



An educational intervention to improve knowledge about prevention against occupational asthma and allergies using targeted maximum likelihood estimation

Daloha Rodríguez-Molina^{1,2} · Swaantje Barth¹ · Ronald Herrera¹ · Constanze Rossmann³ · Katja Radon¹ · Veronika Karnowski⁴

Received: 15 March 2018 / Accepted: 13 December 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Purpose Occupational asthma and allergies are potentially preventable diseases affecting 5–15% of the working population. However, the use of preventive measures is often insufficient. The aim of this study was to estimate the average treatment effect of an educational intervention designed to improve the knowledge of preventive measures against asthma and allergies in farm apprentices from Bavaria (Southern Germany).

Methods Farm apprentices at Bavarian farm schools were asked to complete a questionnaire evaluating their knowledge about preventive measures against occupational asthma and allergies (use of personal protective equipment, personal and workplace hygiene measures). Eligible apprentices were randomized by school site to either a control or an intervention group. The intervention consisted of a short educational video about use of preventive measures. Six months after the intervention, subjects were asked to complete a post-intervention questionnaire. Of the 116 apprentices (70 intervention group, 46 control group) who answered the baseline questionnaire, only 47 subjects (41%; 17 intervention group, 30 control group) also completed the follow-up questionnaire. We, therefore, estimated the causal effect of the intervention using targeted maximum likelihood estimation. Models were controlled for potential confounders.

Results Based on the targeted maximum likelihood estimation, the intervention would have increased the proportion of correct answers on all six preventive measures by 18.4% (95% confidence interval 7.3–29.6%) had all participants received the intervention vs. had they all been in the control group.

Conclusions These findings indicate the improvement of knowledge by the educational intervention.

Keywords Occupational asthma and allergies · Educational intervention · Targeted maximum likelihood estimation · Preventive measures · Causal effect

Introduction

Occupational asthma and allergies are potentially preventable diseases (Tarlo and Lemiére 2014). However, 5–15% of adult asthma cases are due to occupation, and estimates for occupational allergies are similar (Peden and Reed 2010). Occupational asthma is defined as a disease attributable to

the occupational environment and not to stimuli outside the workplace, and characterized by variable airflow limitation as well as hyperresponsiveness under inflammatory conditions (Baur et al. 2012). Similarly, occupational allergies—such as occupational rhinoconjunctivitis—are characterized by variable nasal airflow limitation and hypersecretion due to conditions of the occupational environment (Ameille et al. 2013). Respiratory causes, including asthma and allergies, are the most common occupational diseases in industrialized countries (Peden and Reed 2010), and generate great costs: current estimates from the UK on direct and indirect lifetime costs per average case of occupational asthma range from £120 k to £130 k (Ayres et al. 2011).

The development of work-related allergies or asthma not only has consequences in terms of health status, but can also

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00420-018-1397-1>) contains supplementary material, which is available to authorized users.

✉ Daloha Rodríguez-Molina
daloha.rodriguez_molina@med.uni-muenchen.de

Extended author information available on the last page of the article

make re-training necessary. Unemployment is very common due to the lack of alternatives for a job in the same area without known asthma and allergy risks (Mahmud et al. 2010; Vandenplas et al. 2011). In Bavaria (Southern Germany), at least 21,000 apprentices are trained every year in occupations with an increased risk of allergies, i.e., farmers, hairdressers, bakers and metalworkers (Patuzzi 2012), and for many of these trainees, presenting allergic symptoms means a premature end of training. About 9% of young Bavarians drop out of training (Patuzzi 2012), and approximately one-third of the young people who drop out due to health reasons suffer from skin and respiratory diseases (Muth et al. 2005). To prevent these diseases, technical and organizational measures at the workplace as well as personal protective measures are recommended (Tarlo and Liss 2005; Moscato et al. 2011). Therefore, training about occupational asthma and allergies as well as preventive measures should be included in the curricula in general education and especially vocational schools educating apprentices in high-risk occupations such as farming. Therefore, the Social Security of Agriculture, Forestry and Horticulture in Bavaria (SVLFG) provides vocational schools with the required study curriculum for occupational safety and health. Knowledge and acceptance of preventive measures—especially during training age—are insufficient (Ling and Coulson 2002; Nixon et al. 2006; Crippa et al. 2007; Kim et al. 2012; Levesque et al. 2012; Bonow et al. 2013; Pounds et al. 2014; Cezar-Vaz et al. 2015). This is probably related to a lack of interest and a belief that preventive measures do not affect them personally. Only about a third of employees adhere to the recommended preventive measures (Kütting et al. 2009). Despite the lack of specific data in the farming context, we assume that the use of personal protective equipment (PPE) and other preventive measures between apprentices and workers is similar, and that these figures are also similar between farming and other professions.

Therefore, the aim of this study was to apply an educational intervention designed to improve the knowledge of preventive measures against asthma and allergies in farm apprentices from Bavaria (Southern Germany), and to estimate the average treatment effect of this intervention using double-robust and semi-parametric methods.

Materials and methods

Design and participants

There was an approximate total number of 2167 farming apprentices belonging to 52 vocational schools in the Federal State of Bavaria for the school year 2013–2014 (Bayerisches Landesamt für Statistik und Datenverarbeitung 2014). There are eight main farming schools in Upper Bavaria

distributed in six administrative regions. All six administrative regions granted us permission to invite their students to participate. We only included schools that were willing to participate. Farm apprentices on their second and third year of training in Upper Bavaria were personally invited to participate from March to April 2014 in this 6-month, two-armed, randomized controlled interventional study. Before the beginning of the study, appointments were set for visiting the students at their place of study as part of their training. The recruitment of participants for the control and intervention groups was performed during these visits. Apprentices who had already turned 18 years of age were eligible, which resulted in exclusion of 37 apprentices who were younger at the time of recruitment. In total, we subdivided these six regions into ten main study centers and performed cluster randomization considering the number of participants and the school they belonged to, to rule out systematic error (Fig. 1; Online Resource 1).

During the visits, apprentices were informed about the study and completed the baseline questionnaire about their knowledge of preventive measures against asthma and allergies. For the intervention groups, and following an educational intervention known as “Learning with errors” (Seifried and Baumgartner 2009; Hainich 2010), the completion of the questionnaire was followed by the first version of an educational video in which a farming student did not comply with the recommended measures to protect against asthma and allergies (see Online Resource 3). After that, they were invited to identify errors in a group discussion. In the end, a second version of the video was shown in which the farming student behaved correctly (see Online Resource 4). The time for the training unit was about 20 min. The control group did not receive any training unit. Six months after recruitment (between September and October 2014), all study participants were contacted by phone for the follow-up interview during which their knowledge about preventive measure was again assessed.

All students were asked to fill out a questionnaire and a consent form. Only participants who signed the informed consent with their contact data and successfully completed the baseline questionnaire were eligible for follow-up. Of 277 eligible students at baseline, 40 (14.4%) refused to participate. Of the remaining 237, 120 (50.6%) chose not to provide their email address ($n=29$), their phone number ($n=72$) or neither ($n=19$) due to personal preferences. One additional subject (0.4%) was removed due to missing information on sociodemographic variables. Of the remaining 116 apprentices, 46 belonged to the control group and 70 to the intervention group ($n=70$). Of them, 47 (40.5%; 17 intervention group, 30 control group) completed the follow-up questionnaire (Fig. 1). The study was approved by the Ethics Committee of the University of Munich (LMU) in February 2014.

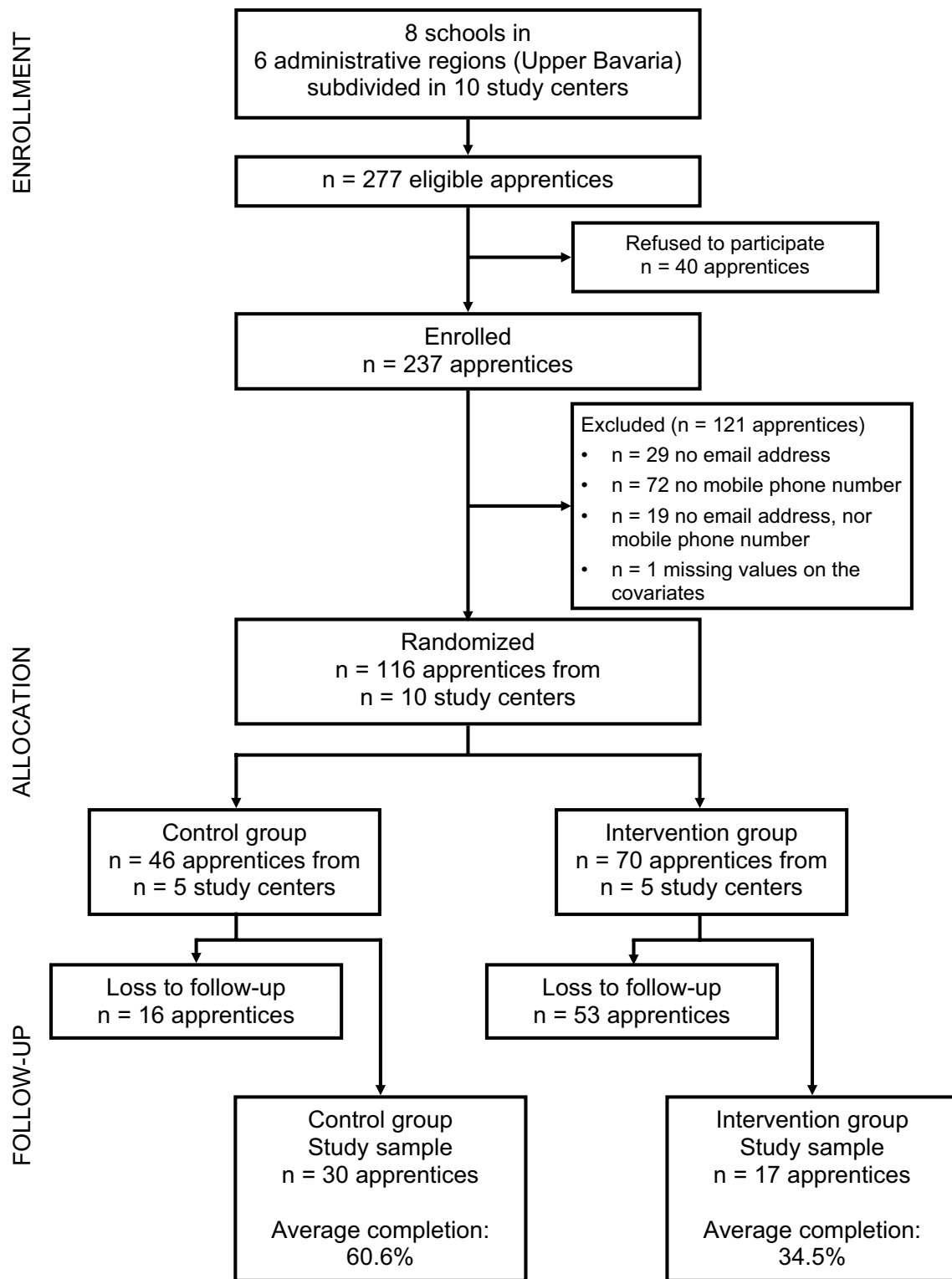


Fig. 1 Flow diagram describing inclusion of study participants

Questionnaire

The baseline questionnaire included sociodemographic factors such as year of birth, schooling years, smoking behavior,

occupational and private contact with animals and plants, as well as personal and parental history of asthma. Additionally, the questionnaire included a section about knowledge regarding prevention of asthma and rhinoconjunctivitis.

The post-intervention survey was similar to the baseline questionnaire.

Variable definition

Sociodemographic variables and medical history

Age was collected as a continuous variable and dichotomized as either younger than 25 years old, or older. According to the European Community Respiratory Health Survey (Burney et al. 1994), smoking was defined as either ever smoked at least 20 packages of cigarettes or 360 g of tobacco, or having smoked at least one cigarette per day or one cigar per week for at least 1 year, or having smoked in the last month.

Risk perception was explored using a 1–5 Likert scale to ask how likely would it be for the subject to present an episode of nasal allergies or asthma in the next 5 years, and how bad would it be for the subject to present an episode of nasal allergies or asthma. Answers such as “very unlikely” and “not bad at all”, respectively, were given the highest score. The scores of these two items were summed up, and a cut-off point corresponding to the 75th percentile was chosen a priori to dichotomize the sum into apprentices with normal level of risk perception (below 75th percentile) and low level of risk perception (≥ 75 th percentile).

Presence of asthma or rhinoconjunctivitis was defined as either self-reported wheezing without having a cold, currently taking asthma medication or symptoms of rhinoconjunctivitis during the 12 months prior to survey.

Outcomes of interest

The main outcome variable of interest was knowledge about six potentially preventive measures against asthma and allergies, based on the subject’s response to the question ‘which of the following measures can be applied in farming to avoid getting asthma and rhinoconjunctivitis?’: washing hair after

work (yes), wearing work shoes (no), keeping work clothes outside of the living environment (yes), disinfecting stables (no), wearing protective goggles (no), and avoid wearing work clothes inside the living environment (yes). Correct answers are given in parenthesis. Each of these six measures was given one point if correct, and zero if not correct. The scores were summed up. Our outcome of interest was to answer all six measures correctly. Sensitivity analyses were carried out using both at least five correct answers, and at least four correct answers.

Statistical analyses

Since previous analyses already show that logistic regression models are not well suited to model the effect of the intervention on the main outcome (Petersen et al. 2012) (see Table 1), we applied Targeted Maximum Likelihood Estimation methods (TMLE) along with Machine Learning (ML) techniques. Logistic regression models did not converge to provide estimates for interpretation because of the low study sample size and high proportion of missing values on the outcome (drop-outs). Preliminary analyses showed that subjects who completed the intervention were not statistically different on confounders of interest from those who did not. TMLE—also found in the literature as targeted minimum loss-based estimation—is a double-robust method providing an effect estimator based on the exposure, on the outcome, or on both, while also allowing to include missing values as a parameter for modeling (Laan and Rose 2011; Colson et al. 2016). TMLE is performed in two stages. In the first stage, it models the outcome Y as a function of both the exposure A (treatment group) and the covariates ($W_1, W_2, \dots, W_n$). At the same time, it performs a joint modeling analysis of the missing values, and includes it in the model as a parameter. The second stage is a bias reduction step, where TMLE iteratively updates parameter estimates using exposure models given the covariates, i.e., modeling the treatment

Table 1 Crude and adjusted effects of the intervention using logistic regression ($n=47$), Bavaria, Germany, 2014

	All six correct measures	At least five correct measures	At least four correct measures
Crude OR			
Parameter	4.31	3.17	8.00
95% CI	0.7–26.6	0.92–10.96	0.92–69.24
Adjusted OR			
Parameter	Not possible to estimate adjusted OR		
95% CI			

Adjusted for sex, age, education level, smoking status, presence of asthma or rhinoconjunctivitis, risk perception, parental asthma, and knowledge about preventive measures against asthma and allergies before the intervention

OR odds ratio, CI confidence interval

mechanism. Within a counterfactual framework (Rubin 1974, 1978), the effect of the exposure on the probability distribution P_0 of the outcome of interest is, ultimately, a difference in effect between the exposed and the unexposed (the target parameter), which could be measured on an additive or a multiplicative scale, e.g., relative risk (RR) or odds ratio (OR) (Laan and Rose 2011). As a consequence, the additive average treatment effect (ATE) as the true target parameter can be defined as

$$\text{ATE} = E_W[E(Y | A = 1, W) - E(Y | A = 0, W)],$$

where $E(Y|A=a, W)$ is the conditional mean of the outcome given the exposure, either in the intervention ($A=1$) or the control group ($A=0$), and the covariates (W) (Gruber and Laan 2010; Laan and Rose 2011). This second step makes use of semi-parametric techniques through ML that, unlike traditional parametric methods such as linear or logistic regression, are able to reduce bias and increase efficiency (Laan et al. 2007; Lee et al. 2010; Pirracchio et al. 2015; Colson et al. 2016). One of the main advantages about using ML techniques is that we allow the data to guide the controlled estimation of the best possible set of terms or interactions to include in the final model because the function ‘learns’ from the data itself (Samii et al. 2016). A full tutorial including a guided implementation of TMLE can be found in the published literature (Luque-Fernandez et al. 2018).

Further, we estimated the average treatment effect on the treated (ATT) and the average treatment effect on the controls (ATC) to have a better understanding of the effect of the exposure only among those who actually receive the intervention, or those who were randomized into the control group, respectively (Lendle et al. 2013b; Lendle 2015b; Samii et al. 2016).

In other words, the ATT is the average difference on the outcome between the exposed subjects had they been exposed, and the same exposed subjects had they been unexposed (Colson et al. 2016):

$$\text{ATT} = E[E(Y|A = 1, W) - E(Y|A = 0, W) | A = 1].$$

In contrast, the ATC is the average difference on the outcome between the control subjects had they been exposed, and the same control subjects had they been unexposed (Samii et al. 2016):

$$\text{ATC} = E[E(Y|A = 1, W) - E(Y|A = 0, W) | A = 0].$$

These results were adjusted for the potential confounders sex, age, education level, smoking status, presence of asthma or rhinoconjunctivitis, risk perception, parental asthma, and knowledge about preventive measures against asthma and allergies before the intervention. In case of statistical testing, an alpha level of ≤ 0.05 was considered significant. All statistical analyses were performed

using R Statistical Software version 3.3.0 (R Core Team 2016). TMLE implementation was performed using the tmle package version 1.2.0–4 (Gruber and Laan 2012). ML implementation within TMLE was performed using the SuperLearner package version 2.0–19 (Polley et al. 2016). Estimation of ATT and ATC was done using the tmle package version 0.3.1 (Lendle 2015a).

Results

The majority of the population ($n=116$) consisted of males (88.8%), 24 years old or younger (94.8%), non-smokers (62.9%), with more than 10 years of schooling (60.3%), with no personal history of asthma or rhinoconjunctivitis (62.9%) or parental history of asthma (75.9%) (Table 2). At baseline, only 23.3% correctly answered all six questions about preventive measures against asthma and allergies, and this proportion increases to 27.7% at follow-up (Table 2). On subjects who completed the intervention phase ($n=47$), the proportion of correct answers in the intervention group increased from 17.7 to 23.5% for the main outcome, and from 52.9 to 64.7% for having at least five correct answers (Table 3). This proportion remained unchanged for having at least four correct answers (Table 3). There were no statistically significant differences before the intervention (Table 3).

The adjusted additive ATE of the educational intervention was 18.4% (95% confidence interval (CI) 7.3–29.6%) for the main analyses (Table 4). In other words, the intervention would have increased the proportion of correct answers on all six preventive measures by 18.4%, had all participants received the intervention vs. had they all been in the control group. The adjusted additive treatment effect on the treated (ATT) was 16.9%, (95% CI 5.4–28.5%), while on the controls (ATC) it was 16.8% (95% CI 5.0–28.6%).

In the sensitivity analyses, the additive ATE was the highest when the cut-off point for knowledge was at least five correct measures (55.5%, 95% CI 37.0–74.1%). The additive ATT was 63.08% (95% CI 46.02–80.13%), while the additive ATC was 32.28% (95% CI 12.84–51.72%). Furthermore, for knowledge of about four correct measures, ATE dropped to 29.60% (95% CI 12.2–47.0%), while ATT was 62.78% (95% CI 41.64–83.93%), and ATC was 18.97% (95% CI 1.91–36.02%).

Discussion

We found that using an instructional video as educational intervention is an effective approach to improve knowledge about preventive measures against occupational asthma and allergies in Bavarian farm apprentices. Our data show that

Table 2 Descriptive data of subjects, and level of knowledge after the intervention, Bavaria, Germany, 2014

	Overall (<i>n</i> = 116) (<i>n</i> , %)		Follow-up (<i>n</i> = 47) (<i>n</i> , %)		Missing values ^a (<i>n</i> , %)		Level of knowledge after the intervention (<i>n</i> = 47)							
							All six measures ^b (<i>n</i> , %)		At least five measures ^c (<i>n</i> , %)		At least four measures (<i>n</i> , %)		Less than four measures (<i>n</i> , %)	
Age														
Younger (18–24 years old)	110	94.83	42	89.36	68	98.55	5	83.33	21	95.45	33	91.67	9	81.82
Older (25–44 years old)	6	5.17	5	10.64	1	1.45	1	16.67	1	4.55	3	8.33	2	18.18
Sex														
Female	13	11.21	8	17.02	5	7.25	4	66.67	7	31.82	7	19.44	1	9.09
Male	103	88.79	39	82.98	64	92.75	2	33.33	15	68.18	29	80.56	10	90.91
Smoking status														
Non-smoker	73	62.93	26	55.32	47	68.12	6	100.0	13	59.09	22	61.11	4	36.36
Smoker	43	37.07	21	44.68	22	31.88	0	0.0	9	40.91	14	38.89	7	63.64
Education level														
9 years of schooling	46	39.66	17	36.17	29	42.03	1	16.67	5	22.73	11	30.56	6	54.55
10+ years of schooling	70	60.34	30	63.83	40	57.97	5	83.33	17	77.27	25	69.44	5	45.45
Risk perception ^d														
High level	86	74.14	36	76.60	50	72.46	4	66.67	16	72.73	26	72.22	10	90.91
Low level	30	25.86	11	23.40	19	27.54	2	33.33	6	27.27	10	27.78	1	9.09
Asthma or rhinoconjunctivitis ^e														
No	73	62.93	30	63.83	43	62.32	3	50.00	13	59.09	22	61.11	8	72.73
Yes	43	37.07	17	36.17	26	37.68	3	50.00	9	40.91	14	38.89	3	27.27
Parental asthma														
No	88	75.86	30	63.83	58	84.06	3	50.00	16	72.73	25	69.44	5	45.45
Yes	28	24.14	17	36.17	11	15.94	3	50.00	6	27.27	11	30.56	6	54.55
Knowledge before the intervention ^f														
Less than cut-off point	89 ^g	76.72	34	72.34	55	79.71	4	66.67	7	31.82	3	8.33	3	27.27
More than cut-off point	27	23.28	13	27.66	14	20.29	2	33.33	15	68.18	33	91.67	8	72.73
Intervention group														
Control	46	39.66	30	63.83	16	23.19	2	33.33	11	50.00	20	55.56	10	90.91
Intervention	70	60.34	17	36.17	53	76.81	4	66.67	11	50.00	16	44.44	1	9.09

^aMissing values of the outcome^bIncludes 4 and 5 correct^cIncludes 4 correct^dA low level of risk perception puts the subject at a higher risk. It results from answering the questions “how likely would you think it is to present an episode of nasal allergies or asthma in the next 5 years” and “how bad do you think it would be to present an episode of nasal allergies or asthma?” For details on how this variable was constructed, see text^eDefined as the 12-month prevalence of either self-reported wheezing without a cold, currently taking medication for asthma or rhinoconjunctivitis^fEach outcome was compared against the same outcome at baseline: all six correct measures at baseline vs. all six correct measures at follow-up, at least five correct measures at baseline vs. at least five correct measures at follow-up, at least four correct measures at baseline vs. at least four correct measures at follow-up^gFor all six correct measures

the intervention would be able to increase the proportion of correct answers by 18% had all participants taken the intervention vs. had nobody taken it. The magnitude of the effect changed depending on the selected cut-off point for knowledge, but remained to be positive and statistically significant for all selected outcomes.

Our results are consistent with the current literature on the topic. Previous studies on apprentices and practicing farmers (Kim et al. 2012; Levesque et al. 2012; Pounds et al. 2014), hairdressers (Ling and Coulson 2002; Nixon et al. 2006; Crippa et al. 2007) and welders (Bonow et al. 2013; Cezar-Vaz et al. 2015) have reported that knowledge

Table 3 Descriptive data by intervention group ($n=47$), Bavaria, Germany, 2014

	Intervention group				<i>p</i> value ^c
	Control <i>n</i> = 30 (<i>n</i> , %)		Intervention <i>n</i> = 17 (<i>n</i> , %)		
Age					
Older (25–44 years old)	2	6.67	3	17.65	0.34
Sex					
Male	24	80.00	15	88.24	0.69
Smoking status					
Smoker	13	43.33	8	47.06	1.00
Education level					
10+ years of schooling	18	60.00	12	70.59	0.68
Risk perception ^a					
Low level	26	86.67	10	58.82	0.07
Asthma or rhinoconjunctivitis ^b					
Yes	10	33.33	7	41.18	0.82
Parental asthma					
Yes	9	30.00	8	47.06	0.39
Knowledge before the intervention					
All six correct answers	10	33.33	3	17.65	0.32
At least five correct answers	19	63.33	9	52.94	0.70
At least four correct answers	25	83.33	16	94.12	0.40
Knowledge after the intervention					
All six correct answers	2	6.67	4	23.53	0.17
At least five correct answers	11	36.67	11	64.71	0.12
At least four correct answers	20	66.67	16	94.12	0.04

^aA low level of risk perception puts the subject at a higher risk. It results from answering the questions “how likely would you think it is to present an episode of nasal allergies or asthma in the next 5 years” and “how bad do you think it would be to present an episode of nasal allergies or asthma?” For details on how this variable was constructed, see text

^bDefined as the 12-month prevalence of either self-reported wheezing without a cold, currently taking medication for asthma or rhinoconjunctivitis

^cChi-squared or Fisher’s exact test, accordingly

about preventive measures varies from poor to average, the use of PPE is insufficient, and several types of interventions tend to increase the level of knowledge and compliance with preventive measures. Kim et al. (2012) demonstrated in a pilot study among farmers that one educational workshop consisting of rotating stations with information on causes of work-related asthma, spirometry and use of PPE, aiming at increasing the knowledge about safety training and occupational health and safety (OHS) was feasible and effective, while Levesque et al. (2012) showed that farmers who received traditional pesticide safety training were more likely to use PPE; however, their confidence intervals were wide due to low number of participants. Meanwhile, Pounds et al. (2014) implemented a different kind of intervention: a social

Table 4 Adjusted average treatment effects of the intervention ($n = 116$), Bavaria, Germany, 2014

	All six correct measures	At least five correct measures	At least four correct measures
Additive ATE			
Parameter	18.44%	55.53%	29.60%
95% CI	(7.3–29.58%)	(36.96–74.09%)	(12.2–47.0%)
Additive ATT			
Parameter	16.9%	63.07%	62.78%
95% CI	(5.38–28.51%)	(46.02–80.13%)	(41.64–83.93%)
Additive ATC			
Parameter	16.8%	32.28%	18.97%
95% CI	(5.02–28.57%)	(12.84–51.72%)	(1.91–36.02%)

Adjusted for sex, age, education level, smoking status, presence of asthma or rhinoconjunctivitis, risk perception, parental asthma, and knowledge about preventive measures against asthma and allergies before the intervention

The adjusted model using TMLE allowed including both observed data ($n=47$) and missing values ($n=69$) as parameters

ATE average treatment effect, ATT average treatment effect on the treated, CI confidence interval, ATC average treatment effect on the controls, TMLE targeted maximum likelihood estimation

media campaign promoting the use of respiratory protection devices (RPD) among farmers, and results showed that communication via e-mail increased knowledge about RPD, as well as intention to use these devices in dusty conditions.

Our study has several strengths. Our educational intervention is easy to apply and to measure. Moreover, all relevant information about correct and incorrect practices is embedded in the videos, and students have the opportunity to discuss these practices and learn from mistakes made by a third party and not themselves (Seifried and Baumgartner 2009; Hainich 2010). A consequence of this dynamic is that the instructor does not need to possess any relevant information about the topic. We would recommend, however, designing an info sheet for the instructor as extra material on how to lead the session. Furthermore, using a double-robust method such as TMLE, we were able to estimate the parameter of interest in the presence of small sample size and with a high proportion of missing values on the outcome. The combination of TMLE and ML (SuperLearner) as semi-parametric techniques allowed us to have more flexibility and to reduce common modeling errors due to incorrect parametric assumptions (Schnitzer et al. 2014), therefore, reducing bias and increasing efficiency (Lendle et al. 2013a; Ahern et al. 2016; Colson et al. 2016). Simulation studies have shown that TMLE has many advantages over other methods such as the G-computation formula, inverse probability of treatment weighting (IPTW) or propensity score matching (PS) (Laan and Gruber 2012; Schnitzer et al. 2014), namely that these other methods are inconsistent if the initial regression

estimator is inconsistent (while TMLE remains consistent), and that TMLE still performs relatively well even if there are violations to the experimental treatment assignment (ETA) assumption. A practical ETA violation occurs when the conditional probability of receiving treatment is below some small ϵ , typically ranging between 0.1 and 0.01, depending on the number of observations (Lendle et al. 2013a). Unlike traditional regression methods (e.g., logistic regression), TMLE is able to provide a measure of effect on the additive scale, which makes interpretation easier. Finally, our study is unique and provides data and analyses on a population of young Bavarian farmers, thus filling a prior gap of knowledge on this specific subpopulation.

Nevertheless, our study also has limitations. We analyzed only knowledge as main outcome, but it would be also interesting to measure and analyze behavioral change before and after the intervention. The cut-off point for the main outcome (all six correct measures) was chosen arbitrarily and might have been too strict. Nevertheless, we decided to use this cut-off point to be conservative and we performed sensitivity analyses considering less strict cut-off points, which confirmed the same results. Further, we modeled our outcome as binary instead of ordinal. Several methods have been proposed for randomized clinical trials with ordinal outcomes, which could increase the relative efficiency of the estimator (Díaz et al. 2016). Hence, the results of our study could be further improved by applying these methods. The main challenge was the large number of losses to follow-up that we faced, which did not permit to use, e.g., logistic regression models (see Table 1). One of the challenges that could explain this high attrition rate was an element of peer pressure during recruitment: apprentices were more likely to participate and remain in the study if they had friends interested in enrolling in the study, and were less likely to participate and comply if the “natural leaders” of the class were not interested in participating. Because there were no statistically significant differences in terms of covariates in between the drop-outs and the remaining study population, we have overcome the limitation of loss-to-follow-up during the analytical phase using TMLE to model these missing values simultaneously with estimators for both the exposure and the outcome. Thus, we were able to obtain interpretable results that allow us to answer our research question.

A popular criticism of the counterfactual model is the possibility of individual treatment effects, which may introduce treatment-effect heterogeneity bias (Angrist 2003; Xie et al. 2012; Imai and Ratkovic 2013; Breen et al. 2015; Grimmer et al. 2017). This type of bias is less common in randomized controlled trials because several assumptions are met: no interference among subjects, one unique version of the treatment, a non-zero probability of treatment assignment for each subject, and independence between treatment level and potential outcomes given observed covariates (Rosenbaum and Rubin 1983; Imai and Ratkovic 2013).

When treatment effects are homogeneous for all subjects, the values for the ATE, ATT and ATC are identical (Xie et al. 2012; Breen et al. 2015). Our results show that the ATE, ATT and ATC on the main outcome are fairly similar. Hence, the presence of treatment-effect heterogeneity bias is unlikely. Additionally, all causal frameworks might still be susceptible to unmeasured confounding, especially in observational studies (Fewell et al. 2007; Vanderweele and Arah 2011). In our study, however, we have made sure to measure and include all possible covariates that might be related to both our exposure and outcome. This approach, along with our study design and randomization strategy, provides us with an adequate amount of confidence to make statements about the results of our data.

It is also important to consider a possible community effect that may ‘contaminate’ the effect of the intervention between those treated and those untreated (Fischer and White 2012). After all, young apprentices take courses and perform daily activities together. However, we reduced the possibility for interaction between treated and untreated subjects when we performed clustered randomization. In other words, all participants of one study center were randomized either to receive the intervention or to be in the control group, thus decreasing the possibility of disturbing the effect on the outcome. Moreover, apprentices in this population tend to meet only occasionally if they have lectures, while the rest of the time they are working on the farms, so they do not see each other as frequently as trainees in other areas. This study was performed in a very well characterized population of young Bavarian farm apprentices. Extrapolation of results to other populations must be done cautiously since risk perception, work practices, and other variables may vary according to strong local cultural patterns (Peres et al. 2013).

Even though we have obtained positive and significant results with our intervention, we believe we are still short in closing the breach of knowledge about preventive measures against asthma and allergies in the workplace. A valid point for discussion is whether or not an 18% increase in the proportion of correct answers is sufficient. As far as we know, there is no proposed cut-off point for what constitutes adequate knowledge regarding preventive measures in the workplace but considering that 23% of our population had a full score on all six preventive measures at baseline (Table 2), and that our intervention would increase this proportion by 18%, we are still left with less than 50% of apprentices correctly answering all six preventive measures, which is less than desirable. Further, it is still unclear if an increase in knowledge would effectively relate to an increase in the correct use of preventive measures, although the knowledge–attitude–behavior continuum model (Bettinghaus 1986) suggests it would. However, our current work is unable to answer such a question.

In summary, our findings suggest that our educational intervention improves the knowledge about preventive

measures against asthma and allergies in about 20% of young Bavarian farm apprentices, and that TMLE is an efficient double-robust and semi-parametric method able to provide causal effect estimates where traditional regression methods cannot.

Acknowledgements Special thanks to Jenny Schlichtiger for her valuable collaboration and hard work in the data collection process, and to Iven-Alex Heim for acting in the educational videos. The authors would also like to thank all farm apprentices who kindly participated in our study. Furthermore, we would like to thank Thomas Brendel and Thomas Bischoff from the Institut für Didaktik und Ausbildungsforschung in der Medizin (LMU) for their help in creating the educational video.

Author contributions Study conception and design: KR, CR, SB and VK; fieldwork and data collection: SB; data analysis: DRM and RH; interpretation of the data: DRM, KR and RH; drafting of the manuscript: DRM and KR. All authors read and approved the final manuscript.

Funding This study was funded by Gesund.Leben.Bayern. <http://www.gesundheit.bayern.de>.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval This study was approved by the Ethics Committee of the University of Munich (LMU) (Project no. 6–14). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

- Ahern J, Karasek D, Luedtke AR et al (2016) Racial/ethnic differences in the role of childhood adversities for mental disorders among a nationally representative sample of adolescents. *Epidemiology* 27:697–704. <https://doi.org/10.1097/EDE.0000000000000507>
- Ameille J, Hamelin K, Andujar P et al (2013) Occupational asthma and occupational rhinitis: the United Airways disease model revisited. *Occup Environ Med* 70:471–475. <https://doi.org/10.1136/oemed-2012-101048>
- Angrist JD (2003) Treatment effect heterogeneity in theory and practice. Institute for the Study of Labor (IZA), Bonn
- Ayres JG, Boyd R, Cowie H, Hurley JF (2011) Costs of occupational asthma in the UK. *Thorax* 66:128–133. <https://doi.org/10.1136/thx.2010.136762>
- Baur X, Sigsgaard T, Aasen TB et al (2012) Guidelines for the management of work-related asthma. *Eur Respir J* 39:529–545. <https://doi.org/10.1183/09031936.00096111>
- Bayerisches Landesamt für Statistik und Datenverarbeitung (2014) Berufliche Schulen in Bayern. In: Schuljahr 2013/14. Bayerisches Landesamt für Statistik und Datenverarbeitung, München
- Bettinghaus EP (1986) Health promotion and the knowledge-attitude-behavior continuum. *Prev Med* 15:475–491. [https://doi.org/10.1016/0091-7435\(86\)90025-3](https://doi.org/10.1016/0091-7435(86)90025-3)
- Bonow CA, Cezar-Vaz MR, Almeida MCV de et al (2013) Risk perception and risk communication for training women apprentice welders: A challenge for public health nursing. *Nurs Res Pract* 2013:386260. <https://doi.org/10.1155/2013/386260>
- Breen R, Choi S, Holm A (2015) Heterogeneous causal effects and sample selection bias. *Sociol Sci* 2:351–369. <https://doi.org/10.15195/v2.a17>
- Burney PG, Luczynska C, Chinn S, Jarvis D (1994) The European Community Respiratory Health Survey. *Eur Respir J* 7:954–960
- Cezar-Vaz MR, Bonow CA, Vaz JC (2015) Risk communication concerning welding fumes for the primary preventive care of welding apprentices in southern Brazil. *Int J Environ Res Public Health* 12:986–1002. <https://doi.org/10.3390/ijerph120100986>
- Colson KE, Rudolph KE, Zimmerman SC et al (2016) Optimizing matching and analysis combinations for estimating causal effects. *Sci Rep* 6:23222. <https://doi.org/10.1038/srep23222>
- Crippa M, Torri D, Fogliata L et al (2007) Implementation of a health education programme in a sample of hairdressing trainees. *Med Lav* 98:48–54
- Díaz I, Colantuoni E, Rosenblum M (2016) Enhanced precision in the analysis of randomized trials with ordinal outcomes. *Biometrics* 72:422–431. <https://doi.org/10.1111/biom.12450>
- Fewell Z, Davey Smith G, Sterne JAC (2007) The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol* 166:646–655. <https://doi.org/10.1093/aje/kwm165>
- Fischer K, White IR (2012) Causal inference in clinical trials. In: Berzuini C, Dawid P, Bernardinelli L (eds) *Causality: statistical perspectives and applications*. Wiley, Ltd, pp 310–326
- Grimmer J, Messing S, Westwood SJ (2017) Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Polit Anal* 25:413–434. <https://doi.org/10.1017/pan.2017.15>
- Gruber S, Laan MJ van der (2010) A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. *Int J Biostat*. <https://doi.org/10.2202/1557-4679.1260> (Article 26)
- Gruber S, Laan MJ van der (2012) Tmle: an R package for targeted maximum likelihood estimation. *J Stat Softw* 51:1–35
- Hainich R (2010) Fehlerkultur in der Ausbildung auf der Intensivstation. *Intensiv* 18:62–69. <https://doi.org/10.1055/s-0030-1251485>
- Imai K, Ratkovic M (2013) Estimating treatment effect heterogeneity in randomized program evaluation. *Ann Appl Stat* 7:443–470. <https://doi.org/10.1214/12-AOAS593>
- Kim J, Arrandale VH, Kudla I et al (2012) Educational intervention among farmers in a community health care setting. *Occup Med (Lond)* 62:458–461. <https://doi.org/10.1093/occmed/kqs129>
- Küttling B, Weistenhöfer W, Baumeister T et al (2009) Current acceptance and implementation of preventive strategies for occupational hand eczema in 1355 metalworkers in Germany. *Br J Dermatol* 161:390–396. <https://doi.org/10.1111/j.1365-2133.2009.09085.x>
- Laan MJ van der, Rose S (2011) Targeted learning: causal inference for observational and experimental data. Springer Science & Business Media, Berlin
- Laan MJ van der, Gruber S (2012) Targeted minimum loss based estimation of causal effects of multiple time point interventions. *Int J Biostat* 8(1)9. <https://doi.org/10.1515/1557-4679.1370>
- Laan MJ van der, Polley EC, Hubbard AE (2007) Super learner. *Stat Appl Genet Mol Biol*. <https://doi.org/10.2202/1544-6115.1309> (Article25)
- Lee BK, Lessler J, Stuart EA (2010) Improving propensity score weighting using machine learning. *Stat Med* 29:337–346. <https://doi.org/10.1002/sim.3782>
- Lendle S (2015a) Tmle: estimates the NDE and ATT with TMLE. R package available at <https://github.com/lendle/tmle>
- Lendle SD (2015b) Targeted minimum loss based estimation: applications and extensions in causal inference and big data. PhD thesis, UC Berkeley

- Lendle SD, Fireman B, Laan MJ van der (2013a) Targeted maximum likelihood estimation in safety analysis. *J Clin Epidemiol* 66:S91–S98. <https://doi.org/10.1016/j.jclinepi.2013.02.017>
- Lendle SD, Subbaraman MS, Laan MJ van der (2013b) Identification and efficient estimation of the natural direct effect among the untreated. *Biometrics* 69:310–317. <https://doi.org/10.1111/biom.12022>
- Levesque DL, Arif AA, Shen J (2012) Effectiveness of pesticide safety training and knowledge about pesticide exposure among Hispanic farmworkers. *J Occup Environ Med* 54:1550–1556. <https://doi.org/10.1097/JOM.0b013e3182677d96>
- Ling TC, Coulson IH (2002) What do trainee hairdressers know about hand dermatitis? *Contact Derm* 47:227–231
- Luque-Fernandez MA, Schomaker M, Rachet B, Schnitzer ME (2018) Targeted maximum likelihood estimation for a binary treatment: A tutorial. *Stat Med* 37:2530–2546. <https://doi.org/10.1002/sim.7628>
- Mahmud N, Schonstein E, Schaafsma F et al (2010) Pre-employment examinations for preventing occupational injury and disease in workers. *Cochrane Database Syst Rev*. <https://doi.org/10.1002/14651858.CD008881>
- Moscato G, Pala G, Boillat MA et al (2011) EAACI position paper: Prevention of work-related respiratory allergies among pre-apprentices or apprentices and young workers. *Allergy* 66:1164–1173. <https://doi.org/10.1111/j.1398-9995.2011.02615.x>
- Muth T, Bahemann A, Voß HJ, Borsch-Galetke E (2005) Gesundheitlich begründete Ausbildungsabbrüche. *Arbeitsmedizin, Sozialmedizin, Umweltmedizin* 40:182
- Nixon R, Roberts H, Frowen K, Sim M (2006) Knowledge of skin hazards and the use of gloves by Australian hairdressing students and practising hairdressers. *Contact Derm* 54:112–116. <https://doi.org/10.1111/j.0105-1873.2006.00790.x>
- Patuzzi M (2012) Ausbildungsreport 2012 Bayern. DGB-Jugend Bayern, Munich
- Peden D, Reed CE (2010) Environmental and occupational allergies. *J Allergy Clin Immunol* 125:S150–S160. <https://doi.org/10.1016/j.jaci.2009.10.073>
- Peres F, Rodrigues KM, Silva Peixoto Belo MS da et al (2013) Design of risk communication strategies based on risk perception among farmers exposed to pesticides in Rio de Janeiro State, Brazil. *Am J Ind Med* 56:77–89. <https://doi.org/10.1002/ajim.22147>
- Petersen ML, Porter KE, Gruber S et al (2012) Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res* 21:31–54. <https://doi.org/10.1177/0962280210386207>
- Pirracchio R, Petersen ML, Laan M van der (2015) Improving propensity score estimators' robustness to model misspecification using super learner. *Am J Epidemiol* 181:108–119. <https://doi.org/10.1093/aje/kwu253>
- Polley E, LeDell E, Laan M van der (2016) SuperLearner: super learner prediction. R package documentation available at <https://cran.r-project.org/web/packages/SuperLearner/index.html>
- Pounds L, Duysen E, Romberger D et al (2014) Social marketing campaign promoting the use of respiratory protection devices among farmers. *J Agromed* 19:316–324. <https://doi.org/10.1080/1059924X.2014.917350>
- R Core Team (2016) R: a language and environment for statistical computing. Available at: <https://www.r-project.org/>
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66:688–701. <https://doi.org/10.1037/h0037350>
- Rubin DB (1978) Bayesian inference for causal effects: the role of randomization. *Ann Stat* 6:34–58
- Samii C, Paler L, Daly SZ (2016) Retrospective causal inference with machine learning ensembles: an application to anti-recidivism policies in Colombia. *Polit Anal* 24:434–456. <https://doi.org/10.1093/pan/mpw019>
- Schnitzer ME, Laan MJ van der, Moodie EEM, Platt RW (2014) Effect of breastfeeding on gastrointestinal infection in infants: a targeted maximum likelihood approach for clustered longitudinal data. *Ann Appl Stat* 8:703–725
- Seifried J, Baumgartner A (2009) Lernen aus Fehlern in der betrieblichen Ausbildung—Problemfeld und möglicher Forschungszugang. In: *bwp@ Berufs- und Wirtschaftspädagogik online*. 17
- Tarlo SM, Lemiere C (2014) Occupational asthma. *N Engl J Med* 370:640–649. <https://doi.org/10.1056/NEJMr1301758>
- Tarlo SM, Liss GM (2005) Prevention of occupational asthma—practical implications for occupational physicians. *Occup Med (Lond)* 55:588–594. <https://doi.org/10.1093/occmed/kqi182>
- Vandenplas O, Dressel H, Wilken D et al (2011) Management of occupational asthma: cessation or reduction of exposure? A systematic review of available evidence. *Eur Respir J* 38:804–811. <https://doi.org/10.1183/09031936.00177510>
- Vanderweele TJ, Arah OA (2011) Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22:42–52. <https://doi.org/10.1097/EDE.0b013e3181f74493>
- Xie Y, Brand JE, Jann B (2012) Estimating heterogeneous treatment effects with observational data. *Sociol Methodol* 42:314–347

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

Affiliations

Daloha Rodríguez-Molina^{1,2}  · Swaantje Barth¹  · Ronald Herrera¹  · Constanze Rossmann³  · Katja Radon¹  · Veronika Karnowski⁴ 

¹ Occupational and Environmental Epidemiology and NetTeaching Unit, Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, University Hospital of Munich (LMU), Ziemssenstr. 1, 80336 Munich, Germany

² Department of Medical Informatics, Biometry and Epidemiology (IBE), Ludwig-Maximilians University Munich (LMU), Marchioninstr. 15, 81377 Munich, Germany

³ Department of Media and Communication Sciences, University of Erfurt, Nordhäuser Str. 63, 99089 Erfurt, Germany

⁴ Department of Communication Studies and Media Research (IfKW), Ludwig-Maximilians University Munich (LMU), Oettingenstr. 67, 80538 Munich, Germany

An educational intervention to improve knowledge about prevention against occupational asthma and allergies using targeted maximum likelihood estimation

Daloha Rodríguez-Molina^{1,2}, Swaantje Barth¹, Ronald Herrera¹, Constanze Rossmann³, Katja Radon¹, Veronika Karnowski⁴

1: Occupational and Environmental Epidemiology & NetTeaching Unit; Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine; University Hospital of Munich (LMU); Ziemssenstr. 1, 80336 Munich, Germany.

2: Department of Medical Informatics, Biometry and Epidemiology (IBE); Ludwig-Maximilians University Munich (LMU); Marchioninstr. 15, 81377 Munich, Germany.

3: Department of Media and Communication Sciences; University of Erfurt; Nordhäuser Str. 63, 99089 Erfurt, Germany.

4: Department of Communication Studies and Media Research (IfKW); Ludwig-Maximilians University Munich (LMU); Oettingenstr. 67, Munich 80538, Germany.

Corresponding author: Daloha Rodríguez-Molina; e-mail: daloha.rodriguez_molina@med.uni-muenchen.de

Online Resource 1

Online Resource 1 Study center, randomization status and study completion in eligible apprentices (n = 117), Bavaria, Germany, 2014.

Study center	Allocation group	Completion (n)	Completion (%)	Average completion (%)
Study center 1	Control	16/22	72.73%	60.55%
Study center 2	Control	8/10	80.00%	
Study center 3	Control	4/6	66.67%	
Study center 4	Control	1/6	16.67%	
Study center 5	Control	2/3	66.67%	
Study center 6	Intervention	7/35	20.00%	34.54%
Study center 7	Intervention	4/17	23.53%	
Study center 8	Intervention	1/8	12.50%	
Study center 9	Intervention	2/4	50.00%	
Study center 10	Intervention	4/6	66.67%	

An educational intervention to improve knowledge about prevention against occupational asthma and allergies using targeted maximum likelihood estimation

Daloha Rodríguez-Molina^{1,2}, Swaantje Barth¹, Ronald Herrera¹, Constanze Rossmann³, Katja Radon¹, Veronika Karnowski⁴

1: Occupational and Environmental Epidemiology & NetTeaching Unit; Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine; University Hospital of Munich (LMU); Ziemssenstr. 1, 80336 Munich, Germany.

2: Department of Medical Informatics, Biometry and Epidemiology (IBE); Ludwig-Maximilians University Munich (LMU); Marchioninstr. 15, 81377 Munich, Germany.

3: Department of Media and Communication Sciences; University of Erfurt; Nordhäuser Str. 63, 99089 Erfurt, Germany.

4: Department of Communication Studies and Media Research (IfKW); Ludwig-Maximilians University Munich (LMU); Oettingenstr. 67, Munich 80538, Germany.

Corresponding author: Daloha Rodríguez-Molina; e-mail: daloha.rodriguez_molina@med.uni-muenchen.de

Online Resource 2

R CODE TO IMPLEMENT THE ANALYSES

```
# R code for implementation
# Author: Daloha Rodriguez-Molina
# The full code is available online at https://github.com/darokun/R---txt2paul/blob/master/23.%20R%20code%20for%20implementation.R
#---
# Loading required packages
library(SuperLearner)
library(tmlr)
library(devtools)
```

```

library(httr)
install_github("lendle/tmlecte")
library(tmlecte)

# Setting up the working directory
setwd("~/Dropbox/R - txt2paul")

# Reading in the data
d <- "/Users/daro/Desktop/txt2PAUL/z_Dokumente fuer
Daloha/txt2PAUL_Gesamt_vers1_130115.csv"
data <- read.csv(d, header=T)
#---

#---
# SOCIODEMOGRAPHIC AND MEDICAL VARIABLES

# handling of variables is shown in full code online at
https://github.com/darokun/R---
txt2paul/blob/master/23.%20R%20code%20for%20implementation.R
#---

#---
# OUTCOME VARIABLE
# handling of variables is shown in full code online at
https://github.com/darokun/R---
txt2paul/blob/master/23.%20R%20code%20for%20implementation.R
#---

#---
# TABLES 1 AND 2
# code to reproduce tables 1 and 2 are shown in full online at
https://github.com/darokun/R---
txt2paul/blob/master/23.%20R%20code%20for%20implementation.R
#---

#---
# LOGISTIC REGRESSION MODELS (APPENDIX - table 4)

# Crude all 6
int.all_wissen_nachher.cat.glm.crude <- glm(all_wissen_nachher.factor ~
      gruppe,
      data = int.data.wo.nas,
      family = binomial,
      na.action = na.omit)

# summary(int.all_wissen_nachher.cat.glm.crude)
exp(coef(int.all_wissen_nachher.cat.glm.crude))
exp(confint.default(int.all_wissen_nachher.cat.glm.crude))

# Adjusted all 6
int.all_wissen_nachher.cat.glm <- glm(all_wissen_nachher.factor ~
      gruppe +
      all_wissen.factor +
      alterkat2 +
      sex.factor +
      smoking.status.factor +
      schulabschluss.factor.twolevels +
      risk_perception.factor +
      asthma.or.rhinoconj.factor +
      par.asthma.factor,
      data = int.data.wo.nas,
      family = binomial,

```

```

na.action = na.omit)
# summary(int.all_wissen_nachher.cat.glm)
exp(coef(int.all_wissen_nachher.cat.glm))
exp(confint.default(int.all_wissen_nachher.cat.glm))

# code to reproduce sensitivity analyses is available online at
https://github.com/darokun/R---
txt2paul/blob/master/23.%20R%20code%20for%20implementation.R

# Logistic regression is not able to provide adequate estimates (due to
small sample size)
# Proposed solution: use TMLE
#---

#---
# TMLE

# variables:
# treatment
A <- as.numeric(int.data.wo.nas$gruppe)

# missings
Delta <- na.vector.all_wissen_nachher

# matrix of covariates
ff <- model.matrix( ~ int.data.wo.nas$sex +
int.data.wo.nas$alterkat2.binary + int.data.wo.nas$schulabschluss.binary +
int.data.wo.nas$smoking.status + int.data.wo.nas$asthma.or.rhinoconj +
int.data.wo.nas$risk_perception + int.data.wo.nas$par.asthma +
int.data.wo.nas$all_wissen)
W <- ff[,-1]
colnames(W) <- paste("W",1:8, sep="")

# SuperLearner libraries
create.SL.knn <- function(k = c(20, 30)) {
  for(mm in seq(length(k))) {
    eval(parse(text = paste('SL.knn.', k[mm], '<- function(..., k = ',
k[mm],
                                ') SL.knn(..., k = k)', sep = '')), envir =
.GlobalEnv)
  }
  invisible(TRUE)
}
create.SL.knn(c(20, 30, 40, 50, 60, 70))
SL.library <- c("SL.glmnet", "SL.knn.20", "SL.knn.40", "SL.knn.40",
"SL.knn.60",
               "SL.mean", "SL.bayesglm", "SL.glm",
               "SL.step.interaction", "SL.randomForest")

# Main outcome: all 6 correct measures
Y <- int.data.wo.nas$all_wissen_nachher
set.seed(1985)
result6 <- tmle(Y=Y,A=A,W=W, Delta=Delta, family= "binomial", cvQinit =
FALSE,
               g.SL.library = SL.library,
               Q.SL.library = SL.library,
               gbound = c(0.025, 0.995))
summary(result6)

# Sensitivity analyses
# at least 5 correct measures

```

```

Y <- int.data.wo.nas$all_wissen_nachher5
set.seed(1985)
result5 <- tmle(Y=Y,A=A,W=W, Delta=Delta, family= "binomial", cvQinit =
FALSE,
               g.SL.library = SL.library,
               Q.SL.library = SL.library,
               gbound = c(0.025, 0.995))
summary(result5)

# at least 4 correct measures
Y <- int.data.wo.nas$all_wissen_nachher4
set.seed(1985)
result4 <- tmle(Y=Y,A=A,W=W, Delta=Delta, family= "binomial", cvQinit =
FALSE,
               g.SL.library = SL.library,
               Q.SL.library = SL.library,
               gbound = c(0.025, 0.995))
summary(result4)
#---

#---
# ATT and ATC
# Main outcome: all 6 correct measures
# outcome variable
Y <- int.data.wo.nas$all_wissen_nachher

# att
set.seed(1985)
a <- 1
att6 <- tmle.cte(Y=Y,A=A,B=W, Delta=Delta, a=a, family= "binomial",
                Q.method = "SL", g.method = "SL", gDelta.method = "SL",
                g.SL.library = SL.library,
                gDelta.SL.library = SL.library,
                Q.SL.library = SL.library)

# atc
set.seed(1985)
a <- 0
atc6 <- tmle.cte(Y=Y,A=A,B=W, Delta=Delta, a=a, family= "binomial",
                Q.method = "SL", g.method = "SL", gDelta.method = "SL",
                g.SL.library = SL.library,
                gDelta.SL.library = SL.library,
                Q.SL.library = SL.library)

att6
atc6

#---
# Sensitivity analyses
# at least 5 correct measures
# outcome variable
Y <- int.data.wo.nas$all_wissen_nachher5

# att
set.seed(1985)
a <- 1
att5 <- tmle.cte(Y=Y,A=A,B=W, Delta=Delta, a=a, family= "binomial",
                Q.method = "SL", g.method = "SL", gDelta.method = "SL",
                g.SL.library = SL.library,
                gDelta.SL.library = SL.library,
                Q.SL.library = SL.library)

```

```

# atc
set.seed(1985)
a <- 0
atc5 <- tmle.cte(Y=Y,A=A,B=W, Delta=Delta, a=a, family= "binomial",
                Q.method = "SL", g.method = "SL", gDelta.method = "SL",
                g.SL.library = SL.library,
                gDelta.SL.library = SL.library,
                Q.SL.library = SL.library)

att5
atc5

#---
# at least 4 correct measures
# outcome variable
Y <- int.data.wo.nas$all_wissen_nachher4

# att
set.seed(1985)
a <- 1
att4 <- tmle.cte(Y=Y,A=A,B=W, Delta=Delta, a=a, family= "binomial",
                Q.method = "SL", g.method = "SL", gDelta.method = "SL",
                g.SL.library = SL.library,
                gDelta.SL.library = SL.library,
                Q.SL.library = SL.library)

# atc
set.seed(1985)
a <- 0
atc4 <- tmle.cte(Y=Y,A=A,B=W, Delta=Delta, a=a, family= "binomial",
                Q.method = "SL", g.method = "SL", gDelta.method = "SL",
                g.SL.library = SL.library,
                gDelta.SL.library = SL.library,
                Q.SL.library = SL.library)

att4
atc4

#-----#
# END OF SCRIPT #
#-----#

```