

Kugler, Philipp

**Article — Published Version**

## The role of wage beliefs in the decision to become a nurse

Health Economics

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Kugler, Philipp (2021) : The role of wage beliefs in the decision to become a nurse, Health Economics, ISSN 1099-1050, Wiley, Hoboken, NJ, Vol. 31, Iss. 1, pp. 94-111, <https://doi.org/10.1002/hec.4442>

This Version is available at:

<https://hdl.handle.net/10419/284770>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

## RESEARCH ARTICLE

# The role of wage beliefs in the decision to become a nurse

Philipp Kugler 

Institute for Applied Economic Research,  
Tübingen, Germany

**Correspondence**

Philipp Kugler, Institute for Applied  
Economic Research, Schaffhausenstr. 73,  
72072 Tübingen, Germany.  
Email: [philipp.kugler@iaw.edu](mailto:philipp.kugler@iaw.edu)

**Funding information**

European Regional Development Fund,  
Grant/Award Number: FEIH ZAFH  
795924; Bundesministerium für Bildung  
und Forschung  
Open Access funding enabled and  
organized by Projekt DEAL

**Abstract**

In light of skilled-labor shortage, the effect of a change in the wage of nurses on their labor supply is intensely discussed in recent literature. Using extensive data of German 14- to 15-year-olds, I analyze the role of the beliefs about a nurse's wage in the decision to become one. To estimate a partial effect, I select controls and their functional form using post-double-selection, which is a data-driven selection method based on regression shrinkage. Highlighting the importance of wages at the extensive margin of labor supply, the wage beliefs play a positive and statistically significant role. Although information is publicly available, educational choices knowingly suffer from misinformation. I find that especially those who do not become a nurse understate the wage. The results lead to two important policy implications. First, increasing the wage may help to overcome the shortage observed in many countries. Second, providing more information on the (relative) wage may be a successful strategy to attract more individuals into this profession. To assess the sensitivity of the results regarding omitted variable bias, I apply a novel approach. It turns out that potential unobserved confounders would have to be strong to overrule the conclusions.

**KEYWORDS**

health professional, machine learning, occupational choice, wage beliefs, wage information

## 1 | INTRODUCTION

Due to demographic change and technological progress in medicine, the demand for skilled nurses has increased in industrialized countries over the past decades (German Employment Agency, 2020a). This trend will continue in the coming years and will further aggravate the lack of nurses. In Germany, on average new job offers were vacant for 118 days before they could be staffed in 2019. However, in nursing, more than 175 days were needed to fill a position, indicating a lack of skilled workers (German Employment Agency, 2020b). Although limits in capacities were never reached in Germany, the lack of nurses has been more tangible than ever in the recent Covid-19 pandemic (Begerow et al., 2020). As a consequence quality in care decreased (Kaltwasser et al., 2021; Krebs et al., 2021). To counteract this development, it is important to analyze and to understand the occupational behavior of nurses. The existing literature discusses a series of factors that might alleviate the lack of skilled workers. These include individual preferences of (future) nurses, improving working conditions and increasing wages, as well as the trade-off between monetary and non-monetary job characteristics (e.g., Eberth et al., 2016; Scott et al., 2015). I contribute to this discussion by analyzing the effect of the beliefs about

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) 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.

© 2021 The Authors. Health Economics published by John Wiley & Sons Ltd.

a nurse's wage of young students on the probability to become one. Wages are the most controversially discussed factor in the literature. Some authors identify it as a very important factor influencing labor supply decisions of nurses (Doiron et al., 2014; Hanel et al., 2014). However, others suggest that the labor supply of nurses is relatively inelastic in terms of wages. Factors such as personal attitude and working conditions seem to play a much larger role (McCabe et al., 2005; Shields, 2004).

Since there are large differences in earnings depending on the occupational choices, the economic literature on the effect of the expected wage is rich (Altonji et al., 2016). The majority of studies agree that the wage has a significant and positive effect on the career choice (e.g., Boudarbat, 2008; Montmarquette et al., 2002). Nonetheless, most studies find that preferences and interests play a larger role in career choice than the wage expectations (Arcidiacono, 2004; Beffy et al., 2012).

In line with the economic literature, the nursing literature suggests preferences and interests to be the most important factors influencing the decision to become a nurse. In particular, caring for people is identified as the key reason for choosing the profession (e.g., Matthes, 2019; Petrucci et al., 2016; Wilkes et al., 2015). Concerning the wage, several studies find that it only plays a minor role in the decision-making process (e.g., Bomball et al., 2010; Cho et al., 2010; McCabe et al., 2005). Based on these results, policy-makers might be tempted to focus on non-monetary factors to attract more young people into nursing. However, this contrasts recent work by Hanel et al. (2014) and Schweri and Hartog (2017). Schweri and Hartog (2017) examine the effect of ex-ante wage expectations on the decision to pursue a nursing degree (tertiary education) by using data on healthcare trainees (upper-secondary education) in Switzerland. Therefore, they analyze the decision on the intensive margin. Their results show that the greater ex-ante wage expectations of a nursing degree, the higher the probability to pursue such a degree later on. This indicates that higher wages may attract more students to become a high-skilled nurse. Hanel et al. (2014) estimate a model of labor supply decisions using data on individuals who hold a nursing qualification. The model accounts for the intensive and extensive margin by allowing individuals to enter and to exit occupations. As a result, they find a considerable high wage elasticity. This differs fundamentally from other work that detect very small elasticities (Andreassen et al., 2017; Shields, 2004). These differences can be fully explained by the frequent neglect of the extensive margin and the exclusive analysis of the intensive margin. Although Hanel et al. (2014) do not account for the choice of becoming a nurse, their results suggest that wages may heavily drive the career choice, that is, a decision on the extensive margin.

The focus of this paper is on beliefs about wages and how they influence the decision to become a nurse. Such beliefs may affect educational choices and could be easily changed by policy-makers—at least compared to other factors such as preferences. For example, Jensen (2010) analyzes perceived returns to secondary schooling of students in the Dominican Republic. He finds that the expected returns are underestimated. By providing information, students completed more years of education.

I use extensive data of former German ninth graders, 14- to 15-year-olds who are about to obtain a lower secondary degree. They have been followed since. It contains information on the wage that young students think a nurse, a hairdresser, a motor vehicle mechanic, a bank clerk, a teacher and a physician earns before occupational choice takes place. This information enables me to estimate the effect of the beliefs about a nurse's wage on the probability to become one. Moreover, I estimate the effect of other factors (e.g., social orientation) on the probability of choosing the profession of a nurse. This allows to assess the magnitude of the impact of the beliefs about a nurse's wage and to fit my results into the recent literature.

In addition, the data contains extensive background information on the individuals measured in ninth grade, that is, before their occupational decision took place. This covers not only educational and parental background but also measures for personality, competencies, interests and attitudes. Overall, the data allows to observe over 150 characteristics. By applying the lasso proposed by Tibshirani (1996), a method that draws coefficients toward zero or exactly to zero, I am able to select the relevant controls and to model non-linearities in confounding. However, the lasso is tailored to choose variables such that an outcome is precisely predicted. Therefore, it cannot be applied directly for variable selection, when the aim is to estimate a partial effect. As a solution, Belloni et al. (2012, 2014b) propose the *post-double-selection* (PDS), which is a two-step procedure to identify relevant controls and their functional form. To interpret the estimated effect as causal, I need to assume that no factors affecting the dependent variable and the variable of interest remain unobserved (unconfoundedness). Despite the ability to condition on a rich set of controls and flexibly model their functional form, this assumption is very strict and likely to be violated in some way. To mitigate the concerns about omitted variable bias and to get an idea about its consequences, I follow a novel approach by Cinelli and Hazlett (2020). For linear models, they propose to assess the minimal strength that unobserved confounding needs to have on the wage beliefs and on the career choice in order to change the conclusion. To this end, Cinelli and Hazlett (2020) propose a procedure for benchmarking

based on observed covariates. The knowledge about main predictors for career choice or the wage beliefs is the crucial premise for the benchmarking to be valuable. Fortunately, literature on determinants of wage expectations and factors driving young people into nursing is rich. Thus, credible benchmarking on observed covariates is possible.

My results show that higher beliefs about a nurse's wage increase the probability to become a nurse. In line with recent literature, individual preferences play a larger role than the beliefs. Since the career choice is a decision on the extensive margin my results are also consistent with those of Hanel et al. (2014). Further, the results show that effects are driven by young people who do not become a nurse and underestimate the wage. This is consistent with recent literature. Although information is publicly available, educational choices are made under uncertainty and incomplete or incorrect information (Finger et al., 2020; Hastings et al., 2016; Oreopoulos & Dunn, 2013). With regard to nursing, Dante et al. (2013) find that especially students who do not become a nurse basically know nothing about it (e.g., understate initial wages). The results of this paper indicate, that the public perception of wages in nursing is too low. Therefore, nursing is less attractive than other occupations for which wages are not systematically understated. To combat the lack of skilled nurses, policy-makers can make the profession more attractive by increasing the beliefs about a nurse's (relative) wage.

The remaining paper is structured as follows. Section 2 outlines the methods applied in the empirical analysis and briefly describes the data, the wage belief measures as well as the control variables. In Section 3, I present and discuss the main results of my analysis. Section 4 concludes.

## 2 | METHODS

### 2.1 | Empirical strategy

#### 2.1.1 | Post-double-selection

The partial effect of the wage belief  $w_i$  on the probability to become a nurse is estimated by a partially linear model

$$y_i = \beta w_i + g(x_i) + \zeta_i, \quad (1)$$

where  $y_i \in \{0,1\}$  denotes the binary choice to become a nurse. The function  $g(x_i)$  is unknown and potentially complicated. I approximate it by a linear combination that may include higher order polynomials and interactions

$$g(x_i) = x_i' \theta_y + r_{yi}, \quad (2)$$

where  $r_{yi}$  is an approximation error. The aim is to estimate  $\beta$ . However, it is a difficult task to define a set of variables to be included in the model and to model their functional form (i.e., what polynomials and interactions to include). Therefore, I rely on data-driven variable selection and follow the PDS approach proposed by Belloni et al. (2012, 2014b). The lasso is a shrinkage method that imposes a penalty on the size of the coefficients, that is, shrinks them toward zero or exactly to zero. This prevents models with many variables that are correlated with each other from overfitting (Hastie et al., 2009). The lasso is defined as

$$\hat{\gamma}^{\text{lasso}} = \underset{\gamma}{\operatorname{argmin}} \left\{ \underbrace{\frac{1}{2} \sum_{i=1}^N \left( y_i - \gamma_0 - \sum_{j=1}^p x_{ij} \gamma_j \right)^2}_{\text{residual sum of squares}} + \underbrace{\lambda \sum_{j=1}^p |\gamma_j|}_{\text{penalty term}} \right\}, \quad (3)$$

where  $\sum_{j=1}^p |\gamma_j|$  imposes the penalty on the size of the coefficients and the parameter  $\lambda \geq 0$  controls the magnitude of the punishment.

A naive approach to estimate  $\beta$  would be to apply the lasso estimator to Equation (1) and to exclude  $\beta$  from the penalty term such that it is enforced to stay in the model. Afterward one might use a least-squares regression of the outcome on  $w_i$  and controls with non-zero coefficients. However, this approach leads to biased estimates because of omitted variables. The lasso is designed to learn a forecasting rule of  $y_i$  given  $w_i$  and  $x_i$  and not to learn about the relationship between  $y_i$  and  $w_i$  given controls  $x_i$  (Belloni et al., 2014a). Therefore, lasso cannot be used off the shelf for the estimation of partial effects. As a solution, Belloni et al. (2012, 2014b) propose an intuitive and easy-to-implement procedure. First, the lasso is

used to estimate a model predicting the outcome given  $x_i$  in Equation (4) and a further model predicting the wage beliefs given  $x_i$  in Equation (5)

$$y_i = x_i' \pi + \epsilon_i, \quad (4)$$

$$w_i = x_i' \theta_w + v_i. \quad (5)$$

Subsequently, all variables with non-zero coefficients in either of the two models are kept as control variables in order to estimate  $\hat{\beta}$  in Equation (1) by an ordinary least-squares regression. This step is known as the “post-lasso.” The crucial assumption under which PDS works is approximate sparsity. It states that the wage belief and the career choice can be approximated by Equations (4) and (5) using only a small number of covariates relative to the sample size. Note, that approximate sparsity is also implicitly assumed in conventional ordinary least squares (OLS) analysis where no double selection by lasso takes place. Additional variables that are considered as important for ensuring robustness, can be included (amelioration set). The condition is that the amelioration set is not substantially larger than the number of variables chosen via the lasso (Belloni et al., 2014b).

The choice of  $\lambda$  is of importance. With the aim of prediction, standard lasso applications choose  $\lambda$  by cross-validation. However, this analysis aims to estimate a partial effect. If  $\lambda$  is too large, only a few variables are selected and omitted variable bias may occur. If  $\lambda$  is too small, the number of variables is very large such that overfitting may become an issue. Therefore, I follow Urminsky et al. (2016) and use  $\lambda = 1.1 \sigma_R \frac{1}{\sqrt{N}} \Phi^{-1} \left( 1 - \frac{0.1}{\ln(N)2p} \right)$ , where  $N$  is the number of observations,  $p$  is the number of potential controls,  $\Phi^{-1}$  denotes the inverse cumulative function of the standard normal distribution and  $\sigma_R$  the standard deviation of the residuals of the model. Finally, it is important to note that the chosen variables are not interpretable since selection depends on the sample (Mullainathan & Spiess, 2017). Hence, when presenting the results, chosen control variables are not reported.

### 2.1.2 | Sensitivity

In order to interpret the partial effect  $\hat{\beta}$  as causal, I need to rely on the assumption of unconfoundedness  $\mathbb{E}[\zeta_i | w_i, r_{yi}, x_i] = 0$ . It states that all factors that affect the choice  $y_i$  and the wage belief  $w_i$  at the same time must be contained in  $g(x_i)$ . Even though I have access to an extensive set of potential controls  $x_i$ , bias due to unobserved confounders may be likely. For example, covariates measuring the interests of the individuals might not fully capture all relevant aspects but only a share of it. Further, it cannot be ruled out, that some factors may remain fully unobserved. Moreover, the assumption of approximate sparsity may be violated. There may exist covariates that are not selected by lasso but affect both, the wage belief and the decision to become a nurse. To analyze the sensitivity of the results due to potentially unobserved (non-)linear confounding factors  $z$ , I make use of a procedure proposed by Cinelli and Hazlett (2020). In a nutshell, they propose to assess the sensitivity of the estimates by analyzing whether a confounder is strong enough to change the conclusion if it is as strong as a very good predictor of  $y$  or  $w$ .

Conventionally, the omitted variable bias can be written as  $\widehat{bias} = \hat{\gamma} \hat{\delta}$ . Hence,  $\hat{\gamma}$  describes the difference in the linear expectation of the outcome if  $z_i$  changes by one unit, holding everything else constant and  $\hat{\delta}$  describes the difference in linear expectation of the confounder if the variable of interest changes by one unit, holding everything else constant (Cinelli & Hazlett, 2020). Arguing that both quantities  $\hat{\delta}$  and  $\hat{\gamma}$  are hard to grasp, Cinelli and Hazlett (2020) write the conventional omitted variable bias formula in terms of partial  $R^2$  measures. Those are easier to interpret and can be exploited for further analysis. Denote  $\hat{\beta}_{obs}$  as the observed estimated effect and  $\hat{\beta}$  as the estimated effect from a model controlling unobserved confounding factors, that is,  $\hat{\beta} = \hat{\beta}_{obs} - \widehat{bias}$ . Then, they show that

$$|bias| = \widehat{se}(\hat{\beta}_{obs}) \sqrt{\frac{R_{y \sim z | w, x}^2 R_{w \sim z | x}^2}{1 - R_{w \sim z | x}^2}} df, \quad (6)$$

where  $df$  defines the degrees of freedom,  $R_{y \sim z | w, x}^2$  stands for the partial  $R^2$  of regressing  $y$  on  $z$  after controlling for  $w$  and  $x$  and  $R_{w \sim z | x}^2$  denotes the partial  $R^2$  of regressing  $w$  on  $z$  after controlling for  $x$ . Further, the standard error of  $\hat{\beta}$  can be written as

$$\widehat{se} = \widehat{se}(\beta_{\text{obs}}) \sqrt{\frac{1 - R_{y \sim z|w,x}^2}{1 - R_{w \sim z|x}^2} \left( \frac{df}{df - 1} \right)}, \quad (7)$$

and the adjusted  $t$ -statistic is defined as  $t_{\text{adj}} = \hat{\beta} / \widehat{se}$ . Applying these definitions,  $\hat{\beta}$ ,  $\widehat{se}$  and  $t_{\text{adj}}$  can be computed by substituting reasonable values for  $R_{y \sim z|w,x}^2$  and  $R_{w \sim z|x}^2$ , that is, the strength of confounding, into Equations (6) and (7). However, actual knowledge about the absolute strength is seldom available. As a solution, Cinelli and Hazlett (2020) argue that the researcher is often able to make a statement on the relative strength of potential unobserved confounding, for example,  $z$  cannot account for as much variation of the outcome as some observed covariate  $x_j$ . There are several ways to formalize such claims. I follow Cinelli and Hazlett (2020) and claim that I measure the key determinant of  $y$  and  $w$  such that the omitted variable cannot explain as much residual variance in  $y$  or  $w$  as this determinant. Define

$$k_w = \frac{R_{w \sim z|x_{-j}}^2}{R_{w \sim x_j|x_{-j}}^2} \quad (8)$$

$$k_y = \frac{R_{y \sim z|x_{-j},w}^2}{R_{w \sim x_j|x_{-j},w}^2}, \quad (9)$$

where  $x_{-j}$  is a vector including all variables contained in  $x$ , excluding  $x_j$ . The ratios  $k_w$  and  $k_y$  show how much of the variance in  $w$  or  $y$  is explained by  $z$  relative to the explanatory power of  $x_j$ , conditional on all other covariates. In this paper  $k_w = k_y = 1$ , that is, I consider the impact of a confounder  $z$  that is as strong as  $x_j$ . Given  $k_w$  and  $k_y$ , Cinelli and Hazlett (2020) show that

$$R_{w \sim z|x}^2 = k_w f_{w \sim x_j|x_{-j}}^2 \quad R_{y \sim z|w,x}^2 \leq k_y \eta^2 f_{y \sim x_j|x_{-j},w}^2, \quad (10)$$

where  $\eta$  is a scalar that depends on  $k_w$ ,  $k_y$ , and  $R_{w \sim x_j|x_{-j}}^2$ . Furthermore,  $f_{w \sim x_j|x_{-j}}^2$  denotes partial Cohen's  $f$  of  $w$  on  $x_j$  and  $f_{y \sim x_j|x_{-j},w}^2$  denotes partial Cohen's  $f$  of  $y$  on  $x_j$ .<sup>1</sup> Cinelli and Hazlett (2020) have shown that these robustness results are exact for a single linear confounder and conservative for multiple, possibly nonlinear, confounding factors.

It is important to emphasize that this bounding procedure heavily relies on the choice of the benchmark variable  $x_j$ . If it is not true that  $x_j$  is a key predictor of the outcome or treatment, the bounding is pointless. Hence, domain knowledge is necessary (Cinelli & Hazlett, 2020). In the following, I choose observed covariates that are often discussed in the literature. First, bounding is based on social orientation, that is, preferring activities to inform, train, educate, cure, or advise other people (Wohlkinger et al., 2011). It is the key characteristic of those who become a nurse (e.g., Matthes, 2019), while preferences are generally a decisive factor in career choice (e.g., Arcidiacono, 2004). In addition, interests also play an important role in the formation of expected wages (Wiswall & Zafar, 2015). Second, the professions of the parents play an important role in the occupational choice (e.g., Knoll et al., 2017). Therefore, the results are bounded by an indicator that indicates whether at least one of the parents is a nurse. Moreover, parents in nursing might inform their children about the wages in nursing. Third, an indicator for gender is considered. Females become nurses much more often than males (Speer, 2019). Moreover, gender also plays a crucial role in wage expectations: females expect lower wages than males (e.g., Brunello et al., 2004; Fernandes et al., 2020). Fourth, (perceived) ability determines the expected wages (Brunello et al., 2004). Therefore, a measure for ability, namely metacognition, is used to bound the results. Note, that these variables have to be part of the model in order to use them as benchmark variables. Hence, the amelioration set contains these four variables, to ensure that they are not excluded by data-driven variable selection.

## 2.2 | Data

This study uses Starting Cohort Four (SC4) of the German National Educational Panel Study (NEPS). The survey collects data on 14- to 15-year-olds who attended the ninth grade in German regular schools in 2010 and has been followed since (Blossfeld & von Maurice, 2011). This includes grammar schools, middle secondary schools, lower secondary schools, comprehensive schools, and schools offering all tracks of secondary education except the grammar school track. Since



becoming a nurse requires a vocational training, all degrees that can be obtained at these schools are sufficient to be admitted. For several reasons the data is highly suitable for investigating the role of beliefs about a nurse's wage in the decision to become one. Since the data is available from 2010 to 2016, the transition from school to further education can be observed in great detail and no retrospective information has to be used. The following analysis is based on a cross section of the panel and focuses on the choice of the first occupational training, which certainly has a relevant impact on the further life course. The German education system is highly complex. Although it is a sequential system, there are many different paths through it that lead to the same degree, that is, to down- or upgrade educational paths (Biewen & Tapalaga, 2017). Upon completion of ninth grade, individuals obtain the lowest educational degree. Afterward, individuals may start a vocational training or, alternatively, stay in the educational system, which in turn may affect the individual characteristics and occupational choice later on. In order not to condition on future outcomes, the cross section contains information of individuals that is measured in ninth grade.<sup>2</sup> Beyond that, the individuals are asked to state their beliefs about the monthly salary of a nurse, a hairdresser, a motor vehicle mechanic, a bank clerk, a teacher and a physician: "Now we are also interested to know how high you think the income is in certain professions. What do you think the monthly income as a [...] is?" Consequently, the question at hand captures knowledge about average wages, knowledge of wages according to collective agreements, but also wrong beliefs due to the lack of information or wrong perceptions of wages. Since a central principle of the human capital theory suggests schooling decisions are made by comparing the benefits with the (opportunity) costs of the decision, the analysis focuses on the relative wage beliefs, that is, the stated nurse's wage relative to stated wages of other occupations (Boudarbat, 2008; Carneiro et al., 2011). To this end, the stated wages of all six occupations are ranked from lowest to highest. If the wage cannot be assigned unambiguously due to ties, the mean rank is assigned such that the sum of ranks is preserved. Formally, I define the  $i$ th individual's rank of a nurse's wage as

$$\text{rank}_i^{\text{nurse}} = 1 + \sum_{w_i \in \{S_i \setminus w_i^{\text{nurse}}\}} \mathbb{1}(w_i < w_i^{\text{nurse}}) + 0.5 \times \mathbb{1}(w_i = w_i^{\text{nurse}}), \quad (11)$$

where  $\mathbb{1}(\cdot)$  denotes an indicator function that takes the value 1 if the expression in the parentheses is true,  $S_i$  is the set of surveyed wage beliefs and  $w_i^{\text{nurse}}$  is the belief about a nurse's wage. Two further measures are defined as the ratio between individual's  $i$  beliefs about a nurse's wage and maximum as well as minimum stated wage

$$\text{relwage}_i^{\text{nurse, max}} = \frac{w_i^{\text{nurse}}}{w_i^{\text{max}}}, \quad (12)$$

$$\text{relwage}_i^{\text{nurse, min}} = \frac{w_i^{\text{nurse}}}{w_i^{\text{min}}}. \quad (13)$$

Further, I also analyze the belief of a nurse's absolute wage.

Based on the ranking measure in Equation (11), I can easily assess how close the relative wage beliefs are to reality by computing the deviation from the true ranks. The median wages reported by German Employment Agency (2018) provide the basis for the true rank. According to this source of information, the following *true* ranking from lowest to the highest wage was established: (1) hairdresser, (2) motor vehicle mechanic, (3) nurse, (4) bank clerk, (5) teacher and (6) physician. The ranking is utilized to construct a measure that captures the knowledge about relative wages by adding the absolute deviations of the stated rank of each occupation

$$\text{rank}_i^{\text{abs. dev.}} = \left| \text{rank}_i^{\text{hairdresser}} - 1 \right| + \left| \text{rank}_i^{\text{mechanic}} - 2 \right| + \left| \text{rank}_i^{\text{nurse}} - 3 \right| + \left| \text{rank}_i^{\text{bank clerk}} - 4 \right| + \left| \text{rank}_i^{\text{teacher}} - 5 \right| + \left| \text{rank}_i^{\text{physician}} - 6 \right|, \quad (14)$$

where the ranks of each occupation are computed the same way as the rank of a nurse's wage. Additionally, I can construct indicators that show whether the wage rank of a nurse is overestimated, correctly estimated or underestimated

$$\text{rank}_i^{\text{nurse, over}} = \mathbb{1}(\text{rank}_i^{\text{nurse}} > 3), \quad (15)$$

$$\text{rank}_i^{\text{nurse, correct}} = \mathbb{1}(\text{rank}_i^{\text{nurse}} = 3), \quad (16)$$

$$\text{rank}_i^{\text{nurse, under}} = \mathbb{1}(\text{rank}_i^{\text{nurse}} < 3). \quad (17)$$

Besides information on wage beliefs, there are other potentially important factors available that may drive young people into or out of nursing (see Wohlking et al., 2011). This enables me to assess the importance of the wage belief by comparing the effect with other effects estimated in the literature. A large share of recent work analyses the effect of working conditions. Hence, I look at the effect of *desired work conditions* using the survey instrument provided by MOW International Research Team (1987). Since literature finds that those who become nurses do not rate the importance of economic factors as important as those who choose another profession, I use a measure of the importance of economic factors (i.e., risk of unemployment and financial aspects) in choosing a career. In addition, I estimate a model that uses an indicator of the desired comfort of the job (i.e., physical working conditions and good working hours). Finally, helping others is considered to be one of the main driving forces in choosing a nursing profession. I use a measure that captures if individuals prefer activities to inform, train, educate, cure, or advise other people. It is surveyed by the Interest Inventory Life Span developed by NEPS (Wohlking et al., 2011).

Extensive background information about the individuals is crucial to identify a causal effect. Due to the complexity of occupational choice and the formation of wage beliefs, factors may remain unobserved. To mitigate this concern, I use many potential control variables from different areas that either directly measure important confounder or could serve as proxies. This includes demography, personal characteristics such as competencies, personality and leisure activities, as well as information on parental background, school and regional labor market. As described above, I use PDS to handle the large set of potential controls. All potential controls are summarized in Table S1. The exclusion of observations with at least one missing value would lead to a substantial loss of observations. Therefore, I impute missing values by chained equations (van Buuren & Groothuis-Oudshoorn, 2011). To estimate the impact of the wage belief via Equation (1), I account for non-linearities in confounding by interacting all variables with each other and by additionally including fifth-order polynomials of non-binary covariates. As a result, 13.878 potential controls are available.

After excluding individuals with extreme beliefs, missing values in variables of interests<sup>3</sup> or with too many missing observations in general,<sup>4</sup> I observe 7089 individuals that transition from school to occupational training, of whom 238 chose nursing.

### 3 | RESULTS

#### 3.1 | The role of wage beliefs

##### 3.1.1 | Descriptive evidence

First, I provide some insights about the univariate relationship between the wage beliefs and the choice whether or not to become a nurse. Table 1 depicts the shares of the nurse's wage ranks reported by nurses and others. Young individuals most often rank a nurse's wage second or third. Compared to other individuals, future nurses much less often rank the nurse's wage lowest or second. In turn, for a rank larger than two, the respective shares of future nurses are larger. This indicates that future nurses have higher wage beliefs than other young individuals. To gain further insights into the differences in wage ranks, Table 2 presents the mean differences of the wage rank by nurses and others. Table 2 reveals that, on average, nurses state a lower rank of the wage earned by hairdressers or mechanics than others but state a larger rank of a nurse's wage. Interestingly, the average ranks of a bank clerk, teacher and physician are not significantly different between nurses and others. This means, that beliefs only differ for occupations with lower wages.

	Rank of a nurse's wage					
	1	2	3	4	5	6
Nurses	5.88	39.07	36.13	13.87	4.62	0.42
Others	14.86	48.65	27.63	6.38	1.96	0.53
All	14.55	48.32	27.92	6.63	2.05	0.52

TABLE 1 Distribution of the wage rank

Note: The table depicts the share of the beliefs about a nurse's wage rank by nurses and others. For the sake of clarity, in the case of ties, the lower rank is reported.



In Figure 1a–c, the remaining measures of wage beliefs are depicted in quantile plots. Figure 1a shows the ratio of the belief of a nurse's wage and the highest stated wage belief as defined in Equation (12). Differences in relative wage beliefs between those who become a nurse and those who do not, are very clear. Except for the lower and upper end, future nurses state higher wages. Similarities in lower and upper ends indicate that extreme beliefs do not differ systematically between groups. In Figure 1b, the distribution of the ratio between the belief of a nurse's wage and the lowest stated wage belief is shown. Except for the lower and upper end, the relative wage belief is larger for future nurses along the whole distribution. In summary, descriptive evidence consistently suggests that future nurses state a higher relative wage than those who do not become a nurse. As Figure 1c reveals, not only the beliefs about the nurse's relative wage is higher for future nurses. At least in quantiles in the middle, the absolute wage belief is also slightly higher.

### 3.1.2 | Results of PDS

The observed descriptive differences may be caused by confounding. For example, those who have no interests in becoming a nurse may state rather low wage beliefs (e.g., Wiswall & Zafar, 2015). However, to draw more valuable policy implications, it is useful to analyze the role of beliefs in becoming a nurse, given the characteristics of the individuals (i.e., equally interested, same background, same skills, etc.). As described in Section 2, I tackle this issue by using PDS to estimate the partial effect of the wage belief on the probability to become a nurse. This effect will be referred to as *the effect* throughout Section 3.1. The results are depicted in Table 3. Each of the three columns shows the results of an unconditional model (single OLS) and the post-lasso (a conditional OLS model with controls chosen by double-selection). In the first column, future nurses are compared to all remaining young people. However, this neglects the heterogeneity of the effect of (relative) wage beliefs. Individuals who are interested in becoming a nurse, for example, chose a similar occupation, may be more responsive to wage beliefs compared to those who have no interest in nursing at all. Hence, in the remaining columns the sample is restricted regarding the career choices. I compare future nurses to (2) young people who opted for vocational training and (3) individuals who chose a social field.<sup>5</sup> Each panel of the table depicts the results of one of the four measures described above. The measures are standardized to have mean 0 and standard deviation 1 such that the results can easily be compared with other factors in the subsequent section.<sup>6</sup>

The first panel shows the estimated effects of a nurse's rank on the probability to become a nurse. Column one compares future nurses to all the remaining individuals in the data. As expected from the descriptive results in Tables 1 and 2, results of the unconditional model show that an increase of the rank by one standard deviation is associated with a statistically significant increase in the probability to become a nurse by 1.4 percentage points. When relevant controls are added, I still observe a statistically significant change by 1 percentage point. The slightly smaller coefficients of the post-lasso compared to the single OLS model shows that those, who are prone to become a nurse (e.g., having parents that are nurses and young people that have a social attitude) have higher relative wage beliefs. The results change only slightly with regard to the comparison group in column two. Comparing future nurses with individuals who chose a social field, I find a much larger effect in both the unconditional and conditional model. These results hint to heterogeneity in the effect, that is, larger effects for those who chose a more similar field. Interestingly, the number of included controls is much smaller when the comparison group only consists of individuals who chose a vocational training or a social field. However, this is expected since the sample size is much smaller. Thus,  $\lambda$  becomes larger and draws the coefficients of the lasso models stronger toward zero. Further, the sample in column 1 is more heterogeneous than the ones in column 2 and 3. Therefore, fewer variables may be required to explain differences.<sup>7</sup>

The second panel contains the results of the effect of the ratio of the beliefs about a nurse's wage and the highest stated wage. Results for the entire sample in the first column show that even after controlling for relevant confounders, I find a statistically significant and positive effect on the probability to become a nurse. Similar to results of the wage rank, the coefficient in the post-lasso model is smaller than in the unconditional model. The effect stays positive and significantly different from zero when the comparison group is changed. For those who chose a social field in column 3, the effect is again much larger. This suggests effect-heterogeneity.

The following panel shows the impact of the ratio of the beliefs about the wage of a nurse and the lowest stated wage on the decision to become a nurse. Comparing those who become nurses with all other individuals, results of the unconditional model show that an increase in the relative wage increases the probability to become a nurse by 0.9 percentage points. When relevant controls are included, the probability increases by 0.8 percentage points. Just like the estimates in the first and second panel, the results also indicate heterogeneity. The effects become even larger when the comparison group consists of young people who chose a social field.

TABLE 2 Differences in mean wage ranks by nurses and others

	Mean of others	Mean of nurses	Difference	p-values of test for differences in means		
				$H_0 : \text{diff.} < 0$	$H_0 : \text{diff.} = 0$	$H_0 : \text{diff.} > 0$
Stated rank of a						
Hairdresser	1.29	1.16	0.12	1.00	0.00	0.00
Mechanic	2.74	2.56	0.18	1.00	0.00	0.00
Nurse	2.46	2.84	-0.39	0.00	0.00	1.00
Bank clerk	4.53	4.49	0.04	0.73	0.54	0.27
Teacher	4.41	4.35	0.06	0.83	0.33	0.17
Physician	5.56	5.58	-0.14	0.40	0.80	0.60

Note: The table depicts the means of the wage ranks by future nurses and others together with their differences. Further, to assess whether differences are statistically significant, *t*-tests are conducted.

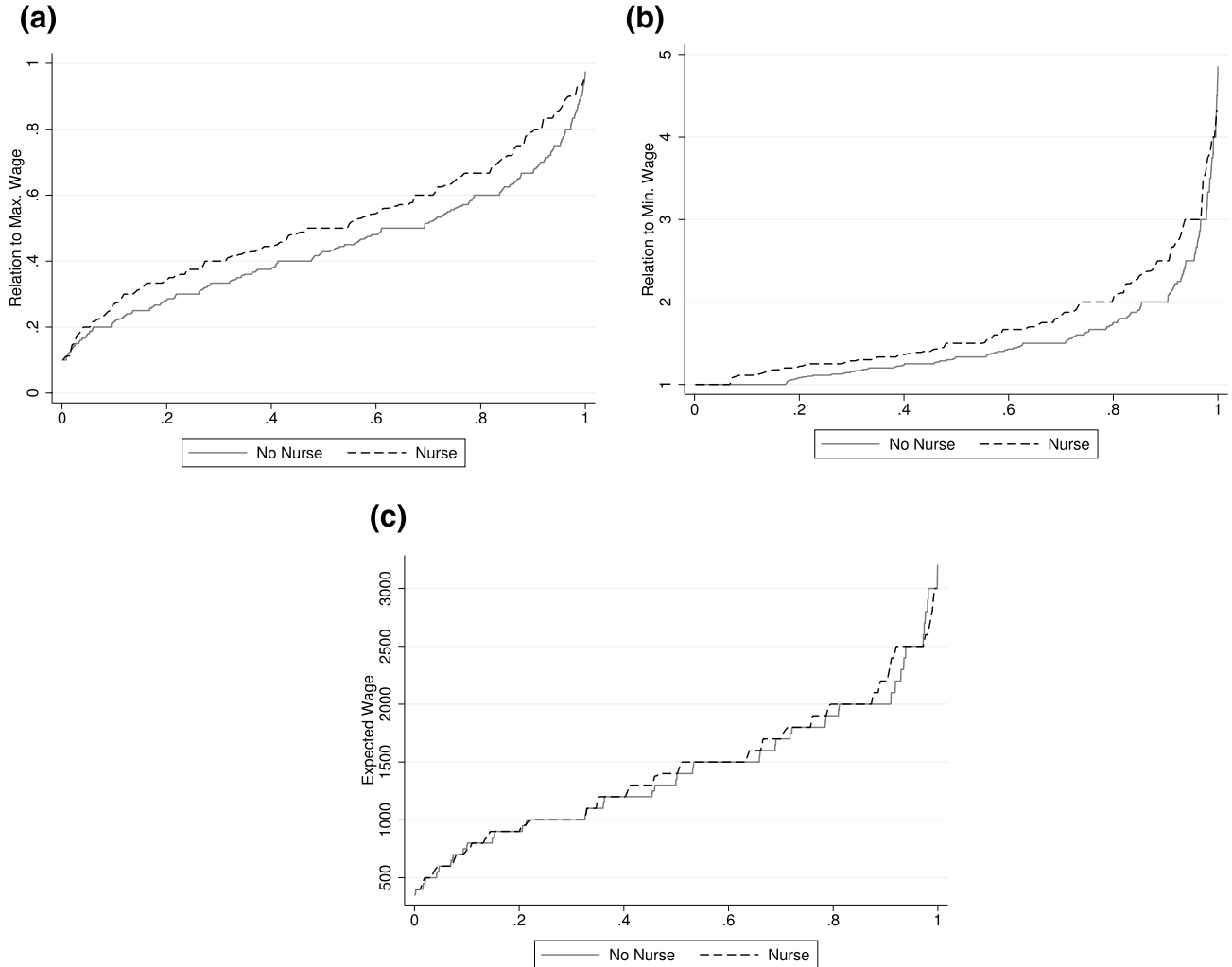


FIGURE 1 Continuous wage belief measures. (a) Ratio of belief of a nurse's wage and maximum stated wage. (b) Ratio of belief of a nurse's wage and minimum stated wage. (c) Beliefs of the absolute wage of a nurse. Each panel depicts a quantile plot of one wage measure. The ratio of the belief of a nurse's wage and maximum stated wage belief is defined as  $relwage_i^{nurse,max} = w_i^{nurse} / w_i^{max}$ , and the ratio of the belief about a nurse's wage and minimum stated wage belief is defined as  $relwage_i^{nurse,min} = w_i^{nurse} / w_i^{min}$ . Extreme values are excluded

TABLE 3 Beliefs about a nurse's wage

	(1)		(2)		(3)	
	Nurse versus all		Nurse versus vocational training		Nurse versus social field	
	Single OLS	Post-lasso	Single OLS	Post-lasso	Single OLS	Post-lasso
<b>Rank of nurse's wage</b>						
	0.014***	0.010***	0.015***	0.012***	0.047***	0.035***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.009)
<i>p</i>	-	41	-	25	-	13
<b>Nurse's wage/highest wage</b>						
	0.011***	0.006***	0.009***	0.006*	0.036***	0.023***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.009)
<i>p</i>	-	38	-	16	-	20
<b>Nurse's wage/lowest wage</b>						
	0.009***	0.008***	0.011***	0.011***	0.025***	0.022***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.007)
<i>p</i>	-	24	-	19	-	9
<b>Nurse's absolute wage</b>						
	0.002	0.008***	0.006**	0.011***	0.021**	0.028***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.009)
<i>p</i>	-	29	-	23	-	13
$N^{\text{nurse}}$	238		238		238	
$N^{\text{comp.}}$	6851		4214		1378	

Note: The table depicts the results of the effect of the wage belief on the decision to become a nurse. The rank of nurse's wage is defined in Equation (11) and the ratio of the nurse's wage and highest/lowest stated wage is defined in Equations (12) and (13), respectively. Standard errors are shown in parentheses. Significance of the coefficient at conventional significance levels 1%, 5%, 10% are indicated by stars \*\*\*, \*\*, \*, respectively.  $N^{\text{nurse}}$  and  $N^{\text{comp.}}$  indicates the number of nurses and individuals in the comparison group, respectively. The number of chosen controls is depicted by *p*.

Abbreviation: OLS, ordinary least squares.

The last panel of Table 3 contains the results of the impact of the absolute wage belief on the probability to become a nurse. After conditioning on relevant controls, I find a statistically significant effect of the absolute wage on the probability to become a nurse that stays significant when the composition of the comparison group is changed. Interestingly, in all columns, the coefficient in the model with no controls is smaller than in the models including controls. Whereas those who are prone to become a nurse expect higher relative wages, they have a lower absolute wage belief. As observed for the relative wage measures, the effect becomes larger when the comparison group is restricted to individuals who chose a similar occupation.

In summary, results in Table 3 show that even after conditioning on an extensive set of relevant controls and accounting for non-linearities in confounding, the beliefs about a nurse's wage affects the probability to become a nurse. This holds true for both the relative and absolute wage. Moreover, I find evidence that effects are heterogeneous. These effects are stronger for individuals who are more prone to choose a social occupation.

### 3.1.3 | Sensitivity regarding omitted variable bias

It is natural to ask whether the positive association between the choice to become a nurse and the wage belief can be interpreted as causal. Despite the large set of potential controls, the presence of omitted variable bias may be likely due to the complexity of occupational choice and the formation of wage beliefs. To analyze the impact of potential unobserved confounding, I conduct a sensitivity analysis as described in Section 2. The results are shown in Table 4. Each panel depicts the results of one wage measure. As discussed above, I follow Cinelli and Hazlett (2020) and make use of

TABLE 4 Sensitivity regarding omitted variables

	(1) Including transformations	(2) Gender: Female		(3) Parents nurses		(4) Social interests		(5) Metacognition		(6) All	
		$\hat{\beta}$	$t_{adj}$	$\hat{\beta}$	$t_{adj}$	$\hat{\beta}$	$t_{adj}$	$\hat{\beta}$	$t_{adj}$	$\hat{\beta}$	$t_{adj}$
		<b>Rank of nurse's wage</b>									
Nurse versus all	No	0.009	4.43	0.009	4.39	0.009	4.38	0.010	4.45	0.009	4.28
	Yes	0.008	3.92	0.008	3.87	0.008	3.97	0.009	4.40	0.005	2.38
Nurse versus vocational training	No	0.012	3.75	0.012	3.78	0.012	3.75	0.012	3.79	0.011	3.62
	Yes	0.010	3.19	0.010	3.19	0.010	3.26	0.011	3.57	0.004	1.20
Nurse versus social field	No	0.035	3.94	0.035	3.97	0.035	3.97	0.035	3.97	0.034	3.85
	Yes	0.035	3.94	0.031	3.59	0.035	3.97	0.035	3.96	0.030	3.43
<b>Nurse's wage/highest wage</b>											
Nurse versus all	No	0.006	2.77	0.006	2.79	0.006	2.74	0.006	2.79	0.006	2.62
	Yes	0.005	2.44	0.005	2.30	0.005	2.29	0.006	2.79	0.002	1.12
Nurse versus vocational training	No	0.006	1.85	0.006	1.90	0.006	1.87	0.006	1.90	0.005	1.70
	Yes	0.005	1.59	0.005	1.51	0.004	1.40	0.006	1.88	0.001	0.30
Nurse versus social field	No	0.023	2.56	0.023	2.57	0.022	2.44	0.023	2.58	0.021	2.40
	Yes	0.023	2.56	0.020	2.27	0.022	2.44	0.023	2.58	0.018	2.06
<b>Nurse's wage/lowest wage</b>											
Nurse versus all	No	0.008	3.57	0.007	3.48	0.008	3.57	0.008	3.58	0.007	3.43
	Yes	0.007	3.50	0.007	3.37	0.007	3.15	0.007	3.54	0.005	2.62
Nurse versus vocational training	No	0.010	3.26	0.010	3.18	0.010	3.27	0.010	3.27	0.010	3.10
	Yes	0.010	3.22	0.010	3.04	0.009	2.84	0.010	3.23	0.008	2.46
Nurse versus social field	No	0.021	2.90	0.022	3.01	0.020	2.77	0.022	2.99	0.019	2.61
	Yes	0.021	2.90	0.021	2.95	0.020	2.77	0.022	2.99	0.017	2.38
<b>Nurse's wage</b>											
Nurse versus all	No	0.008	3.52	0.007	3.49	0.008	3.51	0.008	3.52	0.007	3.43
	Yes	0.006	2.67	0.007	3.21	0.007	3.18	0.007	3.52	0.003	1.20
Nurse versus vocational training	No	0.011	3.43	0.011	3.46	0.011	3.42	0.011	3.43	0.011	3.26
	Yes	0.008	2.54	0.010	3.18	0.010	3.10	0.011	3.42	0.003	0.81
Nurse versus social field	No	0.028	3.12	0.027	3.07	0.028	3.12	0.028	3.12	0.026	2.96
	Yes	0.025	2.84	0.025	2.85	0.028	3.12	0.028	3.12	0.019	2.14

Note: The table depicts the results on the sensitivity of the effect of the wage beliefs on the decision to become a nurse. The adjusted  $t$ -statistic is based on the adjusted estimate  $\hat{\beta}$  and adjusted standard errors  $\widehat{se}$ .  $R^2_{w \sim z|x}$  and  $R^2_{y \sim z|w,x}$  are computed as defined in Equation (10) setting  $k_w = k_y = 1$ , that is, unobserved confounders that are as strong as the considered benchmark variables.

observed covariates that are strong predictors of the occupational choice or the wage belief to analyze the consequence of potential omitted variables. Columns 2 to 5 display the adjusted estimate  $\hat{\beta}$  and  $t$ -statistic  $t_{adj}$ . They are obtained when an unobserved confounder, that explains as much variance in  $y$  and  $w$  as predefined benchmark variables, is additionally controlled for. As mentioned above, the variables used to bound the consequences of omitted variable bias are (1) gender, (2) parents' occupation, (3) social interests and (4) metacognition. The first column indicates whether  $\hat{\beta}$  and  $t_{adj}$  are computed using only one variable or whether it is based on all transformations in the model involving the variable (i.e., interactions and higher order polynomials). For example, the adjusted estimate and  $t$ -statistic with no transformations are computed under the assumption that an unobserved confounder that is as strong as gender exists. In contrast,  $\hat{\beta}$  and  $t_{adj}$  including transformations are computed by assuming that an unobserved confounder exists, that is as strong as gender and all interactions that are included in the model and where gender is involved in (e.g., interaction between gender and math-skills, gender and social interests, etc.). I expect benchmarks that account for transformations to have a much larger impact than benchmarks of single covariates, because many transformations are chosen by the lasso. The last column

shows the results that would have been obtained if an omitted variable that explains as much as all four variables together had been controlled for.

The first panel depicts the sensitivity of the results regarding the rank of a nurse's wage. The adjusted estimate only decreases slightly and equals 0.009, provided that there exists an unobserved confounder as strong as gender for which is additionally controlled. The change in the adjusted t-statistic is very small such that results stay significant at a 1% significance level. Confounders as strong as parent's occupation, social interests and metacognition only lead to minor changes. Even if I additionally control for a confounder that is as strong as all four benchmark variables combined, the conclusion that the rank of a nurse's wage significantly affects the choice to become a nurse is still valid. As expected, adjusted estimates  $\hat{\beta}$  are drawn to zero by a larger amount when transformations are included. Nonetheless, these changes are small. The effect decreases to 0.004 if I control for a confounder that is as strong as all four benchmark variables together and includes all their transformations. It remains significant at the 5%-level. A change in the comparison group leads to similar robust results. The only noteworthy change in the conclusion occurs when the comparison group consists only of those who chose a vocational training. It is caused by a confounder that is as strong as all four benchmark variables including their transformations in column 6. The adjusted t-statistic shows that if such a confounder exists, there is no statistically significant effect anymore.

The second panel displays the sensitivity of the results on the ratio between the beliefs about a nurse's wage and the highest wage. The results for the entire sample show that only controlling for a confounder that is as strong as all four benchmark variables and their respective transformations has an impact that is large enough to change the conclusion. The effect decreases to 0.002 and is not statistically significant. The sensitivity analysis reveals that the estimated effect is sensitive when the comparison group consists of individuals undergoing a vocational training. A confounder as strong as single variables is not strong enough to change the conclusion. However, a cofounder as strong as gender, parental occupation or interests together with their respective transformations leads to an effect that is not statistically significant different from zero. It is evident that a confounder, as strong as all four benchmark variables combined and including their transformation, leads to an insignificant effect too. The result of comparisons between nurses and individuals in a social field are not sensitive to any of the considered strengths of confounding.

The third panel depicts the sensitivity of the ratio between a nurse's wage and the lowest stated wage. The results show that no confounder as strong as the considered benchmark variables is strong enough to change the conclusion. Even a confounder as strong as all four benchmark variables together including their transformation does not lead to remarkable changes in the estimated effect. Similar sensitivity results can also be observed when the comparison group is changed.

The last panel shows how a confounder changes the estimated effect of the beliefs about a nurse's absolute wage. A confounder as strong as a single variable does not have an impact on the estimated effect. Even a confounder as strong as all four benchmark variables together does not change the estimated effect. However, the impact of a confounder as strong as all four benchmark variables including their transformations is considerable. The effect decreases to 0.003 and is not significantly different from zero. The results are similar when the comparison group is changed. Comparing nurses to individuals who chose a vocational training, the impact of a confounder as strong as all four variables combined including their transformations is strong enough to change the conclusion. The effect substantially decreases to 0.003 and is not significantly different from zero. Choosing individuals in a social field as comparison group, none of the considered confounders is strong enough to change the conclusion.

Taking into account that the gender, the parents and the interests are key drivers of occupational choice, it can be concluded that results of the wage rank, the ratio between a nurse's wage and the lowest wage and the absolute wage are only sensitive to a confounder that is strong. Similarly, the results on the effect of the ratio between the nurse's wage and the highest wage are only sensitive regarding a strong confounder when comparing nurses to all other individuals or to those who chose a social field. However, the results are sensitive when nurses are compared to those who chose a vocational training. If a confounder with a certain strength exists, only subgroups are affected by the ratio.

### 3.2 | How much do other factors matter?

In order to assess the size of the effects of the wage beliefs and to obtain some reassurance about the validity of the data, I compare the effect to other factors discussed in the recent literature. More precisely, I estimate three further PDS models

using the self-assessed importance of economic factors, social interests and self-assessed importance of comfort aspects instead of the wage belief. The results are depicted in Table 5. To compare the size of the effects with the effect of the wage belief in Table 3, measures are standardized to have mean 0 and standard deviation 1.

I examine the impact of the importance of economic factors such as unemployment risk and importance of pay, on young people's involvement in nursing or perhaps even their withdrawal from nursing. The results are depicted in the first panel of Table 5. Independent of the composition of the comparison group, I cannot conclude that the importance of economic factors plays a role in the decision to become a nurse. This result replicates findings of recent research: Nurses do not give more weight to economic factors than others. However, it is noticeable that future nurses also do not weight economic factors lower than other individuals.

The next panel presents the results on the role of social interests, like caring for others. As expected, the results suggest that social interests play an important role in the decision to become a nurse. This holds true when the comparison group only consists of those who chose a social field. Compared to the effect of wage beliefs, the effect of social interests is considerably larger (more than twice as large). The finding perfectly fits into both the nursing and the economic literature. It is often shown that preferences matter the most in the choice of training (e.g., Arcidiacono, 2004; Wiswall & Zafar, 2015). Therefore, the result provides some additional reassurance and further supports the results in Table 3.

Nursing is generally known for its rather exhausting tasks and inflexible working hours. Using the desired work conditions regarding the comfort, I can investigate the effect of this reputation. The results are presented in the last panel of Table 5. They suggest that the larger the importance of comfort aspects, the lower the likelihood of becoming a nurse. Interestingly, compared to the coefficients in an unconditional model, the absolute size of the coefficients is larger in conditional models. That is, individuals that are more prone to become a nurse, put less emphasis on comfort aspects in their occupation. In summary, I find that the size of the effect of the wage belief is smaller than the role of individual interests and has about the same size as the importance of comfort aspects.

TABLE 5 Relevance of other factors

	Nurse versus all		Nurse versus vocational training		Nurse versus social field	
	Single OLS	Post-lasso	Single OLS	Post-lasso	Single OLS	Post-lasso
<b>Importance of economic factors</b>						
	-0.002	0.000	-0.006*	0.000	0.008	0.015
	(0.002)	(0.003)	(0.003)	(0.004)	(0.008)	(0.010)
<i>p</i>	-	52	-	31	-	23
<b>Social interests</b>						
	0.028***	0.028***	0.042***	0.041***	0.042***	0.052***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.010)	(0.011)
<i>p</i>	-	79	-	58	-	25
<b>Importance of comfort aspects</b>						
	-0.007***	-0.009***	-0.010***	-0.012***	-0.026***	-0.036***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.009)	(0.010)
<i>p</i>	-	34	-	23	-	14
$N^{\text{nurse}}$	238		238		238	
$N^{\text{comp.}}$	6851		4214		1378	

Note: The table depicts the results of the effect of other factors than the nurse's wage on the decision to become one. Standard errors are depicted in parentheses. Significance of the coefficient at conventional significance levels 1%, 5%, 10% are indicated by stars \*\*\*, \*\*, \*, respectively.  $N^{\text{nurse}}$  and  $N^{\text{comp.}}$  indicates the number of nurses and individuals in the comparison group, respectively. The number of chosen controls is depicted by *p*.

Abbreviation: OLS, ordinary least squares.



### 3.3 | Assessing wage information

In general, there may be three reasons for finding a positive effect of the wage belief. Results can be driven by future nurses who overestimate wages or by non-future nurses who underestimate wages. Moreover, both can occur simultaneously. To this end, I estimate further PDS models. Instead of the wage beliefs, I use measures that capture information relative to actual wages. As discussed above, changing the rank measure to information measures defined in Equations 14–17 is straightforward. The results of the analysis are given in Table 6.

In the first panel, I consider a measure that captures the general level of information defined in Equation (14). The larger the measure, the higher the deviations from the actual relative wage and consequently the lower the level of information. The coefficient is standardized in order to assess its size. The result shows that an increase in the absolute cumulative deviation by one standard deviation decreases the probability to become a nurse by 0.6 percentage points. The effect becomes even larger when nurses are compared to those who chose a more similar occupation. This means that those who become a nurse can rank surveyed wages more precisely than those who do not become a nurse. Hence, I conclude that future nurses are well informed about relative wages.

In the remaining panels the effect of overestimation in Equation (15), correct estimation in Equation (16) and underestimation in Equation (17) on the probability to become a nurse is analyzed. The results show that overestimation of the nurse's rank increases and underestimation of the nurse's rank decreases the probability to become a nurse. Correct estimation does not affect the probability of becoming a nurse. These results remain statistically significant after controlling for an extensive set of confounders chosen by double-lasso-selection (e.g., general interests).

TABLE 6 Information about nurse's wages

	Nurse versus all		Nurse versus vocational training		Nurse versus social field	
	Single OLS	Post-lasso	Single OLS	Post-lasso	Single OLS	Post-lasso
<b>Cumulative absolute deviation to true ranks</b>						
	−0.004**	−0.006***	−0.012***	−0.008**	−0.017**	−0.023**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.009)
<i>p</i>	-	45	-	35	-	17
<b>Nurse's wage rank overestimated</b>						
	0.042***	0.033***	0.041***	0.038***	0.144***	0.111***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.024)	(0.025)
<i>p</i>	-	40	-	17	-	18
<b>Nurse's wage rank correctly estimated</b>						
	0.007	0.002	0.010	0.002	0.008	0.008
	(0.005)	(0.005)	(0.008)	(0.008)	(0.020)	(0.019)
<i>p</i>	-	15	-	10	-	9
<b>Nurse's wage rank underestimated</b>						
	−0.026***	−0.018***	−0.032***	−0.024***	−0.082***	−0.061***
	(0.004)	(0.004)	(0.008)	(0.007)	(0.018)	(0.017)
<i>p</i>	-	27	-	13	-	9
$N^{\text{nurse}}$	238		238		238	
$N^{\text{comp.}}$	6851		4214		1378	

Note: The table depicts the results of the effect of information about nurse's wage on the decision to become one. The measures are defined in Equations 14–17. The true ranking is (1) hairdresser, (2) motor vehicle mechanic, (3) nurse, (4) bank clerk, (5) teacher, and (6) physician. The cumulative absolute deviation to true ranks is standardized. Standard errors are reported in parentheses. Significance of the coefficient at conventional significance levels 1%, 5%, 10% are indicated by stars \*\*\*, \*\*, \*, respectively.  $N^{\text{nurse}}$  and  $N^{\text{comp.}}$  indicates the number of nurses and individuals in the comparison group, respectively. The number of chosen controls is depicted by *p*.

Abbreviation: OLS, ordinary least squares.

On the one hand, the results indicate that future nurses more often overestimate and less often underestimate the wage of a nurse. On the other hand, future nurses rank wages more in accordance with the true wages. Descriptive results in Table 1 give further insights that are crucial for the interpretation of these seemingly contradicting results. The share of individuals who think that nurses earn the least among all six surveyed wages is much higher among non-future nurses than among future nurses (15% vs. 6%). In general, the share of individuals that underestimate the wage of a nurse is large (63%). Even a significant share of future nurses underestimate the wage (45%). In contrast, the share of those who overestimate a nurse's wage rank is low (13%). Furthermore, the comparison of mean wage ranks between future nurses and others in Table 2 shows that there is only a significant difference in the rank of the three occupations with the lowest wages, that is, hairdresser, mechanic and nurse. There are no significant differences concerning occupations with higher wages, that is, the wage rank of a bank clerk, teacher and physician. Therefore, I conclude that future nurses do not have exceptionally high wage beliefs, but individuals who do not become a nurse have beliefs that are too low. Even future nurses often underestimate the relative wages. In summary, the analysis suggests that the perception of a low wage in nursing among young people may be an obstacle to attract more individuals to nursing.

## 4 | DISCUSSION

This paper investigates the policy relevant question of whether and how wage beliefs and information influences individual career choices to become a nurse. To this end, I used state-of-the-art methods for causal machine learning (PDS, Belloni et al., 2014a) and sensitivity analysis (Cinelli & Hazlett, 2020). My analysis does not use retrospective information that is potentially plagued by reverse causation, but longitudinal data following 14- to 15-year-olds up to their decision whether or not to enter nursing training. I report two sets of substantive findings. First, individuals' beliefs about the wages in nursing do influence the probability of taking up nursing. The size of the effect is smaller than the effect of individual preferences but similar to other factors such as comfort aspects. Although different at first sight, this result does not contradict recent research. Most of the work finding a wage elasticity close to zero, only implies that increasing wages does not increase the intensive labor supply of nurses (i.e., working hours). In this paper, I analyze the choice of becoming a nurse, that is, extensive margin, where wages matter much more. Second, I show that understating the true rank of wages in nursing decreases the likelihood of starting a nursing career. These results suggest two important policy implications. First, boosting wages in nursing may help to overcome the shortage observed in many countries. Second, providing more accurate information about actual (relative) wages in nursing would also help to attract more individuals into this profession. Although informing young individuals about wages seems to be a rather inexpensive policy, policy-makers should consider the efficiency. Higher wage beliefs might entail additional costs by attracting more individuals of the "wrong" type, for example, individuals that are less motivated to help people (Heyes, 2005). The study has some limitations which have to be kept in mind. First, the occupational choice and the formation of wage beliefs are complex processes. Despite the fact that I have access to a rich set of controls and carefully choose them using data-driven variable selection, the assumption of unconfoundedness may be violated. To mitigate this concern, I conduct a sensitivity analysis. Although the existence of unobserved confounders cannot be ruled out, results show that potential unobserved confounders would have to be strong to overturn the conclusions. Second, I have access to wage beliefs on six occupations from a wide range of fields. These occupations are well known and are often chosen by young students. However, one could argue that it might be difficult to draw conclusions on the relative wage beliefs that can be generalized since wage beliefs on only six occupations are available. Yet, the absolute wage beliefs also increase the probability to become a nurse. To address these concerns, further research—preferably (quasi-)experimental studies—on the effect of wage beliefs on the probability to become a nurse is needed.

## ACKNOWLEDGMENTS

I would like to thank Martin Biewen, Martin Kroczeck, Tim Ruberg, Jochen Späth, Anne Zühlke, and two anonymous referees for valuable discussions and comments. Yvette Bodry, Manuel Schick, and Susanne Vogegele are acknowledged for excellent research assistance. This paper is an outcome of the ZAFH care4care funded by the EFRE fund and the Ministry for Science, Research and Art Baden Wuerttemberg (grant number: FEIH ZAFH 795924).

Open access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## DATA AVAILABILITY STATEMENT

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 9, <https://doi.org/10.5157/NEPS:SC4:10.0.0>. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg in cooperation with a nationwide network.

## ORCID

Philipp Kugler  <https://orcid.org/0000-0001-9899-7205>

## ENDNOTES

<sup>1</sup> Note that Cohen's  $f^2$  is defined as  $f^2 = \frac{R^2}{1 - R^2}$ .

<sup>2</sup> Note that results do not differ, when I use information that is measured shortly before occupational choice.

<sup>3</sup> That is, wage belief, economic and social orientation and importance of comfort.

<sup>4</sup> Precisely, I drop observations with over 18% missing values—that is the 90% quantile of the share of missing entries.

<sup>5</sup> Note that comparing future nurses with youths who chose a vocational training, that is, did not choose to visit a university, is motivated by the German education system. Education after school is divided into academic and vocational training, whereas nursing belongs to the latter kind.

<sup>6</sup> A discussion of the magnitude of the estimates will be given in the next section.

<sup>7</sup> Note that regarding the choice of variables, mainly interactions are chosen. This hints to strong non-linearities in confounding and stresses the importance of flexible choice of controls.

## REFERENCES

- Altonji, J. G., Arcidiacono, P., & Maurel, A. (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 5, pp. 305–396). Elsevier.
- Andreassen, L., Di Tommaso, M. L., & Strøm, S. (2017). Nurses and physicians: A longitudinal analysis of mobility between jobs and labor supply. *Empirical Economics*, 52(4), 1235–1269.
- Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1–2), 343–375.
- Beffy, M., Fougère, D., & Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *The Review of Economics and Statistics*, 94(1), 334–347.
- Begerow, A., Michaelis, U., & Gaidys, U. (2020). *Wahrnehmungen von pflegenden im bereich der intensivpflege während der covid-19-pandemie*. Pflege.
- Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6), 2369–2429.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), 29–50.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economics and Statistics*, 81(2), 608–650.
- Biewen, M., & Tapalaga, M. (2017). Life-cycle educational choices in a system with early tracking and 'second chance' options. *Economics of Education Review*, 56, 80–94.
- Blossfeld, H.-P., & von Maurice, J. (2011). Education as a lifelong process. *Zeitschrift für Erziehungswissenschaft*, 14(S2), 19–34.
- Bomball, J., Schwanke, A., Stöver, M., Schmitt, S., & Görres, S. (2010). *Imagekampagne für Pflegeberufe auf der Grundlage empirisch gesicherter Daten: Einstellungen von Schüler/innen zur möglichen Ergreifung eines Pflegeberufes*. <http://www.ipp.uni-bremen.de/uploads/IPPSchriften/>
- Boudarbat, B. (2008). Field of study choice by community college students in Canada. *Economics of Education Review*, 27(1), 79–93.
- Brunello, G., Lucifora, C., & Winter-Ebmer, R. (2004). The wage expectations of European business and economics students. *Journal of Human Resources*, XXXIX(4), 1116–1142.
- Carneiro, P., Heckman, J. J., & Vytlačil, E. (2011). Estimating marginal returns to education. *The American Economic Review*, 101(6), 2754–2781.
- Cho, S.-H., Jung, S. Y., & Jang, S. (2010). Who enters nursing schools and why do they choose nursing? A comparison with female non-nursing students using longitudinal data. *Nurse Education Today*, 30(2), 180–186.

- Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1), 39–67.
- Dante, A., Rizzi, L., Ianderca, B., & Palese, A. (2013). Why do university students not choose a nursing degree at matriculation? An Italian cross-sectional study. *International Nursing Review*, 60(1), 129–135.
- Doiron, D., Hall, J., Kenny, P., & Street, D. J. (2014). Job preferences of students and new graduates in nursing. *Applied Economics*, 46(9), 924–939.
- Eberth, B., Elliott, R. F., & Skåtun, D. (2016). Pay or conditions? The role of workplace characteristics in nurses' labor supply. *The European Journal of Health Economics*, 17(6), 771–785.
- Fernandes, A., Huber, M., & Vaccaro, G. (2020). Gender differences in wage expectations. <http://arxiv.org/pdf/2003.11496v1>
- Finger, C., Solga, H., Ehlert, M., & Rusconi, A. (2020). Gender differences in the choice of field of study and the relevance of income information. Insights from a field experiment. *Research in Social Stratification and Mobility*, 65, 100457.
- German Employment Agency. (2018). *Entgeltatlas*. <https://entgeltatlas.arbeitsagentur.de/entgeltatlas>
- German Employment Agency. (2020a). *Arbeitsmarkt im Pflegebereich, Berichte: Blickpunkt Arbeitsmarkt 05/2020*.
- German Employment Agency. (2020b). *Fachkräfteengpassanalyse, Berichte: Blickpunkt Arbeitsmarkt 10/2020*.
- Hanel, B., Kalb, G., & Scott, A. (2014). Nurses' labour supply elasticities: The importance of accounting for extensive margins. *Journal of Health Economics*, 33, 94–112.
- Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The elements of statistical learning: Data mining, inference, and prediction, Springer series in statistics* (2nd ed.). Springer-Verlag New York.
- Hastings, J. S., Neilson, C. A., Ramirez, A., & Zimmerman, S. D. (2016). (Un) informed college and major choice: Evidence from linked survey and administrative data. *Economics of Education Review*, 51, 136–151.
- Heyes, A. (2005). The economics of vocation or 'why is a badly paid nurse a good nurse'? *Journal of Health Economics*, 24(3), 561–569.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics*, 125(2), 515–548.
- Kaltwasser, A., Pelz, S., Nydahl, P., Dubb, R., & Borzikowsky, C. (2021). *Querschnittsstudie zu arbeitsbedingungen und versorgungsqualität in der versorgung von covid-19-patienten* (pp. 1–8). Der Anaesthesist.
- Knoll, B., Riedel, N., & Schlenker, E. (2017). He's a chip off the old block—The persistence of occupational choices across generations. *Labour*, 31(2), 174–203.
- Krebs, S., Lietz, A. L., & Hasseler, M. (2021). Necessary changes in healthcare of German-speaking countries. In E. Al-Aali & M. Masmoudi (Eds.), *Global perspectives on change management and leadership in the post-COVID-19 era* (pp. 1–23). IGI Global.
- Matthes, S. (2019). *Warum werden Berufe nicht gewählt? Die Relevanz von Attraktions- und Aversionsfaktoren in der Berufsfindung, Berichte zur beruflichen Bildung, 1. auflage edn*. Bundesinstitut für Berufsbildung.
- McCabe, R., Nowak, M., & Mullen, S. (2005). Nursing careers: What motivated nurses to choose their profession? *Australian Bulletin of Labour*, 31(4), 380–402.
- Montmarquette, C., Cannings, K., & Mahseredjian, S. (2002). How do young people choose college majors? *Economics of Education Review*, 21(6), 543–556.
- MOW International Research Team. (1987). *The meaning of working*. Academic Press.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- Oreopoulos, P., & Dunn, R. (2013). Information and college access: Evidence from a randomized field experiment. *The Scandinavian Journal of Economics*, 115(1), 3–26.
- Petrucci, C., La Cerra, C., Aloisio, F., Montanari, P., & Lancia, L. (2016). Empathy in health professional students: A comparative cross-sectional study. *Nurse Education Today*, 41, 1–5.
- Schweri, J., & Hartog, J. (2017). Do wage expectations predict college enrollment? Evidence from healthcare. *Journal of Economic Behavior and Organization*, 141, 135–150.
- Scott, A., Witt, J., Duffield, C., & Kalb, G. (2015). What do nurses and midwives value about their jobs? Results from a discrete choice experiment. *Journal of Health Services Research & Policy*, 20(1), 31–38.
- Shields, M. A. (2004). Addressing nurse shortages: What can policy makers learn from the econometric evidence on nurse labour supply? *The Economic Journal*, 114(499), F464–F498.
- Speer, J. D. (2019). Where the girls are: Examining and explaining the gender gap in the nursing major. *Scottish Journal of Political Economy*, 121(4), 1.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
- Urmitsky, O., Hansen, C., & Chernozhukov, V. (2016). Using double-lasso regression for principled variable selection. *SSRN Electronic Journal*, Article 2733374.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67.
- Wilkes, L., Cowin, L., & Johnson, M. (2015). The reasons students choose to undertake a nursing degree. *Collegian*, 22(3), 259–265.
- Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Economics of Education Review*, 82(2), 791–824.
- Wohlkinger, F., Ditton, H., von Maurice, J., Haugwitz, M., & Blossfeld, H.-P. (2011). Motivational concepts and personality aspects across the life course. *Zeitschrift für Erziehungswissenschaft*, 14(2), 155.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Kugler, P. (2022). The role of wage beliefs in the decision to become a nurse. *Health Economics*, 31(1), 94–111. <https://doi.org/10.1002/hec.4442>