Data Report

Study Design and sample

This study employed an observational cross-sectional design. It targeted a sample size of 2,500 individuals, with eligible participants being adults aged 18 and above, currently residing in Lebanon, regardless of their nationality. A multistage cluster sampling approach was employed to ensure comprehensive coverage of the target population across Lebanon's eight governorates. The sampling method used probability-proportional-to-size random selection to choose clusters, allowing for adequate representation of each region. The sampling process involved two stages: first, selecting geopolitical areas (districts or clusters) within each governorate, and second, selecting housing units within the selected areas based on the weight allocated to the district (based on the number of residents in the district as of the Ministry of Public Health Statistics in 2020). All 28 districts/clusters in the country were included, and in each district, the capital was chosen as a primary cluster, with additional villages selected as necessary to meet sample size requirements. Households were selected at random using a systematic sampling technique. Within each selected household, one eligible male and one female were chosen for participation. Each household was visited up to three times to secure participation. If a household was uninhabited, or if the selected individuals were unavailable or declined participation, the next household was selected. To maintain the needed male-to-female ratio, additional nearby households were included as required.

Data Collection Method:

The data collection was conducted through face-to-face interviews using tablets. A team of 24 trained field workers, all holding CITI certification in Social and Behavioral Research, carried out the surveys. These fieldworkers were selected for their competence and experience. The fieldwork took place between February 19, 2024, and April 8, 2024. A total of 1,571 households were approached, with 1,310 agreeing to participate, leading to 2,500 completed questionnaires. In cases of non-response (213 refusals and 48 instances of no one being home after three visits), additional households were selected to maintain the required sample size and balance.

Ethical Considerations:

Ethical approval for the study was obtained from the American University of Beirut’s Institutional Review Board (IRB). Informed consent was verbally obtained from all participants after explaining the study's purpose, procedures, and participants' rights. Participants were assured that their responses would remain confidential and anonymous, with personal identifiers encrypted and access to data restricted to the survey team. Participation in the study was entirely voluntary, and respondents were informed that they could withdraw at any time without any consequences. All data were securely stored in a password-protected database to maintain confidentiality.

Questionnaire:

The questionnaire was developed using insights from previous studies and validated surveys of tobacco use such as the Global Adult Tobacco Survey (GATS), and in consultation with experts from the World Bank. The survey was conducted in both Arabic and English, based on participant preference, and each interview lasted approximately 30-40 minutes. It consisted

of three main components: demographics and income, smoking behaviors, and a volumetric choice experiment. The pilot test of the questionnaire, conducted with 28 participants in February 2024, allowed for refining and adjusting the survey design to enhance clarity and ensure the smooth conduct of the main data collection process. The questionnaire was designed to capture a comprehensive range of information, extending beyond basic demographic details to encompass socio-economic factors, current and retrospective tobacco use behaviors, and responses to price fluctuations.

The first section of the questionnaire includes questions regarding demographic and socio-economic information. This includes basic demographics (age, gender, location of residence, etc), completed education, employment status, individual and household monthly income, household monthly expenditure. The second section focuses on tobacco use and smoking habits. The questionnaire identifies current, past and never smokers. It then conditionally asks a comprehensive set on current and retrospective tobacco consumption, including frequency and volume of smoking, spending on tobacco, brand choice, and product switching. This section also includes information on smoking cessation intentions, health awareness, and views on government policies regarding tobacco regulation and taxation. Additionally, this section includes questions aimed at evaluating the impact of the crisis on tobacco use. Participants were asked to compare tobacco prices before and after a crisis and to assess how these price changes influenced their consumption patterns. The final section of the survey consists of a volumetric choice experiment (VCE). This exercise was restricted to participants who reported being current smokers. The VCE design is explained in detail in the next section.

Volumetric Choice Experiment Design:

The experimental design utilizes two cigarette types: (1) Local brands (e.g., Cedars, Byblos) which are usually cheaper, and (2) Foreign brands (e.g., Marlboro, Davidoff, Gitanes) which are usually more expensive. Each respondent is presented with hypothetical choice scenarios featuring varying price levels for these two types. The design ensures that price combinations change across scenarios to capture the potential substitution or complementarity between the two products. The prices for local and foreign cigarettes varied across a realistic range of market prices at the time the study was conducted and included prices that are typical of each product type. Each product therefore assumed one of 9 possible levels, ranging between $0.30 and $4.50 for local brands, and $1.00 and $6.00 for foreign brands. These price vectors included, at their respective third levels, the typical market prices of $0.60 and $2.00 for local and foreign brands, respectively.

Given the simple choice task format composed of two products with nine possible levels each, we opted for a full factorial design, resulting in 81 unique choice sets derived from all possible two-way price combinations. To reduce respondents' response burden, these 81 choice sets were blocked into nine groups, each containing nine scenarios. Each respondent was randomly assigned to one of the nine blocks, ensuring orthogonal representation of price combinations while limiting fatigue during decision-making. Respondents could choose one or more of each product or none at all, depending on their preferences relative to the products' prices. An example choice task can be found in Figure 1.

SECTION C: CHOICE EXPERIMENT

Description:

- In the next section, you will see hypothetical choice scenarios about buying cigarette packs for yourself. The only difference between the scenarios is the price of the cigarette products
- Each scenario will present you with 3 different tobacco products that you may purchase:
    1. Local brand of cigarettes like Cedars or Byblos
    2. Foreign brand of cigarettes like Marlboro, Davidoff or Gitanes
- Each of the 2 cigarette packs will have its own price, and the combination of prices change in each scenario.
- You will be asked to imagine yourself buying your needs of cigarettes as you would typically in real life and tell us how many items you would buy of each of the 2 products on offer.
- You may choose to buy 1 or more items from some types and none (i.e. 0) from others. You may choose to buy nothing at all, in which case the answer will be 0 for each of the 2 products.
- Please seriously consider the prices (in fresh USD) at which each item is being offered when making your decision about the quantities. Make sure that the sum you spend on these products is in line with your budget for buying cigarettes. Also, keep in mind that the money you spend on cigarettes will not be available for you to spend on other things.
- Here is an example of a choice scenario:

Think about the amount of cigarette packs you would purchase, based on the following prices:

	Local brand Pack of 20	**Foreign brand Pack of 20**
Price	$0.60 (LBP 55,000)	$2.00 (LBP 180,000)
Quantity	|___|	|___|

Now I leave you to complete the choice scenarios. **[INTERVIEWER: LEAVE THE PARTICIPANT TO FILL IN THE CHOICE SETS]**

Sample Descriptives

The study sample comprised 2,500 participants, equally distributed by gender (50% female, 50% male). The mean age of participants was 43.27 years (SD = 3.9), with an average household size of 3.9 members (SD = 0.03). Most participants were Lebanese nationals (90.7%), followed by Syrians (8.1%), Palestinians (0.2%), and other nationalities (1%). Participants were distributed among all the Lebanese governorates as follows: Beirut (9.36%), Mount Lebanon (35.52%), South (11.68%), North (14.48%), El Nabatiyeh (7.60%), Bekaa (6.88%), Baalbek-Hermel (7.44%), and Akkar (7.04%). Employment status varied among participants: 56.7% were employed, 8.56% were unemployed, and 34.16% were inactive in the labor market, including students, homemakers, and retirees. Additionally, 0.56% were not in the labor force. Most participants (32.8%) reported receiving either partial payment in fresh USD or being fully paid in fresh USD (34.7%). Others (3.4%) received partial payment in "Lollars" (bank rate), were fully paid in "Lollars" (4.1%) or were paid in Lebanese pounds (11.9%). The mean monthly individual income was 296.25$ (95%CI: 278.872; 313.635), while the mean monthly household income was 705.26$ (95%CI: 676.372; 734.157). Additional demographic characteristics are summarized in Table 1.

Table 1: Demographic and Socioeconomic Characteristics of the Study Sample

Demographic characteristics		
	n	%
Gender		
Female	1250	50
Male	1250	50
Marital status		
Single	411	16.4
Married or engaged	1,950	78.0
Separated	11	0.4
Divorced	52	2.1
Widow(er)	75	3
Nationality		
Lebanese	2,267	90.7
Syrian	202	8.1
Palestinian	6	0.2
Other	25	1
Highest educational level		
Primary school	443	17.7
Middle school	678	27
High school/equivalent	610	24.4
Undergraduate degree	548	22
Postgraduate degree	105	4.2
No formal education	114	4.6
Employment Status		
Employed	1,418	56.7
Unemployed	214	8.56
Inactive in labor market	854	34.16
Not in labor force	14	0.56
Governorate of Residence		
Beirut	234	9.36
Mount Lebanon	888	35.52
South Lebanon	292	11.68
North Lebanon	362	14.48
El Nabatieh	190	7.60
Bekaa	172	6.88
Baalbek-Hermel	186	7.44
Akkar	176	7.04
	Mean	**SD**
Age	43.27	0.28
Number of individuals in the household	3.9	0.03
Monthly Individual Income ($)	296.25	436.85
Monthly Household Income ($)	705.26	725.25

Data Imputation Approach

This study applied data imputation techniques to address missing values in individual income, household income, and household expenditure. For individual income, out of 2,500 observations, 191 (7.6%) respondents did not provide a continuous income value. Among them, 120 reported their income within predefined intervals, while 71 refused to disclose any information. To impute missing values, we estimate log income regressions based on individual characteristics such as gender, age, marital status, education, nationality, urbanicity, and region of residence. Separate regressions were performed for each income interval, and the fitted values were used to generate imputed data, ensuring alignment with the reported intervals. An alternative approach using software built-in imputation functions was also used to generate another set of imputed values. The average reported income (excluding interval responses) was $276, the average using the primary imputed data was $296 and the average using the built-in imputation function was $282. The empirical distribution of the original and imputed income variables showed no significant aggregate differences, though our primary imputation ensured values remained within the self-reported income ranges.

A similar methodology was applied to household income and household expenditure. For household income, 285 (11.4%) respondents did not report a continuous value, with 208 providing income intervals. The primary imputation approach yielded an average household income of $705 compared to $659 for the pre-imputed variable. Regarding household expenditure, 718 (28.7%) observations were missing continuous values, with 616 providing intervals. The average household expenditure after primary imputation was $683, compared to $657 for the pre-imputed variable.

The imputation marginally increased the averages of the variables. This is expected since higher income individuals are more likely to report an interval value. Which is what we note in the data. For instance, in response to the income interval questions 4 respondents reported earning a monthly income in the more than $5,000 intervals, however, in the continuous data there was not a single person reporting an income higher than $5,000.

How representative is the sample

The absence of a population census, or a clear national sampling framework poses a significant challenge to primary data collection. This section provides an assessment of the representativeness of the survey by comparing key demographic statistics to the ones reported in the latest labor force survey (LFS2022) conducted by the Central Administration of Statistics (CAS) in January 2022.

The demographic distribution in the survey is closely correlated with that in LFS2022. Lebanese nationals are overrepresented in the tobacco survey, with 90% of the participants being Lebanese, in comparison to 80% in the LFS[1]. The Geographic representation in the sample correlates with what is documented in LFS2022. There are two governorates where

[1] It should be noted that LFS oversample Lebanese nationals compared to the proportion of residents. The LFS sampled individuals living in legal dwelling which disproportionately exclude non-Lebanese nationals, most notably Syrian refugees who reside in unofficial settlements, and foreign workers who reside in place of work. The total non-Lebanese residents in Lebanon is highly uncertain with wide range of estimates provided by official Lebanese authorities and multinational institutions supporting refugees and displaced individuals.

we note significant difference, with oversampling in Beirut governorate and under sampling in Mount Lebanon. The age distribution in the survey is significantly different than the LFS2022. This is partly due to restricting the survey to individuals of adult age. The age distribution shows that 95% of survey participants are between the ages of 18 and 68, meaning that in addition to excluding minors, the survey underrepresents the elderly.

The sample in the tobacco survey is significantly over educated, almost 50% of the sample in the survey has completed high school or more, in comparison to 27.3% in the LFS2022. Part of this difference is explained by oversampling Lebanese and under sampling older birth cohorts.

Finally, a comparison with income and employment status shows a higher rate of employment in the survey and significantly higher average income. The rate of employment in the survey is 81.7% and 31.68% for men and women respectively, in comparison to 66.2% and 22.2% in LFS2022. The average individual income in the survey is $296.25, in comparison to $91.36 in LFS2022. Similarly, our survey found that 82.5% of households reported monthly incomes below $600, in comparison to 89.7% of households included in the 2022 Labor Force Survey (LFS). This is not surprising, as wages and employment by 2024 did rebound since the height of the crisis in 2022.