**Analysis plan: Time trends in tobacco purchases in Denmark**

**Target journal:** The Lancet Respiratory Medicine or Tobacco Prevention and Cessation

**Objectives:** It is well-known that smoking is a risk factor for different diseases, cardiovascular disease

amongst others. 73 % of smokers in Denmark wish to quit, still, the smoking prevalence

among adults in Denmark is 17 %. To target preventive as well as smoking cessation interventions, it is important to obtain knowledge about smokers’ behavioral patterns. Consumer purchase data provides a novel approach to assess time trends in tobacco purchases among different population groups. In this paper, we focus on investigating time trends in the amount of tobacco packages bought in Denmark across gender, age, and status of CVD risk.

**Study design:** Longitudinal observational study.

**Data:** Tobacco purchase data from January 2019 until September 2020 (soon October 2021) was collected through an app containing electronic receipts (Storebox). Data consisted of 24,209 tobacco transactions (100,852 tobacco packages) made by 1,857 subjects (out of a total number of 8,800 storebox users by September 2020).

**Methods**

* **Defining a time dynamic population of active storebox users:** For each day, a time dynamic active storebox population (sb.pop) will be defined by the subjects having made at least one transaction within the past month (30 days). Note that the active storebox users per 1st Jan 2019 were based on purchases in Dec 2018. This is done to avoid an unrealistically large storebox population at the end of the period, as some subjects stop using storebox at some point.
  + Whether or not subject i is active at time t (time unit of days) is defined as follows:

* + We then define the active sb population at time t by summing over the n subjects:

* **Defining the tobacco package rate per day and month:** The above population is used to calculate the amount of tobacco packages bought per day (t) per active storebox user:
* This can then be summed by calendar month (M months in total) to get the amount of packages bought per month (m=2019-01, 2019-02,…,2010-09) per active storebox user:
* **Including covariates and grocery shopping history**
  + This parameter () will be investigated for the following subgroups:
    - Gender (male / female)
    - Age (<=40, 40+)
    - CVD risk (hypertension, statins, or diabetes medication).
  + **Graphical approach:** The monthly tobacco package rates will be shown and compared graphically for the different subgroups (including bootstrapped confidence intervals).
  + **Modelling approach:** Assuming that age and CVD risk is time varying, and gender is fixed, we define the set of covariates known just before time t by the filtration:

. Our target parameter is then to estimate the expected number of tobacco packages per active storebox user given :

.

* **Sub-analyses**
  + Compare lockdown periods with different control periods (with January considered as a special case, as literature shows that smoking decreases in January).
    - Jan 2021 (lockdown) with Jan 2020 and Jan 2019.
    - Feb-Mar-Apr 2021 (lockdown) with the preceding 3 month period Sep-Oct-Nov 2020 and the 3 month following period Jun-Jul-Aug 2021.
  + Instead of comparing trends in calendar months, we compare trends in time of tobacco purchase: day (< 16.00 ) vs. evening (> 16.00 ) and weekday (mon-fri 16.00) vs. weekend (fri 16.00-sun).

**Hypotheses**

|  |  |
| --- | --- |
| **Seasonal variation** | **Overall:** Less tobacco in January due to New Year’s resolutions and quit attempts. More tobacco in the summer due to weather and social smoking (active sb users expected to decrease due to summer vacation). General decrease over the years (like for the general population) due to increased tobacco prices and increased focus on risk of smoking.  **Gender:** Same seasonal effect for male / female.  **Age:** Higher increase in January for 40+ due to quit attempts. Higher increase in summer for <40 due to social smoking and summer parties.  **CVD risk status:** Less tobacco purchases for CVD risk group due to fear of disease and counseling about risk of smoking. More tobacco purchases due to stress of disease. |
| **Tobacco toll** | Increase in March 2020 due to hoarding before tobacco toll 01 April 2020.  Same effect expected among the different subgroups. |
| **Lockdown in March-May 2020 and Jan-April 2021** | **Overall:** Increase due to higher stress levels/no smoke policies when working from home. Decrease due to fear of COVID/window of opportunity to quit/less social smoking. Difficult to say overall, subgroup specific. Might decrease initially, followed by an increase back to same level when people realize that the lockdown is “the new normal”.  **Gender:** Decrease for women due to fear of health, unchanged or increase for men mostly due to stress about financial matters caused by COVID-19.  **Age:** Decrease for 40+ due to fear of COVID-19, and decrease for <40 due to attempts to quit and less social smoking. Increase for <40 due to increased stress and increase for 40+ due to remote workplaces/isolation.  **CVD risk status:** Decrease for CVD risk group due to fear of COVID. Increase due to increased stress and isolation. |
|  |  |