propensity score matching

Characteristics	Before PSM (N = 1975)			After PSM (N = 784)			P	
	All	Family history (+)	Family history (-)	P	All	Family history (+)	Family history (-)	
Age, years	56.56 ± 9.01	56.1 ± 9.39	58.39 ± 7.05	<0.0001 ^a	58.61 ± 7.15	58.57 ± 6.85	58.66 ± 7.55	0.865 ^a
Sex (%)				<0.0001 ^b				0.94 ^b
Male	1083	54.90%	136 (34.1%)	947 (60.05%)	517	65.90%	258 (65.8%)	259 (66.1%)
Female	892	45.10%	262 (65.9%)	630 (39.95%)	267	34.10%	134 (34.2%)	133 (33.9%)
BMI	24.32 ± 3.49	23.76 ± 3.36	24.46 ± 3.50	<0.0001 ^a	23.76 ± 3.36	23.76 ± 3.36	23.88 ± 3.56	0.644 ^a
Nodule number	0.63 ± 1.16	1.09 ± 1.53	0.51 ± 1.027	<0.0001 ^a	0.84 ± 1.392	1.1 ± 1.53	0.59 ± 1.17	<0.0001 ^a
History of other cancers				0.023 ^b				0.601 ^b
Present	621	31.40%	144 (36.1%)	477 (30.2%)	275	35.10%	141 (36%)	131 (34.2%)
Absent	1354	68.50%	254 (63.9%)	1100 (69.8%)	509	64.90%	258 (64%)	25 (65.8%)
Category 4 lesion				<0.0001 ^b				0.186 ^b
Present	53	2.68%	21 (5.27%)	32 (2.02%)	36	4.59%	21 (5.3%)	15 (3.82%)
Absent	1922	97.32%	377 (94.73%)	1545 (97.98%)	748	95.41%	371 (94.7%)	377 (96.18%)
Lung cancer				<0.0001 ^b				0.019 ^c
Present	27	1.40%	15 (3.76%)	12 (0.76%)	20	2.60%	15 (3.8%)	5 (1.3%)
Absent	1948	98.60%	383 (96.24%)	1565 (99.24%)	764	97.49%	377 (96.2)	387 (98.7)

^aUsing independent t-test for continuous variables; ^bUsing Chi-square test for categorical variables; ^cUsing Fisher's exact test for categorical variables
Abbreviations: PSM propensity score matching, BMI body mass index

Baseline characteristics before propensity matching

Patients were significantly younger in the family history of lung cancer (+) group compared with the family history of lung cancer (-) group (56.1 ± 9.39 years old versus 58.39 ± 7.05 years old); the BMI in the family history of lung cancer (+) group is lower compared with the family history of lung cancer (-) group ($23.76 \pm 3.36 \text{ kg/m}^2$ versus $24.46 \pm 3.50 \text{ kg/m}^2$); there were more nodules in the family history of lung cancer (+) group compared with the family history of lung cancer (-) group (1.09 ± 1.53 versus 0.51 ± 1.027). There were several parameters of baseline characteristics statistically higher in the family history of lung cancer (+) group, including the percentage of female gender (65.9% vs. 39.95%), the percentage of category 4 lesions (5.27% vs. 2.02%), the percentage of family history of other cancers (36.1% vs. 30.2%), and the percentage of lung cancer (3.76% vs. 0.76%).

Among 27 subjects with non-smoking related lung cancer diagnosed, 8 (29.62%) subjects had a diagnosis of synchronous multiple primary lung cancers (MPLCs) according to the diagnostic criteria proposed by Martini and Melamed before propensity score matching [16]. Among 20 (35%) subjects with non-smoking related lung cancer diagnosed, 7 subjects had a diagnosis of synchronous MPLCs according to the diagnostic criteria proposed by Martini and Melamed after propensity score matching [16].

To further investigate this imbalance, we illustrate histogram of the distribution of the propensity score for both groups before and after propensity matching. Figure 2a presents histograms of unbalanced propensity score distribution for both groups before propensity

matching. Figure 2b presents histograms of balanced propensity score distribution for both groups after the propensity matching.

Baseline characteristics after propensity matching

According to the propensity score matching 1:1 shown in Table 2, 392 patients in the family history of lung cancer (+) group were matched with 392 in the family history of lung cancer (-) group. The matching process eliminated some significant differences that existed between the family history of lung cancer (+) group and the family history of lung cancer (-) group such as age, sex, BMI, the percentage of family history of other cancers, and category 4 lesions, while the nodule numbers and the percentage of lung cancer remained significant different.

Univariate and multivariate logistic regression analysis for lung cancer risk

Table 3 lists the univariate and multivariate logistic regression analyses to determine the predictors of lung cancer. Female gender (univariate model: OR = 10.149, 95% confidence interval (CI) = 1.351–76.227; multivariate model: OR = 11.199, 95% CI = 1.444–86.862), nodule number (univariate model: OR = 1.353, 95% CI = 1.114–1.642; multivariate model: OR = 1.309, 95% CI = 1.066–1.607), and family history of lung cancer (univariate model: OR = 3.08, 95% CI = 1.108–8.557; multivariate model: OR = 2.831, 95% CI = 1.000136–8.015) were significant associated with lung cancer both on univariate and multivariate analysis.

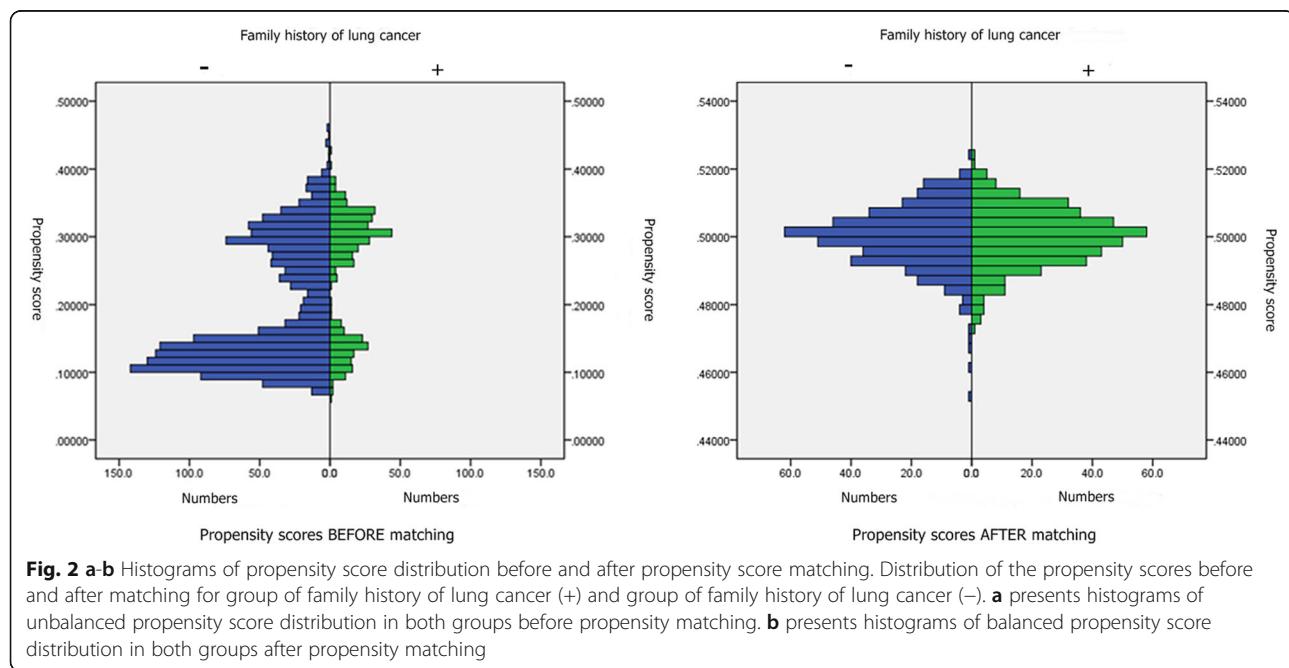


Table 3 Univariate and multivariate logistic regression analyses for predictors of lung cancer in 784 subjects after propensity score matching

Characteristics	Univariate analysis			Multivariate analysis		
	Odds ratio	95% CI	P value	Odds ratio	95% CI	P value
Age, years	1.015	0.953–1.082	0.641	0.994	0.923–1.070	0.871
Sex (female gender)	10.149	1.351–76.227	0.024	11.199	1.444–86.862	0.021
BMI, kg/m ²	1.015	0.895–1.151	0.815	1.079	0.953–1.221	0.23
Nodule number	1.353	1.114–1.642	0.02	1.309	1.066–1.607	0.01
Family history of lung cancer	3.08	1.108–8.557	0.031	2.831	1.000136–8.015	0.05
Family history of other cancer	1.241	0.501–3.073	0.641	1.078	0.425–2.732	0.875

Abbreviations: BMI body mass index, CI confidence interval

Discussion

In this retrospective analysis applying propensity score matching in order to minimize confounding effects and selection bias to estimate the true causal effect, we demonstrated three major findings. The first one is that to utilize the propensity score matching to adjust for selection bias could address the balanced baseline characteristics between exposure and control subjects and improve the internal validity of the study. The second finding is family history of lung cancer and female gender were significantly associated with lung cancer based on univariate or multivariate logistic regression. Previous studies have addressed the issue that family history of lung cancer significantly association with non-smoking related lung cancer, mainly in middle-age women of Asian population. However, these results were based on data available from previous case-control or retrospective cohort studies which more susceptible to the effects of selection bias [7, 8]. The present study demonstrated for the first time that identification two important associated risk factors with lung cancer in an Asian cohort with less smoker using a propensity score matching method to construct quasi-experimental design intended to stimulate randomized controlled trial (RCT) design and minimize the selection bias [17]. Familial risk of lung cancer is attributable to share more complex genetic and environmental factors [18–20]. Our study demonstrated that familial history of lung cancer significantly associated with non-smoking related lung cancer, especially in women. In addition, another study demonstrated that women with a history of lung infection (bronchitis or pneumonia) positively influenced lung cancer development [21]. The third finding, increasing numbers of nodules were significantly associated with lung cancer in an Asian population, mainly non-smoker. The reported incidence of synchronous MPLCs in patients with lung cancer in our study is high up to 35% (one example case shown in Fig. 3). The incidence of synchronous MPLC has been reported to range from 0.7% to 30% of patients with lung cancer in the previous literature reviews [8, 22–24]. This study result

support that high prevalence of Multifocal ground glass/lepidic (GG/L) lung cancer, a kind of lung adenocarcinoma subtype which often occurred in Asian women or non-smoker recently proposed by the International Association for the Study of Lung Cancer (IASLC) Lung Cancer Staging Project in 2016 [9]. In addition, the overall lung cancer prevalence rate was 1.40% (27/1975) in this study cohort. Our results are congruent with other published data from Asian population in the group of non-smokers or lesser smokers (lung cancer prevalence rate 1 ~ 2% at the baseline LDCT screening) [25, 26]. Our study population consists mainly of non-smokers, which is very different from the NLST and other LDCT lung cancer screening studies conducted outside of Asia

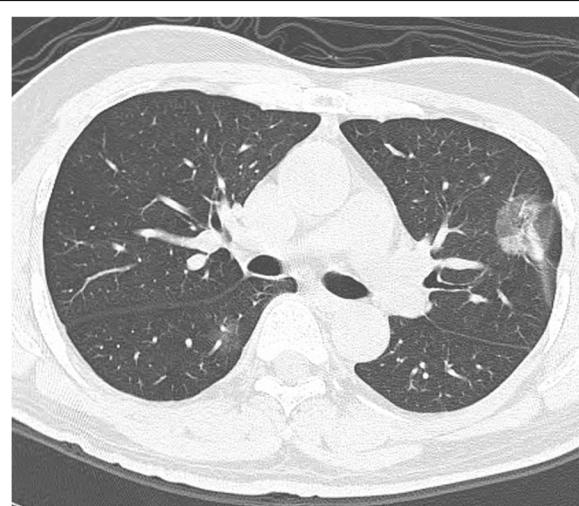


Fig. 3 An example of multifocal GG/L lung cancer, a kind of lung adenocarcinoma subtype which often occurred in Asian women or non-smoker recently according to the IASLC Lung Cancer Staging Project in 2016. A 61-year-old woman had a 2.8 cm part-solid nodule in LUL, and another one pure GGN nodule 1.4 cm in RLL. The patient underwent sequentially video-thoracoscopic wedge resection of RLL and LUL. Further pathologic report demonstrated invasive adenocarcinoma in LUL, and adenocarcinoma in situ in RLL. Synchronous multiple primary lung cancer was diagnosed according to the diagnostic criteria proposed by Martini and Melamed. Abbreviations: RLL = right lower lobe; GGN = groundglass nodule; LUL = left upper lobe

[3, 27]. Recent studies have investigated more detail about the diagnosis, management and prognosis of Multifocal GG/L lung cancer [28–30]. This issue should be more emphasized in Asian lung cancer screening program due to high prevalence of synchronous MPLC reported according to previous and the current studies [8, 22].

There are several limitations to our study. First, propensity score matching can only control for observed covariates such as age, BMI or sex in the study. However, any unobserved covariates (cooking, second-hand smoking and air pollution) cannot be adjusted to balancing baseline characteristics between exposure and unexposed with reducing selection bias [31]. Second, propensity score matching methods resulted in throwing out over half of the subjects in the unexposed group, reducing the overall sample size and negatively affecting statistical power. To maximize our statistical power to detect this effect, it is mandatory to perform a much larger cohort in an Asian population. Third, a large number of subjects are eliminated after propensity scoring matching because of limited numbers within the exposure group despite the algorithm of full matching. Thus further large cohort studies are needed to establish generalizability of these study results because of the loss of study subjects numbers threatening external validity.

Conclusion

In conclusion, in this retrospective analysis applying propensity score matching in order to minimize confounding effects and identify two important risk factors of female gender and family history of lung cancer for non-smoker lung cancer prediction. In the future, risk-based prediction model based on the family history of lung cancer and female gender can potentially improve efficiency of lung cancer screening programs in Taiwan.

Abbreviations

ACR: American College of Radiology; BMI: Body mass index; CI: Confidence interval; CMS: Centers for Medicare & Medicaid Services; CT: Computed tomography; CTDlvol: Volume CT dose index; GG/L: Ground glass/lepidic; IASLC: International Association for the Study of Lung Cancer; LDCT: Low-dose computed tomography; Lung-RADS: Lung Imaging Reporting and Data System; MPLC: Multiple primary lung cancers; NLST: National Lung Screening Trial; OR: Odds ratio; RCT: Randomized controlled trial; SD: Standard deviation

Acknowledgements

The authors thank all doctors who responded to our screening and investigation, and appreciate Miss Taso Shu-Ping for the technical help and preparing this manuscript. This study was supported by a grant from Kaohsiung Veterans General Hospital, VGHKS103-015, VGHKS104-048, VGHKS105-064, Taiwan, R.O.C.

Funding

This study was supported by Grants from Kaohsiung Veterans General Hospital, VGHKS103-015, VGHKS104-048, VGHKS105-064, Taiwan, R.O.C.

Availability of data and materials

All results are available at the Kaohsiung Veterans General Hospital. The database used for the study can be available from the corresponding author under demand if needed.

Authors' contributions

FZW Contribution to concept and design, acquisition of data, analysis, drafting of manuscript. FZW Contribution to concept, design of the study and acquisition of data. FZW Contribution towards concept, design of the study and acquisition of data. KFL, HFW, MTW Contribution towards acquisition of data and initial concept/design. WCH, PLT performed statistical analysis. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The Kaohsiung Veterans General Hospital Institutional Review Board approved the study and waived the requirement for informed consent due to the retrospective nature of this study.

Consent for publication

Not applicable.

Competing interests

None of the authors have a conflict of interest to declare in relation to this work.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan. ²Faculty of Medicine, School of Medicine, Institute of Clinical Medicine, National Yang Ming University, Taipei, Taiwan. ³Research Center of Medical Informatics, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan. ⁴School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung City, Taiwan.

Received: 11 April 2017 Accepted: 30 August 2017

Published online: 06 September 2017

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013;63(1):11–30.
2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin. 2014;64(1):9–29.
3. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011; 365(5):395–409.
4. Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, Thun MJ, et al. Lung cancer in never smokers: clinical epidemiology and environmental risk factors. Clinical cancer research: an official journal of the American Association for Cancer Research. 2009;15(18):5626–45.
5. Wakelee HA, Chang ET, Gomez SL, Keegan TH, Feskanich D, Clarke CA, et al. Lung cancer incidence in never smokers. J Clin Oncol Off J Am Soc Clin Oncol. 2007;25(5):472–8.
6. Ko YC, Lee CH, Chen MJ, Huang CC, Chang WY, Lin HJ, et al. Risk factors for primary lung cancer among non-smoking women in Taiwan. Int J Epidemiol. 1997;26(1):24–31.
7. Lo Y-L, Hsiao C-F, Chang G-C, Tsai Y-H, Huang M-S, Su W-C, et al. Risk factors for primary lung cancer among never smokers by gender in a matched case-control study. Cancer Causes Control. 2013;24(3):567–76.
8. Wu F-Z, Huang Y-L, Wu CC, Tang E-K, Chen C-S, Mar G-Y, et al. Assessment of selection criteria for low-dose lung screening CT among Asian ethnic groups in Taiwan: from mass screening to specific risk-based screening for non-smoker lung cancer. Clinical Lung Cancer. 2016;17:e45–56.
9. Detterbeck FC, Nicholson AG, Franklin WA, Marom EM, Travis WD, Girard N, et al. The IASLC lung cancer staging project: summary of proposals for revisions of the classification of lung cancers with multiple pulmonary sites of involvement in the forthcoming eighth edition of the TNM classification. J Thorac Oncol. 2016;11(5):639–50.
10. Subramanian J, Govindan R. Lung cancer in never smokers: a review. J Clin Oncol Off J Am Soc Clin Oncol. 2007;25(5):561–70.
11. Wu X, Wen CP, Ye Y, Tsai M, Wen C, Roth JA, et al. Personalized risk assessment in never, light, and heavy smokers in a prospective cohort in Taiwan. Sci Rep. 2016;6:36482.

12. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, et al. Comparison of propensity score methods and covariate adjustment. Evaluation in 4 cardiovascular studies. *J Am Coll Cardiol.* 2017;69(3):345–57.
13. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR lung-RADS in a clinical CT lung screening program. *J Am Coll Radiol.* 2015;12(3):273–6.
14. Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, et al. Performance of lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Ann Intern Med.* 2015;162(7):485–91.
15. Kazerooni EA, Austin JH, Black WC, Dyer DS, Hazelton TR, Leung AN, MF MN-G, et al. ACR-STR practice parameter for the performance and reporting of lung cancer screening thoracic computed tomography (CT): 2014 (resolution 4). *J Thorac Imaging.* 2014;29(5):310–6.
16. Martini N, Melamed MR. Multiple primary lung cancers. *J Thorac Cardiovasc Surg.* 1975;70(4):606–12.
17. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res.* 2011; 46(3):399–424.
18. Kanwal M, Ding X-J, Cao Y. Familial risk for lung cancer. *Oncol Lett.* 2017; 13(2):535–42.
19. Hemminki K, Dong C, Vaittinen P. Cancer risks to spouses and offspring in the family-cancer database. *Genet Epidemiol.* 2001;20(2):247–57.
20. Hemminki K, Li X. Familial risk for lung cancer by histology and age of onset: evidence for recessive inheritance. *Exp Lung Res.* 2005;31(2):205–15.
21. Osann KE. Lung cancer in women: the importance of smoking, family history of cancer, and medical history of respiratory disease. *Cancer Res.* 1991;51(18):4893–7.
22. van Rens MT, Zanen P, Brutel de La Riviere A, Elbers HR, van Swieten HA, van Den Bosch JM. Survival in synchronous vs. single lung cancer: upstaging better reflects prognosis. *Chest.* 2000;118(4):952–8.
23. Ferguson MK, DeMeester TR, DesLauriers J, Little AG, Piraux M, Golomb H. Diagnosis and management of synchronous lung cancers. *J Thorac Cardiovasc Surg.* 1985;89(3):378–85.
24. Lam S, MacAulay C, Palcic B. Detection and localization of early lung cancer by imaging techniques. *Chest.* 1993;103(1 Suppl):12s–4s.
25. Yi CA, Lee KS, Shin M-H, Cho YY, Choi Y-H, Kwon OJ, et al. Low-dose CT screening in an Asian population with diverse risk for lung cancer: a retrospective cohort study. *Eur Radiol.* 2015;25(8):2335–45.
26. Chen C-Y, Chen C-H, Shen T-C, Cheng W-C, Hsu C-N, Liao C-H, et al. Lung cancer screening with low-dose computed tomography: experiences from a tertiary hospital in Taiwan. *J Formos Med Assoc.* 2016;115(3):163–70.
27. Fintelmann FJ, Bernheim A, Digumathy SR, Lennes IT, Kalra MK, Gilman MD, et al. The 10 pillars of lung cancer screening: rationale and logistics of a lung cancer screening program. *Radiographics.* 2015;35(7):1893–908.
28. Zhang Z, Gao S, Mao Y, Mu J, Xue Q, Feng X, He J. Surgical outcomes of synchronous multiple primary non-small cell lung cancers. *Sci Rep.* 2016;6:23252.
29. Xue X, Liu Y, Pan L, Wang Y, Wang K, Zhang M, et al. Diagnosis of multiple primary lung cancer: a systematic review. *J Int Med Res.* 2013;41(6):1779–87.
30. Tanvetyanon T, Finley DJ, Fabian T, Riquet M, Voltolini L, Kocaturk C, et al. Prognostic Nomogram to predict survival after surgery for synchronous multiple lung cancers in multiple lobes. *J Thorac Oncol.* 2015;10(2):338–45.
31. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



Is your commute killing you? On the mortality risks of long-distance commuting

Erika Sandow

Department of Geography and Economic History, Umeå University, 90187 Umeå, Sweden;
e-mail: erika.sandow@geography.umu.se

Olle Westerlund

Umeå School of Business and Economics, Umeå University, 90187 Umeå, Sweden; and
Jyväskylä University School of Business and Economics, PO Box 35, FI-40014 University
of Jyväskylä, Finland; e-mail: olle.westerlund@econ.umu.se

Urban Lindgren

Department of Geography and Economic History, Umeå University, 90187 Umeå, Sweden;
e-mail: urban.lindgren@geography.umu.se

Received 28 June 2013; in revised form 2 October 2013

Abstract. There is a general belief that expanding labour-market regions, triggered by increased commuting, have positive economic effects on individuals, firms, and society. Recently, however, scholars have reported possible negative outcomes related to health and well-being. Based on these findings, this study addresses the association between long-distance commuting, and mortality. Using longitudinal individual data from between 1985 and 2008, focusing on 55-year-olds in 1994, we model mortality through propensity score matching and Kaplan–Meyer estimates of survival among long-distance commuters and matched controls from the population travelling short distances to work. The results indicate that women who have experienced long-distance commuting face a significantly higher mortality risk compared with women with short commutes to work. This seems to be driven by variations in income and education: for example, for women with long-distance commuting experience, substantially lower survival rates are found among those with low education and low income. A very different picture emerges for men, for whom mortality risks do not seem to be associated with long-distance commuting. Our findings suggest that men and women are subject to different mechanisms regarding the nexus between commuting and mortality.

Keywords: long-distance commuting, health, mortality, propensity score matching,
survival rates

1 Introduction

Commuting to work over longer distances is an increasingly common phenomenon in most European countries (Hofmeister and Schneider, 2010). While commuting may play an important part in terms of efficient job matching [see Johansson et al, 2002], its potential downsides have received relatively little attention in the scientific literature. In particular, there are few quantitative studies based on larger samples examining the health effects of commuting. This study represents a first attempt to exploit rich longitudinal population data to examine the association between long-distance commuting and mortality.

Many factors drive the increase in long-distance commuting (LDC)—some related to market forces, some to social changes, and some to public policy. Studies using contemporary data show that the importance of localized physical advantages (eg, place-specific raw materials, and proximity to energy such as water power, harbours) for economic activities has decreased substantially (eg, Ellison and Glaeser, 1999). Today, the colocation of firms is

propelled by agglomeration externalities that are nourished by different types of social and economic interactions rather than by physical advantages (eg, Storper and Venables, 2004). However, change in settlement patterns has turned out to be a slow process (eg, Håkansson, 2000)—regarding labour mobility, this is especially the case in Europe compared with the US (Crescenzi et al, 2007)—whereas the geography of jobs has been subject to more rapid transformation. In particular, the employment shift of trade and industry in recent decades from traditional manufacturing to a wide variety of knowledge-intensive production has created new conditions for economic activities. Expanding firms, not only in new sectors but also in innovation systems where firms interact across various sectors (business transactions, labour flows, etc), tend to be located in metropolitan areas and regional centres (eg, Power and Lundmark, 2004). In these regions, agglomeration economies are triggered by the benefits of colocation that enable firms to reduce transaction costs and expenses associated with accessing and transferring knowledge (eg, Maskell and Malmberg, 2007).

However, these centripetal forces operating on economic activities also imply a gradually increasing deviation between the geography of people and the geography of jobs. This development causes imbalances on regional labour markets as reflected in, for example, high unemployment rates and competence-specific labour shortage. Another possible consequence of the spatial mismatch between jobs and people is increased interregional migration. Nonetheless, empirical studies indicate that interregional migration among the gainfully employed is low and that families seem to be unwilling to relocate (Lundholm, 2009). Immobility may be related to location specificity in human capital (eg, Fischer and Malmberg, 2001; Sjaastad, 1962) and place attachment accumulated through long-term continuity (Hay, 1998), which is built up through a slow-paced population redistribution.

The increased divergence between jobs and people has caused politicians and decision makers to create ‘expanding regional labour markets’ by investing in various types of infrastructure (eg, highways, railways, and airports). Improved infrastructure and new technology will make it possible for people to accept job offers further from their place of residence, which is believed to facilitate the skill-matching needed for sustained economic development (eg, Östh and Lindgren, 2012). Naturally, visions of extended labour-market regions would remain merely visions if not for people changing their mobility behaviour. Statistics for the Swedish case show convincingly that commuting across municipality borders has gradually increased during the past decades (Sandow, 2011). There is also some evidence that the average distance between place of residence and workplace is growing (Amcoff, 2007). Thus, there are more commuters than ever, and they travel longer distances. In spite of the likely aggregate positive economic effects of enhanced skill-matching efforts as a consequence of extended regional labour markets, scholars have started to question whether it is really beneficial to the individuals involved. Admittedly, people who start commuting commonly increase their wages (eg, Manning, 2003), but Friberg (2006) argues that more time spent on commuting reinforces traditional gender relations because it is usually men who engage in long-distance journeys to work. Absent men cannot participate in everyday family activities, which affects the life situation of all members of the family. Other studies have found that commuting might be associated with reduced well-being, manifested in stress symptoms (eg, Costa et al, 1988; Gottholmsmeder et al, 2009; Kluger, 1998; Koslowsky et al, 1996; Novaco et al, 1991; Stutzer and Frey, 2008) and illnesses (eg, Hansson et al, 2011; Karlström and Isacsson, 2009).

Based on findings in numerous studies, there seem to be rather clear indications of health-related consequences of commuting. Despite the good merits of these studies, however, they are generally based on data stemming from various types of surveys and interviews that may be regarded as small datasets. Relatively few studies have utilized register data covering entire populations. Moreover, there is a great variety of effect dimensions

(eg, strain, stress, cardiovascular diseases, perceived health situation, subjective well-being, illnesses, or sick leave) in these studies. On the basis of these results, it might be difficult to make an unambiguous overall interpretation of the consequences of commuting on health. As an alternative approach, we propose the use of an indicator that reflects the ultimate consequence of ill health: death. To the best of our knowledge, there is no study focusing on the potential effects of LDC on mortality. Moreover, the use of long-term longitudinal register data covering the entire nation over more than twenty years allows access to a relatively long period with retrospective information as well as a long follow-up period. Register data are particularly suitable for this purpose because they combine a longitudinal trait with the potential to observe rare events. The mortality rate among people not having reached old age is low, especially when the group of interest is narrowed down to individuals with a history of being long-distance commuters.

The aim of this study is to assess the association between LDC and mortality. We stratify the analyses by looking more closely at differences between men and women, and different socioeconomic positions. By focusing on long-distance commuters aged 55 years in 1994, and matched controls drawn from the group of short-distance commuters, we follow the individuals until 2008 and observe their survival during this period. The structure of the paper is as follows. In the next section we provide a theoretical discussion and review some literature focusing on commuting and its consequences for people. In section 3 data and methods are presented. The results of the empirical analyses are laid out in section 4, which is followed by a summary and discussion in section 5.

2 Commuting and health—a background

Theoretically, commuting distance may be perceived as the outcome of sequential or simultaneous choices regarding place of residence and outcomes of job matching. In economic theory, individual choice maximizes subjective utility, given individual preferences and a limited choice set. Differences in preferences and choice sets will produce differences in disutility from commuting and required compensations for commuting time. Possible negative effects on health do not necessarily imply policies counteracting LDC, since benefits of commuting have to be taken into account. Also, LDC may be optimal and health risks correctly discounted by the individuals, although this would require very strong assumptions.

The possible effect of LDC on mortality is an empirical question. There is no lack of arguments regarding why commuting could have negative effects on health. But the alternative to accepting LDC could be associated with higher health risks: for example, insecure employment or jobs with inferior work environments. Also, as has been found in previous research, some commutes are associated with increased utility from the commute itself. Therefore, the potential effect of LDC on mortality is indeterminate on theoretical grounds. However, as indicated by the review of previous research in the next section, there is a clear preponderance of studies pointing at negative effects on health and subjective indicators of well-being.

2.1 Commuting—a stress factor

Commuting has been found to be a major cause of stress impacting on the individual's physical and psychological health and well-being (Evans et al, 2002; Gottholmseder et al, 2009; Hansson et al, 2011; Kluger, 1998). Commuters report experiencing more negative stress than noncommuters and stress levels also increase with age (Gottholmseder et al, 2009; Rüger and Ruppenthal, 2010). Longer commutes are also associated with other negative health outcomes, such as higher blood pressure, obesity, poor sleep quality, fatigue and low self-rated health (Hansson et al, 2011; Hoehner et al, 2012). These indirect health effects of

commuting to work are all prognostic for heart disease, diabetes, and some forms of cancer (Hoehner et al, 2012), and thus also for the mortality risk.

2.2 Commuting costs and life satisfaction

From an urban economic perspective, these perceived commuting burdens and extra monetary costs (eg, travel expenses) are compensated for by a career or financially rewarding job, lower rents, and/or affordable housing prices (eg, Renkow and Hower, 2000; van Ommeren and Rietveld, 2005), or desired housing or neighbourhood characteristics (Plaut, 2006), thus equalizing the commuter's well-being or utility. But, according to some studies, long-distance commuters are not compensated for the stress they experience, and therefore on average have lower life satisfaction than noncommuters (Fults, 2010; Stutzer and Frey, 2008). For example, Stutzer and Frey found that German long-distance (one hour or more) commuters would have to make 40% more in salary to be as satisfied as those with shorter commutes.

While on average stress decreases one's satisfaction with a commute (Abou-Zeid and Ben-Akiva, 2011), commuting in the morning is found to be experienced as especially unpleasant and affects subjective well-being negatively (Kahneman and Kruger, 2006). Having experience of commuting makes it more predictable, and has been found to make the commute less stressful (Kluger, 1998) and to lower the experienced negative subjective life satisfaction due to time pressure (Rüger and Ruppenthal, 2010). This may imply a customization process whereby years of LDC have made it easier to cope with consequences for the daily organization of life and one's health (Rüger and Ruppenthal, 2010).

2.3 The importance of contextual factors

Perceived commuting stress has been found to vary depending on contextual factors, such as the traffic situation and the public transportation system available. The lack of control over and the unpredictability of the commute to work, such as traffic congestion, the behaviour of other drivers, or unreliability of public transport services, increase the perceived stress of commuting (Evans et al, 2002; Gottholmseder et al, 2009; Koslowsky et al, 1996, Kluger, 1998). Active commuting by walking or cycling is related to less negative health outcomes (Hansson et al, 2011) and more satisfaction with the commute (Olsson et al, 2013) than commuting by car, train, or bus. Commuting by car is by far the most common way of commuting in Sweden (62%), while the use of public transportation accounts for only 14% (SIKA, 2007). In a Swedish study (Hansson et al, 2011) the negative health effect of public transport users was found to increase with journey time. However, those commuting by car for 30–60 minutes experienced poorer health than those commuting more than one hour. One possible explanation for this discrepancy between car commuters and public transport users could be that LDC by car in Sweden for the most part does not imply more than one hour in rush-hour traffic. It is more likely that these commutes involve driving in the countryside, which can even offer a positive utility of commuting (see subsection 2.7).

2.4 Various health consequences across genders

Some studies have found significant effects of gender on commuting stress, and report women to be more negatively affected by commuting stress than men (Koslowsky et al, 1996; Novaco et al, 1991). In a study by Roberts et al (2011) in the UK it was found that women were more sensitive to commuting stress than men, even when working hours and occupations were controlled for, and that the stress came from a gendered division of everyday household-related activities, including childcare. Other studies have also found evidence that this stress is a result of too many duties, of which caring for children is the most stressful (Collet and Dauber, 2010). On the other hand, analyses by Gottholmseder et al (2009) on long-distance commuters' stress perception in Austria did not find any significant gender differences, but did find that commuters living in a partnership could handle commuter stress better than

single commuters could. Fults (2010) found that both male and female commuters in Sweden experienced disutility of longer commuting times, affecting their subjective well-being, but if there were young children in the household the disutility was higher for female commuters than their male counterparts.

2.5 Balancing home, family, and long-distance commuting

It is not always unproblematic for commuters to balance work and everyday life when commuting long distances every day. Having a lengthy commute decreases the time available for other daily activities, and can mean insufficient energy and/or time to combine work and family life adequately. Long commuting times to work (over 60 minutes one way) have also been found to be significantly correlated with elevated levels of work–family-life conflict, which itself has important health implications in terms of various physical and mental health problems (Hämming et al, 2009; Jansen et al, 2003). In cross-sectional study in Switzerland, Hämming et al (2009) found a significant correlation between LDC and work–family-life conflict, with associated negative health outcomes for both women and men. However, a Dutch cohort study (Jansen et al, 2003) found this correlation to be only significant in women.

2.6 Indirect health consequences

Other studies also report health outcomes not related to the commuting situation itself but, rather, associated with behaviour patterns that over time may contribute to obesity and other poor health outcomes. As the day is constrained to 24 hours, holding working time constant, increases in commuting time must necessarily mean that one has to trade off this time against other activities. According to a recent American study (Christian, 2012), these trades-offs often result in less engagement in health-related activities. While the greatest share of commuting times comes from sleep reduction, time is also taken from physical activity and food preparation. Consequently, indirect health effects of spending long periods on the road to work are related to higher weight, lower fitness levels, and higher blood pressure, all of which are prognostic for heart disease, diabetes, and some cancers (Hoehner et al, 2012).

That increased absenteeism from work due to sickness is also related to LDC, has been found in Sweden (Hansson et al, 2011) and elsewhere in Europe (Costa et al, 1988; Kluger, 1998). The relationship is not clear or unambiguous, however. The increased sickness absence from work among Swedish long-distance (one hour or more) commuters involves few or short periods of sick leave (Hansson et al, 2011), and being on sick leave for a longer period of time (>15 days) is associated with lower income groups, particularly women (Karlström and Isacsson, 2009).

2.7 Positive aspects of the commute

One line of research focuses on the positive utility of commuting for the individual's personal well-being and on how long commutes can constitute a source of emotional benefit for the commuter him or herself. In contrast to the conventional view that commuting time is a source of disutility that must be minimized, in these studies (eg, Lyons and Urry, 2005; Mokhtarian and Salomon, 2001) it is argued that activities that can be conducted while travelling can add a positive utility to the commute. Examples include reading, listening to music or relaxing, mentally shifting between one's work and home, or using the time to work with modern information technology. Commuters who actively make use of their commuting time also report feeling healthier and less stressed, and experience less disutility from the commute compared with those reporting not making active use of the commuting time (Gottholmseder et al, 2009; Lyons and Urry; 2005; Ory and Mokhtarian, 2005). However, not all long-distance commuters have the opportunity to actively make use of their commuting time. For instance, driving a car does not provide choices to read, work on your laptop, or sleep while commuting, unlike to commuting by public transport.

It has also been argued that the commute itself can be an activity desired for its own sake (Mokhtarian and Salomon, 2001; Mokhtarian et al, 2001), due to factors such as enjoyment of the environment or the speed, or the commute providing an opportunity for driving a car, which projects high social status. Despite various ways of adding potential positive utility to a long commute, the willingness to commute still decreases when commuting times become too long. The majority of commuters would also prefer to decrease their commuting, regardless of the mode used (Páez and Whalen, 2010; Redmond and Mokhtarian, 2001; Sandow and Westin, 2010).

Accordingly, various factors have an influence on long-distance commuters' physical and psychological health, and thus also their mortality risk. Regardless of the motives behind LDC, the present situation on the labour market is that more people than ever are commuting long distances to work and the long-term health consequences, in terms of mortality, are still unknown.

3 Empirical analysis

3.1 Data and design of the study

The analyses are based on longitudinal register data on the entire Swedish population from the ASTRID database, compiled by Statistics Sweden and hosted by the Department of Geography and Economic History at Umeå University. The data contain yearly detailed individual information on socioeconomic and demographic variables, such as family situation, income, education level, and employment status, and are georeferenced with coordinates for place of residence and work at a resolution of 100 m. This information is available for the period 1985 to 2008.

The sample is conditioned to include individuals who in 1994 were aged 55 years and active on the labour market (were employed and had income from work). By choosing this cohort we obtain both relatively long historical and follow-up periods. The age restriction was due to the extremely low mortality in younger cohorts.

Individuals are defined as 'treated' if they were long-distance commuters in 1993 and/or 1994. The 'untreated' individuals in the control group consist of those who did not long-distance commute during 1985–2008. Here, LDC is defined as traveling at least 50 km to work 'one way'. This commuting distance is measured as the Euclidean distance between the coordinates of home and work. Thus it does not measure the actual distance, which is about 30% longer, but is likely to be equivalent to 55–70 km on the ground and correspond to an average of 60 minutes by car in Sweden. While there is no consensus in research on what constitutes a long commuting distance (see overview in Sandow, 2011), a time threshold of around 45 minutes one way has been found to constitute an acceptable daily travel time both nationally and internationally (Sandow and Westin, 2010; van Ham, 2001; van Ommeren, 1996; Wachs et al, 1993). Consequently, the commuters in this study all have in common that they have accepted daily commuting times well above the average and at a level implying considerable disutility as reflected by actual commuting patterns. Unfortunately, the data do not include direct information on travel mode or costs. It is unlikely that people choose an active commuting mode for daily commuting distances exceeding 50 km, which is the focus of this study. However, the dataset at hand provides a comparatively useful proxy for nonobservable characteristics related to lifestyle and attitudes towards taking daily exercise keeping the individual in good physical shape. For the year 1990, prior to the time period over which we observe LDC (1993–94), there is information about mode of travel to work. This variable distinguishes between individuals biking or walking to work compared with those commuting by car or public transports. The variable 'active commuting 1990' is included in the analysis to control for potential health differences between the group of long-distance commuters and the group of people without experience of LDC.

The distribution of job openings varies significantly across different regions. To control for regional diversity regarding the labour market situation, we included a measure of interregional accessibility to employment opportunities ‘access job openings’ in line with the indicator proposed by Eliasson et al (2003).

The total sample size is fairly large, encompassing more than 2700 long-distance commuters (1961 men and 775 women) and 56 800 potential controls (27 462 men and 29 373 women).

3.2 Method

The potential effect of commuting on mortality is estimated using propensity score matching (Rosenbaum and Rubin, 1983), and Kaplan–Meyer estimates of survival for long-distance commuters and matched controls from the population of employed individuals with short commutes to work. For a similar approach see, for example, Stenberg et al (2012). Technically, we estimate the so-called ‘average treatment effect on treated’. However, for reasons discussed below, the results should not be given strong causal interpretations. The sample of long-distance commuters may be nonrandomly selected from the total population of employees. To find a relevant group of controls, the propensity score is estimated as the probability of being a long-distance commuter as a function of individual and regional characteristics. Using nearest-neighbour matching on the propensity score, the group of matched controls is obtained by selecting four individuals from the group of short-distance commuters with the propensity scores closest to those of each long-distance commuter. The outcome of interest is the difference in mortality between long-distance commuters and the matched control group, as indicated by nonparametric estimates of survival functions. Although our data cover a relatively long follow-up period, the analysis includes censored observations: survivors at the last time point of observation. Because of potential differences between the sexes in the underlying process of becoming a long-distance commuter and possible differences in outcomes thereof, the empirical models are estimated separately for men and women.

As is the case with all nonexperimental designs, an unbiased estimate of a causal effect can only be obtained under certain identifying assumptions. In the present context: (i) an individual’s decision to become a long-distance commuter does not affect the mortality of other individuals; (ii) conditional on the covariates used to estimate the propensity score, the probability of being a long-distance commuter is independent of the potential outcome (survival); (iii) conditional on the covariates, the probability of treatment is strictly positive and strictly smaller than 1; and (iv) conditional on the covariates, the censoring mechanism is independent of the outcome.

The variables used for estimating the propensity scores are measures of the following attributes: the individual’s income; recent income changes and unemployment; educational level; whether previously having chosen an active commuting mode; and variables for family situation (partner, children 0–10 years of age in the household, and whether there were children living in another household but within 10 km of the individual’s own place of residence); access to job openings in neighbouring labour markets; and a set of dummy variables for region of residence, using the metropolitan area of Stockholm as the reference region. The regional division is the NUTS 2 classification (EC, 2003). Some employment sectors employing large numbers of long-distance commuting men or women (construction, manufacturing, real estate, and education) were included as dummy variables, as was the health sector, which employs relatively few long-distance commuters.

The explanatory variables in the logit models for estimation of the propensity score are subsets of these variables. This is because overparameterization increases the mean squared error and may cause problems with common support (Caliendo and Kopeinig, 2008; Waernbaum, 2008). For this reason, covariates associated with *p*-values of 0.2 or above have

been excluded—except for cases where the inclusion of a covariate was necessary to balance the LDC and matched comparison groups.

The empirical analysis is based on rich information, and the chosen method provides several advantages compared with parametric regression models. First, the identification of effects relies on less restrictive parametric assumptions. Second, it provides better control for common support in data: that is, the treated are compared with comparable individuals in all measured attributes. Contrary to parametric regression models, identification relying on extrapolation outside support in data is avoided. Third, in combination with highly informative data, it allows examination of differences in effects between subsamples while upholding control over common support in data. As is the case with all nonexperimental designs, the question of possible remaining selection bias in estimated effects cannot be dismissed (eg, Smith and Todd, 2005). To identify the effects of LDC it is crucial that assumption (ii) above is not violated by unobserved factors that affect both the probability of being a long-distance commuter and mortality. If, for some reason, long-distance commuters have shorter longevity than other employees because of unobserved factors, the estimated effect of LDC on mortality will be biased unless the covariates in the propensity score reflect these factors sufficiently. Except for the indicator ‘active commuting 1990’ we have no direct observations of individuals’ habits/lifestyles affecting health, and we have no direct information on poor work environments or illness/morbidity. Theoretically, there may be systematic differences between long-distance commuters and the control group in these respects. However, they are presumably correlated with the aforementioned travel-mode variable and other observed characteristics such as education, income, branch of employment, individual experience of unemployment, or other confounders included in the estimation of the propensity score. It is only the extent to which the covariates do not capture factors (which cannot be directly observed), and the extent to which the distribution of remaining heterogeneity in these attributes differs systematically between long-distance commuters and the matched comparison group, that would lead to biased estimates. Whether this hypothetical case would introduce a positive or negative bias in our results is an open and untestable question. Some health conditions associated with higher mortality may be an obstacle to LDC, and therefore a potential source of negative bias—that is, an underestimation of the effects on mortality.

Even under the assumption of a close correlation between observed indicators of socioeconomic status and unobserved factors, there is particular concern regarding whether *changes* in unobserved conditions may affect the probability of LDC and also mortality [see Heckman and Smith (1999) for a discussion on changes in pretreatment status]. For example, negative labour-market events may cause both increased stress and an increased probability of searching for new jobs involving long-distance commutes. For this reason, the robustness of our results is also checked by controlling for changes in employment status as well as levels of and the dynamics in earnings prior to commuting.

4 Results

In this section we first present the logit estimates for the propensity score using the full sample of women and men. This is followed by a comparison of sample means for unmatched and matched samples, including balancing tests for the matched samples (ie, tests of equality in sample means of the variables used in the logit model). In a third step the estimated survival rates for long-distance commuters and matched comparisons are presented. This is accompanied by log–rank tests for differences in survival. Finally, we examine potential differences in mortality for matched subsamples by socioeconomic status.

Table 1 displays the logit model estimates of the probability of being a long-distance commuter. In line with expectations, the results indicate a higher probability of LDC for

singles persons, the previously unemployed, and individuals with higher education. Overall, most estimates are strongly significant and the signs of coefficients offer no surprises.

It is essential that the samples of matched controls are comparable with the long-distance commuters in all attributes affecting the probability of LDC. Table 2 gives descriptive

Table 1. Logit estimates for the propensity score.

	Men			Women		
	coefficient	standard error	p-value	coefficient	standard error	p-value
<i>Household characteristics</i>						
Partner	-0.201	0.055	0.000	-0.292	0.079	0.000
Child(ren) aged 0–10 years at home ^a	0.418	0.114	0.000			
Income, SEK ^b	0.739	0.054	0.000	0.028	0.066	0.675
<i>Previous employment situation</i>						
Employed 1990 and 1991, unemployed 1992	0.785	0.148	0.000	0.551	0.237	0.020
Unemployed 1990 and 1991, employed 1992	0.487	0.751	0.517	0.995	0.372	0.007
Income changes (1990–92) ^c	-0.120	0.031	0.000	-0.107	0.059	0.069
Access job openings	0.002	0.000	0.000	0.004	0.001	0.000
Active commuting 1990	-0.850	0.080	0.000	-0.727	0.097	0.000
<i>Education level</i>						
High school education	0.164	0.056	0.003	0.346	0.095	0.000
University education	0.280	0.082	0.001	1.004	0.117	0.000
<i>Employment sector</i>						
Real estate, renting, and business activities	-0.188	0.086	0.030	0.691	0.127	0.000
Public administration and defence; compulsory social security	0.162	0.088	0.065	0.075	0.143	0.598
Construction	0.376	0.078	0.000	0.736	0.283	0.009
Manufacturing	-0.508	0.065	0.000	0.063	0.133	0.637
Health and social work	-0.416	0.134	0.002	-0.571	0.102	0.000
<i>Residential region</i>						
East middle Sweden	0.630	0.116	0.000	0.411	0.188	0.029
Småland and the islands	0.787	0.110	0.000	0.364	0.191	0.057
Southern Sweden	0.660	0.097	0.000	0.663	0.153	0.000
West Sweden	0.626	0.094	0.000	0.643	0.149	0.000
North middle Sweden	0.972	0.102	0.000	0.608	0.174	0.000
Middle Norrland	1.342	0.122	0.000	1.501	0.188	0.000
Upper Norrland	1.385	0.121	0.000	1.693	0.181	0.000
Constant	-12.535	0.667	0.000	-5.122	0.797	0.000
Pseudo R ²				0.0524		

Note. Variables measured in 1994 unless stated otherwise. Dependent variable: LDC in 1993 and/or 1994.

^aThere are no small children living at home for women in this sample (age 55 years)

^bIn 1994, €100 was approximately 920 SEK.

^cIncome changes 1990–92 defined as ‘income 1992 – income 1990’.

Table 2. Descriptive statistics of long-distance commuters and comparisons as measured in 1994, unless stated otherwise.

Variable	Men				Women				
	LDC ^a	no LDC	p-value ^b	matched comparisons	p-value ^b	LDC	no LDC	p-value ^b	
N	1961	27 462			775	29 373			
<i>Household characteristics</i>									
Partner	0.732	0.736	0.683	0.733	0.911	0.647	0.698	0.003	0.656
Child(ren) aged 0–10 years at home ^c	0.049	0.033	0.000	0.044	0.920				0.662
Income, SEK ^d	12.332	12.156	0.000	12.296	0.617	11.856	11.801	0.015	—
<i>Previous employment situation</i>									
Employed 1990 and 1991, unemployed 1992	0.031	0.017	0.025	0.727		0.027	0.015	0.011	0.024
Unemployed 1990 and 1991, employed 1992	0.001	0.001	0.689	0.001	0.651	0.011	0.004	0.011	0.590
Income changes (1990–92) ^d	-0.010	0.023	0.054	0.001	0.653	0.044	0.087	0.034	0.009
Access job openings	169.290	164.090	0.009	169.160	0.414	172.780	162.590	0.001	0.058
Active commuting 1990	0.099	0.199	0.000	0.104	0.707	0.186	0.318	0.000	0.761
<i>Education level</i>									
High school education	0.486	0.459	0.022	0.49232	0.830	0.492	0.531	0.035	0.595
University education	0.178	0.128	0.000	0.16301	0.832	0.247	0.129	0.000	0.595
<i>Employment sector</i>									
Real estate, renting, and business activities	0.096	0.097	0.933	0.099	0.912	0.117	0.059	0.000	0.522
Public administration and defence; compulsory social security	0.097	0.066	0.000	0.094	0.699	0.084	0.073	0.261	0.883
Construction	0.131	0.085	0.000	0.122	0.725	0.019	0.010	0.014	0.681
Manufacturing	0.206	0.310	0.000	0.214	0.555	0.105	0.108	0.754	0.983
Health and social work	0.036	0.041	0.318	0.036	0.791	0.213	0.345	0.000	0.848

Table 2 (continued)

Variable	Men				Women				<i>Residential region</i>
	LDC ^a	no LDC	p-value ^b	matched comparisons	p-value ^b	LDC	no LDC	p-value ^b	
N	1961	27462			775	29373			
East middle Sweden									
	0.204	0.174	0.001	0.198	0.405	0.200	0.168	0.024	0.184
Småland and the islands	0.087	0.093	0.362	0.088	0.758	0.063	0.092	0.006	0.064
Southern Sweden	0.139	0.144	0.493	0.143	0.766	0.149	0.141	0.491	0.708
West Sweden	0.198	0.197	0.908	0.203	0.590	0.227	0.197	0.046	0.154
North middle Sweden	0.107	0.096	0.112	0.109	0.536	0.079	0.097	0.096	0.224
Middle Norrland	0.066	0.042	0.000	0.056	0.381	0.069	0.046	0.003	0.083
Upper Norrland	0.076	0.052	0.000	0.025	0.727	0.092	0.057	0.000	0.942
									0.917
									0.908

^aLDC—long-distance commuting.^bt-test of equality in means between long-distance commuters and comparison groups.^cThere are no small children living at home for women in this sample (age 55 years).^dFor 1994 €100 was approximately 920 SEK.

averages for long-distance commuters, the full sample of non-long-distance commuters, and the matched sample of controls. The *p*-values indicate significance levels of differences in sample means. There are clear indications of systematic selection into LDC of, for example, the previously unemployed and individuals with higher education. However, matching on the propensity scores yields matched samples that are comparable in all dimensions measured by the covariates in the matching process. All estimates of differences in survival presented in this section are based on matched samples and the covariate selection follows the criteria stated above. All covariates in the logit model pass the balancing test based on *t*-tests of differences in means.

The estimated survival rates of matched samples are displayed in figures 1 and 2. For males, the survival rates do not differ between long-distance commuters and matched comparisons in the short term. After six years of follow-up and beyond, the estimated survival rates are lower for the sample of long-distance commuters. However, the log-rank test for difference in survival over the entire follow-up period indicates no significant difference in survival.

Turning to the sample of women, the estimated survival rates are consistently lower for long-distance commuters and the log-rank test indicates that the null hypothesis of no difference in survival between groups can be rejected. Table 3 gives hazard ratios based on Nelson–Aalen estimates of cumulative hazards (not reported in the table). At the end of the fourteen-year follow-up, the estimated cumulative hazard (cumulative risk of death) was 0.0967 (*SE* = 0.019) among long-distance commuters and 0.0665 (*SE* = 0.0051) among matched comparisons, which means a hazard ratio of 1.43. The estimated risk of dying within 14 years of follow-up is 43% higher for long-distance commuters. Although the hazard ratio is relatively high, mortality is low as indicated by the survival functions and the cumulative hazards at the end of follow-up.

The use of alternative specifications of the propensity score does not change the main picture: we find no strong evidence of a significant association between LDC and mortality for the full samples of men, but consistent indications of lower survival rates among women who commute long distances.

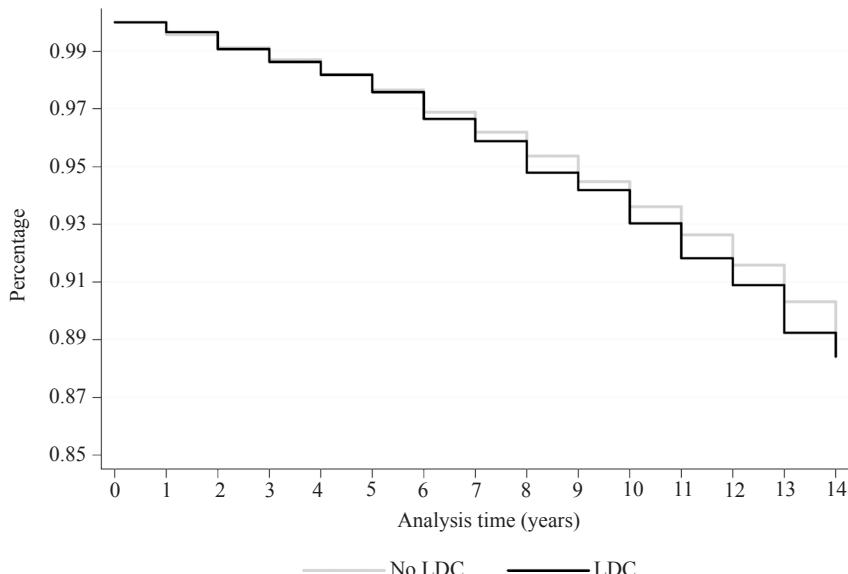


Figure 1. Survival rates, men 1995–2008. Matched samples of long-distance commuting (LDC) men and controls (log-rank test for equality of survivor functions: *p*-value = 0.3933, χ^2 = 0.73).

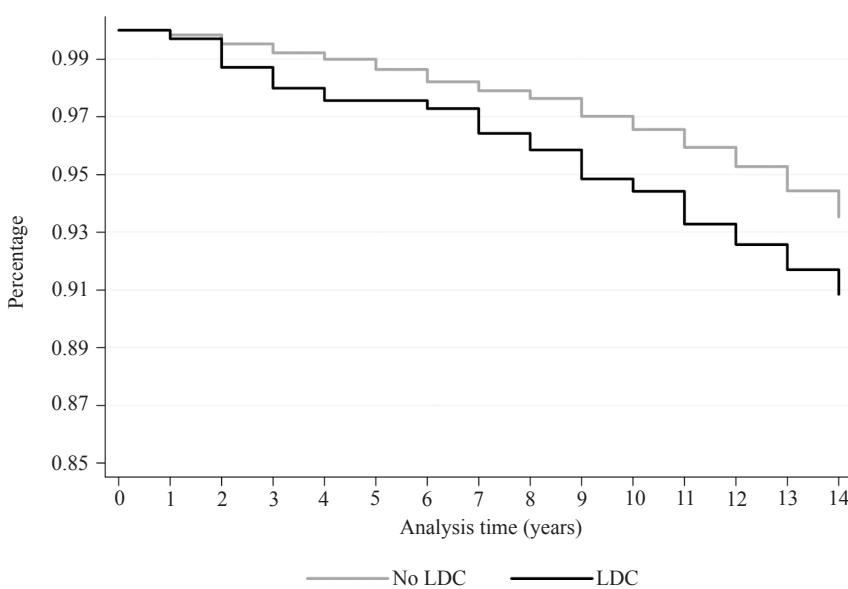


Figure 2. Survival rates, women 1995–2008. Matched samples of long-distance commuting (LDC) women and controls (log–rank test for equality of survivor functions: p -value = 0.0118, χ^2 = 6.34).

Table 3. Hazard ratios of mortality risk between long-distance commuting (LDC) and non-LDC women.

Time (years)	All	Low incomes	Low education
1	1.933	1.222	1.946
2	2.804	4.296	1.419
3	2.577	3.064	1.230
4	2.426	2.942	1.361
5 ^a	na	na	na
6	1.531	1.908	1.210
7	1.716	2.034	1.344
8	1.773	2.071	1.672
9	1.742	1.982	1.742
10	1.636	1.676	1.711
11	1.676	1.801	1.783
12	1.595	1.677	1.728
13	1.508	1.578	1.696
14 ^b	1.436	1.539	1.663

^ana—not applicable; there were no deaths among long-distance commuters in year 5 of follow-up.

^bThe underlying Nelson–Aalen point estimates of cumulative hazards for treated and untreated at the last year of follow-up are significantly different at the 10% level. The overall test of difference in survival rates are provided by the log–rank test reported in the corresponding figures 2–4.

4.1 Survival by socioeconomic status

As shown in previous studies on commuting and health, there is strong reason to expect differences in potential effects on mortality with respect to social/socioeconomic context. Estimation results based on subsamples defined by education and income levels indicate statistically significant associations between LDC and survival rates for the samples of women with low income (below median) and of women with low education (less than high school). The point estimates for women with low income (figure 3) indicate differences

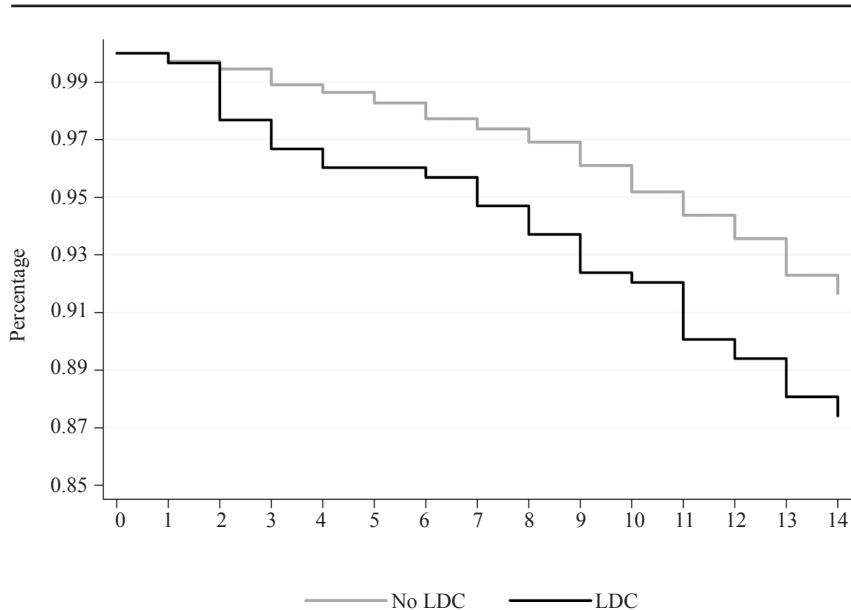


Figure 3. Survival rates for women with low incomes, 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.0203, $\chi^2 = 5.39$; ‘low income’ = annual income below median).

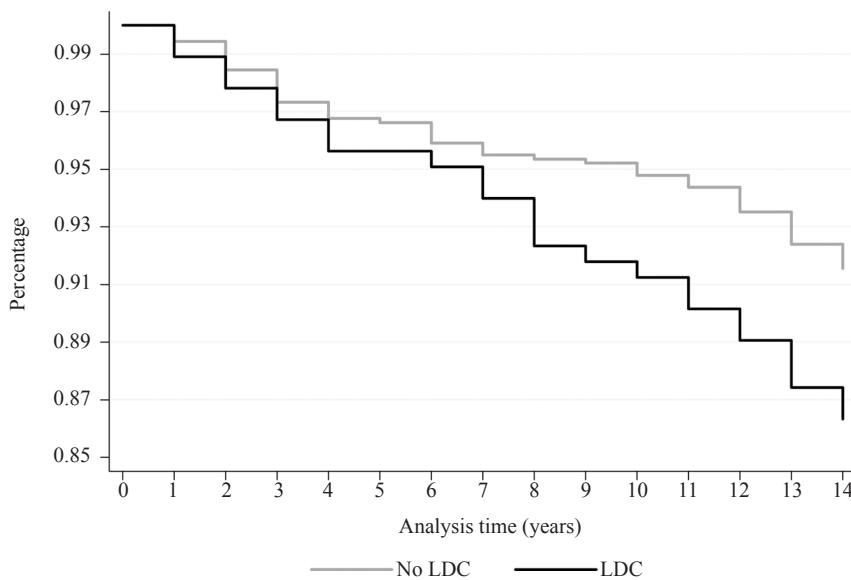


Figure 4. Survival rates for women with low education, 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.0310, $\chi^2 = 4.65$; ‘low education level’ = completed less than high school).

in survival rates amounting to about 1 to 2 percentage points up to around eight years of follow-up, increasing in the following years to approximately 3 to 5 percentage points. At the end of follow-up, the estimated cumulative hazard is 0.1336 (SE = 0.0217) for long-distance commuters and 0.0868 (SE = 0.009) for the matched control group. The hazard ratio indicates a 54% higher risk for long-distance commuter of dying within 14 years compared with matched controls (table 3). Higher mortality is also indicated for long-distance commuters in the sample of women with low education (figure 4). The ratio of cumulative hazards at the end of follow-up is 1.66 (table 3). At that point, the estimated cumulative hazard is 0.146 (SE = 0.0292) for long-distance commuters and 0.0878 for matched controls. Again, the

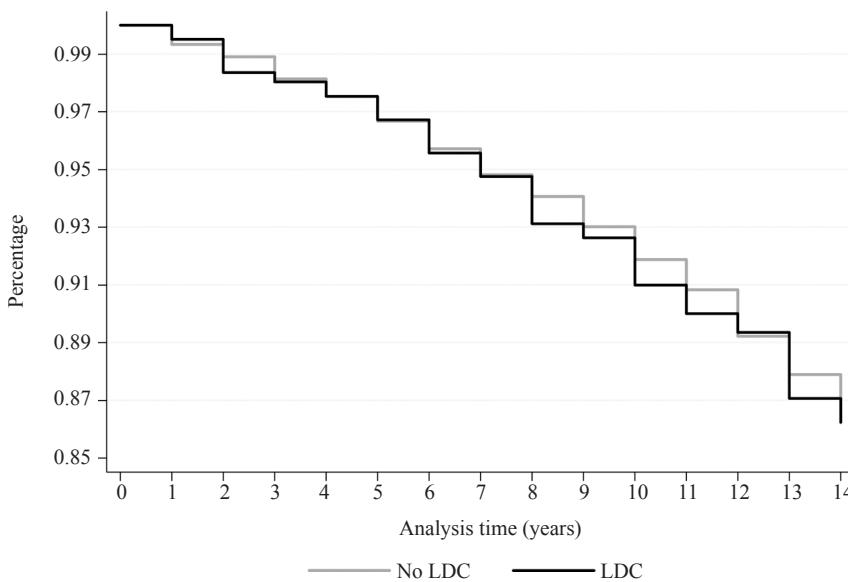


Figure 5. Survival rates for men with low income, 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.9823, $\chi^2 = 0.00$; ‘low income’ = annual income below median).

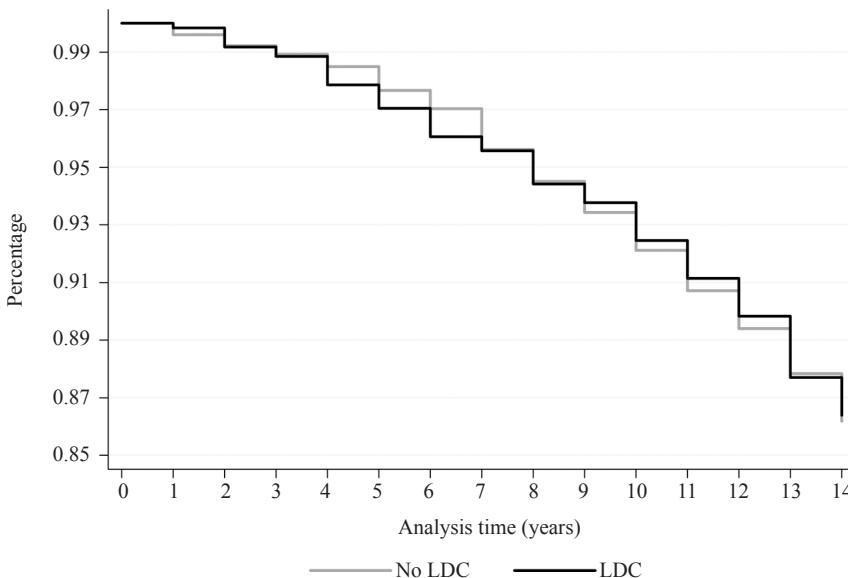


Figure 6. Survival rates for men with low education, 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.9069, $\chi^2 = 0.01$; ‘low education level’ = completed less than high school).

hazard ratio should be viewed against the backdrop of the overall low mortality as displayed by the survival functions and cumulative hazard rates.

The results for men are very different from those for the corresponding subsamples of women (figures 5 and 6): there are no significant differences in estimated survival rates between the long-distance commuters and the comparison groups. This applies for the full sample of men as well as for subsamples of men with low income and men with low education. Figures 7–10 show survival rates by level of education.

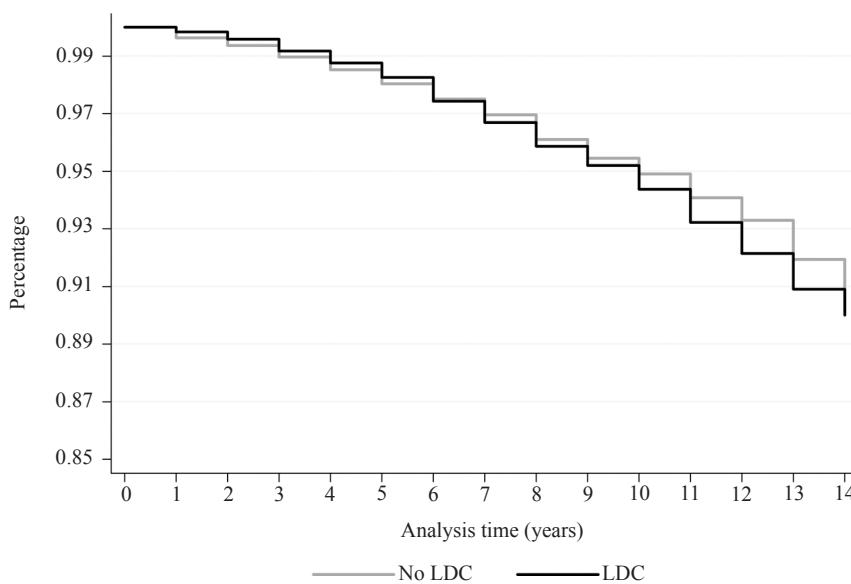


Figure 7. Survival rates for men with high income, 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.4494, $\chi^2 = 0.57$; ‘high income = annual income above median’).

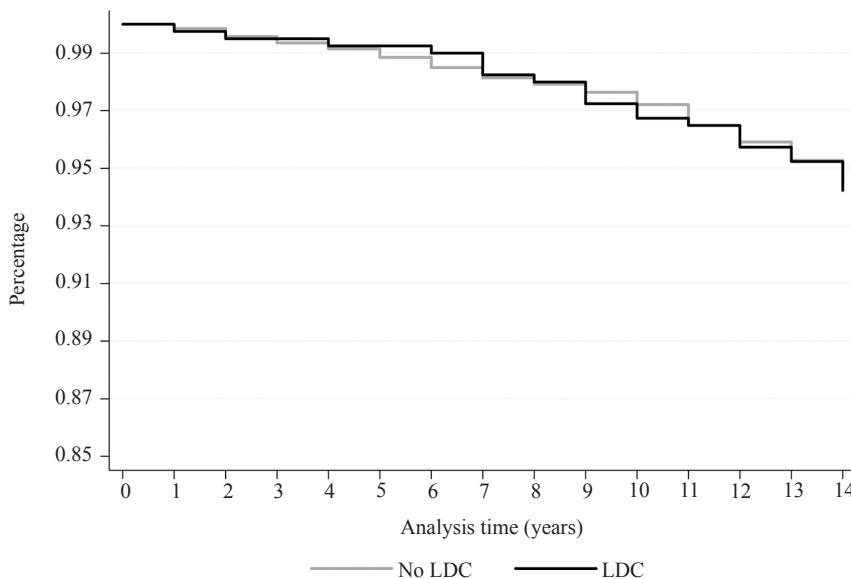


Figure 8. Survival rates for women with high income, 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.9366, $\chi^2 = 0.01$; ‘high income = annual income above median’).

4.2 Robustness checks

The general message of the results presented thus far is that the estimates indicate increased mortality among females who commute long distances, but not for men. For women the results seem to be driven by lower survival rates among long-distance commuters with lower socioeconomic status as measured by income and education. These results are stable with respect to alternative specifications of the logit model for the propensity score, and increasing the trimming of the tails of the distribution of the propensity score from 5% to 7%. Outliers in terms of differences in mortality do not drive our results at ‘young age’. Removing people passing away in the first or second year of follow-up (aged 56 or 57) does not affect

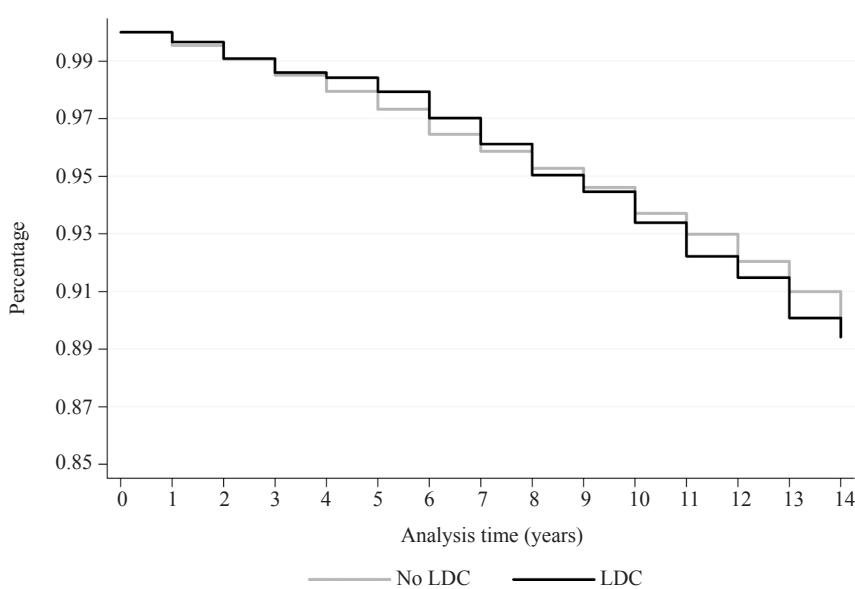


Figure 9. Survival rates for men with high education , 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.7724, χ^2 = 0.08; ‘high education level = completed at least high school’).

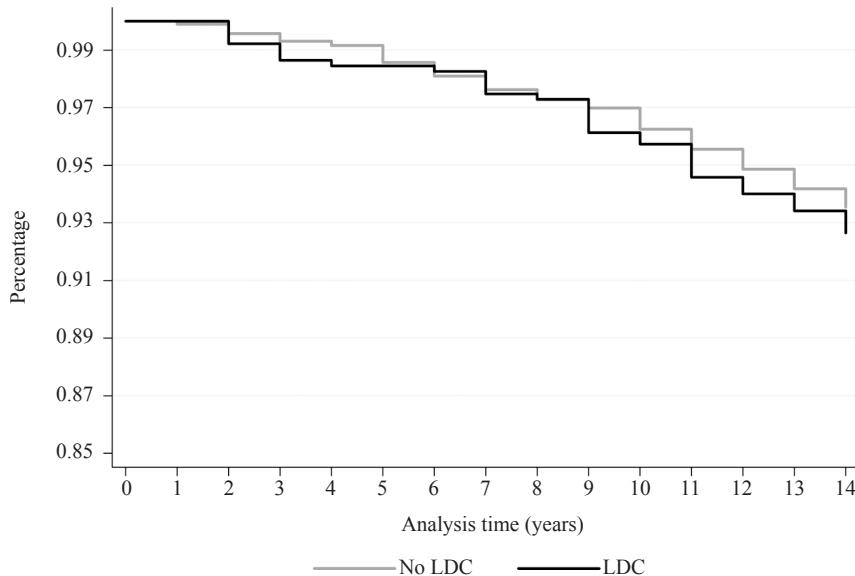


Figure 10. Survival rates for women with high education , 1995–2008. Matched samples of treated and controls (LDC = long-distance commuting; log-rank test for equality of survivor functions: p -value = 0.4627, χ^2 = 0.54; ‘high education level = completed at least high school’).

our main conclusions. This also applies for using the Wilcoxon test of statistical significance as an alternative to the log-rank test.

5 Summary and discussion

Long commutes have been found to have negative health consequences for many commuters, but the long-term health effects of commuting are still unknown. To measure the long-term health implications of LDC, we have analysed the association between LDC and mortality. Using annual register data, a large sample of workers aged 55 years in 1994 have been followed over a period of 14 years (1985–2008). The difference in mortality between LDC

and matched controls has been estimated using propensity score matching and nonparametric estimates of survival.

Conditional on a large set of confounders, the results indicate a statistically significant higher mortality among women with long-distance commutes. Specifically, the estimated association between LDC and mortality for the full sample of women seems to be confined to individuals with lower educational attainment and low incomes. There is no evidence of differences in survival rates between long-distance commuters and matched controls in the corresponding subsamples of men.

The gender differences in mortality risk support results from previous studies suggesting that negative health effects due to commuting are more evident for women than men (eg, Karlström and Isacsson, 2009). Possible explanations for this difference by gender may be related to still existing gender roles in the household. As gender expectations and structural constraints about breadwinning, parenthood, care taking of elderly parents, and other family matters still prevail, women face a double burden: retaining the main responsibility for household-related activities while also being expected to be employed (Coltrane, 2000; William, 2000). This double burden may cause the negative effects of LDC to be worse for women than men. We know from previous studies that female long-distance commuters experience more time pressure and stress induced by commuting than male long-distance commuters, and that this stress is a result of a gendered division of household-related activities (Collet and Dauber, 2010; Roberts et al, 2011). That LDC is particularly related to lower survival rates among women with a lower income or education level may then reflect a high gender inequality in housework. Findings from Sweden, the Netherlands, and the US (Evertsson et al, 2009) show that less educated women spend more time on household work than do more highly educated women, and that this inequality decreases with the level of education—which may reflect more equalitarian work–family arrangements among the well educated. In addition, low education and low income are presumably associated with relatively less flexibility regarding work arrangements (eg, presence at the workplace), which may exacerbate stress from commuting. This, combined with a relatively larger share of household responsibilities, could be an underlying mechanism for negative long-term health consequences, and a higher risk of mortality for women with a lower education or income levels.

Nonexperimental approaches always entail some uncertainty concerning the extent to which the results are influenced by selection bias. In the absence of data from randomized experiments, using rich longitudinal population data still offers excellent possibilities to study the potential health effects of LDC. Although our findings may be interpreted as *indications* of negative effects for women with low socioeconomic status and an absence of effects in corresponding samples of men, they should not be given strong interpretations in terms of causal effects. These results should, rather, inspire further research on this highly relevant and heavily underresearched issue. Adding direct observation of health status and information on causes of death in a large population data setting would allow a substantial step towards causal interpretations of (potentially) statistically significant results. Furthermore, the lack of transit mode data is unfortunate and future research on this subject could focus on possible differences in mortality risk depending on mode of transport. Additional research is also required for statements on possible policy implications. It can provide an understanding of how various socioeconomic factors interact with the long-term effects of commuting, which can be used to identify and mitigate potential negative health outcomes.

Acknowledgements. Financial support was granted by the Swedish Research Council (VR) via the Umeå SIMSAM Network (2008-7491), the Ageing and Living Conditions Programme (ALC), and the Swedish Council for Working Life and Social Research (FAS) (2006–1010).

References

- Abou-Zeid M, Ben-Akiva M, 2011, "The effects of social comparison on commute well-being" *Transportation Research Part A* **45** 345–361
- Amcoff J, 2007 *Regionförstöring—idé, mätparproblem och framtidsutsikter* Institutet för Framtidsstudier, Arbetsrapport 2007:7
- Caliendo M, Kopeinig S, 2008, "Some practical guidance for the implementation of propensity score matching" *Journal of Economic Surveys* **22** 31–72
- Christian T J, 2012, "Trade-offs between commuting time and health-related activities" *Journal of Urban Health* **89** 746757
- Collet B, Dauber A, 2010, "Gender and job mobility", in *Mobile Living Across Europe II. Causes and Consequences of Job-related Spatial Mobility in Cross-national Perspective* Eds N Schneider, B Collet (Barbara Budrich, Oplanden) pp 173–194
- Coltrane S, 2000, "Research on household labor: modeling and measuring the social embeddedness of routine family work" *Journal of Marriage and Family* **62** 1208–1233
- Costa G, Pickup L, Di Martino V, 1988, "Commuting a further stress factor for the working people, evidence from the European Country II. An empirical study" *International Archive of Occupational and Environmental Health* **60** 377–385
- Crescenzi R, Rodríguez-Pose A, Storper M, 2007, "The territorial dynamics of innovation: a Europe–United States comparative analysis" *Journal of Economic Geography* **7** 673–709
- EC, 2003, "Regulation (EC) No 1059/2003 Common classification of territorial units for statistical purposes" *Official Journal of the European Communities* **L 154** 21 May (European Commission, Brussels)
- Eliasson K, Lindgren U, Westerlund O, 2003, "Geographical mobility: migration or commuting?" *Regional Studies* **37** 827–837
- Ellison G, Glaeser E, 1999, "The geographic concentration of an industry: does natural advantage explain agglomeration?" *American Economic Association Papers and Proceedings* **89** 311–316
- Evans G W, Wener R E, Phillips D, 2002, "The morning rush hour: predictability and commuter stress" *Environment and Behaviour* **34** 521–530
- Evertsson M, England P, Mooi-Reci I, HermSEN J, De Bruijn J, Cotter D, 2009, "Is gender inequality greater at lower or higher educational levels? Common patterns in the Netherlands, Sweden and the United States" *Social Politics* **16** 210–241
- Fischer P A, Malmberg G, 2001, "Settled people don't move: on life course and (im-)mobility in Sweden" *International Journal of Population Geography* **7** 357–371
- Friberg T, 2006, "Kommunala utmaningar och genus. Om regionförstöring, pendling, produktion och reproduktion" in *Kommunledning och samhällsförändring* Ed. L Jonsson (Studentlitteratur, Lund) pp 169–194
- Fults K K, 2010 *A Time Perspective on Gendered Travel Differences in Sweden* Licentiate thesis, Department of Transport and Economics, Royal Institute of Technology, Stockholm
- Gottholmseder G, Nowotny K, Pruckner G J Theurl E, 2009, "Stress perception and commuting" *Health Economics* **18** 559–576
- Håkansson J, 2000 *Changing Population Distribution in Sweden—Long Term Trends and Contemporary Tendencies* PhD thesis, GERUM Kulturgeografi, Umeå University
- Hämming O, Gutzwiler F, Bauer G, 2009, "Work–life conflict and associations with work- and nonwork-related factors and with physical and mental health outcomes: a nationally representative cross-sectional study in Switzerland" *BMC Public Health* **9** 435
- Hansson E, Mattisson K, Björk K, Östergren P O Jakobsson K, 2011, "Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden" *BMC Public Health* **11** 834
- Hay R, 1998, "Sense of place in developmental context" *Journal of Environmental Psychology* **18** 5–29
- Heckman J, Smith J, 1999, "The pre-programme earnings dip and the determinants of participation in a social programme. Implications for simple programme evaluation strategies" *Economic Journal* **109** 313–348

-
- Hoehner C M, Carolyn E B, Barlow M S, Allen P, Schootman M, 2012, "Commuting distance, cardiorespiratory fitness, and metabolic risk" *American Journal of Preventive Medicine* **42** 571–578
- Hofmeister H, Schneider N F, 2010, "Job mobilities in Europe: core findings, policy implications future outlook", in *Mobile Living Across Europe II. Causes and Consequences of Job-related Spatial Mobility in Cross-national Perspective* Eds N Schneider, B Collet (Barbara Budrich, Oplanden) pp 337–356
- Jansen N W H, Kant I J, Kristensen T J, Nijhuis F J N, 2003, "Antecedents and consequences of work–family conflict: a prospective cohort study" *Journal of Environmental Medicine* **45** 479–491
- Johansson B, Klaesson J, Olsson M, 2002, "Time distances and labour market integration" *Papers in Regional Science* **81** 305–327
- Kahneman D, Krueger A B, 2006, "Developments in the measurements of subjective well-being" *The Journal of Economic Perspectives* **20** 3–24
- Karlström A, Isacsson G, 2009, "Is sick absence related to commuting travel time? Swedish evidence based on the generalized propensity score estimator", WP 2010:3, Swedish National Road and Transport Research Institute (VTI), Stockholm
- Kluger A N, 1998, "Commute variability and strain" *Journal of Organizational Behaviour* **19** 147–165
- Koslowsky M, Aizer A, Krausz M, 1996, "Stressor and personal variables in the commuting experience" *International Journal of Manpower* **17**(3) 4–14
- Lundholm E, 2009, "Interregional migration propensity and labour market size in Sweden, 1970–2001" *Regional Studies* **44** 455–464
- Lyons G, Urry J, 2005, "Travel time use in the information age" *Transportation Research Part A* **39** 257–276
- Manning A, 2003, "The real thin theory: monopsony in modern labour markets" *Labour Economics* **10** 105–131
- Maskell P, Malmberg A, 2007, "Myopia, knowledge development and cluster evolution" *Journal of Economic Geography* **7** 603–618
- Mokhtarian P, Salomon I, 2001, "How derived is the demand of travel? Some conceptual and measurement considerations" *Transportation Research Part A* **35** 695–719
- Mokhtarian P, Salomon I, Redmond L, 2001, "Understanding the demand for travel: it's not purely 'derived'" *Innovation* **14** 355–380
- Novaco R W, Kliewer W, Broquet A, 1991, "Home environmental consequences of commute travel impedance" *American Journal of Community Psychology* **19** 881–909
- Olsson L E, Gärling T, Ettema D, Friman M, Fujii S, 2013, "Happiness and satisfaction with work commute" *Social Indicators Research* **111** 255–263
- Ory D T, Mokhtarian P L, 2005, "When is getting there half the fun? Modeling the liking of travel" *Transportation Research Part A* **2–3** 97–123
- Östh J, Lindgren U, 2012, "Do changes in GDP influence commuting distances? A study of Swedish commuting patterns between 1990 and 2006" *Tijdschrift voor Economicsche en Sociale Geografie* **103** 443–456
- Páez A, Whalen K, 2010, "Enjoyment of commute: a comparison of different transportation modes" *Transportation Research Part A: Policy and Practice* **44** 537–554
- Plaut P, 2006, "The intra-household choices regarding commuting and housing" *Transportation Research Part A* **40** 561–571
- Power D, Lundmark M, 2004, "Working through knowledge pools: labour market dynamics, the transference of knowledge and ideas, and industrial clusters" *Urban Studies* **41** 1025–1044
- Redmond L, Mokhtarian P, 2001, "The positive utility of the commute: modeling ideal commuting time and relative desired commute amount" *Transportation* **28** 179–205
- Renkow M, Hower D, 2000, "Commuting, migration and rural–urban population dynamics" *Journal of Regional Science* **40** 261–287
- Roberts J, Hodgson R, Dolan P, 2011, "It's driving her mad: gender differences in the effects of commuting on psychological well-being" *Journal of Health Economics* **30** 1064–1076
- Rosenbaum P R, Rubin D B, 1983, "The central role of the propensity score in observational studies for causal effects" *Biometrika* **70** 41–55

-
- Rüger H, Ruppenthal S, 2010, "Advantages and disadvantages of job-related spatial mobility", in *Mobile Living Across Europe II. Causes and Consequences of Job-related Spatial Mobility in Cross-national Perspective* Eds N Schneider, B Collet (Barbara Budrich, Oplanden) pp 69–93
- Sandow E, 2011 *On the Road—Social Aspects of Commuting Long Distances to Work* PhD thesis, GERUM Kulturgeografi 2011:2, Umeå University
- Sandow E, Westin K, 2010, "The persevering commuter—duration of long-distance commuting" *Transport Research Part A* **44** 433–445
- SIKA, 2007, "The national travel survey RES 2005/2006", no. 2007:19, Swedish Institute for Transport and Communication Analysis (SIKA), Stockholm
- Sjaastad, L A, 1962, "The costs and returns of human migration" *The Journal of Political Economy* **70** 80–93
- Smith J, Todd P, 2005, "Does matching overcome LaLonde's critique of non-experimental estimators?" *Journal of Econometrics* **125** 305–353
- Stenberg A, Luna X, Westerlund O, 2012, "Can adult education delay retirement from the labour market?" *Journal of Population Economics* **25** 677–696
- Storper M, Venables A J, 2004, "Buzz: face-to-face contact and the urban economy" *Journal of Economic Geography* **4** 351–370
- Stutzer A, Frey B S, 2008, "Commuting and life satisfaction in Germany" *Scandinavian Journal of Economics* **110** 339–366
- van Ham M, 2001, "Workplace mobility and occupational achievement" *International Journal of Population Geography* **7** 295–306
- van Ommeren J, 1996 *Commuting and Relocation of Jobs and Residences: A Search Perspective* PhD thesis, Vrije Universiteit Amsterdam
- van Ommeren J, Rietveld P, 2005, "Compensation for commuting in imperfect urban markets" *Regional Science* **86** 241–259
- Wachs M, Taylor B D, Levine N, Ong P, 1993, "The changing commute: a case study of the jobs–housing relationship over time" *Urban Studies* **30** 1711–1730
- Waernbaum I, 2008 *Covariate Selection and Propensity Score Specification in Causal Inference* PhD thesis, Statistics, Umeå University
- William J, 2000 *Unbending Gender: Why Family and Work Conflict and What to Do About It* (Oxford University Press, New York)