DA Final Assignment - Smoke Free Laws and Lung Cancer

A. BOGNAR , D. GYEBNAR

4/18/2022

A technical note: the code and data can also be found in this Github repo.

1. Data

The investigation between smoking and cancer was one of the most iconic causal questions of the XX. century. As a result of the research most of public opinion now agrees on the harmful effects of smoking and more and more countries introduce new legislation to combat it. Of the many types of laws, here we try to investigate the efectiveness of banning smoking from closed public areas. Specifically, while most countries in the world have now adopted the basis Framework Convention on Tobacco Control by WHO we wanted to know if actually enforcing the idea indoor protection leads to lower number of lung cancer cases. To answer that we gathered the following data:

1. WHO has been publishing the report on the global tobacco epidemic every two years since 2008. It tries to list for every country the type of institutions where smoking is banned (hospitals, restaurants etc.). Though the data is hard to collect (details in the annex), it provides 12 years of information on the laws.
2. For lung cancer data we used statistics by Gapminder which has data between 1990 and 2016. We later found WHO's own sources which are much more detailed and run from 1990 to 2019 and even to 2020.

Outcome Variable

Our outcome is the rate of new cancer cases in 100 thousand people in a country in a year. We used incidence which is the total number of new cases as opposed to prevalence which is the number of active cases.

Causal variable

We used both a binary and a quantitative variable which we will explain in the relevant sections.

2. Binary Intervention

Under binary intervention we considered countries to be treated starting from the year when they banned smoking in:

1. Restaurants
2. Indoor offices
3. Schools (not including universities)
4. Hospitals
5. Public transportation

We decided on this criteria based the Smoke Free Map website which evaluates European countries based on the same WHO standard and tends to put countries in the best categories only if they have these 5 bans.

Treatment and Control Group

We considered countries to be treated if at some point they met this criteria and continued to do so for the rest of the years recorded in th data. If a country at some point reached the standard but then dropped at least 2 of the 5 bans, we excluded the country since its not clear which group it should belong to. We also had to consider the year of treatment: countries already treated by the first report in 2008 were also dropped since we don't know when they reached the standard. Countries reaching it in 2018 or later were also dropped since we think outcome only 2 years after the change is not enough to show an effect. Therefore, we tried two treatment groups: countries treated between 2010 and 14 and between 2010 and 16. The former group is smaller but has at least 6 years of data after the intervention. Here, we present the results for this first group, in the annex you will see that the results are very similar for the other group. As control countries we included all countries that never reached the standard which is the simplest option but may not be the best as we'll see. We ended up with 29 treated and 109 untreated countries (37-109 for the annex).

Different years of treatment

Since the year of treatment differs, for the treatment group we defined the year of intervention as t0 and compared years to that. For the untreated group, we defined t0 as the average year of treatments i.e. 2012 (2014 in the annex group). Also, note that while the WHO reports are published in odd years 2011,2013, they always refer to the country state in the previous even year. In odd years where no report referred, we considered the previous report's value.

A. Simple OLS

First, let's see the results in a basic cross-section regression for 3 different years.

Model 1 is for 20 years before intervention. Treatment group interventions are between 2010 and 2014. Control Group base year was chosen as 2012. **Model 2** is for year of intervention. Same groups as above. **Model 3** is for 6 years after intervention. Same groups as above.

We see that the new cancer rate of untreated countries increased over the period while treated countries seem to have lower rates by about 7 cases in 100 thousand. However, note that the difference was already there between the two groups way before the interventions and it didn't change much afterwards. In fact, a treated group's difference is not statistically different from 0 after the treatment. This is likely due to the low sample size but also if their is a treatment effect it doesn't seem to be so large that simple OLS could show it.

Table 1: Simple Cross Sectional Regression

	(1)	(2)	(3)
Untreated	20.4***	24.6***	27.7***
	(2.09)	(2.57)	(2.71)
Treated	-7.04*	-7.33	-7.51
	(3.13)	(4.05)	(4.40)
Observations	138	138	138
R2	0.020	0.014	0.013

B. Diff-in-diff

The second type of model compares the treated and untreated countries by taking an avg. of new cancer cases in the 6 years prior to the interventions and calculating the difference from that average to the single year, 6 years after the intervention.

The table and regression show the same result, a very small difference of 0.21 less new cancer in 100 thousand for treated countries and once again it's not statistically significant. Once again we see from the table that the two groups had a big difference which remained relatively stable across the years.

group	before	after	group_diff
untreated	23.2	27.72	-4.53
treated	15.9	20.21	-4.32
total	7.3	7.51	-0.21

Table 2: Diff-in-diff

	(1)
Untreated	4.53***
	(0.618)
Treated	-0.210
	(1.00)
Observations	138
R2	0.0002

C. Long differenc model

So maybe the changes come into effect over a long period of time? We can look at the difference in changes 6 years prior to the intervention compared to the changes seen 6 years after the intervention. Here, we actually see a small, not significant, increase in new cancer rate for treated countries' change.

	(1)
Untreated	-0.157 *
	(0.067)
Treated	0.146
	(0.194)
N	138
R2	0.006

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

3. Summary

In summary we couldn't show a significant treatment effect which may be due to several reasons:

1. The treatment effect may take even longer to realize: it can take an avg. 8 years for lung cancer to develop to point of detection. Thus our sample of 6 year after intervention may not be sufficient. Additionally, one possible mechanism of the treatment is that people may stop smoking due to the bans, this effect may take even longer to realize.
2. Our range of treated countries is very small. To offset this, we tried increasing treatment size and quantitative causal variable. The results were similar, you can find them in the annex.
3. Possible reporting issues/difficulty of defining intervention: the WHO standards seem to be very strict and not always transparent. We found that Hungary is not a treated country as WHO explicitly says in their report that they don't consider Hungary's ban on public transport smoking to be valid. Even though in effect it is not allowed to smoke on transport, some private taxis and buses have designated smoking spaces which excludes Hungary. Such decisions may skew the results. Some countries on either edge of meeting the criteria may actually be closer to other end in reality. One further approach for future research is to incorporate WHO's compliance statistic with the regulations.
4. Relevant control variables: In the annex we touch on possible controls. However, it is difficult to say whether rich countries with better healthcare are more likely to adopt these regulations or if they are also more in favor of personal freedom, e.g. Austria and Germany both allow smoking in Hospitals based on legislation but likely not in practice.

Recommendation

Though we couldn't find an effect, due to the above reasons there may still be one. Therefore, to fairly evaluate such a legislation we have to consider the possible harm it introduces. Some argue that the ban discourages business, others say it removes personal freedom. However, in our opinion any small contribution it may make towards decreasing lung cancer rates is more important than these possible issues. That being said, WHO outlines many other possible laws to combat smoking such as taxation, public education. So until more data is available about smoking bans, it may make sense to do similar tests on the effect of such other measures as we may find on that has a significantly bigger impact.

Annex

1. Data Cleaning decision in details

The code for the full data cleaning steps can be found in the data cleaning file.

WHO reports

As mentioned WHO reports on smoking legislation refer to years 2008, 2010, 2012, 2014, 2016, 2018 and 2020. Only the 2008 report includes country level statistics within the pdf file. Th 2020 file has an excel annex on the website. In all other cases the Annex tables were stored in a previous website which was archived in 2018. We managed to get the Excels for all but 2015, 2017 using Waybackmachine.

The reports can be found via these links:

2008. 2010. 2012. 2018. 2020.

For 2014 and 2016, we checked which countries saw a change between 2012 and 2018 and looked up these countries manually in the individual country level reports which could still be downloaded. See 2014 and 2016.

One more interesting note on WHO categorization: designated smoking places are not considered to be adequate measures.

2. Four year intervention period OLS

As you can see the models with more treated countries included show bigger difference between untreated and treated groups and the difference increases more sharply after the intervention than in the original models. However, the results are still not significant at 95% confidence.

Table 3: Simple Cross Sectional Regression

	(1)	(2)	(3)
Untreated	21.0***	24.8***	27.7***
	(2.16)	(2.58)	(2.71)
Treated	-6.75*	-6.26	-7.65
	(3.11)	(3.98)	(4.22)
Observations	146	146	146
R2	0.020	0.012	0.016

Diff-in-diff

group	before	after	group_diff
untreated	23.63	27.72	-4.09
treated	17.45	20.07	-2.62
total	6.18	7.65	-1.47

Table 4: Diff-in-diff

	(1)
Untreated	4.09***
	(0.570)
Treated	-1.47
	(0.744)
Observations	146
R2	0.014

Long diff

	(1)
Untreated	-0.187 **
	(0.063)
Treated	0.480
	(0.346)
N	146
R2	0.030

*** p < 0.001; ** p < 0.01; * p < 0.05.

3. Quantitative Causal Variable

We mentioned that for the binary causal variable we only looked at 5 institutional ban categories. The full dataset contains 4 more type of bans for:

1. Universities
2. Cafes, pubs, bars
3. Government facilities
4. All other public places

So in total each country in each year can have between 0 and 9 types of smoking bans. For the quantitative variable, we used this number between 0 and 9.

A. Simple OLS

Model 1 shows association between number of places with banned indoor smoking and lung cancer rate of country in 2008. **Model 2** is same for 2016. **Model 3** is same for 2020.

Results are similar overall: more legislation may have some effect, but we can't say based on these results.

Table 5: Simple Cross Sectional Regression

	(1)	(2)	(3)
Intercept	21.1***	23.7***	24.6***
	(2.48)	(3.13)	(3.76)
Number of banned places 2008	-0.094		
	(0.608)		
Number of banned places 2016		-0.299	
		(0.523)	
Number of banned places 2020			0.023
			(0.576)
Observations	174	174	174
R2	0.0001	0.002	8.18e-6

B. Fixed Effect

The model shows that in years where number of places where smoking was banned was increased, basically no change is seen in number of cancer cases. The year of changs however does matter, overall number of cases increases over time as we saw previously.

	(1)
Number of places	-0.002
	(0.070)
Year	0.392 ***
	(0.044)
N	2262
R2	0.987

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

C. First diffrence, with 2 and 6 year lags.

With first difference models and lags, we can investigate how the effect of changes accumulate over the years after the change. In the year of the change we see a slight increase, in years 1, 3 and 5 afterwards a small decrease in years 2 and 4 a small increase. Overall, none of the individual or the cumulative effect is significant, that is why even more years after the changes, at least until 10 years after the chang could be useful.

	(1)	(2)	(3)
Intercept	0.316 ***	0.331 ***	0.404 ***
	(0.045)	(0.053)	(0.071)
Cont Diff of number of bans	0.026	0.043	0.109
	(0.050)	(0.075)	(0.149)
1st Lag of Diff number of bans		-0.025	-0.027
		(0.019)	(0.027)
2nd Lag of Diff number of bans		0.012	0.052
		(0.023)	(0.053)
3rd Lag of Diff number of bans			-0.115
			(0.088)
4th Lag of Diff number of bans			0.034
			(0.037)
5th Lag of Diff number of bans			-0.035
			(0.021)
N	2088	1740	1218
R2	0.000	0.001	0.004

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

D. Controls

As mentioned, we checked a few possible controls from World Bank Indicators:

1. GDP per capita (constant 2015 US$): wealth of a country can effect the types of restrictions introduced and it heavily effects the cancer detection of the country.
2. Current health expenditure per capita (current US$): health consciousness countries are more likely to legislate smoking and will have better overall prevention and treatment.
3. Life expectancy at birth, total (years): since we are not looking at death cases due to cancer and as we saw our time period is way too small for life expectancy to be effected by the change in laws, we can include this control as another proxy for prevention and screening of a country.

These controls are relevant even for first difference model as their effects can change over the years.

Some relevant controls that are hard to include: cultural significance of smoking, perception of health consciousness vs individual freedom in a country. Here, only life expectancy proved to be relevant.

	(1)
Intercept	-3.349 ***
	(0.577)
Cont Diff of number of bans	0.188
	(0.112)
1st Lag of Diff number of bans	-0.058 **
	(0.021)
2nd Lag of Diff number of bans	0.011
	(0.055)
3rd Lag of Diff number of bans	-0.062 ***
	(0.018)
4th Lag of Diff number of bans	0.007
	(0.038)
5th Lag of Diff number of bans	-0.058 ***
	(0.017)
GDP Per capita	-0.000
	(0.000)
Health Expenditure of government	-0.000
	(0.000)
Life Expectancy	0.058 ***
	(0.009)
N	1003
R2	0.036

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.