The Academic and Wage Gap among Healthcare Professionals in Spain

Manuel Salas-Velasco – Universidad de Granada – Department of Applied Economics – Granada (Spain) <msalas@ugr.es>

Abstract

This article investigates the labor market outcomes of health-related degree holders, such as those in medicine and nursing, in their early years of employment. We use nationally representative data from university graduates who completed both degrees at Spanish universities in the 2013-2014 academic year and were interviewed five years after graduation. Economics research has generally focused on the salary that a university graduate commands, but it has not adequately addressed the sample selection bias caused by the university degree choice. The current analysis examines earnings differentials among physicians and nurses and allows for degree choice to be endogenous. Modeling the characteristics that influence the choice of the degree uncorrelated with earnings allows the bias to be corrected. Our results suggest that degree choice can make a substantial difference in early career earnings outcomes. We find an important wage premium for physicians compared to nurses. Inequalities in degree choice based on parents' education and gender persist. Those who are the "first in their family" to attend university (i.e., first-generation university students) are still under-represented in medical schools, whereas women are over-represented in nursing schools.

Keywords: first-generation university students; healthcare workers; interval regression; instrumental variables; logistic regression; wage differentials.

JEL Codes: I26; J31.

1. Introduction

Prospective higher education students in Spain, like in many other countries throughout the world, have a wide range of degree options to choose from when entering university. Medicine and nursing studies are offered as undergraduate degree programs at Spanish universities.[1] However, obtaining a medical degree typically takes six years of study (formal curriculum). After completion, undergraduates must complete a residency program. In contrast, a nursing degree encompasses a four-year bachelor's program in Spain. The formal curriculum contains a considerable number of hours of practice in a teaching hospital. It is worth noting that both degrees attract the best Spanish high school students—university entrance requirements are relatively high when compared to other undergraduate programs—and their academic performance at the university is superior to their peers in other degrees (Michavila, 2012).[2]

If we focus on labor market outcomes, the fields of medicine and nursing both provide excellent career prospects, but physicians generally have higher earning potential compared to nurses due to their advanced education and specialized expertise (Barry, 2021; Britton et al., 2022; Organisation for Economic Co-operation

[1] Spanish university students begin their undergraduate degree programs with their major already selected and take courses that are pre-assigned for the entire program.

[2] In a meta-analysis of international data, O'Neill et al. (2011) also found that attrition rates in health-related degrees are rather low when compared to other university programs.

and Development [OECD], 2023; Salas, 2010).[3] Nursing training is shorter, more occupationally oriented, and generally leads to lower earnings and status. In addition, there is a degree-choice pattern. Rational individuals will choose university degree programs that offer higher profitability, but only when this choice encompasses an acceptable risk level (Jiménez & Salas, 2000). Students from lower social classes tend to opt for shorter educational pathways to reduce the risk of failing in a long-lasting and costly education (Hartlaub & Schneider, 2012). On the contrary, the more advantages students have in terms of social background, the more importance they will give to return and less to risk (Mingat & Eicher, 1982).[4] Moreover, families from upper social classes must opt for the academically most demanding, long-lasting curricula to avoid an intergenerational status loss (Hartlaub & Schneider, 2012).

In this context of degree and career choice among health professionals, the purpose of this paper is to analyze the variation in early career earnings returns among medical doctors and nurses in Spain and investigate the predictability of those returns based on observable individual characteristics. Since the applied research in the literature has not satisfactorily addressed the sample selection bias resulting from university degree choice, the current study dives into the topic of wage differentials,

[3] In all OECD countries, the remuneration of doctors (both family doctors and specialists) is substantially higher than the average wage of a full-time employee across all economic sectors (OECD, 2023).

[4] The better the parent's financial situation is, the lower the constraints on financing long-lasting educational programs are.

addressing the issue of sample selection or the endogeneity of the explanatory variable for degree choice. Modeling the characteristics that influence the choice of the degree uncorrelated with earnings allows the bias to be corrected. In order to achieve our research objectives, this study takes advantage of the opportunity provided by the EILU2019 graduate survey (*Encuesta de Inserción Laboral de titulados Universitarios*), which is a large, nationally representative random sample of Spanish universities and university graduates. A total of 31,651 bachelor's degree holders from the 2013-2014 academic year were surveyed in 2019 (i.e., five years after graduation). As a result, they are young people who have just begun their careers. In the EILU2019 survey, 2.4% of undergraduates studied medicine (n = 767), while 3.9% studied nursing (n = 1,235).[5]

A "limitation" of our data is that the EILU2019 survey contains the net monthly salaries of working undergraduates in interval form. We know the ordered category into which each observation falls, but we do not know the exact value of the observation. Nonetheless, interval regression is used in this article to model outcomes that have interval censoring. In particular, interval wage regression is a valuable econometric tool for studying the determinants of wages and understanding the factors that influence individuals' earnings. The database also enables us to identify factors associated with the choice of university degree, which will be employed as

[5] In the EILU2019 survey, around 91% of medical doctors (80% of nurses) studied at a traditional Spanish public university.

instrumental variables to correct sample self-selection bias and interpret the impact of the degree, medicine versus nursing, as causal. Moreover, because a large portion of health care in Spain is provided by the government, public universities are primarily the institutions in charge of training health professionals such as doctors and nurses. As a result, the basic bias is self-selection based on degree choice rather than a public or private university or sector of employment.

2. The context

2.1 University degree choice

Healthcare professionals are in great demand in all OECD nations (OECD, 2023). The need has been growing in recent years, and as the population ages, it will only get stronger. Several countries are increasingly depending on immigration to address domestic shortages of healthcare workers. Shortages are said to be particularly significant in Canada, the United Kingdom, and the United States of America. So far, in Spain, the supply of health practitioners is keeping up with the growing demand for health workers. The number of registered healthcare professionals in Spain was 877,361 in 2019, according to the Spanish Statistics Institute (Instituto Nacional de Estadística [INE], 2020). The health profession has a predominantly female presence. In 2019, 278,371 were men and 598,990 were women (INE, 2020). The majority were nurses (36.03% of the total), followed by medical doctors (30.55%).[6]

[6] The year 2019 is chosen to connect with the data used in this article, that is, the EILU2019 survey.

Nonetheless, choosing and enrolling in a nursing degree vs. a medical degree and, consequently, selecting a profession is still conditioned by social and cultural factors. For the purposes of this study, we define inequalities in degree choice mainly as the unbalanced distribution of gender and parents with higher education between the two main health sciences bachelor's programs at Spanish universities. An analysis in Table 1 of the gender distribution of students shows that there are large gender differences in the uptake of health science disciplines: 68.5% of women got a medical degree, while this percentage was 81.1% among women who got a degree in nursing studies. Disparities in degree choice according to parents' educational attainment also persist. We note in Table 1 that 64.7% of those who studied medicine have parents with a university degree—at least one of them (father or mother) earned a university education. They are known as non-first-generation students. However, only 35.3% of medical graduates are first-generation university students, that is, they are the first ones in their families to get a university degree. In contrast, only 33.2% of undergraduates in nursing studies are non-first-generation students, meaning that 66.8% of undergraduates in nursing are first-generation going to college (i.e., both parents without university studies). Particularly, 53.5% of undergraduates in medicine have mothers with a university degree; this percentage is only 21.9% among graduate students in nursing (Table 1). Thus, the educational background of a person's mother can have a significant impact on their academic and career choices. Mothers who have obtained higher education degrees may have a higher socioeconomic status

and a greater understanding of the importance of education.[7] This can result in their children perceiving medicine as a prestigious and financially rewarding profession, leading them to choose it as their career path. Furthermore, the educational background of the mother can also impact the access to information and resources available to the individual. A mother with a higher education level may have a wider network of contacts and resources to help their child enter medical school, navigate the admission process, or provide guidance throughout their educational journey. This advantage can increase the likelihood of their child pursuing medicine as a career.

Table 1. Inequalities in students' choice of university degrees in health sciences

	Nurses		Doctors	
	Freq.	Percent	Freq.	Percent
Male	234	19.0	242	31.6
Female	1,001	81.1	525	68.5
	1,235	100.0	767	100.0
Non-first-generation university students	410	33.2	496	64.7
First-generation university students	825	66.8	271	35.3
	1,235	100.0	767	100.0
Mother with university studies				
Yes	271	21.9	410	53.5
No	964	78.1	357	46.5
	1,235	100.0	767	100.0

Source: author's calculations using the EILU2019 survey

[7] Those with a high taste for education and strong motivation for schooling for their children are called "educo-genic" by Anderson (1987).

2.2 Remuneration of health professionals

Governments intervene in the market for health goods and services in all countries, partly because of market failures (e.g., Donaldson & Gerard, 1993; Roemer, 1982). It is possible for the public sector to act as a monopsonist in the labor market for health professionals, particularly in cases where it is the dominant or sole employer of such professionals. So, labor markets for healthcare professionals are imperfect in capturing productivity levels in the wage rate (McPake et al., 2015).

Table 2 displays the salaries of medical doctors and nurses interviewed at the EILU2019 survey who were working in Spain as employees five years after graduation (current job). Besides, we considered those undergraduates who were well-matched to their jobs (i.e., they worked in graduate jobs that required their specific university degree).[8] As a result, the analysis done in Table 2 comprises 81% of the 2014 nursing cohort and 90% of the 2014 medical cohort.

The respondents provided their net monthly salaries at seven intervals (Table 2). Most nurses (around 50%) earn around €1,750, while most doctors (37.3%) earn around €2,250 (mid-interval values). We also calculate an average salary of €1,659 for nurses and €2,367 for doctors; thus, there is an unadjusted wage premium for doctors of around €700. The ratio of 142.6 also tells us that doctors earn 42.6% more than nurses without any controls. We must emphasize that those with nursing degrees enter the

[8] We also eliminated inactive people and a few observations of those living and working in Ceuta and Melilla (two small Spanish cities in the north of Africa).

job market quickly, so their earnings also capture their work experience. In the case of doctors, many respondents (37.7%) were resident doctors. Following completion of their degree, they must take the residency placement test to obtain their medical specialty, including family medicine.[9] As a result, they have far less work experience than nursing graduates. This may explain why, in the early stages of a health professional's career, the compensation premium for doctors over nurses is not as significant.

Table 2. Inequalities in the distribution of health workers' current net monthly salary earnings, 2019

	Nurses		Doctors		
	Freq.	Percent	Freq.	Percent	Mid-interval values
€699 or less	20	2.0	1	0.2	349.50
€700 — €999	59	5.9	7	1.0	849.50
€1,000 — €1,499	234	23.5	41	6.0	1,249.50
€1,500 — €1,999	488	49.0	92	13.4	1,749.50
€2,000 — €2,499	178	17.9	257	37.3	2,249.50
€2,500 — €2,999	14	1.4	156	22.6	2,749.50
More than €3,000	4	0.4	135	19.6	3,000.00
	997	100.0	689	100.0	
Average salary	€1,659.13		€2,366.51		
Ratio	142.6				

Source: author's calculations using the EILU2019 survey

[9] A higher score on the exam means candidates are more likely to get their preferred specialty training spot.

3. Determinants of degree choice: Where are inequalities produced?

3.1 Modeling educational choices

In this section, we analyze the determinants of degree choice. For each individual *i*, the dependent variable takes two values: $Y_i = \{1, 0\}$, depending on whether the individual studied medicine or nursing, respectively.[10] In situations such as this, in which the outcome variable is dichotomous, classic regression, which attempts to explain the level of a continuous outcome variable in terms of a set of explanatory variables, is not the most appropriate econometric method since its optimum properties are based on assumptions that stop being valid when the outcome variable is qualitative (Maddala, 1983). However, by using a qualitative response model, such as the binomial logit model, also known as logistic regression, we can predict the probability that an option will be chosen. In particular, the binomial logit model estimates or predicts the probability that the event defined as $Y_i = 1$ occurs (studying medicine, in the current study) in terms of the values that the independent (explanatory) variables adopt.

McFadden (1974) proved that, in this case, the probability that an individual *i* chooses alternative 1 is:

[10] *Numerus clausus* and other restrictions can condition the choice of degree. Since we do not have information about pre-enrollment student-stated preferences regarding degrees, we must rely on revealed preferences (actual degree choice), implicitly assuming that the degree a student pursues is the student's preferred degree.

$$Prob[Y_i = 1] = \frac{e^{X_i'\beta}}{1 + e^{X_i'\beta}} \quad (1)$$

McFadden (1974) provided a theoretical foundation that linked the logit model to the discrete choice theory from mathematical psychology.[11] Eq. (1) would be the reduced form for the binomial logit model, which can be seen as a special case of a general model of utility maximization (Cramer, 1991). In our degree-choice setup, the X_i row vector of explanatory variables for the *i*-th individual contains factors influencing the choice of degree (including a constant), and β is the vector of parameters to be estimated. Furthermore, it is assumed that the non-observed ε's follow a distribution of logistic probability.

3.2 Results

The logistic regression results—the estimated coefficients, odds ratios (ORs), and marginal effects—are shown in Table 3.[12] A (1-0) degree dummy dependent variable captures the two bachelor's degrees awarded by the Spanish university system: Y = 1, for medicine; Y = 0, for nursing. Table 3 shows the greater or lesser probability of studying medicine by parent's education (base Model I) and according to other individuals' sociodemographic characteristics (full Model II). The values of the ORs are relevant to our analysis. ORs greater than one are associated with an increased probability of pursuing medical studies. ORs between 0 and 1 are associated with a

[11] Seminal works in the educational choice literature are due to Jiménez & Salas (2000).

[12] The estimates were obtained using Stata® 18 software. Descriptive statistics are shown in Appendix Table A1.

reduced probability of studying medicine (or an increased probability of studying nursing). Model I's results reveal that university students who have at least one parent with a university degree (i.e., non-first-generation undergraduates) are the most likely to have developed a professional career as a medical doctor: OR = $e^{1.304} = 3.68$. Another interpretation would be that non-first-generation university students demand 3.68 times more medicine compared to their first-generation peers. Nonetheless, the argument that the performance of children from lower social classes is, on average, lower and therefore they more often decide to undertake lesser amounts of human capital is not valid in our sample because in both degrees we have the best high school performers. The expansion of higher education in Spain has reduced social inequities in access since the 1980s, but social class differences in degree choice may persist if students from lower socioeconomic backgrounds, particularly from less educated families, are less likely to study disciplines with high labor market returns, as this study reveals. True equal opportunities mean that regardless of their socioeconomic origin, all high school graduates have the same likelihood of getting admitted to university and earning more advanced degrees.

Table 3. Results of the logistic regression of degree choice in health sciences

	Model I				Model II			
	Coefficient	Robust Std. Err.	Odds Ratio	Marginal Effect	Coefficient	Robust Std. Err.	Odds Ratio	Marginal effect
Non-first-generation university student (= 1 yes)	1.304 ***	0.113	3.68	0.30	1.118 ***	0.113	3.06	0.26
Gender (=1 female)					-0.577 ***	0.122	0.56	-0.13
Age under 30 years (= 1 yes)					0.709 ***	0.157	2.03	0.16
The main motivation for undertaking the undergraduate program								
Getting training, thinking about my labor future					Ref.			
Broadening knowledge for personal satisfaction					0.412 ***	0.146	1.51	0.10
Other motives					0.109	0.171	1.11	0.03
Mother with secondary or tertiary vocational training (=1 yes)					-0.288 **	0.139	0.75	-0.07
Father born outside the EU (= 1 yes)					0.282	0.209	1.33	0.07
Constant	-1.113 ***	0.136	0.33		-1.077 ***	0.121	0.34	
	Number of obs. = 2,002				Number of obs. = 2,002			
	Wald chi2(1) = 134.20				Wald chi2(7) = 297.65			
	Prob > chi2 = $p < 0.001$				Prob > chi2 = $p < 0.001$			
	Pseudo R2 = 0.0717				Pseudo R2 = 0.1002			
	Log pseudolikelihood = -1236.905				Log pseudolikelihood = -1198.9353			

Dichotomous outcome: university degree (= 1 medicine; = 0 nursing).

Std. Err. adjusted for 17 clusters in the region (*comunidad autónoma*) of the university where the degree was granted.

Marginal effects at the means (MEMs) (Williams, 2012).

** $p < 0.05$

*** $p < 0.01$

Source: author's calculations using the EILU2019 survey

The estimated marginal effects of all independent variables are also shown in Table 3.[13] According to Model I's estimates, non-first-generation university students increase the probability of studying medicine by 0.30 (without any controls); that is, those individuals whose parents went to university are about 30% more likely than first-generation university students (the "first in their family" to attend university) to study medicine. In fact, the probability of studying medicine for a first-generation university student (when $X = 0$) is only 25%:

$$Prob(Y = 1) = \frac{e^{(\alpha+\beta X)}}{1 + e^{(\alpha+\beta X)}} = \frac{e^{-1.113}}{1 + e^{-1.113}} = 0.247$$

However, the probability of studying medicine for a non-first-generation university student (when $X = 1$) is 55%:

$$Prob(Y = 1) = \frac{e^{(\alpha+\beta X)}}{1 + e^{(\alpha+\beta X)}} = \frac{e^{(-1.113+1.304)}}{1 + e^{(-1.113+1.304)}} = 0.547$$

Hence, a non-first-generation university student is 30 percentage points more likely to become a medical doctor than a comparable first-generation student. However, it is also essential to consider that degree choice may not be solely determined by the highest education level attained by the parents. The individual's personal aspirations and interests can also significantly impact the choice between medicine and nursing. It is possible there are unobserved characteristics, such as motivation, which may be associated with degree choice. However, the EILU2019

[13] For categorical variables, the marginal effects show how much P(Y=1) will change if the variable changes from 0 to 1 (with other variables held at mean values).

survey asked questions about the main motivation to undertake their undergraduate programs.[14] According to Model II, broadening knowledge for personal satisfaction as the main motivation in students' choice of degree is associated with a 10% increase in the probability of developing a career as a doctor (the estimated marginal effect shown in the last column of Table 3). Among the further sociodemographic characteristics influencing degree choice, it is worth highlighting the results obtained for two of the variables incorporated in the econometric estimation. First, there is a significant direct gender effect on the actual degree and career choice. Specifically, the results of Model II in Table 3 show that women are less likely to become physicians than men: the OR is equal to 0.56 and the marginal effect is -0.13 for the gender variable. Another interpretation, which is the same, would be that a male is 13 percentage points more likely to become a medical doctor than a comparable female counterpart. Second, those undergraduates whose mothers obtained a vocational education and training (VET) diploma are, all else equal, more likely to become nurses—they reduce the likelihood of studying medicine: OR equal to 0.75 and a marginal effect of -0.07. Parents who did not attend university may try to steer their children toward more vocationally oriented studies with which they are more familiar. This could result in

[14] Previous research has shown that students choose the degree that reports them the maximum utility level (e.g., Artabe & Gardeazabal, 2017; Jiménez & Salas, 2000). In the current study, the observed degree choice is taken to be the alternative that reports the highest utility.

first-generation university students choosing studies with a more clearly defined career path or faster entry into the labor market (Hoxby & Avery, 2013).

3.3 Hypothesis testing

The Chi-square test was used for the overall contrast of the estimated models. The null hypothesis, H_0, is that all the coefficients of the equation (except the constant) are null. In both models, the associated *p*-value is very low (less than 0.001). Thus, the result of this test allows one to reject the null hypothesis and accept the (global) regression as valid. Furthermore, the Hosmer-Lemeshow test was not significant, meaning that the logistic regression model fits the data well.[15]

4. Salaries for medical doctors and nurses in Spain

4.1 The human capital earnings function

To look into the labor market returns of investing in health-related university degrees, the classic Mincerian earnings function (Mincer, 1974) will be employed, which regresses wages on education, experience, and other relevant explanatory variables. In our study, Equation 2 gives a general model of earnings determination where Y_i is the individual *i*'s wage, X_i is a matrix of explanatory variables (mainly work experience and job-related variables), β is the vector of parameters to be estimated, and ε_i is the error term—a random variable with a zero mean and constant variance that reflects non-observable characteristics that influence wages. Our main objective is

[15] Hosmer–Lemeshow chi2(7) = 7.63; Prob > chi2 = 0.3659.

to estimate the wage premium for doctors over nurses. We may achieve this goal by introducing a dummy variable, D, in Eq. 2, which has a value of 1 if the individual is a medical doctor and 0 if they are nurses. The ordinary least squares (OLS) estimation of Eq. 2 provides a measure of the wage premium for medical doctors compared to nurses, denoted by $\hat{\delta}$.

$$Y_i = X_i'\beta + \delta D_i + \varepsilon_i \quad (2)$$

Assuming the rest of the regression model (Eq. 2) is accurately identified, $\hat{\delta}$ would represent the true return of investing in medicine over nursing if the degree were an exogenous explanatory variable. The key assumption here is that there are no variables omitted from this equation that are related to both the higher education choice and subsequent earnings outcomes. However, it seems quite plausible that the same unobserved factors might influence both individuals' educational attainment and their earnings (Bound et al., 1995). Then, D and ε are correlated in Eq. 2: $\text{Cov}(D,\varepsilon) \neq 0$. When explanatory variables are endogenous, OLS regression gives biased and inconsistent estimates of the causal effect of an explanatory variable on an outcome Y (Bound et al., 1995). A common strategy for dealing with this endogeneity is to use instrumental variable estimation, using as "instruments" variables thought to have no direct association with the outcome Y (Bound et al., 1995).

4.2 Econometric strategy

A simple model with endogenous education consists of a two-equation system:

$$Y_i = X_i'\beta + \delta D_i + \varepsilon_i \quad (3)$$

$$D_i = V_i'\alpha + w_i$$

The two-equation system integrates educational attainment and returns to schooling into a unified framework. In the first equation, the variable *D* is an explanatory variable, while in the second equation, it is a dependent variable. Consequently, the university degree must be regarded as an endogenous binary explanatory variable since it is determined from within the model. In Eq. 3, *V* contains exogenous variables that influence the degree choice, called instrumental variables (IVs) or instruments. The latter are used to predict the determinants of the amount of education individuals accumulate. The predicted values for the dependent variable in the second equation will then be included in the first equation, so the estimated coefficient associated with *D* would yield a good measure of the returns to higher education (i.e., the wage premium for medical physicians compared to nurses). Nonetheless, to be effective, the instruments a) should not be correlated with the error term of the wage equation, and b) they should be highly correlated with the endogenous variable. Condition a) guarantees the validity of the instruments; condition b) guarantees their quality (Baltagi, 2011; Bound et al., 1995; Salas, 2006).

4.3 Practical implementation

We use Stata® 18 statistical software to estimate the Mincer-type earnings function while controlling for degree choice. We apply Stata's *eintreg* command (extended interval regression) since respondents in the EILU19 graduate survey reported salaries

in intervals (Appendix Table A2), and *eintreg* also fits an interval regression model that accommodates any combination of endogenous covariates (binary endogenous covariates are allowed). A probit model may be used to account for endogenous sample selection (in our study, degree choice). Specifically, health workers' net monthly salaries are explained by differences in formal education (university degree earned), experience, hours worked, medical specialization, geographic characteristics, and other relevant factors. Descriptive statistics are provided in Appendix Table A3. We use the same subsample of wage-earners that we used in the analysis shown in Table 2.[16]

The challenge in estimating the earnings returns by type of undergraduate program is to account for differences between individuals that might affect their decision of what to study but are unrelated to their salary earnings. Our identification strategy relies on the assumption that selection into different higher education programs is certainly a behavioral result of individuals or their parents (Table 3).[17]

4.4 Results

Table 4 presents the regression results using maximum likelihood, where the intercept and slope estimates have been rounded to three decimal places. At the top of Table 4,

[16] That is, we applied the same filters used in subsection 2.2. Nevertheless, we finally dropped 25 observations with missing values in some of the explanatory variables used in the interval regression analysis.

[17] It is possible to have a binary endogenous explanatory variable and a binary instrumental variable (Wooldridge, 2012).

we show the estimated coefficients for the wage equation, that is, the structural equation. We fit an interval regression model of undergraduates' salaries using education, job characteristics, and several demographic factors as explanatory variables. Because the conditional mean modeled by interval regression is linear, the coefficients are interpreted the same way they are in ordinary least-squares regression. Once we account for differences in pre-university characteristics, we estimate that studying medicine over nursing raises the net monthly salary by €402, which is our estimate of an adjusted wage premium for doctors. When the education variable is dichotomous, the computed coefficient should indeed be understood as the marginal effect over earnings rather than the estimated return to schooling (Psacharopoulos, 1994).[18]

Therefore, the extra years spent in university education have an impact that is both positive and significant on salaries. However, to interpret the latter result as causal, the binary endogenous regressor (i.e., university degree) is instrumented using pre-university variables related to the main motivation for pursuing the specific undergraduate program after finishing high school (also family background

[18] The private internal rate of return on a medical degree would be the discount rate which equates direct and opportunity costs with expected benefits; costs and benefits are discounted to the present moment. The benefits of obtaining a medical degree can be identified with the additional salary that an individual can obtain with a graduate degree, relative to what he or she would have earned with a nursing degree, over the graduate's work life. The additional direct cost of tuition fees for two years and the average wages earned by nurses during their years of employment should also be taken into account.

variables). The parameters for the reduced-form equation showing the relationships between the endogenous variable and the IVs are also displayed in the middle of Table 4.[19] In particular, the respondent's consumption motive for continuing in higher education (personal satisfaction) is an effective instrumental variable since it is highly correlated with the endogenous variable of degree choice (instrument relevance) but does not affect wages directly (instrument exogeneity).[20] In wage equations, labor economists have also used family background variables as IVs for education. In our study, the mother's education (whether the undergraduate's mother got a VET formal education) is positively correlated with the probability of choosing nursing studies, but it is expected to be uncorrelated with the undergraduate's salary.[21]

Let's focus on the other estimation results of the wage equation for which we have found statistically significant estimates. For example, working part-time lowers the undergraduate's monthly salary by €604, and each additional year of experience raises

[19] The selection equation is an example of a reduced-form equation, which means that we have written an endogenous variable in terms of exogenous variables.

[20] The education consumption motive measures the contribution of education to usefulness, separating it from the monetary dimension (Jiménez & Salas, 2000). So, it seems plausible that the education consumption motive is uncorrelated with the error term ε (i.e., the error term of the wage equation).

[21] Unfortunately, we cannot test that the IVs are uncorrelated with ε (Wooldridge, 2012). However, in the context of omitted variables, this means that they should not have a partial effect on Y (outcome). Appendix Table A4 shows a full model of salary determinants, including a proxy variable for ability (number of spoken languages). The estimated coefficients associated with motivation when choosing the university degree and whether the undergraduate's mother got a VET education are not statistically significant. So, our result that at least two variables that likely influence the choices of students are not correlated with earnings (once we control for ability) supports the validity of the instruments.

the salary by €17. There is also an important effect of medical training on subsequent wages. Everything else held constant, medical specialties raised the salary by €515. There is also a raise in the salary for those not working on the peninsula, that is, those healthcare professionals who work in the Balearic Islands (€195) and the Canary Islands (€124). It is worth noting that having earned a master's degree raises the salary by €75. However, there is a gender gap of €82. Note also that the predicted base monthly salary for a fresh male nurse who works full time on the peninsula and did not complete a master's degree is the intercept (constant), which equals €1,619.

4.5 Hypothesis testing

The method of instrumental variables has been used in the current study to solve the problem of endogeneity of the explanatory variable of our interest, that is, degree choice. However, since OLS is preferred to instrumental variable (IV) estimation, we'd like to be able to test for endogeneity to avoid IV estimation.[22] At the bottom of Table 4 is a test of the exogeneity of the endogenous explanatory variable. That is, the test simply asks whether the error terms in the structural equation, the wage equation, and the reduced-form equation for the endogenous variable, the selection equation, are correlated. Under H_0, the correlation is null. We reject the null hypothesis of no endogeneity. See that the predicted value for "rho" turned out to be significantly different from 0.

[22] If we do not have endogeneity, both OLS and IV are consistent, although OLS is more efficient.

For the global contrast of the estimated model, the chi-square test was used. The null hypothesis is that all the coefficients of the equation, except the constant, are null. According to the results shown in Table 4, Wald chi2(17) = 2329.64, and the associated *p*-value was very low (less than 0.001). The result of this test allows us to reject the null hypothesis, accepting the model as good.

Table 4. Health workers' salary determinants

Wage equation (dependent variable: net monthly salary)	Coefficient		Robust Std. Err.
University degree (= 1 medicine; = 0 nursing)	**402.020**	***	62.306
Gender (=1 female)	-82.331	***	20.928
Marital status (= 1 married)	19.329		31.805
Years of working experience	16.909	***	1.907
Part-time job (= 1 yes)	-604.049	***	35.164
Recognized disability >33% (= 1 yes)	-52.421		120.059
Master's degree (= 1 yes)	74.639	***	22.265
Specialist physician (= 1 yes)	515.378	***	56.381
Region (Islas Baleares = 1)	195.425	***	22.692
Region (Islas Canarias = 1)	123.534	***	28.064
Industry of health and social services activities (= 1 yes)	28.116		46.543
Number of training workshops/seminars since graduation paid by SEPE [a)]	2.156		36.808
Fluent in English (=1 yes)	16.228		28.957
Number of spoken languages (excluding mother languages)	17.227		15.638
Age under 30 years (= 1 yes)	15.849		24.914
Factors to get the job: theory [b)]	11.212		31.019
Factors to get the job: management ability [b)]	-15.457		25.610
Constant	1618.509	***	62.319
Selection equation (dependent variable: degree choice)	Coefficient		Robust Std. Err.
Gender (=1 female)	-0.382	***	0.085
Marital status (= 1 married)	-0.044		0.084
Years of working experience	-0.047	***	0.012
Part-time job (= 1 yes)	-0.925	***	0.113
Recognized disability >33% (= 1 yes)	-0.204		0.316
Master's degree (= 1 yes)	-0.159		0.132
Specialist physician (= 1 yes)	1.370	***	0.111
Region (Islas Baleares = 1)	-0.672	***	0.060
Region (Islas Canarias = 1)	-0.127	*	0.076
Industry of health and social services activities (= 1 yes)	0.435	***	0.151
Number of training workshops/seminars since graduation paid by the SEPE [a)]	-5.067	***	0.273
Fluent in English (=1 yes)	0.431	***	0.080

Number of spoken languages (excluding mother languages)	0.314	***	0.064
Age under 30 years (= 1 yes)	0.133	***	0.086
Factors to get the job: theory [b)]	0.754	***	0.123
Factors to get the job: management skills [b)]	-0.342	***	0.077
Non-first-generation university student (= 1 yes)	0.479	***	0.070
The main motivation for undertaking the undergraduate program			
Getting training, thinking about my labor future	Ref.		
Broadening knowledge for personal satisfaction	0.224	**	0.105
Other motives	0.099		0.127
Mother with secondary or tertiary vocational training (= 1 yes)	-0.229	**	0.089
Father born outside the EU (= 1 yes)	0.283		0.317
Constant	-1.758	***	0.199
rho = corr. (error selection equation, error wage equation)	0.199	**	0.095

Log pseudolikelihood = -2839.6101
Number of obs. = 1,661
Uncensored = 0
Left-censored = 21
Right-censored = 138
Interval-cens. = 1,502
Wald chi2(17) = 2329.64
Prob. > chi2 = $p < 0.001$
Std. Err. adjusted for 49 clusters in the province of the current job.
a) SEPE, from its Spanish acronym, is the State Public Employment Service.
b) This variable takes the value of 1 if it was an important or very important factor in obtaining the current job; it takes the value of 0 otherwise.

* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$
Source: author's calculations using the EILU2019 survey

5. Discussion

Social mobility has been a topic of focus throughout OECD countries, as policymakers seek solutions to promote more equitable societies (OECD, 2018). A range of previous literature has highlighted the importance of access to higher education as a pathway to social mobility (Blanden & Macmillan, 2016; Chetty et al., 2017; Rothstein, 2019). According to the literature, socioeconomic status, particularly parental education, has a considerable influence on access to higher education (Blanden & Machin, 2004). In this regard, previous research on first-generation university graduates has attempted to explain how parental education influences their decisions about what (and where) to study at university (Anelli, 2020; Henderson et al., 2020).

The expansion of higher education in Spain has reduced social inequities in access, but social class differences in degree choice may persist if students from higher socioeconomic backgrounds get admitted to university and earn more advanced degrees. In this regard, we find that first-generation students are less likely to study medicine. Since medical doctors are among the highest earners in Spain, and also in most OECD countries, this evidence may have long-term inequality implications for labor market outcomes. This is striking since the first-generation students in our sample are not completely constrained by poor academic performance. Financial constraints are a more plausible reason behind our result. First-generation students come from households with lower levels of education, and thus they are more likely to come from lower-income families. This means that even in a country like Spain with

low tuition fees, they may lack the income to support the other costs, such as opportunity costs and living expenses associated with extra years of university study.

We also find that choosing and enrolling in a higher education degree and, consequently, selecting a profession is still conditioned by social and cultural factors characterized by gender stereotypes that continue to attract more women than men to the nursing profession in Spain. The general perception remains that nursing is "women's work" and that the occupation has a low professional status (OECD, 2023). Indeed, in all OECD countries, the nurse profession continues to be predominantly a female profession. One persistent challenge across OECD countries is the need to attract more male students to nursing (OECD, 2023).[23]

6. Conclusion

This article focuses on the labor market outcomes of physicians and nurses who studied and are working in Spain. We use survey data to explore the variation in early career earnings returns and investigate the predictability of those returns based on observable characteristics of the undergraduates and their jobs. Our results suggest that degree choice is crucial for subsequent earnings outcomes. However, a study examining the wage disparity between doctors and nurses must adequately address the issue of sample selection. Econometrically, self-selection is a problem of the

[23] We also ran a logistic regression that allowed for interactions between parental education and gender in order to account for these combined effects. The dependent variable was binary (Y = 1 for nursing studies). We found that female first-generation students are more likely to become nurses (an estimated OR equal to 3.59).

endogeneity of the explanatory variable, which, in this case, is the dummy variable used to represent the individual's choice of an academic degree. Instrumental variables are important for obtaining consistent estimates when endogeneity is suspected. However, invalid instruments can produce meaningless results. Hence, one of the biggest problems with IV estimation is the choice of the instrumental variables. Our identification strategy relies on the assumption that selection into different higher education programs is certainly a behavioral result of individuals or their parents. In particular, the respondent's consumption motive for continuing in higher education and his/her mother's education are two effective instrumental variables used in our study. Once our earning equation has been adjusted to account for selection bias, we found that finishing medicine over nursing raises the net monthly salary by €402, which is our estimate of an adjusted wage premium for medical doctors. From the results obtained, it can be concluded that degree choice can make a substantial difference in early career earnings outcomes. Nevertheless, inequalities in degree choice based on parents' education and gender persist in Spain.

Declarations

Funding: Not Applicable
Conflicts of interest: Not Applicable
Availability of data and materials: Not Applicable
Ethics approval: Not Applicable
Author contributions: Not Applicable
Informed consent: Not Applicable

References

[1] Anderson, C. A. (1987). *Social selection in education and economic development*. The World Bank.

[2] Anelli, M. (2020). The returns to elite university education: A quasi-experimental analysis. *Journal of the European Economic Association*, *18*, 2824–2868.

[3] Artabe, A., & Gardeazabal, J. (2017). Degree choice evidence from stated preferences. *Empirical Economics*, *52*, 1205-1234.

[4] Baltagi, B. H. (2011). *Econometrics* (5th ed.). Springer.

[5] Barry, J. (2021). Real wage growth in the US health workforce and the narrowing of the gender pay gap. *Human Resources for Health*, *19*(1), 1-9.

[6] Blanden, J., & Machin, S. (2004). Educational inequality and the expansion of UK higher education. *Scottish Journal of Political Economy*, *51*, 230–249.

[7] Blanden, J., & Macmillan, L. (2016). Educational inequality, educational expansion and intergenerational mobility. *Journal of Social Policy*, *45*, 589–614.

[8] Britton, J., van der Erve, L., Belfield, C., Vignoles, A., Dickson, M., Zhu, Y., Walker, I., Dearden, L., Sibieta, L., & Buscha, F. (2022). How much does degree choice matter? *Labour Economics*, *79*, 102268.

[9] Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*(430), 443-450.

[10] Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2017). *Mobility report cards: The role of colleges in intergenerational mobility* (NBER Working Paper No. 23618). National Bureau of Economic Research.

[11] Cramer, J. S. (1991). *The logit model. An introduction for economists*. Edward Arnold.

[12] Donaldson, C., & Gerard, K. (1993). Market failure in health care. In *Economics of health care financing: The visible hand* (pp. 26-48). Palgrave.

[13] Hartlaub, V., & Schneider, T. (2012). *Educational choice and risk aversion: How important is structural vs. individual risk aversion?* (SOEPpapers on Multidisciplinary Panel Data Research, No. 433). Deutsches Institut für Wirtschaftsforschung (DIW).

[14] Henderson, M., Shure, N., & Adamecz-Völgyi, A. (2020). Moving on up: 'First in family' university graduates in England. *Oxford Review of Education*, *46*(6), 734-751.

[15] Hoxby, C. M., & Avery, C. (2013). Low-income high-achieving students miss out on attending selective colleges. Brookings Papers on Economic Activity, Spring, 10.

[16] Instituto Nacional de Estadística. (2020). *Registered health professionals statistics, year 2019*. https://ine.es/en/prensa/epsc_2019_en.pdf

[17] Jiménez, J. D., & Salas, M. (2000). Modeling educational choices. A binomial logit model applied to the demand for higher education. *Higher Education*, *40*(3), 293-311.

[18] Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.

[19] McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

[20] McPake, B., Squires, A., Mahat, A., & Araujo, E. C. (2015). *The economics of health professional education and careers*. The World Bank.

[21] Michavila, F. (Dir.) (2012). *La universidad española en cifras 2012*. Conferencia de Rectores de las Universidades Españolas (CRUE).

[22] Mincer, J. (1974). *Schooling, experience and earnings*. National Bureau of Economic Research.

[23] Mingat, A., & Eicher, J. C. (1982). Higher education and employment markets in France. *Higher Education*, *11*(2), 211-220.

[24] Organisation for Economic Co-operation and Development. (2018). *A broken social elevator? How to promote social mobility*. OECD Publishing.

[25] Organisation for Economic Co-operation and Development. (2023). *Health at a glance 2023: OECD indicators*. OECD Publishing. https://doi.org/10.1787/7a7afb35-en.

[26] O'Neill, L. D., Wallstedt, B., Eika, B., & Hartvigsen, J. (2011). Factors associated with dropout in medical education: A literature review. *Medical Education*, *45*(5), 440-454.

[27] Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, *22*, 1325-1343.

[28] Roemer, M. I. (1982). Market failure and health care policy. *Journal of Public Health Policy*, *3*(4), 419-431.

[29] Rothstein, J. (2019). Inequality of educational opportunity? Schools as mediators of the intergenerational transmission of income. *Journal of Labor Economics*, *37*, S85–S123.

[30] Salas, M. (2006). Private returns to an university education: An instrumental variables approach. *Higher Education*, *51*, 411-438.

[31] Salas, M. (2010). Wage determinants among medical doctors and nurses in Spain. *Higher Education*, *60*, 357-368.

[32] Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal*, *12*(2), 308.

[33] Wooldridge, J. M. (2012). *Introductory econometrics: A modern approach* (5th ed.). South-Western (Cengage Learning).

Appendix

Table A1. Descriptive statistics for students' choice of degree

	Obs.	Mean	Std. dev.
University degree (= 1 medicine)	2,002	0.383	0.486
Non-first-generation university student (= 1 yes)	2,002	0.453	0.498
Gender (=1 female)	2,002	0.762	0.426
Age under 30 years (= 1 yes)	2,002	0.601	0.490
The main motivation for undertaking the undergraduate program			
Getting training, thinking about my labor future	2,002	0.705	0.456
Broadening knowledge for personal satisfaction	2,002	0.160	0.367
Other motives	2,002	0.132	0.338
Mother with secondary or tertiary vocational training (=1 yes)	2,002	0.170	0.376
Father born outside the EU (= 1 yes)	2,002	0.030	0.171

Source: author's calculations

Table A2. Dependent variable: net monthly salary earnings, 2019

	Freq.	Percent
€699 or less	21	1.26
€700 — €999	65	3.91
€1,000 — €1,499	267	16.07
€1,500 — €1,999	572	34.44
€2,000 — €2,499	429	25.83
€2,500 — €2,999	169	10.17
More than €3,000	138	8.31
Total	1,661	100.00

Source: author's calculations

Table A3. Descriptive statistics for explanatory and instrumental variables

	Obs.	Mean	Std. dev.
University degree (= 1 medicine)	1,661	0.413	0.493
Gender (=1 female)	1,661	0.766	0.424
Marital status (= 1 married)	1,661	0.211	0.408
Years of working experience	1,661	3.558	5.625
Part-time job (= 1 yes)	1,661	0.120	0.325
Recognized disability >33% (= 1 yes)	1,661	0.015	0.122
Master's degree (= 1 yes)	1,661	0.460	0.499
Specialist physician (= 1 yes)	1,661	0.285	0.451
Region (Islas Baleares = 1)	1,661	0.035	0.184
Region (Islas Canarias = 1)	1,661	0.063	0.242
Industry of health and social services activities (= 1 yes)	1,661	0.933	0.250

Number of training workshops/seminars since graduation paid by SEPE [a)]	1,661	0.005	0.101
Fluent in English (=1 yes)	1,661	0.269	0.444
Number of spoken languages (excluding mother languages)	1,661	1.494	0.776
Age under 30 years (= 1 yes)	1,661	0.615	0.487
Factors to get the job: theory [b)]	1,661	0.838	0.369
Factors to get the job: management ability [b)]	1,661	0.557	0.497
Non-first-generation university student (= 1 yes)	1,661	0.466	0.499
The main motivation for undertaking the undergraduate program			
Getting training, thinking about my labor future	1,661	0.707	0.455
Broadening knowledge for personal satisfaction	1,661	0.160	0.367
Other motives	1,661	0.131	0.338
Mother with secondary or tertiary vocational training (=1 yes)	1,661	0.169	0.374
Father born outside the EU (= 1 yes)	1,661	0.030	0.169

In italics, continuous variables.

a) SEPE, from its Spanish acronym, is the State Public Employment Service.

b) This variable takes the value of 1 if it was an important or very important factor in obtaining the current job; it takes the value of 0 otherwise.

Source: author's calculations

Table A4. Interval wage regression, including all regressors that are potentially uncorrelated with the error term ε

	Coefficient		Robust Std. Err.
Gender (=1 female)	-124.897	***	24.268
Marital status (= 1 married)	20.002		31.522
Years of working experience	13.224	***	1.992
Part-time job (= 1 yes)	-688.204	***	33.322
Recognized disability >33% (= 1 yes)	-85.499		134.483
Master's degree (= 1 yes)	57.992	**	26.251
Specialist physician (= 1 yes)	685.167	***	48.896
Region (Islas Baleares = 1)	142.776	***	19.540
Region (Islas Canarias = 1)	111.325	***	29.991
Industry of health and social services activities (= 1 yes)	75.819		46.401
Number of training workshops/seminars since graduation paid by SEPE [a)]	-38.743	*	21.185
Fluent in English (=1 yes)	66.058	**	30.949
Number of spoken languages (excluding mother languages)	48.721	***	12.727
Age under 30 years (= 1 yes)	31.661		29.439
Factors to get the job: theory [b)]	88.058	***	32.419
Factors to get the job: management ability [b)]	-53.402	*	27.452
Non-first-generation university student (= 1 yes)	58.048	***	19.554
The main motivation for undertaking the undergraduate program			
Getting training, thinking about my labor future			
Broadening knowledge for personal satisfaction	**48.717**		34.278
Other motives	55.352		42.309
Mother with secondary or tertiary vocational training (=1 yes)	**14.713**		24.080
Father born outside the EU (= 1 yes)	-75.371		71.850
Constant	1603.584	***	62.088
/lnsigma	6.144	***	0.025
sigma	465.706		11.610
	[95% conf. interval]		
	[443.50, 489.03]		

Number of obs. = 1,661
Uncensored = 0
Left-censored = 21
Right-censored = 138
Interval-cens. = 1,502
Wald chi2(21) = 2350.41
Prob > chi2 = $p < 0.001$
Log pseudolikelihood = -2286.3871
Std. Err. adjusted for 49 clusters in the province of the current job.
a) SEPE, from its Spanish acronym, is the State Public Employment Service.
b) This variable takes the value of 1 if it was an important or very important factor in obtaining the current job; it takes the value of 0 otherwise.

* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$
Source: author's calculations