Nudging customers towards healthier food and beverage purchases in a real-life online supermarket: A multi-arm randomized controlled trial

Analysis plan

Josine M. Stuber, et al.

Research questions

What are the real-life effects of single and combined nudging strategies in an online supermarket on the 1) total percentage of healthy purchases (primary question), 2) total percentage of healthy purchases within individual food groups (secondary question), and 3) total retailer revenue (tertiary question).

Population

All customers who purchased groceries in the online supermarket during the five week study period between mid-August and mid-September 2020. Business related customers, and customers who purchased >90% of products in their shop from alcohol, other foods, and snacks, are excluded from the analysis.

Outcome measure (dependent variables)

1. Total percentage of healthy purchases of all purchases (calculated based on purchased grams)
2. Total percentage of healthy purchases within 19 food groups (calculated based on purchased grams within the specified food group):
    a) Fruits
    b) Vegetables
    c) Breads
    d) Bread substitutes
    e) Potatoes
    f) Pastas and rices
    g) Teas and coffees
    h) Sodas, waters and juices
    i) Cheeses
    j) Milk and yogurt products
    k) Meats
    l) Fish

m) Legumes
n) Nuts
o) Fats
p) Other foods
q) Savory snacks
r) Sweet snacks
s) Alcoholic drinks

3. Total retailer revenue in Euros of all purchases

Independent variable

Categorical variable (four trial arms):

1. Arm 1: Control supermarket
2. Arm 2:Information nudges
3. Arm 3:Availability nudges
4. Arm 4:Information and availability nudges

Confounders

Not applicable in a randomized controlled trial.

Effect-modifier

Neighborhood deprivation level (dichotomous variable: high or low deprivation level).

Statistical analysis

Descriptive statistics reported by trial arm for the proportion of females (n(%)), mean age (±SD), and proportion shops from deprived areas (n(%)). Additionally, for the control arm (arm 1), the total percentage healthy grams purchased, and percentage healthy grams per food group are reported (mean (95% confidence interval)).

Continuous outcome data distributions are explored via histograms and descriptive statistics (mean (SD)). A linear mixed model with a random intercept on the store level is used to assess the mean differences between arm 1 and the intervention arms in the percentage of all healthy purchases (primary outcome). Residual plots are used to check the model fit. To test for effect modification by neighborhood deprivation levels, an interaction term will be added to the model (interaction for: group * neighborhood deprivation

level). In the case of an significant interaction term ($p< 0.05$) all results will be stratified by neighborhood deprivation level.

As the sales data of the secondary and tertiary outcomes is U-shaped (percentages healthy products within individual food groups) and/or highly right skewed (total revenue) nonparametric bootstrapping for hierarchical data will be used to estimate the between-arm mean differences in purchases across individual food groups (secondary outcomes) and total revenue (tertiary outcome). The bootstrap analyses will be based on 10,000 non-parametric bootstrap replicates. Shops who did not include purchases of products within a specific food are excluded from the bootstrap analysis for that food group. Statistical significance of all outcomes is defined as the absence of zero in the 95% confidence interval.