Ending the Tobacco era

Analysis of smoking trends and policy impact in OECD countries

Valentin Lacombe (5 CFU)
Flavia Petruso (8 CFU)

Content

- Introduction
- Dataset presentation
- Stakeholders and research questions
- Data preprocessing
- Analysis
- Conclusions

Ending the Tobacco era

m **Monitor** tobacco use and prevention policies

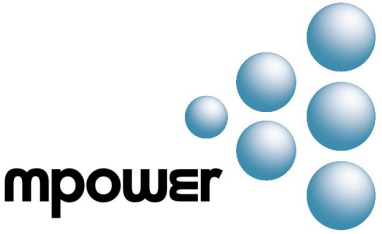
p **Protect** people from tobacco smoke

o **Offer** help to quit tobacco use

w **Warn** about the dangers of tobacco

e **Enforce** bans on tobacco advertising, promotion and sponsorship

r **Raise** taxes on tobacco



Ending the Tobacco era

m **Monitor** tobacco use and prevention policies

p **Protect** people from tobacco smoke

o **Offer** help to quit tobacco use

w **Warn** about the dangers of tobacco

e **Enforce** bans on tobacco advertising, promotion and sponsorship

r **Raise** taxes on tobacco

Introduced by the **World Health Organization** (WHO) in 2008, to assist countries in implementing effective measures to **combat smoking and tobacco use**.

mpower

Dataset

WHO dataset on tobacco for the 38 OECD countries

- **Tobacco smoking prevalence** (total, male, female), 7 years: %
- **Anti-smoking measures**, 6 years:
  - **Factor variables**: MPOWER measures + Campaigns
  - **Continuous variables**: affordability

OECD database

- **GDP per capita** (constant PPP): $
- **Education level**: % of 25-64 population with tertiary education

UNDP database

- **Human Development Index** (HDI): 0-1

Stakeholders and goals

Analysis on behalf of the OECD

BETTER POLICIES FOR BETTER LIVES

Policy makers of OECD nations

1. **Quantify the effect of national policies** on tobacco smoking, to orient policy makers towards the most effective actions
2. **Characterise gender differences** in the factors affecting smoking prevalence
3. Identify OECD countries less likely to **reach the SDG3** (30% decrease from 2010 to 2025) under BAU (business as usual)

General Public

Learn recent trends on smoking in the «**Health at a Glance 2023**» report

Data preprocessing

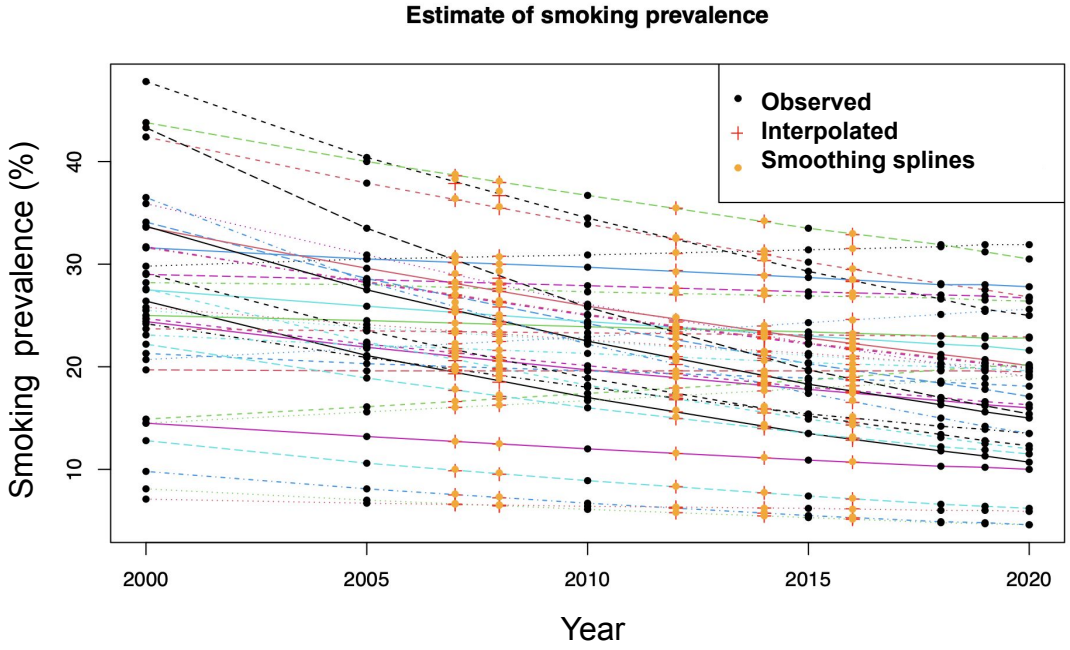
Mismatch between smoking prevalence data and other covariates

Data preprocessing

Mismatch between smoking prevalence data and other covariates

Interpolation using cubic B-splines to estimate the values for the missing years

Smoothing with penalised splines using cubic B-splines with lambda=100

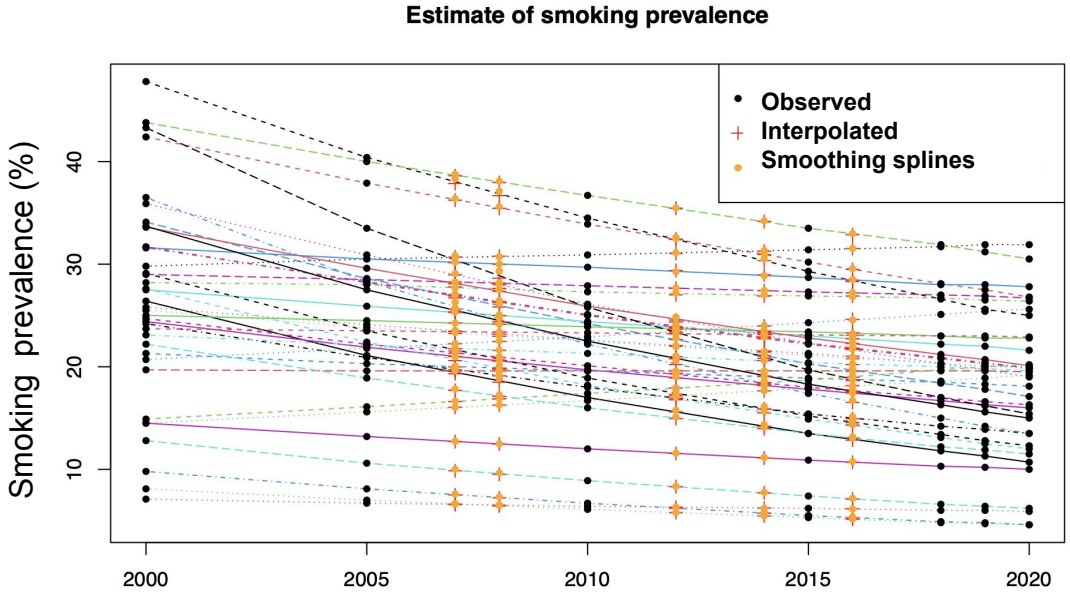


Data preprocessing

Mismatch between smoking prevalence data and other covariates

Interpolation using cubic B-splines to estimate the values for the missing years

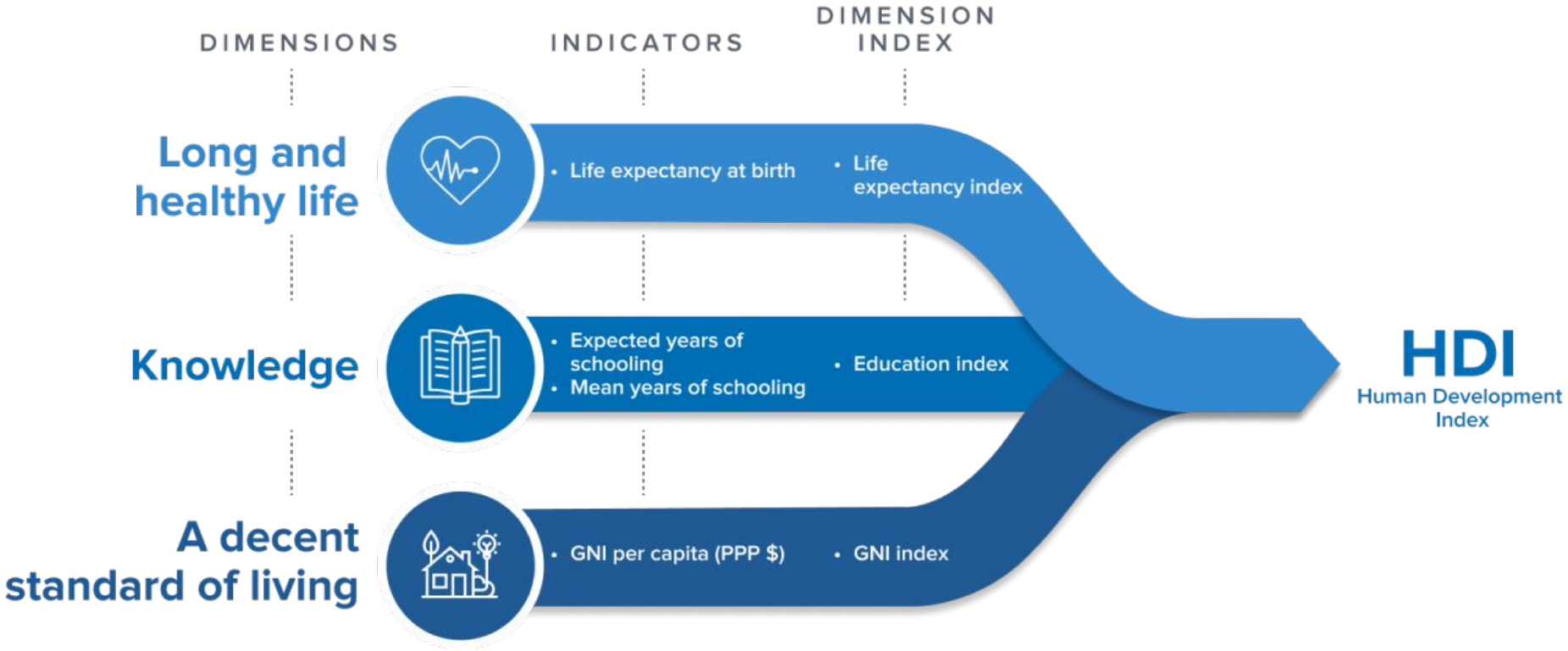
Smoothing with penalised splines using cubic B-splines with lambda=100



Filling of MPOWER NaNs with average values of closest years from the same country

Analysis of prevalence

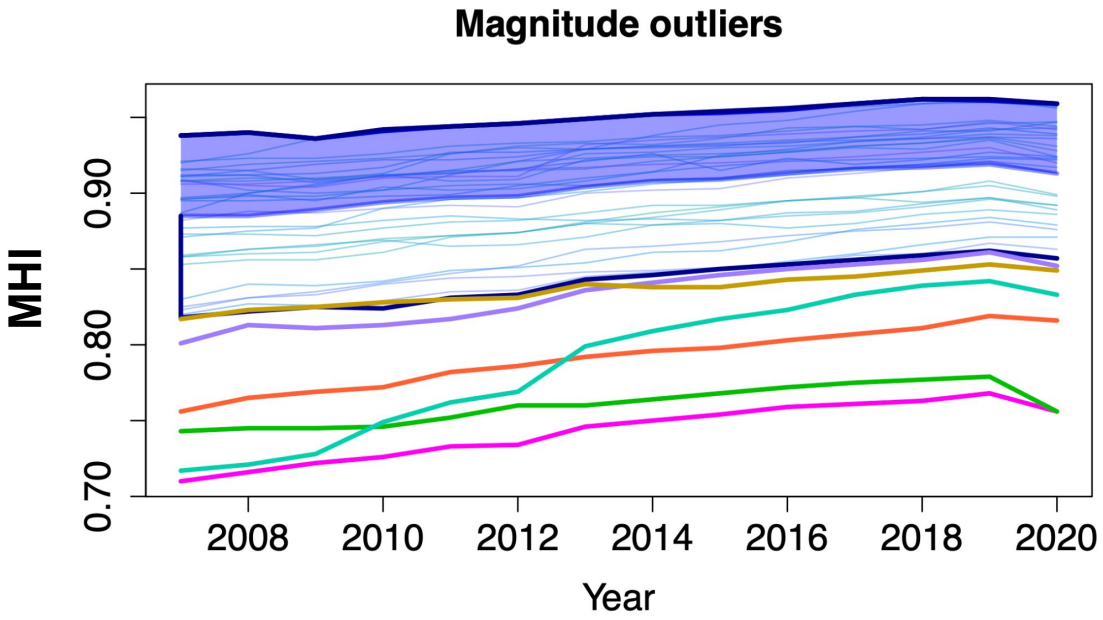
Functional analysis of HDI index



From https://hdr.undp.org/data-center/human-development-index#/indicies/HDI

Functional analysis of HDI index

Computation of Modified Hypograph Index on yearly data of HDI (2007-2020)

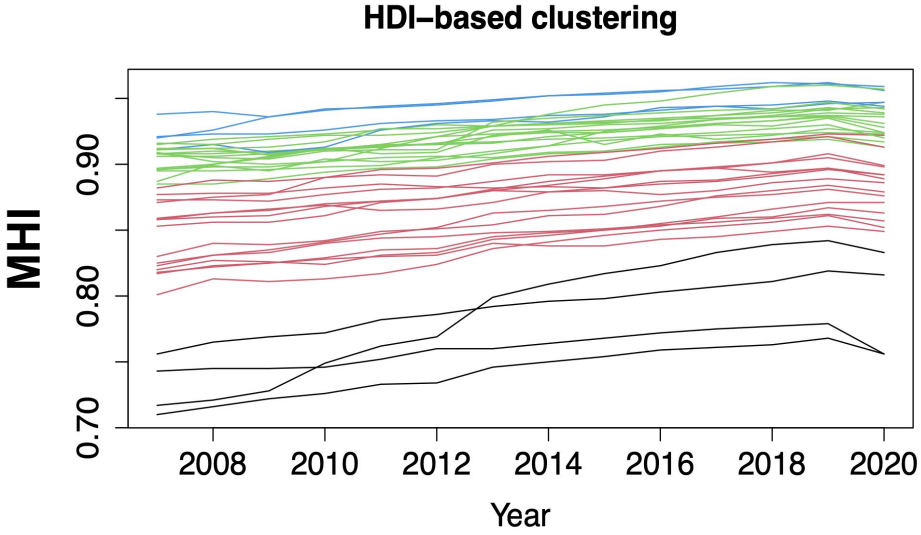


Overall: steady increase of HDI index from all countries.

Outliers: Colombia, Chile, Mexico, Turkey

Functional analysis of HDI index

Clustering based on of Modified Hypograph Index

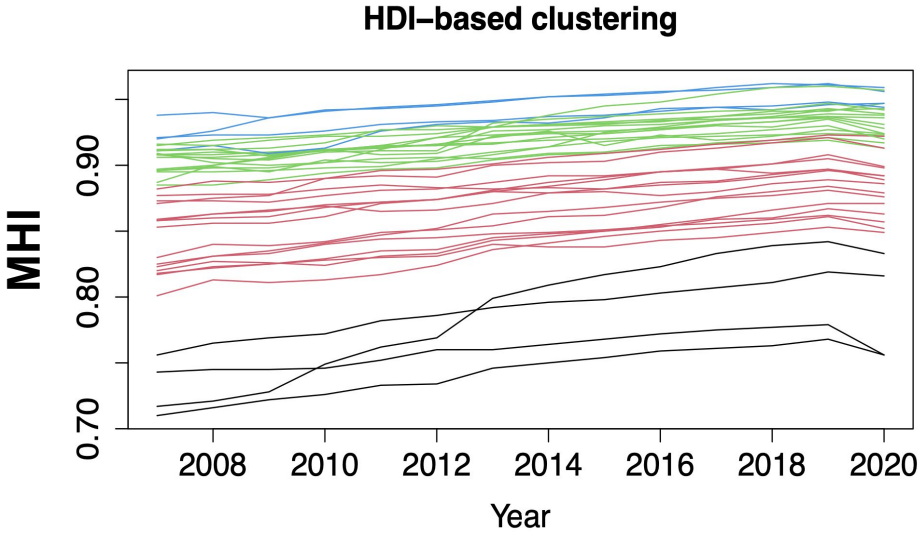


Identification of **4 clusters** based on the 10th, 50th, 90th percentile.

Cluster 1 corresponds to the previous outliers.

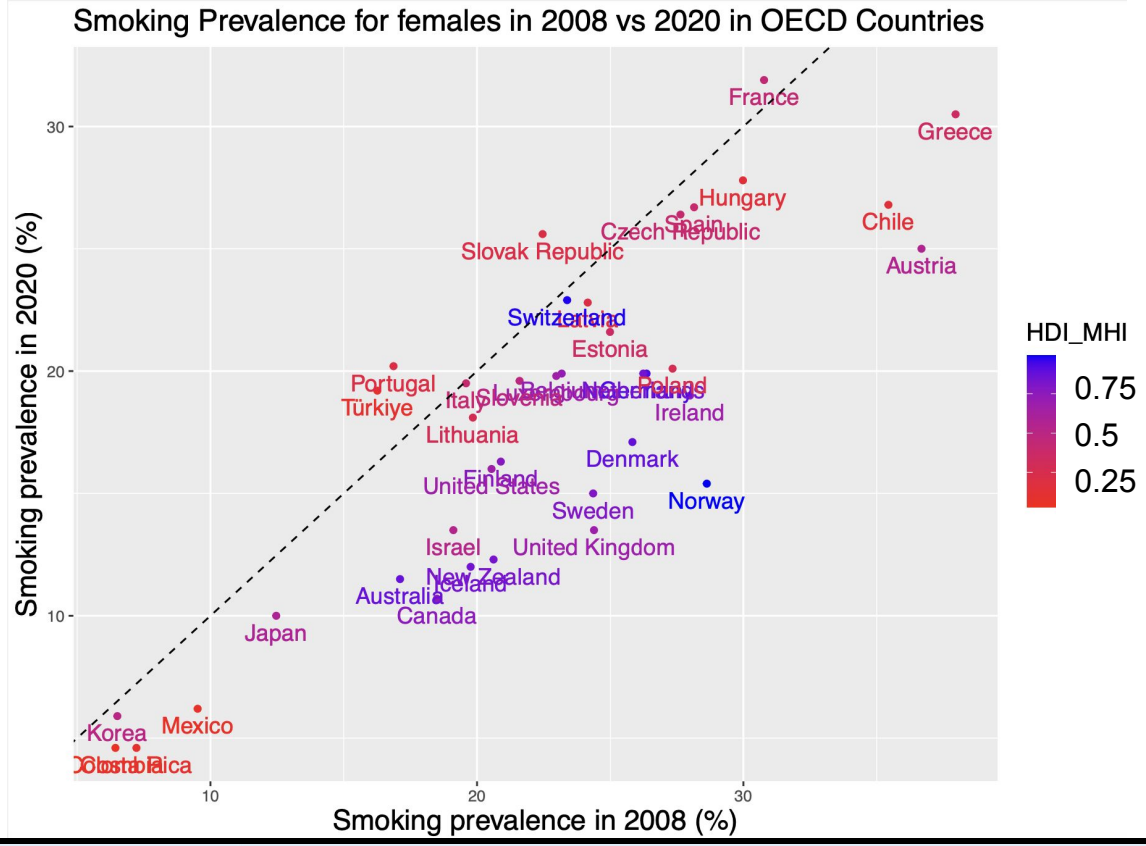
Functional analysis of HDI index

Clustering based on of Modified Hypograph Index



Cluster 1	Cluster 2	Cluster 3	Cluster 4
Colombia	Chile	Australia	Switzerland
Costa Rica	Czeck Republic	Austria	Germany
Mexico	Spain	Belgium	Denmark
Turkey	Estonia	Canada	Norway
	France	Finland	
	Greece	United Kingdom	
	Hungary	Ireland	
	Italy	Iceland	
	Korea	Israel	
	Lithuania	Japan	
	Latvia	Luxembourg	
	Poland	Netherlands	
	Portugal	New Zealand	
	Slovak Republic	Sweden	
	Slovenia	United states	

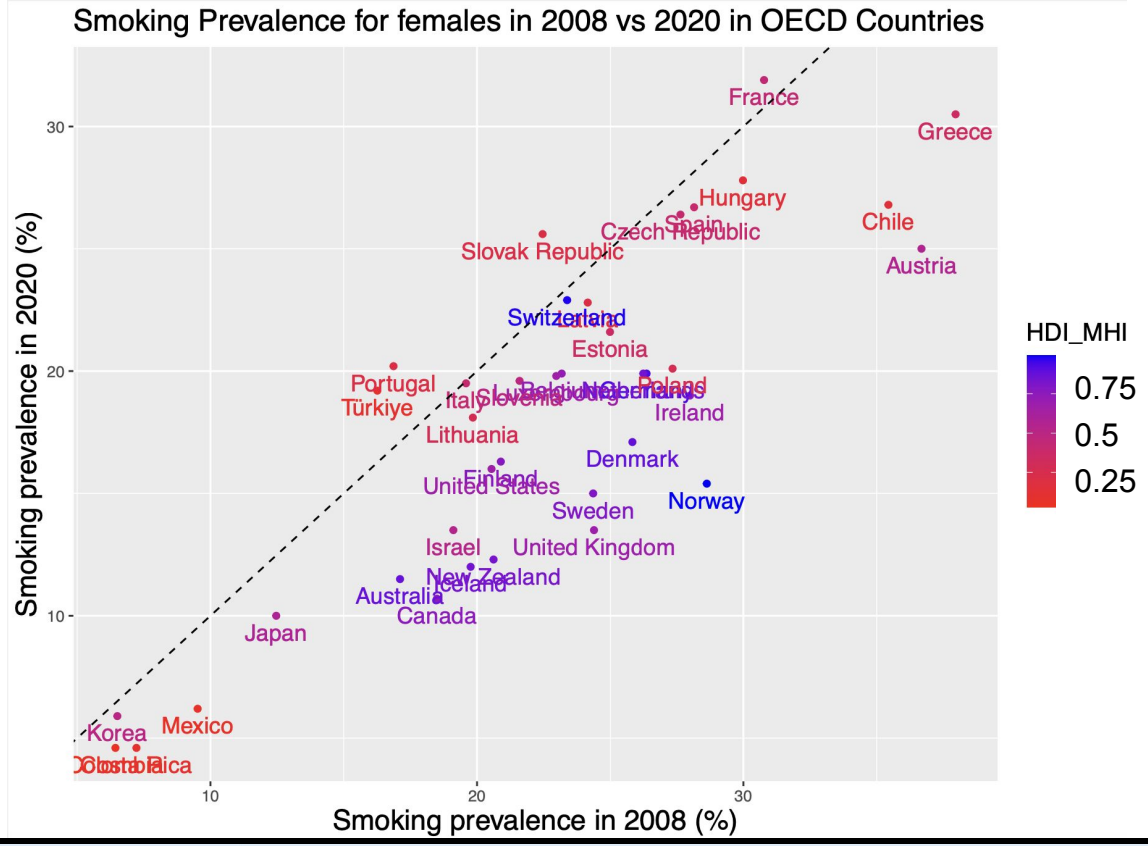
Female smoking prevalence changes 2008-2020



Overall trend: decrease between 2008 and 2020.

Exceptions: France, Slovak Republic, Portugal, Turkey (average/average-low HDI).

Female smoking prevalence changes 2008-2020



Overall trend: decrease between 2008 and 2020.
Exceptions: France, Slovak Republic, Portugal, Turkey (average/average-low HDI).

Overall trend: lower prevalence in more developed countries
Exceptions: Latin American Countries, Korea as well (despite high male prevalence). Why?

Female smoking prevalence changes 2008 -2020

BMC Women's Health

BMC Womens Health. 2014; 14: 156.

PMCID: PMC4319222

Published online 2014 Dec 12. doi: 10.1186/s12905-014-0156-z

PMID: 25495192

Does South Korea have hidden female smokers: discrepancies in smoking rates between self-reports and urinary cotinine level

Myung Bae Park, Chun-Bae Kim,[✉] Eun Woo Nam, and Kyeong Soo Hong

Are female and male prevalence different during the period under consideration?

Univariate permutational test for paired data

$$H_0 : \mu(\text{male}_i - \text{female}_i) = 0 \quad vs \quad H_1 : \mu(\text{male}_i - \text{female}_i) \neq 0,$$

$$\text{Test statistic}_i : |mean(\text{male}_i - \text{female}_i)|,$$

$$\text{Permutation scheme}_i\text{: permutation of the intra-country differences}_i,$$

$$\text{for } i \text{ in } \{2008, 2010, 2012, 2014, 2016, 2018, 2020\}.$$

Are female and male prevalence different during the period under consideration?

Univariate permutational test for paired data

$$H_0 : \mu(\text{male}_i - \text{female}_i) = 0 \quad vs \quad H_1 : \mu(\text{male}_i - \text{female}_i) \neq 0,$$

$$\text{Test statistic}_i : |mean(\text{male}_i - \text{female}_i)|,$$

$$\text{Permutation scheme}_i\text{: permutation of the intra-country differences}_i,$$

$$\text{for } i \text{ in } \{2008, 2010, 2012, 2014, 2016, 2018, 2020\}.$$

Result with alpha=0.05 and 1000 replicas: p-value=0 for each year. We **reject H0**

The prevalence of female smoking is significantly different from the one of males, also in most recent years

Is the difference in prevalence decreasing over years?

Bootstrap confidence intervals for the differences

We operate on the **third quantile of the difference distribution**, as it is may be more useful for our stakeholder.
alpha=0.05, replicas=1000

Is the difference in prevalence decreasing over years?

Bootstrap confidence intervals for the differences

We operate on the **third quantile of the difference distribution**, as it is may be more useful for our stakeholder.
alpha=0.05, replicas=1000

Year	Lower	75^{th} percentile	Upper
2008	15.60	17.29	24.39
2010	14.65	16.08	22.38
2012	8.42	14.55	19.85
2014	10.49	13.20	17.48
2016	7.68	11.87	15.52
2018	3.23	10.63	13.75
2020	4.05	9.73	12.35

Is the difference in prevalence decreasing over years?

Bootstrap confidence intervals for the differences

We operate on the **third quantile of the difference distribution**, as it is may be more useful for our stakeholder.
alpha=0.05, replicas=1000

Differences are decreasing, the **gender gap is closing**

Year	Lower	75^{th} percentile	Upper
2008	15.60	17.29	24.39
2010	14.65	16.08	22.38
2012	8.42	14.55	19.85
2014	10.49	13.20	17.48
2016	7.68	11.87	15.52
2018	3.23	10.63	13.75
2020	4.05	9.73	12.35

Analysis anti-smoking policies

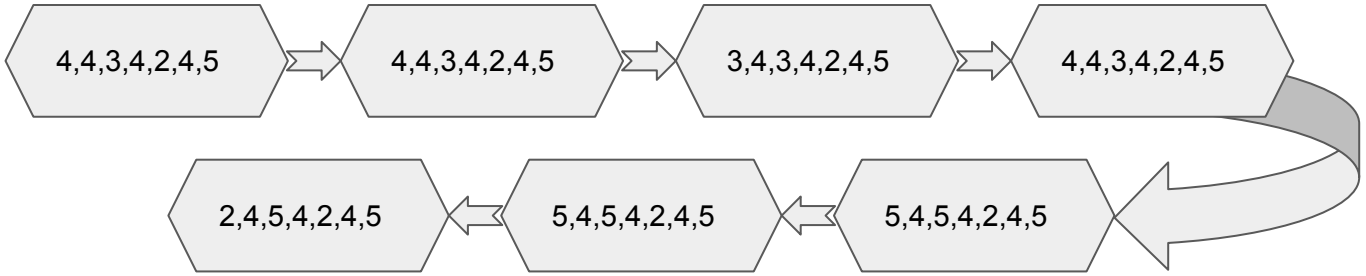
Analysis of MPOWER policies

Example data from Italy

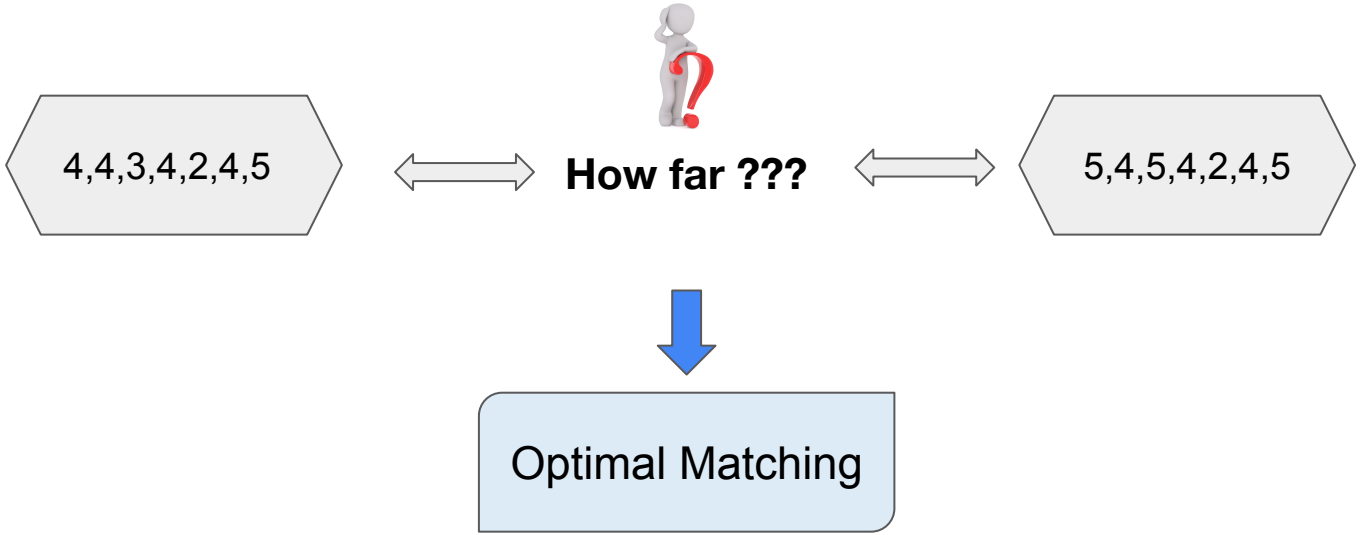
Country	Year	Campaign	Help	Warn	Bans	Protect	Monitor	Taxes
Italy	2008	4	4	3	4	2	4	5
Italy	2010	4	4	3	4	2	4	5
Italy	2012	3	4	3	4	2	4	5
Italy	2014	4	4	3	4	2	4	5
Italy	2016	5	4	5	4	2	4	5
Italy	2018	5	4	5	4	2	4	5
Italy	2020	2	4	5	4	2	4	5

Analysis of MPOWER policies

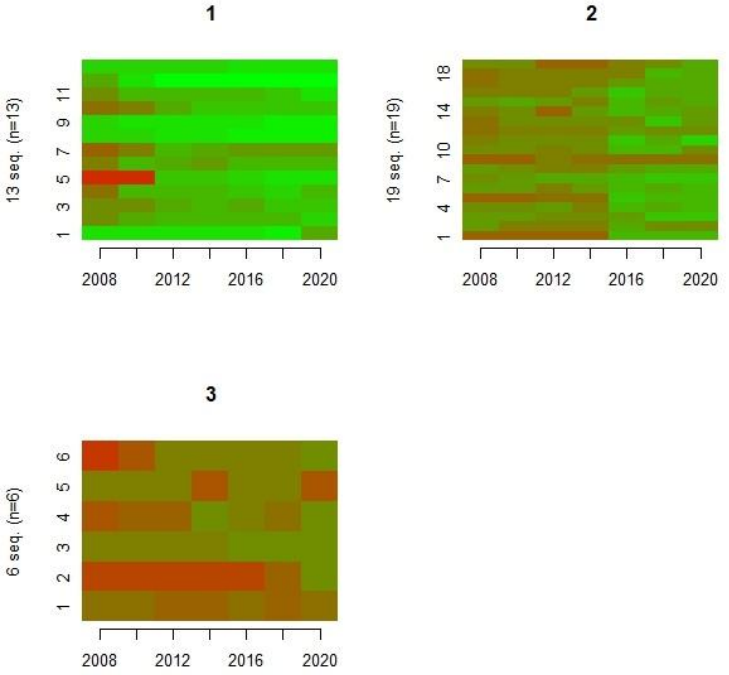
Country	Year	Campaign	Help	Warn	Bans	Protect	Monitor	Taxes
Italy	2008	4	4	3	4	2	4	5
Italy	2010	4	4	3	4	2	4	5
Italy	2012	3	4	3	4	2	4	5
Italy	2014	4	4	3	4	2	4	5
Italy	2016	5	4	5	4	2	4	5
Italy	2018	5	4	5	4	2	4	5
Italy	2020	2	4	5	4	2	4	5

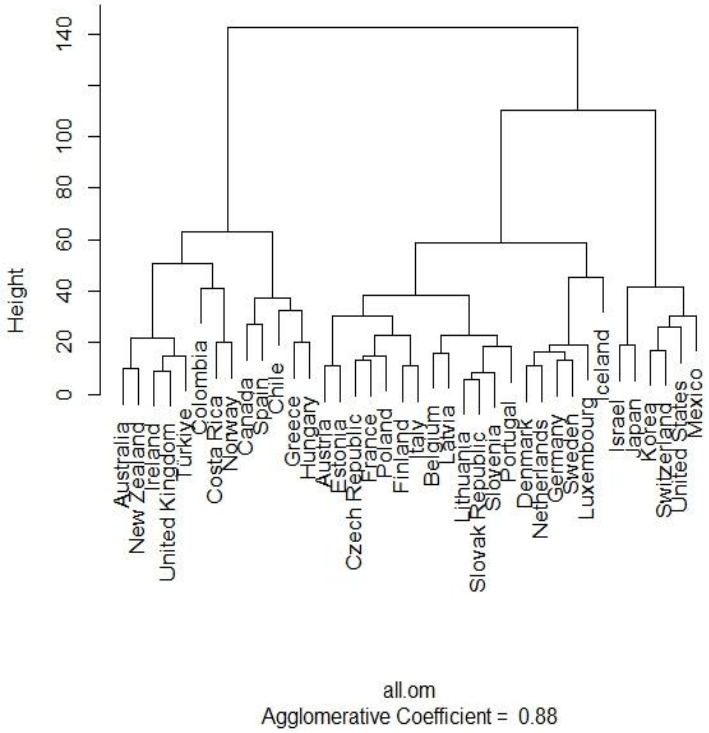


Distance between states



Hierarchical clustering





Distribution of the ANOVA permutational test

$X_i \equiv$ prevalence change between 2008/2020 (%) for a contry in cluster i,

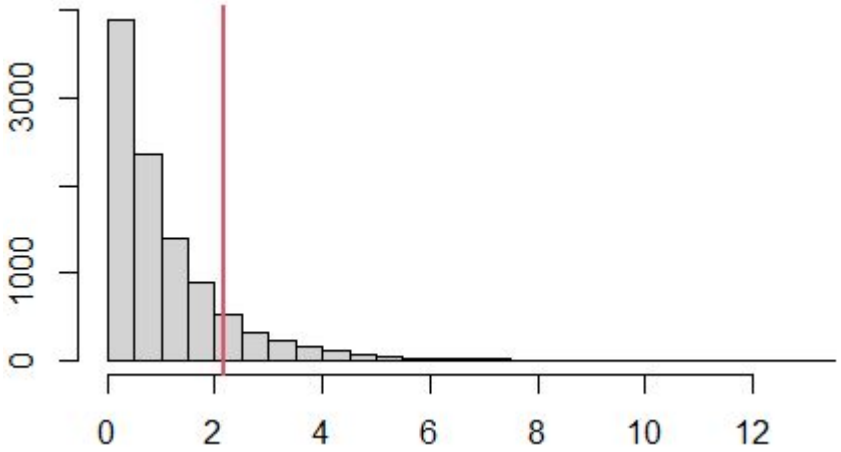
$H_0 : X_1 \overset{d}{=} X_2 \overset{d}{=} X_3 \quad vs \quad H_1 : \exists i, j s.t X_i \overset{d}{=} X_j$,

Test statistic$_i$: $Fstatistics$,

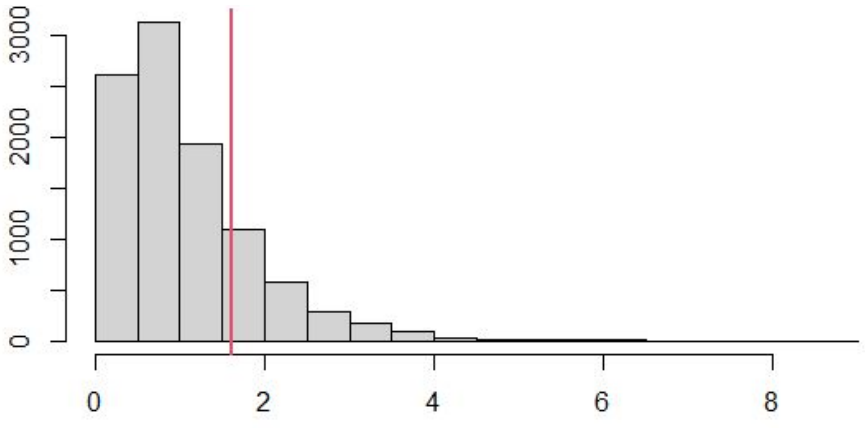
Permutation scheme$_i$: permutation of the cluster$_i$,

for $i \in$ countries .

Distribution of the ANOVA permutational test



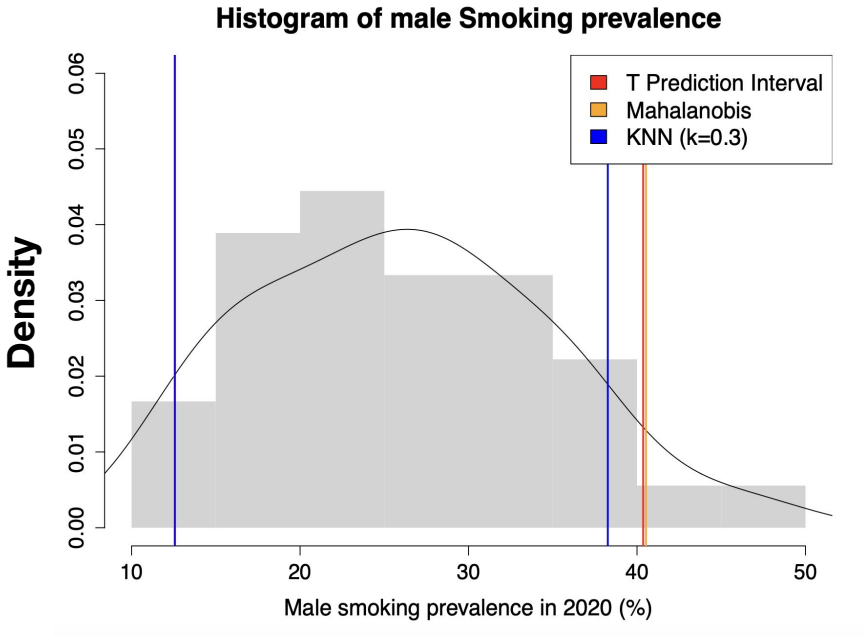
3 clusters case: p-value = 0.128

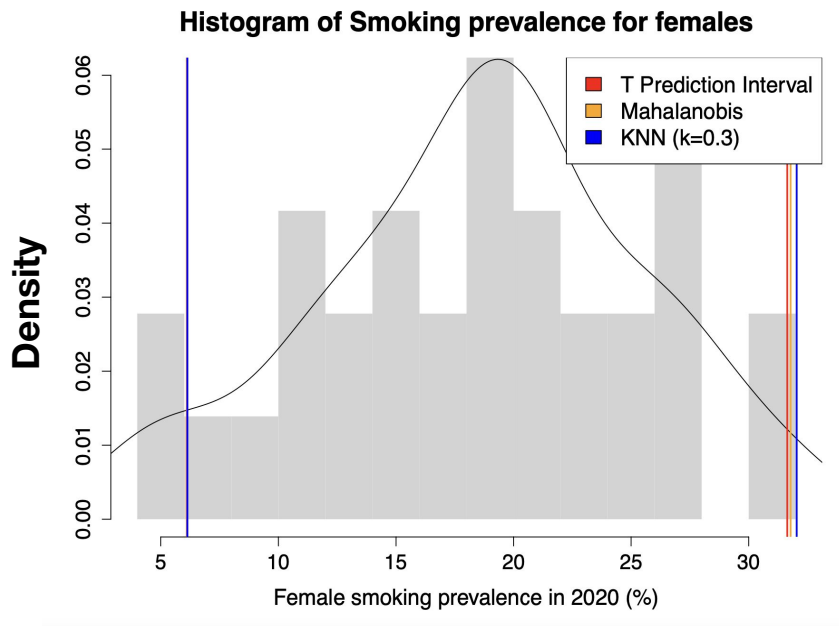


5 clusters case: p-value = 0.1983

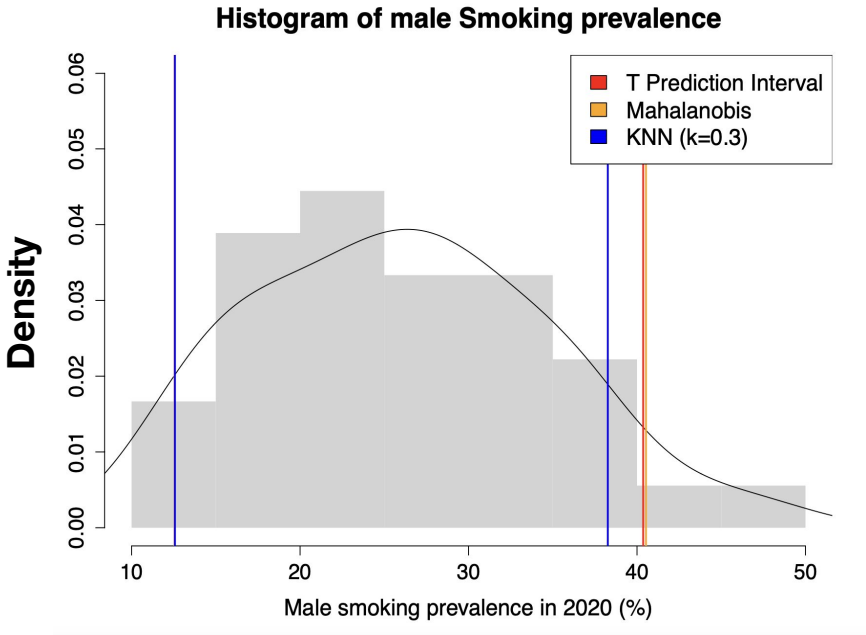
Prediction of prevalence

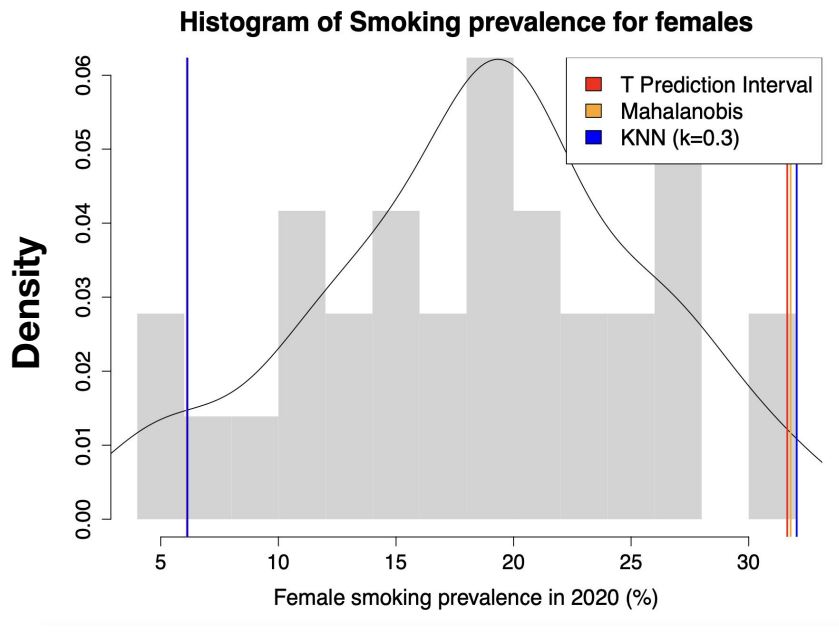
Conformal prediction intervals on smoking prevalence





Conformal prediction intervals on smoking prevalence





Quite large and noninformative… can we do better?

Generalized additive model for female smoking prevalence

Semiparametric model for smoking prevalence (after reduction with pemutational tests)

$$
\begin{aligned}
\text{Prevalence\_females}_i = \beta_0 &+ \beta_{\text{Year}} \cdot \text{Year}_i \\
&+ \sum_{j=1}^{J} \beta_{\text{Country}_j} \cdot \mathbb{1}_{(\text{Country}_i = \text{Country}_j)} \\
&+ f(\text{HDI}_i) + \beta_{\text{Affordability}} \cdot \text{Affordability}_i, + \epsilon_i \\
\textit{with} \quad & \epsilon_i \sim N(0, \sigma^2), \\
& i \in \{1, ..., 222\}, \\
& J = 38.
\end{aligned}
$$

Generalized additive model for female smoking prevalence

Semiparametric model for smoking prevalence (after reduction with pemutational tests)

$$\begin{aligned}\text{Prevalence\_females}_i &= \beta_0 + \beta_{\text{Year}} \cdot \text{Year}_i \\ &\quad + \sum_{j=1}^{J} \beta_{\text{Country}_j} \cdot \mathbb{1}_{(\text{Country}_i = \text{Country}_j)} \\ &\quad + f(\text{HDI}_i) + \beta_{\text{Affordability}} \cdot \text{Affordability}_i, + \epsilon_i \\ \textit{with} \quad \epsilon_i &\sim N(0, \sigma^2), \\ i &\in \{1, ..., 222\}, \\ J &= 38.\end{aligned}$$

cubic b-splines to increase flexibility

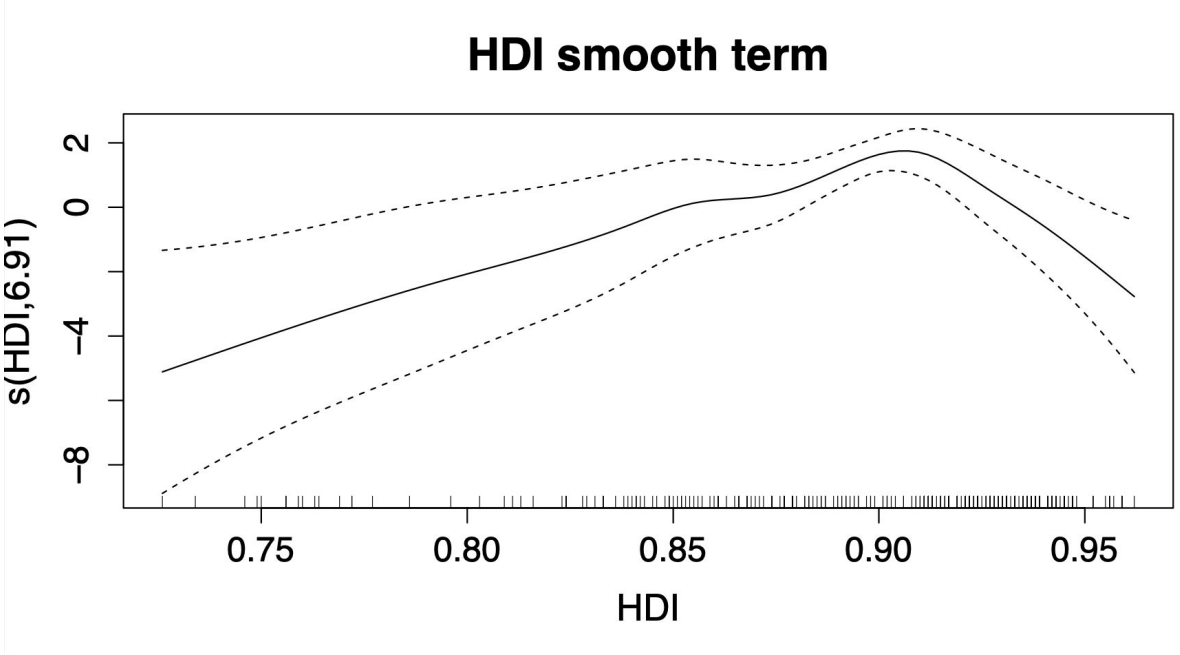
Generalized additive model for female smoking prevalence

Semiparametric model for smoking prevalence (after reduction with pemutational tests)

$$\begin{aligned}\text{Prevalence\_females}_i &= \beta_0 + \beta_{\text{Year}} \cdot \text{Year}_i \\ &\quad + \sum_{j=1}^{J} \beta_{\text{Country}_j} \cdot \mathbb{1}_{(\text{Country}_i = \text{Country}_j)} \\ &\quad + f(\text{HDI}_i) + \beta_{\text{Affordability}} \cdot \text{Affordability}_i + \epsilon_i \\ \textit{with} \quad & \epsilon_i \sim N(0, \sigma^2), \\ & i \in \{1, ..., 222\}, \\ & J = 38.\end{aligned}$$

Percentage of GDP required to buy 2000 cigarette packs

Generalized additive model for female smoking prevalence



Non-monotonic trend for HDI.

Potential “**emancipation effect**”?

Generalized additive model for female smoking prevalence

Table for linear coefficients (excluded Country)

Variable	Coefficient	p-value
Intercept	601.49673	7.77e-09
Year	-0.29076	2.33e-08
Affordability	-0.69530	0.028110

Negative impact of year on prevalence

Negative impact of affordability

Generalized additive model for female smoking prevalence

Table for linear coefficients (excluded Country)

Variable	Coefficient	p-value
Intercept	601.49673	7.77e-09
Year	-0.29076	2.33e-08
Affordability	-0.69530	0.028110

Negative impact of year on prevalence

Negative impact of affordability

Raising taxes to increase the price of tobacco products is the single most effective tobacco control measure

From World Health Organization. (2021).

Can OECD countries reach SDG3 for females?

30% relative reduction of smoking prevalence from 2010 to 2025

We selected 3 countries from different HDI groups and we assessed their prediction for female smoking prevalence in 2025 under “business as usual”. Will they reach their 2025 target?

Can OECD countries reach SDG3 for females?

30% relative reduction of smoking prevalence from 2010 to 2025

Country	Est. HDI	Est. Affordability	Lower	Pred	Upper	Target
France	0.902	3.20	27.67214	28.32308	29.32093	21.63
Turkey	0.850	3.80	15.60299	16.41899	17.38266	11.69
Norway	0.966	2.22	14.57067	15.50573	16.28126	18.06

Can OECD countries reach SDG3 for females?

30% relative reduction of smoking prevalence from 2010 to 2025

Country	Est. HDI	Est. Affordability	Lower	Pred	Upper	Target
France	0.902	3.20	27.67214	28.32308	29.32093	21.63
Turkey	0.850	3.80	15.60299	16.41899	17.38266	11.69
Norway	0.966	2.22	14.57067	15.50573	16.28126	18.06

Neither France nor Turkey will reach their target. **More actions needed**

Conclusions

- **Smoking prevalence:** overall decreasing, but still a gap base on HDI. **Gender differences are decreasing,** extra-attention needed in “developing” countries. Better measures needed in Asian countries for females.

- **Anti-smoking measures:** MPOWER score overall increasing, but difficult to assess their link with prevalence clearly.

- **Prediction:** effect of affordability on smoking prevalence. Possible use in identifying countries needing more actions to reach SDG3.

References

- **WHO database**
  https://apps.who.int/gho/data/node.main.TOBMPOWER?lang=en

- **OECD database**
  https://stats.oecd.org/

- **UNHN Dataset on HDI Index**
  https://hdr.undp.org/data-center/human-development-index#/indicies/HDI

- Park, M. B., Kim, C. B., Nam, E. W., \& Hong, K. S. (2014). Does South Korea have hidden female smokers: discrepancies in smoking rates between self-reports and urinary cotinine level. BMC women's health, 14(1), 1-8.
  https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4319222/

- World Health Organization. (2021). WHO report on the global tobacco epidemic, 2021: addressing new and emerging products. World Health Organization.

Anova model coefficients:

```
(Intercept) data_variation$cluster_32 data_variation$cluster_33
  -0.2965             0.0971                    0.0602
```

```
(Intercept) data_variation$cluster_52 data_variation$cluster_53 data_variation$cluster_54 data_variation$cluster_55
 -0.3310                    0.1556                    0.0896                    0.1099                    0.0947
```

Bagplots on prevalence

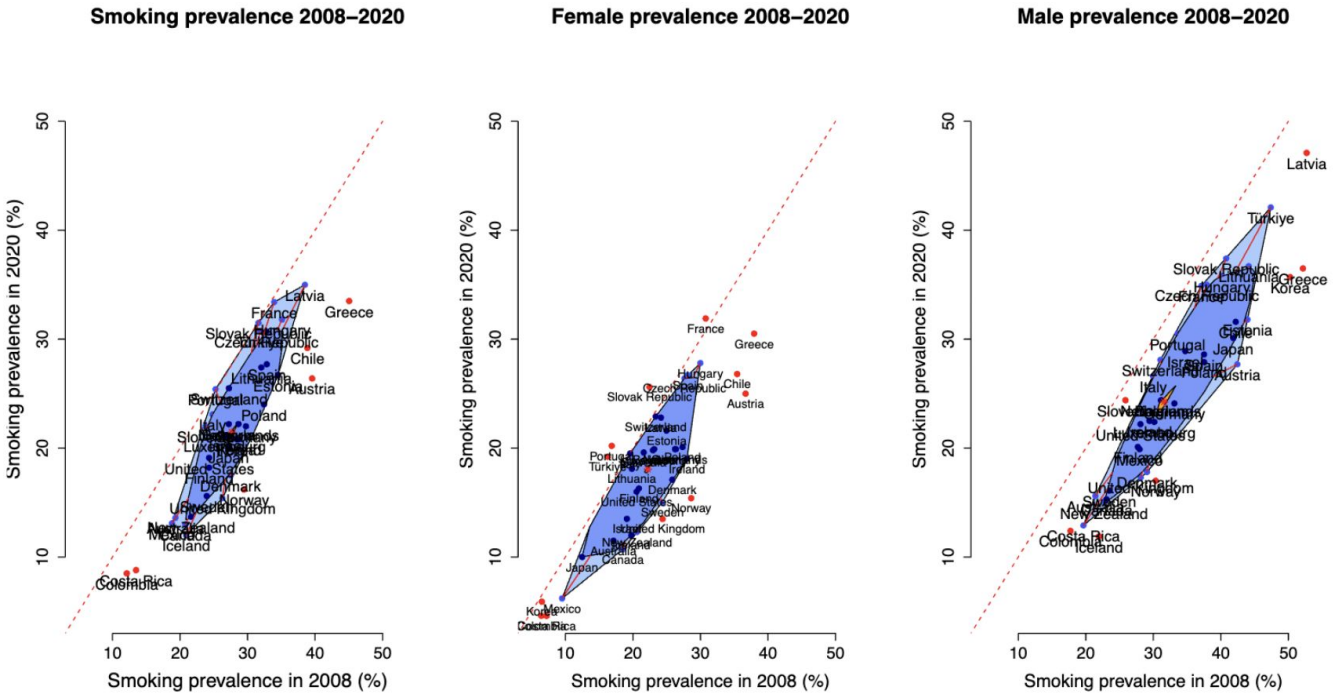
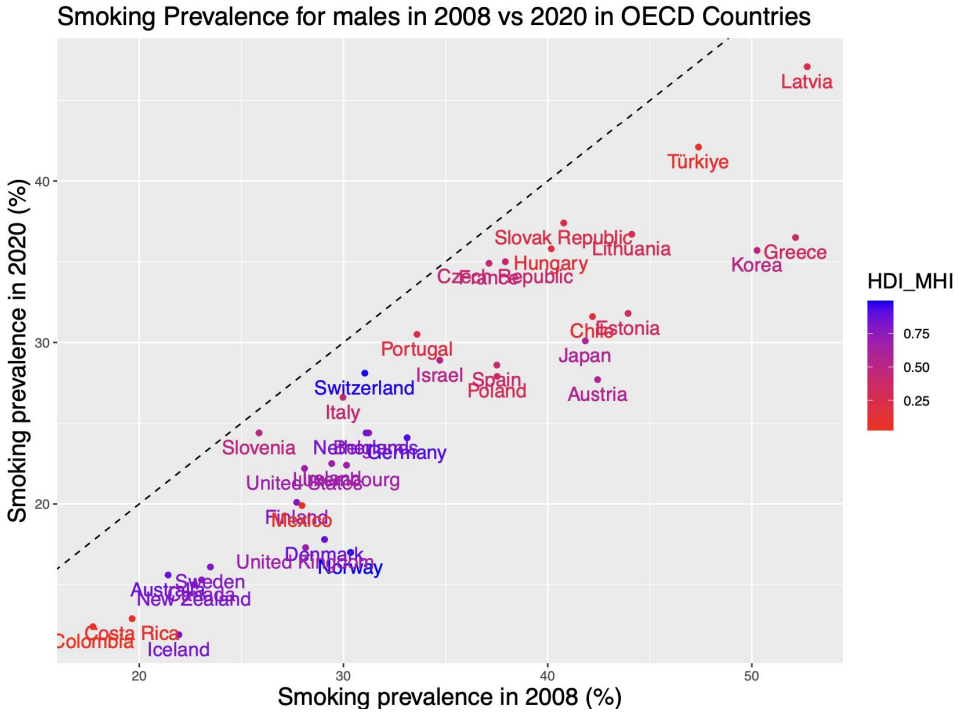


Figure 4: Bagplot of smoking prevalence in 2008 and 2020, factor=1.5. The red dashed line indicates no changes.

Smoking prevalence changes 2008 -2020



Overall trend: decrease between 2008 and 2020. Richer countries overall experiencing lower prevalence

Exceptions: Latin American countries

Test on Functional Spearman correlation coefficient for male and female prevalence

Estimated **rho=0.54**, indicating **moderate positive correlation**

$$H_0 : \rho_{MHI}(\text{male}, \text{female}) = 0 \quad vs \quad H_1 : \rho_{MHI}(\text{male}, \text{female}) \neq 0,$$

$$\text{Test statistic} : |\text{rho}_{MHI}(\text{male}, \text{female})|,$$

$$\text{Permutation scheme: permutation of functional observations}$$

P-value: 0, so we reject H0

MPOWER score

We created a continuous MPOWER score

$$\text{MPOWER}_{score} = \frac{(\text{Monitor} + \text{Protect} + \text{Help} + \text{Warn} + \text{Bans} + \text{Taxes}) - min}{max - min}$$

max=29, min=12

MPOWER score distribution over years

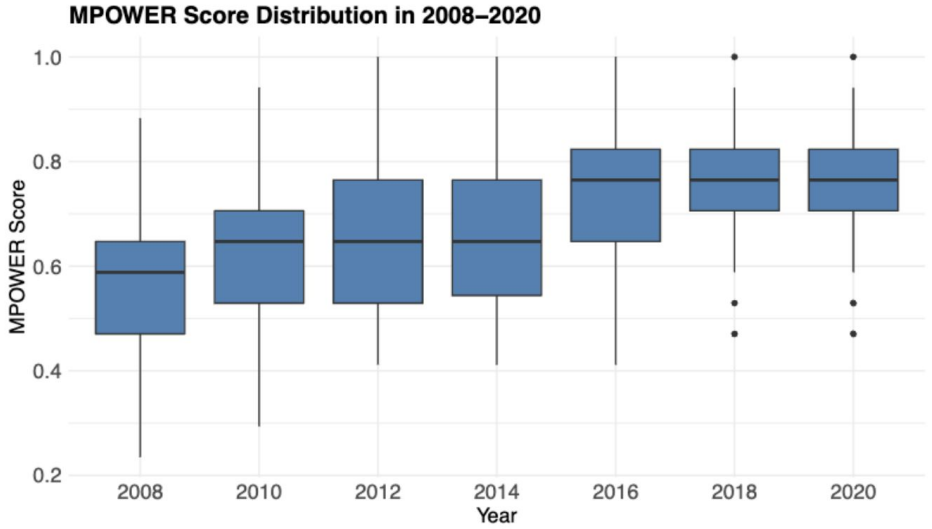


Figure 6: Distribution of overall MPOWER scores on each year.

Permutational tests on the median of the differences

$$H_0 : med(\text{mpower}_{2008} - \text{mpower}_{2020}) = 0 \quad vs \quad H_1 : med(\text{mpower}_{2008} - \text{mpower}_{2020}) \neq 0,$$

$$\text{Test statistic} : |median(\text{mpower}_{2008} - \text{mpower}_{2020})|,$$

Permutation scheme: permutation of the intra-country differences

P-value: 0, so we reject H0