



On the identification of the effect of education on health: a comment on Fonseca et al. (2020)

Pedro Albarrán¹ · Marisa Hidalgo-Hidalgo² · Iñigo Iturbe-Ormaetxe¹ 

Received: 22 March 2021 / Accepted: 13 January 2022 / Published online: 17 February 2022
© The Author(s) 2022

Abstract

In this note we revisit the paper by Fonseca et al. (Series 11: 83-103, 2020) who find that education has a positive effect on health. They use several compulsory schooling reforms as instruments for education. Our objective is to replicate this causal finding, so we start by thoroughly discussing their identification strategy. In particular, we emphasize the importance of carefully defining birth cohort groups and using country-specific time trends. Once we take these issues into account, we show that the instrument they use is too weak.

Keywords Health · Education · Instrumental variables · Compulsory schooling laws

JEL Classification I1 · I2

1 Introduction

There are many works in the literature that find a strong correlation between education and health outcomes. However, there is still no consensus on whether education improves health. Several recent studies as Clark and Royer (2013), Meghir et al. (2018), Albarrán et al. (2020) and Avendano et al. (2020) do not find any causal effect of education on health. Xue et al. (2021) perform a meta-analysis including 99 published papers on this topic. They find a slight publication bias that favors papers that report a positive effect of education on health. Once they correct for this bias, they find that education no longer has a causal effect on health.

In a recent paper in this journal, Fonseca et al. (2020, in the sequel FMZ20) combine data on education and health from three sources, SHARE (continental Europe), ELSA (England) and HRS (USA). They find that education has a positive effect on health,

✉ Iñigo Iturbe-Ormaetxe
iturbe@ua.es

¹ Fundamentos del Análisis Económico (FAE), Universidad de Alicante, Alicante, Spain

² Dpto. Economía, Métodos Cuantitativos e Historia Económica, Universidad Pablo de Olavide de Sevilla, Sevilla, Spain

both with self-reported measures and with various objective measures. The technique they use to prove causality is instrumental variables, exploiting the exogenous variation in education generated by compulsory schooling laws (CSLs) that increase the minimum school leaving age (SLA) as an instrument for education. The intuition is that individuals in year-of-birth cohorts affected by CSLs are forced to stay in school longer than earlier cohorts. Comparing cohorts that are a few years apart, CSLs induce an exogenous change that only affects people's health through increased education.

Given this apparent contradiction between what FMZ20 obtain and what seems to be a growing consensus in this literature—the absence of a causal effect—we have tried to replicate their work.

The main conclusion we reach, using the same data sets and the same countries, is that we cannot prove a causal effect of education on any of the different health outcomes. The reason for this result is not that education has no effect on health, but rather that the instrument (CSL) is too weak to generate an exogenous variation in people's education for these datasets. We want to stress that we cannot exclude the possibility of a causal effect, but simply that this methodology does not work in this specific case. In particular, we do not claim that CSLs are an invalid instrument. There are several papers in this literature that also rely on CSLs and have a convincing identification strategy. Some recent examples are Brunello et al. (2013), Crespo et al. (2014), Brunello et al. (2016), Albarrán et al. (2020), and Hofmarcher (2021). As much as an instrument has proven to be valid before, this does not mean that it will work in all situations. The validity of an instrument (or more generally of an identification strategy) is, to some extent, an empirical matter determined by the specific context, institutional setting and sample. Then, in each new application one must carefully show that it is plausible that it works.

This note is structured as follows. First, we discuss data problems and model specification issues often found when using CSLs as an instrument in a multi-country setting. Next, we present our replication results, without accounting for concerns previously discussed and once we account for them. Finally, we conclude with a simulation exercise focused on the importance of using country-specific trends to avoid getting misleading conclusions about the validity of the instrument.

2 Data and model specification

We construct a database with the same fourteen countries as FMZ20. Since credible causal results rely on the validity of the identification strategy, we focus on potential problems in the first stage. We claim that the problems in FMZ20 lie on this part. However, we have managed to reproduce closely their Summary Statistics and the results of the effect of years of education on health outcomes (OLS), that is, Tables 1 and 5 in FMZ20 (see the Appendix).

To account for the potential endogeneity of education, we use the exogenous variation in education provided by the CSLs and therefore we estimate the following first-stage model (like Eq. 3 in FMZ20):

$$Ed_{icb} = \beta_0 + \beta_1 Y_{Ccb} + \beta_2 W_i + T_{cb} + C_c + u_{icb} \quad (1)$$

Table 1 Impact of education reforms on years of education

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
YC	0.2736*** (0.0705)	0.2744*** (0.0691)	0.3507*** (0.1250)	0.0525 (0.0698)	0.2750*** (0.0714)
Female	− 0.5697*** (0.0385)	− 0.5657*** (0.0383)	− 0.5702*** (0.0382)	− 0.5683*** (0.0381)	− 0.5550*** (0.0381)
Number of cohort blocks	8	10	8	10	–
Country-specific trend	No	No	Yes	Yes	No
Linear/quadratic trend	–	–	–	–	Linear
Observations	52,960	52,960	52,960	52,960	52,960
R-squared	0.2648	0.2649	0.2814	0.2822	0.2639
IVF-stat	15.055	15.775	7.870	0.565	14.837
IVp-value	0.000	0.000	0.005	0.453	0.000
CommonTrend_F-stat			20.154	23.563	
CommonTrend_p-value			0.000	0.000	
	Model 6	Model 7	Model 8	Model 9	Model 10
YC	0.2431*** (0.0671)	0.0763 (0.0603)	0.0889 (0.0566)	− 0.0650 (0.0680)	− 0.0357 (0.0640)
Female	− 0.5694*** (0.0384)	− 0.5620*** (0.0380)	− 0.5707*** (0.0383)	− 0.5727*** (0.0382)	− 0.5702*** (0.0383)
Number of cohort blocks	–	–	40	–	40
Country-specific trend	No	Yes	Yes	Yes	Yes
Linear/quadratic trend	Quadratic	Linear	Linear	Quadratic	Quadratic
Observations	52,960	52,960	52,960	52,960	52,960
R-squared	0.2651	0.2785	0.2802	0.2804	0.2816
IVF-stat	13.116	1.602	2.465	0.914	0.312
IVp-value	0.000	0.206	0.117	0.339	0.577
Common Trend_F-stat		65.439	70.459	41.309	42.820
Common Trend_p-value		0.000	0.000	0.000	0.000

YC is years of compulsory education; IVF-stat is the first-stage F-statistic and IVp-value its corresponding p-value; common trend F-stat is the F-statistic of the test in which the null is that time trends are not country-specific. Standard errors are clustered at the country-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Here, Ed_{icb} denotes years of education of an individual i from a country c born in the cohort year b^1 ; YC_{cb} is the number of years she was required to attend school; W_i is a dummy equal to 1 for women; T_{cb} denotes the “year of birth time trends” (below we discuss in detail how to define them), and C_c is a country fixed effect, captured using

¹ In general, we can potentially observe the outcome for two individuals born in the same cohort b at two different ages. We omit this additional dimension for the ease of exposition at this point.

country dummies for each country:

$$C_c = \sum_{m=1}^M \mu_m D_{cm} \quad (2)$$

where D_{cm} is a dummy that takes value 1 if observation c comes from country m , and 0 otherwise.²

It is worth noting that a person's years of compulsory schooling only varies, in principle, with their country and year of birth (that is, it depends on whether they are affected by a CSL in their country). Table 4 in FMZ20 shows that only eight of the fourteen countries implement a CSL reform.³ For the remaining six countries there is no variation in the instrument across year-of-birth cohorts: YC_{cb} is YC_c , so its effect on education is already captured by the country dummies. In other words, including these six countries does not contribute anything to the identification of the parameter β_1 . However, since we want to focus on the role of time trends for proper identification, we include the fourteen countries of FMZ20 to keep our analysis as close as possible to the original one.

In these types of models, by construction, treated individuals are younger than controls, so we must include time trends to account for secular tendencies. This is particularly relevant when we talk about education or health. Younger cohorts typically have a higher educational level and also better health. Adding time trends to the specification allows us to identify the effect of the reform on those people who, even with the positive trend, would not have acquired more education *without the reform*. When we do not include these trends, we may end up incorrectly attributing improvements in education to school reforms, when in fact those improvements are simply the result of secular trends.

FMZ20 include forty cohorts of individuals in their analysis, including those born between 1917 and 1956. The age of their sample ranges from 50 to 89. The inclusion of a temporal trend, as the authors do, can partially control for the positive trend in education and health in the countries considered. However, there are several potential issues to be aware when specifying time trends.

First, it is very unusual to impose a common time trend for all countries. This has shown too restrictive to capture the great heterogeneity in the temporal evolution of health and education. From Stephens and Yang (2014) we know that it is crucial that these time trends are country-specific, T_{cb} , rather than a single time trend for all countries, T_b . Suppose that different countries exhibit different time trends in education, but we impose a common time trend (as FMZ20 do). This model would have an omitted variable, the country-specific time evolution, which is correlated with the instrument, YC_{cb} , since this has also country-specific time variation. The differential

² This specification includes one dummy for each group in the sample (in this case, M countries), so implicitly we set $\beta_0 = 0$ in (1). We use this type of specification throughout this note instead of the equivalent specification where the model has a constant and a set of dummies for each group but one (the reference group).

³ The countries that implement a CSL reform are Austria, England, Sweden, The Netherlands, Italy, France, Greece, and the Czech Republic. Countries without a reform are Poland, Switzerland, Belgium, Germany, Spain, and the USA.

country-specific trend would be attributed to the (country-specific) reform, YC_{cb} . As a result, its estimated coefficient would be biased and could lead us to wrongly conclude that the instrument is valid. To avoid any potential bias, we use a more general and flexible specification, T_{cb} . Moreover, this allows us to test whether the restrictions implied by the common trend are true, thus giving an empirical answer to the question of which is the correct specification.⁴

Second, a specific formula must be used for time trends. A relatively flexible approach is to group multiple birth cohorts into one group and use a dummy variable for each of these groups:

$$T_b = \sum_{j=1}^J \delta_j D_{bj} \tag{3}$$

where D_{bj} is a dummy that takes value 1 when the birth cohort b belongs to the cohort group j (and 0 otherwise). In the case of country-specific trends if we have M countries and J groups of cohorts:

$$T_{cb} = \sum_{j=1}^J \sum_{m=1}^M \delta_{jm} D_{bj} D_{cm} \tag{4}$$

where D_{cm} is a dummy variable that takes value 1 for country m , and 0 otherwise. As discussed above, testing that $\delta_{jm} = \delta_j$ for all m and for each j implies testing whether a common trend specification is preferred over one with country-specific trends.

This approach requires choosing the number of groups, J , and which cohorts are included in each of them. In the most extreme case, the groups could include a single cohort. In this case, this would involve defining a dummy variable for each of the 40 years of birth. In any case, it is very important to do a robustness analysis considering various alternatives. We guess that FMZ20 use eight groups in their main specification (see Table 1 in FMZ20, Summary Statistics). Each of them corresponds to a set of five birth cohorts. This is very unfortunate, due to the timing of the reforms shown in Table 4. To illustrate this problem, we represent in the figure below all the years of birth included in this analysis, the eight time dummies and the control and treatment groups in each country.

The problem with this specification of time trends is that there are two countries (Italy and Greece) in which the treated cohorts correspond exactly to the years included in the last time dummy. For these two countries, the effect of being a treated cohort cannot be identified separately from the country-specific time effect in these last five years. This is another potential source of bias that can lead to the erroneous conclusion that the instrument is valid. One way to verify this problem is to slightly change the way we define the birth cohort dummies to avoid that the treated cohorts correspond to a single time dummy. For example, suppose we group birth years into ten blocks

⁴ Recent papers that use CSL as instruments and include country-specific time trends are Brunello et al. (2013), Crespo et al. (2014), Mazzona (2014), Gathmann, Jürges, and Reinhold (2015), Brunello et al. (2016), Albarrán et al. (2020), and Hofmarcher (2021).

of four cohorts each instead of eight blocks of five cohorts each.⁵ Intuitively, things should not change much.

An alternative way to define time trends is by using age trends. FMZ20 replicate their analysis with polynomial specifications for age (in particular, age and age squared), rather than birth cohort dummies. For the same reason discussed above, these age trends should be country-specific:

$$T_{bc} = \sum_{m=1}^M \gamma_{0m} D_{cm} + a_b \sum_{m=1}^M \gamma_{1m} D_{cm} + a_b^2 \sum_{m=1}^M \gamma_{2m} D_{cm} \quad (5)$$

where a_b is the age of the individuals of cohort b at the time t of the survey. Again, we can test whether a common trend specification is preferred over one with country-specific trends by using a joint test for all the country-specific coefficients being the same across countries, that is, $\gamma_{0m} = \gamma_0$, $\gamma_{1m} = \gamma_1$, $\gamma_{2m} = \gamma_2$ for all m .

Third, using the forty cohorts of individuals in the sample is an unusual choice in the literature. In the ideal experiment, one would like to use only individuals born on 1st January in the year of the first cohort affected by the reform as the treatment group and those born on 31st December of the previous year as the control group to ensure that they are as similar as possible and that the only source of (exogenous) variation comes from being affected or not by the reform. Of course, this is typically impractical because the sample would be too small. In practice, the literature uses a small number of cohorts around the first affected cohort. Typical window sizes are five or seven cohorts, so no more than a total of fourteen cohorts are included in the analysis. It is a very unrealistic assumption to consider that cohorts born more than 30 years apart are good counterfactuals for each other, even if extremely flexible time trends were used. It is very likely that at least part of the effect captured by the instrument is due to uncontrolled differences between the treatment and control groups.

Another reason for limiting the number of cohorts in the control and treatment groups is that several countries changed the length of compulsory schooling several times during the period considered (Horfmarcher 2021). However, as we said above, since we want to focus on the issue of temporal trends, we are going to include all 40 cohorts in our analysis.

3 Results and discussion

We present our main results in Table 1. We use the fourteen countries and the forty cohorts in FMZ20 to estimate the first stage of ten alternative models. We consider a number of different first-stage specifications to cover all the relevant dimensions

⁵ FMZ20 mention in the first paragraph of Sect. 4 (Empirical Strategy) that they control for “birth cohort dummies for nine age groups” (bold text is ours). However, their Table 1 shows only eight values for the categorical variable “Cohort”. Given this inconsistency and the lack of replication code, it is unclear the final specification used by FMZ20 among the following three: (i) eight cohort groups, (ii) nine cohort groups (but then notice it is not obvious how to evenly split 40 birth years into 9 groups and no comment is made), or (iii) ten cohorts groups (and they meant nine as that would be the number of cohort dummies in a model specification with a constant).

discussed in the literature that uses CSLs as an instrument. In Model 1 we replicate the specification in FMZ20 (Table 6) that has a common time trend and no control for age. To estimate the common time trend, we group birth cohorts into eight blocks of five cohorts each. The model estimates the corresponding dummies for each block. Model 2 is similar to Model 1, with the only change that we group the cohort dummies into ten blocks of four cohorts each. Models 3 and 4 add a country-specific trend to the specifications of models 1 and 2, respectively. To do this, we interact the cohort dummies with the country dummies.

In the rest of the models, we consider polynomial specifications for age to account for time trends. Model 5, instead of including cohort dummies, adds individual age as a linear regressor. The effect of age is assumed to be common across all countries. Model 6 is like model 5, but adding age squared as a regressor. Models 7 to 10 all include interactions of age with the country dummies. In model 7, age enters only linearly. Model 8 is like 7, also adding forty cohort dummies. Models 9 and 10 are like models 7 and 8, respectively, but adding a quadratic term in age. That is, both of them use a quadratic specification for the time trend (interactions of age and its square with the country dummies) rather than a linear one. The only difference is that Model 10 additionally includes the forty cohort dummies. Our preferred specifications are those in models 9 and 10. They correspond to those used by Crespo et al. (2014), Brunello et al. (2016), and Hofmarcher (2021).

Notice that FMZ20 do not provide results on standard errors, nor on the F-statistic or the number of observations (see Table 6 in FMZ20). We can only compare our results in Model 1 with their reported first-stage estimated coefficient, and its magnitude is different (0.5623 vs. 0.2736). This is surprising since we closely reproduce their Tables 1 and 5 (see the Appendix). In any case, the main finding is qualitatively the same: the estimated effect is positive and significant, so the instrument is valid.⁶ This conclusion is maintained in models 2, 3 (although the F-statistic is now clearly lower), 5, and 6. Except for Model 3, every time we include country-specific trends, the effect of the instrument completely disappears. The instrument is too weak and/or its effect cannot be separately identified from the country-specific secular trend. In all these cases the F-statistics here are always below the “safe” value of 10 (see Staiger and Stock, 1997). Furthermore, we find strong evidence against a common time trend in all cases. In summary, a proper specification of time trends, combining both a flexible and careful definition of the birth cohorts blocks and country-specific trends, implies a first stage that does not validate the identification of the causal effect of education on health.

Following the work by Stephens and Yang (2014), which shows that it is crucial to account for country-specific time trends, most (if not all) recent papers in the related literature include country-specific trends. In this case, we have empirically rejected that the time trend is common to all countries. So, they are relevant variables that must be included in the model. Moreover, the country-specific time trends are

⁶ Note that the estimated coefficient of YC in FMZ20 (0.5623) is much larger than the one obtained by Crespo et al. (2014) who also use SHARE data (0.145, s.e. 0.125). Other recent works have also obtained much lower estimates. Brunello et al. (2016) obtain estimates in the range 0.251–0.344, depending on the specification. Albarrán et al. (2020) and Hofmarcher (2021) estimate coefficients of 0.168 and 0.161, respectively.

Year of birth	1917	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56					
Austria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
England	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Sweden	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Netherlands	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Italy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
France	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Greece	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Czech R.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Note: Country-cohorts in the control group are indicated with a zero. Country-cohorts in the treatment group are indicated with a one and highlighted in red.

Fig. 1 Time dummies, control, and treatment groups in each country

obviously correlated with the secular trends in education: younger cohorts are often more educated, but with different patterns across countries. If omitted, the “instrument” exploits a source of variation that is not exogenous, since it does not come from the reform alone but also from the country-specific trend.

We find extremely important the differences between models 3 and 4. As we have said above, both models include country-specific time trends, and the only difference between them is that in model 3 we group the cohorts into eight groups of five, while in model 4 we group them into ten groups of four. A priori, the results should be very similar. The fact that they are not indicates that there is a problem with this specification. As we discuss above, we argue that the reason for this discrepancy is that in Model 3, all the time effect of the last five years in Italy and Greece is erroneously attributed to the instrument.

The fact that the instrument has no bite is also apparent in Fig. 1 of FMZ20, in which there is no clear jump in education around the cut-off. To further illustrate this, we have built a similar graph for each one of the eight reform countries:

In Fig. 2 we see three countries where there is a clear jump up around the threshold (England, the Netherlands, Sweden). However, in Austria, France, Greece, and Italy there is a downward leap around the threshold. Finally, in the Czech Republic, there is no noticeable change.⁷

As an additional check, we present estimates corresponding to the reduced form of the ten models in Table 1. The reduced form estimates the direct effect of the instrument on the outcome variables. We report in Table 2 the results corresponding to the case in which the dependent variable is a dummy variable that takes value 1 when the subject reports to have poor health. Again, we highlight in bold type models with country-specific time trends (Fig. 3).

Again, we see that, except for Model 3, every time we include country-specific trends, the effect of years of compulsory education on the endogenous variable vanishes. This confirms our evidence from the first-stage analysis. In words of Angrist and

⁷ Despite of the publication bias (it is less likely to find published papers where the instrument does not work), some papers do report weak instruments problems when using CSL as an instrument. For instance, Crespo, López-Noval, and Mira (2014), who also work with SHARE data. More generally, there are good reasons to consider that not every compulsory schooling reform is truly increasing the educational attainment. On the one hand, some reforms make legal what is already unofficially implemented (most kids already leave school after the new compulsory age). On the other hand, the type of students typically affected are largely people who end up leaving school as soon as they can.

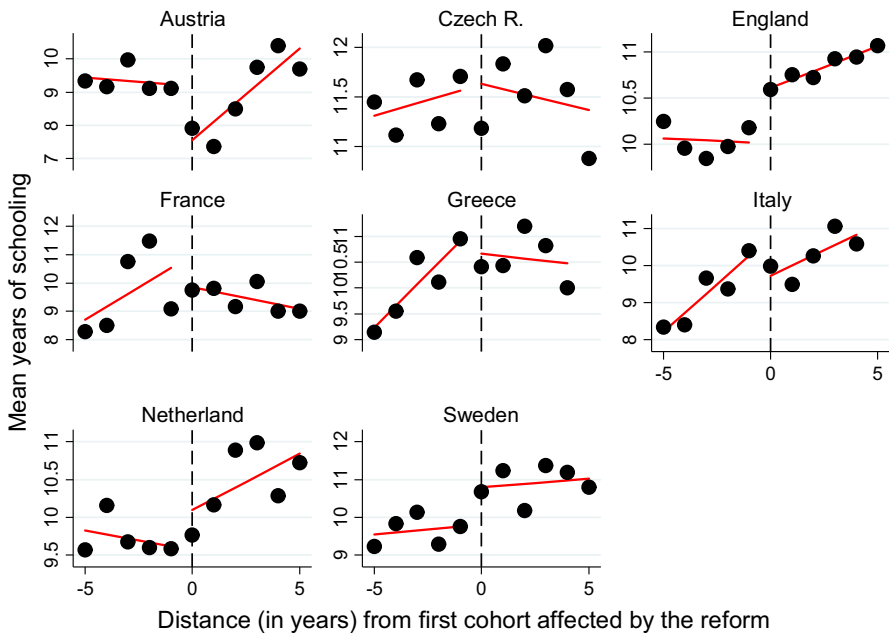


Fig. 2 Control and treatment, by country

Pischke, “If you can’t see it in the reduced form, it ain’t there” (Angrist and Pischke, 2015).

4 Simulations

To further illustrate the problem of time trends, we are going to construct an artificial example in which, by definition, the instrument has no effect on education. We simulate 60,000 observations for five countries. Each observation represents an individual. To keep things simple, there is no gender. We number the years from 0 to 39, and we define age as 89 minus the year, so the age of individuals ranges from 50 to 89 years. In each of the five countries, there is a reform that increases the number of years of compulsory schooling. However, these reforms have no effect on education. The years of schooling of the population follow a linear time trend for the entire period, and we assume that the error term u follows a normal distribution with mean 0 and standard deviation of 2. We summarize the example in Table 3.

As can be seen, the first simulated cohort affected in each of the five countries reproduces the real ones for AT, NL, IT, FR and GR, respectively (see Table 4 in FMZ20). Moreover, the simulated trends in years of schooling are such that the average years of schooling for these same countries and provided in Table 4 coincide with those obtained in the simulation exercise for countries 1 to 5.

The model we estimate is an OLS regression in which the dependent variable is education and the main regressor is the number of years of compulsory education. We

Table 2 Reduced form (poor health)

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
YC	– 0.0247*** (0.0057)	– 0.0246*** (0.0055)	– 0.0866*** (0.0165)	– 0.0102 (0.0077)	– 0.0227*** (0.0053)
Female	0.0305*** (0.0045)	0.0302*** (0.0045)	0.0299*** (0.0045)	0.0291*** (0.0044)	0.0309*** (0.0045)
Number of cohort blocks	8	10	8	10	–
Country-specific trend	No	No	Yes	Yes	No
Linear/quadratic trend	–	–	–	–	Linear
Observations	53,603	53,603	53,603	53,603	53,603
R-squared	0.0745	0.0750	0.0836	0.0843	0.0754
IVF-stat	18.664	19.780	27.454	1.734	18.462
IVp-value	0.000	0.000	0.000	0.188	0.000
CommonTrend_F-stat			11.256	12.523	
CommonTrend_p-value			0.000	0.000	
	Model 6	Model 7	Model 8	Model 9	Model 10
YC	– 0.0243*** (0.0054)	– 0.0029 (0.0057)	– 0.0018 (0.0053)	– 0.0111* (0.0065)	– 0.0069 (0.0066)
Female	0.0302*** (0.0045)	0.0308*** (0.0045)	0.0295*** (0.0045)	0.0293*** (0.0045)	0.0293*** (0.0045)
Number of cohort blocks	–	–	40	–	40
Country-specific trend	No	Yes	Yes	Yes	Yes
Linear/quadratic trend	Quadratic	Linear	Linear	Quadratic	Quadratic
Observations	53,603	53,603	53,603	53,603	53,603
R-squared	0.0756	0.0817	0.0831	0.0831	0.0839
IVF-stat	19.916	0.269	0.112	2.868	1.082
IVp-value	0.000	0.604	0.738	0.091	0.299
Common Trend_F-stat		29.091	35.224	22.462	18.312
Common Trend_p-value		0.000	0.000	0.000	0.000

also include country fixed effects and a time trend. We propose the same ten alternative ways to model time trends as in Tables 1 and 2. We simulate 100 repetitions of our model. We estimate each regression and plot the estimated coefficient of years of compulsory education in each replication.

The only models that correctly detect that the coefficient of years of compulsory education is zero are models 7–10 (see Fig. 3). Not surprisingly, these are the specifications used by Crespo et al. (2014), Brunello et al. (2016), and Albarrán et al. (2020), among other.

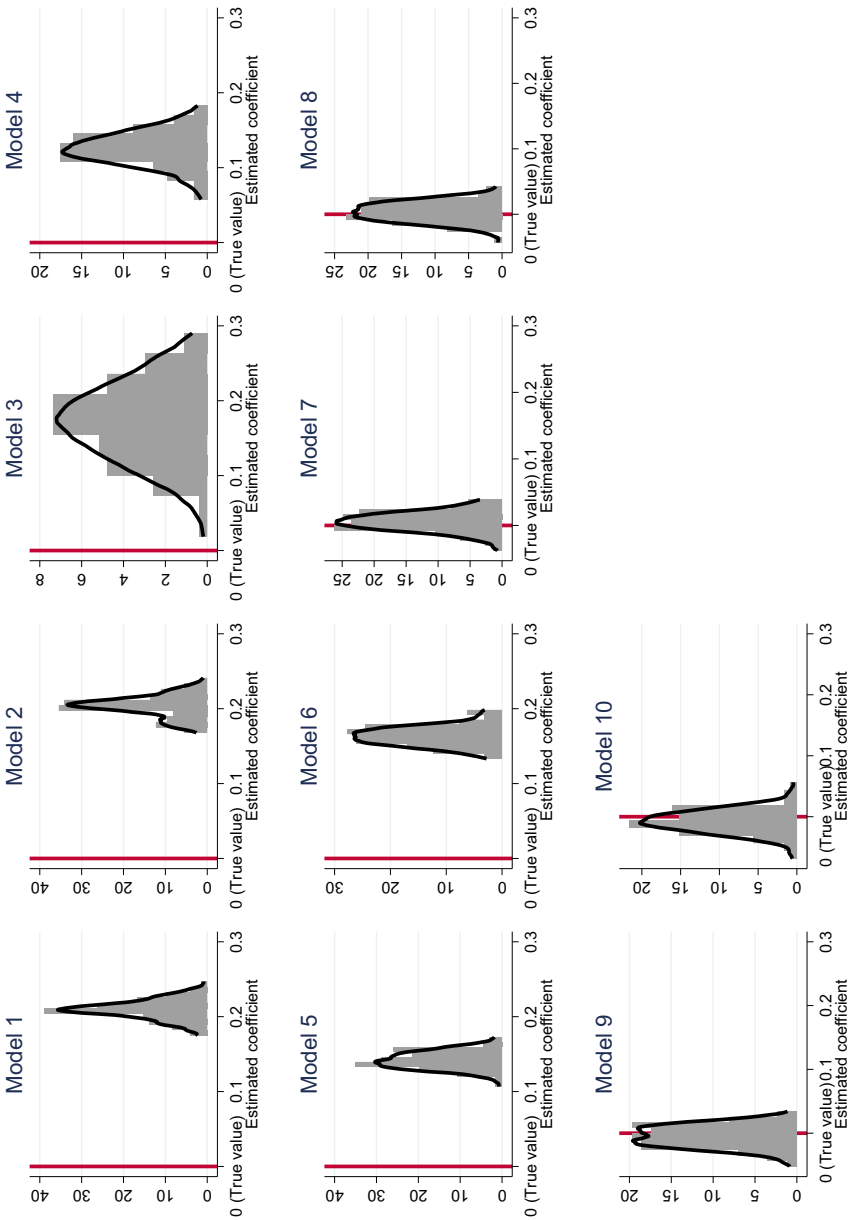


Fig. 3 Histograms and kernel density functions for the estimated coefficient of the instrument in the first stage across replications

Table 3 Data used in the simulations

Countries	First cohort affected	Change in years of compulsory schooling	Trend in education
1	32	From 8 to 9	$8 + 0.03 \cdot \text{year} + u$
2	21	From 6 to 8	$8.5 + 0.13 \cdot \text{year} + u$
3	35	From 5 to 8	$5 + 0.15 \cdot \text{year} + u$
4	6	From 7 to 8	$8.5 + 0.4 \cdot \text{year} + u$
5	35	From 6 to 9	$3.8 + 0.18 \cdot \text{year} + u$

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Tables 4 and 5.

Table 4 Summary Statistics (Table 1 in FMZ20)

Variable	Obs	Mean	Std.Dev	Min	Max
<i>Health variables</i>					
Poor health	53,334	0.334	0.472	0	1
1 + chronic illness	53,522	0.766	0.423	0	1
1 + ADLS	53,539	0.136	0.343	0	1
1 + IADLS	53,539	0.147	0.354	0	1
Cancer	53,506	0.0725	0.259	0	1
Diabetes	53,505	0.128	0.334	0	1
Heart disease	53,507	0.177	0.382	0	1
Hypertension	53,495	0.422	0.494	0	1
Arthritis	53,509	0.346	0.476	0	1
Lung disease	53,504	0.0689	0.253	0	1
Stroke	53,510	0.0470	0.212	0	1
Psychiatric illness	53,341	0.158	0.365	0	1
<i>SES variables</i>					
Years of education	52,696	11.04	3.742	0	25
Age	53,635	65.75	9.894	50	89
Cohort	53,635	5.207	1.991	1	8
Female	53,635	0.557	0.497	0	1

Table 5 OLS models years of education on health outcomes (Table 5 in FMZ20)

	Poor health	1 + Chronic	1 + ADLs	1 + IADLs
Years of education	– 0.0245*** (0.0010)	– 0.0058*** (0.0007)	– 0.0095*** (0.0006)	– 0.0134*** (0.0006)
Observations	52,463	52,635	52,647	52,647
	Cancer	Diabetes	Heart Disease	Hypertension
Years of education	0.0022*** (0.0004)	– 0.0077*** (0.0006)	– 0.0037*** (0.0005)	– 0.0082*** (0.0007)
Observations	52,620	52,618	52,620	52,608
	Arthritis	Lung disease	Stroke	Psychiatric illness
Years of education	– 0.0081*** (0.0007)	– 0.0048*** (0.0004)	– 0.0019*** (0.0003)	– 0.0039*** (0.0006)
Observations	52,622	52,617	52,623	52,473

Standard errors are clustered at the country-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

- Albarrán P, Hidalgo-Hidalgo M, Iturbe-Ormaetxe I (2020) Education and adult health: is there a causal effect? *Soc Sci Med* 249:112830
- Angrist J, Pischke J-S (2015) *Mastering metrics*. Princeton University Press, USA
- Avendano M, de Coulon A, Nafilyan V (2020) Does longer compulsory schooling affect mental health? evidence from a British reform. *J Public Econ* 183:104137
- Brunello G, Fabbri D, Fort M (2013) The causal effect of education on body mass: evidence from Europe. *J Labor Econ* 31(1):195–223
- Brunello G, Fort M, Schneeweiss N, Winter-Ebmer R (2016) The causal effect of education on health: what is the role of health behaviors? *Health Econ* 25(3):314–336
- Clark D, Royer H (2013) The effect of education on adult mortality and health: evidence from Britain. *Am Econ Rev* 103(6):2087–2120
- Crespo L, López-Naval B, Mira P (2014) Compulsory schooling, education, depression and memory: new evidence from sharelife. *Econ Edu Rev* 43:36–46
- Fonseca R, Michaud P, Zheng Y (2020) The effect of education on health: evidence from national compulsory schooling reforms. *Series* 11(1):83–103
- Gathmann C, Jürges H, Reinhold S (2015) Compulsory schooling reforms, education and mortality in twentieth century Europe. *Soc Sci Med* 127:74–82
- Hofmarcher T (2021) The effect of education on poverty: a european perspective. *Ec Edu Rev* 83:102124
- Mazzona F (2014) The long-lasting effects of education on old age: evidence of gender differences. *Soc Sci Med* 101:129–138
- Meghir C, Palme M, Simeonova E (2018) Education and mortality: evidence from a social experiment. *Am Econ J Appl Econ* 10(2):234–256
- Staiger D, Stock JH (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3):557–586
- Stephens M, Yang D (2014) Compulsory education and the benefits of schooling. *Am Econ Rev* 104(6):1777–1792
- Xue X, Cheng M, Zhang W (2021) Does education really improve health? a meta-analysis. *J Econ Surv* 35(1):71–105

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