ITC Japan Survey
Wave 4 Technical Report

May 9, 2022
Updated June 20, 2023

ITC Japan Research Team
Dr. Katanoda Kota[1], Co-Principal Investigator
Dr. Takahiro Tabuchi[2], Co-Principal Investigator

ITC Japan International Research Team
Professor Geoffrey T. Fong[3], Principal Investigator, Canada
Professor Mary E. Thompson[4], Co-Investigator, Canada
Dr. Anne C.K. Quah[3], Co-investigator, Canada
Dr. Christian Boudreau[4], Co-investigator, Canada
Ruth Loewen[3], Survey Management Team, Canada
Dr. Kai Jiang[3], Survey Management Team, Canada
Shelly Jordan[3], Survey Management Team, Canada
Grace Li[3], Data Analyst, Canada
Janine Ouimet[3], Project Manager, Canada
Dr. Steve Xu[3], Project Manager, Canada

[1] Institute for Cancer Control, National Cancer Center, Japan
[2] Cancer Control Center, Osaka International Cancer Institute, Japan
[3] Department of Psychology, University of Waterloo, Canada
[4] Department of Statistics and Actuarial Science, University of Waterloo, Canada

Funding
Canadian Institutes of Health Research Foundation Grant (FDN-148477).

Ethics Statement
The survey protocols and all materials, including the survey questionnaires, were cleared for ethics by the Office of Research Ethics, University of Waterloo, Canada (REB#22508/31428), the Internal Review Board at the Osaka International Cancer Institute, Japan (IRB 21054) and the Internal Review Board at Japan National Cancer Center, Japan (IRB 2021-069). All participants provided consent to participate.

Suggested Citation: ITC Project. (2023, June). ITC Japan Wave 4 (2021) Technical Report. University of Waterloo, Waterloo, Ontario, Canada, Osaka International Cancer Institute, Osaka, Japan, and Japan National Cancer Center, Tokyo, Japan.

CONTENTS

1 Summary and Overview

1.1 Introduction

The International Tobacco Control Policy Evaluation Project (the ITC Project) is the first-ever international cohort study in each participating country, designed to measure the psychosocial and behavioral impact of key policies of the World Health Organization Framework Convention on Tobacco Control (WHO FCTC). The ITC Project is a large global research initiative that currently involves 31 countries (Australia, Bangladesh, Bhutan, Brazil, Canada, China (Mainland), France, Germany, Greece, Hungary, India, Ireland, Israel, Japan, Kenya, Republic of Korea, Malaysia, Mauritius, Mexico, Netherlands, New Zealand, Poland, Romania, Spain, Thailand, United Arab Emirates, United Kingdom, Uruguay, United States of America, Vietnam, and Zambia) inhabited by over 50% of the world's population, over 60% of the world's smokers, and over 70% of the world's tobacco users.

The ITC Project team from the University of Waterloo in Canada partnered with the Japan National Cancer Centre and Osaka International Cancer Institute to develop and manage the ITC Japan Survey. The ITC Japan Project officially commenced in September 2017. The ITC Japan Survey was designed to provide data on tobacco users' and non-users' knowledge, attitudes, beliefs, perceptions, behaviours, and use patterns associated with heated tobacco products (HTPs) from the most important market for these products. Patterns of use and attitudes from Japan may start to appear in other countries in which these products have been or will be released. The ITC Japan Survey is particularly intended to provide evidence for tobacco control policies related to HTPs and their use. Although HTPs are marketed as reduced-harm tobacco products, the effect of HTP use on tobacco cessation, uptake, and/or sustained use is not yet fully understood.

1.2 Main objectives and research questions

The objectives of the Wave 4 of the ITC Japan (JP4) Survey were:

1) To describe how HTP and cigarette use differs among smokers and HTP-only users (some of whom may also be recent ex-smokers) both over time and between important subgroups (e.g., age, gender, income, those planning to quit cigarettes, nicotine dependence level).
2) To examine patterns of tobacco-use behaviour and opinions associated with smoking among adults aged 20 and older in Japan.
3) To examine the effect of a comprehensive smoke-free law on both cigarette smoking and HTP use overall and in public places as well as private places (e.g., homes, cars)4) To examine the impact of recent enhancements of tobacco control policies in Japan.

1.3 Overview of project

- The ITC JP4 Survey was a web-administered survey of behaviours and attitudes related to tobacco and nicotine use among a sample of n=4,500 adult residents of Japan, comprising the following 4 user types: 2,000 cigarette-only smokers, 1,000 cigarette/HTP dual users, 1,000 HTP-only users, and 500 non-users, a total of 4,500. The user type definitions are provided in Table 1 in Section 3.1. Of all 4,500 respondents, the median time of completing the JP4 Survey was 36 minutes. Technical details of the web survey are provided in Appendix 3.
- All sampling and fieldwork were conducted by the same survey firm that conducted the first three waves of the ITC Japan Survey, Rakuten Insight. The sampling frame was an existing Rakuten Insight panel. Further quotas based on the region of residence, sex, and age, were applied to ensure the final sample sizes were proportional to stratum sizes based on the Japan Society and New Tobacco Internet Survey (JASTIS).
- The JP4 Survey was conducted from July 3 to August 22, 2021. Follow-up surveys are planned contingent upon funding.
- Rakuten Insight modified their standard procedures to maximize retention in the panel of the ITC Japan Survey across waves by using a customized ITC-branded background for the survey and survey communications, as well as offering bonus incentives for respondents who complete follow up waves of the survey.
- The fieldwork periods of the four waves are:

ITC Japan Survey Fieldwork Dates	
Wave 1	February 3 to March 2, 2018
Wave 2	December 7, 2018 to January 25, 2019
Wave 3	May 16 to June 17, 2020
Wave 4	July 3 to August 22, 2021

2 Survey Sampling, Protocol, and Process

The ITC Japan Wave 4 (JP4) Survey was a follow-up to the ITC Japan Wave 3 (JP3) Survey. The ITC JP4 Survey included a recontact phase to re-survey respondents from the JP3 Survey, and a replenishment phase to replenish the cohort from the same sampling frame to replace those lost to follow-up.

2.1 Sample

The overall JP4 target sample size was n=4,500. This sample was subdivided along two dimensions:

- Four subsample strata based on cigarette and heated tobacco products (HTPs, also known as heat-not-burn) use status (see Table 1).
- Recontact sample (cohort members who completed the JP3 Survey) vs. Replenishment sample (new recruits at JP4 to replace those either lost to follow up or stayed in the survey but changed their tobacco use behaviours).
- Variables FR309v and HN309v integrate these two dimensions to inform the Recontact sample subgroups and the Replenishment quotas.
- Table 1 provides the four recontact Subsample Groups and further Definition Groups with explanations for the self-reported user types that are counted within each category.

Table 1: JP4 recontact and replenishment Subsample Groups (SG) and Definition (Def) Groups, with explanations

Subsample Group	Definition Groups and explanations	Recontact	Replenishment
SG1. (n=2000)	Def 1: Exclusive Cigarette Smoker • Smokes cigarettes at least weekly AND has smoked 100 cigarettes in lifetime. • Uses HTPs less than weekly or not at all	• FR309v=1-2 and HN309v=3, 4, 5-7	
	Def 2: Cigarette Quitter • Has smoked at least weekly in the past and smoked 100 cigs and doesn't smoke at least weekly now. • Has never used HTP at least weekly. • At Wave 4: recruit Replenishment in this group if quit in the last 2 years	• FR309v=4-8 and HN309v=3, 5-7	• FR309v=4-7, and HN309v=3, 5-7
SG2. (n=1000)	Def 3: Exclusive HTP user • Smokes cigs less than weekly or not at all • Uses HTPs at least weekly.	• FR309v=3-9 and HN309v=1-2	
	Def 4: HTP Quitter (Recontact sample only) • Has never smoked cigs at least weekly. • Has used HTP at least weekly in the past and doesn't use HTP at least weekly now. • At Wave 4: Do not recruit Replenishment in this group.	• FR309v=3, 9 and HN309v=4	• Not eligible
SG3. (n=1000)	Def 5: HTP- Cigarette Dual user • Smokes cigarettes at-least-weekly AND has smoked 100 cigarettes in lifetime • Uses HTPs at least weekly.	• FR309v=1-2 and HN309v=1-2	

Subsample Group	Definition Groups and explanations	Recontact	Replenishment
	Def 6: Dual quitter • Has smoked at least weekly in the past and smoked 100 cigs and doesn't smoke at least weekly now. • Has used HTP at least weekly in the past and does not use HTP at least weekly now. • At Wave 4, recruit Replenishment in this group, if quit in the last 2 years	• FR309v=4-8 and HN309v=4	• FR309v=4-7 and HN309v=4
SG4. (n=500)	Def 7: Never/non-user • Has not smoked cigarettes at least weekly (or, for replenishment, not within the past 2 years) • Has not used HTP at least weekly.	• (FR309v=3 or 9 and HN309v=3, 5-7.)	• FR309v=3 or 8-9 and HN309v=3, 5-7.

Note: Sample targets are assigned for SGs. Each SG consists of 1-2 Defs.

- Within each of the four-subsample stratum, those either lost to follow-up or changed tobacco use status were replaced with new recruits (Replenishment (P) sample) who met the criteria for the subsample stratum.

Table 2: JP4 sample with retention and replenishment targets by subsample based on valid JP3 sample respondents

Subsample group (SG)	JP3 sample (valid)	Target JP4 sample	JP4 C sample (valid)	JP4 P sample (valid)	Total JP4 sample (valid)
SG1. Current exclusive smokers (Def 1) and cigarette quitters (Def 2)	1,977	2,000	1,328	595	1,923
SG2. Current HTP-exclusive users (Def 3) and Recontact HTP quitters (Def 4)	943	1,000	392	511	903
SG3. Current cigarette-HTP dual users (Def 5) and cigarette-HTP dual quitters (Def 6)	985	1,000	649	295	944
SG4. Never or Non-users (Def 7)	482	500	310	174	484
Total sub-sample	-	-	**2,679**	**1,575**	-
Total sample	**4,387**	**4,500**	**4,254**		**4,254**

Note: P = Replenishment, C= Recontact/cohort

The retention rate between JP3 and JP4 is 61.1%.

A detailed description of the transitions of the JP3 recontact sample (lost, successfully recontacted under the same or a different user type) and the replenishment sample of new respondents at JP4 by user type is shown with a flowchart that can be found in Appendix 4.

2.2 Survey development process

The survey development process comprises four main phases:

1. Determining survey content

2. Operationalization of survey content
3. Translation into Japanese
4. Translation review and checking

2.2.1 Survey Development – content and operationalization

- During Phase 1 of the survey development process, the research investigators, project management team, and the survey management team determined which topics were most important to include in the survey, and then developed the detailed survey questions necessary to measure relevant constructs using the existing framework of the ITC database of questions. Questions were adapted to the Japanese context and new questions were designed as necessary. The resultant draft survey was then sent to the ITC Survey Management Group (SMG) for operationalization of the survey (Phase 2).
- In Phase 2, the operationalization of survey development involved comprehensively and iteratively reviewing and revising the survey to ensure that routing, question wording, response options, and all other survey elements were refined and cross-referenced for consistency, clarity, and accuracy. At the conclusion of Phase 2, the final draft of the survey was generated by SMG and sent to Rakuten Insight for programming and testing.
- During the period when the survey firm programmed and tested the survey, additional revisions were made in consultation between Rakuten Insight and SMG, until a fieldwork version of the survey was achieved. The fieldwork version of the survey was sent to SMG by the firm and is retained in the SMG database. The updated last version of the survey in the database was later used in cross-reference with the data set.

2.2.2 Survey Development – translation and review/verification

- The team first developed the JP4 survey content and specifications in English. The final English JP4 Survey was then translated into Japanese by Rakuten Insight per specifications provided by the research team.
- After the initial translation from English to Japanese had been completed and checked internally by Rakuten Insight's translator(s), the Japanese translation of JP4 was then checked by the Japan team, and issues were identified, discussed, and resolved to confirm that the Japanese translation met the research team's standards for the highest possible degree of accuracy.

2.3 Survey content

The JP4 Survey content was developed to assess the research objectives described in Section 1.2 as well as to measure other constructs necessary to meet the survey objectives. These included demographic, social, and psychological factors relevant to models of behaviour change, as well as content to meet logistical requirements for the survey.

The specific JP4 content included the following:

- Information about the survey, time commitment, contact information for ethical concerns or survey-related concerns, and an explicit consent screen.
- A screening section that assesses age, gender, region of residence, smoking status, HTP use status.
- HTP questions: brand/device choice, duration of use, dependence, reasons for using, warning labels, purchase, environmental exposure, advertising/promotion, and perceived risk.
- Cigarette questions: brand choice, perceptions of light/mild, dependence, quitting attempts and aids used, knowledge of health effects of tobacco, warning labels, smoke-free places, advertising and promotion, purchase, beliefs about quitting, psychosocial beliefs, regulation, perceived risk.
- Other questions: anti-tobacco campaigns, e-cigarette use, moderators (i.e., factors not asked in any of the previous categories that are important in models of behaviour change and/or policy evaluation such as time perspective, stress, co-morbidities).
- Demographic questions (e.g., age, gender, education, income, socio-economic status).
- COVID-19 questions about the pandemic and its impact on tobacco use.

2.4 Web survey programming, testing and translation

2.4.1 Unique ID and smart data

- Unique ID:

    - For Recontact records, the SAME unique ID number as in JP3 was used for each respondent's JP4 record.
    - For Replenishment records, a NEW unique ID was assigned at JP4.
- Smart data: Although both Recontact and Replenishment Survey pathways were programmed within one single Computer-assisted Web Interview (CAWI) instrument, the Recontact records (but not the Replenishment records) included 'smart data':
    - The 'smart data' for the Recontact sample records included information from the JP3 Survey that informed the JP4 CAWI program from a back-end data file.
    - ITC UW was responsible for collating and providing the smart data to Rakuten Insight.

2.4.2 Web survey programming and testing

- Rakuten Insight used the ITC JP4 Survey specifications (provided in Microsoft Word format) to program the English Survey using Confirmit CAWI software.
- The Recontact and Replenishment survey pathways were specified within the same single CAWI instrument.
- 'Smart data' extracted from JP3 were programmed into the JP4 Survey Recontact records.
- Rakuten Insight worked closely with the research team to test the English survey CAWI program and survey quotas. The research team provided signoff on the JP4 English Survey program that met pre-determined standards prior to beginning data collection.

2.4.3 Translation, overlay into the English CAWI program, and testing

- Rakuten Insight followed the ITC Translation Manual to translate the English Survey into Japanese
- After the Japanese translation was accepted, Rakuten Insight overlaid the Japanese into the CAWI program (originally programmed in English)
    - Rakuten Insight internally tested the Japanese CAWI program to ensure accuracy.
    - ITC UW and the Japan team tested the Japanese CAWI program to ensure the program was functioning as intended and was free of errors. ITC UW provided signoff on the Japanese CAWI program prior to fieldwork.

2.4.4 Programming and testing other technical components and quotas

- Rakuten Insight worked with ITC UW to ensure quotas and all technical components necessary for fieldwork met ITC standards.

3 Fieldwork Procedures

3.1 Ethics clearance

- Prior to any fieldwork procedures/communications with panelists for the ITC JP4 Survey, ethics clearance had been provided by the Office of Research Ethics, University of Waterloo (REB#22508/31428), the Internal Review Board at the Osaka International Cancer Institute, Japan (IRB 21054) and the Internal Review Board at Japan National Cancer Center, Japan (IRB 2021-069).

3.2 Inclusion/exclusion criteria for Rakuten Insight panelists

- Recontact respondents identified from the JP3 Smart Data File were eligible to participate in the JP4 Survey
- The inclusion criteria for the Replenishment sample web panelists were as follows:
    - All respondents must be adults aged 20 years or older.
    - Respondents must be either a current cigarette only smoker (smokes at least weekly), a current HTP only user (uses HTP at least weekly), a current dual user (smokes a least weekly and uses HTP at least weekly), or a tobacco non-user, per the definitions in Table 2.
    - The quota for the panelist's specifications (i.e., user type, region of residence/urbanization, age, gender) was still open.
- Web panelists excluded from the Replenishment sample were:
    - Those younger than 20 years old.
    - Those who did not meet the definition for one of the four user or non-user subsample types specified in Table 2.
    - Those for whom the quota corresponding to the panelist's specifications was full.

3.3 Description of sampling frame

- The sampling frame of the JP4 Survey was Rakuten Insight's web panel. All of the user types were recruited from Rakuten Insight's panel(s).
- Rakuten Insight provided the following description of their panel(s): The ITC Japan Surveys were conducted with Rakuten Insight's proprietary online panel in Japan. The online panel is actively managed in-house with a dedicated panel management team in Tokyo, and utilized for market research purposes only. Recruitment for the panel is conducted daily, tapping into users of Rakuten services (e.g., e-commerce, credit cards, insurance, mobile services, etc.), as well as other online resources such as affiliates, and email and banner recruits in order to maintain a panel as consistent as possible with the general population. Panelists are pre-profiled with a series of questions which in turn can be used as pre-targeting variables (e.g., smoking, HTP usage). Panelists receive email invitations and also have the option of logging into their proprietary panel site to access the survey they are invited to participate in. Details available at: https://insight.rakuten.com.

3.4 Survey invitations (Phase 1, 2, and 3)

3.4.1 Phase 1 – Recontact sample email invitations

- Invitation emails were sent to JP3 sample members identified as eligible for the ITC JP4 Survey.
- Rakuten Insight invited panel members to the ITC JP4 Survey by sending them a standard email invitation that informed the panelists of the survey length and that they would receive the standard incentive for a 25- to 30-minute Rakuten Insight survey.
- Per standard procedures, Rakuten Insight sent one email invitation and two reminders to these recontact sample panelists.
- Once an appropriate level of retention had been achieved or survey activity had plateaued, Rakuten Insight consulted with the ITC UW team to confirm if the replenishment sample recruitment should be started.
- Panelists were able to ignore the emails, or contact Rakuten Insight to refuse the study or unsubscribe from the panel at any time.

3.4.2 Phase 2 – Replenishment tobacco user quotas vs. Phase 3 – Replenishment non-user quota email invitations

- Phase 1 replenishment invitation emails: Invitation emails were sent strategically to replenishment sample identified as probable tobacco users in order to fill the three tobacco user subsample quotas (i.e., quotas for cigarette smokers, HTP users, and dual users were filled first, and those respondents turned out to be non-users accepted as part of the non-user quota).
- Phase 2 replenishment invitation emails: After the three tobacco user subsample quotas (i.e., quotas for cigarette smokers, HTP users, and dual users) were filled, then replenishment invitations strategically targeted non-users to fill the remaining open positions in the non-user quota.
- Rakuten Insight invited panel members to the JP4 Survey by sending them a standard email invitation that informs the panelists of the survey length and that they would receive the standard incentive for a 25-to 30-minute Rakuten Insight survey.
- Per standard procedures, Rakuten Insight sent one email invitation and two reminders to panelists who have been pre-identified as being potentially eligible for the ITC JP4 Survey. Once the quotas had been achieved, the web survey would be closed.
- Panelists were able to ignore the emails, or contact Rakuten Insight to refuse the study or unsubscribe from the panel at any time.

3.4.3 Panelists' survey experience and ITC Survey features

- The ITC JP4 Survey was designed to have the look and feel of a typical Rakuten Insight survey, with some branding to identify the survey as an ITC survey.
- Consent screens required by Canada research guidelines provided information about the survey, time commitment, contact information for ethical concerns or survey-related concerns (the country investigator for content; or Rakuten Insight for technical issues), and asked the panelist to provide his/her consent to complete the survey.
- The recontact and replenishment samples were surveyed within one CAWI instrument, but experienced a tailored survey according to their status.
- 'Smart data' from the respondents' JP3 Survey were used to provide the recontact sample panelists with a tailored survey within a single programmed instrument.
- The ITC JP4 Survey began with a screening section that assessed panelists' eligibility (based on user type and possibly region/gender/age-based quotas) and determined which survey questions (related to user type) would be asked throughout the survey.
- ITC provided the ITC JP4 Specifications and the quota specifications to Rakuten Insight as separate files.
- Rakuten Insight was responsible for programming the survey and the quota system(s) to interact and achieve the appropriate outcome (i.e., selecting panelists for the survey or terminating panelists if their respective quota was already full).
    - Rakuten Insight included a data field that uniquely identifies each record's quota line. The records for every survey respondent as well as any records for panelists who were terminated due to a full quota included the quota line identifier.
- Respondents could navigate back to previous questions to change a response.
- Respondents were able to stop the survey and log in to finish at a later time without losing any data.
- There were several questions (but not many) with 'Other-specify' response options, which required open-text responses by the panelists.
- The survey contained a few encouragement screens that indicated the respondents were getting close to the end of the survey.
- Respondents were required to submit their completed survey in order for their survey record to be considered 'complete' and be provided with the remuneration.
- Item non-response was acceptable, provided that the majority of questions were answered, that 'essential questions' used for eligibility were all answered, and the panelist had submitted his/her survey.
- The median length of the survey interview was 37 minutes for the valid complete records.

3.5 Fieldwork timeline and interview length

- The ITC JP4 Survey was conducted from July 3 to August 22, 2021.

- The median survey interview length is given in Table 3.

Table 3: Median survey interview length (minutes)

Sample description	N	Median survey interview length (minutes)
Final valid completes	4,254	37
Completes with speeders*	4,466	36

3.6 Assigning disposition codes

- Disposition codes were used to track the outcomes of survey respondents.
- Temporary disposition codes were applied to respondents who did not complete the survey within one session
- A final disposition code was assigned to each record (see Section 5 Disposition Codes).
- Three types of disposition codes were used in the study: 1) disposition codes programmed into the survey script, 2) disposition codes entered by the survey firm, and 3) dispositions derived at the end of fieldwork (see Section 5 Disposition Codes).
- Each completed survey record was further sub-coded as being completed on a desktop/tablet device vs. completed on a mobile device vs. being undefined (not possible to classify the device).

3.7 Maximizing retention between waves

- Maximizing retention across waves is extremely important to the longitudinal cohort design of this study. Procedures to improve retention in both the Rakuten Insight panel and the ITC JP4 Survey included: 1) inter-wave reminder emails; and 2) bonus remuneration for follow up surveys.

3.7.1 Inter-wave reminders

- During the 14-month interval between waves, respondents who completed ITC JP3 Survey were provided with 2 reminder(s) requesting that they stay in the panel and look for the ITC JP4 Survey invitation.
- Panelists were able to ignore the inter-wave reminders or contact Rakuten Insight to refuse the study or unsubscribe from the panel at any time.
- The inter-wave reminder(s) included any of or all the following formats: emails, pop-up messages on the panelist's member site, and a survey page that populates when the panelist clicks the link from his/her member site, or from an email invitation, and includes a visual cue – the ITC Project, Japan National Cancer Center, and Osaka International Cancer Institute logos.

3.7.2 Study incentives and bonus payments for follow up surveys

- After a recontact sample panelist completed his/her survey, Rakuten Insight provided the standard number of points for a survey of this length, plus a bonus of $14 USD regardless of tobacco use status.
- Upon submitting a completed survey, all Wave 4 Survey respondents were informed that if they would stay in the panel and complete the JP5 Survey about 12 months later, then they would be provided with a standard bonus to total $14 USD regardless for completing the JP5 Survey. The phrasing for this request will be contingent upon whether funding has been secured for additional waves.
- The bonuses were intended to convey appreciation for completing the survey and to appeal to the participant to stay in the panel and accept the invitation to complete subsequent survey waves.

4 Quality Control

4.1 Fieldwork monitoring and progress reports

- At the beginning of fieldwork, the initial sample invitations were released carefully at deliberate intervals and survey activity was closely monitored to ensure that all aspects were working as intended. This method is termed a 'soft launch' and occurred from July 2-4, 2021.
- The 'soft launch' data were systematically reviewed by both Rakuten Insight and the research team. No major issues of concern were determined.
- Throughout fieldwork Rakuten Insight closely monitored survey activity and ensured a smooth implementation.
- Rakuten Insight provided weekly fieldwork reports and analysis of next steps concerning the survey recruitment strategy.

4.2 Survey completes vs. partial completes

- For the purposes of the fieldwork sample and subsample targets, the definition of a "survey complete" is the survey record for a panelist who started the survey, completed the survey questions, perhaps endorsing "prefer not to answer" for a reasonable proportion of questions, and then chose to 'submit the survey' after the last survey question.
- Survey response data for survey completes were checked using the criteria defined in Section 4.3. Speeders and cases with conflicting answers were considered as invalid completes. Records that passed the checks were considered valid completes.
- Survey response data for partially completed survey records (defined as records for which the panelist answered at least one question but did not complete the survey) were not included in the final data set.

4.3 Identification and removal of 'satisficers/speeders' from the data set

- Respondents who completed the survey in an extremely short time and/or skipped (refused or said 'don't know'. to) a certain number of questions were considered as speeders. Speeder data were removed from the final data set.

4.3.1 Criteria for assessing poor-quality data

- Two criteria were used to assess poor-quality data:
  1. Seconds per question (SecperQ)
  2. % of responses that were either Refused or Don't Know (%RDK)

- Six user types were created to assess data quality since the respondents were required to answer different series of questions
  1. Cigarette-only smokers
  2. HTP-only users
  3. Dual users of cigarette and HTP
  4. Non-users
  5. Short-term cigarette quitters (QT) using HTP at least weekly (quitted cigarettes within two years)
  6. Short-term cigarette quitters using HTP less than weekly (quitted cigarettes within two years)

4.3.2 Removing poor-quality data to create 'normal' group

- The initial criteria for creating the normal groups for each user type were high and somewhat arbitrary, as the goal was to remove users who might not have been answering responsibly.

Table 4: Percentile cut-offs used to determine users ineligible for the "normal" group

User Group	SecperQ Cutoff (seconds)[a]	% RDK Cutoff [b]	Users Ineligible for Normal Group (N)
Cig- only smoker	0.08	0.2941	1,294
HTP-only user	0.08	0.1884	553
Dual user	0.07	0.2181	620
Non-user	0.09	0.2451	400
Short term QT using HTP at least weekly	0.07	0.2119	91
Short term QT using HTP less than weekly	0.08	0.2651	129

[a] SecperQ cutoff are users with less than or equal to the 20th percentile of their user-type. Because the response and typing speed varied substantially by age, some participants less than or equal to the 20th percentile of their user-type might simply be the younger participants who typed faster than the older participants.

[b] %RDK cutoff are users with greater than the 85th percentile of their user-type

- Frequency distributions by normal user groups were used to create more precise cut-offs for 'poor-quality' responses. 'Faulty points' were assigned based on the normal group percentiles. Respondents were removed from the dataset if they scored too many points.

4.3.3 Assigning points to user types, identifying speeders

- Points were assigned on these bases:

1. SecperQ -- The calculated value was time taken to complete the survey divided by the number of questions answered by the respondent. Very short times suggest poor data quality.

- **5 points** were assigned if the respondent's value was lower than the normal group's 0.5*minimum.
- **3 points** were assigned if the respondent's value fell between the normal group's 0.5*minimum (inclusive) to the 0.75*minimum.
- **2 points** were assigned if the respondent's value fell between the normal group's 0.75*minimum (inclusive) to minimum.
- **All other** values would have a point of 0.

Because SecperQ (time per question) was considered the most important of the 2 data-quality criteria, it was weighted more heavily, so the points assigned here were 2, 3 and 5, rather than 1, 2 and 3 as for the %RDK criterion.

2. %RDK -- The calculated value was the number of questions answered with either "Refused" or "Don't know", divided by the number of questions answered by the respondent. Large numbers represent poor data quality

- **3 points** were assigned if the respondent's value was ≥ to the normal group's 2*maximum.
- **2 points** were assigned if the respondent's value fell between the normal group's 1.33*maximum (inclusive) to 2*maximum.
- **1 point** was assigned if the respondent's value fell between the normal group's maximum to 1.33*maximum
- All other values of %RDK would have a point of 0.

- Points results: The range of possible points was 0 – 8. At the applied cut-off of 4 points, 212 respondents were considered speeders.

Table 5: Number of speeders and non-speeders by user group

User Group	Speeders	Non-speeders	Total
Cigarette-smoker Only	87	1,761	1,848
HTP user Only	47	755	802
Dual user of cigarette and HTP	41	825	866
Non-user	21	595	616
Short term cigarette quitter using HTP at least weekly	4	125	129
Short term cigarette quitter using HTP less than weekly	12	193	205
Total	**212**	**4,254**	**4,466**

- Comparison between respondents speeders and non-speeders:
  - There were more speeders in the 20-29 (9.32%) and 30-39 (7.12%) age groups than in the 40-59 (3.69%) and 60+ (2.39%) age groups.
  - The distribution of speeders across gender was similar with males at 4.62% and females at 4.54%.

4.4 Data cleaning and topline frequencies

- After fieldwork was completed, Rakuten Insight cleaned the data and then transferred the cleaned data to the ITC Project.
- ITC Project analysts completed further data cleaning, weights construction, and initial descriptive analyses, including generating topline frequencies.

4.5 Translation review and verification

- Standard procedures at ITC include validating the translation against the fieldwork-version of the survey. This process is conducted by an independent reviewer fluent in Japanese and English.

4.6 CHERRIES checklist

- The Checklist for Reporting Results of Internet E-Surveys (CHERRIES) is an assessment tool endorsed by the Journal of Medical Internet Research to promote complete and accurate standard reporting guidelines for authors describing internet-based surveys.
- Appendix 3 provides the CHERRIES checklist for the JP4 Survey.

5 Disposition Codes

5.1 Recontact disposition codes for ITC JP4 Survey

Recontact/Cohort Disposition Codes
ITC Japan Survey – Wave 4
Last updated by C. Boudreau on Mar 3, 2021

Notes:
1) Disposition codes are in column B
2) See legend at bottom for type of disposition code (column C)

Recontact/Cohort Disposition Code			
DMC Code	Type*1	Description	Comment
A – Interviewed			
C-A1	P	Selected respondent completes the entire survey; maybe skipping or refusing to answer a few questions	
B – Eligible, non-interview			
C-B10	P or E	Respondents closes his/her web browser or get disconnected during screener	This is an important disposition code, and we expect that a few individuals will fall into this category; distinguishing between C-B10 and C-B19 is not essential
C-B11.1	P	Respondent refuses, can't answer or doesn't know his/her age	
C-B11.2	P	Respondent refuses, can't answer or doesn't know his/her gender	
C-B11.5	P	Respondent refuses, can't answer or doesn't know smoking status	
C-B11.6	P	Respondent refuses, can't answer or doesn't know HTP status	
C-B11.7	P	Respondent refuses, can't answer or doesn't know how long ago he/she quit smoking or use of cigarettes while quit	
C-B13	P	Respondent refuses at consent	
C-B15	P	Age entered by respondent is not close enough to age on file for that same respondent	
C-B19	P or E	Respondent consented and started to answer the survey, but did not complete the survey; closed his/her web browser, got disconnected or quitted	This is an important disposition code, and we expect that a few individuals will fall into this category; distinguishing between C-B10 and C-B19 is not essential
C-B70	S	Invalid email or email bounce back	
C-B72	E	Respondent never logged into system to start the survey (but there was no email bounce back/invalid)	This is an important disposition code, and we expect that quite a few individuals will fall into this category
C-B80	P	Respondent refuses (or doesn't know) to provide required information to derive their stratum	
C-B90	S	Any other reason why interview was not completed	Unlikely to be used, but left in as a precaution
D – Not eligible			
C-D10	P	Respondent is out of the target population	For example, respondent has move and now resides outside the 47 Prefectures of Japan
C-D72	P	Respondent doesn't meet the eligibility criteria on smoking or tobacco use	This disposition code was created for inexperienced smokers (ie, smoked less than 100 cigs) that were recruited at wave 1 or 2, but are no longer eligible at wave 3 because we've changed our criteria.
C-D90	S	Any other reason why respondent is not eligible	Unlikely to be used, but left in as a precaution

Notes:
*1 Type of disposition codes:
P = disposition code programmed into the script
S = Dispositions to be entered by supervisor or other person responsible for fieldwork
E = Dispositions to be derived at the end of fieldwork
HTP = heated tobacco product

5.2 Replenishment disposition codes for ITC JP4 Survey

Replenishment/New Recruits Disposition Codes
ITC Japan Survey – Wave 4
Last updated by C. Boudreau on Mar 3, 2021

Notes:

1) Disposition codes are in column B
2) See legend at bottom for type of disposition code (column C)

Replenishment/New Recruits Disposition Codes			
DMC Code	Type *1	Description	Comment
A – Interviewed			
P-A1	P	Selected respondent completes the entire survey; maybe skipping or refusing to answer a few questions	
B – Eligible, non-interview			
P-B19	P or E	Respondent completed eligibility questions and was deemed to be eligible, then started to answer the survey but did not complete the survey	This is an important disposition code, and we expect that a few individuals will fall into this category; distinguishing between P-B19 and P-C10 **is essential**
P-B90	S	Any other reason why interview was not completed, but eligibility was confirmed by respondent	Unlikely to be used, but left in as a precaution
C – Unknown eligibility, non-interview			
P-C10	P or E	Respondents closes his/her web browser or get disconnected during screener; thus unknown if he/she is eligible	This is an important disposition code, and we expect that a few individuals will fall into this category; distinguishing between P-B19 and P-C10 **is essential**
P-C11.1	P	Respondent refuses, can't answer or doesn't know his/her age; thus unknown if he/she is eligible	
P-C11.2	P	Respondent refuses, can't answer or doesn't know his/her gender	
P-C11.5	P	Respondent refuses, can't answer or doesn't know his/her cigarette smoking status; thus unknown if he/she is eligible	
P-C11.6	P	Respondent refuses, can't answer or doesn't know his/her HTP status; thus unknown if he/she is eligible	
P-C11.7	P	Respondent refuses, can't answer or doesn't know how long ago he/she quit smoking or use of cigarettes while quit; thus unknown if he/she is eligible	
P-C13	P	Respondent refuses at consent	
P-C70	S	Withdrawal and/or unsubscribe	Normally this would also include invalid email and email bounce back, but Rakuten is unable to provide that information at this wave
P-C72	S	Respondent never logged into system to start the survey (but there was no email bounce back/invalid)	This is an important disposition code, and we expect that many individuals will fall into this category
P-C80	P	Respondent refuses (or doesn't know) to provide required information to derive their stratum, and thus sampling weights cannot be computed; since this question is before key screening questions, it is unknown if respondent is eligible	
P-C90	S	Other reason why unknown eligibility	Unlikely to be used, but left in as a precaution
D – Not eligible			
P-D10	P	Respondent is out of the target population	For example, respondent does not reside in any of the 47 Prefectures of Japan
P-D70	P	Respondent is too young (i.e., < 20 years old)	
P-D72	P	Respondent doesn't meet the eligibility criteria on smoking or tobacco use	
P-D80.1	P	Quota for cigarette smokers only is full	
P-D80.2	P	Quota for HTP only users is full	
P-D80.3	P	Quota for dual users is full	
P-D80.4	P	Quota for non-users is full	
P-D90	S	Any other reason why respondent is not eligible	Unlikely to be used, but left in as a precaution

Notes:

*1 Type of disposition codes:

P = disposition code programmed into the script
S = Dispositions to be entered by supervisor or other person responsible for fieldwork
E = Dispositions to be derived at the end of fieldwork
HTP = heated tobacco product

A2.3 *Sub-code for participants coded P-A1: Desktop/Tablet vs. Mobile vs. Undefined

- All records coded P-A1 were further sub-coded per the definitions below.
    - Desktop/Tablet (DskTab) = Respondent completed survey on desktop computer/tablet defined as a device on which the survey appears large-screen formatting, e.g. grid questions show as grids.
    - Mobile (Mob) = Respondent completed survey on mobile device defined as a smaller-screen device on which the JP4 survey questions are reformatted for a mobile experience, e.g. grids are shown as a list of items that can be scrolled down vertically instead of in a grid that populates horizontally.
    - Undefined (Undef) = Respondent completed survey on device not classified as either a Desktop/Tablet or a Mobile.

Notes:
*1 Type of disposition codes:
P = disposition code programmed into the script
S = Dispositions to be entered by supervisor or other person responsible for fieldwork
E = Dispositions to be derived at the end of fieldwork
HTP=Heated tobacco product

6 Cooperation and Response Rates

Remarks:
1) Response rates, cooperation rates and other figures/numbers in this spreadsheet are for the JP4 replenishment sample only.
2) Because respondents were recruited via Rakuten's web panel, the response and cooperation rates in this spreadsheet are not comparable to those from ITC surveys conducted via phone/RDD or face-to-face.

	Freq	%
A – Interviewed		
Total (interviewed)*1	1,729	5.1%
B – Eligible, but not interviewed		
Refusal/breaks off	100	0.3%
Other	0	0.0%
Total (eligible but not interviewed)	100	0.3%
C – Unknown if eligibility (not interviewed)		
Logged into system to start survey (once or more)	3,336	9.9%
Estimated number of eligible and quota not full*	666	2.0%
Estimated number of not eligible or quota full*3	2,670	7.9%
Never logged into system to start survey	19,151	56.8%
Estimated number of eligible and quota not full	3,823	11.3%
Estimated number of not eligible or quota full	15,328	45.5%
Total (unknown if eligible)	22,487	66.7%
D – Not eligible		
Out of sample	7	0.0%
Respondent is not eligible	248	0.7%
Quota full	6,239	18.5%
Other	2,904	8.6%
Total (not eligible)	9,398	27.9%
Total sample with final disposition	33,714	100%
Estimated eligibility rate*4	88.1%	
Estimated proportion for which quota was full*5	77.3%	
Response rate*6	27.4%	
Cooperation rate*7	94.5%	

Notes:

1.The total number of completed interviews (i.e., 1,729) includes some respondents that were then removed from the final dataset because they were deemed to be fraudulent/speeder
If those respondents were excluded from the above table, both the response rate and cooperation would be slightly lower than their current values
2.Estimated number of respondents that would have been eligible and for which the corresponding quota would not have been full
Formula: row 18 x row 35 x (1 - row 36), rounded to the nearest integer
3.Formula: row 18 - row 19
4.Estimated proportion of individuals that were eligible (i.e., age 20 & older and meet smoking/tobacco use criteria)
Formula: 1 - row 28 / (row 10 + row 15 + row 28)
5.Estimated proportion of individuals that were terminated because the corresponding quota was full
Formula: row 29 / (row 10 + row 15 + row 29)
6.The response rate is the proportion of selected respondents who complete the survey; i.e., the number of eligible respondents who completed the survey divided by the estimated number of eligible respondents that were selected. In other words, the response rate accounts for the fact that numerous individuals could not be contacted or screened for eligibility, whereas the cooperation rate does not.
Formula: row 10 / (row 10 + row 15 + row 19 + row 22)
7.The cooperation rate is the proportion of eligible respondents (i.e., those who have completed all eligibility questions and have been found to be eligible) who complete the survey.
Formula: row 10 / (row 10 + row 13)
The above formula for the cooperation rate is the same as the AAPOR COOP4 formula; see https://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx
Contrary to the response rates, the cooperation rates are comparable across countries.

APPENDIX 1: ALLOCATION (PER STRATUM)

Quotas - ITC Japan - Wave 4

Created by C. Boudreau and Y. Li on May 12, 2021

Last updated by C. Boudreau on Jun 2, 2021

Notes:

1) All quotas in this spreadsheet are for individuals that are to be recruited at wave 4; individuals recruited at waves 1-3 that are recontacted at wave 4 do not count towards the quota
2) All quotas must be meet or slightly exceeded by the end of fieldwork
3) The quotas to be programmed are in orange; the other numbers in the same column as quotas are totals that do not need to be programmed
4) Cells in gray are meant to be easily modified (just type in a new number); all other cells/calculations are automated and thus those cells shouldn't be modified
5) Definitions: In the ITC Japan Survey,
   a) An exclusive cigarette smoker is defined as someone who i) has smoked at least 100 cigarettes in their lifetime, ii) who smokes cigarettes at least weekly, AND iii) uses heated tobacco products (HTP) less than weekly or not at all.
      Alternatively, an exclusive cigarette smoker is also defined as someone who i) has smoked at least 100 cigarettes in their lifetime, ii) who quit smoking cigarettes within the last 2 years, AND iii) uses heated tobacco products (HTP) less than weekly or not at all.
   b) An exclusive HTP user is defined as someone who i) uses HTP at least weekly, AND ii) smokes cigarettes less than weekly or not at all.
   c) A dual user is defined as someone who i) has smoked at least 100 cigarettes in their lifetime, ii) smokes cigarettes at least weekly, AND iii) uses
      Alternatively, a dual user is also defined as someone who i) has smoked at least 100 cigarettes in their lifetime, ii) has smokes cigarettes at least weekly in the past, but quit doing so within the last 2 years, AND iii) has used HTP at least weekly in the past, but quit doing so within the last 2
   d) A non-user is defined as someone who i) has not smoked cigarettes at least weekly within the last 2 yers, AND ii) has never used HTP at least

				Initial estimates and quotas						Updated/revised quotas		
				W3 eligible respondents *3	timated # of transition		Estimated eligible after transitions *5	Estimated # of recontacts *6	Quotas *7	Actual # recontacted *8	Actual # of replenishment *9	Quotas *10
		% *1	Total *2		Transit out	Transit in						
Exclusive cigarette smokers												
Males	20-39	12.8	255	293	76	101	318	190	65	140	115	255
	40 & older	57.1	1,142	1,118	291	160	987	592	550	869	273	1,142
	Total	69.9	1,397	1,411	367	261	1,305	782	615			
Females	20-39	6.2	125	140	36	38	142	85	40	63	62	125
	40 & older	23.9	478	475	123	53	405	243	235	329	149	478
	Total	30.1	603	615	159	91	547	328	275			
Total Japan		100.0	2,000	2,026	526	352	1,852	1,110	890			2,000
Exclusive HTP users												
Males	20-39	40.3	403	377	211	51	217	130	273	142	261	403
	40 & older	16.2	162	355	199	107	263	157	5	162	0	162
	Total	56.5	565	732	410	158	480	287	278			
Females	20-39	19.3	193	123	69	20	74	44	149	51	142	193
	40 & older	24.2	242	130	73	37	94	56	186	91	151	242
	Total	43.5	435	253	142	57	168	100	335			
Total Japan		100.0	1,000	985	552	215	648	387	613			1,000
Dual users												
Males	20-39	24.2	242	278	132	160	306	183	59	161	81	242
	40 & older	52.0	520	487	231	276	532	319	201	369	151	520
	Total	76.2	762	765	363	436	838	502	260			
Females	20-39	10.4	104	103	49	62	116	69	35	61	43	104
	40 & older	13.4	134	138	65	119	192	115	19	109	25	134
	Total	23.8	238	241	114	181	308	184	54			
Total Japan		100.0	1,000	1,006	477	617	1,146	686	314			1,000
Non-users												
Males	20-39	16.1	80	68	6	113	175	105	0 *11	39	41	80
	40 & older	22.0	110	91	8	185	268	160	0 *11	79	31	110
	Total	38.1	190	159	14	298	443	265	0			
Females	20-39	18.1	90	91	8	41	124	74	16	42	48	90
	40 & older	43.8	219	215	19	71	267	160	59	185	35	220
	Total	61.9	310	306	27	112	391	234	75			
Total Japan		100.0	500	465	41	410	834	499	75			500
Grand total Japan			4,500	4,482	1,596	1,594	4,480	2,682	1,892			4,500

Notes

*1 Percentages of the population by gender/age groups; obtained from the 2020 Japan Society and Tobacco Internet Study (JASTIS), see file quotasJP-w3-JASTIS.sas

*2 This is the total number of respondents to be interviewed at wave 4
Formula: column D x Total Japan (i.e., rows 29, 38, 47 & 56)

*3 Number of wave 3 respondents (by gender/age groups) that are eligible to be recontacted at wave 4; some will change smoking/HTP use status and some will be lost to follow-up

*4 Estimated number of respondents that are expected to transition (i.e., change smoking/HTP use status) between waves 3 and 4

*5 Formula: column F - column G (transitions out) + column L (transitions in)

*6 Estimated number of respondents that will be successfully recontacted at wave 4
This is based on an expected/estimated retention 60.00%

*7 Estimated number of respondents (by gender/age groups) to be recruited at wave 4

*11 Based on how many respondents completed the wave 3 survey, the estimated number of transitions and the estimated retention rate, it is expected that number of recontact respondents will already exceed the total number

APPENDIX 2: ITC JP4 Email Invitation and Reminder Templates
23 February 2021, ORE #22508/31428

2.1 ITC JP4 Recontact Email Invitation Template

Email Subject: Follow Up to Special "5-Year Survey": Users and non-users of tobacco products [PROJECT NUMBER]

Email Body:
[NAME],

Thank you very much for participating in the special "5-Year Survey": Users and non-users of tobacco products in **May to June 2020**.

We would like to ask you a follow-up survey today.

□□
You are part of a special group of participants across Japan (and in 28 other countries throughout the world) that is invited to participate in this research that will help policymakers and health professionals in forming future policies and programs on tobacco.

The International Tobacco Control (ITC) Japan Survey is a research of people's knowledge, attitudes, and behaviours about tobacco products. Both "users" and "non-users" of tobacco products are needed for this study — a survey that will be conducted about once a year over five years.
□□

This is a continuing research study that involves:

- participating in this 25- to 30-minute survey in July to August 2021, and
- if you complete the survey this time, being invited to participate in follow-up surveys every 12- months.

If you qualify and fully complete this survey, it's extremely important that you come back and join our Special "5-Year Survey" again in 12 months!

Please join the below survey.

****

- **Survey title:**
  Special 5 Year Survey

- **Survey URL: [LINK TO SURVEY]**
  - Answering a survey can only be done once per person.

- Survey length: About 25-30 minutes. (Actual length may vary depending on contents)
  - Incentive: [**Rakuten min – max incentive**].
- Closing Date: Once the required number of survey completions is reached.
- Incentive Awarded Date：Max 60 days after completing the survey
  - Depending on the survey, it may take longer than the above timing.

****

Please take note while answering surveys:

We ask that you answer truthfully

- Please make sure to review each question and answer choice carefully to make sure your answers are accurately selected.

Internet Environment

- We suggest the below internet environment: [xx]

Privacy Policy

- We strictly adhere to the Privacy Policy in the Member Agreement. Please do not disclose any information you acquired in the survey to any third party, including posts to bulletin boards and blogs.

Surveys that are not shown of your My Page

- Surveys that were emailed more than 2 weeks ago will not appear on your My Page. Please click the URL to answer the survey.
    - If the URL is cut off based on your email setting, please copy the whole URL and paste it into your browser address bar.
    - You may not enter the survey if the survey is already closed.

Other

- Answering a survey can only be done once per person
- You will not be able to reply to this email.
- Please go to FAQ for any questions.

2.2 ITC JP4 Replenishment Email Invitation Template

Email Subject Invitation to a Special “5-Year Survey”: Users and non-users of tobacco products [PROJECT NUMBER]

Email Body:
[NAME],

Thank you very much for participating in the Rakuten surveys.

Warm welcome to the Special "5-Year Survey”!

□□
You are part of a special group of participants across Japan (and in 28 other countries throughout the world) that is invited to participate in this research that will help policymakers and health professionals in forming future policies and programs on tobacco.

The International Tobacco Control (ITC) Japan Survey is a research of people’s knowledge, attitudes, and behaviours about tobacco products. Both "users" and "non-users" of tobacco products are needed for this study — a survey that will be conducted about once a year over five years.
□□

This is a continuing research study that involves:
- participating in this 25- to 30-minute survey during the next few days, and
- if you complete the survey this time, being invited to participate in follow-up surveys every 12months.

If you qualify and fully complete this survey, it's extremely important that you come back and join our Special “5-Year Survey" again in 12 months!

Please join the below survey.

****
- **Survey title:**
  Special 5 Year Survey
- **Survey URL: [LINK TO SURVEY]**
  - Answering a survey can only be done once per person.

- Survey length: About 25-30 minutes. (Actual length may vary depending on contents)
  - Incentive: [**Rakuten min – max incentive**].

- Closing Date: Once the required number of survey completions is reached.
- Incentive Awarded Date： Max 60 days after completing the survey
  - Depending on the survey, it may take longer than the above timing.

****

2.3 Rakuten email reminder Template #1:
Email Subject: Invitation to a Special “5-Year Survey”: Users and non-users of tobacco products [PROJECT NUMBER] - Reminder

Email Body:

[NAME],

Thank you very much for participating in the Rakuten surveys.

We recently sent you an invitation to the Special "5-Year Survey”.

Your participation is important to us -- we hope you will agree to participate in our study!

□□
You are part of a special group of participants across Japan (and in 28 other countries throughout the world) that are invited to participate in this research that will help policymakers and health professionals in forming future policies and programs on tobacco.

The International Tobacco Control (ITC) Japan Survey is a research of people’s knowledge, attitudes, and behaviours about tobacco products. Both "users" and "non-users" of tobacco products are needed for this study — a survey that will be conducted about once a year over five years.
□□
This is a continuing research study that involves:

- participating in this 25- to 30-minute survey during the next few days, and
- if you complete the survey this time, being invited to participate in follow-up surveys every 12 months.

If you qualify and fully complete this survey, it's extremely important that you come back and join our Special “5-Year Survey" again in 12 months!

We'd appreciate you taking the survey at your earliest convenience by using the web address below. If you have already completed the survey, thank you and please discard this email.

Please join the below survey.

****

- **Survey title:**
  Special 5 Year Survey

- **Survey URL: [LINK TO SURVEY]**
  - Answering a survey can only be done once per person.

- Survey length: About 25-30 minutes. (Actual length may vary depending on contents)
  - Incentive: [**Rakuten min – max incentive**].

- Closing Date: Once the required number of survey completions is reached.
- Incentive Awarded Date：Max 60 days after completing the survey
  - Depending on the survey, it may take longer than the above timing.

****

2.4 Rakuten email reminder Template #2:
Email Subject: Special Survey "5 Year Tobacco Panel" [PROJECT NUMBER] - Reminder

Email Body:

[NAME],

Thank you very much for participating in the Rakuten surveys.

We recently sent you an invitation to the Special "5-Year Survey”.

Please participate in our study – this study will be closing soon!

□□
You are part of a special group of participants across Japan (and in 28 other countries throughout the world) that are invited to participate in this research that will help policymakers and health professionals in forming future policies and programs on tobacco.

The International Tobacco Control (ITC) Japan Survey is a research of people’s knowledge, attitudes, and behaviours about tobacco products. Both "users" and "non-users" of tobacco products are needed for this study — a survey that will be conducted once a year over five years.
□□

This is a continuing research study that involves:

- participating in this 25- to 30-minute survey during the next few days, and
- if you complete the survey this time, being invited to participate in follow-up surveys every 12 months.

If you qualify and fully complete this survey, it's extremely important that you come back and join our Special “5-Year Survey" again in 12 months!

We'd appreciate you taking the survey at your earliest convenience by using the web address below. If you have already completed the survey, thank you and please discard this email.

Please join the below survey.
****

- **Survey title:**
- Special 5 Year Survey

- **Survey URL: [LINK TO SURVEY]**
  - Answering a survey can only be done once per person.

- Survey length: About 25-30 minutes. (Actual length may vary depending on contents)
  - Incentive: [**Rakuten min – max incentive**].

- Closing Date: Once the required number of survey completions is reached.
- Incentive Awarded Date：Max 60 days after completing the survey
  - Depending on the survey, it may take longer than the above timing.

****

APPENDIX 3: ITC JP4 Survey Summary – Checklist for Reporting Results of Internet E-Surveys (CHERRIES)

<table>
<tr><td></td><td>Checklist for Reporting Results of Internet E-Surveys (CHERRIES)</td></tr>
<tr><td>Item Category & Checklist Item</td><td>ITC Japan Wave 4 (JP4) Survey compliance with CHERRIES</td></tr>
<tr><td>Design</td><td></td></tr>
<tr><td>Target population</td><td>The ITC Japan Wave 4 sample was designed to be nationally representative of Japanese smokers, HTP users, dual users (of cigarettes and HTP), and never or non-smokers.</td></tr>
<tr><td>Sampling frame</td><td>The ITC Japan Wave 4 recontact sample comprised the following user types.

JP4 recontact subsamples with definitions
<table>
<tr><td>Subsample group</td><td>Definition</td></tr>
<tr><td rowspan="2">Exclusive Cigarette Smoker and Cigarette Quitter</td><td>Cigarette Smoker
• Smokes cigarettes at least weekly AND has smoked 100 cigarettes in lifetime.
• Uses HTPs less than weekly or not at all</td></tr>
<tr><td>Cigarette Quitter
• Has smoked at least weekly in the past and smoked 100 cigs and doesn’t smoke at least weekly now.
• Has never used HTP at least weekly.</td></tr>
<tr><td rowspan="2">Exclusive HTP User and HTP Quitter</td><td>HTP user
• Smokes cigs less than weekly or not at all
• Uses HTPs at least weekly.</td></tr>
<tr><td>HTP Quitter
• Has never smoked cigs at least weekly.
• Has used HTP at least weekly in the past and doesn’t use HTP at least weekly now.</td></tr>
<tr><td rowspan="2">Cigarette-HTP Dual User and Dual Quitter</td><td>Dual user
• Smokes cigarettes at-least-weekly AND has smoked 100 cigarettes in lifetime)
• Uses HTP at least weekly.</td></tr>
<tr><td>Dual quitter
• Has smoked at least weekly in the past and smoked 100 cigs and doesn’t smoke at least weekly now.
• Has used HTP at least weekly in the past and does not use HTP at least weekly now.</td></tr>
<tr><td>Never or Non-user</td><td>Never-/non-user
• Has not smoked cigarettes at least weekly
• Has not used HTP at least weekly.</td></tr>
</table>
Rakuten Insight provided the following description of their panel(s): The JP4 survey was conducted with Rakuten’s proprietary online panel in Japan. The online panel is actively managed in-house with a dedicated panel management team in Tokyo, and utilized for market research purposes only. Recruitment for the panel is conducted daily, tapping into users of Rakuten services (e.g., e-commerce, credit cards, insurance, mobile services, etc.)</td></tr>
</table>

	Checklist for Reporting Results of Internet E-Surveys (CHERRIES)
Item Category & Checklist Item	***ITC Japan Wave 4 (JP4) Survey compliance with CHERRIES***
	as well as other online resources such as affiliates, email and banner recruits in order to maintain a panel as consistent as possible with the general population. Panelists are pre-profiled with a series of questions which in turn can be used as pre-targeting variables (e.g. smoking, HTP usage, etc.). Panelists receive email invitations and also have the option of logging into their proprietary panel site to access the survey they are invited to participate in. Details available at: https://insight.rakuten.com//
Sample	The ITC Japan Wave 4 sample is provided in Table 2 in Section 2.1.
IRB (Institutional Review Board) approval and informed consent process	
IRB approval	All survey procedures and materials were cleared by a University of Waterloo Research Ethics Committee (REB#22508/31428).
Informed consent	All participants were part of a Rakuten Insight panel and had an existing agreement with Rakuten Insight to complete surveys for awards. For the ITC JP4 Survey, there were two survey phases: recontact phase - pre-identified respondents from the JP3 Survey were invited by email invitation to complete the survey; replenishment phase - pre-identified panelists from the same (JP1,2&3) sampling frame were invited to complete the survey to replace JP3 cohort members either lost to follow up or having changed their tobacco use status. The email invitation described the topic of the survey and that the overall project involved one or more follow-up survey(s) in the future, the length of the survey, and the incentive value. Upon entering the survey, the first few screens reiterated the information above and also explained who was conducting the research, that respondents could skip questions or withdraw, and that data were strictly confidential. The ethics contact information was provided.
Data protection	No personal identifying information data for panelists was stored in the survey data. All survey response data files were treated as confidential and maintained on secure servers. Data were transferred using the University of Waterloo secure system. Rakuten employed strict data protection & security measures to safeguard panelist information. Examples include protecting data by Encryption Communication Technologies (SSL), utilizing Site Authentication Systems to prevent unauthorized access, appointing a privacy information officer to manage panelist information, secure data transfer, and adherence to local laws with respect to privacy and confidentiality as stipulated in the Japan Marketing Research Association's guidelines of personal information protection and codes of conduct.
Development and pre-testing	
Development and testing	The survey content and logic were developed by a team of international tobacco control and survey design experts through a structured and iterative process of consultation and revision. Survey content was developed in English and then translated into Japanese. The Japanese translation was reviewed and verified by bilingual members of the research team. The survey specifications were then sent to an experienced ITC Survey Management team for operationalization and further extensive systematic review and

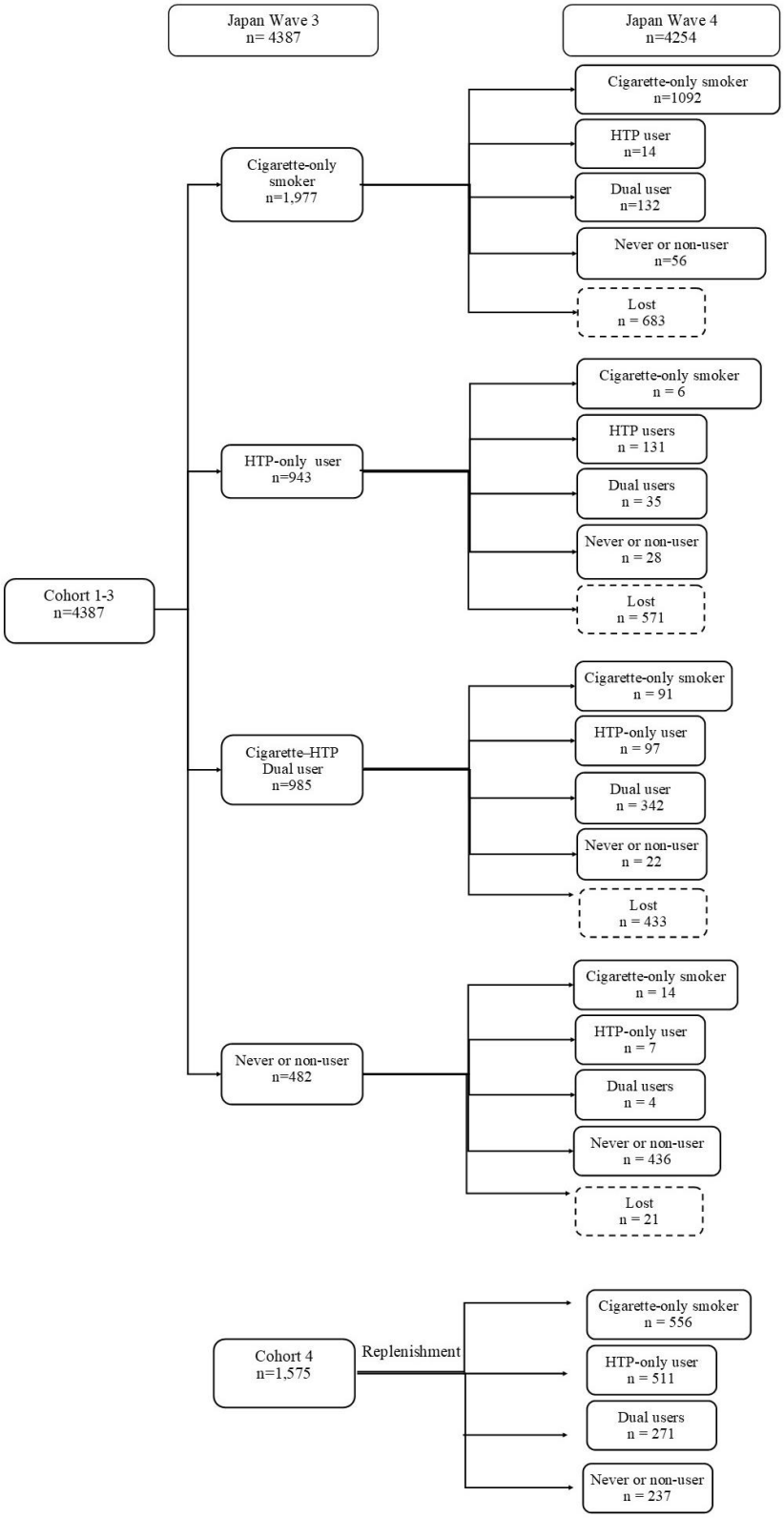
	Checklist for Reporting Results of Internet E-Surveys (CHERRIES)
Item Category & Checklist Item	***ITC Japan Wave 4 (JP4) Survey compliance with CHERRIES***
	refinement to ensure survey logic, question wording, response options, and all other survey elements were refined and cross-referenced for consistency, clarity, and accuracy. The ITC JP4 Survey was programmed using Confirmit software by Rakuten Insight via a collaborative process between the programmer/fieldwork team and the investigator team to refine survey content as needed to ensure a good survey experience and accurate data collection. The programmed instrument was then systematically and comprehensively tested for usability and technical functionality by the survey firm, as well as by the ITC team.
Recruitment process and description of the sample having access to the questionnaire	
Open survey versus closed survey	The survey was a closed, password-protected survey, in which the respondents could access only their own unique survey record via a predetermined link and/or via the members' private account on the Rakuten Panel site. Each record was associated with a unique ID number. Email invitations included the name of the intended respondent.
Contact mode	Panelists were contacted in the standard method (standard Rakuten Insight survey invitation with some minor adaptations for the ITC JP4 study) from Rakuten.
Advertising the survey	Only panelists who were pre-identified as being candidates for the survey were invited via a study-specific email invitation.
Survey administration	
Web/E-mail	The ITC JP4 Survey was a web survey hosted by Rakuten Insight on their secure server.
Context	The survey could only be accessed by participants who had been specifically invited to participate. All of the persons invited to the survey were Rakuten Insight Japan panelists.
Mandatory/voluntary	The survey was voluntary. Up to two reminder emails were sent to those who had not submitted their completed survey in order to maximize the response.
Incentives	All panelists were given the Rakuten Insight standard number of points upon completion of the survey. Recontacts who completed the JP4 Survey were given an additional bonus of total $14 USD regardless of tobacco use status.
Time/Date	Data collection: July 3 to August 22, 2021
Randomization of items or questionnaires	None of the questionnaire items were randomized. This was intentional to prevent any differential priming of respondents.
Adaptive questioning	The ITC JP4 Survey used adaptive questioning based on information that the respondent had provided during the survey or to ITC/Rakuten Insight.
Number of Items	For most of the survey, one item (i.e., question) was asked per page, with the exception of 'question series'. Each 'question series' consisted of multiple items on one screen (i.e., as grids or a scroll-down format for mobile devices).
Number of screens (pages)	The **TOTAL** number of screens (pages) applicable to the JP4 web survey was over 368 screens. However, no respondent would ever have been exposed to all these screens due to routing and filtering of questions for different samples, user types, and response patterns throughout the survey. Each respondent would have seen significantly fewer screens in total, and the number of screens seen across respondents would have varied significantly depending on their sample source, user types, and response patterns.

	Checklist for Reporting Results of Internet E-Surveys (CHERRIES)
Item Category & Checklist Item	***ITC Japan Wave 4 (JP4) Survey compliance with CHERRIES***
Completeness check	Respondents had to select a response to every survey item in order to progress to the next screen but were able to choose 'Prefer not to answer' for any question. The survey contained essential questions that were necessary to determine eligibility. If a respondent was unwilling or unable to answer an essential question, then the respondent was shown a note explaining that they would be unable to continue the survey if they did not provide a response. This explanatory note appeared on the same screen as the essential question.
Review step	Respondents were able to review and change their answers at any point, up until they formally submitted their survey. At the end of the survey, the respondent was prompted to hit the submit button in order to submit their completed responses.
Response rates	
Unique site visitor	A unique visitor was defined based on the respondent's unique ID number. Each record had a unique ID number and was assigned to one predetermined respondent.
View rate (Ratio of unique survey visitors/unique site visitors)	Not provided
Response rate	27.4%
Cooperation rate	94.5%
Preventing multiple entries from the same individual	
Cookies used	Cookies were not used to identify multiple entries, but they were used as part of the survey software.
IP check	The IP address of the respondent computer was not used to identify potential duplicate entries from the same user. Respondents could access their own unique survey record only (via a direct link with an embedded unique ID number). Respondents were allowed to stop and re-start the survey during the period that fieldwork was open up until the point of formally choosing to submit their data (by choosing this option at the end of the survey).
Log file analysis	The log file was not used to identify multiple entries.
Registration	As described in the "IP check" field above, respondents could access their own unique survey record only (via a direct link with an embedded unique ID number). Respondents were allowed to stop and resume the survey during the period that fieldwork was open up until the point of formally submitting their data (by choosing this option at the end of the survey). Thus, respondents were able to return to where they last left off, unless there was a specific section in the survey where they can over-ride their previous answers.
Analysis	
Handling of incomplete questionnaires	Only completed surveys were included in the final data set released for analyses.

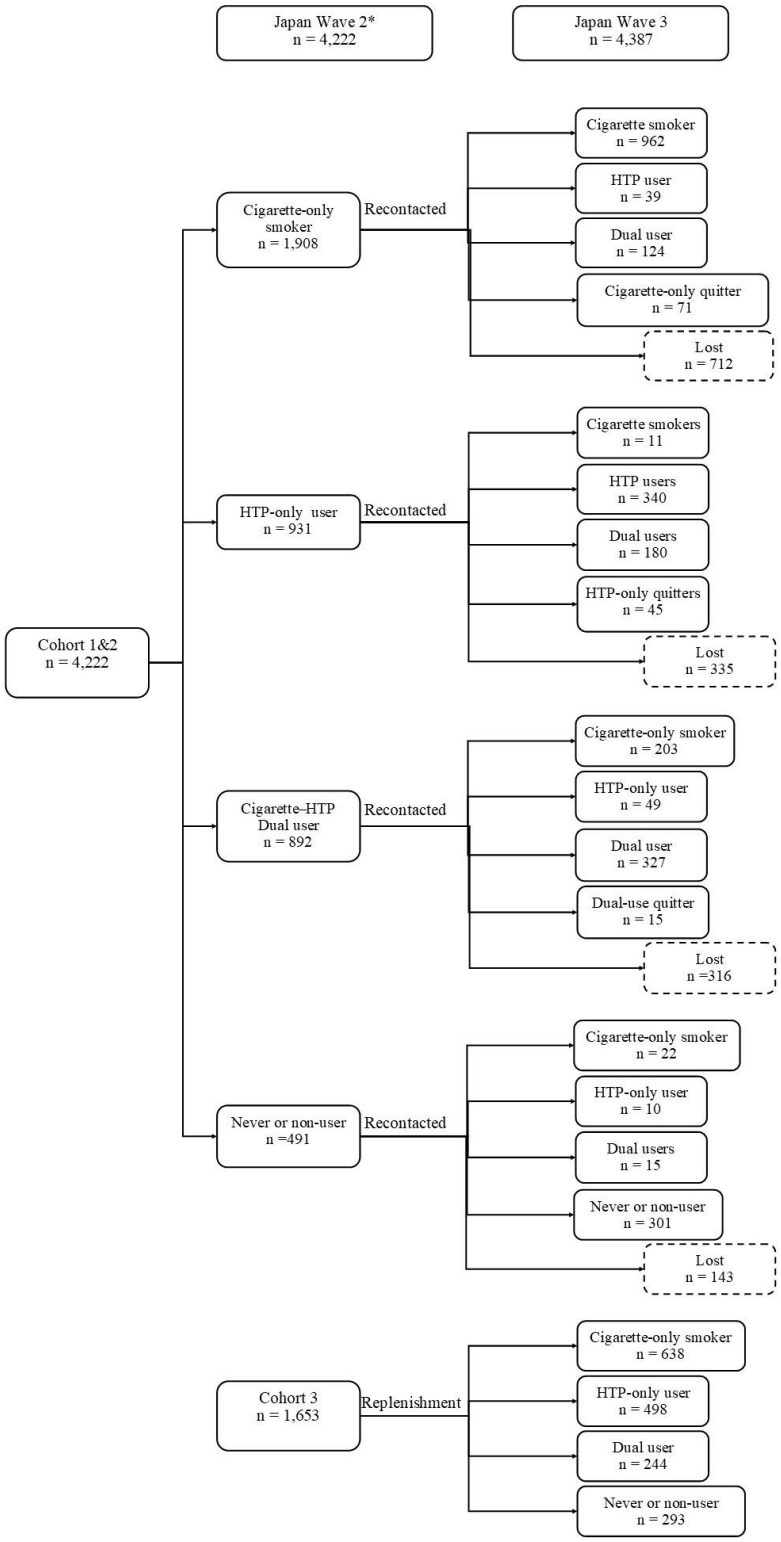
	Checklist for Reporting Results of Internet E-Surveys (CHERRIES)
Item Category & Checklist Item	***ITC Japan Wave 4 (JP4) Survey compliance with CHERRIES***
Questionnaires submitted with an atypical timestamp	Two criteria were used to assess poor-quality data: 1) Seconds per question (SecperQ) 2) % of responses that were either Refused or Don't Know (%RDK). 212 respondents were deleted from the final dataset due to very extreme values for both of these variables: SecperQ and/or %RDK of their user-group. Please refer to Section 5.3 in the ITC JP1 Technical Report for further details on criteria for poor data quality.
Statistical correction	Cross-sectional survey weights for different analyses were constructed for the final data set. For the cross-sectional weights, respondents were first divided into four broad user types: 1) cigarette-only users, 2) HTP-only users, 3) HTP/cigarette dual users, and 4) non-users. The weights were then calibrated on the following cross-tabs using a raking algorithm: user group x gender, user group x age group, user group x geographic region. This weight calibration was done using data from the 2020 JASTIS Survey. Please refer to Appendix 6 in the ITC JP4 Technical Report for further details.

APPENDIX 4: Flowchart of Number of Recontact, Lost, or Replenished Respondents between JP3&JP4, JP2&JP3, and JP1&JP2

4a. JP3-JP4

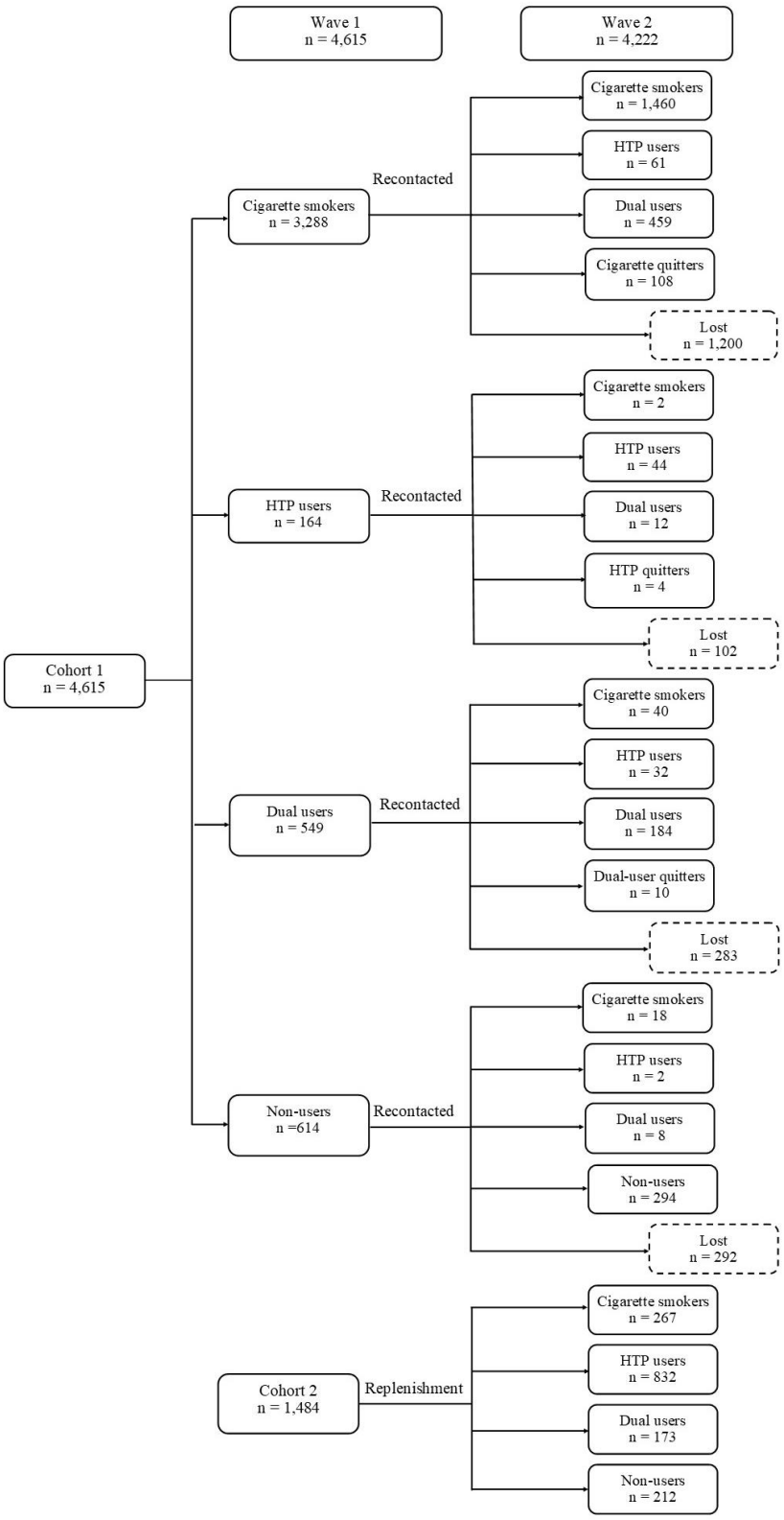


4b. JP2-JP3



* At JP2, we defined cigarette smokers as those smoke cigarettes at least monthly AND has smoked 100 cigarettes in lifetime, cigarette quitter as those smoked cigarrettes at least monthly and has smoked 100 cigarrettes in life time. When preparing the JP2 recontact data, the JP3 defintion of smokers and quitter were applied to reclassify JP2 exclusive smoker subsample to endure the difintion is the same between JP2 and JP3.

4c. JP1-JP2


Wave 1
n = 4,615
Wave 2
n = 4,222
Cohort 1
n = 4,615
Cigarette smokers
n = 3,288
Recontacted
Cigarette smokers
n = 1,460
HTP users
n = 61
Dual users
n = 459
Cigarette quitters
n = 108
Lost
n = 1,200
HTP users
n = 164
Recontacted
Cigarette smokers
n = 2
HTP users
n = 44
Dual users
n = 12
HTP quitters
n = 4
Lost
n = 102
Dual users
n = 549
Recontacted
Cigarette smokers
n = 40
HTP users
n = 32
Dual users
n = 184
Dual-user quitters
n = 10
Lost
n = 283
Non-users
n =614
Recontacted
Cigarette smokers
n = 18
HTP users
n = 2
Dual users
n = 8
Non-users
n = 294
Lost
n = 292
Cohort 2
n = 1,484
Replenishment
Cigarette smokers
n = 267
HTP users
n = 832
Dual users
n = 173
Non-users
n = 212

Appendix 5: JP4 Recontact Sample Subgroup Classification and Methodology for Determining Replenishment Quota Specifications

Preamble: This appendix summarizes the sampling plan for the ITC JP4 Survey, by subgroup and Recontact vs. Replenishment status. It then provides the methodology and specifications for classifying the Recontact respondents (using FR309v and HN309v) in order to determine the number of Replenishment recruits necessary to achieve the target sample.

Part A: Definition of recontact and replenishment sample subgroups, please review Table 1.

Part B: Classifying Replenishment sample into Definition Groups (Def 1, Def 2, etc.) and Subsample Groups (SG).
Replenishment respondents will be classified into Definition Groups based on their responses to FR309v and HN309v.

Table 6: JP4 Replenishment statuses by Definition Groups (Def 1, Def 2, etc.), where

- SG1 includes Def 1 (exclusive cigarette smoker) and Def 2 (exclusive cigarette quitter),
- SG2 includes Def 3 (exclusive HTP user) (Def 4 not applicable to Replenishment sample),
- SG3 includes Def 5 (dual cigarette smoker-HTP user) and Def 6 (dual cigarette-HTP quitter), and
- SG4 includes Def 7 (non-user).

			HTP status (HN309v)			
			User	**LTW* (past unknown)**	**Quitter**	**Always LTW**
			1-2	**3**	**4**	**5-7**
Cig status (FR309v)	**User**	**1-2**	Dual user (Def 5)	Cigarette only (Def 1)	Cigarette only (Def 1)	Cigarette only (Def 1)
	LTW *(past unknown)	**3**	HTP only (Def 3)	Non-user (Def 7)	X Not eligible (Def 4)	Non-user (Def 7)
	Quitter ≥ 2 Years	**4-7**	HTP only (Def 3)	Cigarette quitter (Def 2)	Dual quitter (Def 6)	Cigarette quitter (Def 2)
	Quitter >2 Years Always LTW	**8-9**	HTP only (Def 3)	Non-user (Def 7)	X Not eligible (Def 4)	Non-user (Def 7)

*Note: LTW = Less than weekly

Part C: Classifying Replenishment sample into Definition Groups (Def 1, Def 2, etc.) and Subsample Groups (SG).
Replenishment respondents will be classified into Definition Groups based on their responses to FR309v and HN309v .

Table 7: JP4 Replenishment statuses by Definition Groups (Def 1, Def 2, etc.), where

- SG1 includes Def 1 (exclusive cigarette smoker) and Def 2 (exclusive cigarette quitter),
- SG2 includes Def 3 (exclusive HTP user) (Def 4 not applicable to Replenishment sample),
- SG3 includes Def 5 (dual cigarette smoker-HTP user) and Def 6 (dual cigarette-HTP quitter), and
- SG4 includes Def 7 (non-user).

			HTP status (HN309v)			
			User	**LTW* (past unknown)**	**Quitter**	**Always LTW**
			1-2	**3**	**4**	**5-7**
Cig status (FR309v)	**User**	**1-2**	Dual user (Def 5)	Cigarette only (Def 1)	Cigarette only (Def 1)	Cigarette only (Def 1)
	LTW *(past unknown)	**3**	HTP only (Def 3)	Non-user (Def 7)	X Not eligible (Def 4)	Non-user (Def 7)
	Quitter ≥ 2 Years	**4-7**	HTP only (Def 3)	Cigarette quitter (Def 2)	Dual quitter (Def 6)	Cigarette quitter (Def 2)
	Quitter >2 Years Always LTW	**8-9**	HTP only (Def 3)	Non-user (Def 7)	X Not eligible (Def 4)	Non-user (Def 7)

*Note: LTW = Less than weekly

Part D: Classifying Replenishment respondents into sample subgroups

Replenishment respondents will be classified into subsample groups based on their current tobacco user status determined by both FR309v and HN309v . These two elements will inform a Replenishment subgroup variable.

Tables 8 and 9 show the subgroup classification as a function of respondents' JP4 sample subgroup status.

Table 8: JP4 Recontact statuses by Definition Groups (Def 1, Def 2, etc.), where

- SG1 includes Def 1 (exclusive cigarette smoker) and Def 2 (exclusive cigarette quitter),
- SG2 includes Def 3 (exclusive HTP user) and Def 4 (exclusive HTP quitter),
- SG3 includes Def 5 (dual cigarette smoker-HTP user) and Def 6 (dual cigarette-HTP quitter), and
- SG4 includes Def 7 (non-user).

			HTP status (HN309v)			
			User	**LTW* (past unknown)**	**Quitter**	**Always LTW**
			1-2	**3**	**4**	**5-7**
Cigarette status (FR309v)	**User**	**1-2**	Dual user (Def 5)	Cigarette only (Def 1)	Cigarette only (Def 1)	Cigarette only (Def 1)
	LTW* (past unknown)	**3**	HTP only (Def 3)	Non-user (Def 7)	HTP quitter (Def 4)	Non-user (Def 7)
	Quitter	**4-8**	HTP only (Def 3)	Cigarette quitter (Def 2)	Dual quitter (Def 6)	Cigarette quitter (Def 2)
	Always LTW	**9**	HTP only (Def 3)	Non-user (Def 7)	HTP quitter (Def 4)	Non-user (Def 7)

*Note: LTW = Less than weekly

Part F: Calculating the Replenishment subgroup targets:

If the sample subgroup targets cannot be achieved with the Recontact sample, then the Replenishment target count for Subgroup 1 =2000-[Recontact # of SG1 ; and the Replenishment target count for Subgroup 2 =1000 -[Recontact # of SG2]; etc. See Table 9.

Table 9. JP4 Replenishment targets for each subgroup as a function of FR309v and HN309v

JP4 Sample Subgroup (SG) description (Column 1)	Sample SG target (Column 2)	Recontact completes per FR309V and HN309v response (Column 3)	Replenishment Target (Column 4)
SG1: Exclusive smokers /cigarette quitters	2000	Def 1: Smokers: # of Recontact completes FR309V=1-2 and HN309v=3-7 Def 2: Cig quitters:# of Recontact completes with FR309v=4-8 and HN309v=3, 5-7	=2000-[count in Column 3]; allocated to 4 strata as described in "quotasJP-w4-[date]". Initial predicted sample size =706
SG2: Exclusive HTP users/ HTP quitters	1000*	Def 3: HTP only users: # of Recontact completes with FR309v=3-9 -and HN309v=1-2 Def 4: HTP quitters: # of Recontact completes FR309v=3, 9 and HN309v=4	=1000-[count in Column 3]; allocated to 4 strata as described in "quotasJP-w4-[date]". Initial predicted sample size =322.
SG3: Dual cigarette-HTP users/ dual quitters	1000	Def 5: Dual users: # of Recontact completes with FR309v=1-2 and HN309v=1-2 Def 6: Dual quitters: # of Recontact completes with FR309v=4-8 and HN309v=4	=1000-[count in Column 3]; allocated to 4 strata as described in "quotasJP-w4-[date]". Initial predicted sample size = 364.
SG4: Non or never users	500	Def 7: # of Recontact completes with FR309v=3 or 9 and HN309v=3, 5-7	=500-[count in Column 3]; allocated to 4 strata as described in "quotasJP-w4-[date]". Initial predicted sample size = 0.

* Important Note:

There is a possibility that the originally targeted age or gender quotas for HTP-only users set by ITC cannot be satisfied. Rakuten Insight is to inform UW immediately. UW and Rakuten Insight should discuss and reach an agreement on ways to move forward—the agreement has to be decided in terms of support for the research questions, feasibility, and cost—the early days of the survey fieldwork will probably show what can be reasonably expected.

On ways to move forward, it is possible to redistribute the numbers, e.g., if there are not enough males aged 20-39, more males who are 40 and older, or more females aged 20-39 could be recruited. If even with a reasonable redistribution, if it is still not possible to obtain the required overall number of HTP-only users, increasing the sample size of other user group(s) could be contemplated, e.g., dual (HTP-cigarette) users, accept a lower overall sample size, etc.

Part G: Cigarette (FR23309v) and HTP (HN23309v) self-reported status variables from the JP4 Survey

	JP4 Variables
FR23309v	Derived variable -- Cigarette smoking status For both Rtype=C and P (Recontact and Replenishment): 01 Daily smoker: (FR225=1 or QA711=1) and BI345v=1. 02 Weekly smoker: (FR225=2 or QA711=2) and BI345v=1. 03 Less than weekly smoker: (FR225=3-4 or QA711=3) and BI345v=1. 04 Quitter < 30 days ago: [((QA342=1-2) or FR225=5) and FR143v=1 and (QA701=2 or QA706=1 or QA711=4) and BI345v=1 and … QA439=1]. 05 Quitter 1-6 months ago: . . . QA439=2-3. 06 Quitter 7-11 months ago: . . . QA439=4. 07 Quitter 1-2 years ago: . . . QA439=5. 08 Quitter more than 2 years ago: . . . QA439=6-8. 09 Never-a-smoker: [FR225=6 or [(FR225=5 OR QA342=1 or 2) and FR143v=2]]. 01 Daily smoker 02 Weekly smoker 03 Less than weekly smoker 04 Quitter <30 days ago 05 Quitter 1-6 months ago 06 Quitter 7-11 months ago 07 Quitter 1-2 years ago 08 Quitter more than 2 years ago 09 Never-a-smoker
FR23305	Derived variable: simplified cigarette status. 1 Smoker (FR309v=1-3) 2 Quitter (FR309v=4-8) 3 Non-smoker (FR309v=9)
HN23309v	Derived variable - heated tobacco status: 1 Current Daily HTP user (HN140=1) 2 Current Weekly HTP user (HN140=2) 3 Current Less than weekly HTP user (HN140=3 or 4) 4 Quitter: past at-least-weekly user of HTP ((HN195=1-2 or HN196=1) and HN140=5) 5 Trier: used HTP once or a few times ((HN195=4-5) or ((HN195=3 or HN196=2) and HN140=5)) 6 Never Tried HTP (HN106=2) 7 Never Heard of HTP (HN106=3)

Sampling Weights of the International Tobacco Control Japan (ITC JP) Survey

4th Edition (Waves 1–4)

C. Boudreau [†,1,2], Y. Li[2,3], M. Yan [2,3], M. Grey [2,3] and M.E. Thompson[1,2]

This document describes the various weights for wave 1 (section 1), wave 2 (section 2), wave 3 (section 3) and wave 4 (section 4) of the ITC Japan (JP) Survey. It also provides some guidance on which set of weights should be used depending on the analysis being performed, as well as cautionary notes when analyzing ITC Japan data (section 5). All sampling weights adjust for unequal selection probabilities, sample mis-representation, non-response and other biases. It is thus essential to use weighted data, when preforming any analyses using ITC Japan data.

Contents

[†]Correspondence: cboudreau@uwaterloo.ca
[1]Dept. of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada.
[2]Data Management Core (DMC) – ITC Project, University of Waterloo.
[3]Dept. of Psychology, University of Waterloo.
[4]This document was created using LaTeX, and last updated on Jan. 6, 2022 (with minor updates to the wave 4 benchmark figures of appendix A.2 on Apr. 3, 2023)

List of Tables

1 Wave 1 weights

Eight sets of cross-sectional weights were computed for the 4615 respondents interviewed at wave 1 of the ITC Japan Survey; see summary in table 1 and detailed description (including how and when they should be used) below: All sampling weights for the ITC Japan Survey were computed using the statistical software R (http://www.r-project.org). As mentioned at the beginning of this document, these weights adjust for unequal selection probabilities (in particular, the different sampling rates used for each of the 6 tobacco user groups described in section 1.1), sample mis-representation, non-response and other biases. **It is thus essential to use weighted data, when preforming any analyses using ITC Japan data**.

Weight	Variable Names
Wave 1 cross-sectional inflation weights for all respondents	aWTS23100v
Rescaled wave 1 cross-sectional weights for all respondents	aWTS23101v
Rescaled wave 1 cross-sectional weights for cigarette smokers	aWTS23201v
Rescaled wave 1 cross-sectional weights for dual users	aWTS23401v
Rescaled wave 1 cross-sectional weights for recent quitters †	aWTS23501v
Rescaled wave 1 cross-sectional weights for all tobacco users	aWTS23601v
Rescaled wave 1 cross-sectional weights for HTP users	aWTS23701v
Rescaled wave 1 cross-sectional weights for non-users ‡	aWTS23801v

† Those who have quit smoking cigarettes within the last 2 years (note that some of those quitters are now HTP users)
‡ This group also includes long-term quitters (i.e., those who quit more than 2 years ago)

Table 1: Cross-sectional sampling weights for wave 1 of the ITC Japan Survey.

1.1 User groups

For all cross-sectional weights, respondents were first divided into 6 user groups (variable `aUserGrp`[1] in the dataset): *i*) cigarette only users, *ii*) dual users, *iii*) exclusive HTP users, *iv*) recent quitter (i.e., quit cigarette smoking within the last two years) using HTP at least weekly, *v*) recent quitter using HTP less than weekly or not at all, and *vi*) non-users & long-term quitters.

The numbers of respondents in each of those user groups are given in table 2, and a detailed description of these groups is provided below. All respondents had to be 20 years or older at the time they completed the survey.

- To be classified as a cigarette only smoker (group *i*), a respondent had to (at the time of data collection) smoke cigarettes at least monthly, smoked at least 100 cigarettes in their lifetime, and use HTP less than weekly or not at all.
- To be classified as a dual user (group *ii*), a respondent had to smoke cigarettes at least monthly, smoked at least 100 cigarettes in their lifetime and use HTP at least weekly
- To be classified as exclusive heated tobacco product (HTP) user, (group *iii*), a respondent had to use HTP at least weekly and smoke cigarettes less than monthly or not at all.
- To be classified as a recent quitter (groups *iv* and *v*), a respondent had to have quit cigarette smoking within the last two years, and use HTP at least weekly (group *iv*), or use HTP less than weekly or not at all (group *v*).
- Finally, to be classified as a non-user/long-term quitter (group *vi*), a respondent had to smoke cigarettes less than monthly or not at all, use HTP less than weekly or not at all, and either never have been a cigarette smoker (or smoked less than 100 cigarettes in their lifetime) or quit more than 2 years ago. See appendix A.3 for outline of computer code used to create these 6 user groups.

[1] Where 1 = cigarette only, 2 = exclusive HTP, 3 = dual, 4 = non-user or long-term quitter, 5 = recent quitter using HTP at least weekly and 6 = recent quitter using HTP less than weekly; see appendix A.3.

User group⋆	n
Cigarette smokers	
Cigarette only	3288
Dual users	549
Total	3837
Exclusive HTP users	74
Recent quitters†	
Using HTP	90
Not using HTP	78
Total	168
Non-users and long-term quitters‡	536
Total	4615

⋆ See section 1.1 for the description of those user groups
† Those who have quit smoking cigarettes within the last two years
‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 2: Wave 1 respondents by user group.

1.2 Cross-sectional sampling weights

1- Variable `aWTS23100v` contains the wave 1 cross-sectional inflation weights for all 4615 respondents interviewed at wave 1. This total excludes the 229 respondents that were deemed to be fraudulent (also referred to as speeders or falsifiers in some documentation), the 37 daily smokers who had smoked fewer than 100 cigarettes in their lifetime, and 32 respondents who confused HTP and e-cigarettes. In other words, 298 were excluded because of data quality issues.

Within these 6 user groups, respondents were further subdivided based on gender, age ([20, 25), [25, 35), [35, 45) , [45, 55) & [55, 85]) and education (low, medium & high). This yielded the following 3 cross-tabs: user group × gender, user group × age group, and user group × education. Note that some cells in the user group × age group and user group × education cross-tabs were collapsed because they contained too few respondents. In particular, all age groups were collapsed into a single group for HTP only users (group *iii*); likewise for education. Finally, respondents were divided based on the 8 geographic regions of japan. However, region was not crossed with user groups.

Data from the 2017 Japan Society and Tobacco Internet Study (JASTIS) was then used to obtain benchmark/calibration figures (e.g., estimated number of individuals that are dual users) for each of these 3 cross-tabs and 8 geographic regions. These figures are given in appendix A.2. A raking procedure (see appendix A.1) was then applied to calibrate the weights based on user group × gender, user group × age group, user group × education and the 8 geographic regions.

These weights are designed to make respondents in each of the 6 groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 1 data collection. For example, the `aWTS23100v` weights of the 549 dual users are designed to make them representative of the Japanese population of dual users at the time of data collection; likewise for the other 5 groups.

If interest lies in a target population that consists of two or more of the 6 user groups, the `aWTS23100v` weights are still appropriate. For example, when studying cigarette smokers, one can simply combine

the aWTS23100v weights of the 3288 cigarette only users with those of the 549 dual users (for a total of 3837 respondents in the analysis), and assigned a weight of 0 to respondents in the other user groups. However, it simpler and, likely, more appropriate to use variable aWTS23201v in this situation. The same also applies to most of the possible combinations of the 6 user groups, and the weights listed below are likely to be more practical than these cross-sectional inflation weights.

Last but not least, since these are inflation/un-rescaled weights, they should not be used in analyses involving two or more countries. The various rescaled weights (i.e., variables aWTS23201v to aWTS23101v) described below were created especially for such multi-country analyses; see section 5.2 for more information on inflation versus rescaled weights.

2- Variable aWTS23201v contains the rescaled wave 1 cross-sectional weights for the 3837 (see table 2) respondents who were at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection. These are simply the wave 1 cross-sectional inflation weights (variable aWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 3837$). These weights are designed to make these 3837 cigarette smokers representative of the Japanese population of at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection.

3- Variable aWTS23401v contains the rescaled wave 1 cross-sectional weights for the 549 (see table 2) respondents who were dual users at the time of wave 1 data collection. These are simply the wave 1 cross-sectional inflation weights (variable aWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 549$). These weights are designed to make these 549 dual users representative of the Japanese population of dual users at the time of wave 1 data collection.

4- Variable aWTS23501v contains the rescaled wave 1 cross-sectional weights for the 168 (see table 2) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 1 data collection. These are simply the wave 1 cross-sectional inflation weights (variable aWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 168$). These weights are designed to make these 168 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 1 data collection.

5- Variable aWTS23601v contains the rescaled wave 1 cross-sectional weights for the 4001 (see table 2) respondents who were tobacco users (i.e., smoking cigarettes at least monthly and/or using HTP at least weekly) at the time of wave 1 data collection. These are simply the wave 1 cross-sectional inflation weights (variable aWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 4001$). These weights are designed to make these 4001 tobacco users representative of the Japanese population of tobacco users at the time of wave 1 data collection.

6- Variable aWTS23701v contains the rescaled wave 1 cross-sectional weights for the 713 (see table 2) respondents who were at least weekly HTP users at the time of wave 1 data collection. These are simply the wave 1 cross-sectional inflation weights (variable aWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 713$). These weights are designed to make these 713 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 1 data collection.

7- Variable aWTS23801v contains the rescaled wave 1 cross-sectional weights for the 536 (see table 2) respondents who were non-users or long-term quitters at the time of wave 1 data collection. Recall

that non-users are defined as smoking cigarettes less than monthly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the wave 1 cross-sectional inflation weights (variable aWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 536$). These weights are designed to make these 536 respondents representative of the Japanese population of non-users and long-term quitters.

8- Variable aWTS23101v contains the rescaled wave 1 cross-sectional weights for all 4615 respondents. These are simply the wave 1 cross-sectional inflation weights (variable aWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 4615$). As with the inflation weights, these weights are designed to make these 4615 respondents representative of the adult (20 years & older) Japanese population at the time of wave 1 data collection.

These rescaled wave 1 cross-sectional weights preserve the ratios between the 6 user groups. In other words, they preserve the ratios of: cigarette only smokers to exclusive HTP users, cigarette only smokers to non-users & long-term quitter, etc. Hence they are designed for analyses involving multiple or all 6 user groups; see section 5.2 for more information. However, it should be noted that tobacco users, recent quitters and non-users (and long term quitters) are ultimately distinct populations. Hence, great care must be taken when deciding to analyse them together using the aWTS23101v weights. This is probably fine when the goal is to carry out descriptive inference about the joint population of tobacco users, recent quitters and non-users. However, carrying out analytical inference (e.g., linear regression and logistic regression) from that same joint population is probably much more questionable.

It should be noted that it is **not appropriate** to combine variables aWTS23201v – aWTS23801v to jointly analyse multiple user groups. Each of those variables was rescaled to sum to its respective sample size, and thus combining them will not preserve the ratios between the user groups resulting in erroneous results. This is why variable aWTS23101v was computed.

2 Wave 2 weights

Eight sets of cross-sectional weights and 8 sets of longitudinal weights were computed at wave 2 of the ITC Japan Survey. The cross-sectional weights are listed in table 3, and their computation and how/when they should be used are detailed section 2.1. Likewise, the longitudinal weights are listed in table 4, and their computation and use are detailed in section 2.2. All sampling weights for the ITC Japan Survey were computed using the statistical software R (http://www.r-project.org). As mentioned at the beginning of this document, these weights adjust for unequal selection probabilities (in particular, the different sampling rates used for each of the 6 tobacco user groups described in section 1.1), sample mis-representation, non-response and other biases. **It is thus essential to use weighted data, when preforming any analyses using ITC Japan data**.

Weight	Variable Names
Wave 2 cross-sectional inflation weights for all respondents	bWTS23100v
Rescaled wave 2 cross-sectional weights for all respondents	bWTS23101v
Rescaled wave 2 cross-sectional weights for cigarette smokers	bWTS23201v
Rescaled wave 2 cross-sectional weights for dual users	bWTS23401v
Rescaled wave 2 cross-sectional weights for recent quitters †	bWTS23501v
Rescaled wave 2 cross-sectional weights for all tobacco users	bWTS23601v
Rescaled wave 2 cross-sectional weights for HTP users	bWTS23701v
Rescaled wave 2 cross-sectional weights for non-users ‡	bWTS23801v

† Those who have quit smoking cigarettes within the last 2 years (note that some of those quitters are now HTP users)
‡ This group also includes long-term quitters (i.e., those who quit more than 2 years ago)

Table 3: Cross-sectional sampling weights for wave 2 of the ITC Japan Survey

2.1 Cross-sectional sampling weights

As mentioned at the beginning of section 2, 8 sets of cross-sectional weights were computed at wave 2 of the ITC Japan Survey. These are listed in table 3, and their computation as well as how and when they should be used are detailed in this section. Those weights were computed for the 4228 respondents interviewed at wave 2. This total excludes the 117 respondents that were deemed to be fraudulent (also referred to as speeders or falsifiers in some documentation), 150 daily smokers who had smoked fewer than 100 cigarettes in their lifetime and/or confused HTP and e-cigarettes, and 5 other respondents with data quality issues. In other words, 272 were excluded because of data quality issues. Those 4228 individuals consist of 1484 respondents newly recruited at wave 2, and of 2744 cohort respondents that were recruited at wave 1 and successfully retained at wave 2. It should be noted that all cross-sectional weights are designed to make the sample representative at the time of wave 2 data collection. If the aim is to ensure that the sample is representative at the time of wave 1 data collection, then the longitudinal sampling weights computed in section 2.2 should be used instead.

1- Variable bWTS23100v contains the wave 2 cross-sectional inflation weights for all 4228 respondents interviewed at wave 2. Computation of those weights followed the same steps and used the same raking algorithm (see appendix A.1) as the wave 1 cross-sectional inflation weights (variable aWTS23100v of section 1.2). The only differences being:

   i) The wave 2 cross-sectional inflation weights are designed to make the sample representative of the target population at the time of wave 2. Consequently, for cohort respondents, age at wave 2 was used instead of their age when they first recruited. Likewise, variable bUserGrp was used instead of aUserGrp, and similarly for education and geographic region. It is thus possible for a respondent transition from being a cigarette only user at wave 1 and to being classified as a dual user at wave 2. Other types of transitions are also possible. Likewise, it is possible for respondents to move between geographic regions.

   ii) Fewer user group × age and user group × education cells needed to be collapsed, as less of them contained too few respondents.

Weight	Variable Names
Waves 1–2 longitudinal inflation weights for all respondents	bWTS23120v
Rescaled waves 1–2 longitudinal weights for all respondents	bWTS23121v
Rescaled waves 1–2 longitudinal weights for cigarette smokers	bWTS23221v
Rescaled waves 1–2 longitudinal weights for dual users	bWTS23421v
Rescaled waves 1–2 longitudinal weights for recent quitters †	bWTS23521v
Rescaled waves 1–2 longitudinal weights for all tobacco users	bWTS23621v
Rescaled waves 1–2 longitudinal weights for HTP users	bWTS23721v
Rescaled waves 1–2 longitudinal weights for non-users ‡	bWTS23821v

† Those who have quit smoking cigarettes within the last 2 years (note that some of those quitters are now HTP users)
‡ This group also includes long-term quitters (i.e., those who quit more than 2 years ago)

Table 4: Longitudinal sampling weights for wave 2 of the ITC Japan Survey

User group⋆	n
Cigarette smokers	
Cigarette only	1810
Dual users	885
Total	2695
Exclusive HTP users	621
Recent quitters†	
Using HTP	304
Not using HTP	149
Total	453
Non-users and long-term quitters‡	459
Total	4228

⋆ See section 1.1 for the description of those user groups
† Those who have quit smoking cigarettes within the last two years
‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 5: Wave 2 respondents by user group.

iii) Data from the 2019 JASTIS, instead of the 2017 JASTIS, was used to obtain benchmark/calibration figures; see the "Wave 2" column of the various tables in appendix A.2.

These weights are designed to make respondents in each of the 6 groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 2 data collection. For example, the bWTS23100v weights of the 885 dual users are designed to make them representative of the Japanese population of dual users at the time of data collection; likewise for the other 5 groups.

If interest lies in a target population that consists of two or more of the 6 user groups, the bWTS23100v weights are still appropriate. For example, when studying cigarette smokers, one can simply combine the bWTS23100v weights of the 1810 cigarette only users with those of the 885 dual users (for a total of

2695 respondents in the analysis), and assigned a weight of 0 to respondents in the other user groups. However, it simpler and, likely, more appropriate to use variable bWTS23201v in this situation. The same also applies to most of the possible combinations of the 6 user groups, and the weights listed below are likely to be more practical than these cross-sectional inflation weights.

Last but not least, since these are inflation/un-rescaled weights, they should not be used in analyses involving two or more countries. The various rescaled weights (i.e., variables bWTS23201v to bWTS23101v) described below were created especially for such multi-country analyses; see section 5.2 for more information on inflation versus rescaled weights.

2- Variable bWTS23201v contains the rescaled wave 2 cross-sectional weights for the 2695 (see table 5) respondents who were at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 2 data collection. These are simply the wave 2 cross-sectional inflation weights (variable bWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 2695$). These weights are designed to make these 2695 cigarette smokers representative of the Japanese population at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 2 data collection.

3- Variable bWTS23401v contains the rescaled wave 2 cross-sectional weights for the 885 (see table 5) respondents who were dual users at the time of wave 2 data collection. These are simply the wave 2 cross-sectional inflation weights (variable bWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 885$). These weights are designed to make these 885 dual users representative of the Japanese population of dual users at the time of wave 2 data collection.

4- Variable bWTS23501v contains the rescaled wave 2 cross-sectional weights for the 453 (see table 5) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 2 data collection. These are simply the wave 2 cross-sectional inflation weights (variable bWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 453$). These weights are designed to make these 453 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 2 data collection.

5- Variable bWTS23601v contains the rescaled wave 2 cross-sectional weights for the 3620 (see table 5) respondents who were tobacco users (i.e., smoking cigarettes at least monthly and/or using HTP at least weekly) at the time of wave 2 data collection. These are simply the wave 2 cross-sectional inflation weights (variable bWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 3620$). These weights are designed to make these 3620 tobacco users representative of the Japanese population of tobacco users at the time of wave 2 data collection.

6- Variable bWTS23701v contains the rescaled wave 2 cross-sectional weights for the 1810 (see table 5) respondents who were at least weekly HTP users at the time of wave 2 data collection. These are simply the wave 2 cross-sectional inflation weights (variable bWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 1810$). These weights are designed to make these 1810 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 2 data collection.

7- Variable bWTS23801v contains the rescaled wave 2 cross-sectional weights for the 459 (see table 5) respondents who were non-users or long-term quitters at the time of wave 2 data collection. Recall that non-users are defined as smoking cigarettes less than monthly or not at all, and using HTP less

than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the wave 2 cross-sectional inflation weights (variable bWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 459$). These weights are designed to make these 459 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 2 data collection.

8- Variable bWTS23101v contains the rescaled wave 2 cross-sectional weights for all 4228 respondents. These are simply the wave 2 cross-sectional inflation weights (variable bWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 4228$). These weights are designed to make these 4228 respondents representative of the adult (20 years & older) Japanese population at the time of wave 2 data collection.

These rescaled wave 2 cross-sectional weights preserve the ratios between the 6 user groups (see section 1.1). In other words, they preserve the ratios of: cigarette only smokers to exclusive HTP users, cigarette only smokers to non-users & long-term quitter, etc. Hence they are designed for analyses involving multiple or all 6 user groups; see section 5.2 for more information. However, it should be noted that tobacco users, recent quitters and non-users (and long term quitters) are ultimately distinct populations. Hence, great care must be taken when deciding to analyse them together using the bWTS23101v weights. This is probably fine when the goal is to carry out descriptive inference about the joint population of tobacco users, recent quitters and non-users. However, carrying out analytical inference (e.g., linear regression and logistic regression) from that same joint population is probably much more questionable.

It should be noted that it is **not appropriate** to combine variables bWTS23201v – bWTS23801v to jointly analyse multiple user groups. Each of those variables was rescaled to sum to its respective sample size, and thus combining them will not preserve the ratios between the user groups resulting in erroneous results. This is why variable bWTS23101v was computed.

2.2 Longitudinal sampling weights

As mentioned at the beginning of section 2, 8 sets of longitudinal weights were computed at wave 2 of the ITC Japan Survey. These are listed in table 4, and their computation as well as how and when they should be used are detailed in this section. All of those weights were computed for respondents recruited at wave 1, then successfully retained and interviewed at wave 2. It should be noted that all longitudinal weights are designed to make the sample representative at the time of baseline wave (i.e., wave 1 in the present case), and not at the time of the current wave (i.e., wave 2 in the present case). If the aim is to ensure that the sample is representative at the time of wave 2 data collection, then the cross-sectional sampling weights computed in section 2.1 should be used instead.

1- Variable bWTS23120v contains the waves 1–2 longitudinal inflation weights for all 2744 respondents recruited at wave 1, and then successfully recontacted at wave 2. These weights are designed to make respondents in each of the 6 user groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 1 data collection.

These waves 1–2 longitudinal inflation weights are the wave 1 cross-sectional inflation weights (i.e., variable aWTS23100v computed in section 1.2) adjusted for attrition between waves 1 and 2. Consequently, computation of those longitudinal inflation weights followed the same steps as described

User group⋆	n
Cigarette smokers	
Cigarette only	2102
Dual users	268
Total	2370
Exclusive HTP users	31
Recent quitters†	
Using HTP	31
Not using HTP	26
Total	57
Non-users and long-term quitters‡	286
Total	2744

⋆ See section 1.1 for the description of those user groups
† Those who have quit smoking cigarettes within the last two years
‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 6: Wave 1 respondents successfully recontacted at wave 2 by user group.

in the computation of variable aWTS23100v. Since these weights are meant to make these 2744 respondents representative of the population at the time of wave 1 data collection, the benchmark/-calibration figures obtained from the 2017 JASTIS and used to compute the wave 1 cross-sectional weights were used here as well; as opposed to the 2019 figures used to compute the wave 2 cross-sectional weights (see section 2.1). Likewise, respondents were divided into the 6 user groups based on their wave 1 status; similarly, age, education and geographic region at wave 1 were used.

2- Variable bWTS23221v contains the rescaled waves 1–2 longitudinal weights for the 2370 (see table 6) respondents who were at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection, and then successfully recontacted at wave 2. These are simply the waves 1–2 longitudinal inflation weights (variable bWTS23120v) of those respondents rescaled to sum to sample size (i.e., $n = 2370$). These weights are designed to make these 2370 cigarette smokers representative of the Japanese population of at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection.

3- Variable bWTS23421v contains the rescaled waves 1–2 longitudinal weights for the 268 (see table 6) respondents who were dual users at the time of wave 1 data collection, and then successfully recontacted at wave 2. These are simply the waves 1–2 longitudinal inflation weights (variable bWTS23120v) of those respondents rescaled to sum to sample size. These weights are designed to make these 268 dual users representative of the Japanese population of dual users at the time of wave 1 data collection.

4- Variable bWTS23521v contains the rescaled waves 1–2 longitudinal weights for the 57 (see table 6) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 1 data collection, and then successfully recontacted at wave 2. These are simply the waves 1–2 longitudinal inflation weights (variable bWTS23120v) of those respondents rescaled to sum to sample size. These weights are designed to make these 57 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 1 data collection.

5- Variable `bWTS23621v` contains the rescaled waves 1–2 longitudinal weights for the 2432 (see table 6) respondents who were tobacco users (i.e., smoking cigarettes at least monthly and/or using HTP at least weekly) at the time of wave 1 data collection, and then successfully recontacted at wave 2. These are simply the waves 1–2 longitudinal inflation weights (variable `bWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 2432 tobacco users representative of the Japanese population of tobacco users at the time of wave 1 data collection.

6- Variable `bWTS23721v` contains the rescaled waves 1–2 longitudinal weights for the 330 (see table 6) respondents who were at least weekly HTP users at the time of wave 1 data collection, and then successfully recontacted at wave 2. These are simply the waves 1–2 longitudinal inflation weights (variable `bWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 330 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 1 data collection.

7- Variable `bWTS23821v` contains the rescaled waves 1–2 longitudinal weights for the 286 (see table 6) respondents who were non-users or long-term quitters at the time of wave 1 data collection, and then successfully recontacted at wave 2. Recall that non-users are defined as smoking cigarettes less than monthly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the waves 1–2 longitudinal inflation weights (variable `bWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 286 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 1 data collection.

8- Variable `bWTS23121v` contains the rescaled waves 1–2 longitudinal weights for all 2744 respondents recruited at wave 1, and then successfully recontacted at wave 2. These are simply the waves 1–2 longitudinal inflation weights (variable `bWTS23120v`) of those respondents rescaled to sum to sample size (i.e., $n = 2744$). These weights are designed to make these 2744 respondents representative of the adult (20 years & older) Japanese population at the time of wave 1 data collection. The remarks and notes listed for variable `bWTS23101v` also apply here.

3 Wave 3 weights

Eight sets of cross-sectional weights and 16 sets of longitudinal weights were computed at wave 3 of the ITC Japan Survey. These cross-sectional weights are listed in table 7, and their computation and how/when they should be used are detailed section 3.1. Likewise, the longitudinal weights are listed in table 8, and their computation and use are detailed in sections 3.2 and 3.3. All sampling weights for the ITC Japan Survey were computed using the statistical software R (`http://www.r-project.org`). As mentioned at the beginning of this document, these weights adjust for unequal selection probabilities, sample mis-representation, non-response and other biases. **It is thus essential to use weighted data, when preforming any analyses using ITC Japan data**.

3.1 Cross-sectional sampling weights

As mentioned at the beginning of section 3, 8 sets of cross-sectional weights were computed at wave 3 of the ITC Japan Survey. These are listed in table 7, and their computation as well as how and when

Weight	Variable Names
Wave 3 cross-sectional inflation weights for all respondents	cWTS23100v
Rescaled wave 3 cross-sectional weights for all respondents	cWTS23101v
Rescaled wave 3 cross-sectional weights for cigarette smokers	cWTS23201v
Rescaled wave 3 cross-sectional weights for dual users	cWTS23401v
Rescaled wave 3 cross-sectional weights for quitters †	cWTS23501v
Rescaled wave 3 cross-sectional weights for all tobacco users	cWTS23601v
Rescaled wave 3 cross-sectional weights for HTP users	cWTS23701v
Rescaled wave 3 cross-sectional weights for non-users ‡	cWTS23801v

† Those who have quit smoking cigarettes within the last 2 years (note that some of those quitters are now HTP users)
‡ This group also includes long-term quitters (i.e., those who quit more than 2 years ago)

Table 7: Cross-sectional sampling weights for wave 3 of the ITC Japan Survey

they should be used are detailed in this section. Those weights were computed for the 4482 respondents interviewed at wave 3. This total excludes the 151 respondents that were deemed to be fraudulent (also referred to as speeders or falsifiers in some documentation). Those 4482 individuals consist of 1687 respondents newly recruited at wave 3, and of 2795 cohort respondents (1891 recruited at wave 1 and 904 recruited at wave 2) that were successfully retained at wave 3. It should be noted that all cross-sectional weights are designed to make the sample representative at the time of wave 3 data collection. If the aim is to ensure that the sample is representative at the time of wave 1 or 2, then the longitudinal sampling weights computed in sections 3.2 or 3.3 should be used instead.

One important change was made at wave 3. It was decided that to be as classified a cigarette smoker a respondent had to (at the time of data collection) smoke cigarettes at least weekly and smoked at least 100 cigarettes in their lifetime. This weekly requirement is different than at the previous two waves, where a respondent had to smoke at least monthly (and smoked at least 100 cigarettes in their lifetime) to be as classified a cigarette smoker. This change obviously affects how respondents are categorized as cigarette only smokers (group *i*) and dual users (group *ii*), but also user groups *iii* and *vi*. For example, a respondent smoking cigarette less than weekly but at least monthly and using HTP less than weekly or not at all was classified as a non-user at wave 3, whereas that same individual would have been classified as a cigarette only smoker at the previous two waves. Likewise, a respondent smoking cigarette less than weekly but at least monthly and using HTP at least weekly was classified as an exclusive HTP user at wave 3, whereas that same individual would have been classified as a dual user at the previous two waves.

Last but not least, this new classification only applies for the wave 3 cross-sectional sampling weights. As mentioned before, longitudinal weights are designed to make the sample representative at the time of data collection for that earlier wave. Hence, the user group assigned at wave 1 (using the old user group definition of section 1.1) is used when computing the waves 1–3 longitudinal sampling weights (section 3.2). Likewise, the user group assigned at wave 2 (again, using the old user group definition) is used when computing the waves 2–3 longitudinal sampling weights (section 3.3),

1- Variable cWTS23100v contains the wave 3 cross-sectional inflation weights for all 4482 respondents interviewed at wave 3. Computation of those weights followed the same steps and used the same rak-

Weight	Variable Names
Waves 1–3 longitudinal inflation weights for all respondents	cWTS23120v
Rescaled waves 1–3 longitudinal weights for all respondents	cWTS23121v
Rescaled waves 1–3 longitudinal weights for cigarette smokers	cWTS23221v
Rescaled waves 1–3 longitudinal weights for dual users	cWTS23421v
Rescaled waves 1–3 longitudinal weights for quitters †	cWTS23521v
Rescaled waves 1–3 longitudinal weights for all tobacco users	cWTS23621v
Rescaled waves 1–3 longitudinal weights for HTP users	cWTS23721v
Rescaled waves 1–3 longitudinal weights for non-users ‡	cWTS23821v
Waves 2–3 longitudinal inflation weights for all respondents	cWTS23122v
Rescaled waves 2–3 longitudinal weights for all respondents	cWTS23123v
Rescaled waves 2–3 longitudinal weights for cigarette smokers	cWTS23223v
Rescaled waves 2–3 longitudinal weights for dual users	cWTS23423v
Rescaled waves 2–3 longitudinal weights for quitters †	cWTS23523v
Rescaled waves 2–3 longitudinal weights for all tobacco users	cWTS23623v
Rescaled waves 2–3 longitudinal weights for HTP users	cWTS23723v
Rescaled waves 2–3 longitudinal weights for non-users ‡	cWTS23823v

† Those who have quit smoking cigarettes within the last 2 years (note that some of those quitters are now HTP users)
‡ This group also includes long-term quitters (i.e., those who quit more than 2 years ago)

Table 8: Longitudinal sampling weights for wave 3 of the ITC Japan Survey

ing algorithm (see appendix A.1) as the wave 1 cross-sectional inflation weights (variable aWTS23100v of section 1.2). The only differences being:

i) The wave 3 cross-sectional inflation weights are designed to make the sample representative of the target population at the time of wave 3. Consequently, for cohort respondents, age at wave 3 was used instead of their age at waves 1 or 2. Likewise, variable cUserGrp was used instead of aUserGrp, and similarly for education and geographic region. Recall that it is possible for a respondent to transition from being a cigarette only user at wave 1 to being classified as a dual user at wave 3. Other types of transitions are also possible. Likewise, it is possible for respondents to move between geographic regions.

ii) More cells needed to be collapsed because they contained too few respondents.

iii) Data from the 2020 JASTIS, instead of the 2017 JASTIS, was used to obtain benchmark/calibration figures; see the "Wave 3" column of the various tables in appendix A.2.

These weights are designed to make respondents in each of the 6 groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 3 data collection. For example, the cWTS23100v weights of the 909 dual users are designed to make them representative of the Japanese population of dual users at the time of data collection; likewise for the other 5 groups.

User group⋆	n
Cigarette smokers	
Cigarette only	1848
Dual users	909
Total	2757
Exclusive HTP users	795
Recent quitters†	
Using HTP	174
Not using HTP	177
Total	351
Non-users and long-term quitters‡	579
Total	4482

⋆ See section 1.1 for the description of those user groups, as well as beginning of section 3.1 for important change made on the classification of cigarette smokers

† Those who have quit smoking cigarettes within the last two years

‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 9: Wave 3 respondents by user group.

If interest lies in a target population that consists of two or more of the 6 user groups, the `cWTS23100v` weights are still appropriate. For example, when studying cigarette smokers, one can simply combine the `cWTS23100v` weights of the 1848 cigarette only users with those of the 909 dual users (for a total of 2757 respondents in the analysis), and assigned a weight of 0 to respondents in the other user groups. However, it simpler and, likely, more appropriate to use variable `cWTS23201v` in this situation. The same also applies to most of the possible combinations of the 6 user groups, and the weights listed below are likely to be more practical than these cross-sectional inflation weights.

Last but not least, since these are inflation/un-rescaled weights, they should not be used in analyses involving two or more countries. The various rescaled weights (i.e., variables `cWTS23201v` to `cWTS23101v`) described below were created especially for such multi-country analyses; see section 5.2 for more information on inflation versus rescaled weights.

2- Variable `cWTS23201v` contains the rescaled wave 3 cross-sectional weights for the 2757 (see table 9) respondents who were at least weekly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 3 data collection. These are simply the wave 3 cross-sectional inflation weights (variable `cWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 2757$). These weights are designed to make these 2757 cigarette smokers representative of the Japanese population at least weekly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 3 data collection.

3- Variable `cWTS23401v` contains the rescaled wave 3 cross-sectional weights for the 909 (see table 9) respondents who were dual users at the time of wave 3 data collection. These are simply the wave 3 cross-sectional inflation weights (variable `cWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 909$). These weights are designed to make these 909 dual users representative of the Japanese population of dual users at the time of wave 3 data collection. Recall that, at wave 3, dual users are defined as smoking cigarettes at least weekly, smoked at least 100 cigarettes in their lifetime, and using HTP at least weekly.

4- Variable `cWTS23501v` contains the rescaled wave 3 cross-sectional weights for the 351 (see table 9) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 3 data collection. These are simply the wave 3 cross-sectional inflation weights (variable `cWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 351$). These weights are designed to make these 351 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 3 data collection.

5- Variable `cWTS23601v` contains the rescaled wave 3 cross-sectional weights for the 3726 (see table 9) respondents who were at least weekly tobacco users (i.e., smoking cigarettes at least weekly and/or using HTP at least weekly) at the time of wave 3 data collection. These are simply the wave 3 cross-sectional inflation weights (variable `cWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 3726$). These weights are designed to make these 3726 tobacco users representative of the Japanese population of at least weekly tobacco users at the time of wave 3 data collection.

6- Variable `cWTS23701v` contains the rescaled wave 3 cross-sectional weights for the 1878 (see table 9) respondents who were at least weekly HTP users at the time of wave 3 data collection. These are simply the wave 3 cross-sectional inflation weights (variable `cWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 1878$). These weights are designed to make these 1878 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 3 data collection.

7- Variable `cWTS23801v` contains the rescaled wave 3 cross-sectional weights for the 579 (see table 9) respondents who were non-users or long-term quitters at the time of wave 3 data collection. Recall that non-users are defined as smoking cigarettes less than weekly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the wave 3 cross-sectional inflation weights (variable `cWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 579$). These weights are designed to make these 579 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 3 data collection.

8- Variable `cWTS23101v` contains the rescaled wave 3 cross-sectional weights for all 4482 respondents. These are simply the wave 3 cross-sectional inflation weights (variable `cWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 4482$). These weights are designed to make these 4482 respondents representative of the adult (20 years & older) Japanese population at the time of wave 3 data collection.

   These rescaled wave 3 cross-sectional weights preserve the ratios between the 6 user groups. In other words, they preserve the ratios of: cigarette only smokers to exclusive HTP users, cigarette only smokers to non-users & long-term quitter, etc. Hence they are designed for analyses involving multiple or all 6 user groups; see section 5.2 for more information. However, it should be noted that tobacco users, recent quitters and non-users (and long term quitters) are ultimately distinct populations. Hence, great care must be taken when deciding to analyse them together using the `cWTS23101v` weights. This is probably fine when the goal is to carry out descriptive inference about the joint population of tobacco users, recent quitters and non-users. However, carrying out analytical inference (e.g., linear regression and logistic regression) from that same joint population is probably much more questionable.

It should be noted that it is **not appropriate** to combine variables cWTS23201v – cWTS23801v to jointly analyse multiple user groups. Each of those variables was rescaled to sum to its respective sample size, and thus combining them will not preserve the ratios between the user groups resulting in erroneous results. This is why variable cWTS23101v was computed.

3.2 Waves 1–3 longitudinal sampling weights

As mentioned at the beginning of section 3, 16 sets of longitudinal weights were computed at wave 3 of the ITC Japan Survey. Half of those sets (see top half of table 8) were computed for respondents recruited at wave 1, and then successfully retained and interviewed at waves 2 and 3. This section details the computation of those weights, as well as how and when they should be used. It should be noted that all waves 1–3 longitudinal weights are designed to make the sample representative at the time of wave 1. If the aim is to ensure that the sample is representative at the time of wave 3 data collection, then the cross-sectional sampling weights computed in section 3.1 should be used instead.

User group⋆	n
Cigarette smokers	
Cigarette only	1453
Dual users	180
Total	1633
Exclusive HTP users	22
Recent quitters†	
Using HTP	21
Not using HTP	18
Total	39
Non-users and long-term quitters‡	197
Total	1891

⋆ See section 1.1 for the description of those user groups
† Those who have quit smoking cigarettes within the last two years
‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 10: Wave 1 respondents successfully recontacted at wave 3 by user group.

1- Variable cWTS23120v contains the waves 1–3 longitudinal inflation weights for all 1891 respondents recruited at wave 1, and then successfully recontacted at waves 2 and 3. These weights are designed to make respondents in each of the 6 user groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 1 data collection.

These waves 1–3 longitudinal inflation weights are the waves 1–2 longitudinal inflation weights (i.e., variable bWTS23120v computed in section 2.2) adjusted for attrition between waves 2 and 3. Consequently, computation of those longitudinal inflation weights followed the same steps as described in the computation of variable bWTS23120v. Since these weights are meant to make these 1891 respondents representative of the population at the time of wave 1 data collection, the benchmark/-calibration figures obtained from the 2017 JASTIS and used to compute the wave 1 cross-sectional

weights were used here as well; as opposed to the 2020 figures used to compute the wave 3 cross-sectional weights (see section 3.1). Likewise, respondents were divided into the 6 user groups based on their wave 1 status[1]; similarly, age, education and geographic region at wave 1 were used.

2- Variable `cWTS23221v` contains the rescaled waves 1–3 longitudinal weights for the 1633 (see table 10) respondents who were at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection, and then successfully recontacted at waves 2 and 3. These are simply the waves 1–3 longitudinal inflation weights (variable `cWTS23120v`) of those respondents rescaled to sum to sample size (i.e., $n = 1633$). These weights are designed to make these 1633 cigarette smokers representative of the Japanese population of at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection.

3- Variable `cWTS23421v` contains the rescaled waves 1–3 longitudinal weights for the 180 (see table 10) respondents who were dual users at the time of wave 1 data collection, and then successfully recontacted at waves 2 and 3. These are simply the waves 1–3 longitudinal inflation weights (variable `cWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 180 dual users representative of the Japanese population of dual users at the time of wave 1 data collection.

4- Variable `cWTS23521v` contains the rescaled waves 1–3 longitudinal weights for the 39 (see table 10) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 1 data collection, and then successfully recontacted at waves 2 and 3. These are simply the waves 1–3 longitudinal inflation weights (variable `cWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 39 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 1 data collection. However, given that there are only 39 such respondents, any analyses focusing on those recent quitters is at best questionnable.

5- Variable `cWTS23621v` contains the rescaled waves 1–3 longitudinal weights for the 1676 (see table 10) respondents who were tobacco users (i.e., smoking cigarettes at least monthly and/or using HTP at least weekly) at the time of wave 1 data collection, and then successfully recontacted at waves 2 and 3. These are simply the waves 1–3 longitudinal inflation weights (variable `cWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 1676 tobacco users representative of the Japanese population of tobacco users at the time of wave 1 data collection.

6- Variable `cWTS23721v` contains the rescaled waves 1–3 longitudinal weights for the 223 (see table 10) respondents who were at least weekly HTP users at the time of wave 1 data collection, and then successfully recontacted at waves 2 and 3. These are simply the waves 1–3 longitudinal inflation weights (variable `cWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 223 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 1 data collection.

7- Variable `cWTS23821v` contains the rescaled waves 1–3 longitudinal weights for the 197 (see table 10) respondents who were non-users or long-term quitters at the time of wave 1 data collection, and then successfully recontacted at waves 2 and 3. Recall that non-users are defined as smoking cigarettes

[1]Since the user group at wave 1 (i.e., variable `aUsergrp`) is used to compute the longitudinal weights, respondents smoking at least monthly were classified as cigarette smokers; as opposed to the at least weekly requirement for the wave 3 cross-sectional weights of section 3.1.

less than monthly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the waves 1–3 longitudinal inflation weights (variable `cWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 197 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 1 data collection.

8- Variable `cWTS23121v` contains the rescaled waves 1–3 longitudinal weights for all 1891 respondents recruited at wave 1, and then successfully recontacted at waves 2 and 3. These are simply the waves 1–3 longitudinal inflation weights (variable `cWTS23120v`) of those respondents rescaled to sum to sample size (i.e., $n = 1891$). These weights are designed to make these 1891 respondents representative of the adult (20 years & older) Japanese population at the time of wave 1 data collection. The remarks and notes listed for variable `cWTS23101v` also apply here.

3.3 Waves 2–3 longitudinal sampling weights

As mentioned at the beginning of section 3, 16 sets of longitudinal weights were computed at wave 3 of the ITC Japan Survey. Half of those sets (see bottom half of table 8) were computed for respondents recruited at either wave 1 or wave 2, and then successfully retained and interviewed at wave 3. This section details the computation of those weights, as well as how and when they should be used. It should be noted that all waves 2–3 longitudinal weights are designed to make the sample representative at the time of wave 2. If the aim is to ensure that the sample is representative at the time of wave 3 data collection, then the cross-sectional sampling weights computed in section 3.1 should be used instead.

User group*	n
Cigarette smokers	
Cigarette only	1226
Dual users	646
Total	1872
Exclusive HTP users	385
Recent quitters†	
Using HTP	180
Not using HTP	58
Total	238
Non-users and long-term quitters‡	300
Total	2795

* See section 1.1 for the description of those user groups
† Those who have quit smoking cigarettes within the last two years
‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 11: Waves 1 and 2 respondents successfully recontacted at wave 3 by user group.

1- Variable `cWTS23122v` contains the waves 2–3 longitudinal inflation weights for all 2795 respondents interviewed at wave 2, and then successfully recontacted at wave 3. These weights are designed to

make respondents in each of the 6 user groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 2 data collection.

These waves 2–3 longitudinal inflation weights are the wave 2 cross-sectional inflation weights (i.e., variable bWTS23100v computed in section 2.1) adjusted for attrition between waves 2 and 3. Consequently, computation of those longitudinal inflation weights followed the same steps as described in the computation of variable bWTS23100v. Since these weights are meant to make these 2795 respondents representative of the population at the time of wave 2 data collection, the benchmark/-calibration figures obtained from the 2019 JASTIS and used to compute the wave 2 cross-sectional weights were used here as well; as opposed to the 2020 figures used to compute the wave 3 cross-sectional weights (see section 3.1). Likewise, respondents were divided into the 6 user groups based on their wave 2 status[1]; similarly, age, education and geographic region at wave 2 were used.

2- Variable cWTS23223v contains the rescaled waves 2–3 longitudinal weights for the 1872 (see table 11) respondents who were at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 2 data collection, and then successfully recontacted at wave 3. These are simply the waves 2–3 longitudinal inflation weights (variable cWTS23122v) of those respondents rescaled to sum to sample size. These weights are akin to variable cWTS23221v, but are designed to make these 1872 respondents representative of the population at the time of wave 2 data collection instead of wave 1 data collection.

3- Variable cWTS23423v contains the rescaled waves 2–3 longitudinal weights for the 646 (see table 11) respondents who were dual users at the time of wave 2 data collection, and then successfully recontacted at wave 3. These are simply the waves 2–3 longitudinal inflation weights (variable cWTS23122v) of those respondents rescaled to sum to sample size. These weights are akin to variable cWTS23421v, but are designed to make these 646 respondents representative of the population at the time of wave 2 data collection instead of wave 1 data collection.

4- Variable cWTS23523v contains the rescaled waves 2–3 longitudinal weights for the 238 (see table 11) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 2 data collection, and then successfully recontacted at wave 3. These are simply the waves 2–3 longitudinal inflation weights (variable cWTS23122v) of those respondents rescaled to sum to sample size. These weights are akin to variable cWTS23521v, but are designed to make these 238 respondents representative of the population at the time of wave 2 data collection instead of wave 1 data collection.

5- Variable cWTS23623v contains the rescaled waves 2–3 longitudinal weights for the 2437 (see table 11) respondents who were tobacco users (i.e., smoking cigarettes at least monthly and/or using HTP at least weekly) at the time of wave 2 data collection, and then successfully recontacted at wave 3. These are simply the waves 2–3 longitudinal inflation weights (variable cWTS23122v) of those respondents rescaled to sum to sample size. These weights are akin to variable cWTS23621v, but are designed to make these 2437 respondents representative of the population at the time of wave 2 data collection instead of wave 1 data collection.

6- Variable cWTS23723v contains the rescaled waves 2–3 longitudinal weights for the 1211 (see table 11) respondents who were at least weekly HTP users at the time of wave 2 data collection, and then

[1]Since the user group at wave 2 (i.e., variable bUsergrp) is used to compute the longitudinal weights, respondents smoking at least monthly were classified as cigarette smokers; as opposed to the at least weekly requirement for the wave 3 cross-sectional weights of section 3.1.

successfully recontacted at wave 3. These are simply the waves 2–3 longitudinal inflation weights (variable cWTS23122v) of those respondents rescaled to sum to sample size. These weights are akin to variable cWTS23721v, but are designed to make these 1211 respondents representative of the population at the time of wave 2 data collection instead of wave 1 data collection.

7- Variable cWTS23823v contains the rescaled waves 2–3 longitudinal weights for the 300 (see table 11) respondents who were non-users or long-term quitters at the time of wave 2 data collection, and then successfully recontacted at wave 3. Recall that non-users are defined as smoking cigarettes less than monthly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the waves 2–3 longitudinal inflation weights (variable cWTS23122v) of those respondents rescaled to sum to sample size. These weights are akin to variable cWTS23821v, but are designed to make these 300 respondents representative of the population at the time of wave 2 data collection instead of wave 1 data collection.

8- Variable cWTS23123v contains the rescaled waves 2–3 longitudinal weights for all 2795 respondents interviewed at wave 2, and then successfully recontacted at wave 3. These are simply the waves 2–3 longitudinal inflation weights (variable cWTS23122v) of those respondents rescaled to sum to sample size (i.e., $n = 2795$). These weights are akin to variable cWTS23121v, but are designed to make these 2795 respondents representative of the population at the time of wave 2 data collection instead of wave 1 data collection. The remarks and notes listed for variable cWTS23101v also apply here.

4 Wave 4 weights

Eight sets of cross-sectional weights and 24 sets of longitudinal weights were computed at wave 4 of the ITC Japan Survey. These cross-sectional weights are listed in table 12, and their computation and how/when they should be used are detailed section 4.1. Likewise, the longitudinal weights are listed in table 13, and their computation and use are detailed in sections 4.2–4.4. All sampling weights for the ITC Japan Survey were computed using the statistical software R (http://www.r-project.org). As mentioned at the beginning of this document, these weights adjust for unequal selection probabilities, sample mis-representation, non-response and other biases. **It is thus essential to use weighted data, when preforming any analyses using ITC Japan data**.

4.1 Cross-sectional sampling weights

As mentioned at the beginning of section 4, 8 sets of cross-sectional weights were computed at wave 4 of the ITC Japan Survey. These are listed in table 12, and their computation as well as how and when they should be used are detailed in this section. Those weights were computed for the 4408 respondents interviewed at wave 4. This total excludes the 212 respondents that were deemed to be fraudulent (also referred to as speeders or falsifiers in some documentation). Those 4408 individuals consist of 1598 respondents newly recruited at wave 4, and of 2810 cohort respondents (1436 recruited at wave 1, 493 recruited at wave 2 and 881 recruited at wave 3) that were successfully retained at wave 4. It should be noted that all cross-sectional weights are designed to make the sample representative at the time of

Weight	Variable Names
Wave 4 cross-sectional inflation weights for all respondents	dWTS23100v
Rescaled wave 4 cross-sectional weights for all respondents	dWTS23101v
Rescaled wave 4 cross-sectional weights for cigarette smokers	dWTS23201v
Rescaled wave 4 cross-sectional weights for dual users	dWTS23401v
Rescaled wave 4 cross-sectional weights for quitters †	dWTS23501v
Rescaled wave 4 cross-sectional weights for all tobacco users	dWTS23601v
Rescaled wave 4 cross-sectional weights for HTP users	dWTS23701v
Rescaled wave 4 cross-sectional weights for non-users ‡	dWTS23801v

† Those who have quit smoking cigarettes within the last 2 years (note that some of those quitters are now HTP users)
‡ This group also includes long-term quitters (i.e., those who quit more than 2 years ago)

Table 12: Cross-sectional sampling weights for wave 4 of the ITC Japan Survey

wave 4 data collection. If the aim is to ensure that the sample is representative at the time of waves 1, 2 or 3 then the longitudinal sampling weights computed in sections 4.2–4.4 should be used instead.

1- Variable dWTS23100v contains the wave 4 cross-sectional inflation weights for all 4408 respondents interviewed at wave 4. Computation of those weights followed the same steps and used the same raking algorithm (see appendix A.1) as the wave 1 cross-sectional inflation weights (variable aWTS23100v of section 1.2). The only differences being:

   i) The wave 4 cross-sectional inflation weights are designed to make the sample representative of the target population at the time of wave 4. Consequently, for cohort respondents, age at wave 4 was used instead of their age at waves 1, 2 or 3. Likewise, variable dUserGrp was used instead of aUserGrp–cUserGrp, and similarly for education and geographic region. Recall that it is possible for a respondent to transition from being a cigarette only user at wave 1 to being classified as a dual user at wave 4. Other types of transitions are also possible. Likewise, it is possible for respondents to move between geographic regions.

   ii) A few more cells needed to be collapsed because they contained too few respondents.

   iii) Data from the 2021 JASTIS, instead of the 2017 JASTIS, was used to obtain benchmark/calibration figures; see the "Wave 4" column of the various tables in appendix A.2.

   These weights are designed to make respondents in each of the 6 groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 4 data collection. For example, the dWTS23100v weights of the 863 dual users are designed to make them representative of the Japanese population of dual users at the time of data collection; likewise for the other 5 groups.

   If interest lies in a target population that consists of two or more of the 6 user groups, the dWTS23100v weights are still appropriate. For example, when studying cigarette smokers, one can simply combine the dWTS23100v weights of the 1780 cigarette only users with those of the 863 dual users (for a total of 2643 respondents in the analysis), and assigned a weight of 0 to respondents in the other user groups. However, it simpler and, likely, more appropriate to use variable dWTS23201v in this situation. The

Weight	Variable Names
Waves 1–4 longitudinal inflation weights for all respondents	dWTS23120v
Rescaled waves 1–4 longitudinal weights for all respondents	dWTS23121v
Rescaled waves 1–4 longitudinal weights for cigarette smokers	dWTS23221v
Rescaled waves 1–4 longitudinal weights for dual users	dWTS23421v
Rescaled waves 1–4 longitudinal weights for quitters †	dWTS23521v
Rescaled waves 1–4 longitudinal weights for all tobacco users	dWTS23621v
Rescaled waves 1–4 longitudinal weights for HTP users	dWTS23721v
Rescaled waves 1–4 longitudinal weights for non-users ‡	dWTS23821v
Waves 2–4 longitudinal inflation weights for all respondents	dWTS23122v
Rescaled waves 2–4 longitudinal weights for all respondents	dWTS23123v
Rescaled waves 2–4 longitudinal weights for cigarette smokers	dWTS23223v
Rescaled waves 2–4 longitudinal weights for dual users	dWTS23423v
Rescaled waves 2–4 longitudinal weights for quitters †	dWTS23523v
Rescaled waves 2–4 longitudinal weights for all tobacco users	dWTS23623v
Rescaled waves 2–4 longitudinal weights for HTP users	dWTS23723v
Rescaled waves 2–4 longitudinal weights for non-users ‡	dWTS23823v
Waves 3–4 longitudinal inflation weights for all respondents	dWTS23124v
Rescaled waves 3–4 longitudinal weights for all respondents	dWTS23125v
Rescaled waves 3–4 longitudinal weights for cigarette smokers	dWTS23225v
Rescaled waves 3–4 longitudinal weights for dual users	dWTS23425v
Rescaled waves 3–4 longitudinal weights for quitters †	dWTS23525v
Rescaled waves 3–4 longitudinal weights for all tobacco users	dWTS23625v
Rescaled waves 3–4 longitudinal weights for HTP users	dWTS23725v
Rescaled waves 3–4 longitudinal weights for non-users ‡	dWTS23825v

† Those who have quit smoking cigarettes within the last 2 years (note that some of those quitters are now HTP users)
‡ This group also includes long-term quitters (i.e., those who quit more than 2 years ago)

Table 13: Longitudinal sampling weights for wave 4 of the ITC Japan Survey

same also applies to most of the possible combinations of the 6 user groups, and the weights listed below are likely to be more practical than these cross-sectional inflation weights.

Last but not least, since these are inflation/un-rescaled weights, they should not be used in analyses involving two or more countries. The various rescaled weights (i.e., variables dWTS23201v to dWTS23101v) described below were created especially for such multi-country analyses; see section 5.2 for more information on inflation versus rescaled weights.

2- Variable dWTS23201v contains the rescaled wave 4 cross-sectional weights for the 2643 (see table 14) respondents who were at least weekly cigarette smokers (and have smoked at least 100 cigarettes in

User group⋆	n
Cigarette smokers	
Cigarette only	1780
Dual users	863
Total	2643
Exclusive HTP users	797
Recent quitters†	
Using HTP	127
Not using HTP	198
Total	325
Non-users and long-term quitters‡	643
Total	4408

⋆ See section 1.1 for the description of those user groups, as well as beginning of section 3.1 for important change made on the classification of cigarette smokers

† Those who have quit smoking cigarettes within the last two years

‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 14: Wave 4 respondents by user group.

their lifetime) at the time of wave 4 data collection. These are simply the wave 4 cross-sectional inflation weights (variable `dWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 2643$). These weights are designed to make these 2643 cigarette smokers representative of the Japanese population at least weekly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 4 data collection.

3- Variable `dWTS23401v` contains the rescaled wave 4 cross-sectional weights for the 863 (see table 14) respondents who were dual users at the time of wave 4 data collection. These are simply the wave 4 cross-sectional inflation weights (variable `dWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 863$). These weights are designed to make these 863 dual users representative of the Japanese population of dual users at the time of wave 4 data collection. Recall that, at waves 3 and 4, dual users are defined as smoking cigarettes at least weekly, smoked at least 100 cigarettes in their lifetime, and using HTP at least weekly.

4- Variable `dWTS23501v` contains the rescaled wave 4 cross-sectional weights for the 325 (see table 14) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 4 data collection. These are simply the wave 4 cross-sectional inflation weights (variable `dWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 325$). These weights are designed to make these 325 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 4 data collection.

5- Variable `dWTS23601v` contains the rescaled wave 4 cross-sectional weights for the 3567 (see table 14) respondents who were at least weekly tobacco users (i.e., smoking cigarettes at least weekly and/or using HTP at least weekly) at the time of wave 4 data collection. These are simply the wave 4 cross-sectional inflation weights (variable `dWTS23100v`) of those respondents rescaled to sum to sample size (i.e., $n = 3567$). These weights are designed to make these 3567 tobacco users representative of the Japanese population of at least weekly tobacco users at the time of wave 4 data collection.

6- Variable dWTS23701v contains the rescaled wave 4 cross-sectional weights for the 1787 (see table 14) respondents who were at least weekly HTP users at the time of wave 4 data collection. These are simply the wave 4 cross-sectional inflation weights (variable dWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 1787$). These weights are designed to make these 1787 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 4 data collection.

7- Variable dWTS23801v contains the rescaled wave 4 cross-sectional weights for the 643 (see table 14) respondents who were non-users or long-term quitters at the time of wave 4 data collection. Recall that non-users are defined as smoking cigarettes less than weekly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the wave 4 cross-sectional inflation weights (variable dWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 643$). These weights are designed to make these 643 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 4 data collection.

8- Variable dWTS23101v contains the rescaled wave 4 cross-sectional weights for all 4408 respondents. These are simply the wave 4 cross-sectional inflation weights (variable dWTS23100v) of those respondents rescaled to sum to sample size (i.e., $n = 4408$). These weights are designed to make these 4408 respondents representative of the adult (20 years & older) Japanese population at the time of wave 4 data collection.

   These rescaled wave 4 cross-sectional weights preserve the ratios between the 6 user groups[1]. In other words, they preserve the ratios of: cigarette only smokers to exclusive HTP users, cigarette only smokers to non-users & long-term quitter, etc. Hence they are designed for analyses involving multiple or all 6 user groups; see section 5.2 for more information. However, it should be noted that tobacco users, recent quitters and non-users (and long term quitters) are ultimately distinct populations. Hence, great care must be taken when deciding to analyse them together using the dWTS23101v weights. This is probably fine when the goal is to carry out descriptive inference about the joint population of tobacco users, recent quitters and non-users. However, carrying out analytical inference (e.g., linear regression and logistic regression) from that same joint population is probably much more questionable.

   It should be noted that it is **not appropriate** to combine variables dWTS23201v – dWTS23801v to jointly analyse multiple user groups. Each of those variables was rescaled to sum to its respective sample size, and thus combining them will not preserve the ratios between the user groups resulting in erroneous results. This is why variable dWTS23101v was computed.

4.2 Waves 1–4 longitudinal sampling weights

As mentioned at the beginning of section 4, 24 sets of longitudinal weights were computed at wave 4 of the ITC Japan Survey. A third of those sets (see top third of table 13) were computed for respondents recruited at wave 1, and then successfully retained and interviewed at wave 4 (as well as at every wave in between). This section details the computation of those weights, as well as how and when they should be used. It should be noted that all waves 1–4 longitudinal weights are designed to make the sample

[1] See section 1.1 for the description of those user groups, as well as beginning of section 3.1 for the important change made on the classification of cigarette smokers.

representative at the time of wave 1. If the aim is to ensure that the sample is representative at the time of wave 4 data collection, then the cross-sectional sampling weights computed in section 4.1 should be used instead.

User group*	n
Cigarette smokers	
Cigarette only	1102
Dual users	129
Total	1231
Exclusive HTP users	11
Recent quitters†	
Using HTP	10
Not using HTP	13
Total	23
Non-users and long-term quitters‡	156
Total	1421

* See section 1.1 for the description of those user groups
† Those who have quit smoking cigarettes within the last two years
‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 15: Wave 1 respondents successfully recontacted at wave 4 by user group.

1- Variable `dWTS23120v` contains the waves 1–4 longitudinal inflation weights for all 1421 respondents recruited at wave 1, and then successfully recontacted at wave 4 (as well as at every wave in between). These weights are designed to make respondents in each of the 6 user groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 1 data collection.

   These waves 1–4 longitudinal inflation weights are the waves 1–3 longitudinal inflation weights (i.e., variable `cWTS23120v` computed in section 3.2) adjusted for attrition between waves 3 and 4. Consequently, computation of those longitudinal inflation weights followed the same steps as described in the computation of variable `bWTS23120v`. Respondents were thus divided into the 6 user groups based on their wave 1 status[1]; similarly, age, education and geographic region at wave 1 were used. Since these weights are meant to make these 1421 respondents representative of the population at the time of wave 1 data collection, the benchmark/calibration figures obtained from the 2017 JASTIS and used to compute the wave 1 cross-sectional weights were used here as well; as opposed to the 2021 figures used to compute the wave 4 cross-sectional weights (see section 4.1).

2- Variable `dWTS23221v` contains the rescaled waves 1–4 longitudinal weights for the 1231 (see table 15) respondents who were at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection, and then successfully recontacted at wave 4 (as well as at every wave in between). These are simply the waves 1–4 longitudinal inflation weights (variable `dWTS23120v`) of those respondents rescaled to sum to sample size (i.e., $n = 1231$). These

[1]Since the user group at wave 1 (i.e., variable `aUsergrp`) is used to compute the longitudinal weights, respondents smoking at least monthly were classified as cigarette smokers; as opposed to the at least weekly requirement for the wave 3 cross-sectional weights of section 3.1.

weights are designed to make these 1231 cigarette smokers representative of the Japanese population of at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 1 data collection.

3- Variable `dWTS23421v` contains the rescaled waves 1–4 longitudinal weights for the 129 (see table 15) respondents who were dual users at the time of wave 1 data collection, and then successfully recontacted at wave 4 (as well as at every wave in between). These are simply the waves 1–4 longitudinal inflation weights (variable `dWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 129 dual users representative of the Japanese population of dual users at the time of wave 1 data collection.

4- Variable `dWTS23521v` contains the rescaled waves 1–4 longitudinal weights for the 23 (see table 15) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 1 data collection, and then successfully recontacted at wave 4 (as well as at every wave in between). These are simply the waves 1–4 longitudinal inflation weights (variable `dWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 23 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 1 data collection. However, given that there are only 23 such respondents, any analyses focusing on those recent quitters is at best questionnable.

5- Variable `dWTS23621v` contains the rescaled waves 1–4 longitudinal weights for the 1252 (see table 15) respondents who were tobacco users (i.e., smoking cigarettes at least monthly and/or using HTP at least weekly) at the time of wave 1 data collection, and then successfully recontacted at wave 4 (as well as at every wave in between). These are simply the waves 1–4 longitudinal inflation weights (variable `dWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 1252 tobacco users representative of the Japanese population of tobacco users at the time of wave 1 data collection.

6- Variable `dWTS23721v` contains the rescaled waves 1–4 longitudinal weights for the 150 (see table 15) respondents who were at least weekly HTP users at the time of wave 1 data collection, and then successfully recontacted at wave 4 (as well as at every wave in between). These are simply the waves 1–4 longitudinal inflation weights (variable `dWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 150 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 1 data collection.

7- Variable `dWTS23821v` contains the rescaled waves 1–4 longitudinal weights for the 156 (see table 15) respondents who were non-users or long-term quitters at the time of wave 1 data collection, and then successfully recontacted at wave 4 (as well as at every wave in between). Recall that non-users are defined as smoking cigarettes less than monthly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the waves 1–4 longitudinal inflation weights (variable `dWTS23120v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 156 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 1 data collection.

8- Variable `dWTS23121v` contains the rescaled waves 1–4 longitudinal weights for all 1421 respondents recruited at wave 1, and then successfully recontacted at wave 4 (as well as at every wave in between).

These are simply the waves 1–4 longitudinal inflation weights (variable dWTS23120v) of those respondents rescaled to sum to sample size (i.e., $n = 1421$). These weights are designed to make these 1421 respondents representative of the adult (20 years & older) Japanese population at the time of wave 1 data collection.

4.3 Waves 2–4 longitudinal sampling weights

As mentioned at the beginning of section 4, 24 sets of longitudinal weights were computed at wave 4 of the ITC Japan Survey. A third of those sets (see middle third of table 13) were computed for respondents who completed the wave 2 survey (i.e., recruited at either wave 1 or 2), and then successfully retained and interviewed at both waves 3 and 4. This section details the computation of those weights, as well as how and when they should be used. It should be noted that all waves 2–4 longitudinal weights are designed to make the sample representative at the time of wave 2. If the aim is to ensure that the sample is representative at the time of wave 4 data collection, then the cross-sectional sampling weights computed in section 4.1 should be used instead.

User group⋆	n
Cigarette smokers	
Cigarette only	934
Dual users	467
Total	1401
Exclusive HTP users	174
Recent quitters†	
Using HTP	83
Not using HTP	42
Total	125
Non-users and long-term quitters‡	229
Total	1929

⋆ See section 1.1 for the description of those user groups
† Those who have quit smoking cigarettes within the last two years
‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 16: Wave 2 respondents successfully recontacted at wave 4 by user group.

1- Variable dWTS23122v contains the waves 2–4 longitudinal inflation weights for all 1929 respondents who completed the wave 2 survey (i.e., recruited at either wave 1 or 2), and were then successfully recontacted at both waves 3 and 4. These weights are designed to make respondents in each of the 6 user groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 2 data collection.

These waves 2–4 longitudinal inflation weights are the waves 2–3 longitudinal inflation weights (i.e., variable cWTS23122v computed in section 3.3) adjusted for attrition between waves 3 and 4. Consequently, computation of those longitudinal inflation weights followed the same steps as described in the computation of variable bWTS23100v. Respondents were thus divided into the 6 user groups

based on their wave 2 status[1]; similarly, age, education and geographic region at wave 2 were used. Since these weights are meant to make these 1929 respondents representative of the population at the time of wave 2 data collection, the benchmark/calibration figures obtained from the 2019 JASTIS and used to compute the wave 2 cross-sectional weights were used here as well; as opposed to the 2021 figures used to compute the wave 4 cross-sectional weights (see section 4.1).

2- Variable `dWTS23223v` contains the rescaled waves 2–4 longitudinal weights for the 1401 (see table 16) respondents who were at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 2 data collection, and then successfully recontacted at both waves 3 and 4. These are simply the waves 2–4 longitudinal inflation weights (variable `dWTS23122v`) of those respondents rescaled to sum to sample size (i.e., $n = 1401$). These weights are designed to make these 1401 cigarette smokers representative of the Japanese population of at least monthly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 2 data collection.

3- Variable `dWTS23423v` contains the rescaled waves 2–4 longitudinal weights for the 467 (see table 16) respondents who were dual users at the time of wave 2 data collection, and then successfully recontacted at both waves 3 and 4. These are simply the waves 2–4 longitudinal inflation weights (variable `dWTS23122v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 467 dual users representative of the Japanese population of dual users at the time of wave 2 data collection.

4- Variable `dWTS23523v` contains the rescaled waves 2–4 longitudinal weights for the 125 (see table 16) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 2 data collection, and then successfully recontacted at both waves 3 and 4. These are simply the waves 2–4 longitudinal inflation weights (variable `dWTS23122v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 125 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 2 data collection. However, given that there are only 125 such respondents, any analyses focusing on those recent quitters is questionnable.

5- Variable `dWTS23623v` contains the rescaled waves 2–4 longitudinal weights for the 1658 (see table 16) respondents who were tobacco users (i.e., smoking cigarettes at least monthly and/or using HTP at least weekly) at the time of wave 2 data collection, and then successfully recontacted at both waves 3 and 4. These are simply the waves 2–4 longitudinal inflation weights (variable `dWTS23122v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 1658 tobacco users representative of the Japanese population of tobacco users at the time of wave 2 data collection.

6- Variable `dWTS23723v` contains the rescaled waves 2–4 longitudinal weights for the 724 (see table 16) respondents who were at least weekly HTP users at the time of wave 2 data collection, and then successfully recontacted at both waves 3 and 4. These are simply the waves 2–4 longitudinal inflation weights (variable `dWTS23122v`) of those respondents rescaled to sum to sample size. These weights are designed to make these 724 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 2 data collection.

[1]Since the user group at wave 2 (i.e., variable `bUsergrp`) is used to compute the longitudinal weights, respondents smoking at least monthly were classified as cigarette smokers; as opposed to the at least weekly requirement for the wave 4 cross-sectional weights of section 4.1.

7- Variable dWTS23823v contains the rescaled waves 2–4 longitudinal weights for the 229 (see table 16) respondents who were non-users or long-term quitters at the time of wave 2 data collection, and then successfully recontacted at both waves 3 and 4. Recall that non-users are defined as smoking cigarettes less than monthly or not at all, and using HTP less than weekly or not at all; whereas, long-term quitters are defined as having quit smoking cigarette more than two years ago (and smoked at least 100 cigarettes in their lifetime prior to that) and using HTP less than weekly or not at all. These are simply the waves 2–4 longitudinal inflation weights (variable dWTS23122v) of those respondents rescaled to sum to sample size. These weights are designed to make these 229 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 2 data collection.

8- Variable dWTS23121v contains the rescaled waves 2–4 longitudinal weights for all 1929 respondents who completed the wave 2 survey, and were then successfully recontacted at both waves 3 and 4. These are simply the waves 2–4 longitudinal inflation weights (variable dWTS23122v) of those respondents rescaled to sum to sample size (i.e., $n = 1929$). These weights are designed to make these 1929 respondents representative of the adult (20 years & older) Japanese population at the time of wave 2 data collection.

4.4 Waves 3–4 longitudinal sampling weights

As mentioned at the beginning of section 4, 24 sets of longitudinal weights were computed at wave 4 of the ITC Japan Survey. A third of those sets (see bottom third of table 13) were computed for respondents who completed the wave 3 survey (i.e., recruited at either wave 1, 2, or 3), and then successfully retained and interviewed at wave 4. This section describes those weights, and is akin the other sub-sections of section 4. It should be noted that all waves 3–4 longitudinal weights are designed to make the sample representative at the time of wave 3. If the aim is to ensure that the sample is representative at the time of wave 4 data collection, then the cross-sectional sampling weights computed in section 4.1 should be used instead.

1- Variable dWTS23124v contains the waves 3–4 longitudinal inflation weights for all 2810 respondents who completed the wave 3 survey (i.e., recruited at either wave 1, 2 or 3), and were then successfully recontacted at wave 4. These weights are designed to make respondents in each of the 6 user groups representative (with respect to gender, age, education and region) of the corresponding population at the time of wave 3 data collection.

   These waves 3–4 longitudinal inflation weights are the waves 3 cross-sectional inflation weights (i.e., variable cWTS23100v computed in section 3.1) adjusted for attrition between waves 3 and 4. Consequently, computation of those longitudinal inflation weights followed the same steps as described in the computation of variable cWTS23100v. Respondents were thus divided into the 6 user groups based on their wave 3 status; similarly, age, education and geographic region at wave 3 were used. Since these weights are meant to make these 2810 respondents representative of the population at the time of wave 3 data collection, the benchmark/calibration figures obtained from the 2020 JASTIS and used to compute the wave 3 cross-sectional weights were used here as well; as opposed to the 2021 figures used to compute the wave 4 cross-sectional weights (see section 4.1).

2- Variable dWTS23225v contains the rescaled waves 3–4 longitudinal weights for the 1922 (see table 17)

User group⋆	n
Cigarette smokers	
Cigarette only	1320
Dual users	602
Total	1922
Exclusive HTP users	330
Recent quitters†	
Using HTP	61
Not using HTP	115
Total	176
Non-users and long-term quitters‡	382
Total	2810

⋆ See section 1.1 for the description of those user groups, as well as beginning of section 3.1 for important change made on the classification of cigarette smokers

† Those who have quit smoking cigarettes within the last two years

‡ This group also includes long-term quitters (i.e., those who have quit smoking cigarettes more than 2 years ago)

Table 17: Wave 3 respondents successfully recontacted at wave 4 by user group.

respondents who were at least weekly[1] cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 3 data collection, and then successfully recontacted at wave 4. These weights are designed to make these 1922 cigarette smokers representative of the Japanese population of at least weekly cigarette smokers (and have smoked at least 100 cigarettes in their lifetime) at the time of wave 3 data collection.

3- Variable `dWTS23425v` contains the rescaled waves 3–4 longitudinal weights for the 602 (see table 17) respondents who were dual users[2] at the time of wave 3 data collection, and then successfully recontacted at wave 4. These weights are designed to make these 602 dual users representative of the Japanese population of dual users at the time of wave 3 data collection.

4- Variable `dWTS23525v` contains the rescaled waves 3–4 longitudinal weights for the 176 (see table 17) respondents who had recently (i.e., within the last 2 years) quit smoking cigarettes (and had smoked at least 100 cigarettes in their lifetime prior to that) at the time of wave 3 data collection, and then successfully recontacted at wave 4. These weights are designed to make these 176 recent quitters representative of the Japanese population of recent quitters (i.e., within the last 2 years) at the time of wave 3 data collection. However, given that there are only 176 such respondents, any analyses focusing on those recent quitters is questionnable.

5- Variable `dWTS23625v` contains the rescaled waves 3–4 longitudinal weights for the 2313 (see table 17) respondents who were tobacco users (i.e., smoking cigarettes at least weekly and/or using HTP at least weekly) at the time of wave 3 data collection, and then successfully recontacted at wave 4. These weights are designed to make these 2313 tobacco users representative of the Japanese population of tobacco users at the time of wave 3 data collection.

[1] Recall (see beginning of section 3.1) that, starting at wave 3, to be as classified a cigarette smoker a respondent had to smoke cigarettes at least weekly; instead of at least monthly at waves 1 and 2.

[2] Note that the switch from at least monthly to at least weekly in the classification of cigarette smokers also carries to dual users.

6- Variable dWTS23725v contains the rescaled waves 3–4 longitudinal weights for the 993 (see table 17) respondents who were at least weekly HTP users at the time of wave 3 data collection, and then successfully recontacted at wave 4. These weights are designed to make these 993 HTP users representative of the Japanese population of at least weekly HTP users at the time of wave 3 data collection.

7- Variable dWTS23825v contains the rescaled waves 3–4 longitudinal weights for the 382 (see table 17) respondents who were non-users or long-term quitters at the time of wave 3 data collection, and then successfully recontacted at wave 4. These weights are designed to make these 382 respondents representative of the Japanese population of non-users and long-term quitters at the time of wave 3 data collection.

8- Variable dWTS23125v contains the rescaled waves 3–4 longitudinal weights for all 2810 respondents who completed the wave 3 survey, and were then successfully recontacted at wave 4. These are simply the waves 3–4 longitudinal inflation weights (variable dWTS23124v) of those respondents rescaled to sum to sample size (i.e., $n = 2810$). These weights are designed to make these 2810 respondents representative of the adult (20 years & older) Japanese population at the time of wave 3 data collection.

5 Remarks and cautionary notes

This section provides remarks and cautionary notes when analyzing ITC Japan data. These notes echo some of the limitations of the ITC Japan Survey. In addition to this section, users should also look at the cautionary notes included in the descriptions of variables aWTS23101v, bWTS23101v, cWTS23101v and dWTS23101v.

5.1 Estimating prevalences

Though it offers rich possibilities for analyses, the ITC Japan Survey is not suited for estimating prevalences. This is in large due to its sampling design, where respondents are sampled at different sampling rates depending on which of the 6 user groups (section 1.1) they belong to. For example, at wave 1, cigarette only smokers were sampled at a rate for 19 per 100,000, dual users at a rate of 80 per 100,000, and non-users/long-term quitters at a rate just below 1 per 100,000. This sampling design was chosen to ensure sufficiently large samples sizes for the key user groups, while keeping the overall sample size and cost reasonable.

The downside of such a sampling design is that it requires the use of external benchmark/calibration data to compute sampling weights. The Japan Society and Tobacco Internet Study (JASTIS) was used for this purpose; yielding the benchmark/calibration figures found in appendix A.2. A raking algorithm (see appendix A.1) was then used to compute the weights of the ITC Japan Survey based on these figures. The consequence and drawback of using external benchmark/calibration figures to compute the weights is that the various prevalence estimates obtained from the ITC Japan data are actually those of the JASTIS benchmark survey. For example, the estimated prevalence of cigarette smoking obtained from ITC Japan data is actually that of the JASTIS; likewise for the prevalences of HTP, dual use, quitting, etc. Furthermore, since the weights were calibrated on user group × gender, user group × age

group, and user group × education, this drawback also holds true when looking at prevalences within gender, age group and education.

5.2 Inflation versus rescaled weights

A key reason for rescaling the weights is to facilitate joint analyses involving data from multiple ITC countries. The 2017 JASTIS survey was used to calibrate the wave 1 cross-sectional inflation weights (variable aWTS23100v). Based on that survey, there were about 18.3 million cigarette smokers (ages 20 & older) in Japan at the time of wave 1 data collection. Data from the 2016 National Health Interview Survey (NHIS) was used to calibrate the weights of the US sample for wave 1 of the ITC Four Country Smoking and Vaping Survey (4CV) Survey, which was conducted in 2016. According to the 2016 NHIS, there were about 39.8 million cigarette smokers (ages 18 & older) in the United States at the time of data collection. Hence, any joint analysis using data from ITC Japan and the US sample of the 4CV Survey will be dominated by the US if the inflation weights (variable aWTS23100v) are used.

On the other hand, all of the rescaled weights sum to the appropriate sample size, as described in the previous sections; and likewise for the 4CV Survey. Hence, if the rescaled weights are used, Japan will have a slightly greater impact on the results (e.g., the ITC Japan wave 1 sample of cigarette smokers consists of 3837 respondents; whereas the 4CV sample of US cigarette smokers consists of 2327 respondents), but no country will dominate the analysis. In summary, rescaling the weights to sum to the sample size is a simple and efficient way to make countries with different population sizes comparable. This also holds true when comparing ITC Japan data to other ITC countries.

Last but not least, it should be mentioned that rescaling the weights will not affect the results when estimating population means and proportions/percentages, as well as when fitting various statistical models (e.g., logistic and linear regressions). However, the rescaled weights should not be used to estimate population totals (e.g., the total number of daily smokers or HTP users).

5.3 Covariates to include in statistical modelling

As with other surveys, it is good practice to include the survey design variables and the variables used in the weight construction, when fitting statistical models (e.g., linear or logistic regression models) using ITC Japan data. Hence, we highly recommend that any statistical model includes the following covariates: user group, gender and age (either as a continuous or as a categorical covariate). Though somewhat less critical, users should also strongly consider adding education to their statistical model(s). The geographic region should also be used as the stratification variable in the statistical software. Table 18 below lists those variables.

5.4 Weekly vs. monthly smokers

As explained at the beginning of section 3.1 (page 13), an important change was made at wave 3 in how respondents are classified into the 6 user groups of section 1.1. Starting at wave 3, respondents had to smoke at least weekly (and to have smoked at least 100 cigarettes in their lifetime) to be classified as cigarette smokers; whereas, the requirement was to smoke at least monthly at waves 1 and 2. Consequently, as described in section 3.1, only those who smoked cigarettes at least weekly were assigned rescaled wave 3 cross-sectional weights for cigarette smokers (i.e., variable cWTS23201v) and, depending

	Variable name			
Covariate	Wave 1	Wave 2	Wave 3	Wave 4
gender		GENDER		
age (continuous)	aAGE	bAGE	cAGE	dAGE
age (categorical)	aAGEGRP	bAGEGRP	cAGEGRP	dAGEGRP
user group	aUsergrp	bUsergrp	cUsergrp	dUsergrp
education	aDE23312v	bDE23312v	cDE23312v	dDE23312v
geographic region/strata	aStrata	bStrata	cStrata	dStrata

Table 18: Covariates to be include in statistical models.

on their use of HTP, rescaled wave 3 cross-sectional weights for dual users (i.e., variable cWTS23401v). Likewise, only those who smoked cigarettes at least weekly were treated as cigarette only smokers/dual users, when computing the wave 3 cross-sectional inflation weights (i.e., variable cWTS23100v). The same logic was also applied when computing the various wave 4 cross-sectional weights (section 4.1) and the waves 3–4 longitudinal weights (section 4.4). This is different than in most ITC countries, where respondents need to smoke cigarettes at least monthly (and to have smoked at least 100 cigarettes in their lifetime) to be classified as cigarette smokers. This difference raises comparability issues when comparing waves 3 and 4 to waves 1 and 2, as well as when comparing waves 3 and/or 4 data with that of other ITC countries.

Before looking at solutions, it is important to note that the overall eligibility criterion for the ITC Japan Survey were not changed at wave 3 (and subsequent waves). Rather, respondents were simply classified slightly differently between the 6 user groups at wave 3 (and subsequent waves) as opposed to the previous waves. This change impacted not only the sampling weights, but also the quotas and thus the sampling rates as well. However, the routing of the ITC Japan survey questionnaire is such that less than weekly but at least monthly cigarette smokers were asked the relevant smoking questions. It is thus possible to include less than weekly but at least monthly cigarette smokers when carrying out statistical analyses of smokers. Their weights will tend to be higher than those of at least weekly smokers because at waves 3 and 4 they were sampled at a lower rate.

If the aim is to compare smokers at waves 3 and/or 4 with those at waves 1 and/or 2, then the simplest and probably best approach is to restrict the analysis to at least weekly smokers at all waves. In other words, exclude the less than weekly but at least monthly cigarette smokers from the analysis. This will ensure that respondents from all waves are comparable. Note that this approach of ensuring that inclusion criterion are the same for all waves applies to all analyses. For example, any analysis of non-users and long-term quitters would need to exclude the waves 3 and 4 respondents who are smoking cigarettes less than weekly but at least monthly and using HTP less than weekly or not. This is required as those same individuals would have been classified as a cigarette only smoker at the previous two waves, and thus excluded from analyses involving non-users and long-term quitters.

If the aim is to compare smokers in Japan at wave 3 with those of other ITC countries (where less than weekly but at least monthly smokers are classified as smokers), then using the rescaled wave 3 cross-sectional weights for all respondents (i.e., variable cWTS23101v of section 3.1) is recommended. Under this approach, a less than weekly cigarette smoker who uses HTP at least weekly is treated as an exclusive HTP user when computing his/her sampling weights; echoing the fact that his/her behaviour is closer to those individuals than that of a cigarette smoker. Likewise, a less than weekly cigarette smoker who uses HTP less than weekly or not at all is treated as an non-user when computing his/her sampling weights; again, echoing the fact that his/her behaviour is closer to those individuals than

that of a cigarette smoker. Nevertheless, this is not a perfect solution, and care should be taken when performing such analyses. An alternative approach would be to exclude less than weekly but at least monthly cigarette smokers from the other ITC countries being compared to wave 3 of ITC Japan. The same approach (i.e., using variable dWTS23101v of section 4.1) can also be used at wave 4.

Although this causes some comparison issues to earlier waves and with other ITC countries, as outlined above, these small changes to the definition of the 6 user groups were necessary to adapt to the evolving landscape of tobacco use.

Funding & Acknowledgements

The ITC Japan Survey was supported by the Canadian Institutes of Health Research (CIHR) (grant # FDN-148477).

We are both indebted and grateful to Dr. Takahiro Tabuchi and his research team for allowing us to use data from their Japan Society and Tobacco Internet Study (JASTIS) to compute the benchmark/-calibration figures used in the calculations of all sampling weights of the ITC Japan Survey.

A Appendix

A.1 Raking algorithm

This section details the raking algorithm used to calibrate the wave 1 cross-sectional inflation weights (i.e., variable aWT23S100v of section 1.2) to the benchmark figures of section A.2. However, the same logic applies to the other inflations weights.

Step 1: Let $w_i^{(0)}$ be the start weight of the i^{th} respondent. If this is the very first iteration of the raking algorithm, then $w_i^{(0)} = 1$ for all respondents; otherwise, $w_i^{(0)} = w_i^{(4)}$, as computed in step 4 below.

The $w_i^{(0)}$ weights are calibrated to the user groups $\times$ gender benchmark figures of the first table of section A.2. To this end, the respondents were divided into the 12 user groups/gender cells of that table. For respondents in cell k ($k = 1, \ldots, 12$), this calibration/post-stratification adjustment consists in multiplying their $w_i^{(0)}$ weights by $\widehat{N}_k/t_k$ to produce calibrated $w_i^{(1)}$ weights. These $w_i^{(1)}$ weights are such that their sum over all respondents in cell k is equal to $\widehat{N}_k$, the estimated number of individuals in that cell (as obtained from the 2017 JASTIS). Let k be the cell to which the i^{th} respondent belongs to, the $w_i^{(1)}$ weight of that respondent is given by

$$w_i^{(1)} = w_i^{(0)} \times \frac{\widehat{N}_k}{t_k} = w_i^{(0)} \times \frac{\widehat{N}_k}{\sum\limits_{i \in C_k} w_i^{(0)}}$$

where $\widehat{N}_1, \ldots, \widehat{N}_{12}$ are given in column 3 of the first table of section A.2 and C_k is the set of all respondents in cell k.

Step 2: Using the same post-stratification technique described in step 1, the $w_i^{(1)}$ weights were then calibrated to the user groups $\times$ age benchmark figures of the second table of section A.2. The $w_i^{(1)}$ weights of respondents in cell ℓ ($\ell = 1, \ldots, 20$) were then multiply by a factor, $\widehat{N}_\ell^{(2)}/t_\ell^{(2)}$, to produce calibrated $w_i^{(2)}$ weights. Let ℓ be the cell to which the i^{th} respondent belongs to, the $w_i^{(2)}$ weight of that respondent is given by

$$w_i^{(2)} = w_i^{(1)} \times \frac{\widehat{N}_\ell^{(2)}}{t_\ell^{(2)}} = w_i^{(1)} \times \frac{\widehat{N}_\ell^{(2)}}{\sum\limits_{i \in C_\ell^{(2)}} w_i^{(1)}}$$

where $\widehat{N}_1^{(2)}, \ldots, \widehat{N}_{20}^{(2)}$ are given in column 3 of the second table of section A.2 and $C_\ell^{(2)}$ is the set of all respondents in cell ℓ.

It is important to recognize that this second calibration partially destroys the calibration done in step 1; in other words, we no longer necessary have that

$$\sum_{i \in C_k} w_i^{(2)} = \widehat{N}_k \quad \text{for } k = 1, \ldots, 12$$

where C_k and $\widehat{N}_k$ were defined in step 1 above. Because of this, step 1 will need to be repeated (most likely multiple times) after calibrating to the other benchmark figures of section A.2; see step 5 below.

Step 3: The $w_i^{(2)}$ weights were then calibrated to the user groups × education benchmark figures of the third table of section A.2. This was done using the same post-stratification technique as detailed above, and yielded the $w_i^{(3)}$ weights. As in step 2, this third calibration partially destroys the calibration done in steps 1 and 2, and those two steps will need to be repeated.

Step 4: The $w_i^{(3)}$ weights were then calibrated to the geographic benchmark figures of the fourth table of section A.2. This was done using the same post-stratification technique as detailed above, and yielded the $w_i^{(4)}$ weights. As in step 2, this third calibration partially destroys the calibration done in the previous steps, and those two steps will need to be repeated.

Step 5: Repeat steps 1–4 until convergence; that is until,

$$\sum_{i \in C_k} w_i^{(4)} = \widehat{N}_k \quad \text{for } k = 1, \ldots, 12 \qquad \sum_{i \in C_\ell^{(2)}} w_i^{(4)} = \widehat{N}_\ell^{(2)} \quad \text{for } \ell = 1, \ldots, 20$$

and likewise for the calibrations on education (step 3) and geographic region (step 4). In other words, repeating steps 1–4 until convergence ensures that the weights are calibrated to all of the benchmark figures of section A.2.

Note that weight trimming (and redistribution) was done at some of the above mentioned steps. This was done to prevent extreme weight variation arising from a few respondents having very large sampling weights.

A.2 Benchmark/calibration figures

The benchmark/calibration figures used to compute the weights are given in the tables below. The 2017 Japan Society and Tobacco Internet Study (JASTIS) was used for the wave 1 cross-sectional weights[1]; whereas, the 2019 JASTIS was used for the wave 2 cross-sectional weights[2], the 2020 JASTIS was used for the wave 3 cross-sectional weights[3], and the 2021 JASTIS was used for the wave 4 cross-sectional weights. For waves 1 and 2, the figures are for the user groups defined in section 1.1; whereas, the revised/updated user groups (see beginning of section 3.1 on page 13) were used for the figures at waves 3 and 4. It should be noted that the 2019, 2020 and 2021 JASTIS surveys did not ask respondent how long ago they had quit smoking. It was thus impossible (using solely data from those surveys) to divide quitters between recent (i.e., within the last 2 years) and long-term. To get around this, data from the 2017 JASTIS (where respondents were asked how long ago they had quit smoking) was combined to the 2019/20/21 JASTIS surveys to produce the tables below.

We are grateful to Dr. Takahiro Tabuchi and his research team for allowing us to use the JASTIS data to compute those benchmark figures.

[1]The 2017 JASTIS was also used to compute the waves 1–2 longitudinal weights, the waves 1–3 longitudinal weights and the waves 1–4 longitudinal weights.

[2]The 2019 JASTIS was also used to compute the waves 2–3 longitudinal weights and the waves 2–4 longitudinal weights.

[3]The 2020 JASTIS was also used to compute the waves 3–4 longitudinal weights.

User group	Gender	# Individuals Wave 1	Wave 2	Wave 3	Wave 4
cigarette only	male	11,952,484	7,962,600	7,349,263	7,617,433
	female	5,265,064	3,421,906	3,170,657	3,650,312
dual users	male	886,869	4,381,968	4,070,523	4,798,115
	female	221,274	1,341,241	1,273,668	1,777,070
HTP only	male	257,299	255,370	378,734	338,457
	female	110,247	84,341	276,584	101,601
quitters using HTP	male	419,839	419,839	75,836	101,316
	female	139,738	139,738		
quitters not using HTP	male	743,091	1,067,435	956,130	1,085,022
	female	367,556	594,146	559,639	578,847
non users	male	32,568,301	33,307,290	33,643,010	32,559,858
	female	41,597,269	43,037,367	41,925,616	41,103,971

User group	Age	# Individuals Wave 1	Wave 2	Wave 3	Wave 4
cigarette only	[20, 25)	578,386	514,586	231,869	352,856
	[25, 35)	2,235,165	1,073,741	895,994	957,570
	[35, 45)	3,653,059	2,172,151	1,820,290	2,043,784
	[45, 55)	3,793,415	2,815,215	2,762,256	2,869,139
	[55, 85]	6,957,523	4,808,813	4,809,511	5,044,396
dual users	[20, 25)	75,463	496,868	284,064	369,590
	[25, 35)	314,564	1,025,905	941,754	1,092,814
	[35, 45)	307,334	1,427,590	1,331,461	1,715,903
	[45, 55)	171,500	1,114,222	1,362,351	1,683,831
	[55, 85]	239,282	1,658,624	1,424,561	1,713,047
HTP only	[20, 35)	367,546	115,286	655,318	440,058
	[35, 85]		224,425		
quitters using HTP	[20, 35)	115,878	115,878	75,836	101,316
	[35, 85]	443,699	443,699		
quitters not using HTP	[20,25)	204,627	99,619	98,187	69,978
	[25,35)		214,128	211,100	234,532
	[35,45)		392,000	370,594	398,917
	[45,55)	906,020	165,206	160,272	168,990
	[55,81]		790,628	675,616	791,452
non users	[20, 25)	5,404,781	5,127,183	5,574,577	5,451,860
	[25, 35)	10,667,215	10,635,477	10,675,936	10,573,679
	[35, 45)	13,246,315	12,404,429	12,609,321	11,938,486
	[45, 55)	12,885,728	13,708,458	13,879,392	13,487,095
	[55, 85]	31,961,531	34,469,110	32,829,400	32,212,709

User group	Education	# Individuals Wave 1	Wave 2	Wave 3	Wave 4
cigarette only	low	5,563,844	3,870,491	3,658,426	4,234,591
	medium	3,841,440	2,694,777	2,393,975	2,488,591
	high	7,812,264	4,819,238	4,467,519	4,544,563
dual users	low	317,422	1,653,500	1,464,287	2,084,159
	medium	220,537	1,021,681	1,158,091	1,247,652
	high	570,184	3,048,028	2,721,813	3,243,374
HTP only	low & medium	367,546	178,210	350,393	181,251
	high		161,501	304,925	258,807
quitters using HTP	low & medium	210,018	210,017	75,836	101,316
	high	349,559	349,560		
quitters not using HTP	low	659,447	613,271	551,888	690,852
	medium		340,961	326,343	315,448
	high	451,200	707,349	637,538	657,569
non users	low	20,562,672	20,367,408	19,838,052	20,533,482
	medium	20,308,807	19,778,452	20,257,523	17,703,180
	high	33,294,091	36,198,797	35,473,051	35,427,167

Region†	# Individuals Wave 1	Wave 2	Wave 3	Wave 4
Hokkaido	4,007,804	4,046,891	3,893,949	3,916,290
Tohoku	6,558,000	6,563,035	6,339,793	6,334,897
Kanto	32,980,907	33,660,468	33,069,460	33,072,456
Chubu	15,744,074	16,002,778	15,595,701	15,582,516
Kansai	16,782,730	16,986,371	16,590,336	16,582,487
Chugoku	5,384,000	5,429,747	5,276,265	5,277,529
Shikoku	2,777,000	2,797,400	2,679,210	2,692,148
Kyushu & Okinawa	10,294,515	10,526,549	10,234,945	10,253,680

† These are the 8 regions of Japan; see figure 1.

Figure 1: Regions of Japan.

A.3 Pseudo code

Pseudo code detailing how variable `aUserGrp` (abbreviated `UserGrp`) was created.

```
if {smokes cigarettes at least monthly (ie, FR23309v in (1,2,3))} then
   do;
      if {uses HTP at least weekly (ie, HN23309v in (1,2))} then
         UserGrp=3; /* dual */
      else then UserGrp=1; /* cig only */
   end;

else if {non-smoker or quitter (ie, FR23309v=9)} and {QA23439=n/a} then
   do;
      if {uses HTP at least weekly (ie, HN23309v in (1,2))} then
         UserGrp=2; /* HTP only */
      else then UserGrp=4; /* non-user */
   end;

else if {non-smoker or quitter (ie, FR23309v=9)}
         and {quit within the last 2 years (ie, QA23439 in (1:4))} then
   do;
      if {uses HTP at least weekly (ie, HN23309v in (1,2))} then
         UserGrp=5; /* quitter using HTP at least weekly */
      else then UserGrp=6; /* quitter using HTP less than weekly */
   end;

else if {non-smoker or quitter (ie, FR23309v=9)}
        and {quit more than 2 years ago (ie, QA23439 in (5,6,7,88,99))} then
   do;
      if {uses HTP at least weekly (ie, HN23309v in (1,2))} then
         UserGrp=2; /* HTP only */
      else then UserGrp=4; /* non-user */
   end;
```

Note: the response option for variable `QA23439` changed between waves 1 and 2; at wave 1, response

options 1–4 corresponds to those who quit within the last 2 years.