Consumer Reported Outcome Measures Consortium Task Force

Technical Report

Best Practices and Guidelines with Respect to Descriptive CROM for Research on Tobacco and Nicotine Containing Products

April 2024

Study Coordinator and Authors:

Lai Wei*, Mohamadi Sarkar, Altria Client Services LLC, Richmond, Virginia, USA

Emilie Clerc*, Christelle Chrea, PMI R&D, Philip Morris Products S.A., Switzerland

Stacey McCaffrey*, Nicholas Goldenson , Juul Labs, Inc., USA

Krishna Prasad, Mandara Shetty, British American Tobacco, Southampton, UK

**Lead Author*

ACKNOWLEDGEMENTS

This document was initiated by the Cooperation Centre for Scientific Research Relative to Tobacco (CORESTA) Consumer-Reported Outcome Measure (CROM) Task Force Descriptive CROM working group in 2021. Lai Wei, Krishna Prasad, and Mohamadi Sarkar are co-leaders of the working group with support from team members including Emilie Clerc, Nicholas Goldenson, Mandara Shetty, and Mimi Kim. The Descriptive CROM working group also includes an advisory board consisting of members with various backgrounds in tobacco research, including Pierpaolo Magnani, Esther Afolalu, Xavier Cahours, Lesley Giles, and Ryan Black.

Lai Wei, Emilie Clerc, and Stacey McCaffrey are the primary authors and have laid the substantive groundwork for the project. Nicholas Goldenson and Mimi Kim have also contributed to the writing of this document. All authors, working group members, and advisory board members conducted a throughout review of the recommendations in this document. Saul Shiffman and Sandra Sulsky are external subject matter expert (SME) reviewers, who reviewed this document in the latter stages of guideline development. During the final stage of development, we invited additional reviewers, who provided valuable feedback to further refine the recommendations discussed in this guideline. The reviewers include Sade Johns, Deena Battista, Edward Largo, Raheema Muhammad-Kah, Jonathan Gallegos, and Sarah Baxter. Last, but not least, Jill Serrano provided editorial support with timely technical editing and writing assistance to ensure the quality of the guideline.

We would like to acknowledge with gratitude the authors, working group, and advisory board members, SME reviewers and editorial support for their contributions.

Table of Contents

LIST OF ABBREVIATIONS

Abbreviation	Definition
3MC	Multinational, Multiregional, and Multicultural
ABOUT	Assessment of Behavioral Outcomes related to Tobacco and nicotine products
ACASI	Audio computer-assisted self-interview
APA	American Psychological Association
BRFSS	Behavior Risk Factor Surveillance System
CASI	Computer-assisted self-interview
CAWI	Computer-assisted web-interview
CDC	Centers for Disease Control and Prevention
CORESTA	Cooperation Centre for Scientific Research Relative to Tobacco
CROM	Consumer reported outcome measure(s)
CTP	Center for Tobacco Products
CV	Coefficient variation
DEBRA	Deutsche Befragung zum Rauchverhalten
EBS	Eurobarometer Survey
EHIS	European Health Interview Survey
ENDS	Electronic nicotine delivery systems
ENNDS	Electronic non-nicotine delivery systems
ERIC	European Research Infrastructure Consortium
ESCS	Economic, social, and cultural status
ESS	European Social Science
EU LGBT	European Union Lesbian Gay Bisexual Transgender
FCTC	Framework Convention on Tobacco Control
FDA	US Food and Drug Administration
GATS	Global Adult Tobacco Survey
GHPSS	Global Health Professions Student Survey
GLAAD	Gay and Lesbian Alliance Against Defamation
GSK	GlaxoSmithKline
GSPS	Global School Personnel Survey
GTSS	Global Tobacco Surveillance System
GYTS	Global Youth Tobacco Survey
HET	Health on Equal Terms
HnB	Heat-not-burn
HTP	Heated tobacco product
ICC	Intraclass correlation coefficient
ISCO-08	International Standard Classification of Occupations
ITC	International Tobacco Control

MOP	Modern oral product
MPOWER	Monitor tobacco use and prevention policies
MRTPA	Modified Risk Tobacco Product Application
MS	Member states
MTSS	Motivation to Stop Scale
NCI	National Cancer Institute
NHIS	National Health Interview Survey
NHNS	National Health and Nutrition Survey
NIH	National Institutes of Health
NRT	Nicotine replacement therapy
NSDUH	National Survey on Drug Use and Health
NTN	Non-tobacco nicotine
ONP	Oral nicotine product
PAPI	Paper and pencil interview
PATH	Population Assessment of Tobacco and Health
PhenX	Consensus measures for Phenotypes and eXposures
PMTA	Premarket Tobacco Product Application
RSE	Relative standard error
SAMHSA	Substance Abuse and Mental Health Services Administration
SAP	Statistical analysis plan
SE	Substantial equivalence
SES	Socioeconomic status
SHP	Swiss Household Panel
SHS	Swiss Health Survey
SME	Subject matter expert
SOGI	Sexual Orientation and Gender Identity
STS	Smoking ToolKit Study
TF	Task force
THS	Tobacco heating systems
TNP	Tobacco and Nicotine-Containing Product (TNP refers to tobacco products, as well as nicotine-containing products that do not contain tobacco.)
TPPI	Tobacco product perception and intention
TRR	Tobacco Regulatory Toolkit
TUS-CPS	Tobacco Use Supplement – Current Population Survey
UK	United Kingdom
US	United States
WG	Working group
WHO	World Health Organization

DEFINITION OF TERMS

Term	Definition
Construct	A measurable, complex idea or concept formed from a synthesis of simpler ideas.
Consumer Reported Outcome Measure (CROM)	A measurement instrument where data are collected by self-report from the subject of research.
Descriptive CROM	CROM intended to measure observable characteristics and behaviors.
Domain	A related set of behaviors, concepts, or constructs.
TNP Use States	**Never Use:** Having never used the TNP, even once. **Ever Use:** Having ever tried or used the TNP, even once. **Current Use:** Having used the TNP in the past 30 days OR having reported using the product 'every day' or 'some days' now. **Former Use:** Having ever used the product but having not used the TNP in the past 30 days OR having reported using the product 'not at all' now. **Dual Use:** Concurrently using two TNPs from different TNP categories or subcategories. **Poly Use:** Concurrently using three or more TNPs from different TNP categories or subcategories. **Lifetime Established Use:** Having reached the lifetime criterion (summarized in Table 3) for a TNP category. *Note:* Lifetime established use criteria can be applied to ever, current, former, and dual/poly users to further characterize user categories.
Transition Behavior	A transition between TNP Use States, as defined above. **Initiation:** First use of a given TNP. **Cessation:** Stops using the TNP after having used the product to its lifetime established use criterion[1]. **Quit attempt**: Stops using a TNP for longer than 1 day during a specified time frame (e.g., past 12 months) because they were trying to quit using the product. **Relapse/Re-initiation:** Starts using a TNP again after a period of abstinence. **Switching or Transition:** Change of use state in terms of the TNPs being used between two time points.
Psychometric CROM	CROM intended to measure underlying individual psychological attributes/unobservable latent constructs.
Psychometric Property	Validity and reliability of the measurement tool. For detailed definitions of **Validity** and **Reliability**, please refer to Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with Respect to Psychometric CROM for Use in Research on Tobacco and Nicotine Containing Products [1].

[1] When the lifetime established use criterion has not been reached, we refer to this state as an experimentation state.

1. INTRODUCTION

CORESTA is an organization developed with the purpose of promoting international cooperation in scientific research relative to tobacco and its derived products. Its vision is "to be recognized by our members and relevant external bodies as an authoritative source of publicly available, credible science, and best practices related to tobacco and its derived products."
(https://www.coresta.org/who-we-are-29290.html)

In 2018, CORESTA approved the formation of a new TF to establish best practices and guidelines for the development and use of CROM[2] in research on TNPs[3]. This TF defined a CROM as a measurement instrument where data are collected by self-report from the research subject[4].

The CROM TF consists of members from seven contributing manufacturers, and its primary objectives are 1) to provide guidance on the development, modification, and application of CROM, and 2) to facilitate the identification of and access to recommended CROM. The CORESTA Scientific Commission provides oversight of the consortium to ensure conformity of the work with CORESTA standards. The best practices and guidelines developed by the CROM TF focus on CROM for adult consumers who are above the legal age to purchase TNPs. A consortium approach, with contributions from manufacturers and industry partners, has been taken to develop a scientific framework based on the following shared vision:

- To work together to create a paradigm shift in the way CROM are conceptualized and implemented in research on TNPs,
- To work with SMEs to establish guidance for developing and validating new measures,
- To establish consensus on existing survey measures and research methods,
- To use a core set of concepts and tools to facilitate sharing, comparing, and replicating findings, and integrating data from multiple sources.

The CROM TF distinguishes between ***Psychometric* CROM**, which are intended to measure underlying (unobservable) attributes of an individual, and ***Descriptive* CROM**, which are intended to measure observable characteristics and behaviors. To achieve its primary objective, the CROM TF created several WGs (see **Figure 1** for an overview of the purpose of each WG). Two separate best practices and guidelines were developed by the WG, as listed below.

A. "Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with Respect to Psychometric CROM for Use in Research on TNPs" [1]

B. "Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with Respect to Descriptive CROM for Research on TNPs"

This document represents the final deliverable of the WG focused on Descriptive CROM (i.e., guidelines articulating best practices for the selection, development and validation, modification, and implementation of Descriptive CROM for use in research on TNPs).

[2] Within this document, "CROM" can refer to "measure" (singular) or "measures" (plural), which can be inferred through context.

[3] Within these guidelines, "TNPs" refer to tobacco products, as well as nicotine-containing products that do not contain tobacco.

[4] Although not common practice in the field of TNP research, in theory, a CROM could also be completed by someone other than the subject of research. For example, a parent could be asked about their child's use of tobacco products.

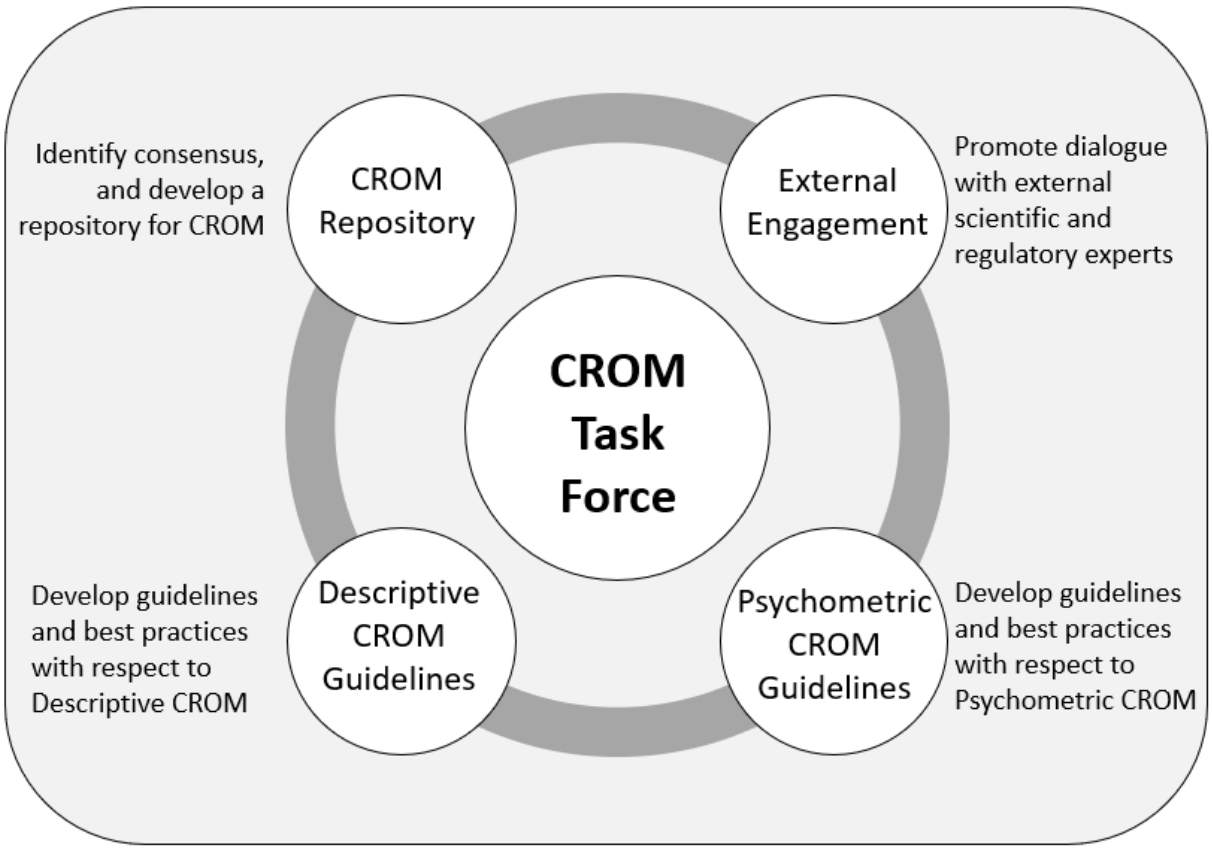


Figure 1 - Governance Structure of CROM Task Force

OVERVIEW OF DESCRIPTIVE CROM BEST PRACTICES AND GUIDELINES

Descriptive CROM best practices and guidelines include the following six main chapters:

Chapter 2 discusses the rationale for developing Descriptive CROM best practices and guidelines. It also presents definitions of descriptive CROM and related terms and discusses the selection of Descriptive CROM based on research objectives.

Chapter 3 presents the Descriptive CROM Conceptual Domain Framework and describes the CROM TF Descriptive CROM WG review process and consensus approach for the selection of Descriptive CROM.

Chapter 4 provides foundational definitions, including TNP use states (e.g., established use, exclusive, ***dual use***, or ***poly use***) and classification of TNPs, to facilitate survey instrument development and secondary analysis of survey data.

Chapter 5 summarizes Descriptive CROM recommendations based on a review of existing Descriptive CROM from national/international surveys. Recommendations are provided for various domains, including demographics and TNP use prevalence or consumption.

Chapter 6 provides recommendations for the development, modification, and adaptation of Descriptive CROM. We recommend modifying an existing Descriptive CROM whenever possible before developing a new Descriptive CROM. We present a multi-stage process for CROM development, including item generation, qualitative and quantitative assessments, etc.

Chapter 7 outlines best practices for study design and development, data analysis, and reporting of Descriptive CROM data.

The following are important points concerning this guideline document:

- These guidelines and the CROM TF do not pretend to represent authoritatively the views of regulatory bodies and any guidance they may publish. This document is intended to serve as a guide for those conducting TNP research using or considering the use of Descriptive CROM.
- This document describes the current thinking of this CROM working group and should be viewed only as recommendations. The use of the word "should" simply means that something is suggested or recommended. The recommendations in this document are grounded in scientific rationale and what may currently be considered best practices regarding the use of Descriptive CROM in research on TNPs. However, best practices may also evolve over time with advances in research on TNPs.
- The guidelines are not intended to reflect unattainable standards; researchers in the field of TNPs should be knowledgeable about these guidelines, and then make an informed decision as to what extent they are applicable or necessary for a particular study. The researcher is ultimately responsible for defending their research.
- The intended audience of these guidelines are individuals who not only have appropriate knowledge of behaviors associated with the use of TNPs, but who also have basic familiarity or experience with how CROM can be used as endpoints in research studies. Additionally, recommended reading/references are incorporated throughout the guidelines for readers interested in learning more.
- The recommendations presented in this document do not reflect the views of individual companies whose members are part of the CROM TF and are not intended to have binding implications on past, current, or future research conducted by individual companies or research that may be conducted in support of TNP regulatory applications or scientific publications.
- Based on consensus within the research community [2-4], throughout the guidelines, we adopt the use of person-first language (e.g., "people who smoke") rather than commonly used labels (e.g., "smokers") to promote greater respect and convey dignity for people who use TNPs. It has been suggested that the use of precise and bias-free language to describe people who use TNPs has the potential to reduce tobacco-related stigma and may enhance the precision of scientific communication [4]. Words or phrases found in the Definition of Terms appear in ***bold italics*** at first mention.

2. PROPOSED BEST PRACTICES AND GUIDELINES FOR DESCRIPTIVE CROM IN RESEARCH ON TNPs

CROM are data collected by self-report from the subject of research pertaining to perceived states, behavior, and/or understanding of messages. Alignment in the development and use of descriptive CROM is necessary to evaluate the prevalence and use patterns of TNPs for pre- and post-market research and regulatory compliance. As more research on TNPs incorporates established CROM, a reliable basis for monitoring and evaluating changes in population use patterns over time will emerge.

2.1 Need for Best Practices and Guidelines for Descriptive CROM

Research on TNPs is a continuously evolving field. The wide variety and diversity of available TNPs and the continued development of new products require a constant evolution of measures to adequately assess TNP use and related ***constructs*** [5, 6]. Many of the currently available measures are based on or adapted from questionnaires/items initially developed for people who smoke conventional cigarettes [7]. New CROM are needed to build the foundation for standards of measurement in research on TNPs in the changing tobacco product landscape. For example, the use of ENDS has increased in the US and internationally over time. ENDS are a diverse product category that varies in how they produce aerosol as well as the levels of nicotine and flavorings contained within the liquids [8]. A lack of common definitions of patterns of use and types of users, in addition to high variability in the ENDS products themselves, limits consistency in measurement and therefore, comparisons among research studies [8, 9]. Similarly, the composition of other non-cigarette TNPs varies within category and may require unique definitions of use and user states.

Some measurement and standardization initiatives have been proposed to assess consumer perception and behavior associated with TNPs [6, 7, 10], which will allow for comparison of results across similar studies in research on TNPs. For example, the PhenX Tobacco Regulatory Research Toolkit, funded by the US NIH and the US FDA CTP, was developed to expand the breadth and depth of tobacco product-related measures in order to enhance cross-study analysis in large-scale research [6]. The ABOUT Toolbox was developed to allow for comparisons of consumer perceptions and behaviors across various TNPs and across research studies, to facilitate informed decision-making regarding regulation of TNPs, and to improve surveillance associated with the impact of TNPs on public health [7]. In 2022, the US FDA published final guidance for designing and conducting TPPI studies that may be submitted as part of a MRTPA, a PMTA, or a SE report [11]. Despite these ongoing efforts, as TNPs continue to evolve, consensus on survey measures in research on TNPs is especially challenging, particularly for emerging TNP categories (e.g., measurement of established use behavior or product consumption).

2.2 Descriptive CROM and Related Definitions

Descriptive CROM are self-reported survey outcome measures that are intended to measure observable characteristics and behaviors. Some examples of Descriptive CROM include demographic variables and product use behaviors. In comparison, ***Psychometric*** **CROM** are intended to measure underlying individual psychological attributes/unobservable latent constructs. Examples of Psychometric CROM commonly used in TNP regulatory research include but are not limited to the following: product risk perception and behavioral intentions, reactions to product use, and believability. Psychometric CROM best practices and guidelines

for use in tobacco regulatory research are discussed in a separate document (i.e., Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with Respect to Psychometric CROM for Use in Research on TNPs [1]) as mentioned in the Introduction.

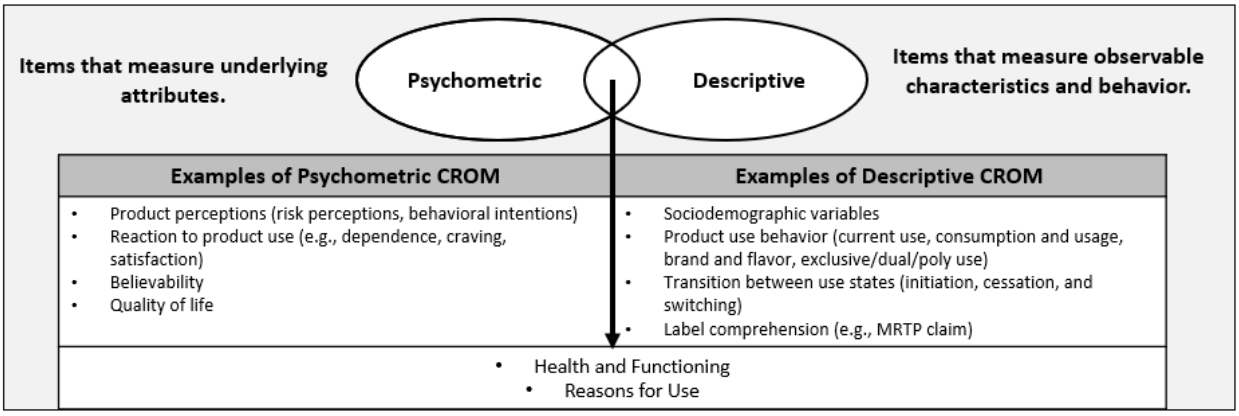


Figure 2 - Psychometric and Descriptive Constructs

A guideline of best practices in the selection, development, implementation, and analysis of Descriptive CROM in research on TNPs will serve to 1) provide recommendations to reduce sources of measurement error applicable to Descriptive CROM, such as comprehension, recall bias, and/or social desirability bias, 2) facilitate consensus building around tools to assess use of novel TNPs that may be introduced in the future, and 3) promote data comparability and cross-study analyses.

2.3 Selection of CROM based on Research Objective and Target Population

The selection of CROM should be based on the target study population, study objectives, and research hypothesis, and/or driven by regulatory requirements. Chapter 2 of the Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with Respect to Psychometric CROM for Use in Research on TNPs describes the framework for identifying optimal characteristics of a Psychometric CROM within the context of a particular study [1]. Such a framework could be applied to a study on Descriptive CROM, which helps determine whether (1) an existing CROM may be appropriate (with or without any additional testing) for the study, (2) an existing CROM might be modified to meet the study's needs, or (3) a new CROM needs to be developed to meet the study's needs.

3. METHODOLOGY

3.1 Descriptive CROM Conceptual Domain Framework

The Descriptive CROM conceptual domain framework was based on a comprehensive review of the extant literature and existing TNP surveillance surveys conducted at a national and international level and is used to guide the development of Descriptive CROM recommendations. The framework includes three main ***domains***: 1) population-level, 2) product category-level, and 3) poly-/cross-category-level (Figure 3). The population-level domain includes survey measures that are typically posed to all survey respondents to evaluate demographics and SES (e.g., income, education, etc.). The product category-level domain includes survey measures that evaluate TNP consumption, brand and flavor preferences, initiation and cessation of use, and reasons for use for each TNP category among users of a particular TNP. Lastly, the poly-/cross-category-level domain includes survey measures that evaluate dual/poly use and ***switching*** between TNP categories.

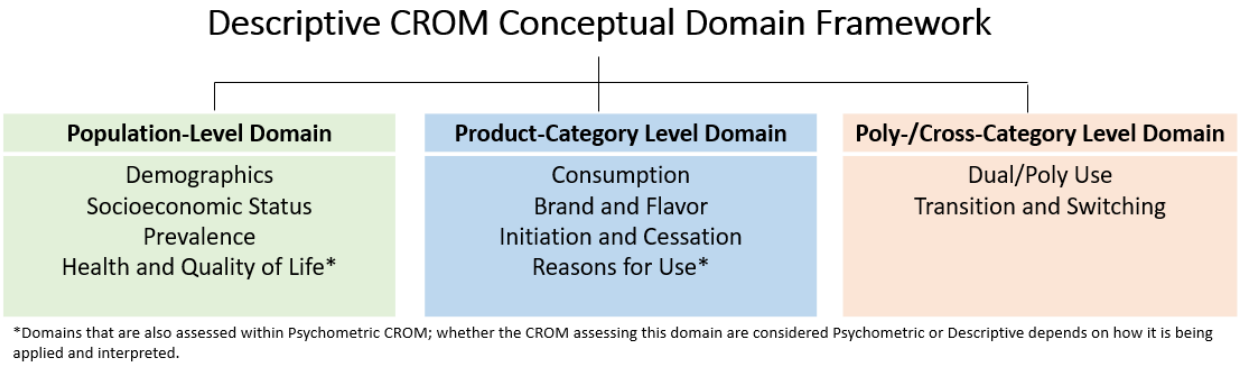


Figure 3 - Descriptive CROM Conceptual Domain Framework

3.2 Review of Existing Descriptive CROM based on National/International Surveys

CROM TF Descriptive CROM WG conducted a comprehensive review of existing Descriptive CROM from surveys that include modules to assess the use of TNPs. We selected fifteen surveys conducted at a national or international level and cover a wide range of Descriptive CROM for adults who use TNPs. An overall summary of the fifteen surveys is shown in **Table 1** (see also Appendix Table 1 for a summary of survey methodology).

CROM TF Descriptive CROM WG selected the surveys based on their representativeness, diversity in domains, geographic coverage, and accessibility of questionnaire data. All selected surveys are currently ongoing, with a regular data collection schedule in place. In general, we focused our review on the most recent survey questionnaires being administered among adult populations. Most of the selected surveys were designed to monitor TNP use trends and participant health status at a national level, and hence, were conducted among nationally representative samples of the respective adult populations. Only publicly available surveys were included, and some surveys (e.g., the BRFSS survey) were not included due to overlap in Descriptive CROM domains with other national surveys (e.g., NHIS) resulting from existing consensus on survey measures between national surveys. Survey items from the most recent survey questionnaires were discussed and grouped into domains based on the CROM Conceptual Domain Framework, including key demographics, prevalence, TNP usage, consumption, initiation, and cessation, etc.

3.3 Consensus Approach for Selection of Descriptive CROM and Related Definitions

CROM TF Descriptive CROM WG first identified two areas to facilitate the selection and use of Descriptive CROM: TNP classifications and ***TNP use state*** definitions. As survey measures typically evaluate individual categories of TNPs, clear and consistent classification of TNPs would facilitate accurate assessment of product use behavior and would allow for comparisons of results across surveys. Additionally, the definitions of TNP use states would affect the development of survey conditional branching (i.e., skip logic). Consistency of definitions is essential for data analysis and reporting to improve harmonization in research findings and to make research findings comparable.

CROM TF Descriptive CROM WG then utilized a consensus-based approach to recommend Descriptive CROM and related definitions based on a review of selected surveys and existing literature. The best practices and guidelines on Descriptive CROM were proposed by core team members and reviewed by advisory board members. CROM TF Descriptive CROM WG also collaborated with other CORESTA subgroups, including the CORESTA Product Use Behavior subgroup, the In Vitro Toxicity Testing subgroup, and the Tobacco and Tobacco Products Analysis subgroup, to align product category-specific measures and definitions. Additional SMEs were invited to provide written feedback or to participate in discussions via virtual meetings. Furthermore, best practices and guidelines are posted on an open-access platform to allow for additional input from the broader scientific TNPs research community. Lastly, best practices and guidelines may be updated as the TNP landscape evolves to address CROM for emerging TNP categories.

Table 1 - Summary of Survey Information

Abbreviation	Full Name	Objective	Funding Agency
International Surveys			
GATS	Global Adult Tobacco Survey	"To enhance country capacity to design, implement, and evaluate tobacco control interventions, and monitor key articles of the WHO FCTC and components of the WHO MPOWER technical package". (Link) GTSS includes the collection of data through four surveys: GYTS, GSPS, GHPSS, and the GATS.	WHO, CDC
ITC Survey	International Tobacco Control Survey	"The ITC Project (link) has established a research platform to guide strong, evidence-based implementation of FCTC policies to: - Evaluate FCTC policies at the level of the individual smoker - Identify the determinants of effective tobacco control policies - Disseminate research findings to the global tobacco control community, including researchers, policy makers, and advocates."	30+ agencies around the world
Surveys in European Countries			
DEBRA	Deutsche Befragung zum Rauchverhalten	"Representative survey on the use of tobacco and alternative nicotine delivery systems in the German population. Baseline questions cover smoking status and ever-use of e-cigarettes. Depending on the response behavior, current tobacco smokers (cigarettes or other tobacco products), recent ex-smokers (<12 months since quitting tobacco), and ever-users of e-cigarettes or a similar product (e.g., e-hookah, e-cigar, or e-pipe) will answer on further detailed questions about smoking behavior, quit attempts, exposure to health professionals' advice on quitting, and use of cessation aids". (Link) The methodology is closely aligned to the STS, which will allow comparisons with data from England."	German Ministry of Health, the Ministry for Innovation, Science and Research of the German Federal State of North Rhine-Westphalia.

Abbreviation	Full Name	Objective	Funding Agency
EBS	Eurobarometer Survey	"To monitor the public opinion of the EU member and candidate countries. The standard modules ask for attitudes towards European unification, institutions, and policies, complemented by measurements for general socio-political orientations, as well as by respondent and household demographics. Intermittently, Eurobarometer extensively addresses special topics, such as environment, technology, health (e.g., tobacco use behavior)". (Link)	European Commission, the European Parliament
EHIS	European Health Interview Survey	"To measure on a harmonized basis and with a high degree of comparability among MS the health status (including disability), health determinants (lifestyle) of the EU citizens and use of health care services and limitations in accessing it". (Link)	European Commission
ESS	European Social Survey	"To measure the attitudes, beliefs, and behavior patterns of diverse populations in more than thirty nations. The main aims of the ESS are to chart stability and change in social structure, conditions, and attitudes in Europe and to interpret how Europe's social, political, and moral fabric is changing; to achieve and spread higher standards of rigor in cross-national research in the social sciences". (Link)	European Research Infrastructure Consortium (ERIC)
HET	Health on Equal Terms - Sweden	"To investigate the health of the population and to show changes in the population's health over time. The questions in the national public health survey cover physical and mental health, consumption of pharmaceuticals, contact with healthcare services, dental health, living habits, financial conditions, work and occupation, work environment, safety, and social relationships". (Link)	Public Health Agency of Sweden
SHP	Swiss Household Panel	"To observe social change, in particular the dynamics of changing living conditions and representations in the population of Switzerland. The survey covers a broad range of topics and approaches in the social sciences". (Link)	Swiss National Science Foundation

Abbreviation	Full Name	Objective	Funding Agency
SHS	Swiss Health Survey 2017: Tobacco Consumption	"To monitor Swiss population health status, to identify principal epidemiological trends in Switzerland and to assess prevention projects and health promotion programs effectiveness. Information collected on population general health state, diseases, resources and competencies in the health domain, situation in the health insurance domain, lifestyle and life conditions which may have an influence on health". (Link)	Swiss Confederation
STS	Smoking ToolKit Study	"To provide monthly nationally representative data on key indicators of smoking behavior, cessation, and tobacco control initiatives. Key assessments (relevant to Descriptive CROM) are: smoking status (daily; non-daily; quit within the last year; quit more than a year ago; never smoked for a year or more; use of noncigarette tobacco), amount smoked and nicotine intake (cigarettes or other tobacco products used per day, week, or month), harm reduction prevalence of attempts to cut down but not quit, use of nicotine replacement therapy when cutting down and/or prohibited from smoking, and demographics". (Link)	Cancer Research UK, Pfizer, and GSK
Survey in Asian Countries			
Japan - NHNS	National Health and Nutrition Survey	"To understand the status of people's health, nutritional intake, and lifestyle habits and to obtain basic data necessary for comprehensive health promotion". (Link)	Ministry of Health, Labour and Welfare
Surveys in the US			
NHIS	National Health Interview Survey	"To monitor the health of the US population through the collection and analysis of data on a broad range of health topics". (Link)	CDC
NSDUH	National Survey on Drug Use and Health	"To provide accurate data on the level and patterns of alcohol, tobacco, and illegal substance use and abuse, track trends in the use of alcohol, tobacco, and various types of drugs, assess the consequences of substance use and abuse, and identify those groups at high risk for substance abuse". (Link)	SAMHSA

Abbreviation	Full Name	Objective	Funding Agency
PATH	Population Assessment of Tobacco and Health	"To monitor and assess behaviors, attitudes, biomarkers, and health outcomes associated with tobacco use in the United States." (Link)	FDA, NIH
TUS-CPS	Tobacco Use Supplement - Current Population Survey	"To serve as a key source of national, state, and sub-state data on tobacco use behaviors, attitudes, and policies in the United States." (Link)	FDA, NCI

4. FOUNDATIONAL DEFINITIONS

Clear classification of TNP categories and TNP use states is essential to evaluate consumer-reported outcomes. Here, we introduce TNP classifications with detailed descriptions of each category and TNP use state definitions to facilitate survey instrument development and secondary analysis of survey data.

4.1 Tobacco and Nicotine-Containing Product Classification

Over the past several decades, a wide range of new TNPs have emerged in the global market. This product diversity poses a challenge to researchers and regulators in the TNPs space due to the complexity in defining and differentiating among various primary use states and differences in each product's contents. Based on recommended definitions from the WHO, US FDA, and CORESTA Tobacco and Tobacco Products Analysis and Product Use Behavior CORESTA subgroups, we developed a classification system for TNPs that separates products into two main categories: combustible and non-combustible (**Table 2**). This classification aims to provide a clear overview of existing TNPs on the market and to support the development of CROM that assess TNP use behavior by categorizing and describing each product. A brief description of the product categories with example product images in questionnaires will enhance the clarity of the surveys. When addressing combustible tobacco products, it may be necessary to include descriptions that provide the respondent with further clarification when there is the possibility that some products may be confused with others. For example, when addressing hookah use, the 2019 NHIS included the following instructions in the product description for further clarification: 'Do not include electronic hookah or e-hookahs' when assessing hookah use. In the 2018-19 TUS-CPS survey, the pipe tobacco category description included the following description: 'It does not include smoking hashish, marijuana, crack, or other substances in a pipe. Do not include water pipes/hookahs.'.

Table 2 - Classification of TNPs

Category		Subcategory	Category/Subcategory Description
Combustible Products	**Cigarette**	• Manufactured Cigarette • Roll-Your-Own Cigarette	A cigarette is a tube-shaped tobacco product that is made of finely cut, cured tobacco leaves wrapped in thin paper. A cigarette is lit on one end, and the smoke is inhaled. Roll-your-own cigarettes are made of loose tobacco that is placed inside rolling paper. As with manufactured cigarettes, one end is lit, and the smoke is inhaled. (Source: Cigarettes | NCI (content as of Apr 11, 2022), Cigarettes | FDA (content as of Apr 29, 2021, accessed Dec 27, 2021), and Roll-Your-Own Tobacco | FDA (content as of Dec 21, 2019, accessed Dec 27, 2021))
	Cigar/ Cigarillo	• Traditional Cigar • Cigarillo • Little Filtered Cigar	A cigar is a roll of tobacco wrapped in leaf tobacco or in a substance that contains tobacco. They vary in size—from smaller cigars, such as little filtered cigars or cigarillos, to larger ones, such as large so-called premium cigars. The cigar is lit on one end and smoked, but the smoke is usually not inhaled into the lungs. (Source: Cigars, Cigarillos, Little Filtered Cigars | FDA (content as of Jun 11, 2021, assessed Dec 27, 2021), Cigars | NCI (content as of October 23, 2023))
	Pipe		Pipe tobacco is generally loose-leaf tobacco burned in a traditional smoking pipe with a bowl. A pipe is a device with a mouthpiece at one end of a tube, and a small bowl at the other end that is filled with tobacco, which is lit and smoked. The smoke from a pipe is usually not inhaled into the lungs. (Source: Pipe Tobacco | FDA (content as of Oct 06, 2020, accessed Dec 27, 2021), Pipe (NCI) (accessed Dec 27, 2021))
	Hookah (Shisha or Waterpipe Tobacco)		Hookah tobacco (also known as waterpipe tobacco, maassel, shisha, narghile, or argileh) is smoked with a hookah (waterpipe). A form of moist tobacco is placed in the head of the hookah with charcoal placed on top (often separated by perforated aluminum foil) to provide a heat source. The heated air, passing over the charcoal, contains charcoal combustion products, passes through the tobacco, and the mainstream smoke aerosol is produced. The smoke then passes through the waterpipe body, bubbles through the water in the bowl, and is carried through the hose and inhaled or puffed by users via a mouthpiece. (Source: Hookah Tobacco (Shisha or Waterpipe Tobacco | FDA (content as of Jan 03, 2020, accessed Dec 27, 2021), Water pipe | NCI (accessed Dec 27, 2021), Sutfin, McKelvey [12], Waterpipe Tobacco Smoking | WHO (accessed Jun 07, 2022))

Category		Subcategory	Category/Subcategory Description
Non-Combustible Products	**Electronic Nicotine Delivery Systems (ENDS)**	• E-cigarette • E-cigar • E-pipe • E-hookah	ENDS are battery-powered devices that are designed to electrically heat a liquid (may also be called an e-liquid), to produce an inhalable aerosol. The most common ENDS are 'electronic cigarettes', also known as 'e-cigarettes'. There are currently four major types of ENDS products: disposable ENDS products, ENDS products with replaceable pre-filled cartridges or pods, tank systems that can be filled with liquids, and modular systems that can be filled with liquids. Several terms and acronyms are used to describe this product category, including e-vapor, vapes, vaporizers, vape pens, etc. Other subcategories of ENDS could include e-cigar, e-pipe and e-hookah. Some ENDS products are manufactured with non-tobacco nicotine (i.e., synthetic nicotine)[5]. Additionally, the WHO refers to electronic non-nicotine delivery systems as ENNDS[6]. (Source: Vaporizers, E-Cigarettes, and other Electronic Nicotine Delivery Systems (ENDS) | FDA (content as of Sep 17, 2020, accessed Dec 27, 2021), World Health Organization [13] and CORESTA [14])
	Heated Tobacco Products (HTPs)		HTPs contain a tobacco substrate that is designed to be heated and not combusted by a separate source (e.g. electrical, aerosol, carbon, etc.) to produce a nicotine-containing aerosol. Regulatory agencies, researchers, and manufacturers use a variety of terms and acronyms to describe this product category, such as tobacco heating systems (THS), heat-not-burn tobacco products (HnB), etc. (Source: CORESTA Product Use Behavior Subgroup Heated Tobacco Products (HTPs): Standardized Terminology and Recommendations for the Generation and Collection of Emissions (content as of Oct 24, 2023, accessed Oct 24, 2023), Heated Tobacco Products | CDC (content last reviewed, 2020 Dec 16, accessed Dec 27, 2021))

[5] FDA has begun to regulate tobacco products containing nicotine from any source including tobacco products containing NTN, that is, nicotine not made or derived from tobacco, such as synthetic nicotine (Effective Date: April 14, 2022).

[6] ENNDS: The WHO report 'Why electronic non-nicotine delivery systems ENNDS are included in this report" explains ENNDS are included because they are almost indistinguishable from ENDS. They often have enhanced flavors that appeal to young people and may be perceived as being a safer, less addictive option. Although ENNDS are marketed to not contain nicotine, many e-liquids have been found to contain nicotine when tested. Furthermore, depending on the device used, users may be able to select e-liquids that contain nicotine or not. [13]World Health Organization, *WHO report on the global tobacco epidemic 2021: addressing new and emerging products*. 2021.

Category		Subcategory	Category/Subcategory Description
	Smokeless Tobacco Products	• Chewing Tobacco	Chewing tobacco is cured tobacco in the form of loose leaf, plug, or twist. The product is chewed during use and subsequently discarded. Loose-leaf chewing tobacco typically consists of loosely packed, cut, or granulated stem-free tobacco leaf to which additional ingredients may be added. Plug chewing tobacco typically contains flaked tobacco leaves to which additional ingredients may be added. The product has the appearance of a compressed tobacco brick wrapped inside a natural tobacco leaf. Twist chewing tobacco has the appearance of thick rope-like twists of tobacco. (Source: Smokeless Tobacco Products, Including Dip, Snuff, Snus, and Chewing Tobacco | FDA (content as of Jun 23, 2020, accessed Dec 27, 2021), CORESTA Tobacco and Tobacco Products Analysis Sub-group (i.e., Smokeless Tobacco Sub-group)
		• Moist snuff/Dip	Moist snuff/Dip is cut tobacco that can be loose or pre-portioned (i.e., pouched), placed in the mouth, and discarded after use. Moist snuff/Dip is finely ground tobacco packaged in cans or pouches. It may have flavorings added. Moist snuff is commonly placed between the cheek and gum during use and discarded after use. (Source: Smokeless Tobacco Products, Including Dip, Snuff, Snus, and Chewing Tobacco | FDA (content as of Jun 23, 2020, accessed Dec 27, 2021), CORESTA Tobacco and Tobacco Products Analysis Sub-group (i.e., Smokeless Tobacco Sub-group)
		• Dry Snuff	Dry snuff is loose, finely cut, or powdered dry tobacco that is typically sniffed through the nostrils. (Source: Smokeless Tobacco Products, Including Dip, Snuff, Snus, and Chewing Tobacco | FDA (content as of Jun 23, 2020, accessed Dec 27, 2021), CORESTA Tobacco and Tobacco Products Analysis Sub-group (i.e., Smokeless Tobacco Sub-group)
		• Snus	Snus is cut tobacco that is processed into fine particles. The products are usually placed between the upper lip and gum and are discarded after use. Products are available as loose tobacco or as individually portioned pouches. (Source: CORESTA Tobacco and Tobacco Products Analysis Sub-group (formally known as Smokeless Tobacco Sub-group)

Category		Subcategory	Category/Subcategory Description
		• Dissolvable Tobacco Products	Dissolvable tobacco products are finely ground tobacco pressed into shapes such as tablets, sticks, or strips. Dissolvable tobacco products can be sold as lozenges, orbs, strips, or sticks. Lozenges resemble pellets or tablets, orbs resemble small mints, sticks have a toothpick-like appearance, and strips are thin sheets that work like dissolvable breath strips or medication strips. Dissolvable tobacco products are placed in the mouth and allowed to dissolve during use. (Source: Dissolvable Tobacco Products | FDA (content as of Jun 14, 2018, accessed Dec 27, 2021), Smokeless Tobacco | CDC (content as of May 14, 2021, accessed Dec 27, 2021), CORESTA Tobacco and Tobacco Products Analysis Sub-group (i.e., Smokeless Tobacco Sub-Group)
	Nicotine-Containing Oral Products	• Nicotine Pouches • Gums • Lozenges	Nicotine-containing oral products contain a base substrate, nicotine, and added flavors, but not tobacco leaf. The nicotine can either be derived from tobacco, or synthetic nicotine. These products are exclusively intended for oral use. The products come in a variety of forms, such as pouches, gums, and lozenges. Regulatory agencies, researchers, and manufacturers use a variety of terms to describe this product category, such as ONPs and MOPs. (Source: Taljout [15], CORESTA Product Use Behavior Sub-Group)

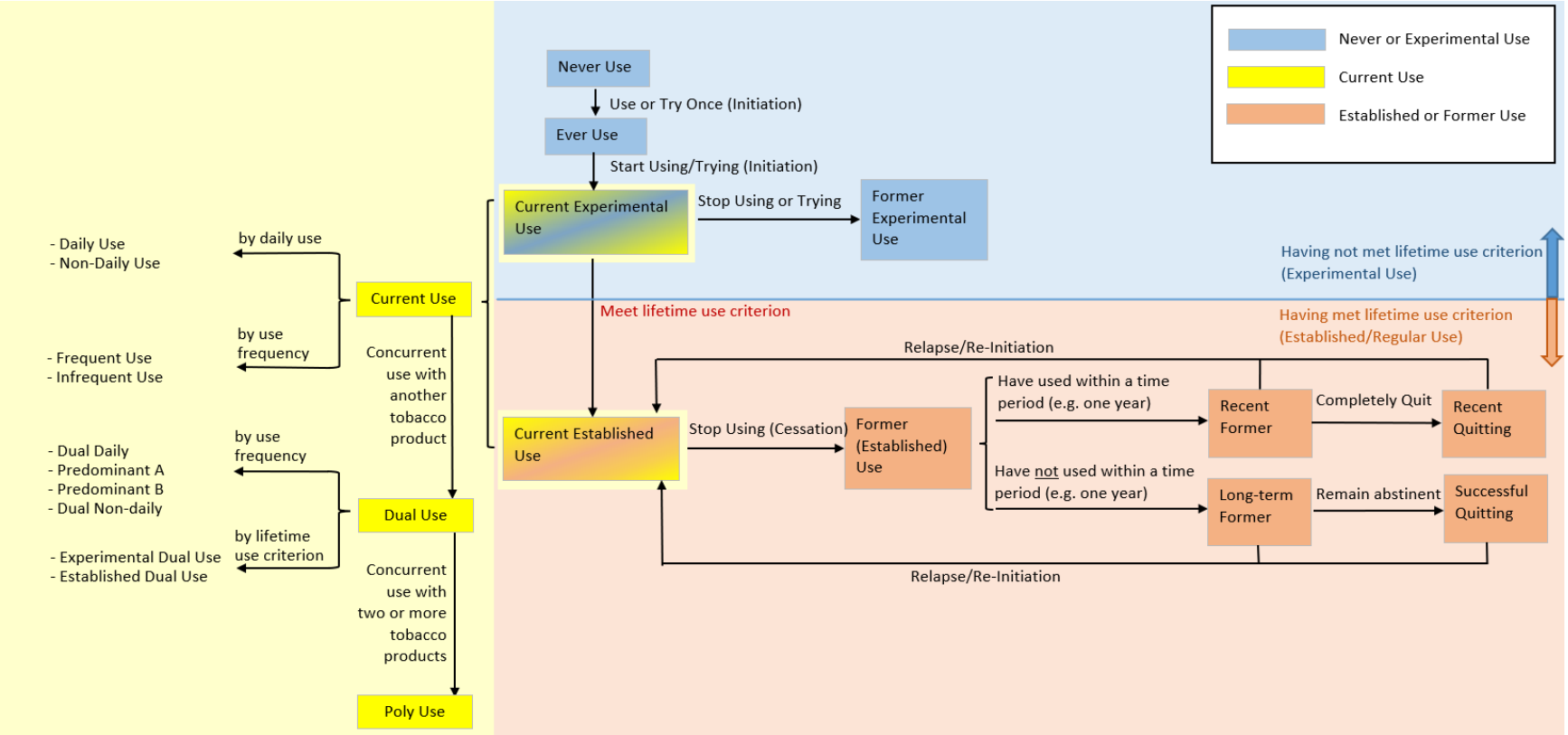
Additionally, NRT is a class of nicotine-delivering pharmaceutical products designed to help people stop using TNPs. There are two classes of NRT products: the transdermal patch and several oral dosing systems, including chewing gum, inhalers, sprays, tablets, and lozenges [16].

The main difference between NRT and TNPs is that NRT includes drug products that are subject to a different set of regulations and regulatory requirements than those governing TNPs. NRT has been developed and is designed and indicated for use to support the cessation of TNPs. NRT is usually evaluated in survey questionnaires with tobacco product usage but is not the subject of this guideline.

4.2 Tobacco and Nicotine-Containing Product Use State

It is often the case that only certain Descriptive CROM are posed to a specific subgroup of the general population, such as ever, current, or former users of a TNP. Therefore, consistent definitions of TNP use states are needed to define study groups or to develop conditional branching (i.e., skip logic) for survey design based on TNP use behaviors. After CROM are collected, consistent definitions of TNP use states are required for data analysis and reporting to allow for comparisons among research findings. While the tobacco landscape has evolved, there is a lack of consistency in definitions of people who use TNPs and TNP use states.

Here, we provide our recommended definitions of TNP use for adults who consume tobacco. We classify survey respondents into various TNP use states based on lifetime, past, and current TNP use. While it may not be possible to list every possible combination of tobacco use states, we address key transitions in a simplified conceptual flow diagram, which only shows changes in use behaviors within a single TNP category (**Figure 4**). Interactions between two product categories (e.g., switching behaviors) are not included in Figure 4 but will be discussed in later sections.



Note: Either a numerical (e.g., having used X times) or non-numerical (e.g., having used fairly regularly) threshold can be used to define established use.

Figure 4 - TNP Use State Conceptual Framework

The conceptual framework includes the following use states: ***Never or Experimental use*** (blue), ***Current use*** (yellow), ***Established use*** (orange), and ***Former use*** (orange). The flow chart begins with a '***Never Use***' state. After trying a TNP once, the individual moves to the '***Ever Use***' state (i.e., becomes an ever user of a TNP), which is often called initiation. TNP initiation generally refers to the first use of a given TNP. When the individual starts using the product, the individual moves to a 'Current Use' state. While 'Current Use' is often defined as the use of the product 'every day' or 'some days', we can also define 'Current Use' based on past 30-day usage, specifically, 'having used the TNP in the past 30 days'. 'Current Use' can be further categorized as 'Current Experimental Use' and 'Current Established Use'. If the individual has not reached the predefined criterion for ***lifetime established use*** (e.g., 100 cigarettes), the individual may be classified into the 'Current Experimental Use' state. If an individual reaches the lifetime established use criterion while using the product, the individual could be classified into the 'Current Established Use' state. 'Current Use' state can be further characterized by daily versus non-daily use based on use being reported every day or some days (or reported being used 30 days out of the past 30 days), or frequent versus infrequent use based on number of days used in the past 30 days (e.g., greater than versus less than 20 days).

We recommend using the suggested lifetime established use criterion (**Table 3**) to distinguish individuals who are experimental users from individuals who are established users, as distinct differences have been found between these two groups of users for various TNP categories [17-20]. While the lifetime established use criterion of 'having smoked 100 or more cigarettes' has been widely adopted in research on TNPs [21-23], lifetime established use criteria for other TNP categories are less definitive [24-28]. Sánchez-Romero et al. [29] demonstrated lower variability in cigarette smoking prevalence by different current use frequency thresholds compared to other TNP categories, which may be due, in part, to the adoption of the '100-cigarette lifetime' criterion in prevalence estimation. Wei et al. [20] conducted analyses of PATH study data to examine the level of agreement between non-numerical (i.e., 'having smoked/used the product fairly regularly') and numerical (i.e., 'having smoked/used the product 20/50/100 times' or use occasions or product units) lifetime established use criteria and proposed a set of criteria based on the level of agreement (kappa coefficients ranging from 0.3 to 0.6) while also taking existing consensus into consideration. It was concluded that, as new TNPs emerge in the market, non-numerical criteria may be a good approach to identify individuals who use TNPs regularly at an early stage [19]. In comparison, the numerical lifetime use criterion provides a more objective characterization of lifetime established use when the new TNP becomes more widely used. The suggested numerical lifetime use thresholds should be evaluated further as the tobacco landscape evolves through qualitative and quantitative studies to ensure the thresholds are appropriate to distinguish between individuals who are experimental users from those who are established users. Current use frequency has also been adopted by other studies to define established use [9, 17, 29, 30]. With the focus of this section being lifetime ever use, which is the basis of the conceptual flow of TNP state, we will discuss the utilization of current use frequency to estimate the prevalence and established use in Section 5.1.2.

Table 3 - Adult Lifetime Established Use Criterion for TNP Categories

Category	Threshold Type	Suggested Criterion for Established Use	References
Cigarette	Numerical	Having smoked **100** cigarettes	[21-23, 26]
Cigar	Numerical	Having smoked **50** cigarillos/ traditional cigars/ filter cigars	[31-34]
	Non-numerical	Having smoked cigarillos/traditional cigars/filter cigars fairly regularly	[32, 34]
Pipe	Numerical	Having smoked **50** bowls filled with pipe tobacco	[20, 26]
	Non-numerical	Having smoked pipe tobacco products fairly regularly	[35]
Hookah	Numerical	Having smoked hookah **20** times[a]	[34]
	Non-numerical	Having smoked hookah products fairly regularly	[20, 35, 36]
ENDS	Numerical	Having used ENDS products **20** times[a]	[34]
	Non-numerical	Having used ENDS products fairly regularly	[9, 20, 37]
HTP	Numerical	Having used **100** or more heatsticks	[38]
	Non-Numerical	Having used HTP fairly regularly	-
Smokeless	Numerical	Having used smokeless tobacco **20** times[a]	[20, 39, 40]
	Non-numerical	Having used smokeless tobacco fairly regularly	[20, 41]
Snus	Numerical	Having used snus **20** times[a]	[20, 34, 42]
	Non-numerical	Having used snus tobacco fairly regularly	[20, 41]
Dissolvable	Numerical	Having used dissolvable TNPs **20** times[a]	[34]
	Non-numerical	Having used dissolvable TNPs fairly regularly	[20, 35]
Nicotine-Containing Oral Products	Numerical	Having used nicotine-containing oral products **20** times[a]	-
	Non-numerical	Having used nicotine-containing oral products fairly regularly	-

[a] 'One time' refers to a typical session when the participant picks up the product to use it. For example, the description in the PATH Wave 5 questionnaire is "the participant picks up the ENDS product to use it. Multiple puffs can be taken within one session."

Additionally, dual and poly use states can be met if an individual is classified as a current user of two or more TNPs from different TNP categories or subcategories. Context should be provided for studies that focus on dual usage, including TNP categories being assessed, the definition of use (e.g., current, past 30-day, past year, or ever use), and inclusion or exclusion of other TNPs. Some existing research has further categorized individuals who are dual users into four segments depending on their use frequency (daily or frequent use) of the two products due to the heterogeneity found within these individuals [43-47]. For example, Borland et al. proposed to classify dual use into dual daily (i.e., daily use of both products) use, predominant product A use (i.e., daily use of product A and non-daily use of product B), predominant product B use (i.e., daily use of product B and non-daily user of product A), and dual non-daily (i.e., non-daily use of both products) use states [44].

Lastly, cessation is defined as stopping the use of the TNP after having used the product to at least its lifetime use criterion. Individuals who are former users need to report having not currently used the product for a predefined timeframe and can be further classified into 'Recent Former' and 'Long-Term Former' states based on when the TNP was last used (e.g., < 1 year versus ≥ 1 year, based on predefined timeframe). If the individual is classified into a 'Recent Former' user state and reports 'having completely quit' the TNP, the product use state is characterized as ‘Recent Quitting’. Individuals initially characterized as ‘Long-Term Former’ are further characterized as ‘Successful Quitting’ if the individual remains abstinent for a predefined period of time (e.g., 1, 2, or 3 years) based on research objectives. Lastly, relapse/re-initiation has been used in the literature to define restarting the TNP after a period of abstinence (e.g., 1 year) and may occur during the former use states. Existing research demonstrates that such ‘relapse/re-initiation’ are less likely after being abstinent for longer than one year [48, 49].

5. RECOMMENDATIONS BASED ON EXISTING DESCRIPTIVE CROM

This chapter provides recommendations based on our review of existing Descriptive CROM from national/international surveys. Items in existing surveys may be referred to, and the source survey is referenced[7].

5.1 Population-Level Domain

5.1.1. Demographics and Socioeconomic Status

Key socio-demographic concepts of interest for research on TNPs that may influence product use patterns. Examples of demographic and socioeconomic variables include age, sex at birth, gender, race, ethnicity, level of education, and income. Other socio-demographic concepts that may be considered according to study objectives and endpoints include occupation, work status, nationality, and/or residency. Tobacco use among minority groups and people of various sexual orientations and/or religion may also be of interest in some studies.

In this section, we provide recommendations for assessing each of these concepts by summarizing existing consensus. The information obtained through these survey items may be considered sensitive information by individuals in some countries, therefore, we suggest alternate methods of collecting the information. In addition to potentially sensitive information, there are considerations to be made for issues around diversity and inclusion and measurement equivalence.

Assessing Age

Age may be evaluated in several ways. In US-based surveys, participants are often asked to provide their date of birth (e.g., MM/DD/YYYY), which should be considered as Protected Health Information [50]. Participants may also be given the option to refuse to answer or to respond with "I don't know". If either of the two latter options are chosen, a participant may be asked to provide age in a numerical format. Confirmation of participant age may also be added after the date of birth is given [51, 52]. A response range (e.g., 1 to 120 for age) has also been used [53]. In European and international surveys, participants may be more likely to be asked their age (EBS) or year of birth [54, 55] because the exact date of birth may be considered sensitive information or may be perceived as highly confidential.

According to the US FDA and CDC, although age ranges for youth and young adults vary across studies, in general, individuals aged 11-17 years are described as "youth" or "adolescent", individuals aged 18-25 years are described as 'young adults' (although, developmentally, the period between 18-20 years of age is often labeled 'late adolescence'), and individuals 26 years of age or older are described as 'adults' or 'older adults' [56, 57]. Since the US federal minimum age for the sale of TNPs was raised from 18 years old to 21 years old [58], the terms 'underage adults' or 'underage young adults' have also been used to describe young adults aged 18-20 years in the US.

Assessing Sex at Birth and/or Gender

Sex at birth and gender are two distinct constructs. Our recommendations are developed based on several existing guidelines [59, 60]. Sex at birth refers to the sex recorded on a person's birth certificate. Sex at birth is based on biological attributes, commonly external genitalia, and typically consists of two categories: male and female. Inter-sex is a third potential category,

[7] Where applicable, copyright clearance should be obtained.

which corresponds to people born with biological characteristics diverging from the male and female categories [60, 61]. However, this category is not commonly assessed in surveys. Recommended survey items include 'What is your sex?' [52, 62] or "What sex were you assigned at birth?" [63]. In addition to the typical "Male" and "Female" response options, "I don't know" and "I prefer not to answer" should also be included. Participants should choose only one answer for this question.

Gender is a multidimensional construct that has psychological, social, and behavioral dimensions, including gender identity and gender expression. Gender identity refers to a person's internal sense of gender, and gender expression refers to how a person expresses their identity through appearance and behavior [61]. The most common gender identities are man and woman, matching the sex they were assigned at birth. On the contrary, the gender identity of people who are transgender does not match the sex they were assigned at birth [64]. Also, transgender is an umbrella term that includes a wide range of gender identities, including transgender men, transgender women, queer, gender variant, transsexual, and cross-dresser [63].

Recommended survey items for gender identity include "How would you describe your gender?" [65], or "How do you describe yourself?" [60, 61]. The recommended response options include "Male", "Female", "Transgender", or "I do not identify as female, male, or transgender" [60, 61]. An "I prefer not to say" response option may also be included. Participants should choose only one answer for either of these questions.

Sex at birth and/or gender may be assessed in a TNP use survey depending on the specific interests of the study. In a two-step approach, both questions may be posed to participants, starting with the sex at birth question, followed by the gender identification question [61]. For studies interested in evaluating the beliefs, perceptions, and behaviors associated with the use of tobacco and/or nicotine products in sexual minorities, additional questions aimed at identifying gender identity and gender expression may be considered. The GLAAD Association provides a comprehensive glossary of terms that can be used to help understand and differentiate gender identities and expressions [64], and the EU LGBT Survey – Technical Report is a resource for methodology and survey questions that can be used to properly identify and express transgender identities [63].

Assessing Sexual Orientation

Sexual orientation has three main dimensions: sexual attraction, sexual behavior, and sexual identity [60, 66]. We recommend that sexual identity be assessed using a single item, i.e., "Do you consider yourself to be..." with the following response options: "Straight", "Lesbian or Gay", "Bisexual", "Something else" [52]. Additional response options may include "Not sure" and "I prefer not to say". Participants should choose only one answer to this question.

If sexual attraction and/or sexual behavior are areas of interest, the Federal Interagency Working Group on Measuring SOGI provides an overview of current measures of sexual orientation in US federal surveys [66] and is a relevant resource to identify survey items that can be used to further investigate sexual orientation. In European and other countries, sexual orientation may be considered highly confidential information.

Assessing Race and/or Ethnicity

Race and ethnicity are two distinct concepts, that have evolved over time [67] and have been extensively defined and re-defined in social science and epidemiologic research. While ethnicity is defined as *"a group of people that identify with each other based on shared ancestry"*, race is more ambiguous and has been described as "*a sociopolitical construct used to categorize individuals into social groups"* [60, 68]. Different racial categories are anchored

in a historical context of colonialism, in which individuals sharing a common race are perceived as a homogeneous group with respect to biological inheritance. However, biological inheritance is not observable, therefore, it would be more accurate to refer to the assessment of race as a perception based on external features and phenotypes [67].

Methods to evaluate race and ethnicity may vary by geographic location, which may be due to perceived sensitivity associated with survey items about race and/or ethnicity and/or differences in reporting requirements by country, traditions and customs within each country, and/or national census classifications. For example, in US national surveys, participants may be asked about race directly using a single item (e.g., "What is your race?") that is followed by a second item asking about ethnicity, in addition to asking if the participant is Hispanic (Latino), or of Spanish origin [52]. However, race may be assessed using more elegant phrasing (e.g., "Which of these groups describe you?" [51]). In US national surveys, racial response categories are commonly asked as multiple-choice questions with responses that include "White", "Black or African American", "American Indian or Alaska Native", "Asian", "Native Hawaiian or other Pacific Islander", while ethnicity response categories include Spanish origin categories such as "Mexican", "Puerto Rican", "Cuban", etc. Surveys in countries with large Asian populations can also differentiate Asian populations, such as "Asian Indian", "Chinese", "Filipino", "Japanese", etc. In the US, a high non-response rate has been observed among Hispanic respondents when questions about race and ethnicity are separated into two survey items. As a result, Weinberger et al. suggested combining questions about race and ethnicity into one survey item [60]. They also emphasize that standard categories may not be sufficient to comprehensively capture racial/ethnic backgrounds and may be supplemented by additional categories relevant to the population of interest and the study outcomes.

In European-based surveys, race and ethnicity are not assessed under these terms. Instead, survey items ask how the respondent would describe their ancestry [54]. Response categories vary according to the prevalence of race/ethnicity in the country of interest. Most European-based surveys reviewed for this project do not assess race and/or ethnicity (e.g., SHP [69], SHS [70], EBS [71], EHIS [72]). In international-based surveys, such as GATS, race and ethnicity are merged into one item that asks about the respondent's racial/ethnic background, in which response categories are country-specific [65].

Assessing SES

SES is typically assessed by asking about the level of education, occupation, and income and can be measured at the individual or household level. Additional indicators may include wealth and savings [73]. Indexes have also been developed and further revised to assess SES, including Duncan's SEI and the ESCS index. The description hereafter focuses on the assessment of the three most used indicators (income, level of education, and work status/occupation).

Income

Income is typically assessed over the past 12 months and may be measured at the individual or household level. Response categories cover income ranges (e.g., less than $10,000, $10,000-$14,999, $15,000-$19,999, etc., [52]), and it should be specified whether the respondent should consider taxes and compulsory deductions in their response. Responses to this item may be inaccurate, depending on who in the household is answering the survey. An option to answer with either of "I don't know" or "I prefer not to say" should be included. In addition to income range, other items may include the source(s) of income (e.g., salaries, self-employment, pension, etc.), number of people living in the household, and how many family members contribute to the household income. Assessments of income may be tailored to specific countries to account for region-specific variability in economy, salary, and cost of living. Methods for developing items to assess income and for harmonization across countries are

available and have been applied in multi-national surveys, such as the ESS survey [54, 74, 75]. Recommended survey questions include “Which of the following categories best describes your total household income in the past 12 months?” [52] or “Please tell me which letter describes your household's total income, after tax and compulsory deductions, from all sources? If you don't know the exact figure, please give an estimate.” [54]. The ESS survey also asks respondents how they feel about their income (e.g., “living comfortably” to “finding it very difficult”) to allow for comparability among surveys.

Level of Education

Education systems and degrees vary considerably among countries. Therefore, the assessment of education level needs to be tailored based on regional differences. Typically, we recommend assessing education level using a single choice item asking about the highest level of education achieved. Response options are usually based on a particular country’s educational system, are ordered from lowest level to highest level, and include a “I prefer not to say” option. This item may be complemented by a second item asking about the number of years of education, which would be provided in a numerical format.

Level of education may be difficult to assess in the context of multi-national research because it requires a harmonization step to make the data comparable. Section 6.3 of the present guideline describes the assessment and harmonization of education level in the ESS survey, as an example of achieving measurement equivalence in multi-national surveys.

Work Status and Occupation

Work status, type of work, and work environment are correlated with TNPs use status. Social or cultural effects related to occupation are important determinants of smoking [76], and substantial differences in smoking prevalence have been observed across industry and occupation groups [77].

Survey items assessing work status may ask about employment status [52] and alternatives to working for profit status categories, such as student, unemployed, and in compulsory military service [72]. Surveys may also include categories indicating other activities associated with work, such as currently seeking employment (ESS [54]). NHIS has an employment module, which includes additional questions about work status. TUS-CPS includes questions about the workplace environment and policy to assess the use and exposure to smoking at the workplace. The 2017-18 US NHANES Occupation Questionnaire also collects data on employment and variables related to the work environment [78]. The survey recall period may vary from the past year (most of the surveys evaluated by CROM TF Descriptive CROM WG), to the past 7 days (e.g., GATS). Survey items assessing current occupation may be organized by labor group. The ISCO-08 provides a four-level hierarchically structured system for classifying and aggregating occupational information, which allows all jobs in the world to be classified into 436 unit groups [79]. These unit groups form the most detailed level of the classification structure, which are then aggregated into 130 minor groups, 43 sub-major groups, and 10 major groups, based on their similarity in terms of skill level and skill specialization required for each job. We recommend using this classification system, which allows for the compilation of detailed and internationally comparable data and provides summary information for only 10 groups at the highest level of aggregation.

Residency

Residency is often used as an eligibility criterion. Residency may be assessed to confirm the place of residence in a specific country using a dichotomous response option (yes/no) or may suggest a list of countries from which the respondent should select one option. Country-specific items may be developed to establish the state and/or region of residence.

Religion

Responses to items about religion may be country specific. If assessment of religion is of interest to the study outcomes, the most prevalent religion(s) in the country in which the study takes place should be identified and listed as response options. Other response options should include "Other", with the possibility to enter text only if the study budget and timeline allows for qualitative data analysis, and a "I prefer not to say" response option.

Tobacco-Related Disparities

Items that assess demographics and SES allow for identification of "vulnerable populations" [80] with regards to use of TNPs, such as minoritized sex, non-traditional gender, and sexual orientation identities; persons with minoritized racial and ethnic backgrounds; persons with lower SES; persons with lower health literacy; and persons with mental health concerns. These groups present a higher prevalence of TNP use, are under-represented in TNP research, and may experience tobacco-related health disparities [60, 81]. Therefore, if the research plans to study tobacco-related health disparities, we recommend conducting subgroup analysis for these groups (discussed further in Section 5.4).

5.1.2. TNP Use Prevalence

TNP prevalence (or product use rate) is the proportion of individuals in a population of interest who use the TNP at a specified point in time or over a specified period. Prevalence can be evaluated at a product category level (e.g., ENDS) or at a product subcategory level (e.g., e-cigarette, e-hookah, e-cigar, etc.) (see Section 4.1 for classification of TNPs). There are various ways to measure TNP prevalence depending on the timeframe, use frequency, and other use behaviors.

- **Lifetime use prevalence** is the proportion of individuals in a population of interest who at some point in their lives have ever used TNPs. Examples of lifetime use prevalence include 'ever use' and 'lifetime established use' (as discussed in Section 4.2) among survey respondents or subpopulations. 'Ever use' prevalence is usually determined based on the question 'Have you ever used [TNP] even one or two times?'. For emerging TNP categories, it is common to ask an awareness question first, such as 'Have you seen or heard of [TNP] before this study?' [52]. For lifetime established use, a typical question for cigarette smoking is 'Have you smoked at least 100 cigarettes in your ENTIRE LIFE?' [53]. In some surveys (e.g., NHIS), the cigarette lifetime established use criterion (i.e., having smoked 100 or more cigarettes) is used to define 'ever smoking' [53], which should be viewed as ever established cigarette smoking based on our definition of 'ever use' and 'established use'. The PATH Study also includes a non-numerical threshold measure of 'Have you ever used [TNP] fairly regularly?' to define 'ever established use'. In addition to this non-numerical threshold of established use, suggested numerical thresholds to define 'lifetime established use' for other TNP categories are summarized in Section 4.2. We recommend including lifetime established use questions, either with numerical thresholds or with non-numerical criteria (see Wei et al. [34], for detailed discussions), in surveys to distinguish experimental users and established users.
- **Point use prevalence** is the proportion of individuals in a population of interest who use a TNP at a specific point in time. An example is 'current use', which is usually assessed by asking 'Do you now use the [TNP]' with 'every day', 'some days' and 'not at all' [52, 53] or 'daily',' less than daily', and 'not at all' [65] as response options.

- **Period use prevalence** is the proportion of individuals in a population of interest who have been using a TNP during a given period of interest. Examples of period prevalence include past 7-day, past 30-day, and past 12-month TNP usage. Typical questions to assess period prevalence include 'In the past [time period], have you used [TNP] even one or two times' [52] and 'How long has it been since you last smoked part or all of a cigarette?' [51] with response options such as 'Within the past 30 days', 'More than 30 days ago but within the past 12 months', etc. We recommend measuring usage in the past 30 days as a recency of use to increase harmonization in research on TNPs. However, the recency of TNP use measures should also be selected based on the study objectives. For example, usage in past 7 days would be an important measure for studies that involve biomarker assessments.

Prevalence estimates for a TNP category may vary due to differences in product category descriptions, survey measures (current vs. past 30-day), and mode of survey administration [25, 82-84]. Point and period prevalence estimates are usually assessed among those who report 'ever use' of a TNP. 'Current (every day or some days) use' and 'past 30-day use' are the most used measures for current TNP use prevalence. Studies have also shown considerable variability in prevalence when it is estimated by different frequencies of use in the past 30 days [9, 29, 30, 85]. Amato et al. showed that a threshold of current use as '≥ 5 days during the past 30 days' for ENDS could restrict prevalence estimates to non-experimenters because experimenters are more likely to use the product infrequently and to discontinue use [30, 85]. Either a lifetime ever use criterion (as discussed in Section 4.2) or a threshold of current use frequency can be used to exclude experimenters from individuals who are currently using the TNP. We recommend applying the lifetime established use criteria shown in **Table 3** to identify experimental and established users of the TNP for use prevalence estimations.

5.2 Product Category Level Domain

Domains and subdomains in this section are usually evaluated at a product category or subcategory level. Participants should be provided with a clear description of each category of TNPs, which should be supported by product category images. The description and image(s) allow participants to differentiate among TNP categories and to prevent potential measurement errors, which may occur if respondents lack knowledge about product attributes and confuse a given product with other TNP categories. This may occur in the assessment of use behaviors associated with novel TNPs, such as HTPs, which may be confused with e-cigarettes.

5.2.1. Consumption

The primary subdomains under consumption are the number of days used in a pre-specified time frame (e.g., in the past 30 days), units used per day on days used, and type/form of products used (e.g., disposable ENDS products, ENDS products with replaceable pre-filled cartridges or pods, tank or modular systems that can be filled with liquids). Questions about consumption can be asked for each subtype within each TNP category. The questions corresponding to subcategories, subtypes, or type/form of TNP used are discussed for each TNP category.

Number of Days Used in the Past 30 Days

The number of days used in the past 30 days measure is used to evaluate TNP use frequency. Responses to this question typically include an option of '0-30' days among current or past 30-day users. To avoid inconsistency in responses when the question is asked in conjunction with the current use question, it can be assumed that the respondent's answer is '30 days' for those who report using a TNP 'every day'. Therefore, only respondents who report the use of a TNP product on 'some days' should be asked about the number of days used in the past 30 days.

Type/Form of the TNP(s) Used and Units Used per Day on Days Used

The number of units used per day on days used measure can be asked in conjunction with a number of days used in the past 30 days to obtain the past 30-day use of a TNP. The unit used in the measure is based on the subcategory or type of TNP.

Implementation of an upper limit for response options that use an interval scale (i.e., a range of units) should be considered to prevent an invalid response. For example, in the NHIS survey, a response option of '95' is coded for smoking 95 or more cigarettes per day on days smoked [53]. Additionally, instead of asking for numerical inputs of units used, the response option can be changed into categorical scale. For example, response options in the NSDUH survey include "less than one cigarette per day", "1 cigarette per day", "2 to 5 cigarettes per day", "6 to 15 cigarettes per day (about ½ pack)", "16 to 25 cigarettes per day (about 1 pack)", "26 to 35 cigarettes per day (about 1½ packs)", and "more than 35 cigarettes per day (about 2 packs or more)" [51].

As the product measurement unit would differ per product category, we summarize the commonly used measures for each category below. The selection of measurement unit may depend on study objectives.

[**Conventional Cigarettes**]

Manufactured cigarettes and roll-your-own cigarettes are two common cigarette types included in survey questionnaires. The unit commonly used is the individual cigarette stick. Many surveys also remind respondents how many cigarettes are in a pack (20 in most jurisdictions) because people who smoke sometimes think of their cigarette consumption in packs.

[**Cigars**]

Cigar subcategories include traditional cigars, cigarillos, and filter cigars. Blunts, which are modified cigars of any type in which the tobacco is removed and replaced with marijuana, are sometimes of interest when studying cigar usage. The unit commonly used is the individual cigar or cigarillo.

[**ENDS Products**]

Development of standardized self-report survey measures of ENDS product consumption is challenging due to the several forms of ENDS products that are available [9, 10, 86-88]. The type of ENDS product used is critical to understanding how ENDS products are consumed [10]. There are currently three major types of ENDS products: disposable ENDS products, ENDS products with replaceable pre-filled cartridges or pods, and tank or modular systems that can be filled with liquids. In addition to these 3 major types, there are rechargeable devices and non-rechargeable devices. Other relevant and important measures for ENDS consumption include use of products with or without nicotine and nicotine concentration, if applicable [10] as different product use behaviors have been observed when comparing nicotine-containing and non-nicotine-containing products [89].

As of now, there is little research to demonstrate the ***reliability*** and ***validity*** of the various unit measurements for ENDS products. Liu et al. conducted a qualitative assessment of e-cigarette use and concluded that 'number of times and/or puffs taken in a day' is the most common approach to describe quantity used compared to device-specific terms (i.e., replacement of disposable devices, cartridges/pods, use of e-liquid) and perceived equivalence to a quantity of traditional cigarettes [88]. In the PATH Wave 4 Survey Questionnaire, the 'number of times' used measure was defined as the 'number of times one picks up the ENDS product to use it' [52]. The questions in PATH Wave 4 were asked as follows: "On average, on the days that you use, how many times each day do you pick up your electronic nicotine product to use it, whether you take one puff or several?" and "Each time you pick up your electronic nicotine product to

use it, about how many puffs do you take?". The combination of responses to these two questions accounts for potential differences in daily use patterns. Additional device-specific unit measures for ENDS products include number of disposable ENDS products used, number of replaceable prefilled ENDS cartridges used, the frequency of filling the ENDS product with e-liquid, and the number of milliliters of e-liquid the device holds. We recommend the number of times or use occasions and puffs per time or use occasion measure if the research objective is to study overall ENDS category-level consumption and/or to report usage patterns. However, it is worth noting that a recent study has shown that the number of puffs per use occasion may be underestimated. The number of puffs may indicate relative heaviness of use across individuals but may not be a reliable measure to quantify the amount of nicotine taken over the course of several days [90]. Device-specific unit measures should be considered if the research objective is to study a specific ENDS product or subcategory.

[**Smokeless Tobacco Products**]

Smokeless tobacco product subcategories include moist snuff/dip, dry snuff, snus, and chewing tobacco. Two general types of smokeless tobacco include loose smokeless tobacco products and smokeless tobacco products in pouches. The units commonly used for smokeless tobacco products include number of times used, the number of pouches used for smokeless tobacco in pouches, and the number of cans used for loose smokeless tobacco products. We recommend the number of times or occasions of use measure if the research objective is to study smokeless category-level consumption and report general usage. However, subcategory-specific items should be specified if the research objective is to study a specific smokeless product or a subcategory, such as use of a pouched product.

[**Nicotine-Containing Tobacco-Free Oral Products]**

Tobacco-free oral nicotine products are available in pouches and other forms. We recommend the number of times/occasions measure if the research objective is to study general usage of tobacco-free nicotine-containing oral products. However, form-specific units (numbers of pouches, chewable pieces, lozenges, etc.) should be used if the research objective is to study a specific product or a subcategory.

[**Heated Tobacco Products**]

The unit commonly used for HTPs is the tobacco stick, which is specifically engineered to be heated to temperatures below the point of combustion by a battery-powered holder [91, 92].

5.2.2. Brand Usage

TNP users who have a regular brand or who own a TNP are typically asked about the brand of TNP they use most often or used last. We recommend asking the brand that is used most often in the past 30 days to assess brand usage. When applicable, product images can be provided to facilitate the selection.

5.2.3. Flavor Usage

Like usual brand used, respondents can be asked about first flavor used, flavor(s) used most often, usual use, or last used, in addition to what flavor(s) they used in the past 30 days. First flavor(s) used upon ***initiation*** of use of a TNP(e.g., ENDS) has been studied to evaluate trends associated with TNP experimentation, subsequent tobacco use, and TNP use progression [93, 94]. We recommend asking flavor(s) used in the past 30 days as individuals using TNPs are likely using multiple flavors, especially for emerging TNP categories. When applicable, product images can be provided to facilitate the selection of flavor usage. When conducting secondary analysis of survey data, researchers should be aware that flavors may be misclassified depending on the brand selected by survey participants. For example, Villanti et al. [95]

observed inconsistencies among people who smoke cigarettes and their reporting of use of menthol vs. non-menthol flavored products where individuals may report that their usual brand was non-menthol while the brand selected could be a non-menthol brand and for which at least 99 % of sales for that brand were menthol.

5.2.4. Initiation, Cessation, and Relapse

Initiation

Initiation of use of a TNP generally refers to the first use of that product. Commonly used survey measures to study TNP initiation include age/year of first use, age/year of first daily/regular use, and length of time as a daily/regular user of the product. Information on the first TNP a respondent tried can be used to understand an individual's TNP use trajectory. Current and established use, subcategories, types, and flavors as mentioned in previous sections are also relevant measures to understand TNP use initiation.

Cessation

TNP ***cessation*** occurs when an individual stops using a TNP after having used the product to at least its lifetime use criterion (i.e., after established use). Former established users are often asked how long it has been since they last used the product and if they have completely stopped using it. Based on when the product was last used, as discussed in Section 4.2, respondents could be categorized as 'recent former users' or 'long-term former users.' Respondents could be asked about TNP use patterns before cessation and alternate TNPs used to further understand events that may lead up to cessation of TNPs.

Quit Related Measures

Attempts to quit using a TNP ("***quit attempts***") refer to having stopped using the product for > 1 day during a specified time frame (e.g., past 12 months) because they were trying to quit using the product [96]. Quit attempts are considered an important intermediate step in TNP cessation [97]. Additional relevant survey items include ever tried to quit, interested in quitting, number of quit attempts over a timeframe, and duration and recency of quit attempt(s), methods used in quit attempts or in successful quitting, and attempts to decrease consumption during the (recent) quit attempt.

There have been various instruments developed to measure intention to quit smoking, such as a Stages of Change measure [98], the MTSS, and a Likert scale measure [99]. The MTSS developed by Kotz, Brown, and West has been shown to provide a strong prediction of attempts to quit smoking and is a candidate to monitor a user's level of intention to quit smoking [100]. The MTSS has also been shown to have comparable construct and predictive validity compared to other instruments [99]. Additionally, readiness to quit can be studied using a pre-specified time frame of planning to quit the product or based on the stages of change in the process of quitting [101].

Relapse/Re-Initiation

Relapse/re-initiation are terms often used to refer to the use of a TNP after a period of abstinence (e.g., 1 year).In general, use in this context refers to current use of the TNP.

5.3 Poly-/Cross-Category Level Domain

Survey items discussed in this section usually involve two or more TNP categories, like assessment of dual/poly usage and switching or transitioning from one product to another.

Dual/Poly Usage

The growing diversity of TNPs available in the market has led to increased prevalence of concurrent use of two or more TNPs [102, 103]. In epidemiological studies, dual and poly use are typically derived variables and are often operationalized based on measures of current use of TNPs. Dual use is typically defined as concurrent use of two TNPs from different TNP categories or subcategories (e.g., dual use of cigarette and ENDS, dual use of ENDS and HTPs, etc.). Similarly, poly use is usually defined as concurrent use of three or more TNPs. Due to the heterogeneity found among dual users, dual users can be further categorized into subgroups, such as 'Dual Daily', 'Predominant A', 'Predominant B', 'Concurrent Non-Daily' states [43-47] (see discussions in Section 4.2).

Switching and Transitions

Transition refers to a change in a use state based on the TNPs used before and currently.

Particularly, 'switching' refers to completely transitioning from the current TNP to another TNP. Individuals who switch may be a subpopulation of quitters who no longer use the product that they used before. Transitions or switching behaviors are sometimes directly evaluated using retrospective measures in cross-sectional surveys or they may be derived based on use of one or more TNPs at each time point in longitudinal surveys.

5.4 Descriptive CROM Domains that Overlap with Psychometric CROM Domains

Health and health-related quality of life domains and reasons for TNP use may be assessed with either Descriptive or Psychometric CROM. For health and health-related quality of life domains, for example, items that evaluate the respondent's perceptions of health-related quality of life are Psychometric CROM, whereas survey items that ask about medical diagnoses are Descriptive CROM. We will provide recommendations for survey measures that fall under the overlapping domains in the CROM repository.

5.5 Recommendations for Selecting an Existing Descriptive CROM

Sections 5.1 to 5.4 provide recommended Descriptive CROM to measure sociodemographic variables, TNP usage, and other characteristics and behaviors. When considering the most appropriate Descriptive CROM for a study whether the Descriptive CROM are part of a participant screener for use in a clinical study, a survey of TNP use administered daily during an actual use study, or a national survey assessing prevalence of TNP use, it is always recommended that the researcher start by clearly defining what needs to be measured. This may be particularly important when measuring health and functioning or reasons for product use, which may require either Psychometric or Descriptive CROM, depending on what the researcher intends to measure[8]. (If Psychometric CROM are needed, the reader is referred to the "Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with

[8] For example, if the researcher intends to assess the presence or absence of cough during the past week (yes/no), a Descriptive CROM would be used. Conversely, if a researcher is looking to assess severity of respiratory symptoms over the past week, this would be estimated using a Psychometric CROM because the researcher is trying to estimate an underlying construct (respiratory symptomatology).

Respect to Psychometric CROM for Use in Research on TNPs" [1]). When defining what needs to be measured, the researcher should be as specific as possible, considering the particular study in which the Descriptive CROM will be used. To illustrate, a researcher may need to measure *the number of days that participants who smoke cigarettes in an actual use study used the candidate product during the past week.* Importantly, this definition refers to several study-specific factors, including:

- the target population, or end-users of the Descriptive CROM ("participants who smoke cigarettes", which would be further defined in the study protocol)
- the timeframe (recall period) ("over the past week")
- the behavior to be measured ("use of the candidate product")
- the units of measurement ("number of days").

Having a clear and specific definition will help facilitate CROM selection, or, if an existing CROM does not need the researcher's needs, the definition will guide CROM modifications or the development of a new CROM. Generally speaking, the researcher should first consider the consensus measures recommended in Chapter 5 of these guidelines. If Chapter 5 does not include a recommended measure to assess the characteristic or behavior of interest (e.g., health and functioning, claim comprehension, reasons for product use), or if the recommended Descriptive CROM from Chapter 5 are not appropriate for purposes of the current study, then it is recommended that the researcher select Descriptive CROM that have evidence of validity from peer-reviewed literature or national/international surveys. If the researcher is still unable to identify an existing Descriptive CROM appropriate for the study, then an existing CROM may need to be modified or a new Descriptive CROM may need to be developed. It is not uncommon that existing Descriptive CROM need to be modified to fit the study's requirements (e.g., modifying the recall period, modifying the CROM to reference a different product category, etc.). The next sections of these guidelines discuss recommendations and best practices for modifying existing Descriptive CROM, as well as recommendations for developing and validating a new Descriptive CROM.

6. DEVELOPMENT AND MODIFICATION OF DESCRIPTIVE CROM

As discussed in Chapter 5, there are many well-established Descriptive CROM to assess various domains for various categories of TNPs. We recommend relying on or modifying an existing Descriptive CROM, when suitable, before developing a new Descriptive CROM.

6.1 Modifying an Existing Descriptive CROM

Descriptive CROM modifications can vary in terms of the *type* of modification (i.e., changes to content, administration, and/or application) and the *extent* of the modification (i.e., minor vs. substantial). In this section, we present definitions and examples of the types and extent of modifications that a researcher might make to an existing Descriptive CROM. Then, we discuss qualitative and quantitative strategies that can be used to gather evidence to support the modification, as well as the factors that influence the type and extent of evidence recommended to support the modifications. Portions of this chapter come directly from or are modified from the "Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with Respect to Psychometric CROM for Use in Research on TNPs", and additional details can be found in the guidelines [1].

As described above in Section 5.5, prior to modifying an existing CROM, the researcher should have clearly defined what they intend to measure. This definition will guide what modifications are needed for the CROM to fit the study's needs. Depending on the type and extent of the modifications and the content of the CROM, it may be helpful for the researcher to consult literature, SMEs, or individuals representing the end-users of the CROM (the intended population of respondents to whom the CROM will be administered) when revising the CROM. Section 6.2 includes recommendations for drafting the content of a new CROM (e.g., using simple, direct, unambiguous language) which should also be considered when modifying the content of an existing CROM.

Types of Descriptive CROM Modifications

The researcher may choose to modify an existing Descriptive CROM in various ways to make it fit the needs of their study, such as making modifications to the CROM content, administration, and/or application. These *types* of CROM modifications are further defined in Table 4. In practice, a CROM modification may impact multiple areas; for example, if modifying a Descriptive CROM pertaining to consumption of cigarettes to reference consumption of ENDS and administering it to people who use ENDS (as opposed to people who smoke), this would include modifications to both Content and Application. See Table 4 for examples of each type of modification[9].

[9] This table was adopted from Psychometric CROM Guidelines and modified as needed for application to Descriptive CROM.

Table 4 - Types of Descriptive CROM Modifications

Type of Modification	Illustrative Examples (Non-Exhaustive)
Content: Modifying the instructions, items, and/or response options	• Removing or introducing a response option of "I don't know" • Changing the number of response categories (e.g., increasing the granularity of an item asking about household income) • Changing response category labels • Changing instructions and/or item content to reference a different product category (e.g., "ENDS" instead of "cigarettes") or a specific brand • Changing language/terminology (e.g., changing "e-vapor" to "e-cigarettes") • Adding product images to items asking about use of that product • Changing the recall period (e.g., "in the past 30 days" to "in the past 7 days")
Administration: Changing the mode, method, and/or format of administration	• Administering a descriptive CROM developed for paper-and-pencil electronically • Changing the method of administration from self-completed to interviewer-administered • Changing the order of item administration (items asking about the use of different TNPs are presented in a random order instead of fixed)
Application: Applying the CROM in a new way, such as to a new population or product (from which it was originally developed / validated)	• Modifying and applying measures of cigarette consumption to the consumption of a new TNP category[a] • Translating a descriptive CROM into a different language and administering it to a new population (i.e., individuals whose primary language differs from languages the CROM has been validated for) • Administering a CROM to individuals from another culture (i.e., individuals whose cultural background differs from the background of individuals for whom the CROM was originally validated for)

[a] This would be an example of modifying CROM content as well; as stated above, it is not uncommon for different types of modification to occur in tandem.

Extent of CROM Modifications

CROM modifications also differ with respect to the *extent* of the modification. In theory, modifications fall on a continuum and range from very minor to substantial (Figure 5), with some substantial modifications departing so grossly from the original CROM that the modified CROM should be considered a new CROM (in these circumstances, the researcher is advised to follow Section 6.2). Within the context of these guidelines, we adopt definitions of "Minor" and "Substantial" modifications taken from "Consumer-Reported Outcome Measure (CROM) Best Practices and Guidelines with Respect to Psychometric CROM for Use in Research on TNPs" [1].

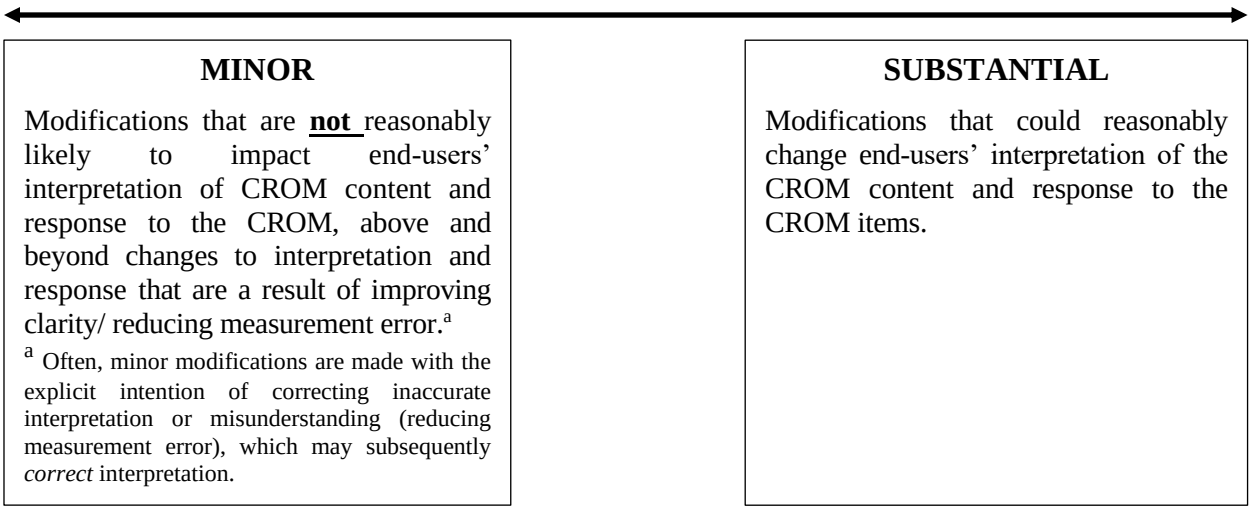


Figure 5 – Extent of CROM Modifications

As indicated in Figure 5, Minor modifications are modifications that are not reasonably likely to impact end-users' interpretation of CROM content and response to the CROM, above and beyond changes to interpretation and response that are a result of improving clarity/ reducing measurement error. Conversely, Substantial modifications could reasonably change end-users' interpretation of the CROM content and response to the CROM items. Illustrative, non-exhaustive examples of Minor and Substantial modifications that could be made to Descriptive CROM are presented in Table 5. Distinguishing between Minor vs. Substantial modifications is important as these classifications are linked to different recommendations pertaining to the need for empirical evidence to support the modification(s). That is, for Minor modifications, no additional evidence is needed to support the modification. That said, additional qualitative evidence may still be helpful in some circumstances to support the modification, such as evidence from cognitive interviews (discussed below in Section 6.2). Conversely, qualitative evidence is typically recommended to support Substantial modifications. In some instances, the researcher may also decide to collect quantitative evidence to help support substantially modified CROM (see Section 6.2 below). However, in most cases, quantitative evidence is not necessary, and qualitative evidence is sufficient to support the modification.

The responsibility of determining the classification of a modification as Minor vs. Substantial and defending the decision to collect or not to collect evidence to support the modification ultimately falls on the researcher and should be justified. While in theory some modifications may seem to fall between Minor and Substantial, into a "Moderate" category, this classification is not meaningful as the researcher will ultimately need to decide whether to take the more conservative approach and collect evidence to support the modification, consistent with the recommendations in Table 5 for Substantial modifications, or decide that additional evidence to support the modification is not warranted (following the recommendations for Minor modifications in Table 5). It is also worth noting that if modifying an existing CROM from literature or elsewhere that does *not* have any evidence of validity, collecting qualitative evidence to support the modified CROM is generally recommended.

Table 5 - Recommendations Pertaining to CROM Modifications

Modification	Minor	Substantial
Examples	• Making the text bold and underlining the recall period in the instructions ("In the **past 7 days"**) for visibility and emphasis • Changing font size or font style • Adding additional clarifying language to an item or instruction • Adding an image of the product being referenced • Adding an "I don't know" response option • Administering a paper-and-pencil CROM electronically without changing the presentation of the CROM • Changing the item to reference a different brand	• Substantially changing the content of the CROM (e.g., changing the consumption measure of moist smokeless tobacco from times per day to cans per day) • Applying the CROM to TNPs for which it was not developed • Modifying and administering the CROM to a population for which it was not developed • Translating a CROM into a new language and administering it to this new cultural population
Recommended Approach(es) to Support Modification	• Generally, no evidence is needed • In certain circumstances, qualitative evidence may be helpful (e.g., to ensure that new clarifying language added to instructions is clear) • Usability testing may be helpful in some circumstances, such as when modifying a paper-and-pencil CROM for electronic administration	Qualitative evidence would likely be helpful and is generally recommended to support the modification. Qualitative evidence may be particularly helpful in the following circumstances: • If CROM content is substantially changed[b] • If responses from two versions of a CROM are being directly compared in a study • When administering a CROM to a new population and/or applying the CROM to a new product, and such changes could reasonably impact respondents' interpretation of the CROM and response to the CROM. • When translating a CROM into a new language[c] In some instances, quantitative evidence may also help support the modification.

[a] Often, Minor modifications are made with the explicit intention of correcting inaccurate interpretation or misunderstanding (reducing measurement error), which may subsequently *correct* interpretation.

[b] Depending on the modification, qualitative evidence is generally helpful to ensure that participants understand the new content. For example, response options may be too granular for participants to respond accurately (e.g., the exact number of cigarettes smoked in the past 30 days, the household income from last year), or recall periods may be inappropriate (asking participants to recall their product use from several years ago may yield inaccurate responses due to limitations with memory).

[c] It is generally recommended that the researcher work in close in collaboration with an expert or organization specialized in linguistic services to determine and execute the most appropriate linguistic and cultural validation strategy for developing or modifying an existing CROM.

6.2 Development and Validation of a New or Substantially-Modified CROM

Drafting CROM Content

In some instances, the researcher may determine that, after clearly defining what they intend to measure, a new Descriptive CROM is needed to meet the needs of their study. This definition will guide the content of the new fit-for-purpose CROM. Depending on the CROM content, consulting literature, SMEs, and/or individuals representing the end-users of the CROM (the intended population of respondents to whom the CROM will be administered) may be helpful when drafting the new CROM content. The following examples are presented for illustration:

- A researcher wants to assess whether respondents have been intubated in the past 5 years due to a respiratory condition. The researcher interviews pulmonologists to understand how patients describe intubation, so that this language can be used to define "intubated" in the Descriptive CROM. Including a definition of "intubated" that is understood by participants reduces measurement error (i.e., a respondent selecting "no" or "yes" incorrectly because they did not accurately understand "intubated"). Alternatively, the researcher might conduct a focus group of individuals with a history of having been intubated to determine the most appropriate language to describe intubation. After drafting the CROM with input from pulmonologists and/or individuals with a history of having been intubated, the researcher could conduct cognitive interviews with end-users to verify that the language of the new CROM is clear and understood as intended (qualitative and quantitative strategies for testing a CROM once drafted are described in greater detail below).
- A researcher is developing a Descriptive CROM to assess what strategies respondents used to help them quit smoking. To develop an appropriately comprehensive list of strategies, the researcher might conduct an electronic survey among individuals representing end-users of the CROM that asks respondents to describe how they quit smoking with an open-ended question; responses are then coded into categories, recording the frequency with which participants' responses fall into each of the categories. The researcher then uses the most commonly endorsed categories as response options for the new Descriptive CROM, and includes a response of "another way not listed above" to capture less popular cessation strategies.

The following recommendations should be considered when drafting a new CROM[10]:

Global recommendations

- Use simple language (be cognizant of reading level[11] [12]) and avoid technical terminology, slang, idiomatic expressions, or colloquialisms (if possible)
- Use direct, unambiguous language
- Avoid leading questions and biasing language
- Use of images can be helpful to aid comprehension/reduce confusion

With respect to CROM instructions/item content

- Each item should communicate a single concept (no "double-barreled" items)
- Avoid hypothetical questions
- Recall period should be relevant and appropriate

With respect to response options

- Response option labels should be appropriately labeled and relevant
- Response option labels should be appropriately granular (to meet the study's needs, while also balancing limitations in participant's memory, for example)
- Response options should cover the full range of potential responses (a response of "another reason not listed" [or another similar response option] may be helpful)
- Avoid response option labels that may bias the direction of the responses
- Use of "not applicable" should be avoided when possible (items should be applicable for participants, and skip patterns can be used to avoid administering items to participants for whom they are truly not applicable)
- "I don't know" (or other similar response options) should be visually distinct from the other response options, and should be placed last in the response set

Qualitative and Quantitative Strategies to Collect Validity Evidence of a New CROM

Once the CROM is drafted, a number of qualitative and quantitative strategies can be used to evaluate the psychometric functioning of the new Descriptive CROM. There is no single "correct" approach for validating Descriptive CROM; the most appropriate approach for collecting validity evidence of a new CROM will depend on the CROM content and should be determined in consultation with experts in measurement. Of note, it is not always necessary to formally evaluate the psychometric functioning of a new Descriptive CROM. Generally speaking, Descriptive CROM with simple, direct, unambiguous language measuring straight-forward characteristics or behavior that are not likely to be misunderstood by the respondent may not need validity evidence. For example, a question asking whether the participant has ever used "X" product even once (with an image of the product) is likely "face valid" and collecting validity evidence for this item is likely unnecessary. Conversely, it is best to collect validity evidence when an item asks about characteristics or behaviors that could potentially be misinterpreted / interpreted differently across participants (e.g., whether participants have used

[10] Adapted from Psychometric CROM Guidelines

[11] The researcher can assess reading level (Flesch-Kincaid grade level) using a feature in Microsoft Word or another program.

[12] FDA TPPI Guidance (2022) recommends that the reading level be "appropriate for those with less than a high school education" (p. 14).

a product “fairly regularly” or some other qualitative descriptor of product use behavior). Researchers are encouraged to consult measurement experts when deciding whether qualitative and/or quantitative psychometric validation is necessary for a new Descriptive CROM, and this decision will need to be justified. When in doubt, the more conservative approach is to collect qualitative and/or quantitative data to support the new Descriptive CROM.

Cognitive debriefing interviews are a commonly used approach that are helpful in many circumstances to evaluate the content validity of a new CROM (or a substantially modified CROM). These interviews are conducted for purposes of verifying that the CROM language is clear and is understood as intended (e.g., for a Descriptive CROM that asks about using a product “fairly regularly,” is the phrase “fairly regularly” understood as intended?), that the recall period is appropriate, that the response options are appropriate, etc. Issues with the CROM or opportunities for improving the CROM identified during cognitive testing can also be used to improve the CROM. Cognitive testing participants should represent the end-users of the CROM, and should be appropriately diverse with respect to demographics, TNP use history, etc., as appropriate. The researcher may choose to oversample specific groups of interest (e.g., individuals with low health literacy, individuals with particular TNP use histories) to ensure that the perspective of individuals from these groups are captured during cognitive interviews. For example, a Descriptive CROM may be accurately understood among those with normal health literacy but may be misunderstood among those with limited health literacy, who represent an important segment of end-users of the CROM. The mode and method of administration should also be considered and the new or modified CROM should be tested using all available modalities to avoid needing to modify it for alternative methods of administration in the future. Many books, articles, and guidance documents exist to provide interested readers with greater detail regarding the conduct of these interviews and analysis of cognitive interviewing data and interested readers are referred elsewhere (See [104, 105]).

In some circumstances, the researcher may decide to (also) collect quantitative validity evidence of a new Descriptive CROM. There is no “standard” approach that is appropriate for all Descriptive CROM, and the particular psychometric properties to be evaluated and analyses chosen to evaluate these properties will depend on various factors, such as the content and purpose of the CROM. For example, a participant’s response to a Descriptive CROM asking about whether the participant has ever been diagnosed with asthma should not vary over a brief period of time (e.g., 1 week), unless the participant received a diagnosis between assessment points. Conversely, a participant’s response as to the number of cigarettes smoked each day may indeed vary over time. In the first example, evaluating test-retest reliability of this asthma CROM over a 1-week period would be appropriate; as participants’ responses would largely be expected to be stable over this brief time period, high consistency between responses would support reliability of the new CROM. Conversely, in the later example (cigarettes smoked per day), the researcher would not be interested in evaluating test-retest reliability of this CROM because participants’ responses are expected to fluctuate (and would yield a low stability coefficient). Other examples of psychometric properties of Descriptive CROM that might be evaluated include convergent (evidence that the new Descriptive CROM is related to other measures that it should theoretically be related to) and discriminant validity (evidence that the new descriptive CROM is *not* related to other measures that it should *not* theoretically be related to). Convergent validity includes both concurrent validity (evidence that the new Descriptive CROM is related to another measure that it should be related to, and these two measures are completed at the same time [e.g., within the same survey]) and predictive validity (evidence that the new Descriptive CROM is related to another measure that it should be related to, assessed at a later point in time). Known-groups validity (when responses to the new Descriptive CROM vary as expected between groups of respondents) is another type of quantitative validity evidence that may be helpful to evaluate to support a new Descriptive CROM.

As another strategy to support the validity of a new Descriptive CROM, participants' responses to the new Descriptive CROM can be compared against "real-world" data (results from medical tests, diagnoses from medical charts, amount of product that the participant was directly observed to use during an in-clinic product use assessment, etc.). Of course, whether this is relevant or appropriate will depend on the new CROM, and this approach is not always feasible for many reasons (e.g., confidentiality of medical records, the additional cost that is often associated with collecting or accessing real-world data, etc.).

Although not a strategy to collect validity evidence, usability testing may be helpful in certain circumstances, such as when modifying CROM formatting developed for a full-sized computer screen or a paper survey to fit a small-screen electronic device (smartphone).

6.3 3MC Research

Large-scale surveys are sometimes conducted at multi-national, multi-cultural, and/or multi-regional levels. The main objective of 3MC research is to compare data and outcomes among countries, "cultures", and/or regions. A primary challenge of 3MC research lies in maximizing comparability, quality, and equivalence of items and constructs across surveys, as in general, languages, social systems, "cultures", and society are substantial and relatively complex and differ across countries and within countries.

"Measurement equivalence implies that the instrument measures the same concept in the same way, across various subgroups of respondents […] it does not mean that there are no differences between the populations regarding a measured construct. Rather, it implies that respondents from different groups that have the same position on a trait of interest should provide a similar response" [106]. Ensuring measurement equivalence in 3MC survey research is necessary to produce data that are comparable across contexts (e.g., countries, and cultures). Comparison errors arise when comparing survey data collected in several countries/regions if the questionnaire content (instruction, item stem, items, response options), and/or methods of conducting the survey differ. This affects data quality and the interpretability of the results and related research conclusions.

The goal of linguistic equivalence is to make data comparable by ensuring that each aspect of a questionnaire (i.e., questionnaire name, instructions, item stems, items, response options) when presented in two or more languages conveys the same meaning and can be understood equally by respondents from each language. Simple literal translations, even when correct, may not be enough to ensure equivalence. Basic principles for linguistic equivalence include ensuring consistency in question formulation across different languages, using simple vocabulary, conveying the same meaning in the questions across different languages, being consistent in the structure and layout of the questionnaire across different languages, and ensuring that the structure of the response scales and continuum.

If the participants in 3MC research are asked to provide data that may differ across groups, it is necessary to use crosswalks to ensure comparability. See references [75, 107, 108] for best practices for ensuring data comparability.

The ESS survey has achieved high standards in multi-national research methodology [54]. ESS multi-national survey has been conducted every two years across Europe since 2001. As an example of harmonizing international data, the ESS assesses country-specific education categories, which are then mapped to a cross-national scheme based on the international statistical classification for education-related data [109]. The ISCED is maintained by the UNESCO Institute for Statistics in Montreal, Canada [110, 111]. This step aims to harmonize educational levels across countries, to standardize information, and to improve cross-national comparability and coding consistency over time. The country-specific educational level response options and harmonization guidelines can be found on the ESS website [110].

7. DESCRIPTIVE CROM DATA COLLECTION, ANALYSIS, AND REPORTING

7.1 Descriptive CROM Data Collection

Proper design of survey instruments and post-survey data processing can help reduce participant dropout and measurement error, which is essential to collecting high quality Descriptive CROM data.

Sampling Strategies and Recruitment

The objective of sampling is to generate a sample that is sufficiently representative of the population of interest to allow for statistical inference and generalizability of findings [112]. Two major approaches are probability and non-probability sampling. The primary objective of probability sampling is to ensure that each element of a population has a known probability of being included in the sample [113]. With probability sampling, the quality of the sampling frame is the dominant feature to ensure that there is adequate coverage of the desired population to be surveyed, including subpopulations of interest. In non-probability sampling, the chances of being included are unknown, so modeling, weighting or other adjustments are necessary to project from the sample to the larger population [114]. Statistical power analysis should be conducted to determine sufficient sample size to detect the hypothesized effect size(s) based on the primary research question(s) or objectives.

Modes of Data Collection

Survey questions may be interviewer-administered (i.e., a trained researcher asks the survey questions to participants and records their answers) or self-administered (i.e., participants access and answer the survey questions by themselves). The most common modes of interviewer-administered survey data collection include face-to-face interviews and telephone interviews. The most common modes of self-administered survey data collection include mailed paper surveys and web surveys (e.g., CAWI, CASI, and ACASI).

Similar to FDA's TPPI Guidance [11], we recommend selecting the mode of data collection best suited to the research objective, study design, method, and study population. Several factors should be considered when choosing a mode of data collection, including but not limited to: cost and budget, study purpose and objectives, survey design (e.g., structure, complexity, content, length), infrastructure, availability of key personnel, geographical distribution of the study population, and intended sample size. Thus, selecting a mode of data collection may require trade-offs between available resources, complexity of questionnaire content, and data quality.

In some circumstances, it is advisable to conduct a survey using a mixed mode design, which refers to a survey being conducted using several modes of data collection, and which *"offers the possibility of off-setting the weaknesses of one approach with the strengths of another"* ([115], p.9). For example, using mixed modes of data collection can help reach a larger panel of respondents such as those with and without internet access. It can reduce non-response rates e.g., by including people who choose not to answer telephone surveys, and it may optimize costs because some methods of data collection are more expensive than others. If a mixed mode approach is selected, it is important to minimize the mode effects on data quality and completeness. Dillman, Smyth, and Christian [116] discussed methods for implementing internet, phone, mail and mixed-mode surveys. FDA's TPPI Guidance provides additional recommendations on quantitative survey methods, study design, and experimental design [11].

Survey Structure

Generally, survey content should be structured to facilitate the respondents' ability to properly comprehend and respond to the questions [117]. Without the appropriate structure, a poorly designed survey can lead to low response rates and/or inaccurate and/or incomplete data. Different methods can be applied to ensure a suitable survey structure.

First, pre-testing implemented to validate the survey instrument and its measurements should include assessments of survey structure and design to ensure the visual appeal and clarity of the survey instrument. This may increase response rates and reduce the likelihood of collecting inaccurate or incomplete data. Second, surveys can incorporate multiple breaks throughout the questionnaire, to capture partial or incomplete responses so that subsets of the information collected can be used even when the survey in its entirety is not completed [117]. Third, the survey should be organized into logical categories (e.g., by theme) to help the respondents in organizing their thoughts to promote complete and accurate answers and to reduce frustration while completing the survey [117].

When developing a questionnaire, one must also be mindful of order bias. For example, randomizing the order of answer choices can help mitigate order bias [118]. Another method that can improve survey structure is the addition of skip or display logic to streamline the respondents' experience and help improve the quantity and quality of data. However, when applying such methods, it is vital to perform extensive testing with a wide variety of sample respondents to ensure that the correct questions are being displayed to the correct respondents, and that questions are not being missed [117].

While the recommendations for statistical analysis methods are not in the scope of this guideline, we discuss some of the key issues that affect CROM data processing, analysis, and reporting in the sections that follow.

Post-CROM Data Processing

Data coding, editing, and quality checks are necessary to ensure data integrity [119]. Some data validation checks/editing can be implemented in the computer-assisted interviewing software during the data collection step if that mode of data collection has been implemented. Preparatory analysis may be done to process the raw study data, including reformatting and correcting and combining data sets. Actions taken during editing, after raw data collection, should be coded and documented in the data set [120, 121]. We recommend that at least the main analyses be completed both with the original data and with the cleaned data to identify differences. Data cleaning includes identification and management of invalid responses, outliers, and missing data.

If probability sampling has been used to recruit respondents, respondent weights must be generated to correct for any disproportionality of the sample with respect to the target population of interest so that the study findings can be generalized. Weights can be viewed as adjustment factors assigned to each individual that account for their probability of selection, as well as other factors including non-response, and post-stratification.

Missing data may occur due to nonresponse and the rate of missing data may vary by data item. Nonresponse analysis should be conducted to obtain unweighted and weighted response rates and to examine if the data are missing at random [121, 122]. Missing responses may be ignored or imputed based on the type of analyses and the pattern of missingness [121, 123]. In cases where variables (e.g., demographic variables) are needed to create survey weights, statistical imputation methods should be applied to assign missing values.

Data protections should be implemented throughout the production process [121]. For surveys that include confidential data, procedures should be established to ensure that information is protected during production, use, storage, transmittal, and disposition of the survey data.

7.2 Descriptive CROM Data Analysis and Reporting

For primary survey studies, an SAP should be developed during the survey development stage and prior to data collection and analysis [124]. According to the FDA E9 Statistical Principles for Clinical Trials, "the statistical analysis plan may be written as a separate document to be completed after finalizing the protocol [124]. In this document, a more technical and detailed elaboration of the principal features stated in the protocol may be included". See FDA E9 Statistical Principles for Clinical Trials for the development of the SAP [124].

For secondary survey analyses, a SAP is also recommended to outline the analysis objectives and analytical methods.

Respondent Weights

When analyzing publicly available data from national/international surveys, recommended data analysis specifications (e.g., stratifications, clusters, weights, and replicate weights) to account for complex survey designs can often be found in the survey user guide. Appropriate use of survey weights can compensate for probabilities of selection, non-response, and other characteristics of the actual survey sample [125]. Proper use of survey statistics, including application of weights, is essential for the correct estimation of sample variance [125]. It is good practice to compare weighted and unweighted results on key study variables and weighting variables before conducting a full statistical analysis [126, 127]. Only variance estimates, not point estimates of effect, should differ between weighted and unweighted analysis results.

In general, the variability associated with descriptive statistics is affected more by the use of survey weights than is variability associated with multivariate models (e.g., regression models) [128]. Survey weights are also expected to affect analytical results to a greater extent when the study sample size is small compared to when it is large. When weights are not used to adjust for the sample selection process, the resulting estimates will be biased. If necessary, unweighted multivariate analyses can be completed if the variables that would be used to construct the survey sample weights are included as independent variables. All appropriate weights should be used in the analysis [129]. When accessing publicly available data from national/ international surveys, the data documentation should provide guidance about variance estimation for that sample. When producing an original data set, the scientific rationale for the weighting procedure should be included in the SAP along with the calculation methods needed to produce the correct weights.

Missing Data

The patterns and rates of missing data should be explored prior to undertaking statistical analyses. When the rate of missing data on key analytic variables is very low (e.g., < 1 %-2 % of cases), the effect on the analytical results will probably be small. However, when the rate of missing data is sizeable, there may be biases introduced by ignoring missing data. Bias will be introduced into the results if the missing data are not missing at random. Various missing data imputation methods can be implemented to correct the underlying processes that led to the missing information [123]. For large national/international survey programs, the data producer may perform missing data imputations and/or provide weights to account for nonresponse on key variables before the survey data are released for public use.

Reporting Data with Small Numbers

Small data sets or covariate patterns with small numbers of observations should be identified or suppressed for reasons including lack of reliability and the protection of respondent confidentiality, respectively. These problems can be resolved by aggregating data, for example, by collapsing smaller categories to form larger ones or by combining data for multiple years of data collection. We recommend suppressing (i.e., not reporting) results and/or noting their unreliability when the RSE or CV is greater than a certain threshold (e.g., 25 % or 30 %) or when the cell count is less than a certain threshold (e.g., $n < 50$); thresholds should be prespecified in the SAP [130]. For secondary analysis of complex survey data, follow data suppression rules as required by the survey sponsor.

Lastly, to deal with data with invalid responses, outliers, and missing data, we recommend conducting the analysis with original data and cleaned data to compare the differences in outcome measures under primary and secondary research objectives/hypotheses.

Reporting Research Findings

Thorough reporting of study methods, including the features of the survey design and sample selection, is essential. The guiding principle is for study design and analysis methods to be reported completely enough not only for their quality, correctness, and scientific rigor to be evaluated by a reader, but for the study to be replicated by others. If the survey research results are to be disseminated in a scientific journal, page limitations on publications may necessitate liberal use of supplementary files or appendices. A study archive, however, should completely document all features of the survey and its analysis and should include the final version of all analytical results and their interpretations.

8. SUMMARY AND CONCLUSIONS

The guideline provides a conceptual introduction to Descriptive CROM, with a focus on providing consistent definitions and descriptions of product and exposure categories and patterns of use. The WG engaged consultants from various sectors who have subject matter expertise related to research methods in general, as well as research methods specific to TNPs. With these engagements, we developed recommendations of Descriptive CROM based on existing Descriptive CROM from surveys that include modules to assess use of TNPs. The consensus Descriptive CROM measures, along with the use of consistent definitions, could facilitate comparisons across studies, aggregation of data sets, and eventually, improve harmonization in research findings.

With the changing TNP landscape, developing and identifying optimal Descriptive CROM becomes challenging, especially for emerging TNPs. We recommend modifying an existing Descriptive CROM, when suitable, before developing a new Descriptive CROM. When modifying an existing Descriptive CROM, depending on the type and extent of modifications needed, qualitative and/or quantitative evidence should be gathered to support the modification. The development and validation of Descriptive CROM should follow a well-established framework with an initial, qualitative assessment phase followed by a quantitative phase to assess the validity, reliability, and other properties of the CROM as deemed necessary. Lastly, proper survey instrument design and post-survey data processing are also essential components to ensure Descriptive CROM data quality.

These recommendations on the development, modification, and application of Descriptive CROM are grounded in scientific rationale and developed with consensus from the TNP research community. Readers of this guideline are advised to obtain appropriate technical training or to engage technical consultants, and to refer to guidance documents referenced in this report, to properly implement the guideline. Finally, as the best practices and guidelines may evolve over time, the CROM TF will update best practices and guidelines based on the dynamic TNP landscape and regulatory requirements to advance TNP research.

9. BIBLIOGRAPHY

[1] CORESTA, *Best Practices and Guidelines with Respect to Psychometric CROM for Use in Tobacco Regulatory Research*, Consumer Reported Outcome Measures Consortium, Editor. 2023 (Drafted).

[2] Corrigan, P.W., AC., *Understanding the impact of stigma on people with mental illness.* World Psychiatry, 2002. **1**(1): p. 16.

[3] Volkow, N.G., JA; Koob, GF., *Choosing appropriate language to reduce the stigma around mental illness and substance use disorders.* Neuropsychopharmacology, 2021. **46**(13): p. 2230-2.

[4] Williamson, T.R., KE; Carter-Harris, L; Ostroff, JS., *Changing the language of how we measure and report smoking status: implications for reducing stigma, restoring dignity, and improving the precision of scientific communication.* Nicotine and Tobacco Research, 2020. **22**(12): p. 2280-2.

[5] PhenX. *PhenX Measures for Tobacco Regulatory Research (TRR).* [cited 2023 July 16]; Available from: https://www.phenxtoolkit.org/collections/trr.

[6] Swan, G.E., et al., *The PhenX Toolkit: measures for tobacco regulatory research.* Tob Control, 2020. **29**(s1): p. s1-s4.

[7] Chrea, C., et al., *Developing fit-for-purpose self-report instruments for assessing consumer responses to tobacco and nicotine products: the ABOUT™ Toolbox initiative.* F1000Res, 2018. **7**.

[8] Glasser, A.M., et al., *Overview of Electronic Nicotine Delivery Systems: A Systematic Review.* Am J Prev Med., 2017. **52**(2): p. e33-e66.

[9] Weaver, S.R., et al., *Establishing consensus on survey measures for electronic nicotine and non-nicotine delivery system use: current challenges and considerations for researchers.* Addictive behaviors, 2018. **79**: p. 203-212.

[10] Pearson, J.L., et al., *Recommended core items to assess e-cigarette use in population-based surveys.* Tobacco control, 2018. **27**(3): p. 341-346.

[11] United States Food and Drug Administration, *Tobacco Products: Principles for Designing and Conducting Tobacco Product Perception and Intention Studies*. 2022, United States Food and Drug Administration.

[12] Sutfin, E.L., et al., *Measuring waterpipe tobacco smoking in survey research.* Tobacco control, 2020. **29**(5): p. 593-600.

[13] World Health Organization, *WHO report on the global tobacco epidemic 2021: addressing new and emerging products*. 2021.

[14] CORESTA, *E-Cigarettes: A Brief Description of History, Operation and Regulation*. 2014.

[15] Taljout, P.B.-t.H.H.B.B., *Nicotineproducten zonder tabak voor recreatief gebruik*. 2021, Rijksinstituut voor Volksgezondheid en Milieu.

[16] World Health Organization, *Proposal for Inclusion of Nicotine Replacement Therapy in the WHO Model List of Essential Medicines*, 17th Expert Committee on the Selection and Use of Essential Medicines, Editor. 2008: Geneva.

[17] Brouwer, A.F., et al., *The impact of current tobacco product use definitions on estimates of transitions between cigarette and ENDS use.* Nicotine & Tobacco Research, 2022.

[18] Wei, L., et al., *The impact of cigarette and e-cigarette use history on transition patterns: a longitudinal analysis of the population assessment of tobacco and health (PATH) study, 2013–2015.* Harm reduction journal, 2020. **17**(1): p. 1-12.

[19] Wei, L., et al., *Examination of Lifetime Established Use Criteria for Adult Tobacco Product Users (Drafted).* TBD, 2022.

[20] Wei, L.S., M; Hannel, T; Largo, E; Muhammad-Kah, R., *Examination of Lifetime Established Use Criteria for Adult Tobacco Product Users* F1000Res, 2023. **12**: p. 225.

[21] Bondy, S.J., J.C. Victor, and L.M. Diemert, *Origin and use of the 100 cigarette criterion in tobacco surveys.* Tobacco control, 2009. **18**(4): p. 317-323.

[22] Klemperer, E.M., et al., *Tobacco and Nicotine Use Among US Adult "Never Smokers" in Wave 4 (2016–2018) of the Population Assessment of Tobacco and Health Study.* Nicotine and Tobacco Research, 2021. **23**(7): p. 1199-1207.

[23] Cornelius, M.E., et al., *Tobacco product use among adults—United States, 2019.* Morbidity and Mortality Weekly Report, 2020. **69**(46): p. 1736.

[24] Johnson, S.E., et al., *Tobacco product use among sexual minority adults: findings from the 2012− 2013 national adult tobacco survey.* American journal of preventive medicine, 2016. **50**(4): p. e91-e100.

[25] Agaku, I.T., A.J. Awopegba, and F.T. Filippidis, *The impact of inter-survey differences in the definition of current smokeless tobacco use on comparability of US national and state-specific prevalence estimates, 2009–2011.* Preventive medicine, 2015. **74**: p. 86-92.

[26] Agaku, I.T., et al., *Tobacco product use among adults—United States, 2012–2013.* MMWR. Morbidity and mortality weekly report, 2014. **63**(25): p. 542.

[27] Giovenco, D.P., M.J. Lewis, and C.D. Delnevo, *Factors associated with e-cigarette use: a national population survey of current and former smokers.* American journal of preventive medicine, 2014. **47**(4): p. 476-480.

[28] Andrews, J.A., et al., *Perceptions and Use of E-cigarettes across Time among Emerging Adults.* Tobacco Regulatory Science, 2016. **2**(1): p. 70-81.

[29] Sánchez-Romero, L.M., et al., *A comparison of tobacco product prevalence by different frequency of use thresholds across three US surveys.* BMC Public Health, 2021. **21**(1): p. 1-11.

[30] Amato, M.S., R.G. Boyle, and D. Levy, *How to define e-cigarette prevalence? Finding clues in the use frequency distribution.* Tobacco control, 2016. **25**(e1): p. e24-e29.

[31] Corey, C.G., et al., *Little filtered cigar, cigarillo, and premium cigar smoking among adults—United States, 2012–2013.* Morbidity and Mortality Weekly Report, 2014. **63**(30): p. 650.

[32] Corey, C.G., et al., *US adult cigar smoking patterns, purchasing behaviors, and reasons for use according to cigar type: findings from the Population Assessment of Tobacco and Health (PATH) study, 2013–2014.* Nicotine and Tobacco Research, 2018. **20**(12): p. 1457-1466.

[33] Rostron, B.L., C.G. Corey, and R.M. Gindi, *Cigar smoking prevalence and morbidity among US adults, 2000–2015.* Preventive Medicine Reports, 2019. **14**: p. 100821.

[34] Wei, L., et al., *Examination of Lifetime Established Use Criteria for Adult Tobacco Product Users.* F1000Research, 2023. **12**(225): p. 225.

[35] Wheldon, C.W., A.R. Kaufman, K.A. Kasza, and R.P. Moser, *Tobacco use among adults by sexual orientation: findings from the population assessment of tobacco and health study.* LGBT health, 2018. **5**(1): p. 33-44.

[36] Robinson, J.N., et al., *Characteristics of Hookah tobacco smoking sessions and correlates of use frequency among us adults: findings from wave 1 of the Population Assessment of Tobacco and Health (PATH) study.* Nicotine and Tobacco Research, 2018. **20**(6): p. 731-740.

[37] Rodu, B. and N. Plurphanswat, *E-cigarette use among US adults: Population Assessment of Tobacco and Health (PATH) study.* Nicotine and Tobacco Research, 2018. **20**(8): p. 940-948.

[38] Afolalu, E.F., et al., *Prevalence and patterns of tobacco and/or nicotine product use in Japan (2017) after the launch of a heated tobacco product (IQOS®): a cross-sectional study.* F1000Research, 2021. **10**(504): p. 504.

[39] Mazurek, J.M., G. Syamlal, B.A. King, and R.M. Castellan, *Smokeless tobacco use among working adults—United States, 2005 and 2010.* Morbidity and Mortality Weekly Report, 2014. **63**(22): p. 477.

[40] Rodu, B. and N. Plurphanswat, *Mortality among male smokers and smokeless tobacco users in the USA.* Harm Reduction Journal, 2019. **16**(1): p. 1-9.

[41] Cheng, Y.-C., et al., *Patterns of use of smokeless tobacco in US adults, 2013–2014.* American journal of public health, 2017. **107**(9): p. 1508-1514.

[42] Biener, L., et al., *Snus use and rejection in the USA.* Tobacco control, 2016. **25**(4): p. 386-392.

[43] Baig, S.A. and D.P. Giovenco, *Behavioral heterogeneity among cigarette and e-cigarette dual-users and associations with future tobacco use: Findings from the Population Assessment of Tobacco and Health Study.* Addict Behav, 2020. **104**: p. 106263.

[44] Borland, R., et al., *A new classification system for describing concurrent use of nicotine vaping products alongside cigarettes (so-called 'dual use'): findings from the ITC-4 Country Smoking and Vaping wave 1 Survey.* Addiction, 2019. **114 Suppl 1**(Suppl 1): p. 24-34.

[45] Smith, D.M., et al., *Exposure to Nicotine and Toxicants Among Dual Users of Tobacco Cigarettes and E-Cigarettes: Population Assessment of Tobacco and Health (PATH) Study, 2013–2014.* Nicotine & Tobacco Research, 2021. **23**(5): p. 790-797.

[46] Lizhnyak, P.N., et al., *Not All Dual Users of Cigarettes and E-vapor Products Are the Same: Biomarkers and Tobacco Use Behavior in Subpopulations of Dual Users From the PATH Wave 1 Data.* 2021.

[47] Buu, A., et al., *Subtypes of dual users of combustible and electronic cigarettes: longitudinal changes in product use and dependence symptomatology.* Nicotine & Tobacco Research, 2022.

[48] Hughes, J.R., E.N. Peters, and S. Naud, *Relapse to smoking after 1 year of abstinence: a meta-analysis.* Addictive behaviors, 2008. **33**(12): p. 1516-1520.

[49] Herd, N., R. Borland, and A. Hyland, *Predictors of smoking relapse by duration of abstinence: findings from the International Tobacco Control (ITC) Four Country Survey.* Addiction, 2009. **104**(12): p. 2088-2099.

[50] United States Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. 2012.

[51] NSDUH. *National Survey on Drug Use and Health*. [cited 2023 July 16]; Available from: https://nsduhweb.rti.org/respweb/homepage.cfm.

[52] PATH. *The Population Assessment of Tobacco and Health (PATH) Study*. [cited 2023 July 16]; Available from: https://pathstudyinfo.nih.gov.

[53] NHIS. *National health interview survey data.* [cited 2023 July 16]; Available from: https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm.

[54] European Social Survey, *Source Questionnaire*.

[55] Centers for Disease Control and Prevention. *Global Tobacco Control*. [cited 2023 August 27]; Available from: https://www.cdc.gov/tobacco/global/index.htm.

[56] United States Department of Health and Human Services, *Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General*. 2012.

[57] Food and Drug Administration, *Tobacco Product Standard for Menthol in Cigarettes*. 2022.

[58] United States Food and Drug Administration. *Tobacco 21*. 2021 09/01/2021 [cited 2023 July 16]; Available from: https://www.fda.gov/tobacco-products/retail-sales-tobacco-products/tobacco-21#:~:text=On%20Dec.,from%2018%20to%2021%20years.

[59] ESOMAR, *Gender: Best Practice Recommendations for Multi-country Work*. 2023.

[60] Weinberger, A.H., et al., *Assessing Sex, Gender Identity, Sexual Orientation, Race, Ethnicity, Socioeconomic Status, and Mental Health Concerns in Tobacco Use Disorder Treatment Research: Measurement Challenges and Recommendations From a Society for Research on Nicotine and Tobacco Pre-conference Workshop.* Nicotine & Tobacco Research, 2021.

[61] Badgett, M.V., N. Goldberg, K.J. Conron, and G.J. Gates, *Best Practices for Asking Questions about Sexual Orientation on Surveys (SMART)*. 2009.

[62] United States Census Bureau. [cited 2023 August 27]; Available from: https://www.census.gov/programs-surveys/cps.html.

[63] EU_LGBT_Survey_Technical_Report, *EU LGBT Survey - Technical Report*. 2013, European Union Agency for Fundamental Rights.

[64] GLAAD. *GLAAD Media Reference Guide 10th Edition*. 2016 [cited 2023 July 16]; Available from: https://www.glaad.org/reference.

[65] GATS. *Global Adult Tobacco Survey*. 2007 [cited 2023 July 16]; Available from: https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/global-adult-tobacco-survey.

[66] FIWG, *Current Measures of Sexual Orientation and Gender Identity in Federal Surveys*. 2016.

[67] Lin, S.S. and J.L. Kelsey, *Use of race and ethnicity in epidemiologic research: concepts, methodological issues, and suggestions for research.* Epidemiologic reviews, 2000. **22**(2): p. 187-202.

[68] Kaplan, J.B. and T. Bennett, *Use of race and ethnicity in biomedical publication.* Jama, 2003. **289**(20): p. 2709-2716.

[69] Swiss National Science Foundation, *Swiss Household Panel*.

[70] Swiss Federal Statistics Bureau, *Swiss Health Survey*.

[71] Eurobarometer Data Service, *Eurobarometer Surveys*.

[72] EHIS. *European Health Interview Survey*. 2019 [cited 2023 July 16]; Available from: https://ec.europa.eu/eurostat/cache/metadata/en/hlth_det_esms.htm.

[73] Berzofsky, M., H. Smiley-McDonald, A. Moore, and C. Krebs, *Measuring Socioeconomic Status (SES) in the NCVS: Background, Options, and Recommendations*. 2014, United States Department of Justice, Office of Justice Programs.

[74] Harkness, J.A. *Conducting Cross-National and Cross-Cultural Surveys*. in *International Workshop on Comparative Survey Design and Implementation (CSDI)*. 2005.

[75] Hoffmeyer-Zlotnik, J., *Methodological Aspects in Cross-National Research*, J.A. Harkness, Editor. 2005.

[76] Ham, D.C., et al., *Occupation and workplace policies predict smoking behaviors: analysis of national data from the current population survey.* Journal of occupational and environmental medicine/American College of Occupational and Environmental Medicine, 2011. **53**(11): p. 1337.

[77] Centers for Disease Control and Prevention, *Current cigarette smoking prevalence among working adults--United States, 2004-2010.* MMWR. Morbidity and mortality weekly report, 2011. **60**(38): p. 1305-1309.

[78] NHANES. *National Health and Nutrition Examination Survey 2017-2018*. 2017 [cited 2023 July 16]; Available from: https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017.

[79] ISCO. *International Standard Classification of Occupations*. [cited 2023 July 16]; Available from: https://www.ilo.org/public/english/bureau/stat/isco/.

[80] United States Food and Drug Administration. *CFR - Code of Federal Regulations Title 21*. 2021 October 1, 2021 [cited 2023 July 16]; Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=56.111.

[81] Centers for Disease Control and Prevention. *Smoking and Tobacco Use: Tobacco-Related Disparities*. 2021 December 3, 2021 July 16, 2023]; Available from: https://www.cdc.gov/tobacco/disparities/index.htm.

[82] Li, C., et al., *A comparison of prevalence estimates for selected health indicators and chronic diseases or conditions from the Behavioral Risk Factor Surveillance System, the National Health Interview Survey, and the National Health and Nutrition Examination Survey, 2007–2008.* Preventive medicine, 2012. **54**(6): p. 381-387.

[83] Ryan, H., A. Trosclair, and J. Gfroerer, *Adult current smoking: differences in definitions and prevalence estimates—NHIS and NSDUH, 2008.* Journal of environmental and public health, 2012. **2012**.

[84] Levy, D.T., et al., *An Examination of the Variation in Estimates of E-Cigarette Prevalence among US Adults.* International journal of environmental research and public health, 2019. **16**(17): p. 3164.

[85] Amato, M.S., R.G. Boyle, and D. Levy, *E-cigarette use 1 year later in a population-based prospective cohort.* Tobacco Control, 2017. **26**(e2): p. e92-e96.

[86] Cooper, M., M.B. Harrell, and C.L. Perry, *A qualitative approach to understanding real-world electronic cigarette use: implications for measurement and regulation.* 2016.

[87] Kim, H., A.H. Davis, J.L. Dohack, and P.I. Clark, *E-cigarettes use behavior and experience of adults: qualitative research findings to inform e-cigarette use measure development.* Nicotine & Tobacco Research, 2017. **19**(2): p. 190-196.

[88] Liu, S.T., et al., *Qualitative insights on how adult e-cigarette users describe quantity of e-cigarettes used–PATH Study 2018.* Preventive Medicine Reports, 2021: p. 101421.

[89] Nguyen Zarndt, A., et al., *Adult Use of and Transitions From Nicotine and Non-nicotine-Containing E-cigarettes: Data From the Population Assessment of Tobacco and Health (PATH) Study, 2013–2016.* Nicotine and Tobacco Research, 2020. **22**(7): p. 1225-1229.

[90] Addicott, M.S., EL; Reynolds, LM; Donny, EC; Matich, EK; Hsu, PC., *Biochemical validation of self-reported electronic nicotine delivery system and tobacco heaviness of use.* Exp Clin Psychopharmacol., 2023. **31**(3): p. 715-723.

[91] Fischer, K.B., M; Mainy, N; AlMoosawi, S; Sieverding, M; Zwisele, B; Camille, N; Magnani, P; Roulet, S., *Trends in prevalence and patterns of use of a heated tobacco product (IQOS™) in Japan: A three-year repeated cross-sectional study.* F1000Res, 2022. **11**: p. 720.

[92] Glantz, S., *Heated tobacco products: the example of IQOS.* Tob Control, 2018. **27**(Suppl. 1): p. s1-s6.

[93] Villanti, A.C., et al., *Association of flavored tobacco use with tobacco initiation and subsequent use among US youth and adults, 2013-2015.* JAMA network open, 2019. **2**(10): p. e1913804-e1913804.

[94] Rose, S.W., et al., *Flavour types used by youth and adult tobacco users in wave 2 of the Population Assessment of Tobacco and Health (PATH) Study 2014–2015.* Tobacco control, 2020. **29**(4): p. 432-446.

[95] Villanti, A.C., et al., *Changes in the prevalence and correlates of menthol cigarette use in the USA, 2004–2014.* Tobacco control, 2016. **25**(Suppl 2): p. ii14-ii20.

[96] Creamer, M.R., et al., *Tobacco product use and cessation indicators among adults—United States, 2018.* Morbidity and mortality weekly report, 2019. **68**(45): p. 1013.

[97] United States Department of Health Human Services, *The health consequences of smoking—50 years of progress: a report of the Surgeon General*. 2014, Atlanta, GA: US Department of Health and Human Services, Centers for Disease.

[98] Prochaska, J.D., CC, *Stages and processes of self-change of smoking: Toward an integrative model of change.* Journal of Consulting and Clinical Psychology, 1983. **51**(3): p. 390-395.

[99] Hummel, K., et al., *Construct and predictive validity of three measures of intention to quit smoking: Findings from the International Tobacco Control (ITC) Netherlands Survey*. Nicotine and Tobacco Research, 2018. **20**(9): p. 1101-1108.

[100] Kotz, D., J. Brown, and R. West, *Predictive validity of the Motivation To Stop Scale (MTSS): a single-item measure of motivation to stop smoking.* Drug and alcohol dependence, 2013. **128**(1-2): p. 15-19.

[101] Clark, M.A., W. Rakowski, F.J. Kviz, and J.W. Hogan, *Age and stage of readiness for smoking cessation.* The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 1997. **52**(4): p. S212-S221.

[102] Chen, D.G., C; Mechili, EA; Millett, C; Filippidis, FT., *Global Patterns and Prevalence of Dual and Poly-Tobacco Use: A Systematic Review*. Nicotine Tob Res., 2021. **23**(11): p. 1816-1820.

[103] Rubenstein, D.P., LR; McClernon FJ., *Multiple Tobacco Product Use Conceptual Framework: A 2021 Update on Evidence.* Nicotine Tob Res., 2022. **24**(8): p. 1208-1217.

[104] Beatty, P.C., *Cognitive interviewing: the use of cognitive interviews to evaluate ePRO instruments.* , in *Pro: Electronic Solutions for Patient-Reported Data* B. Byrom and B. Tiplady, Editors. 2010, Gower: United Kingdom. p. 2-48.

[105] Office of Management and Budget, *Statistical Policy Directive No. 2: Standards and Guidelines for Statistical Surveys; Addendum: Standards and Guidelines for Cognitive Interviews*. 2016.

[106] Davidov, E., et al., *Measurement equivalence in cross-national research.* Annual Review of Sociology, 2014. **40**: p. 55-75.

[107] Harkness, J.A., *Conducting cross-national and cross-cultural surveys.* Mannheim: ZUMA Nachrichten Spezial, 2006. **12**.

[108] Smith, T.W., *Improving multinational, multiregional, and multicultural (3MC) comparability using the total survey error (TSE) paradigm.* Advances in comparative survey methods, 2018: p. 13-43.

[109] ISCED. *International Standard Classification of Education*. 2011 [cited 2023 July 16]; Available from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf.

[110] Schneider, S., *Guidelines for the measurement of educational attainment in the European Social Survey*. 2020, GESIS – Leibniz Institute for the Social Sciences.

[111] Schneider, S.L., *The classification of education in surveys: a generalized framework for ex-post harmonization.* Qual Quant, 2021.

[112] Fricker, S., B. Kopp, L. Tan, and R. Tourangeau, *A Review of Measurement Error Assessment in a U.S. Household Consumer Expenditure Survey.* Journal of Survey Statistics and Methodology, 2015. **3**(1): p. 67-88.

[113] Lee, E.S., R.N. Forthofer, and R.J. Lorimor, *Analyzing complex survey data*. 1989, Newbury Park, CA: Sage.

[114] AAPOR. *Best Practices for Survey Research*. 2023 [cited 2023 July 16]; Available from: https://aapor.org/standards-and-ethics/best-practices/.

[115] Roberts, C., *Mixing modes of data collection in surveys: A methodological review*. 2007.

[116] Dillman, D.A., J.D. Smyth, and L.M. Christian, *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. 2014: John Wiley & Sons.

[117] Clayton, J., *Best Practices in Survey Design,* O.o.I.R. Planning, Editor. 2021, UW - Whitewater.

[118] Redmiles, E.M., Y.G. Acar, S. Fahl, and M.L. Mazurek. *A Summary of Survey Methodology Best Practices for Security and Privacy Researchers*. 2017.

[119] United Nations, *Designing Household Survey Samples: Practical Guidelines*, in *Studies in methods*, D.o.E.a.S. Affairs, Editor. 2008: New York.

[120] CESSDA Training Team, *Cessda Data Management Expert Guide.* CESSDA ERIC, Bergen, Norway, 2020.

[121] SAMSHA, *Standards and Guidelines for Statistical Surveys*, O.o.M.a. Budget, Editor. 2006.

[122] Groves, R.M., et al., *Survey methodology*. Vol. 561. 2011: John Wiley & Sons.

[123] Heeringa, S.G., B.T. West, and P.A. Berglund, *Applied survey data analysis*. 2017: chapman and hall/CRC.

[124] United States Food and Drug Administration, *E9 Statistical Principles for Clinical Trials*, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, Editors. 1998.

[125] Levy, P.S. and S. Lemeshow, *Sampling of Populations: Methods and Applications*. Fourth ed. 2008, Hoboken: John Wiley & Sons.

[126] Bollen, K.A., et al., *Are survey weights needed? A review of diagnostic tests in regression analysis.* Annual Review of Statistics and Its Application, 2016. **3**: p. 375-392.

[127] Ciol, M.A., et al., *Understanding the use of weights in the analysis of data from multistage surveys.* Archives of physical medicine and rehabilitation, 2006. **87**(2): p. 299-303.

[128] Young, R. and D.R. Johnson. *To Weight or Not to Weight, That is the Question: Survey Weights and Multivariate Analysis*. 2012.

[129] Krewski, D. and J.N. Rao, *Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods.* The Annals of Statistics, 1981: p. 1010-1019.

[130] Wasserman, C.O., Eric, *Department of Health Agency Standards for Reporting Data with Small Numbers*. 2018, Washington State Department of Health.

APPENDIX - SURVEY INFORMATION SUMMARY

APPENDIX TABLE 1: SUMMARY OF EXAMPLE SURVEYS DISCUSSED IN THIS GUIDELINE

Abbreviation	Survey Type	Survey Frequency	Survey Mode	Sample Size	Age in Years	Link to Survey Website
Global						
GATS	Cross-sectional	Annual	CAPI	Varies by country, 380,000 overall total	≥ 15	GATS
ITC	Cross-sectional /Longitudinal	Varies	Varies (CAPI, CATI, web, etc.)	Varies by country	≥ 13 (varies by country)	ITC
Europe						
DEBRA	Cross-sectional	Every 2 months + 6 months follow-up	CAPI (baseline), CATI (follow-up)	2,000	≥ 14	DEBRA
EBS	Cross-sectional	Every 2-3 years	CAPI	~1000*28 countries	≥ 15	EBS
EHIS	Cross-sectional	Every 5 years	varies (CATI, CAPI, CAWI, PAPI)	Varies by country	≥ 15	EHIS
ESS	Cross-sectional	Biennial	CAPI	~1000*28 countries	≥ 15	ESS
HET		Annual	Web-based	20,000	16 to 84	HET
SHP	Longitudinal	Annual panel study (with three waves)	CATI	~29,000	≥14	SHP
SHS	Cross-sectional	Every 5 years	CATI / CAPI (<1 %) / Proxy (4 %), completed by a web-based survey or a paper survey	~22,000	≥ 15	SHS
STS	Cross-sectional	Monthly + 3 months and 6 months follow-up	CAPI (baseline), postal (follow-ups)	~1,800	≥ 16	STS

Abbreviation	Survey Type	Survey Frequency	Survey Mode	Sample Size	Age in Years	Link to Survey Website
Asia						
Japan - NHNS		Annual	In-Person (Health Centers)	18,000		NHNS
U.S.						
NHIS	Cross-sectional	Annual	CAPI	~30,000	≥ 18 (adult interview on tobacco use)	NHIS
NSDUH	Cross-sectional	Annual	ACASI and CAPI	~50,000	≥ 12	NSDUH
PATH	Cross-sectional /Longitudinal	Every 1-2 years	ACASI and CAPI	~30,000	≥ 12	PATH
TUSCPS	Cross-sectional	Every 3-4 years	CAPI/CATI	~240,000	≥ 18	TUSCPS