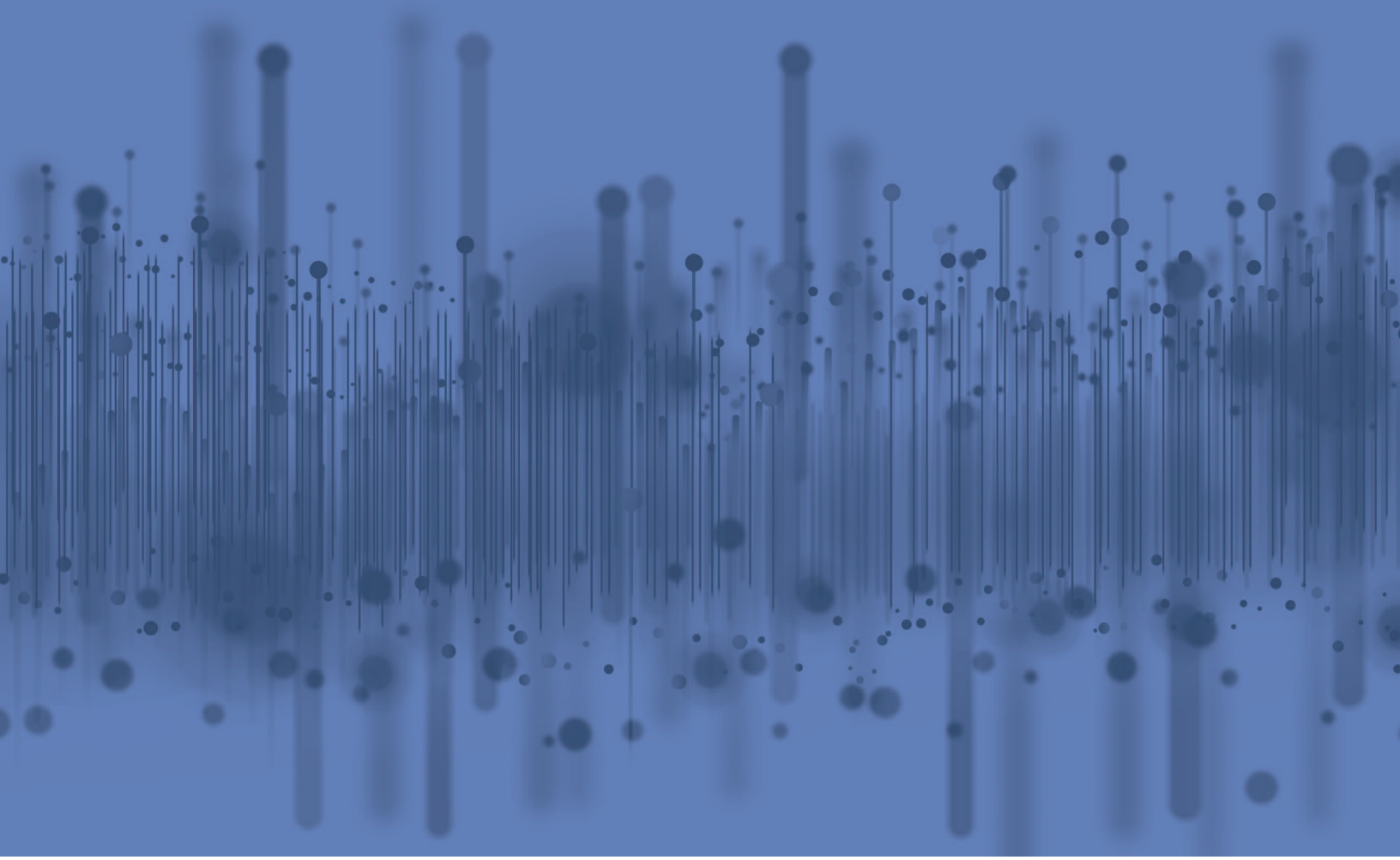
American Cancer Society®

tobacconomics
Economic Research Informing
Tobacco Control Policy

A Toolkit on

Measuring Illicit Trade in Tobacco Products



INSTITUTE FOR
HEALTH RESEARCH
AND POLICY
UIC

Suggested Citation: Stoklosa M., Paraje G., Blecher E., A Toolkit on Measuring Illicit Trade in Tobacco Products. A Tobacconomics and American Cancer Society Toolkit. Chicago, IL: Tobacconomics, Health Policy Center, Institute for Health Research and Policy, University of Illinois at Chicago, 2020. www.tobacconomics.org

Authors: This Toolkit was written by Michal Stoklosa, PhD, American Cancer Society, Atlanta, United States; Guillermo Paraje, PhD, Professor, Business School, Universidad Adolfo Ibáñez, Santiago, Chile; and Evan Blecher, PhD, Senior Economist, Health Policy Center, University of Illinois at Chicago. It was peer-reviewed by Roberto Iglesias, PhD, Technical Officer, Fiscal Policies for Health, World Health Organization, Geneva, Switzerland; and Hana Ross, PhD, Principal Research Officer, Research Unit on the Economics of Excisable Products, University of Cape Town, Cape Town, South Africa.

This Toolkit was funded by Bloomberg Philanthropies.

About Tobacconomics: Tobacconomics is a collaboration of leading researchers who have been studying the economics of tobacco control policy for nearly 30 years. The team is dedicated to helping researchers, advocates, and policy makers access the latest and best research about what's working—or not working—to curb tobacco consumption and its economic impacts. As a program of the University of Illinois at Chicago, Tobacconomics is not affiliated with any tobacco manufacturer. Visit **www.tobacconomics.org** or follow us on Twitter **www.twitter.com/tobacconomics**.

About Economic and Health Policy Research (EHPR): The EHPR program at the American Cancer Society seeks to address cancer worldwide by conducting research on the economic and policy aspects of risk factors to cancer. These risk factors include areas of tobacco, healthy eating and active living (HEAL), and harmful alcohol use. In tobacco control, the team has developed world-leading expertise in economic issues, particularly around fiscal policies such as taxation, trade, and investment. The team has also led the production of the recent editions of The Tobacco Atlas, one of the most influential and most cited resources in tobacco control. Visit the book's companion website: **tobaccoatlas.org**.

Improving Our Toolkit: The Tobacconomics and EHPR teams are committed to making this toolkit as clear and useful as possible. We would like your feedback on whether you found this toolkit useful in your research and, if so, we would appreciate learning about your experience on any successful implementation. We would also like to hear whether you have encountered any issues in applying the methodologies presented in the toolkit and your thoughts on how we could improve it.

For any comments or questions about the toolkit and its content, please email us at info@tobacconomics.org. We very much look forward to hearing from you.

Acknowledgements
The authors would like to thank Dr. Jeffrey Drope, Dr. Liora Sahar, and Dr. Robert Stephens at the American Cancer Society for their input regarding the methods for primary data collection.

Table of Contents

Introduction

1

The most effective way to reduce tobacco use is to increase the price of tobacco products through higher tobacco taxes (IARC 2011 and NCI/WHO 2016). Price increases reduce tobacco use in multiple ways: increased smoking cessation, reductions in smoking intensity among continuing users, and reduced smoking initiation, particularly by youth. Increases in tobacco taxes also generate additional government revenues. However, tobacco tax increases are often opposed by the tobacco industry using the primary argument that taxes are ineffective at reaching the government's policy goals due to increases in illicit trade. The threat of growing illicit trade is also used by the industry as a strategy to oppose other tobacco control measures, including menthol cigarette bans, plain packaging, minimum-pack-size laws, and point-of-sale display bans.

In light of the tobacco industry's use of illicit trade to oppose tobacco tax increases, it is important to understand the true scope and nature of the illicit tobacco trade. Such knowledge is crucial to the formulation of new tobacco control regulations. Illicit trade makes tobacco products more affordable and accessible. The increase in affordability and accessibility leads to greater total tobacco use, which results in higher mortality and morbidity among users. This, in turn, hinders economic growth through higher health care costs and reduced worker productivity. In addition, illicit trade in tobacco products increases crime rates, undermines tax collection, and debases administrative systems and tax policy.

In 2012, parties to the World Health Organizations (WHO) Framework Convention on Tobacco Control (FCTC) addressed the issue of illicit trade in tobacco products at the global level by adopting the Protocol to Eliminate Illicit Trade in Tobacco Products. This new treaty came into effect in 2018. The Protocol provides tools for preventing illicit trade by securing the supply chain (including implementation of tracking and tracing systems), requiring strict law enforcement measures, and strengthening international cooperation.

Along with this Toolkit, Tobacconomics has published a white paper entitled "Illicit Trade in Tobacco Products Need Not Hinder Tobacco Tax Policy Reforms and Increases." The white paper, together with the accompanying policy brief and country case study fact sheets are designed to inform tobacco control researchers, practitioners, and policy makers and to provide a better understanding of illicit trade in tobacco products. The documents critically evaluate the tobacco-industry-led opposition to tobacco tax reforms and conclude that reforms in tobacco tax policy should not be discouraged by the industry argument of an increase in illicit tobacco trade.

Key Messages from the white paper, "Illicit Trade in Tobacco Products Need Not Hinder Tobacco Tax Policy Reforms and Increases":

1. The tobacco industry uses illicit trade to oppose tax increases, specifically arguing that increases in tobacco taxes will lead to increases in illicit trade, which, in turn will undermine public health and fiscal policy objectives.

2. The tobacco industry exaggerates the scale and extent of illicit trade as a means of advocating against tobacco tax increases.

3. Taxes and prices are not the key driver and determinant of illicit trade; many other factors are more likely to drive illicit trade.

4. Even in the presence of illicit trade, experience from a wide range of countries finds that increases in tobacco taxes have consistently produced significant fiscal and health benefits through increased revenue and reduced tobacco use.

5. If governments are concerned about the levels and/or extent of illicit trade, there are many policies, and administrative and enforcement measures that they can undertake to reduce illicit trade, even while increasing tobacco taxes.

Given the tobacco industry's exploitation of illicit trade in the tobacco tax policy space, the need for credible estimates of the level and trends of illicit trade has grown. The illicit tobacco trade is difficult to estimate since the underlying activities are illegal and are often undertaken in an unobservable black market. Vendors hide their illicit products, while users conceal their illegal smoking habits. Due to the unobservable nature of illicit trade, original research, including primary data collection, is often necessary.

Few governments have illicit tobacco trade estimates available. Most governments do not conduct or commission their own studies on illicit trade. Instead, the tobacco industry provides governments with illicit tobacco trade estimates. Unsurprisingly, in most cases, the tobacco-industry-funded studies systematically and strategically overstate illicit trade. Thus, it is necessary to conduct credible, transparent, replicable, and independent studies. Independent or academic estimates that use reliable methods are subject to peer review, and importantly, are not funded by the tobacco industry. Such research is paramount to estimating the level and trends of illicit trade to contrast biased tobacco industry results and provide the public, civil society, and government with a more accurate understanding of illicit trade.

Several methods, using both primary and secondary data, have been developed to estimate the level and/or trends in illicit trade. In a previous Tobacconomics publication, Ross (2015) provided a thorough review of existing methods, including their advantages and disadvantages. Instead of replicating Ross's review, this toolkit provides detailed technical guidance for estimating illicit trade in tobacco products using the two most commonly used approaches. This guidance includes discussion of issues in primary data collection, data analysis, and relevant econometric techniques with statistical codes. Elements of research planning, survey design, sampling, and pack identification are also outlined.

1.1 Purpose

The primary purpose of this toolkit is to guide research on illicit tobacco trade. The toolkit provides step-by-step technical guidance on applying the two most commonly used approaches in estimating illicit trade: first, estimating illicit trade using primary data collection through physical pack collections and smoker surveys; and second, estimating illicit trade using secondary data through gap analysis. This toolkit reviews statistical and economic tools and techniques that can be used to analyze data and discusses the theoretical background and rationale of each method. The toolkit also proposes methods of estimation and the use of the statistical software Stata®.

This toolkit is especially geared towards research for low- and middle-income countries where independent illicit trade estimates may not exist. In many countries, secondary data, collected as part of smoking prevalence or household surveys, can be used to estimate the size of the illicit tobacco market. In addition, cost-effective primary data collection is also available. The primary and secondary data enable a cost-effective, reliable, and independent assessment of the size of the problem.

This toolkit is one of several developed by the World Bank, WHO, and Tobacconomics to provide guidance on conducting economic analysis of tobacco demand and the impacts of tobacco consumption on employment, equity, illicit trade, and economic costs. This Tobacconomics toolkit is the second in the Tobacconomics series designed to build capacity on economic analysis of tobacco taxation. The first toolkit was published in 2019 with the primary focus on using household expenditure surveys to estimate demand for tobacco products (John et al., 2019).

1.2 Who should use this toolkit

The toolkit does not presume knowledge of tobacco taxation, economics of tobacco control, or methods in survey design on the part of the reader. However, a background in economics, econometrics/statistics, and a basic understanding of Stata is required. Armed with the knowledge this toolkit provides, researchers will be prepared to engage in independent studies of the illicit trade in tobacco products through the discussion of methods and step-by-step guides with Stata. Policy makers, government analysts, and civil society organizations can also benefit from the policy discussions, rationale of different economic concepts in tobacco control, and interpretations of results provided. Finally, this toolkit can help all users further their understanding of estimating the scope of the illicit tobacco trade, conduct an independent estimation of illicit trade, and identify and understand problems with estimates provided by the tobacco industry.

1.3 How to use this toolkit

The chapters are structured as follows: Each chapter begins with an introduction and an explanation of foundational principles along with the rationale for engaging in the analysis. A brief technical discussion on data collection and econometric methods follow. The discussion of econometric methods is kept intentionally brief as much of the details are available in standard econometric textbooks and other published sources. Additional references are provided should the reader wish to seek additional information on these theoretical concepts. After the methods section, a short discussion is presented on preparing data for analysis. The discussion includes using Stata as an analysis tool and the relevant Stata code. Case studies from various countries are presented. These case studies demonstrate the methodological variations often required when assessing the size of illicit tobacco market. The toolkit discusses relevant analysis methods for all tobacco products, but the case studies focus predominantly on cigarettes.

It should also be noted that there is no one-size-fits-all approach when it comes to generating estimates of illicit trade. Each country will have its own unique circumstances concerning data availability, composition of the tobacco market, etc. Therefore, apart from drawing knowledge and inspiration from this toolkit, researchers who are novices in measuring the illicit cigarette trade are encouraged to seek advice from more experienced researchers. This toolkit provides an overview of research conducted by some of the most experienced investigators in the field.

The toolkit is organized as follows: Chapter 2 provides a background discussion on the definitions of illicit trade. Illicit trade encompasses both tax avoidance and tax evasion, and within the realm of tax avoidance and tax evasion, there are different types. Specific methodologies may be more appropriate than others to estimate different types of tax avoidance and tax evasion. It is, therefore, imperative for the application of the toolkit that readers begin with a foundational understanding of the various types and attributes of illicit trade.

Chapter 3 discusses the use of primary data collection in estimating illicit tobacco trade. It focuses on two methods. The first method deals with inspection of cigarette packs through littered cigarette pack collection. Data collection from single-stick vendors is also described. The second method describes the smoker pack examination survey method. These two methods provide the most reliable estimates for illicit tobacco trade for two reasons. First, the methods can measure the scope and trends of illicit trade. Second, they can ascertain the nature and origin of illicit trade as well as, in the case of the smoker survey, the characteristics of consumers who purchase these goods.

Chapter 4 reviews the use of secondary data in estimating illicit tobacco trade utilizing the gap analysis method. The gap analysis method is often a cost-effective and quick way to estimate the trend and sometimes scope of illicit trade. There are several issues with this method, however, and each is thoroughly described in the chapter along with suggestions on how to circumvent these issues. When correctly addressed through cross validation using primary data collection, the gap analysis method is a reliable and accurate way of estimating illicit tobacco trade. A step-by-step Stata code for estimating illicit trade using the gap analysis method is also included in this chapter.

Chapter 4 and the Methodology and Code Appendices (Chapter 6) demonstrate specific examples of methods for studies to estimate illicit trade in tobacco products. Chapter 4 provides a step-by-step Stata code for estimating illicit trade using the data gap analysis method, while Chapter 6 provides a sample study instrument and step-by-step Stata code to show the appropriate techniques for estimating illicit trade using the pack examination approach. The code sections in the two chapters build upon the Stata background discussed in the previous toolkit's appendix (John et al., 2019). The individual Stata commands are placed in angle brackets < > and are italicized. This is for illustrative purposes only. The command itself must be used without brackets for the code to run in Stata. The variable names used in the examples are italicized.

Some Descriptions and Definitions

2

Inconsistent definitions of illicit trade create confusion, which can lead practitioners or policymakers to draw incorrect conclusions and negatively affect policy outcomes. The WHO FCTC Protocol to Eliminate Illicit Trade in Tobacco Products (WHO, 2013) defines illicit trade as "any practice or conduct prohibited by law and which relates to production, shipment, receipt, possession, distribution, sale or purchase, including any practice or conduct intended to facilitate such activity." Working from the WHO's broad definition of illicit trade, this toolkit focuses on illicit trade as it pertains to trade with only partial or without full payment of taxes.

Before a discussion on illicit trade definitions, though, it may be helpful to define what is *not* considered illicit trade in tobacco products:

- Domestic production for which all domestic taxes are paid (for example, VAT, general sales tax, tobacco taxes, etc.).
- Foreign production that is legally imported into the domestic market with fully paid import duties (if mandated) and domestic taxes.
- Duty-free purchases within the legal allowance and used only for personal consumption (not to be resold).

In contrast, illegal methods of circumventing tobacco taxes are called tax evasion, as they intend to evade paying some or all tobacco taxes. Tax evasion often occurs due to poor enforcement. There are various types of tax evasion. One of the most common forms is smuggling tobacco products across borders without paying tax in the jurisdiction of consumption. This may also occur when goods are diverted during transit. In many cases, taxes may have even been paid in another jurisdiction, albeit one with lower taxes. While cross-border smuggling is the most common form of illicit trade globally, tax evasion also occurs with domestic production.

Counterfeit cigarettes are cigarettes manufactured without the trademark owner's authorization. The intent is to deceive consumers about the origin of the cigarettes as well as avoid paying taxes. Illicit or "cheap white" cigarettes are brands manufactured in one jurisdiction, often legally in the jurisdiction of manufacture, which are then smuggled and sold in another jurisdiction where applicable duties are not fully paid in the jurisdiction of sale. Unbranded tobacco products are often sold as finely cut, loose tobacco. Unbranded cigarettes are sold in clear plastic bags and called "baggies." Illicit manufacturing may involve the misrepresentation of the quality and origin, failure to obtain a license to grow and produce tobacco, and/or failure to register as an importer/exporter/distributor.

Illicit trade is undertaken by entities that are not properly registered with government agencies or by legitimate entities operating outside of legal and regulatory boundaries. The underlying motivation to engage in illicit tobacco trade is often linked to the size of the illegal operation. Small-scale tax evasion operations usually occur between neighboring countries or at the regional level. This involves moving products across the border in excess of the allowable limits. Bootlegging can also occur when products

purchased “for personal consumption” in one country are sold for profit in another without paying the appropriate taxes (known as “ant smuggling”). Small-scale operations may still pay taxes on their products but in a lower-taxed jurisdiction. Bootlegging generally takes place on a small scale. However, in cases where daily border crossing limits do not exist or borders can be crossed at low cost, bootlegging may contribute significantly to the overall size of illicit tobacco trade. Therefore, when estimating illicit trade, it is important to account for small-scale tax evasion.

Large-scale tax evasion generally involves the nonpayment of all taxes and often occurs at the international level. Individual or corporate greed, money laundering, or financing of criminal activities motivate these actions. Large-scale tax evasion operations can involve counterfeits, genuine products with counterfeit tax stamps, illicit white cigarettes, or domestic production beyond declared amounts. They often take advantage of “in-transit” regimes and/or tax-free zones (Ross, 2015).

Tax avoidance differs from tax evasion in that it refers to legal mechanisms to avoid paying taxes. Tax avoidance takes advantage of poor policy or administration. A prominent form of tax avoidance by individuals is cross-border shopping. Consumers, often individual tobacco users, purchase tobacco products from a lower-tax (or duty-free) jurisdiction within the allowable amount.

Tobacco companies exploit loopholes in legislation through clever accounting practices in order to reduce their overall tax liability (Ross et al., 2017). Cigarette manufacturers engage in forestalling, a practice of producing a significantly larger number of cigarettes before a tax increase takes effect. In this way, manufactures circumvent paying a higher future tax rate (Ross et al., 2017). Manufacturers may also change product attributes to bypass a tax increase. These changes include reducing the weight of tobacco present in a cigarette or changing various characteristics to shift their products to a lower tax bracket.

These types of tax avoidance activities are technically legal. However, they deprive the government of tax revenues and continue to contribute to tobacco product affordability. Hence, these tax avoidance activities undermine public health policies and fiscal measures. Governments should especially pay attention to large-scale tax avoidance schemes because they account for the greatest amount of forgone revenue and contribute the most to the public health burden. In many cases, the government can directly address the issue of tax avoidance through administrative oversight or improved policy design (Ross et al., 2017). Both WHO (2010) and World Bank (2018) have published useful guides to assist countries in addressing the issue of tax avoidance.

Although tax avoidance and tax evasion operate differently, they are often exploited for the same purpose. Both benefit from tax liability reduction and undermine the public health and fiscal policy objectives of tobacco taxation. Throughout this toolkit, careful descriptions are provided for each of the specific types of tax evasion or avoidance that is being measured.

Pack Examination Studies

3

3.1 Introduction

As mentioned previously, it is difficult to obtain a direct measure of illicit tobacco trade due to the inherent secrecy around the purchase and use of illicit tobacco products. However, researchers can circumvent this problem in two ways. The first is by observing product packaging directly. Alternately, researchers can ask smokers about their product's features—rather than asking about illicit cigarette smoking directly. Legal and illegal cigarette packaging often differ by observable characteristics, including pack brand, presence of a proper tax stamp, presence of proper health warnings, missing price or content information, and duty-free markings. Researchers can collect cigarette pack information by evaluating either littered packs or packs in the smokers' possession.

Human consumption and other behaviors are analyzed through garbage in several different social sciences. Trash reveals a lot about consumption, including illicit consumption. Thus, illicit tobacco use is easily observed through garbage analysis. Littered packs are directly pulled from streets or garbage. Similarly, researchers can observe the types of packs smokers have, either by public random observation or privately, in the smoker's household. Due to the stigma associated with illicit cigarette smoking, consumers are unlikely to willingly admit that they use illicit cigarettes. They may not think twice, however, about stating the brand and price of their cigarettes or showing the pack. This information can provide researchers with valuable insight into the prevalence of tax noncompliant packs.

This chapter focuses on two approaches to pack examination: littered packs and smoker surveys. It provides step-by-step guidance for researchers who wish to implement these methods to estimate the prevalence of noncompliant cigarette products. A reference to studies that have used either or both of these methodologies is provided. The chapter refers mostly to estimation of tax noncompliance for cigarettes. Cigarettes are the most common tobacco product, and therefore they are of greatest interest to policy makers. However, the strategies used for cigarettes can also be applied to estimating the scope of illicit trade in other tobacco products (see section 3.4.5).

The chapter discusses each stage of the research. It first describes issues in research planning, including formulating research question and selecting pack characteristics for examination. It then briefly discusses advantages and disadvantages of littered pack collection and smoker pack examination methods to aid researchers in choosing the proper method for their research project. Next, issues in sample selection are thoroughly explained, as there are many aspects in which sampling packs differs from sampling individuals. Finally, the chapter considers issues in questionnaire construction, data collection, and analysis.

3.2 Research planning

Researchers tend to focus their attention heavily on study design and analysis. While these elements of research are undoubtedly of great importance, data collection merits careful planning as well. Particularly when it comes to measuring and estimating the scope of illicit trade, thoughtful data collection is essential to accurate estimates. The covert nature of illicit trade makes accurate estimates even more valuable to policy makers as they develop policies for curbing illicit trade. Moreover, if the goal is to aid policy makers in creating effective policies, then the research should be centered around questions that are of paramount importance to policy makers. Time and financial constraints also need to be considered in the research and data collection process. A well-devised research plan should balance all research needs and ensure careful study design and analysis as well as valuable and accurate estimates for policy makers. These issues are explored further below.

3.2.1 Formulating the research question

A fundamental understanding of smoking habits and illicit trade in tobacco is essential background necessary to formulate a relevant research question and hypothesis. It is imperative that researchers inform themselves prior to engaging in research. A baseline understanding of these issues can be gained by reviewing media reports, academic literature, and government policies related to illicit tobacco trade. Key information to enhance study design includes understanding consumer behavior as it relates to cigarette purchase and consumption—both legal and illegal—as well as understanding the origins of illicit cigarettes. Additionally, it is necessary to understand the target audience for the research—be it policy makers, practitioners, researchers, or other parties—and to adjust the research questions according to the audience.

The information collected during the preparation process greatly determines the research question and method. Several examples of important circumstances that should be considered in the preparation phase include the following: If a large proportion of cigarettes sold is in the form of single sticks, then an inspection of packs from single-stick vendors might be necessary. If cigarettes are mostly sold in soft packs that degrade easily when littered, then a survey of smokers might be preferable over a littered pack inspection. If it is unsafe for enumerators to enter certain areas, then surveying smokers in public places might be preferable to in-home surveys.

During this preparation phase, potential biases need to be considered. This includes accounting for the potential for information biases which can easily sneak into the studies. In this regard, it is important to note that the tobacco industry is the primary provider of information on illicit tobacco trade. Cross-examination of this information shows that the tobacco industry results are often biased (Gilmore et al., 2015). These studies are also rarely independently verified or peer-reviewed (Gallagher et al., 2019). Moreover, the tobacco industry overestimates or falsifies the threat of illicit trade for the strategic purpose of halting the implementation of new tobacco control measures (Gilmore et al., 2015). Therefore, information provided by the industry should be taken with caution due to the conflict of interest. The media should be especially cautious as their audience is far reaching and the provided information not only influences public opinion, but also has public health impacts.

Nevertheless, it is necessary to keep the industry provided reports in mind. When developing a new study design, the pre-existing tobacco research is a key factor to incorporate (e.g., a biased sample). Since tobacco industry reports are generally the only ones available on illicit tobacco trade, they become a benchmark reference. These reports are used by the government, the media, and the general public. Academic research on the illicit tobacco trade must identify the principle flaws in the tobacco industry studies. In this way, the tobacco industry results can be properly addressed and corrected.

Information provided by the government can also be incomplete. For example, customs controls usually have a specific view on the illicit tobacco trade as they interact with illicit tobacco crossing their borders. According to the customs officials, the number of seizures at the border would be the measurement metric for illicit tobacco control. Although valuable information, this measure fails to capture the full extent of the illicit tobacco trade. The information gathering phase also involves clarifying definitions around illicit trade in tobacco products, an issue discussed in the previous chapter.

Once this groundwork has been laid, researchers can move on to defining their research question(s). At this point it is important to consider the target audience of the research. If the study is aimed at policy makers, the research questions will likely focus on the size of the illicit cigarette market or the possible impact of new regulations, such as cigarette taxation. Policy makers would best be served with research on the prevalence of tax avoidance and evasion. Research on other aspects of illicit tobacco trade would best be targeted to other parties. For example, the origin of an illicit cigarette pack or the characteristics of illicit cigarette pack users may be of interest to entities dealing with customs and border control or public health departments.

If researchers can pursue questions freely without grant or funding limitations, they should do so. However, if resources are limited, as is often the case, it may be helpful to discuss the research question(s) with stakeholders. A good research question should be explicit and clearly stated (Kelley et al., 2003). However, stakeholders may demand answers to a wide variety of questions, which may not be feasible within funding, data, or time constraints. Researchers must pull the best possible question(s) from the array; otherwise the study may be weakened. It may not be feasible to answer certain questions because the data do not exist or obtaining the data would be prohibitively expensive. Thus, the researchers' expertise on the subject matter should guide the process. In some cases, it may be sufficient to provide a single national estimate of illicit cigarette trade to refute the industry's claims.

3.2.2 Selecting pack characteristics for examination

Once the research question is established and the primary data collection method is determined, researchers must consider which pack characteristics to focus on. Pack characteristics that aid in distinguishing legal from illegal packs may include pack brand, the presence of a valid tax stamp, health warning, price or content information, duty-free markings, and other features. Each country will have a separate set of pack characteristics, which will be used to distinguish compliant and noncompliant packs. It is important that these definitions are created before data collection begins. During the preparation phase, researchers should familiarize themselves with pack characteristics required in their country and create a protocol for the pack examination process. As demonstrated later in this section, it might not be necessary to collect data on all possible pack characteristics, but limit the focus to the characteristics that best predict pack legality.

In addition, these characteristics may change over time. For this reason, a regular review of laws dealing with characteristics of legal packs is advised throughout the project. For example, the tobacco control bill in Poland provides a set template of warnings that cigarette manufacturers are required to place on all packs. The bill also provides strict guidelines regarding size and placement of warning labels (Sejm of the Republic of Poland, 2019). Additionally, Poland's excise tax bill requires the packs to show excise tax stamps (Sejm of the Republic of Poland, 2008).[1] Reviewing these two bills would be essential for research conducted in Poland, as they detail the appropriate criteria for evaluation of health warnings and tax stamps in Poland and help researchers determine the legality of the cigarette packs (Figure 1).

[1] Template stamps are also provided by the Ministry of Finance of Poland (2019).

Figure 1 Health warning and tax stamp required on cigarette packs in Poland

Source: Sejm of the Republic of Poland (2019); Polish Ministry of Finance (2019)

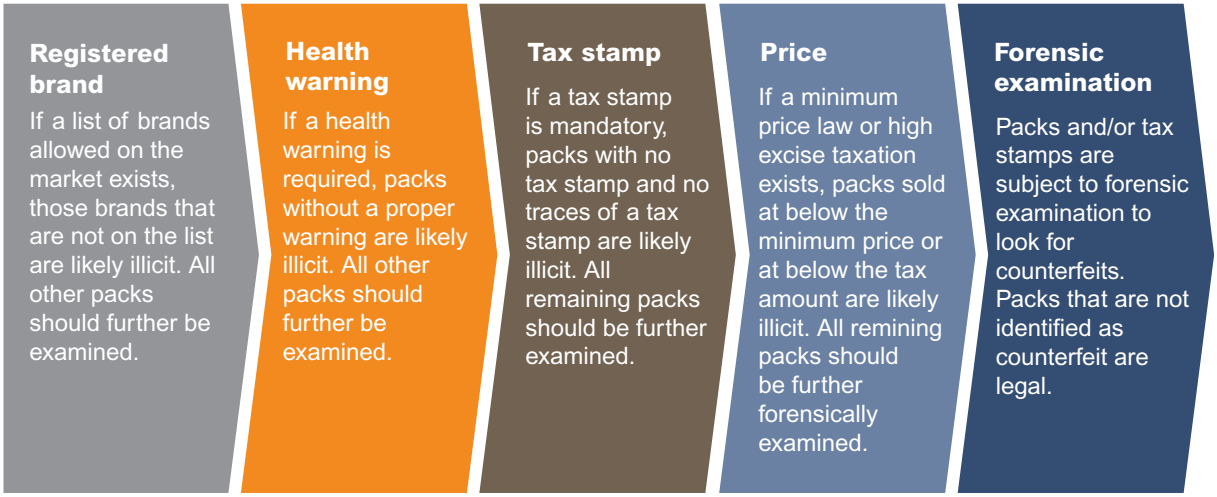
Most pack characteristics are easily visually identifiable. Less-obvious pack characteristics require additional expert attention or equipment. Illicit cigarette manufacturers may place fake or foreign tax stamps on their packs. A forensics expert would be able to determine if the tax stamp is authentic. Alternatively, researchers can return the packs to the stamp-printing authorities, who can evaluate the authenticity of the tax stamp. With specialized equipment, some counterfeit packs can be distinguished through an ultraviolet irradiation and light microscopy scanner that can identify distinctive features. These features include the type of ink used or the package's production method (Kurti et al., 2017). More information on forensic pack examination is provided later in the chapter in section 3.4.4.

Figure 2 shows an example of a pack examination process. The list of pack characteristics for examination presented in this figure is not exhaustive and will likely differ by country. Apart from health warnings, other required pack markings can include a quitline number, unique identification number for the pack's tracking and tracing, as well as the price, tar content, and destination market information. Researchers may encounter several other issues during the pack legality identification process. A few examples are outlined below.

In the United States (U.S.), the majority of states require tax stamps on cigarette packs. However, the tax stamp is placed on a clear cellophane packaging instead of directly on the pack. This makes it difficult to determine if the appropriate taxes have been paid, since most smokers remove the cellophane packaging immediately after opening the pack (Barker et al., 2016). It is highly unlikely to find littered packs that also have an intact cellophane packaging. Therefore, illicit trade research conducted in the U.S. requires additional evaluations of pack features to determine pack legality beyond tax stamps.

Brazil implemented a sophisticated tax stamp tracking and tracing system. It should be straightforward to determine whether the proper Brazilian tax was paid for a given cigarette pack. However, when the pack is opened, the excise tax stamp, which contains the pack's unique identifying number, is torn. Unfortunately, this design makes the use of the tracking and tracing features unsuitable for researchers identifying noncompliant packs, unless the packs are obtained before opened.

Figure 2 Pack examination process



Pack health warnings may not be suitable for identifying legality either. For example, the European Union (EU) Tobacco Product Directive requires packs in all member states to carry both pictorial and text health warnings (European Union, 2014). The EU provides the complete list of possible texts and pictures (European Commission, 2016). Countries sharing the same language have packs that are nearly indistinguishable from one another (for example, Germany and Austria), as these packs display identical health and pictorial warnings. Researchers in those circumstances must pay extra close attention to other pack details (for example, in the case of Germany and Austria, the packs carry different quitline messages) to distinguish between packs.

The above examples serve to illustrate that statistical tests are not inherently incorrect, but the application of them may be. The research must choose the correct statistical test, as these are simply tools. Sensitivity (ability to detect true positive outcomes) and specificity (ability to detect true negative outcomes) are important concepts in the statistician's toolbox. These concepts measure the performance of binary classification. In a test for illegal packs, specificity refers to the test's ability to detect whether a pack is legal, and sensitivity refers to the test's ability to detect whether the pack is illegal.

Intuitively, a tax stamp might be a good indicator of pack legality. However, as mentioned above regarding tax stamps in the U.S., tax stamps can be missing from legal packs (see also Stoklosa and Ross, 2014). During the littered pack collection process, researchers may encounter missing stamps that were either torn off during pack opening or washed away by rain. A test classifying each pack without a tax stamp as an illegal pack would have low specificity. The low specificity comes from mis-specifying all packs as illegal if they do not show a tax stamp, which is not an accurate way of measuring pack legality.

Alternatively, the low cost of illicit cigarettes could be used as marker for legality. Illegal packs are likely significantly cheaper than their legally manufactured counterparts. However, a test classifying only cheap packs as illegal, without taking other pack characteristics into consideration, could miss illegal packs that are sold for a premium price. This test has low sensitivity, since the proportion of correctly identified illicit packs is low.

Application of each test involves trade-offs. Ideally, researchers are able to select a test that yields both high sensitivity and specificity. In practice, however, this is not often the case. If researchers want to estimate illicit cigarette trade as a worst-case scenario and are willing to accept the possibility that some legal packs are mis-specified, then they should use a test with low specificity and high sensitivity. This example highlights the necessity of collecting accurate data, as tests can only be as good as the data they are being tested on.

It should also be noted that pack noncompliance is not always equivalent to tax avoidance or evasion. For example, if a new health warning has been introduced and is now required on all packs in the country, but some old packs with the previous versions of the warning are still in circulation, those old packs may be tax-paid and, at the same time, noncompliant with the health warning requirements. In such case, the results for the health warning compliance should be evaluated and reported separately from the tax laws compliance.

When selecting pack characteristics for examination, it is also important to note that data collection is time consuming and costly. The best survey questions concisely target responses that help in answering the research question(s). Well-crafted survey questions can answer the research question(s) with the minimal amount of available information. It may be enticing to collect as much information from cigarette packs as possible, but not all information on packs may be relevant to answering the research question(s). Collecting irrelevant data can waste valuable resources without adding value to the research.

An illustrative example of this principle comes from Mexico, where the General Law for Tobacco Control in Mexico defines a list of ten pack features that each legal cigarette pack should meet (Sáenz de Miera Juárez et al., 2020). When researchers began an illicit trade study, however, they noticed it would take a prohibitively long time to collect information on all ten pack features. Moreover, increasing the number of collected packs features also increases the probability of misspecification due to pack damage. The researchers collected information on all pack features for a subsample (approximately 10 percent of collected packs). Based on information from that subsample, the researchers determined that three features were sufficient to distinguish legal and illegal cigarette packs with high accuracy. Therefore, for the rest of the sample (approximately 90 percent), data on only three pack features were collected (Sáenz de Miera Juárez et al., 2020). In this way, the researchers significantly decreased research costs by reducing the amount of data that needed to be collected from each cigarette pack without sacrificing effectiveness.

3.2.3 Choosing the research method

Researchers have two main methods at their disposal. Depending on the specific circumstances, researchers may want to employ the smoker pack examination approach or the littered pack collection method or use them both concomitantly. The smoker pack examination approach involves collecting data from the tobacco product pack during a survey of tobacco users. The survey can also ask smokers about their product, such as product price and place of purchase. During a littered pack collection, the data is collected from packs that were previously discarded on the ground in public places. The data collected from the pack can be augmented by information about the localities where the pack was found.

Understanding the advantages and disadvantages of each method is necessary to properly apply the methodologies to the research question. The advantages and disadvantages of both methods are discussed in detail below. Other research methods are also briefly discussed. These are the vendor pack survey, where the packs are obtained from retail sellers, as well as the collection of properly disposed of packs from offices, households, and public trash bins. All four research methods are defined and thoroughly explained in Ross (2015).

3.2.3.1 Advantages of the smoker pack examination survey

Using the smoker pack examination has the advantage of a breadth of available information during the collection phase. Since this method requires direct interaction with the smoker, smoker characteristics are collected at the same time as pack information is obtained. This direct contact feature is absent in the littered pack collection method. Researchers can collect information about the smoker's socioeconomic characteristics, smoking patterns, and price and place of pack purchases. The additional information can aid researchers in distinguishing between tobacco tax avoidance and evasion. The direct observation of smokers also helps researchers determine who illicit cigarette smokers are and where these types of cigarettes are purchased.

A distinct advantage of the smoker pack examination survey method is that researchers can more easily and accurately determine the legality of the cigarette pack. It is often the case that tourists and foreign workers bring cigarettes with them for personal consumption. These imported cigarettes are legal as long as their quantity remains within legal limits. The littered pack collection method cannot directly distinguish between packs imported legally or illegally by tourists. Since the smoker pack examination survey involves direct interaction with the product users, researchers are able to ascertain the origin and price of the pack. Having this information helps researchers distinguish between legal tax avoidance and illegal tax evasion.

The littered pack collection method may be potentially biased, since its method depends on examining packs that were dropped on the ground rather than being properly disposed of. The results from the littered pack collection method are biased when smokers of legal and illicit packs litter at significantly different probabilities. For instance, smokers in some localities may not feel comfortable smoking illicit packs in public. Although studies that used both the smoker pack examination and the littered pack collection did not find significant differences in the littering rate (Merriman, 2010; Stoklosa and Ross, 2014; Sáenz de Miera Juárez et al., 2020), there is a chance for illegal packs to be littered nonrandomly. Alternately, the smoker pack examination survey method allows for all packs to be investigated.

Cultural norms associated with littering should also be considered. If smokers with certain characteristics are more likely to litter, littered pack collection might not represent packs smoked by the general population. This makes the smoker pack examination survey method preferable, as it allows the analysis to control for smoker characteristics, such as age, gender, and socioeconomic status (for example, through stratified sampling as explained later in this chapter). Controlling for such factors with the smoker pack examination survey method can eliminate bias from smoker characteristics that cannot be eliminated from littered pack collection.

Finally, while it is difficult to conduct a survey of littered packs covering the area of a whole country, especially rural areas, the implementation of a nationally representative survey of smokers is still feasible. A prominent example of a nationally representative survey that incorporates cigarette pack examination is the Global Adult Tobacco Survey (for example, in Poland and Ukraine) (CDC, 2019).

3.2.3.2 Advantages of the littered pack collection method

The littered pack collection method generates an objective estimate of illicit trade, which is one of its main advantages. Unlike smoker pack examination surveys, discarded pack collections are free from problems associated with self-reporting. Smokers may be unwilling to disclose their smoking habits either due to social or cultural norms. Smokers may also be unwilling to admit that they smoke illicit cigarettes, or they may not want to show their cigarette pack. In this way, the littered pack collection method has the potential to report more accurate information on smoking habits.

The littered pack examination method is also easy to implement. Even in clean cities, pack collectors usually do not have problems with collecting the desired amount of discarded packs. Littered pack collection studies do not involve human subjects, and therefore a lengthy review by an ethics committee is not required. If time is of the essence—such as when researchers want to measure illicit trade pre- and post- a tobacco control intervention or a tobacco excise tax increase—the littered pack collection method allows for greater flexibility with study implementation. Even in the (highly unlikely) event that an ethics review is required, the review should be straightforward. Some countries may require an ethics review for the purpose of assessing the risks to which researchers could be exposed.

Another advantage with littered pack examinations is that researchers are able to spend as much time on a discarded pack as is necessary to collect the desired data. Littered packs can be collected and sent to a research facility with the proper equipment and resources to evaluate pack legality. For example, an expert forensic investigation to determine counterfeit packs (see section 3.4.4 for a description on how to perform such a test) would be impossible without a physical pack. Moreover, the collected packs can be stored and revisited at a later time. This is particularly helpful because the same packs can be reused for a subsequent study that may be examining different pack characteristics. The ability to store the collected packs may also increase the transparency of the study as packs can be reexamined when proof of results is requested. Although it is possible to purchase or request packs from smokers in a smoker pack inspection, not all smokers would agree to sell or give their packs to the enumerators. Moreover, even though it may be possible to take a photograph of the pack during the smoker survey, these photographs may not be sufficient if additional tests are needed.

In addition to being more time sensitive, littered pack collection studies are also less costly than smoker survey pack inspections. In the pack examination survey, skilled survey enumerators are required. Their required skills include the ability to approach an individual and to establish rapport, to read and speak clearly in the native language, and to carefully listen and remain neutral throughout the interview. These enumerators need training in the use of the respondent booklet, visual materials used in the interview, and—in the case of computer-assisted personal interviewing (CAPI) or other tools—training to use the survey software. Market research companies with existing teams of trained enumerators can charge premium prices.

The littered pack collection method, on the other hand, does not require the same skill level. The skills required of pack collectors are limited to the ability to walk along a designated path, collect the littered packs, and record the location of the collection, or sort through garbage and record its origin. Therefore, the services of market research companies are often not needed for the discarded pack collection method. Enumerators for this type of study can be recruited among lower-skilled laborers at a lower cost. For example, in some previous studies, university students were recruited to collect littered packs.

3.2.3.3 Other pack examination methods

There are other pack examination methods available to researchers, which provide helpful cross-validation of the results from the smoker survey and littered pack collection methods and reveal more information on smoker behavior. The first alternative method is referred to as the vendor pack survey. In many countries, a large percentage of cigarettes are sold in the form of single sticks. Single-stick cigarette smokers do not have a pack for researchers to examine. If single-stick cigarette vendors litter their cigarette packs at the same rate as pack-purchasing smokers, then a littered pack examination is suitable to estimate illicit cigarette prevalence in these countries. However, this is not always the case. Since stalls are set up in the morning, and—in some countries—vendors remain at the same location throughout the day, vendors frequently have a proper disposal system for their used packs. In such cases, the littered pack collection method would miss the properly disposed packs from single-stick vendors.

In the vendor pack survey, the vendor of loose cigarettes sets aside emptied cigarette packs for collection by enumerators at the end of the day. The vendor pack survey was first implemented in India, where 55 percent of cigarettes are sold as loose cigarettes (John and Ross, 2018). In this survey, in exchange for a small monetary reward, single-stick vendors along predetermined routes in eight Indian cities deposited empty cigarette packs into provided bags, which were collected at the end of the day.

Although direct consumer engagement is not part of this type of study, enumerators may be able to collect information from vendors about cigarette brands, quantities, and prices. However, it is important to remember that information collected from vendors might be subject to bias related to the vendor's willingness to respond and the truthfulness of those responses. If selling illegal cigarettes carries a penalty or is a crime, the vendor could be especially reluctant to give their empty packs to the enumerator or to answer the enumerator's questions. For this reason, information provided by the vendors on cigarette brands, price, and quantity is secondary, and serves only as a complement or a robustness check to the primary goal of collecting cigarette packs from the vendors. The vendor-provided information should not be used to inform the prevalence of illicit packs prior to pack examination, because the share of illicit packs among the packs disposed of by the vendor provides the most accurate information. Finally, as the vendor could be reluctant to provide the disposed packs to the enumerators, researchers should consider incentivizing survey participation. It is important that the monetary reward be small and uniform. In other words, the monetary reward cannot be based on number of packs collected, as it may incentivize vendors to add extra packs. This would substantially and nonrandomly bias the results.

Although the vendor pack survey method is not the focus of this toolkit, several aspects of the study design described later in this chapter directly apply to this method as well. Due to the similarity between this method and the littered pack collection method, the study planning, execution, and data analysis are directly related to the vendor pack disposal survey. For example, in both methods, the enumerators must follow predetermined routes, which are selected based on where smokers reside and purchase their cigarettes. The main difference is that in the littered pack collection method all littered packs are collected along the route, while in the vendor pack survey, the packs are collected from vendors along the route.

Combining the littered pack collection and the vendor pack survey methods provides a more complete view of overall smoker behavior. Single-stick consumers are different from pack-purchasing consumers, either because of frequency of consumption or income level. If 30 percent of cigarettes smoked are from single-stick vendors while 70 percent are purchased in packs, then the illicit trade prevalence can be estimated by calculating the weighted average of illicit cigarette prevalence from both sources.

The final method discussed in this chapter is the properly disposed pack collection method. It is possible that properly disposed packs capture a different consumer than the littered pack collection method, and thus, makes the two methods complementary. Differentiating between the two may provide some information about the types of smokers. The littered pack collection method collects the packs outside, after the smoker has finished their pack. This means that littered pack collection misses people who predominantly smoke—and dispose of their packs—inside of residences, offices, or other establishments. To address this oversight, researchers can focus on collecting appropriately disposed packs found in trash containers and receptacles located near workplaces and residences. The method has been used in several studies, including in Chicago (Merriman, 2010), New York City (Consroe et al., 2016), and Paris (Lakhdar, 2008).

One drawback of the properly disposed pack collection method is the difficulty in generating a representative sample. For the sample to be representative, trash needs to be collected from a wide range of places. Although theoretically easy, in practice this is difficult to achieve. Trash collection companies differ for public places and residences, and several different companies may operate in the same area covering different sectors. This makes obtaining a representative sample, or a biased sample for which researchers can explain the bias, challenging. Collaboration with trash collection groups is needed, since they generally have exclusive access to trash bins. When working with these companies, researchers intercept the trash at the transfer station before the trash goes into the landfill. Larger trash collection companies collect larger quantities of trash, which further complicates matters as the precise origin of the trash is unknown.

Choosing the properly disposed collection method means having to sort through the collected trash since it is generally not presorted. Proper storage of the trash, such as coolers and freezers would also be required as sorting through trash bags is time consuming. Since trash bags also contain perishables, the organic material will begin to break down rather quickly. Minimizing trash odors is necessary for the comfort of the enumerators. Several other precautions need to be taken as the enumerators may encounter potentially dangerous or hazardous materials such as medical needles and sharp objects during the sorting process.

The properly disposed pack collection method is the most involved pack examination approach, and it is questionable if this method is more accurate than others. The main benefit of using this method is to establish whether the characteristics of those who litter are similar to those who do not. The high cost and the unknown value gained from this method do not make it viable, and therefore it is not discussed further in this toolkit, but some parts of this toolkit are also applicable to the properly disposed pack collection method. These include the sections on pack characteristics, designing a sample of geographical areas, and data collection from the packs.

Researchers conducting studies using these methods should keep their strengths and weaknesses in mind. As mentioned above, if possible, researchers should use the other method(s) to cross-validate their findings for the most robust and precise approach. Cross-validation is useful because each method captures aspects of illicit trade that other methods cannot. For example, the smoker pack survey data collection process is more expensive than the littered pack collection method. Used in conjunction, however, the littered pack collection method can be used on a wide geographical area, while the smoker pack examination survey can be implemented in a few subsections. This way, the benefits of both methods can be used and the downsides of either method are mitigated through cross-validation.

This mixed approach was used recently by researchers in Brazil. Littered packs were collected in five cities (Belo Horizonte, Campo Grande, Joao Pessoa, Rio de Janeiro, and Sao Paulo), and the smoker pack examination survey was conducted in two of those cities (Rio de Janeiro and Sao Paulo). Additionally,

researchers conducted a properly disposed pack survey from residence garbage in Rio de Janeiro. Together, the researchers were able to estimate robust results (Carvalho Figueiredo et al., forthcoming).

3.3 Sampling

Sample design is an important aspect of the study design and requires some additional commentary. Researchers with previous survey experience are likely to feel comfortable with survey execution and data analysis, but the sampling methods are relatively different between the two main methods presented above, and this may require clarification.

A sample is defined as a subset of a population. Therefore, before considering a sample, researchers need to clearly define the study population, which is discussed below.

3.3.1 Study population

Researchers should explicitly define the study population and then consider how the sample fits within it. When the study population is poorly defined, the sample is unlikely to be representative of the true population (Sudman, 1983). Defining the study population in illicit trade surveys especially requires careful planning and is essential to arriving at accurate estimates. This section includes important points to consider when defining the study population in illicit trade surveys.

First, an important fact to remember in defining the study population is that adults have differing smoking rates. In most places, adults with higher incomes tend to have lower smoking rates (Drope et al., 2018). Therefore, a representative sample takes into consideration the distribution of adult smokers instead of polling subjects uniformly. Polling uniformly will create an overrepresentation of people in higher socioeconomic brackets, who historically have lower smoking rates. Researchers designing illicit tobacco study samples should focus on the typical smoker and where smokers live and work.

Another important point to consider is that the ultimate focus of the study is to gather information on cigarettes. Smokers are the source of this information. Smokers are surveyed to provide information on their smoking habits to estimate the illicit tobacco trade. The smoked cigarettes constitute the study population, not smokers. This can be an important distinction when smoking behavior differs between illicit cigarette smokers and regular smokers. It is possible, for example, that smokers of illicit cigarettes smoke more cigarettes per day than the average smoker. Estimates of the prevalence of illicit cigarettes need to take these smoking behavior differences into consideration.

In addition, when conducting research on the illicit tobacco trade, researchers should keep in mind that the study population is all cigarettes that are purchased and consumed, not only illicit cigarettes. Frequently, illicit trade researchers tend to focus exclusively on illicit cigarettes and often incorrectly focus on collecting data in areas known to be high traffic areas for illicit cigarettes. While sampling hot spots can be a part of a study design, collecting data in this fashion would overestimate the proportion of illicit cigarettes. Publishing research conducted in this fashion will incorrectly inflate estimates of illicit cigarette consumption.

The next point to consider is that illicit trade studies should focus on current cigarette consumption. Past behavior is not necessarily informative about current behavior. Ideally, the survey focus should be on the last-purchased or last-consumed pack of cigarettes. Questions about the past should not be used unless the information about the year and month of that behavior is collected. The ambiguity introduces bias if these questions are not time specific. Ambiguously stated questions will collect inaccurate data. When asking smokers about past smoking behavior, questions need to be asked very carefully. For example, the Eurobarometer survey asks smokers if they have been offered black market cigarettes. Possible answers

are the following: rarely (less than one a month); occasionally (once to three times per month); regularly (once per week or more frequently); or never (European Commission, 2016). To answer this question truthfully, a person who was offered a pack of illicit cigarettes a decade ago cannot choose the "never" option on the questionnaire. Therefore, the Eurobarometer survey is not properly designed. Similarly, some surveys ask if smokers have ever purchased illicit cigarettes. That is also not precise and such questions should be avoided.

Narrowing down the appropriate study population also requires careful consideration. As mentioned above, illicit tobacco trade estimates should focus on all cigarettes being purchased and consumed. Studies of illicit cigarette availability, however, focus on a different population than studies estimating the level of illicit trade of cigarettes. An incorrectly defined study population will lead to biased results. In the abovementioned Eurobarometer survey, the study aim was to measure black market penetration (European Commission, 2016). The research focused on measuring the prevalence of being offered an illicit cigarette. The Eurobarometer survey found that Luxembourg (94 percent of respondents declared never being offered illicit cigarettes) ranked highest in terms of legal tobacco trade, and Lithuania and Bulgaria (both 64 percent) ranked lowest. The report concludes that "geographically, black market cigarettes have larger penetration in Eastern European countries than elsewhere."

The survey likely yielded biased responses, driven by poorly defined study population parameters. The probability of being offered an illicit cigarette is generally different in local markets (bazaars) versus chain supermarkets. Additionally, food shopping behavior may be markedly different between countries and therefore may be a confounding factor. The prevalence of shopping in chain supermarkets is likely higher in Luxembourg while the prevalence of shopping in local markets in Lithuania and Bulgaria is likely higher. If the probability of illicit cigarette offering is higher in local markets, then the study will be biased if the data are not appropriately weighted to account for this difference. As a result, the data obtained do not necessarily confirm that black market penetration is higher in Eastern European countries.

3.3.2 Geography

The next research planning stage is selecting the study area to be covered. The goal of each cigarette pack study is to produce a representative sample of smoked cigarette packs within a larger administrative unit (for example, city, province, or country). Funding for independent research studies is usually more limited. The tighter budget constraints often make sampling the entire country infeasible. Pack examination surveys are rarely nationally representative (rare exceptions include Fix et al., 2014) or have national coverage (for example, Barker et al., 2016; Little et al., 2019).

A nationally representative littered pack collection is difficult to implement. For this method to be nationally representative, the researchers would have to collect packs from several different areas of the country based on a carefully designed sampling framework that would assure representativeness of the sample. This would include collections in cities, where population is denser, as well as rural areas. The likelihood of finding littered packs in less-densely populated areas decreases significantly as fewer people are spread across a larger area. To cover rural areas appropriately, study costs would increase drastically as more labor hours are required to cover these areas.

Instead of aiming for a nationally representative sample, littered pack collection studies are usually limited to a number of cities. Although this limits the scope of the study, as long as representative samples are collected from the cities, the study remains valid. Additional requirements for littered pack collection studies include that the cities must be spread across the country; there must be a sufficient number of people living in each city; and the tobacco users of each city should mirror the average users in the country as a whole.

The estimates from such a study provide an unbiased view of illicit tobacco trade in the country, especially if the urban population studied represents a large portion of the total population.

For example, a recent study from Colombia estimates the level of tobacco tax avoidance and evasion. The researchers utilized the smoker pack examination method in several cities to obtain a nationally representative sample (Maldonado et al., 2018). The survey was conducted in five Colombian cities, representing 63 percent of the country's cigarette market. The sample was cleverly designed to match the age and sex distribution of smokers from a different nationally representative survey. Another example is a Mexican study that used a similar approach (Sáenz de Miera Juárez et al., 2020). Both the smoker pack examination survey and the littered pack collection were conducted in eight cities across the country. Additionally, another vendor pack survey in India collected empty packs from single-stick vendors in eight cities (John and Ross, 2018). Both the Mexican and the Indian studies covered a large share of the total population and had a vast geographic spread.

3.3.3 Age and other characteristics of individuals

Another factor to consider is that researchers must establish a minimum cutoff age for survey subjects. This is not a concern for the littered pack collection method, because it involves collecting packs discarded by all smokers (since it is impossible to discern smoker characteristics from packs alone). Parental consent is usually required to survey minors because it is not ethically appropriate to inquire about smoking behaviors of minors. When a legal smoking age is set (usually 18), the cutoff age for the survey should abide by the same age.

A smoker's residency status requires a cutoff value as well. Researchers may want to identify tourists and temporary workers in their sample, either to exclude them or to measure a different research aspect. Tourists and temporary workers are not part of the country's general population. A similar consideration should be given to people who live outside of the defined city or jurisdiction limits. If a substantial percentage of workers commute into the studied city, then appropriate parameters should be set for this kind of city.

3.3.4 Sample size

The next step in the sampling process is to establish the sample size. Often the type of study question will determine the required sample size (Sudman, 1983). There is no hard and fast rule on the appropriate sample size for illicit tobacco trade studies. For an overview study, estimates do not have to be particularly accurate since these studies report descriptive statistics. This kind of study serves as an exploratory study and essentially provides a first glimpse into the issue. For studies providing more in-depth analysis, including statistical inference, larger and more carefully selected samples are needed. Studies require larger, representative samples if they intend to test whether their estimates are statistically different from the tobacco industry's estimates. Similarly, city-level estimates of tobacco avoidance and evasion also require a sufficiently large sample. A measure of tax noncompliance may require an even larger sample to appropriately measure the levels across different cities.

Sample attrition also needs to be considered when setting the sample size. Low response rates in smoker pack surveys are a source of attrition. Smokers may be unable or feel uncomfortable answering questions on their smoking habits or producing their packs for examination. A preliminary small-scale pilot study gives researchers a general idea of the sample attrition rate. Based on this information, researchers can get a better sense of the required sample size as well as the amount of time needed to complete each survey.

A pilot study is also important for the littered pack collection method. The pilot study gives researchers a general idea of how many packs can be collected in each area. Since there is large variance in population density, smoking prevalence, and pack discarding behavior, the number of packs discarded may also have large variance. In a littered pack collection study in Chicago, for example, collectors needed to walk roughly 135 miles (217 km) (Merriman, 2010) to collect the required sample. A similar study in Mexico only required collectors to walk around 52 miles (83 km) per city (Sáenz de Miera Juárez, 2020). Surprisingly, both studies collected nearly the same number of packs per city (around 1,000). A small-scale, pilot study gave the researchers the necessary information to set the required distance to collect the appropriate number of packs.

The following statistical formula helps to estimate the minimum sample size for a study, given the desired precision level:

$$\hat{n} = \frac{Z^2_{1-\alpha/2} \cdot P(1-P) \cdot DEFF}{\delta^2}$$

where n is the sample size, Z is the Z statistic for a given level of confidence (for example, the critical Z score is $Z_{1-0.05/2} = 1.96$ for the standard 95 percent confidence level), P is the expected prevalence of illicit packs (for example, P = 0.2 for an expected prevalence of 20 percent), DEFF is a design effect accounting for the complex sample design (see below), and δ is the assumed precision. The assumed precision is also called the margin of error (for example, δ = 0.05 if the assumed margin of error is 5 percent). This n is the minimum sample size needed for accurate and reliable inference. The formula for n assumes an infinite (very large) population. Thus, for large populations, the sample size does not depend on the size of the study population. A smaller n will not have the necessary precision for accurate inference testing at the given statistical confidence level. Although larger samples will not hurt the study, the study costs will unnecessarily increase without additional benefit.

Several details about this formula are worth noting. First, the desired sample size depends on illicit cigarette prevalence (P). The pilot study will give researchers a rough estimate for the size of P. Mathematically, the value of P(1 - P) is greatest when P = 0.5. The sample size will be largest when P = 0.5 (half of the packs are illicit). Second, the DEFF adjustment factor is expressed as the ratio of the variance for the given sampling design divided by the variance of the effective sample of the same size. The most effective sample is obtained through random sampling. For simple random samples, the numerator of the DEFF ratio is equal to its denominator. Thus, DEFF is equal to one. For more complicated sample structures, such as those described in the next sections of this chapter, the variance in the numerator of the DEFF ratio is different from the variance in the denominator, so the DEFF factor will not equal one.

A downside to the DEFF factor is that calculating it requires the sample variance. However, the sample variance cannot be determined until the study has been performed. This is impractical for many research projects. A rule of thumb is often used when selecting the DEFF factor. For cluster sampling, which is a sampling design that is often used in pack examination studies, the design effect of 2 or greater is usually assumed when no estimates of DEFF are available (Bostoen and Chalabi, 2006; Maas and Hox, 2005).

Mexican researchers implementing the littered pack collection study used the above formula to determine the appropriate sample size. The study assumed: Z = 1.96 (the critical Z score value for the 95 percent confidence level), P = 0.02 (based on the expectation that 2 percent of packs are illicit), DEFF = 2 (allowing for variability in the number of collected packs per sampling unit and the prevalence of illicit packs in each sampling unit), and δ = 0.015 for the assumed margin of error at 1.5 percent, yielding the sample size of

670 packs per city. The pilot study determined that collectors should find about 8 packs per kilometer. Thus, the minimum travel distance required to collect the necessary amount is 52 miles (83 km) per city (Sáenz de Miera Juárez, 2020). The goal of the Mexican study was to select a sample size that would accurately represent the illicit cigarette market at the city level. If the researchers had wanted to compare illicit cigarette trade levels across city districts, appropriate district-level sample sizes would also have had to be chosen. When researchers want to perform subpopulation studies (for example, by ethnic group or gender), the formula would have to be used for each subpopulation. The sample size would need to be inflated for stratified sample testing.

A sample size can also be taken from a previously conducted study. While using the formula to determine sample size provides researchers with a precise sample number, adopting a sample size from a previous study does not. In studies with national coverage, the sample size is typically larger than 2,000 (for example, Barker et al., 2016; Little et al., 2019) and sometimes as large as 8,000 (Sáenz de Miera Juárez et al., 2020). The city-level estimates typically each have about 1,000 observations (for example, Merriman, 2010; Sáenz de Miera Juárez et al., 2020). Again, researchers who wish to obtain estimates at the sub-city or sub-sample levels must increase the sample size.

3.3.5 Sample design

Sample design defines the set of rules that specify how to select the sample. Survey research, such as smoker pack examination studies, usually uses probability sampling design. The underlying principle in probability sampling is that selection of each element in the population has a non-zero probability. Simple random samples are the most basic type of samples. Each element of the population has equal probability of being selected into the sample. Stratification divides the sample into mutually exclusive and exhaustive subpopulations (for example, male/female, child/adult, etc.) called strata. Samples are drawn from each subgroup. Finally, cluster sampling breaks the population into clusters, from which a sample of clusters is selected. The following section presents each of these sampling techniques in the context of the pack examination studies.

3.3.5.1 Simple random sampling

Simple random sampling is the most straightforward design. An attribute of this method is the low sample variance. In practice, simple random sampling is challenging to implement in large population surveys. A key feature of the simple random sample is the equal probability of each observation to be selected into the sample. Since people are not uniformly distributed across the country and cigarette packs are not uniformly distributed among people, this criterion is difficult to meet. A random sample is unlikely when people are approached on the street. The time of day determines what type of people can be polled. Early in the morning, people are on their way to work. Later in the day, researchers may encounter a higher proportion of university students who are not tied to a strict nine-to-five schedule. This type of approach may also completely miss people who rarely leave their homes.

The proper implementation of pure random sampling would require researchers to possess a full list of population elements (population frame) from which the sample is randomly drawn. This information allows researchers to uniformly sample across the population elements and therefore satisfy the uniform probability criterion. However, in the case of smoker pack examination surveys, it is highly unlikely that a full list of smokers exists.

In the unlikely event that a complete list of households with at least one smoker exists, it still may not be feasible to implement a simple random sample. Since each selected household would have to be visited individually, visiting all randomly selected households could be logistically challenging or costly, especially if

they are spread across large geographical areas. The attrition rate may also be high because residents may not be present at the time visited. In addition, tobacco use behaviors may vary among smokers.

Simple random sampling is easy to implement in theory, but to adequately satisfy the criterion is challenging in practice. Researchers conducting survey studies tend to resort to other sample-selection methods: stratification and clustering.

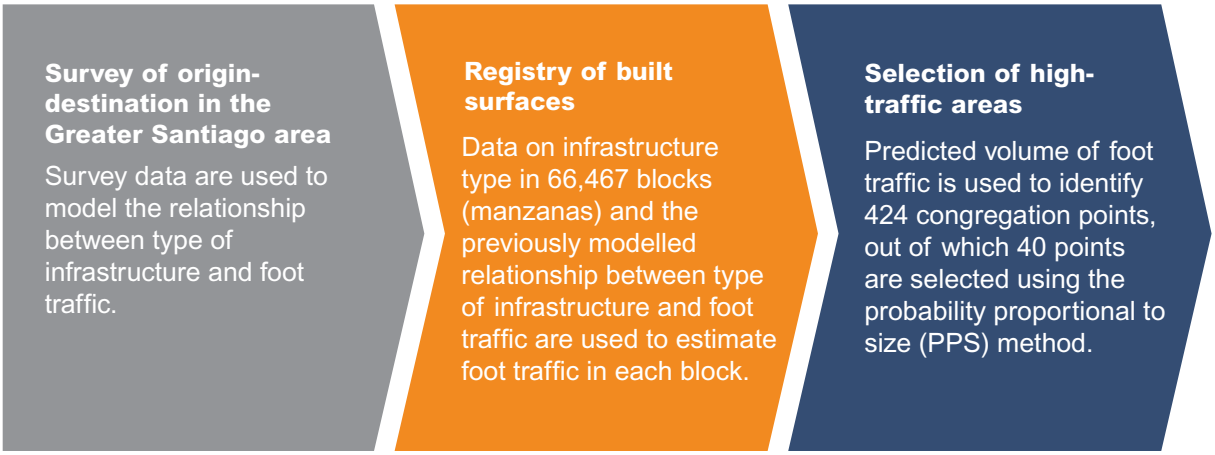
3.3.5.2 Stratified sampling

The basic notion of stratified sampling is to divide the entire population into separate and distinct subpopulations (strata). Independent samples are drawn within each stratum (Frankel, 1983). Stratified sampling ensures that the key groups are appropriately represented in the sample. Additionally, researchers may want to study certain population subgroups. The sample size for that strata can be increased to ensure sufficient sample size for the separate subgroup analysis. In the full sample, the larger strata can be properly weighted within the total sample (Frankel, 1983). The larger strata do not need to be altered when using weights.

Stratified sampling is particularly useful for estimation of illicit cigarettes using the smoker pack examination survey method. Tobacco use surveys—such as the Global Adult Tobacco Survey (GATS) (CDC, 2019) or other national surveys—provide information on age, gender, and socioeconomic characteristics of smokers. The information collected by GATS can be used to inform the strata in other research projects. For example, if women aged 60 and older represent five percent of all smokers in the country, then the researchers can use five percent for this age-gender subgroup in their stratification.

The International Tobacco Control Policy Evaluation Project (ITC Project) has some of the best examples of stratified sampling in smoker surveys (ITC, 2017). The ITC Project surveys—several of which are nationally representative—conveniently stratify the population into tobacco users and non-users. The sample sizes for both users and non-users are representative, so that when they are combined the survey analysis yields precise estimations for the entire country population. However, the tobacco-user stratum is considerably

Figure 3 Sampling of congregation points using predicted foot traffic in Santiago de Chile



increased (oversampled). This allows researchers to perform additional analyses with greater precision on this subgroup. Due to the size of the smoker stratum, some ITC surveys are able to break down the stratum even further. The subgroup is categorized by cigarette users, users of other tobacco products, as well as dual/mixed users (that is, more than one tobacco product). Analysis can be performed on an even more granular level. The survey-provided weights can be used to combine differently sized strata in order for analysis to be performed on the total sample (ITC, 2017).

Stratified sampling has two important caveats. First, although the method ensures that subgroups are properly represented in the sample, each subgroup (stratum) still needs to be properly sampled. For example, if ten women aged 60 and older are to be sampled, but all of them are sampled in the same location, then the sample will likely be biased. People in the same location are likely to share similar characteristics (for example, income, education, marital status, etc.). To preserve randomness within the sample, it should also be geographically controlled. Secondly, researchers should be aware of potential issues with a low response rate and stratification. It would not be statistically appropriate—and thus bias the sample—to approach 100 women aged 60 and older to collect information on only 10 of them due to attrition. Usually, with low cooperation, there are also significant biases in subject characteristics (for example, smokers with higher income are less likely to show their packs) (Sudman, 1983). Therefore, in places with a low expected response rate to the smoker pack examination survey, the littered pack collection method may be a more appropriate way to collect data.

An illicit trade study in Santiago de Chile employed a geographically controlled stratified sample (Paraje et al., 2018). The study used estimates from a previous national survey on drug use to stratify their sample by age and gender. The researchers collected data in high-traffic areas. The study used a sophisticated process to select these areas (Figure 3). Section 6.1 in the appendix provides more information on the congregation point selection. Figure 4 presents the geographical distribution of the 424 congregation points (red dots and blue stars) as well as the 40 congregation points selected for the sample (blue stars).

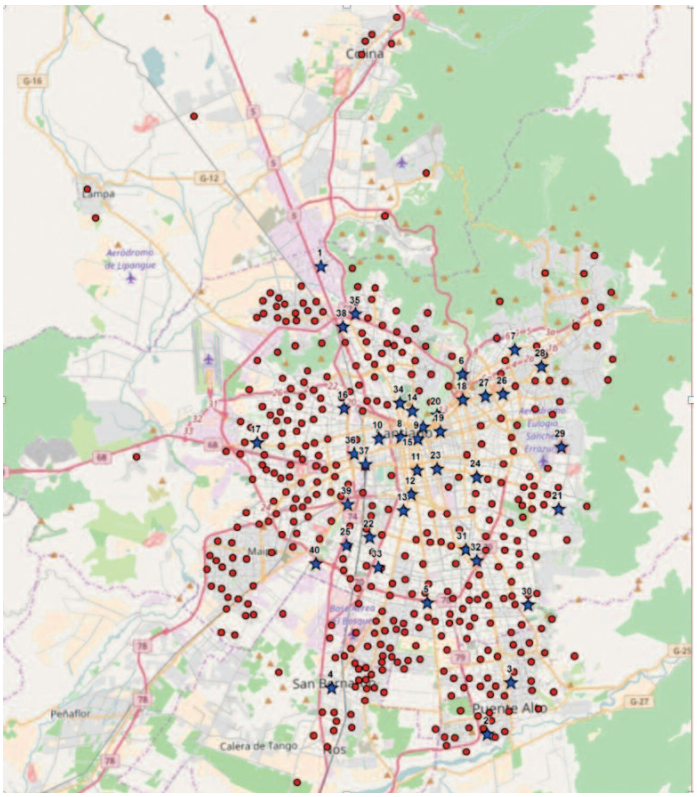
A Colombian illicit cigarette trade study similarly approached an age- and gender-stratified sample using geographically controlled sampling (Maldonado et al., 2018). The study collected data in the areas surrounding educational institutions, public transportation stations, city parks, etc. The study survey spots were distributed across each city, so that individuals from varying city districts were equally represented.

3.3.5.3 Cluster sampling

As discussed in earlier sections, simple random sampling focuses on individual elements selected from the whole population. Stratified sampling divides the population into subgroups (strata) based on observable characteristics (for example, gender, age, and race). Cluster sampling is in between a simple random sample and a stratified sample. The population is broken into groups (clusters). A sample of those groups is selected for the study. A simple random sample of the study subjects is drawn from each selected cluster. An example of clustered sampling is a study that divides a city into blocks (clusters) and then randomly samples from those blocks. The analysis is performed on only the sampled blocks.

Due to the nature of the littered pack collection method, cluster sampling is the most appropriate sampling approach. The traditional simple random sample (that is, randomly selecting packs from a list or database) would be nearly impossible to implement. In cluster sampling, the region of interest is divided into clusters, from which random clusters are chosen. Littered packs can be collected only from those chosen clusters. There are several examples of studies using the cluster sampling technique for the littered pack collection method: Merriman (2010), Stoklosa and Ross (2014), Barker et al. (2016), and Ross et al. (2019). This sampling method was also used in several smoker pack examination surveys (for example, Joossens et al., 2014; Little et al., 2019).

Figure 4 Congregation points used in the study of illicit trade in the metropolitan area of Santiago de Chile



Source: Authors' own map

The first stage of cluster sampling with the littered pack collection method requires researchers to select geographical clusters. Clusters are referred to as primary sampling units (PSUs) and should be sufficiently small. During the second stage, a sample of packs is collected in the selected PSU. Stoklosa and Ross (2014) estimated city-level prevalence of tax avoidance and evasion in Warsaw, Poland. Researchers selected a sample of 30 PSUs out of 783 voting districts within the boundaries of the city. Barker and colleagues (2016), who estimated national-level prevalence of tax avoidance and evasion in the U.S., selected a sample of 160 school districts areas (PSUs) out of tens of thousands of school districts in the country. Apart from voting districts and school districts, other geographical divisions, such as census tracts, postal codes, zones, and city districts are potential candidates for PSUs in a littered pack collection study.

Two factors contribute to the decision of which geographical division to use for a PSU: 1) the information available for each geographical division; and 2) the size of the geographical division. The sampling process requires smoker information, such as where they live, smoke, and discard their packs. This is vital information to ensure a representative sample. The best PSU candidates are those for which this information is readily available. It is important to note that a smaller census tract will have lower variance. Smaller PSUs are preferred, except in cases where visiting multiple smaller PSUs is logistically challenging or costly. As an alternative, researchers should consider several larger PSUs and collect more data from each.

Although theoretically preferred, in practice, information on tobacco product use and discarding behavior is often not available for PSUs. In this case, smoking and discarding behaviors must be estimated using existing PSU information. The researchers can use pre-existing information on socioeconomic

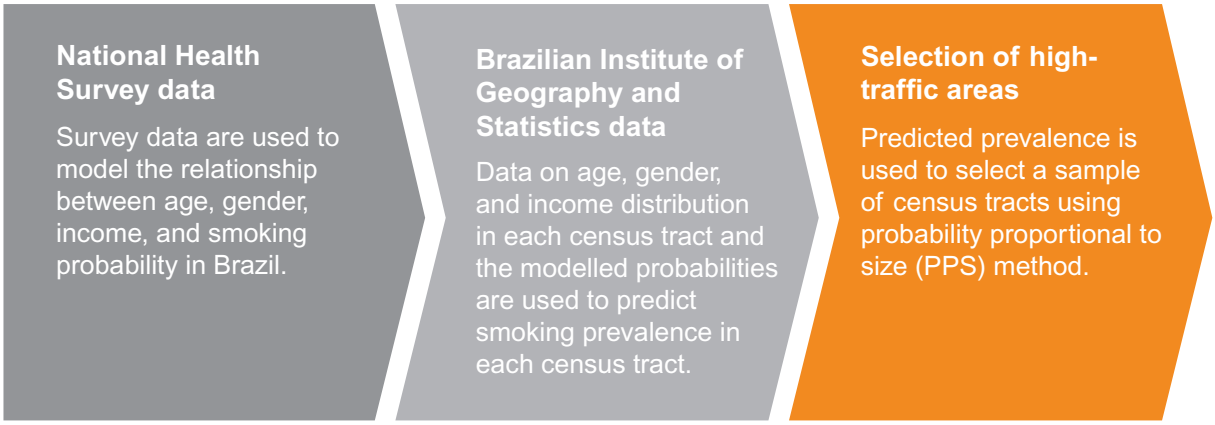
characteristics of each PSU population. These estimates can be used to infer the number of smokers of each PSU. Gender, age, and income characteristics are the best tobacco smoking predictors. In general, men are more likely to smoke than women. In terms of age, the probability of smoking is bell-shaped: at first, smoking rates rise with age until they peak, after which they decline. Smoking rates also have a strong correlation with income. Whether smoking rates rise or fall with income depends on the country's tobacco epidemic or the stage of the country's economic and educational development.

A Brazilian study estimated the smoking population in each PSU for a littered pack collection method. The study aimed to estimate tobacco tax avoidance and evasion in the cities of Rio de Janeiro, São Paulo, Belo Horizonte, Campo Grande, and João Pessoa. The researchers used census tracts as the PSUs. In order to select the best census-tract-based samples, the researchers needed to evaluate the number of smokers in each PSU. The Brazilian Institute of Geography and Statistics (IBGE), which conducts the census, provides information on age, gender, and income distributions of residents in each census tract. These data were combined with data from the National Health Survey on smoker characteristics to estimate the number of smokers in each census tract. Age, gender, and income were used to predict the probability of cigarette smoking in each census tract, using a logistics regression model for each census tract. The predicted smoking probabilities were used to determine the number of smokers (Figure 5) (Carvalho Figueiredo et al., forthcoming). Appendix 6.4 provides more detail on this study..

The number of included PSUs depends on the desired final sample size. The study described in chapter 3.3.4 estimates the desired number of packs per Mexican city at 670 packs, requiring collectors to walk approximately 52 miles (83 kilometers). The Basic Geostatistical Areas (Área Geoestadística Básica or AGEBs) were selected as PSUs. Researchers walked along all major roads in each AGEB. Since the average length of roads in each AGEB was approximately 5 miles (8 kilometers), the researchers determined that they needed to visit on average 11 AGEBs per city. The number of AGEBs ranged from 9 to 14 and the variability depended on the street density in each city (Sáenz de Miera Juárez et al., 2020).

Once the number of study elements in each PSU (cluster) is determined and the number of PSUs is chosen, the selection of PSUs can occur in two ways. If the size of the pack population in each PSU is expected to be similar among all PSUs, the sample of the PSUs can be selected randomly. However, if the size of the pack population differs significantly among the PSUs, the PSUs need to be selected using

Figure 5 Sampling of geographical regions using predicted smoking prevalence and probability proportional to size (PPS) method in Brazil

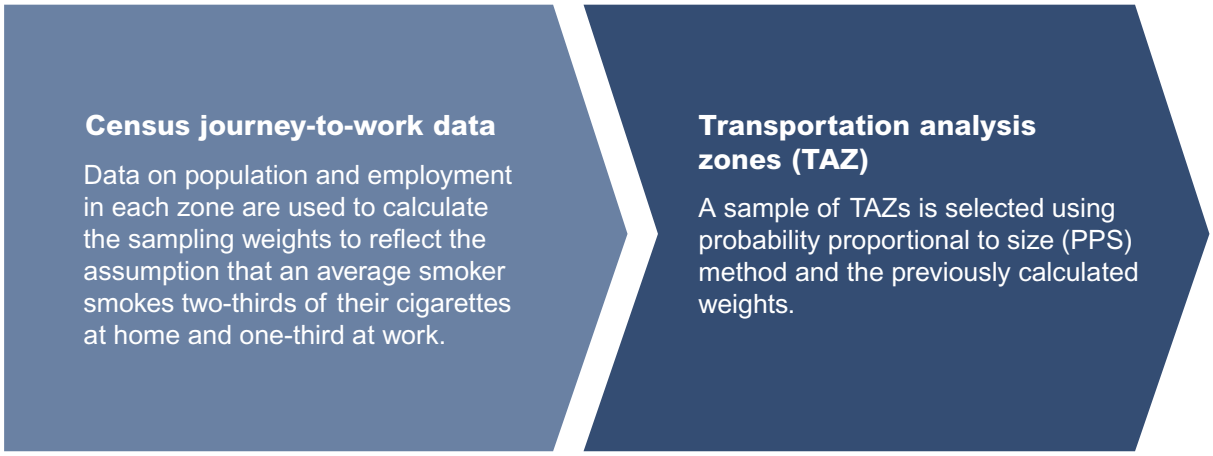


probability proportional to size (PPS). The simple random sampling of PSUs is straightforward. Appendix 6.5 provides an example of random PSU selection using Stata code. The PPS sampling is only slightly more complicated. Each PSU is given an appropriate weight based on the number of smokers and/or estimated discarded packs. The PSUs are then sampled based on the weights. Appendix 6.4 has the code for PPS selection from the aforementioned Brazilian study.

A PPS sampling with two or more weight variables is more complex. A study conducted in Chicago (Figure 6) used such an approach to estimate cigarette tax avoidance and evasion (Merriman, 2010). Transportation zones were used as PSUs. Information on cigarette littering behavior was not available in the transportation zones, so researchers needed to estimate the number of smokers using data on available characteristics. They used population and number of people working in each area as predictors. Areas with a higher population were given a weight of 100 percent. A weight of 49 percent was given to higher-employment regions (Figure 6).[2] The weights used reflect the assumption that an average smoker smokes two-thirds of their cigarettes at home and one-third at work. The aim was to account for smokers who discard their packs near their residence as well as their employment location. The transportation zones were weighted using the PPS method to select a sample of zones where packs were collected. Appendix 6.6 presents the Stata code that can be used for this kind of sampling.

An important distinction between the Brazilian and the Chicago study is that everyone in Chicago was assumed to have the same probability of being a smoker. The Chicago study made no distinction between age, gender, income, or other observable characteristics in their smoking probabilities.

Figure 6 Sampling using probability proportional to size (PPS) method in Chicago



[2] Merriman uses the weight of 49 percent and not of 50 percent because of a technical constraint. Employment is highly concentrated in some parts of Chicago. The highest weight that allowed for selection of 100 TAZs from the total of 930 Chicago TAZs without replacement was 49.

3.3.6 Sampling stages

Only in the case of simple random sampling is the study design a single-stage sampling. All other sampling methods require multistage sampling. A multistage sampling approach refers to drawing subsequent samples in ever-smaller sampling units. The studies mentioned in previous sections provide good examples. A Chilean study on illicit trade sampled in Santiago de Chile selected 40 congregation points (first stage), after which smokers were selected using the stratified sampling method (second stage) (Paraje et al., 2018). Similarly, the Mexican study selected on average 11 AGEBs per city (first stage), after which all littered cigarette packs were collected along major streets (second stage) (Sáenz de Miera Juárez et al., 2020).

Sometimes it is necessary to use more than two sampling stages. For example, a Georgian study on illicit trade selected PSUs by using the PPS method (first stage). The households were subsequently selected within each PSU (second stage). Finally, smokers were randomly selected within each household using the Kish grid method (third stage) (Little et al., 2019).

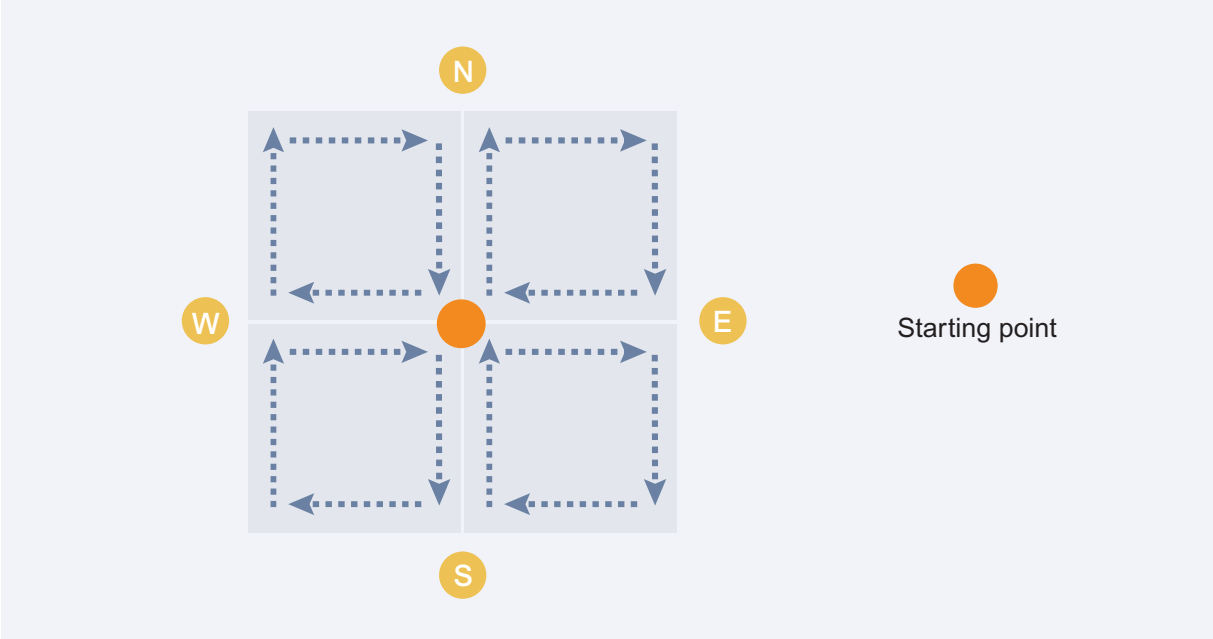
For littered pack collections, it is almost always necessary to use more than two sampling stages. It is rarely feasible to collect all discarded packs within a PSU. When PSUs are large, researchers need to designate routes (second stage) where littered packs will be collected. Packs will be systematically collected from only parts of the PSU. The length of the route is determined by several factors: the number of packs expected to be found per unit distance (miles or kilometers), the number of PSUs selected into the sample, and the desired total pack sample size.

A walking protocol for pack collection will ensure that proper sampling occurs. The protocol should be strictly adhered to during the data collection process in all PSUs. A study on illicit trade conducted in Buenos Aires, Argentina conducted by Pizarro et al. implemented the following walking protocol: enumerators walk 100 meters north from the starting point. Every 100 meters, the collector turns right until they have completed a square with a 400-meter circumference. This pattern is repeated four times, each time choosing a different starting direction. The end result is a two-by-two grid of 400-meter squares (Figure 7). All discarded packs along the route are collected (Pizarro et al., in review). The sample is unbiased when collectors walk the pre-planned routes. A best practice is to select all packs along the pre-planned routes. Different numbers of packs collected per PSU reflect the number of smokers more accurately as well as smoking intensity, which cannot be measured with collection by quota.

Researchers should also consider the appropriate starting point in the scenario described above. Since the starting point is at the epicenter of the larger 1600-meter square, the randomness of the starting point location is crucial. Handpicking a starting point will not guarantee an accurate representation of the total smoking population within a PSU. For example, if illicit packs are more prevalent around market squares, then pack collections that always start at the market square will overrepresent illicit cigarette packs. Choosing a starting point randomly for each data collection route ensures that the collected packs better represent the packs typically smoked in a given area.

Starting locations can be chosen in several different ways. In studies where PSUs are based on census tracts or other official geographical divisions, a list of physical addresses within each of those PSUs typically exists. One of these addresses can be randomly selected from the list for each PSU. Alternatively, a starting point can be selected spatially at random. If a digital map (for example, shapefiles) exists, PSU starting points can be selected at random. In the ArcGIS program (a geographical information systems software), a “Create Random Points” function can be utilized. The collectors can use the GPS coordinates to locate their starting points. Figure 8 below presents the randomly selected starting points in the census tracts in one of the neighborhoods of Rio de Janeiro, Brazil (Carvalho Figueiredo, forthcoming).

Figure 7 A 1600-meter route walked in each of the selected census tracts in the illicit pack study in Argentina



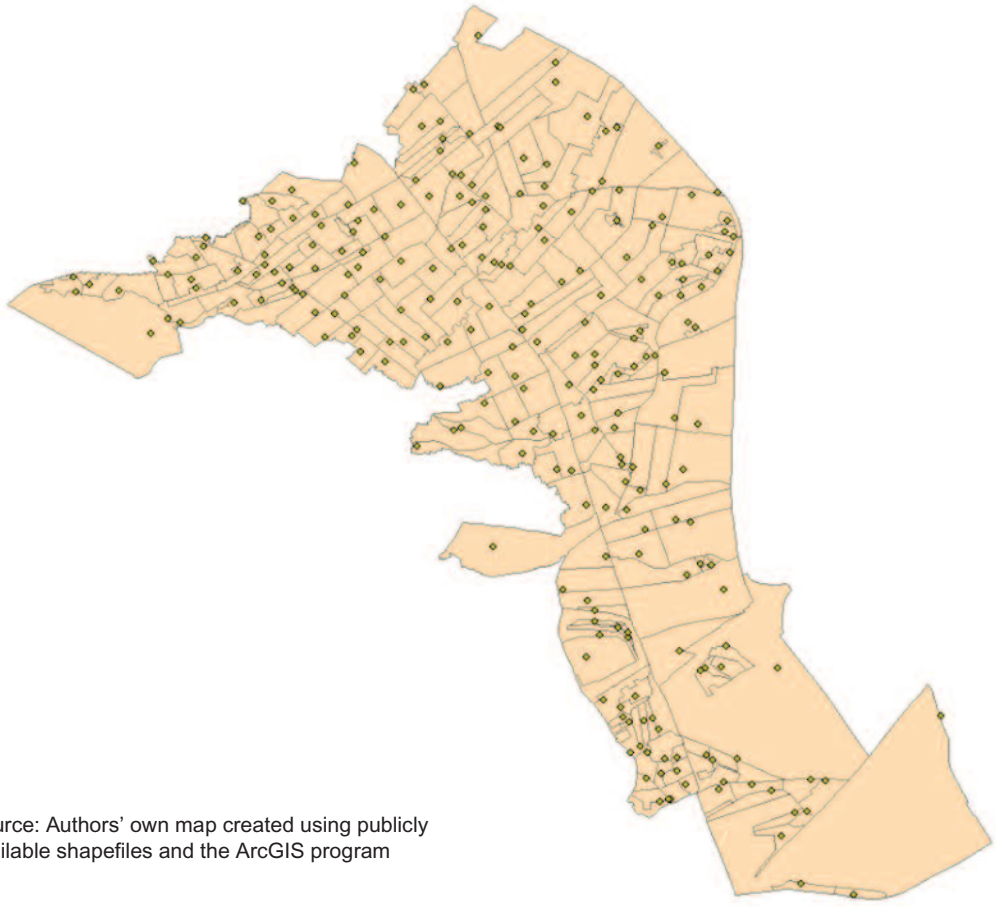
Source: Pizarro et al., in review

Apart from ArcGIS, there are several free online random point generators. These can generate random points within a circular region (for example, centered at the middle of the PSU) or within a rectangular region (for example, defined as north, south, west, and east limits of each PSU). Although any map projection distorts the shape of large areas (for example, continents and countries), the PSUs researched in the discarded pack collections are always much smaller and are not significantly affected by map projection. Thus, selecting random starting points on the map should still work in a littered pack collection.

When choosing the survey starting point, it is imperative not to target places where illicit tobacco product packs are more likely to be found. In fact, littered pack collections paid for by the tobacco industry are often criticized for focusing on areas where consumption of tax noncompliant products is more likely (Rowell et al., 2014). United Kingdom industry surveys are conducted at sporting events, where foreign/illicit packs have a higher probability of being consumed (Gilmore et al., 2014). Similarly, an industry-funded pack collection study in Germany systematically overrepresented geographical regions along the country's border and around U.S. military bases, where more foreign cigarettes are to be expected (Adams et al., 2011). Objectively measuring tobacco tax avoidance and evasion requires collecting data from a random sample of all cigarettes consumed.

Researchers should not send enumerators to areas with high percentages of illicit cigarette users, unless it is a part of a study design to collect information from these hot spots. Instead, the focus should be on all cigarettes consumed. If a venue with a high occurrence of illicit cigarettes falls in the collectors' route, then it should be part of the data collection. Similarly, the study should not focus on high-income areas only. The goal is to not over- or under-sample certain areas due to preconceived notions, as this will introduce bias into the analysis and make the results unreliable.

Figure 8 Randomly selected points in each census tract in the Ramos neighborhood of Rio de Janeiro, Brazil



Source: Authors' own map created using publicly available shapefiles and the ArcGIS program

3.4 Data collection

3.4.1 Questionnaire construction

Developing instruments for the smoker pack examination survey is relatively straightforward. To begin with, the enumerator should recite a brief explanation and obtain consent to be interviewed. The explanation should include information on the study objective and the participant's rights. After consent is given, the first several questions on a survey should be screening questions. These are designed to determine if the subject is eligible to participate in the study. The screening questions should include the criteria for inclusion in the study, such as participant age or smoking status. Stratified sampling surveys should also include information on the participants' gender, income, etc.

Next, the enumerator should ask the smokers to show their cigarette pack. Whenever smokers have several cigarette packs available (for example, there are multiple packs in the household), research guidelines should determine how the enumerator should proceed. A best practice would be to include only the last-purchased pack (e.g., Maldonado 2018), or, if several packs were purchased, the pack from which the last cigarette was smoked (e.g., Paraje 2018). This way, particular brands or illicit cigarettes are not specifically

targeted. Including some cigarette packs and not others can lead to a biased sample. Representing only the most recently purchased or smoked cigarette pack does not discriminate subjectively and most accurately reflects smoking behavior. The goal is to have an unbiased sample in order to provide the most accurate representation of population behavior.

The following hypothetical example serves to demonstrate the benefit of asking for the last-purchased pack. Assume that a sample consists of one hundred smokers, smoking brands X and Z. All smokers purchase either brand and alternate their purchasing behavior between X and Z. Each smoker buys one pack of X and subsequently purchase two packs of Z. Brand X is the preferred brand, but brand Z is cheaper because it is the illicit brand and is therefore purchased at a higher frequency. Each smoker has at least one pack of either cigarette brand. If an enumerator asks the consumer to show only their preferred brand, all one hundred members would produce the brand X cigarettes. As a result, the estimated prevalence of illicit cigarettes would be zero for this household. Alternatively, if the enumerator asks each smoker if they have illicit cigarettes, then the estimated prevalence would be 100 percent for the population. Neither scenario captures the true smoking behavior in the population. A survey asking each smoker to produce their last-purchased pack, however, will arrive at the illicit cigarette prevalence of 67 percent, which is the true purchasing rate of the illicit brand Z.

The survey should accurately capture the illicit cigarette smoking rate. This is referred to as consumer smoking intensity. An important question to capture consumer smoking intensity is how many cigarettes are smoked per day or week. Estimation of illicit cigarette prevalence is impossible without this information. The following hypothetical example serves to demonstrate the necessity of asking about the consumer's smoking intensity. Assume that there are ten smokers, each smoking only one brand. Eight of the ten smokers consume a legal brand X at a rate of five cigarettes per day. The other two smokers consume an illegal brand Z at a rate of 20 cigarettes per day (a full pack). Since the smoking rates are different between the two types of consumers, a survey failing to include a question on smoking rate would not accurately portray the illicit cigarette prevalence in the group. The correct illicit cigarette prevalence among the ten smokers is 50 percent (for every 40 cigarettes of brand X smoked by the eight smokers, there are another 40 cigarettes of brand Z smoked by the two smokers). Thus, the true illicit cigarette prevalence can only be captured by including appropriate weights on participants' smoking intensity.

In order to decrease the attrition rate, enumerators should explain the purpose of conducting the survey. In order for the respondents to understand why they should devote time to take the survey, enumerators should stress that the survey is conducted to improve public health. If the survey is conducted in collaboration with a public health organization, the name of the organization should be mentioned. People across all demographic groups generally know and trust public health organizations and, therefore, mentioning the name of the organization will increase the likelihood of response.

Once the last-purchased pack is produced, the enumerator needs to collect data on the pack characteristics. The survey should include all necessary questions about the pack features that aid researchers in distinguishing between compliant and noncompliant packs (see section 3.2.2). Questions should focus on the pack brand and presence of appropriate tax stamp as well as the presence, type, size, and language of the health warning, and other pack markings required by law.

The information will be more objective if it is collected through pack observation during the survey or by obtaining a pack from the respondent (for example, for future forensic examination) rather than a respondent answering questions. A good practice is to take pictures of each side of the presented pack. These pictures can help to identify coding errors, since researchers can refer back to the pack images if issues arise. These pictures can also serve as study documentation in case the results of the study are questioned. Pictures taken with a cell phone will have the appropriate date, time, and place (GPS

coordinates) associated with each picture. This facilitates matching to the appropriate survey, ease of searching, and is useful information during the data analysis phase.

Multiple choice is the best format for survey questions because they standardize responses—mitigating potential coding errors—and aid the enumerator in swift survey completion. For example, the cigarette brand question should include a list of cigarette brands available on the market. The option to enter the brand name manually should also be present in the event the brand name is not on the list. Finally, if the survey has a tax stamp question then the question should be carefully formulated. In some countries, tax stamps on exterior packing or on cellophane will be torn off while opening the pack. This feature is designed to avoid reuse of tax stamps. The tax stamp question should be formulated in such a way that it can distinguish between packs with tax stamp residue and packs with no tax stamp (for example, researchers in Mongolia looked for glue residue that would indicate that the stamp had been removed (Ross et al., 2019)).

Two pieces of information may be sufficient to estimate the prevalence of noncompliant cigarette packs. These two pieces of information are consumer smoking intensity and features of the last-purchased cigarette pack. There are three exceptions, however, where additional information may be required. First, when a significant portion of responders are unable/unwilling to produce their cigarette packs, the researchers may want to follow up with several additional questions. As discussed in earlier sections, the inability (for example, single-stick smokers) or unwillingness (for example, cultural norms) to produce cigarette packs is likely correlated with the smoker's demographic and socioeconomic characteristics. Dropping these individuals from the sample introduces bias into the illicit trade estimation. The additional questions asked can be used later to impute the missing data on tax compliance. Additional questions should include information on the last-purchased cigarette brand, quantity, and price, as well as the location of the last purchase and whether the respondent smokes packs or single cigarettes. When interviewing a single-stick user, the enumerator should ask the smoker to show the cigarette, as the stick often has the brand printed on it.

The second circumstance requiring additional smoker information is when an in-depth analysis of illicit trade is needed. For instance, collecting data on the smoker's education, income, and occupation might help in determining whether certain socioeconomic groups or occupations are more prone to smoking illicit cigarettes. Including a question regarding purchasing locations could help to understand the supply chain of noncompliant cigarettes. Additionally, information on the type of retail outlet from which the cigarettes were purchased is helpful. Distinguishing cigarettes bought during a trip abroad, in a duty-free store, or from a local market helps distinguish illegal tax evasion and legal tax avoidance.

Third, it may be necessary to obtain packs from the smokers if additional forensic examination of the pack or the stamp is needed. In such cases, smokers could be given a small remuneration for providing their pack to the enumerators.

There is, of course, a trade-off between gathering more information and the time required to complete the survey. Although gathering more information will help in obtaining better estimates, increasing the number of questions could possibly discourage people from participating in the survey. Researchers need to weigh the additional information gained against the potentially higher attrition rate.

A sample questionnaire is presented in appendix 6.2 as a guideline. The questionnaire should be tailored to the needs of each study, taking into consideration the issues discussed in this chapter. Most importantly, the survey must be compliant with local ethics standards. Researchers are also encouraged to contact authors of research discussed in this toolkit who used surveys (for example, Argentina (Prizarro et al., in review), Brazil (Carvalho Figueiredo, forthcoming), Chile (Paraje et al., 2018), Colombia (Maldonado et al., 2018), Georgia (Little et al., 2019), Mexico (Sáenz de Miera Juárez, 2020), Mongolia (Ross et al., 2019), Poland (Stoklosa and Ross, 2014) and six countries in Southeastern Europe: Albania, Bosnia and Herzegovina,

Kosovo, Montenegro, North Macedonia, and Serbia (Zubovic et al., forthcoming). The authors of this toolkit can facilitate contact if needed.

3.4.2 Ethics approval

The survey must be approved by the Institutional Review Board (IRB). The littered pack collection method will most likely not require approval from the IRB because human subjects are not directly involved. Some countries may require an ethics review, however, to assess the risks to which researchers are exposed. Whenever human subjects are involved, IRB approval is required. Although the smoker pack examination survey's main focus is the cigarette pack, IRB approval is required because human subjects are involved.

IRB needs to provide clearance before data collection can begin. Optimally, the IRB clearance is obtained in the country where the study is being conducted. The clearance process can take up to several weeks. Pack examination surveys should build additional time for IRB clearance into the project timeline. An important consideration is that while the sale of illicit tobacco products may be illegal, consumption of these tobacco products tends not to be. When smokers are not engaging in illegal activity, it is possible for researchers to obtain an exemption from the IRB. An IRB sample exemption application is shown in appendix 6.3. A complete packet needs to be sent to the IRB for review including the full study documentation, the sampling methodology, the survey instrument, the agreement with the interviewing company, etc. Failure to obtain the proper IRB clearance may result in an unpublishable study in a reputable peer-reviewed journal and inability to present the results in academic settings.

3.4.3 Survey execution

The goal of the data collection process is to gather the information in a manner that is consistent so it can be replicated. While conducting the data collection process for these surveys in-house is manageable, delegating the data collection process to a market research company is a viable alternative. Research groups with limited time, scarce human resources, and sufficient funding may prefer the market research company option.

Several well-known international market research companies exist as well as smaller, domestic enumerators. This toolkit does not endorse any particular company. As a general guideline, it is crucial that the market research firm be ready to cooperate well with the research team. Market research companies specialize in working with small- and medium-sized businesses. These businesses often lack the capacity to conduct research studies and analysis on their own and generally look for a finished product including a report or presentation of the study findings. Market research companies are generally unfamiliar with clients who are involved in the study, including at the design stage, survey construction, and analysis of raw data. The market research company should know upfront that the researchers will be actively involved in each stage of the study.

The task assigned to the market research company should be clearly defined and outlined, since research groups are an atypical client for them. However, it has been successfully done. Recently, a market research firm conducted a survey in six countries of the Southeastern European region, called the Survey of Tobacco Consumption-Southeastern Europe (STC-SEE) (Zubovic et al., forthcoming). A consortium of think tanks in the region, partnering with Tobacconomics team at the University of Illinois at Chicago developed the questionnaire for the STC-SEE, which included a smoker pack examination survey, and then analyzed the data collected by the firm. The roles of the market research firm and the research consortium partners were clearly defined, and the survey data is being used in analysis of tobacco use prevalence and tax evasion and avoidance in the region.

Whether the data collection process occurs in-house or is outsourced, a pilot study is essential to test the survey instruments. As mentioned before, pilots can be useful in approximating the response rate and—in the case of littered pack collections—the number of packs found along a predefined route. The pilot also serves to provide an estimate of the interview duration and identify potential problems with the survey. At the very least, researchers should ask for feedback from other researchers not involved in the project.

The timing of the data collection requires additional consideration. In the case of the littered pack collection, data collection timing is of the utmost importance. Streets are regularly cleaned, removing the littered packs from the environment. How frequently and when streets are cleaned will be necessary information for the researchers to ascertain. For the smoker pack examination surveys, timing needs to be considered when surveying in public locations or visiting households. Different settings will be affected by different timing.

Before implementation of the data collection process, some administrative work is required. For the littered pack collection method, researchers need to create walking routes on maps for the collectors. Other equipment—such as plastic bags with a unique identifying number and gloves—needs to be organized. Collectors should use best practices and take a picture of the littered pack and the collection bag with the identification number. Alternatively, the bag identification number can be marked on the walking map. Although they may seem tedious, these preparations help researchers identify the pack collection location, which may be relevant later. Further, coordination among the pack collection sites and research facilities requires attention, especially in cases where packs are collected across large regions. Finally, cigarette pack storage needs to be secured. Although storing empty packs can be both costly and troublesome, a good practice is to keep the collected packs as long as possible, in case the study results are questioned or in case additional pack examination is needed.

For smoker pack examination surveys, an enumerator should be given the survey protocol, including the data collection procedural information, an enumerator manual to reinforce skills learned during basic training, and the reference materials needed for the enumerators to complete the pack examination (for example, a hardcopy of all approved health warnings should be included in the packet of information to help the enumerator distinguish between domestic and foreign health warnings). For both the littered pack collection and the smoker pack examination survey, enumerators should have access to a functional cell phone with sufficient battery power and available photo storage.

Having researchers on hand during the training process is advised even when data collection is outsourced to market research companies. Researchers who are intimately involved in the specifics of the research study also benefit the training process. Researchers can convey the importance of the study and instill the same enthusiasm in the enumerators. Providing enumerators with the study context can enhance motivation and ensure a well-trained staff. Having researchers present also guarantees that the necessary information and materials are disseminated.

Good data collection management requires managers to work closely with enumerators, especially during the first collection days. It is imperative that managers thoroughly check the data collection quality. The research manager should cross-validate the packs and the pictures taken. In this way, errors can be addressed early on and fixed accordingly without wasting resources. As an example, in the Brazilian study it became apparent during the first days that the enumerators could not enter the buildings of the wealthiest households to conduct the pack examination survey. High security prevented the enumerators from entering. The research manager decided that enumerators ought to wait for residents outside of the building. With this modification in place the data collection continued with little interruption, and the proper number of the wealthiest households were still included in the survey (Carvalho Figueiredo, forthcoming).

As discussed in section 3.2.3, some illicit trade studies use both the smoker pack examination and littered pack collection methods simultaneously, which allows for cross-validation of findings. When the probability

proportional to size (PPS) is used to sample areas (PSUs), both methods can be conducted concurrently. The enumerator can collect littered packs and conduct surveys along the route. However, since littered pack collection requires less skill, implementing this strategy is only efficient when the enumerator does a lot of walking between households. Otherwise, it may be more efficient to have two sets of enumerators, one focusing on littered pack collection and the other conducting the surveys. For example, an illicit trade study in Warsaw, Poland hired a market research company to conduct interviews with smokers, while college students were used to collect littered packs (Stoklosa and Ross, 2014). When the two methods are conducted separately, they should be implemented within a similar timeframe to allow for cross-validation.

Data entry is the final stage of the data collection process. Data managers oversee the creation of a workable dataset. In this dataset, each pack represents one entry. There are several ways to impute the information. These datasets can be created in a database management system in a statistical software (for example, Stata), or a spreadsheet program (for example, Microsoft Excel). The littered pack collection method requires each pack to be individually assessed and the information entered into the database. The smoker pack examination survey requires the information to be transcribed into the database. Smoker characteristics, including smoking intensity will also be included for each pack. For both methods, time and place (for example, PSU) of data collection needs to be added to each entry. If additional information—such as GPS coordinates and pictures—is available, this information should also be included. A good practice is to include a reference to each pack picture in the data entry for ease of identification later.

Not all pack features need to be entered into the database. Figure 2 in section 3.2.2 presents a sequential approach to identifying illicit packs. Specifically, in an example presented in that figure, a brand of cigarettes is first examined. Packs with brands that are not registered for legal sales are coded as illicit. Only those packs with brands that are officially allowed on the market are then subjected to a detail examination. In this approach, the brand information is used to quickly weed out all brands that are illicit. The sequential approach saves coders time, as they do not need to examine other pack characteristics for all the packs with illicit brands.

Be careful. Errors will sneak in at each stage of the process. It is especially vital to check the database for errors. As a check with the littered pack collection method it is advisable to have two separate coders enter the information for a subset of the collected packs. If for some reason the data entered does not match, then researchers can investigate further. Otherwise it can be challenging to find errors. With the pack examination survey, a subset of the sample can be cross-checked by comparing the information entered into the database with the pack photos that were taken by enumerators.

3.4.4 Identifying counterfeit packs

It is possible that cigarette packs appear legal but are in fact produced illegally. These types of cigarettes fall in the counterfeit pack category. Counterfeit packs are relatively rare. Tobacco industry reports state that only 0.5 percent of cigarettes consumed worldwide are counterfeit (Philip Morris International, 2019). Nevertheless, several countries face a cigarette counterfeiting problem.

When counterfeiting is an issue, ultraviolet (UV) irradiation and light microscopy can identify counterfeit cigarettes (Kurti et al., 2017). The method was developed by Kurti. The equipment needed includes a light microscope with a USB camera, a UV Lamp, a UV viewing cabinet and, optionally, a UV USB sensor. The equipment is relatively inexpensive, and the application is straightforward. The Kurti method uses a diagnostic test performed on counterfeit packs provided by law enforcement.

Researchers need to obtain confirmed counterfeit packs (for example, from law enforcement authorities) that can then be compared to confirmed legal packs (for example, obtained from reliable stores) to calibrate

a test with optimal specificity and sensitivity. With the calibration complete, the second phase of the project can be implemented. In this phase, packs are collected through the littered pack collection or through the smoker pack examination survey. Those packs are sorted by legality. A pack deemed legal after the initial examination will be subject to further tests to definitively determine that it is not counterfeit. The counterfeiting test is used to establish pack legality.

Counterfeit cigarette pack characteristics can differ between countries. The Kurti method was specifically designed to identify counterfeit Marlboro cigarettes. This method has also proven adequate for verifying other cigarette brands (Kurti et al, 2017). The Kurti method remains appropriate for identification of counterfeit packs in a multitude of settings. In Figure 9 below, the Kurti method guidelines are described in detail.

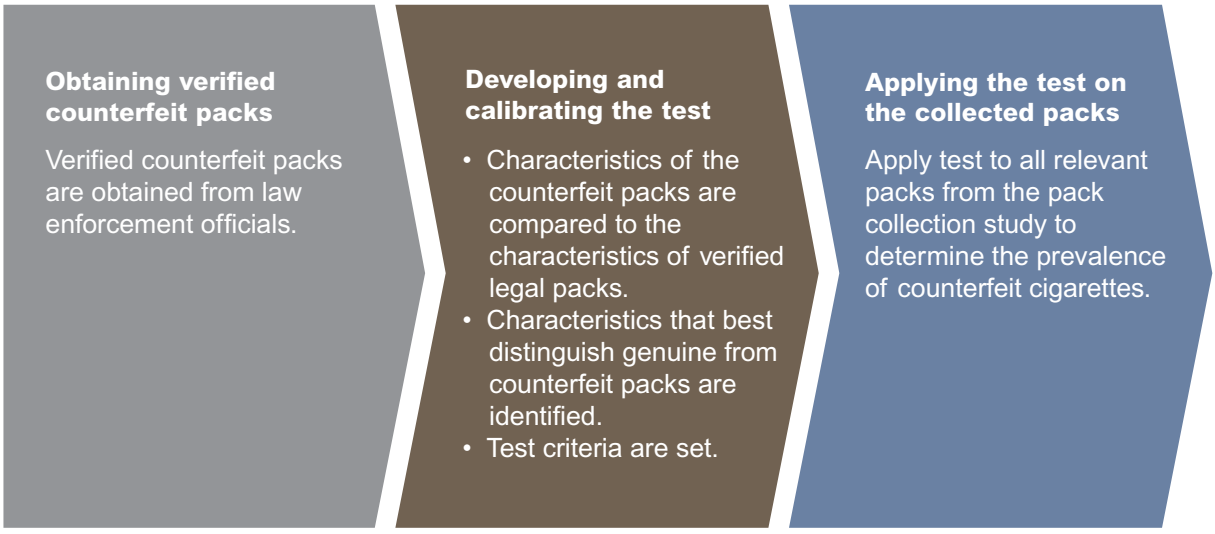
3.4.4.1 Creating and calibrating the test

The first stage of the Kurti method requires collaboration with law enforcement agencies. Researchers need to identify the most frequent counterfeit cigarette brands in their study area. Law enforcement agencies would have access to this information as well as sample counterfeit packs. The counterfeit cigarette packs should be the most recently obtained packs. Most counterfeit packs will have been seized through raids on illegal manufacturing facilities. It is best if the counterfeit pack samples for each brand come from several different seizures. These packs obtained from law enforcement agencies constitute the sample of independently verified counterfeit packs.

Researchers must also use a control group for comparison. These legal packs also need to be verified as legal. Each counterfeit brand present in the sample should have a legal counterpart. Legal packs can be obtained through large, trusted retail chains or directly from legally operating tobacco manufacturers.

Once all necessary packs are obtained, data need to be collected on the pack characteristics from both samples. Ultra-violet radiation and light microscopy are used to examine each pack. Following the description of the Kurti method (Kurti et al., 2017), the pack outside, inner frame, and—if the pack is wrapped in foil—the inside foil all need to be exposed to a long-wave UV light. Researchers need to collect data on the presence of optical brightening agents and fluorescence for both groups. The UV sensor can

Figure 9 Counterfeit pack test development



objectively measure the presence of both features. Second, researchers should look for signs of poor-quality printing at several pre-defined places on each counterfeit pack using the light microscope. These areas are then compared to the printing on genuine packs. Each brand will have unique elements on the print design that distinguish counterfeit and legal packs.

With the microscopic verification and the UV irradiation data, researchers can identify specific brand characteristics that best distinguish genuine from counterfeit packs. The ideal test will have the highest specificity (that is, ability to detect whether a pack is genuine, or true negatives) as well as the highest sensitivity (that is, ability to detect whether the pack is counterfeit, or true positives). Researchers should use the receiver operating characteristic (ROC) curve analysis. This will help to identify optimal characteristics for each brand.

It is important to note that this test is not 100 percent accurate in distinguishing between legal and counterfeit packs. The test specificity and sensitivity will provide the research with a margin of error. The Kurti method has been found to have high accuracy between brands since the test accuracy is similar between brands. The test correctly classified 100 percent of the Newport and Marlboro Gold packs using only two pack characteristics (sensitivity and specificity of 100 percent). The test correctly identified 96 percent of the regular Marlboro sample using four characteristics (sensitivity of 94.12 percent and specificity of 100 percent) (Kurti et al., 2017). It is essential that pack raters are well-trained so they can effectively collect the necessary information from cigarette packs using this test.

3.4.4.2 Determining the extent of cigarette counterfeiting using the test

Once the counterfeiting test is developed, the test can be applied on a large sample of packs obtained from a littered pack collection or from a smoker pack examination survey where smokers provide their packs. The test is to be applied to packs deemed as legal. These packs are identified as legal because they present clearly visible and easily recognizable appropriate health warnings and/or tax stamps. The test will help determine whether a pack classified as legal is in fact genuine.

The test will only be able to distinguish between brands that have been analyzed. For example, if only Pall Mall cigarettes were used to develop the test, then the test can only check whether Pall Mall cigarette packs are legal or counterfeit. A good practice is to have 10 percent of the legal subsample independently evaluated by another rater. Test reliability and accuracy can be confirmed using this technique.

3.4.5 Use of pack examination studies for other tobacco products than cigarettes

Although the methods described in this chapter refer to cigarettes only, both the smoker pack examination survey and the littered pack collection can be used to measure the prevalence of illicit trade in other tobacco products. In places where the packaging of the other products is regulated to require for example, health warnings or tax stamps, researchers can collect data and evaluate the packs using the techniques described in this chapter. Thus far, however, there has been no large-scale study of illicit trade in tobacco products other than cigarettes.

3.5 Data analysis

The most important identifying characteristic of each pack is whether it is compliant or noncompliant. This can easily be represented with a dummy variable (zero for compliant, one for noncompliant). Assigning the correct dummy may seem trivial, but it is not always straightforward which packs are compliant and which

are noncompliant. Occasionally, multiple pack features need to be taken into account to determine pack legality. Section 3.2.2 has a thorough discussion on how to use pack features to distinguish legal (tax-paid) packs from illegal (tax-unpaid) packs.

Smoker pack examination surveys in countries with a high proportion of single-stick users must apply a different compliance status as smokers may not have a pack available. In this case, price information on the last-purchased cigarettes, cigarette brand/name, and other characteristics can be used to calculate the compliance status for these smokers. The Mexican illicit cigarette trade study, for example, found that the cigarette brand name was a sufficient indicator of legality of single sticks (Sáenz de Miera Juárez, 2020). When brand and price information is insufficient to determine compliance status, the status can be imputed from smokers who produce packs.

An econometric model is a more rigorous approach to assigning compliance status for smokers who do not show their packs. In this case, a predictive model is implemented to estimate the missing values. Consumer characteristics and cigarette information are necessary to estimate these missing values. Variables that best predict compliance are usually country specific and can vary. Appendix 6.7 provides Stata code that predicts the compliance status for smokers who do not produce their packs.

The next step is to estimate the prevalence of noncompliant packs in a given geographical region. As demonstrated by Merriman, in littered pack collections, the simple mathematical average of the probabilities of compliance for each of the PSUs is an unbiased estimator of the true regionwide mean probability of compliance (Merriman, 2010). The prevalence of noncompliant packs in the sample is a good representation of prevalence of noncompliant packs in the region. Calculating the prevalence of noncompliant packs from the smoker pack examination survey is only slightly more complicated. First, researchers must weigh the smoker pack compliance status by the respondent's smoking intensity. This ensures that the weight is proportional to the number of cigarettes smoked per day. Second, if stratified sampling resulted in some strata being relatively smaller, then weights can be applied to adjust for the smaller subsamples. Noncompliance prevalence can then be accurately computed.

Apart from estimating illicit cigarette prevalence, researchers might want to conduct additional analyses on the data. Tests of proportions (<*prtest*> function in Stata) can be used to determine whether the probability of smoking noncompliant cigarettes differs among gender, age, or socioeconomic groups. Pearson $\chi 2$ test (<*tabulate, chi2*> function in Stata) can be used to determine whether the prices are different among compliant and noncompliant cigarettes. Finally, multivariate analysis can be used to study more complex relationships in the data. For example, Merriman uses a probit model (<*probit*> function in Stata) to analyze whether the probability of tax compliance in a given PSU depends on the PSU's proximity to a lower-tax jurisdiction. These calculations control for household income, race, and other socioeconomic characteristics of PSU inhabitants (Merriman, 2010).

3.6 Conclusion

This chapter presents pack examination methods to estimate illicit tobacco products trade prevalence. It focuses on two methods commonly applied in illicit trade studies: smoker pack examination survey and littered pack collection. The content of this chapter should equip researchers with knowledge on all aspects of the study from research planning and sample design, through data collection, to data analysis.

Pack examination through primary data collection is the most reliable way to measure illicit tobacco trade, but it is also costly and time-consuming. The next chapter focuses on analyzing secondary data to estimate illicit tobacco trade.

4 Gap Model

4.1 Introduction

A gap model estimates the evolution of the trend in illicit tobacco trade. A gap model estimates the difference between the total cigarette market (combination of tax-paid and tax-unpaid) as measured by surveys and reported tax-paid cigarette volumes.

Most countries collect survey information on smoking prevalence (the percentage of the population that smokes in a given period) and smoking intensity (the number of cigarettes each smoker smokes within a given period). Countries should also have precise figures, or at least adequate estimates, of total population size. The gap can be measured using these estimates. The number of cigarettes consumed can be estimated by combining the smoking prevalence, smoking intensity, and population estimates. As is generally the case with estimation techniques, the results only serve as an approximation of total cigarette consumption; measurement error that is present in the survey will carry through to the approximation. In addition to survey measurement error, people may not accurately report their smoking habits. This introduces a potentially non-random bias.

While the gap analysis application is not theoretically difficult, gap analysis implementation can often be challenging due to data and analysis constraints. Furthermore, analysis and interpretation can be arduous. Understanding this approach's strengths and weaknesses is important during the analysis and interpretation stage. This chapter covers both the theoretical concepts as well as implementation, and also addresses the finer nuances related to interpretation and presentation of results.

First, the conceptual underpinnings of the gap model, based on consumption identities, are explained. Second, the strengths and limitations of the model are spelled out conceptually and with empirical cases. Third, data sources needed to use this method are discussed and examples are given. Fourth, a number of challenges that the data may present and should be considered are also explained with specific examples. Fifth, examples of Stata codes to apply the methodology are given, along with different ways of presenting the results. Finally, a number of case studies where this method has been applied are provided.

4.2 Conceptual framework of the gap model

Let Q^T be the total number of cigarettes consumed in a certain period in a country or relevant jurisdiction (with a common tobacco tax policy), Q^L is the total amount of cigarettes paying tobacco taxes in the same period/country, and Q^E represents cigarettes that are domestically produced and taxed but smuggled out of the country:

$$Q^T = (Q^L - Q^E) + Q^I$$

where Q^I is the amount of non-tax-paying cigarettes consumed in such a period/country. The non-tax-paying cigarettes are not necessarily illicit cigarettes, as they can include, for instance, duty-free allowances for personal consumption. It is assumed that the proportion of licit cigarettes that do not pay taxes out of Q^I remains constant across time. Fluctuations in Q^I are solely due to fluctuations in illicit cigarette consumption. Q^I is the variable of interest and is the gap that will be estimated and compared over time:

$$Q^I = Q^T - (Q^L - Q^E)$$

In turn, Q^T can be defined as the total domestic consumption market (Q^D) and is equal to:

$$Q^T \equiv Q^D = \sum_{i=1}^{S} N_i$$

where S is the total number of smokers in the relevant period in a country/jurisdiction and N is the number of cigarettes consumed by smoker i in such a period/country. When working with aggregate data:

$$Q^D = P^T \times M^{Prev} \times \bar{N}$$

P^T is the total population in a country in a certain period, M^{Prev} is the proportion of smokers in the population in such a period/country, and $\bar{N}$ is the average number of cigarettes smoked in such a period/country. M^{Prev} and $\bar{N}$ have to be measured over the same time period, such as per month. Then M^{Prev} would be the monthly prevalence of smoking and $\bar{N}$ the average number of cigarettes smoked monthly by such a population. If the average intensity is measured weekly but the prevalence is measured monthly, then it must be assumed that the weekly intensity is constant within the month. The average number of cigarettes smoked per month is 4.43 times the weekly average.

The identity expressed in (1) is unobserved in practice due to smokers underreporting their smoking habits (in N). Quite often the result is that Q^L is larger than Q^D, a nonsensical result. Therefore, the relationship in (1) has to be assessed over time to assess whether the underreporting (and the proportion of legal duty-free cigarettes consumed) is constant. Equation (1) must be transformed to:

$$\frac{\Delta Q^I}{Q^I_t} = \frac{\Delta(Q^D)}{(Q^D_t)} \; S - \frac{\Delta Q^L}{Q^L_t} \; (s-1)$$

where Δ is the difference of relevant quantity in periods t+1 and t (base year). Each ratio is the percentage increase over time t. Variable s represents the ratio between the total domestic consumption market quantity and the illicit market at time t (Q^D_t / Q^I_t). Thus, s is a constant greater than 1 and, naturally, unknown. Equation (2) solves for s and presents the evolution of the illicit market.

If the percentage increase in the total domestic market (Q^D) is larger (or smaller) than the percentage increase in Q^L, it would indicate that Q^I has proportionally increased (or decreased) with respect to the base year. Assuming that underreporting is constant over time, such an increase (or decrease) would be indicative of an increase (or decrease) in the illicit trade market share (defined as illicit trade and as a percentage of the total market).

All things remaining the same, a relative increase in the illicit trade market share can occur in a domestic consumption market that is increasing over time. Thus, there would be an absolute increase in the illicit market share. On the other hand, a decrease in the illicit trade market share and domestic consumption market would imply a reduction in the volume of illicit cigarettes being sold. Finally, if the total domestic consumption market is increasing (decreasing) and the illicit trade market share is decreasing (increasing), the evolution of absolute volume sold cannot be calculated in the illicit market.

Only one illicit market size estimation (in volume) is required to extrapolate the illicit market size at other time points. The estimation conditions necessarily require that the underreporting is constant over time. In a Brazilian study, researchers used the estimate for the 2013 total cigarette volume sold in the illicit market to deduce the volume for 2012, 2014, 2015, and 2016 (Szklo et al., 2017).

When the actual illicit trade volume—measured using primary data (pack examination survey)—is unavailable, then a certain fixed percentage of underreporting can be assumed. The fixed percentage can be added to the trend estimated through the gap method (Blecher, 2010). Indirect indicators must suggest that the underreporting proportion is within approximate and reasonable bounds, or the results will be unreliable. Sensitivity analyses would provide robustness on the underreporting percentage.

4.3 Strengths, weaknesses, and limitations

The gap method's greatest strength lies in its data and resource economy. On the data side, only two data sets are needed. The first necessary data set is the quantity of tax-paid cigarettes. These data are usually available from national tax authorities and/or from national customs offices in countries where cigarettes are entirely imported.

The second data set that is needed is cigarette consumption patterns. These data are generally available through user surveys. Governments collect these data regularly. The information required is the prevalence of smoking for a certain period (for example, the month before the study) and average smoking intensity over the same or a shorter period (for example, the week before the study). It is relatively easy to access surveys containing the two variables. Often, governments archive their data. This makes estimation of historical trends feasible, although tax-paid sales data will also be necessary to estimate historical trends.

On the resources side, the gap method does not require primary data collection and can be performed using solely secondary data. Special statistical software or extensive econometric abilities are also not required. The gap method only requires simple worksheet software and basic statistical knowledge.

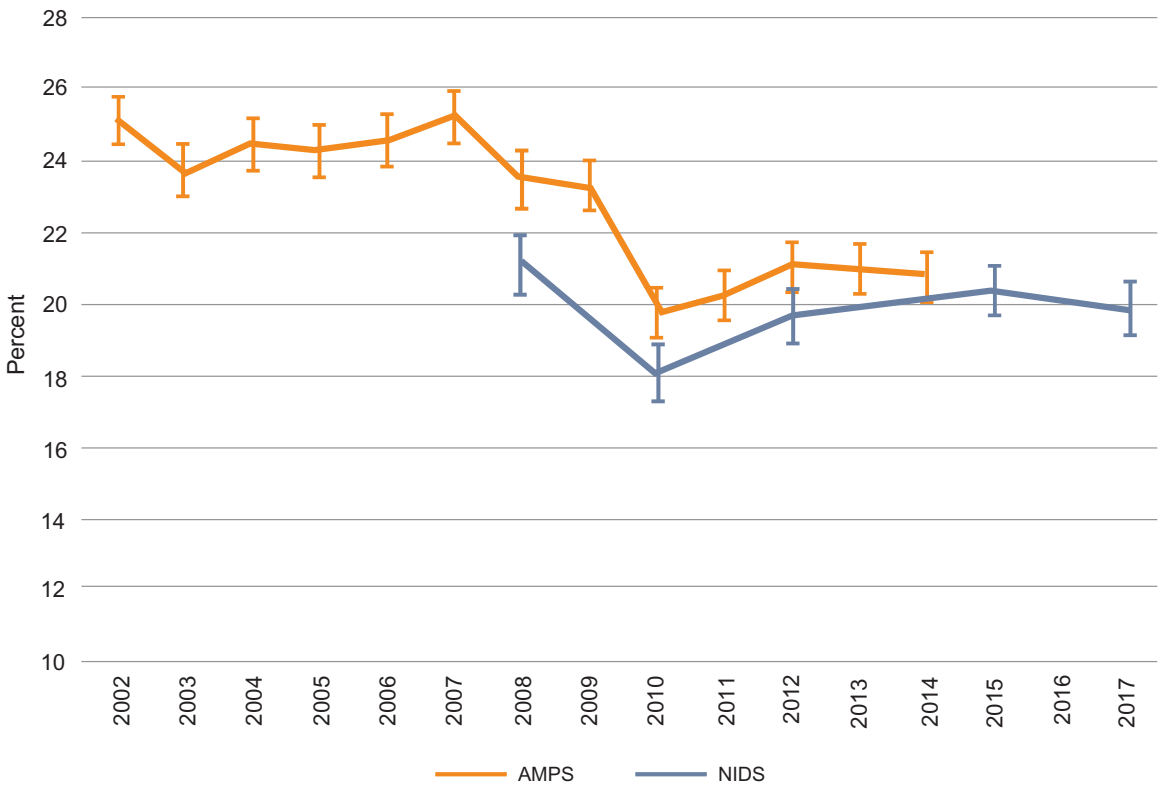
The gap method has two main weaknesses. First, the method will not provide a precise estimate of either volume or proportion. The main issue in estimation is underreporting. Underreporting is assumed to be constant over time and is itself an estimate. Both assumptions are strong ones to make. It is possible that the size of underreporting is significant. For example, in an Argentinian study, the reported cigarette quantities from general population surveys (aged 18 and older) are between 25 percent and 35 percent lower than the recorded tax-paid sales (Paraje, 2018). In the case of Brazil, they can be up to 17 percent lower (Paraje, 2018).

It is often impossible to gauge the extent of underreporting. Therefore, the gap method does not provide an exact figure or percentage of the illicit market size. If cigarette consumption underreporting can be assumed to be a constant or proportion of registered consumption, then gap analysis can estimate the cigarette consumption trends where taxes have not been paid. If such an assumption cannot be reasonably made, then gap analysis should not be used.

Second, gap analysis should only be used if the fixed underreporting proportion (and constant proportion for duty-free shopping, etc.) is realistic. Since underreporting is an unknown, a priori, the assumption should be backed by reasonable claims. The assumption is more credible when consistent and regular measurements of population cigarette consumption (for example, user surveys) are available. Consistency within the survey sampling framework and collection methods is necessary for comparison purposes. The assumption of smuggled cigarettes leaving the country must also be assumed constant. As with cigarettes smuggled into the country, this amount is rarely known.

The gap method provides the most reliable results when consistent surveys, or surveys with similar methodologies, are used over some period of time. It is not advisable to use surveys with different sampling frames, survey designs, seasonal variations, varying definitions of tobacco use, tobacco products, or ages. Using surveys that vary based on these criteria is likely to result in inconsistency or difficulty in comparing estimates over time. For example, the South African All Media and Products Survey and the National Income Dynamics Study estimate smoking prevalence in overlapping years. Figure 10 shows the results of smoking prevalence for three overlapping years (2008, 2010, and 2012). The National Income Dynamics Study consistently estimates lower smoking prevalence in the overlapping years.

Figure 10 Smoking prevalence in South Africa, 2002–2017



Source: Vellios et al., (2019)

The method should define what is being measured. The estimated total consumption (often underreported) is usually estimated from surveys on tobacco or drug use with a tobacco module. The surveys are polled from the general population or specific groups (for example, adults). It is rare that the surveys have information on cigarette origin. This information would be helpful because it would provide data on whether the smokers consume licit or illicit cigarettes. It is possible that consumers themselves are unaware of the status of cigarettes they are consuming, for example with single-stick smokers. The total amount of cigarette consumption provided includes licit and illicit cigarettes.

When cigarette consumption is available for only a non-representative subset of the population (for example, youths, females/males, inhabitants of a particular region, etc.), tax-paid data should correspond to that group (for example, tax-paid cigarettes purchased by the adult population). If tax-paid cigarette information for that subset of the population is unavailable (quite often), then gap analysis cannot be conducted.

4.4 Data sources

4.4.1 Estimating tax-paid sales using official data

Many governments publish data on cigarette sales volume and/or cigarette tax revenues. Volume data may be collected and presented directly or indirectly through measuring of fiscal marks like tax stamps.

Many other countries do not have data on sales volume and only publish data on revenues. Data on volume can easily be imputed when the system relies on specific taxes. For example, if $1 billion in revenue is received in a year and a specific tax of $0.50 per pack is applied, then the cigarette volume is two billion packs per year (1,000,000,000 / 0.50). When tiered specific tax systems are used, imputed sales data can only be derived if revenues are specified by tier. Alternatively, external market share data can be used as an alternative if it is available.

Ad valorem tax systems present challenges since the tax base may be difficult to understand or estimate. Also, revenues cannot be used to determine volume. Instead, it is possible to use the average tax base on which taxes are applied. From the tax base, the taxes collected can be used to extrapolate the number of cigarettes. Assuming that the ad valorem revenue is $1 billion, the average retail price is $1, and the average tobacco tax share is 50 percent, it would mean that the average tax base is $0.50. The ratio of the tax base and the total revenue could give an estimate of total legal cigarettes consumed. Using this approach, an especially strong assumption needs to be made. The tax base composition, in terms of brands, remains fixed over time. This can potentially be an unrealistic assumption.

Tax changes during the study period will require adjustments. Government data will specify the time coverage. If yearly data is unavailable, then researchers should be careful in imputing annual data due to seasonal fluctuations, specific stocking policies of the tobacco companies, etc.

4.4.2 Estimating tax-paid sales using external sources

In many countries, aggregate cigarette sales data are available from private sector market research firms, including Euromonitor International or GlobalData. These yearly data are usually presented as a historical time series. These data do not come without challenges, however. They cover a limited number of countries, although data on large countries is generally available. Furthermore, data vintages vary. Revisions to present and historical data suggest using the most up-to-date vintage. The data sources can vary as well

and could be sourced from government or industry data. The data can often be adjusted to fit calendar years, rather than fiscal years, or the timeframe of survey data.

Other government or international institutions may also be useful in finding external consumption data, including the U.S. Department of Agriculture and the Food and Agriculture Organization of the United Nations. However, the data source should be kept in mind when using these types of data. Consumption data may be imputed from leaf production and trade data or manufactured tobacco trade data. With these sources, as with the United Nations Comtrade database, national tax-paying cigarette consumption is estimated using apparent consumption relationship, defined as:

$$C_t = P_t - X_t + M_t$$

where C_t is consumption (tax-paying) at time t, P_t is manufactured tobacco at time t, X_t is exports at time t, and M_t is imports (tax-paying) at time t. A time-invariant conversion rate to sticks, typically measured in tons, has to be applied to the data. Researchers could assume that a stick weighs one gram. One ton of cigarettes means that there are 1,000,000 cigarettes. This assumption may not be realistic over long periods of time, however, since the quantity of tobacco in manufactured cigarettes has fallen over time.

Apparent consumption should be estimated on quantities of manufactured cigarettes. Other tobacco products such as leaves, raw tobacco, etc., are not appropriate to be used. The Harmonized Commodity description and Coding System (HS) codes for these products are: (240210) Cigars, cheroots, and cigarillos containing tobacco, including the weight of every band, wrapper, or attachment thereto; and (240220) Cigarettes containing tobacco, etc.

4.4.3 Surveys of smoking prevalence

Surveys on tobacco consumption or on drug use with a tobacco module are used to retrieve data on smoking prevalence (in a certain period, usually the month previous to the survey) and on smoking intensity. In South America, for instance, surveys ask about smoking in the month previous to the survey and the average daily smoking in the same period. In at least five countries, these surveys have been conducted at least twice, with comparable questions, allowing them to be used to make comparisons over time (Paraje, 2018).

One issue that can create challenges is the population covered by a survey. For instance, the Chilean General Population Study on Drug Use, conducted every two years, surveys people aged 12 to 65 years (Observatorio Chileno de Drogas, 2017). The Chilean School Children Study on Drug Use, also conducted every two years, collects information on tobacco use in children typically 12 to 18 years old (Observatorio Chileno de Drogas, 2015). Naturally, these two sets of surveys cannot be used jointly, as the sampling frame is different. It would be advisable to use the survey with the most comprehensive coverage of the smoking population. In this case, the General Population Survey would be preferable. This survey, however, does not cover smokers aged 65 and older. It is generally assumed that the excluded group's consumption prevalence and intensity are constant over time.

On the other hand, in the Argentinean National Risk Factor Survey, the sample framework includes the population of 18 and older (Instituto Nacional de Estadísticas y Censos, 2015). The population below age 18 is excluded. The average age for smoking onset in Argentina, however, is close to 16 (Guindon et al., 2017). It must be assumed that the 16- to 18-year-olds have identical and time-invariant consumption patterns for smoking prevalence and intensity. This assumption may seem innocuous for the population

above 65 (as they are long-term smokers, potentially less influenceable by tobacco control policies), but it could be riskier for the population below 18. To minimize risk, the surveys ought to be taken close together in time, which makes it less likely that any policy changes will take full effect and affect prevalence or intensity.

These surveys are frequently available from National Statistical Offices websites and can be freely downloaded after filling out short forms. Occasionally, surveys are available from Ministries of Health websites and/or Drug Use Observatories, as they are responsible for collecting these data. If unavailable online, it is advisable to first identify the institution implementing and collecting the survey and, second, to directly contact the responsible institution.

There are other surveys, not directly focusing on health or drug use, that include consumer smoking information. The Income and Expenditure Survey is such an example. This survey, collected every five to ten years, is used to estimate representative expenditure patterns in a number of countries. The goal is to estimate a representative basket of goods used for calculation of the General Consumption Price Index. They collect a large number of household socioeconomic characteristics. More importantly, the survey collects information on the quantity and value of goods and services consumers purchase or receive over a time period. Such periods usually vary by good/service type (less frequently consumed goods/services, such as education fees, purchase of durable goods, etc., have a longer recall period). Frequently consumed goods, such as cigarettes, have a relatively short recall period (for example, previous week, previous month, etc.). The household cigarette consumption is used to estimate the general population consumption. Sometimes these surveys do not collect information on purchased quantities but on item expenditures. To transform such information into purchased quantities, an average price for cigarettes has to be considered. The caveats for this strategy are similar to the ones for transforming ad valorem tax revenues into tax-paid cigarette sales, discussed above.

The Income and Expenditure Surveys collect data on household expenditures and individual expenses. The data are generally available as two separate data sets. To estimate population cigarette consumption, researchers should make use of the individual expenses data set, which usually contains quantities purchased. The data set on household expenditures usually contains information on expenditures for products purchased by one person (that is, the head of the household) but not necessarily consumed by him/her (for example, food, clothing, etc.), or on “public goods” within households, such as electricity, heating, housing costs, etc.

A limitation to the Income and Expenditure Surveys is that minors are likely to misrepresent their cigarette consumption. They are likely to either not report or underreport their cigarette consumption in an effort to conceal their habit from their parents.

When countries lack the human resources or funding to implement national cigarette consumption surveys themselves, international agencies may collect these data instead. The Demographic and Health Survey (DHS) is financed by the U.S. Agency for International Development (now collected by ICF). This survey has collected more than 400 surveys in 93 low- and middle-income countries since the early 1990s. The survey is administered typically every three to five years. Although not all countries are equally polled, most have at least two surveys available. The DHS primarily focuses on maternal and child health. Recently, the DHS started collecting other aspects related to household smoking. The survey incorporates information on the frequency of cigarette consumption and other tobacco products among males and females aged 15 to 49. Information for smoking intensity, however, is only available for males. Although this is a limitation, it is only a minor one as smoking prevalence is low among women in low- and middle-income countries. If it can be reasonably assumed that the smoking prevalence (and smoking intensity) is constant over time, then DHS is a viable data source.

The Global Adult Tobacco Survey (GATS) is another tobacco-consumption data source for low- and middle-income countries. The survey is funded by the U.S. Centers for Disease Control and Prevention (CDC) as part of its global tobacco surveillance system (GTSS). GATS are nationally representative household surveys for the population aged 15 years and older. The instrument has questions on the frequency and intensity of use of tobacco products. GATS data are comparable across time. Thus far, GATS data are available for 25 countries. However, not all countries have two or more surveys available. This makes gap analysis infeasible for countries that have fewer than two surveys available.

4.4.4 Population data

Government statistical agencies regularly publish national population data. Often, these data are reported annually. These data are reported in the census data and are the official country estimates. In most cases, the estimates are credible. However, caution is advised for countries known to meddle with their data for political reasons. If there are doubts about the data integrity, researchers can rely on independent estimates published by international organizations. Reliable population estimates are available and released by various United Nations agencies, including the World Bank or the United Nations Department of Economic and Social Affairs Population Division. Alternative population estimates can be used as a robustness check.

If population-level data are used in the analysis, they must be disaggregated to ensure the age structures for the population and smoking prevalence are similar. Young children, typically represented in population-level estimates, do not smoke and need to be excluded from the population-level estimates. Otherwise, the basis for comparison between the population and prevalence estimates is not comparable.

For example, if the adult smoking prevalence applies to the population aged 16 and older, the population data need to be disaggregated to account for the same age range. If youth smoking prevalence is applicable to the population aged 12 to 16, then the population data need to reflect the same age range as well.

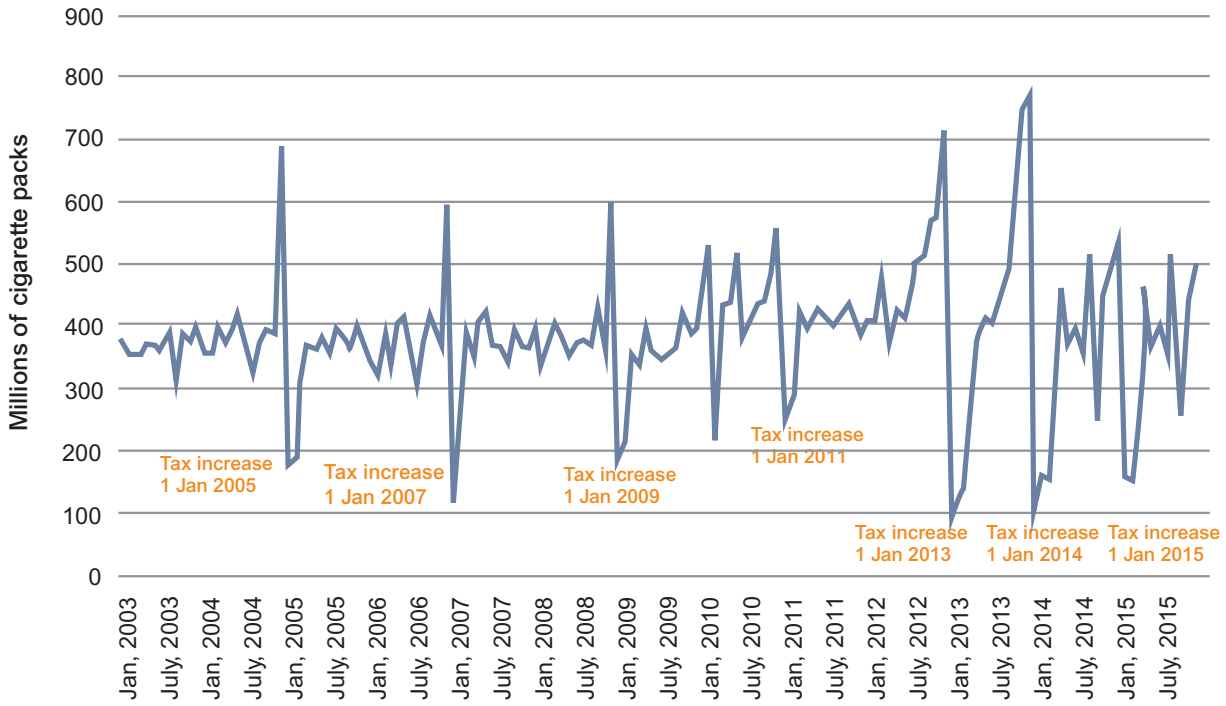
In South Africa, for example, the 2018 population estimate was 57,725,606. Youth aged one to 14 accounted for 17,043,517, suggesting that the adult population represented 40,682,089 (StatsSA, 2018). It would be incorrect to apply an adult smoking prevalence rate of 20 percent to the total population. Without disaggregation, this would suggest that there are 11,545,121 smokers. This would grossly overrepresent the number of actual adult smokers (8,136,418 smokers). If smoking intensity is 10 cigarettes per smoker per day, using both estimates, the total market size would be estimated at 42.1 billion cigarettes and 29.7 billion cigarettes, respectively. Researchers would overestimate the true market size by 12.4 billion cigarettes, a significant difference. Therefore, the population represented in data needs to be taken into consideration when using two different data sets.

4.5 Challenges with the data

4.5.1 Forestalling tax changes

The tobacco industry strategically forestalls (stockpiling or front-loading) cigarettes when a tax increase is expected or announced. The tobacco industry floods the market with cigarettes before the tax change takes effect to avoid paying the new tax. A temporary increase in sales is to be expected, followed by a sudden fall in tobacco sales post-tax implementation. Figure 11 shows this phenomenon for the Philippines.

Figure 11 Volume of cigarette removals in Philippines (domestic production)



Source: Ross et al., (2017)

Forestalling can compromise the use of the gap method, as the sudden increase (decrease) in tax-paid sales can be wrongly attributed to a decrease (increase) in illicit trade, when the fluctuation in sales is just an artificial, industry-driven process. If forestalling occurs, the gap method should be used with care. Long-term sales trends should be used to smooth erratic consumption patterns. Time-series averages, long enough to compensate for the forestalling, should be used as total market estimates. Alternatively, periods with unnatural growth/decline in sales (that is, more than 10 percent, 20 percent, etc.) can be excluded from the calculations. For the gap method, the key is the evolution of the tax-paid market, rather than actual quantities sold.

4.5.2 Inconsistency in time periods

The gap method requires that the tax-paid sales period and consumption period be the same. However, consumption estimates are generally collected for a limited period of time (for example, a specific month). Comparisons made between monthly consumption and monthly sales data may miss the effect of seasonal sales fluctuations for tax-paid cigarettes. Ideally, consumption data are collected for a year, which is rarely the case. Researchers must assume that consumption patterns are constant for the year, even if only monthly data are available. This may be a strong assumption to make, as seasonal sales effects would suggest that consumption is higher in some months than others. But without this necessary assumption, the gap method cannot be used. The sales data should be at least a year long.

4.5.3 Underreporting of smoking prevalence and intensity

Underreporting of cigarette consumption is an issue. However, it is not unusual for estimates of total consumption to be lower than tax-paid sales. This occurs because some population groups (institutionalized populations, for example) may not be covered in the sample, or that information is simply not available (duty-free sales). Without introducing bias into the sample, it can be assumed that these types of omissions are constant and can be reasonably ignored. The greater issue arises when prevalence and smoking intensity are underreported by people in the sample, because this type of omission is non-random. Prevalence underreporting occurs when smokers misreport whether they smoke. Intensity underreporting occurs when they misreport the quantity consumed in the relevant period (for example, they report a lower quantity than actually smoked).

The magnitude of underreporting is unknown. Some authors have documented its frequent existence by comparing different user data sources (that is, two different consumer surveys) (Liber and Warner, 2018; Graham and Owen, 2003) or when comparing user responses with cotinine or other chemical-level analyses (Stelmach et al., 2015). A meta-analysis found that self-reporting accuracy depends on the survey setting, study population, measurement methods, and study purpose (Patrick et al., 1994).

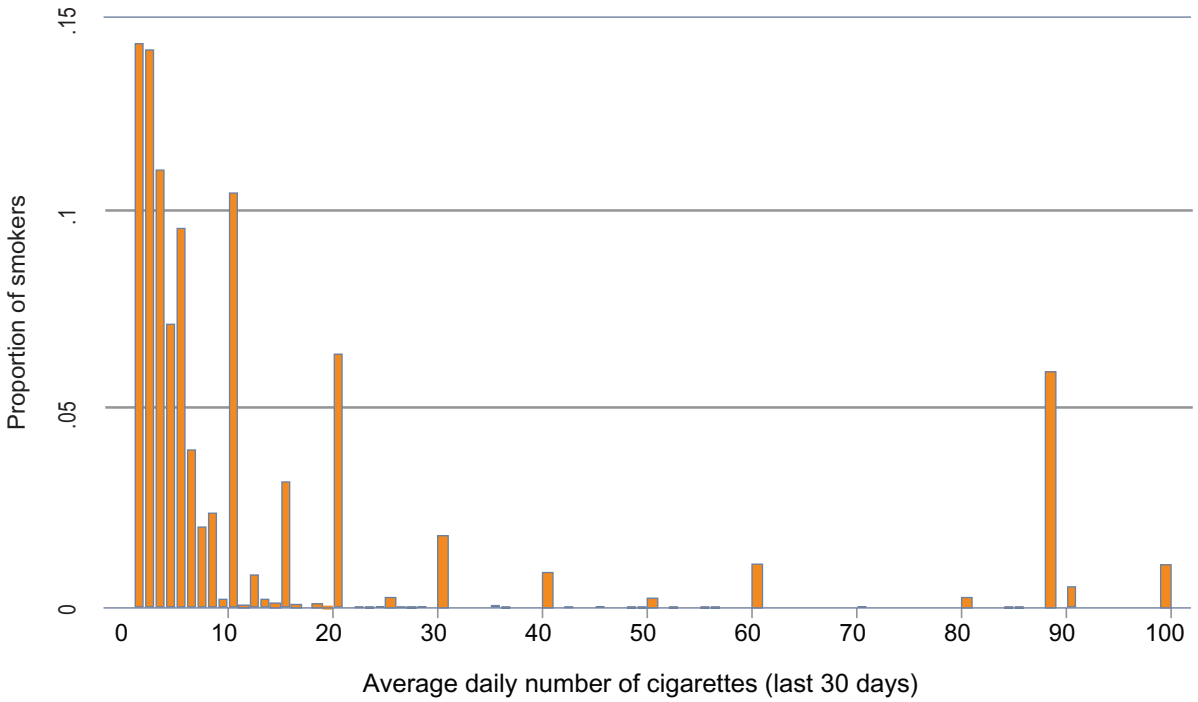
The gap method can still be used when underreporting is present. The most important aspect is how realistic the constant underreporting assumption is over time. Proving this assumption may be difficult to do as data for comparison are either scarce or nonexistent. For the U.S., it has been demonstrated that in a context of declining prevalence, underreporting of prevalence and intensity remained fairly constant across time (Liber and Warner, 2018).

4.5.4 Recall bias in smoking intensity

Smoking intensity is measured by asking survey respondents how many cigarettes they smoked in a preceding time period, usually a day or a week, while other surveys ask respondents to estimate the average number cigarettes smoked each day or week. For example, the World Health Organization Tobacco Questions for Surveys model recommends the following question: “On average, how many of the following products do you currently smoke each (day/week)?”

Respondents are likely to provide an estimate of their consumption and are often subject to underreporting (see above). Researchers are unable to validate the self-reported answer. Past research suggests that the number of cigarettes consumed tends to cluster in tens (for example, 10, 20, 30, etc., cigarettes per day). This is referred to as recall bias. A likely explanation is that cigarette consumers may think more in terms of pack sizes (20 cigarettes in a pack) than in number of consumed cigarette sticks. Smokers may also round their response to tens out of convenience. These recall biases introduce inaccuracies into smoking intensity data. Figure 12 shows the smoking intensity distribution answers for the 2016 Chilean General Population Survey on drug use (Observatorio Chileno de Drogas, 2017). Spikes are clearly visible at 10, 20, 30, and 40 cigarettes (pack sizes in Chile come in 10 and 20 cigarettes). Recall bias does not suggest that survey respondents are intentionally misrepresenting the truth but merely that consumers tend to group their answers in a way that is most convenient for recollection. Unfortunately, there is no satisfactory way to deal with this problem, though it should be properly acknowledged as a limitation of the data.

Figure 12 Histogram of smoking intensity in general population in Chile, 2016



Source: Authors' estimations from Observatorio Chileno de Drogas (2017)

4.6 Estimating the gap and presenting the data

The first step is to obtain total reported consumption from surveys. Then the gap method can be used to estimate the difference between tax-paid cigarette sales and total consumption. When population-level surveys are used, the sample structure needs to be identified first. This is important because the relative representation of each age group needs to be considered. Generally, surveys use multistage survey designs in the process of collecting data. Documentation, provided along with the survey data, usually gives detailed information on the specific sampling design used. Statistical packages, such as Stata, can handle such survey designs to make accurate inferences about the population.

The Stata command <*svyset*> declares that a survey design is being used. The command allows for the number of sampling stages, methods, and weights that were used in the survey to be defined. The strata, primary sampling units (PSU), or clusters can easily be defined. Strata are broad geographic units where PSUs are located and define the geographic representativeness of surveys. PSUs are geographic units where data are collected and that are defined and chosen, usually randomly. Sampling weights account for the original importance of each population subgroup within the total population, given the sampling structure. In other words, each observation in the sample represents many individuals/households in the population, and such a relationship is summarized by sampling weights.

For instance, the Chilean Décimo Segundo Estudio Nacional de Drogas en Población General (12th Study of Drug Use in General Population) uses a three-stage sampling design. First, it considers PSUs based on census data. Such PSUs are a group of close blocks (manzanas) within 108 urban communes in Chile. The PSU selection is probabilistic. PSUs are assigned probabilities based on the relative size of each commune.

The second stage is the selection of households within PSUs. The number of households within each PSU are assigned based on the proportion of households within the PSU. For the third stage, respondents (aged 12 to 65) are randomly selected within each household.

The command <*svyset*> has a sequence of steps. It is advisable to clear any previous design by typing <*svyset, clear*>. Second, the <*svyset*> syntax is: <*svyset psu [weight] [, design options] [|| ssu, design options] … [options]*>, where psu is the variable name identifying the primary sampling unit in the data, weight identifies the sampling weight, ssu identifies the sampling units in the second stage, and so on. Third, once the sample design is declared, the prefix <*svy:*> is required with each subsequent command to account for the sample design.

Assume a survey dataset was conducted using a two-stage design. The variable designated to indicate the clustering of households in a small geographic area is psu. The weighting variable, weight, gives the relative importance within the population of individuals selected in the second stage. Thus, the command to set the sample design is:

```
svyset psu [pw=weight]
```

Also assume that in that survey the variable q10 gathers the responses to the question "Have you smoked in the last 30 days?", with 0 = no; 1 = yes; 2 = do not know/not answering. To construct a variable that signals if the individual is a regular smoker (having smoked in the previous month) the Stata code would be:

```
gen smoker=(q10==1)
replace smoker=. if smoker > 1
```

Individuals who did not want to identify their smoking status will not be registered as either smoker or nonsmoker. To obtain the monthly smoking prevalence in the population, the following command can be used:

```
svy: mean smoker
```

This gives the proportion of the population who regularly smoke. To get the actual number of smokers in the population, the following command can be used:

```
summarize weight
. su weight
```

variable	Obs	Mean	Std. Dev.	Min	Max
weight	41,392	554.0998	1144.936	2	32745

The "Obs" column gives the number of observations in the dataset, and the "Mean" column gives the mean value of the weight in the dataset. Multiplying both columns gives the number of people in the sample. Multiplying this number by the proportion of smokers gives the number of regular smokers in the sample.

Assuming that the variable q11 gathers the responses to the question "How many cigarettes have you smoked daily in the last 30 days?" then

```
svy: mean q11 if smoker==1
```

represents the average daily intensity per smoker. Multiplying this average intensity by 365 gives the total cigarette quantity consumed in a year per smoker. (While almost every survey estimates daily intensity, it may also be presented as weekly or monthly. In these cases, researchers would multiply by 52 or 12, respectively.) In turn, multiplying the total cigarettes smoked per smoker-year by the number of smokers gives the total number of cigarettes consumed in a year.

Alternately, the code below generates a new variable:

```
summarize q11 if smoker==1 [w=weight]
gen total = r(mean)*365
```

The first line provides the average daily smoking intensity, while the second line creates a new variable with the total number of cigarettes smoked per individual-year.

In some cases, there is information on infrequent smokers that is not collected by questions like the ones presented above. It is rare that people smoking 10 cigarettes in a month report "0.33" when asked their average daily consumption. For them, there may be information on the number of days they had smoked ("In the past month, how many days have you smoked?") and on the quantity smoked per day ("On the days you have smoked during the past month, how many cigarettes on average have you smoked?"). So, for an individual who reports smoking 5 days in the past month and smoking 3 cigarettes on average in those days, the number of cigarettes smoked for him/her in a year would be 180 (3 * 5 * 12). If such information is available, these consumptions can be easily added to obtain total number of cigarettes smoked for those individuals in a year.

Assuming that the variable q12 collects information on the number of days smoked in the past month, while q13 collects information on the number of cigarettes smoked on average in those days, then the codes for estimating the total annual consumption are:

```
summarize q12 if smoker==1 [w=weight]
gen days = r(mean)
summarize q13 if smoker==1 [w=weight]
gen cigs = r(mean)
gen total = days*cigs*12
```

Survey data are generally subject to statistical errors. Confidence intervals around the total consumption estimates are advisable by bootstrapping the sample:

```
bootstrap sumboot=r(sum), reps(1000) force cformat(%9.0f): sum total [fw=weight]
```

Bootstrapping is a nonparametric method that randomly samples (with replacement) a number of subsamples from the study population (in the previous line, such a number is 1,000). One thousand artificial samples are constructed using the original data (1,000 is sufficiently large to produce reliable standard

errors). The requested statistics can be computed (number of cigarettes smoked in a population in a year) along with 95 percent confidence intervals for the statistics.

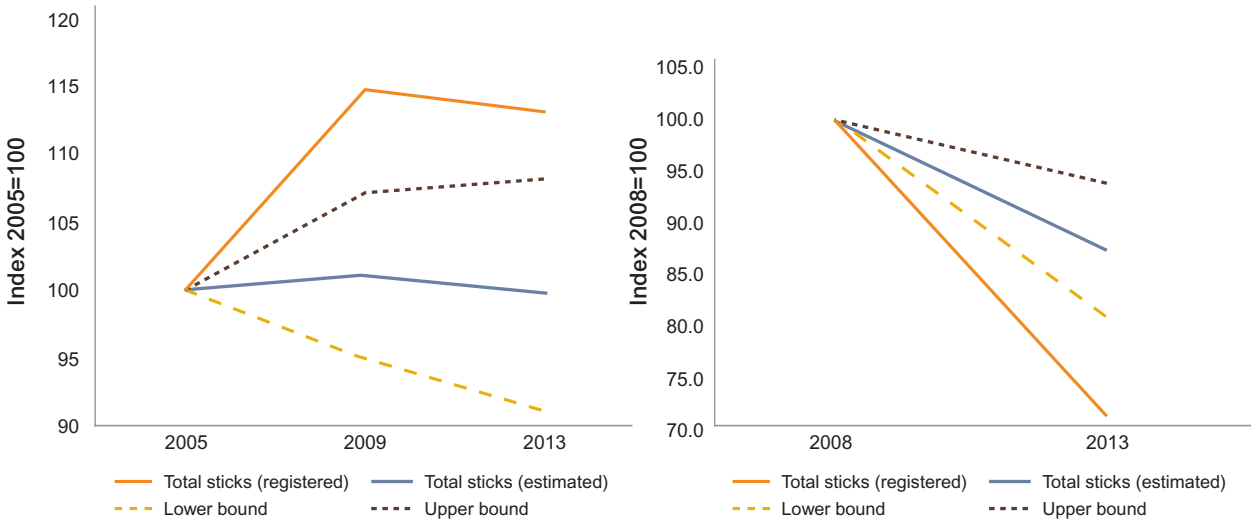
These calculations need to be done for each year in the survey. Once completed, the gap analysis can be estimated using only three data points (the total market and its 95 percent confidence interval bounds) and tax-paid cigarette sales. Due to underreporting, it is likely that reported consumption or total market sales are lower than tax-paid sales, which would be a nonsensical result (suggesting that illicit trade is negative).

The gap method estimates the evolution of the non-tax-paid market over time, where the illicit market is expected to be its main component. Thus, a base year should be defined, and the evolution of different markets can be estimated from the base year. The total market (and its confidence intervals) and the tax-paid market can be indexed to 100 for the base year. The other year's values can be imputed using the simple rule of three. Hence, if the tax-paid market for the base year is 1.5 billion cigarettes and estimated consumption for the same year is 1.3 billion, it can be assumed that both figures are equal to 100. If the following year, the tax-paid market and estimated consumption are 1.6 billion and 1.35 billion, respectively, then, in terms of the base year they are equivalent to 106.7 (1.6 billion * 100 / 1.5 billion) and 103.8 (1.35 billion * 100 / 1.3 billion), respectively.

Figure 13 shows the Brazilian and Argentinian gap analysis. For Argentina (left), the tax-paid sales evolution lies above the total market 99 percent confidence interval limits. This suggests that the licit market's growth rate was greater than the total market growth rate. The illicit market decreased in proportional terms under the time-invariant underreporting assumption. It is likely that the illicit market shrank since the smoking prevalence decreased over this period.

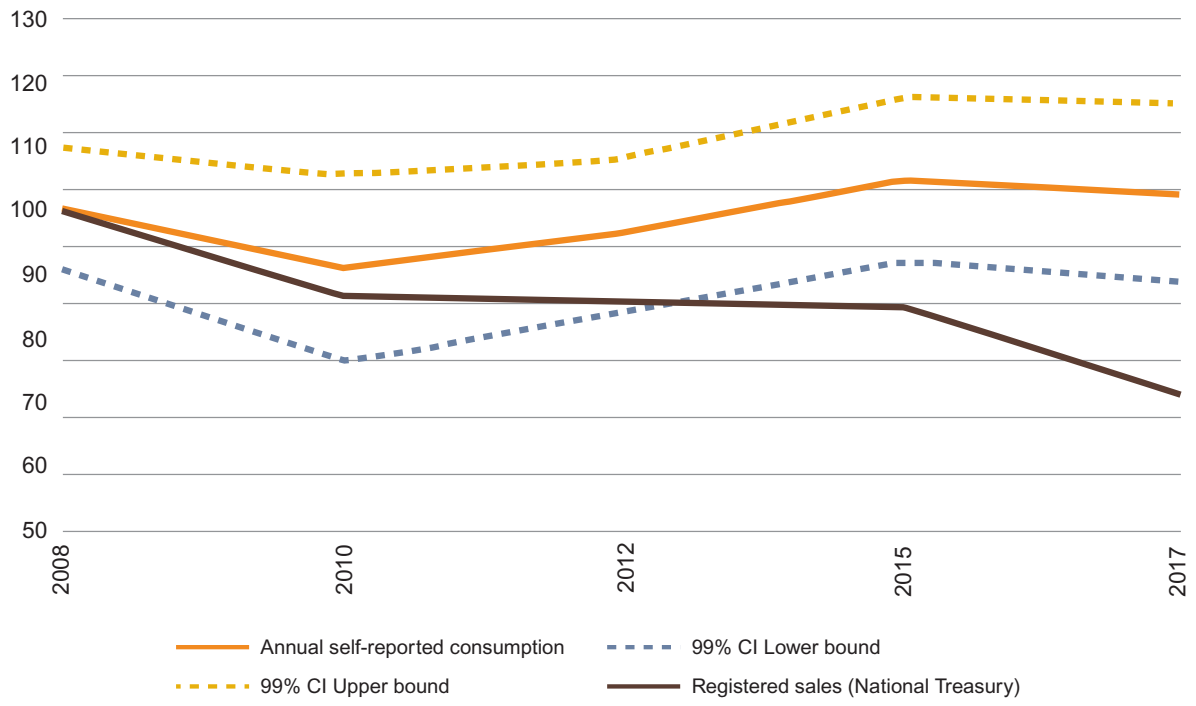
For Brazil (right), the tax-paid market's growth rate fell relative to the total market's growth rate. The tax-paid market's growth rate lies below the 99 percent confidence interval. This result implies that the non-tax-paid market grew in relative terms. The figure suggests that total cigarettes consumed decreased in the period.

Figure 13 Gap analysis for Argentina (left) and Brazil (right)



Source: Paraje (2018)

Figure 14 Gap analysis for South Africa (2002=100)



Source: Vellios et al., (2019)

However, legal sales also decreased. The relative change in the size of the illicit market could not normally be determined in such a case. In the specific case of Brazil, there are strong indications that the size of the illicit market grew, as consumption indicated by survey in 2013 was higher than tax-paid sales.

Figure 14 shows the gap analysis for South Africa. As in the case of Brazil, the registered market fell below the 99 percent confidence interval, showing an increase in the illicit market for cigarettes. Contrary to the Brazilian case, in South Africa such an increase occurred with a statistically constant smoking prevalence.

4.7 Case studies

4.7.1 The case of Canada

A Canadian study investigated contraband cigarette trends using gap analysis with a special focus on the provinces of Ontario and Quebec (Guindon et al., 2016). The gap analysis contrasts tax-paid cigarette estimates with consumption data from smoker behavior surveys. The study authors use the tobacco tax-paid cigarette numbers reported to Health Canada. In addition, the authors consider two large national surveys (the Canadian Community Health Survey (CCHS) and the Canadian Tobacco, Alcohol, and Drugs Survey (CTADS)) to estimate cigarette consumption. The authors determined the ratio of self-reported cigarette consumption from the CCHS and CTADS to tax-paid cigarette sales. Assuming that the pattern of underreporting is time invariant, if the ratio is constant (increases) it implies a constant (increased) contraband cigarette proportion. These ratios were calculated from 1999 to 2013 with biannual frequency.

The analyses show “relatively little change in the early 2000s interrupted by a clear upward jump around 2005, followed with a decreasing trend from about 2007 to 2009” (Guindon et al., 2016). The authors also found that estimates for Ontario and Quebec followed similar trends but with greater magnitudes. The researchers were able to determine which parts of the trends were significant due to the inclusion of confidence intervals.

4.7.2 Gap analyses for five Latin American countries

A recent study assessed the illicit cigarette share evolution in Argentina, Brazil, Chile, Colombia, and Peru (Paraje, 2018). Tax-paid cigarette sales were compared to consumption estimates from nationally representative surveys. Tax-paid cigarette sales data were obtained from official sources (Argentina, Brazil, and Chile) or from estimates from consulting firms (Colombia and Peru). The Argentinian used two surveys: National Survey of Risk Factors (ENFR) and the National Survey on the Prevalence of Psychoactive Substance Consumption (ENPRECOSP). The surveys have different sample frameworks. The ENFR collects data on the 18-and-older population while the ENPRECOSP collects data on the population aged 16 to 65. This study estimated two different illicit trade trends based on each data set.

The researchers found that illicit trade in Argentina decreased between 2005 and 2009 and remained constant until 2013. For Brazil, there are statistically significant changes in the difference between consumption trend estimates and tax-paid sales between 2008 and 2013. This implies an increase in the contraband cigarette share. Other Brazilian studies, using the same methodology, confirm these results (Szklo et al., 2017; Iglesias et al., 2016).

The Chilean study did not find any statistically significant difference between tax-paid cigarettes and consumption trends (population aged 12 to 65 years) for 2008 to 2014. This implies that the Chilean illicit trade share did not change over this period.

The study of Colombia did not find any statistically significant difference between tax-paid cigarettes and consumption trends (population aged 12 to 65 years) for 2008 to 2013. This implies that the Colombian illicit trade share did not change over this period.

Finally, the Peruvian study did not find statistically significant differences between tax-paid cigarettes and consumption trends (population aged 12 to 65 years) for 2006 to 2010. This implies that the Peruvian illicit trade share did not change over this period.

4.7.3 Estimating illicit trade in South Africa

A South African study estimated tax-paid cigarette sales using fiscal revenues and consumption. The study aimed to estimate prevalence and the size and evolution of illicit trade (Blecher, 2010). The study could estimate the number of tax-paid cigarettes because South Africa taxes cigarettes using a specific excise tax. Dividing the annual tobacco tax revenue by the unitary specific tax amount provides the total number of licit cigarettes. Smoking prevalence data is obtained from the All Media and Product Survey (AMPS). Smoking intensity data comes from three data sources: previous independent research, AMPS data, and private consulting firm estimates. The total smoking population is simulated using total population estimates from Statistics South Africa. By simulating different levels of consumption underreporting (from 5 percent to 10 percent of reported consumption), the author provides not only trends but figures of the illicit cigarette market penetration.

The author finds that the illicit trade market penetration grew rapidly from 1997 until 2000 at between 9.4 percent and 11.5 percent of the total market. The market declined slowly between 2000 until 2007.

In a recent study, Vellios et al. (2019) produced the illicit cigarette market trend and size for the period from 2002 to 2017. Using a similar methodology to Blecher (2010), Vellios et al. (2019) estimate cigarette consumption from AMPS (2002–2014) and the National Income Dynamics Study (NIDS) for 2008, 2010, 2012, 2015, and 2017. The authors assume that smoking underreporting in these surveys is 5 percent and 10 percent for AMPS and 15 percent and 20 percent for NIDS. NIDS consistently shows lower smoking prevalence than AMPS, which is why a higher underreporting percentage was used in these calculations. The researchers found that the illicit cigarette market increased sharply since 2009, reaching between 30 to 35 percent of the total market by 2017.

4.7.4 Measuring tax gaps in the United Kingdom

The HM Revenue & Customs (HMRC) service, the fiscal authority in the UK, regularly provides UK tax evasion estimates. These estimates are provided for tobacco and several other goods, such as alcohol and hydrocarbon oils, using gap analyses (HM Revenue & Customs, 2019). The HMRC includes cigarettes and hand-rolling tobacco in their tobacco numbers.

In the 2019 edition, cigarette estimates were calculated using the smoking prevalence and consumption estimates provided by the General Lifestyle Survey (GLF), the Opinions and Lifestyle Survey (OPN), and Health Survey for England (HSE). Adult population estimates are pulled from the Office for National Statistics (ONS). A cigarette consumption underreporting or "uplift factor" is also used. This factor "is calculated by taking estimates of total consumption from the GLF in a base year, comparing with consumption based on actual clearances to HMRC and an estimate of legitimately purchased cigarettes from abroad." Confidence intervals are estimated to account for uncertainties in the survey process. In this particular case, confidence intervals are especially important because dual smokers (those who smoke both manufactured and hand-rolled cigarettes) need to be properly accounted for. The methodology for hand-rolled tobacco follows a similar methodology to manufactured cigarettes.

The HMRC estimates that tax-paid cigarette consumption has "declined steadily from 49.5 billion cigarettes in 2005–06 to 27.5 billion cigarettes in 2017–18," and "[since] 2010–11 the central estimate of the illicit market has been fairly stable, ranging between 3.0 billion and 5.5 billion cigarettes" (HM Revenue & Customs, 2019). This implies an illicit market share of 9 percent of the total market in 2017–18.

For hand-rolling tobacco the situation is different, as "tax-paid consumption volume of hand-rolling tobacco increased steadily between 2005–06 and 2013–14 and has been relatively stable since," while "illicit market volume of hand-rolling tobacco has shown a long-term decline since 2005–06 and is estimated to be 3.1 million kg (tons) in 2017–18." Here, the illicit market represents 32 percent of the total market.

4.8 Conclusion

Gap analysis is an easy-to-implement, quick, and inexpensive method that provides estimates of the underlying trends in illicit cigarette trade. Though it cannot provide estimates of the level of the illicit market at one point in time, it allows researchers to infer whether illicit trade has increased, remained constant, or decreased over time.

In this chapter the main requirements in terms of data and challenges surrounding this method have been presented. This method is not particularly data-intensive but at least two series of data are required: tax-paid cigarette sales and consumption for at least two points of time. Consumption can be estimated using consumer surveys, usually collected by official agencies. The other key element, tax-paid sales of cigarettes, is also usually available at no or low cost.

Preparing the data and interpreting the results may imply some challenges, and most of them addressed in this chapter using empirical examples.

Overall, the method offers a rapid diagnosis of the trend of the cigarette illicit trade in a country and can be used to provide a quick, sound, easy-to-understand estimate of the trend to authorities and the general public. As with all methods, limitations have to be considered to avoid over-interpretation (for example, by giving estimates of the size of the market).

5 References

Adams, M., & Effertz, T. (2011). Tabaksteuern, Schmuggel und unversteuerte Zigaretten – Zur Glaubwürdigkeit der "Entsorgungsstudie" der Tabakindustrie [Excise Taxes on Tobacco and the Problem of Smuggling – Concerning the Credibility of the Tobacco Industry's "Discarded-Cigarette-Packages Study"]. *Gesundheitswesen*, 73, 705–12.

Barker, D.C., Wang, S., Merriman, D., et al. (2016). Estimating cigarette tax avoidance and evasion: evidence from a national sample of littered packs. *Tobacco Control*, 25, i38–43.

Blecher, E. (2010). A mountain or a molehill: is the illicit trade in cigarettes undermining tobacco control policy in South Africa? *Trends in Organized Crime*, 13(4), 299–315.

Bostoen, K., & Chalabi, Z. (2006). Optimization of household survey sampling without sample frames. *Int J Epidemiol*, 35, 751–755.

Centers for Disease Control and Prevention. (accessed 2019, June 11) *Global Tobacco Surveillance System Data*. https://www.cdc.gov/tobacco/global/gtss/gtssdata/index.html

Consroe, K., Kurti, M., Merriman, D., et al. (2016). Spring Breaks and Cigarette Tax Noncompliance: Evidence from a New York City College Sample. *Nicotine Tob Res Off J Soc Res Nicotine Tob*, 18, 1773–9.

Gallagher, A.W.A., Evans-Reeves, K.A., Hatchard, J.L., et al. (2019). Tobacco industry data on illicit tobacco trade: a systematic review of existing assessments. *Tobacco Control*, 28, 334.

Carvalho Figueiredo, V., Szklo, A.S., Iglesias R.M., et al. (In progress). Measuring illicit cigarette trade in Brazil.

Drope, J., Schluger, N., Cahn, Z., et al. (2018). *The Tobacco Atlas* (6th ed). Atlanta: American Cancer Society and Vital Strategies 2018.

European Commission. (accessed 2019, Aug 21). *Special Eurobarometer 443: Public perception of illicit tobacco trade.* Brussels, Belgium: 2016.
http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/73943

European Parliament and the Council of the European Union. (accessed 2014, Aug 28). *Directive 2014/40/EU of the European Parliament and the Council of 3 April 2014 on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products and repealing Directive 2001/37/EC.*
http://ec.europa.eu/health/tobacco/docs/dir_201440_en.pdf

European Commission. (accessed 2019, Aug 6). *Health Warnings.* https://ec.europa.eu/health/tobacco/law/pictorial_en

Fundación Interamericana del Corazón – Argentina (FIC Argentina). (Manuscript in progress). *Littered pack data to estimate illicit cigarette trade in in the city of Buenos Aires.*

Frankel, M. (1983). *Handbook of Survey Research: Sampling Theory.* New York: Academic Press, Inc.

Gilmore, A.B., Rowell, A., Gallus, S., et al. (2014). Towards a greater understanding of the illicit tobacco trade in Europe: a review of the PMI-funded "Project Star" report. *Tobacco Control,* 23, e51–61.

Gilmore, A.B., Fooks, G., Drope, J., et al. (2015). Exposing and addressing tobacco industry conduct in low-income and middle-income countries. *The Lancet*, 385, 1029–43.

Graham, H., & Owen, L. (2003). Are there socioeconomic differentials in under-reporting of smoking in pregnancy? *Tobacco Control*, 12(4), 434.

Guindon, G.E., Burkhalter, R., & Brown, K.S. (2016). Levels and trends in cigarette contraband in Canada. *Tobacco Control*.

Guindon, G.E., Paraje, G.R., & Chávez, R. (2017). Prices, inflation, and smoking onset: The case of Argentina. *Economic Inquiry.*

IARC. IARC Handbooks of Cancer Prevention, Tobacco Control, Vol. 14: Effectiveness of Tax and Price Policies for Tobacco Control. Lyon: International Agency for Research on Cancer, 2011.

Kurti, M., He, Y., Silver, D., et al. (2019). Presence of Counterfeit Marlboro Gold Packs in Licensed Retail Stores in New York City: Evidence from Test Purchases. *Nicotine Tob Res*, 21, 1131–4.

Kelley, K., Clark, B., Brown, V., et al. (2003). Good practice in the conduct and reporting of survey research. *Int J Qual Health Care*, 15, 261–6.

HM Revenue & Customs. (2019). *Measuring tax gaps 2019 edition: Tax gap estimates for 2017–18 London, UK: HM Revenue & Customs.*

Iglesias, R.M., Szklo, A.S., Souza, M.C., de Almeida, L.M. (2016). Estimating the size of illicit tobacco consumption in Brazil: findings from the global adult tobacco survey. *Tobacco Control.*

International Tobacco Control (ITC) Policy Evaluation Project. (accessed 2019, Sep 8). *About ITC.* 2017. https://itcproject.org/about

John, R.M., & Ross, H. (2018). Illicit cigarette sales in Indian cities: findings from a retail survey. *Tobacco Control*, 27, 684–8.

John, R., Chelwa, G., Vulovic, V., & Chaloupka, F. (2019). *Using Household Expenditure Surveys for Research in the Economics of Tobacco Control. A Tobacconomics Toolkit.* Chicago, IL: Tobacconomics, Health Policy Center, Institute for Health Research and Policy, University of Illinois at Chicago.

Joossens, L., Lugo, A., Vecchia, C.L., et al. (2014). Illicit cigarettes and hand-rolled tobacco in 18 European countries: a cross-sectional survey. *Tobacco Control*, 23, e17–23.

Kurti, M., He, Y., von Lampe, K., et al. (2017). Identifying counterfeit cigarette packs using ultraviolet irradiation and light microscopy. *Tobacco Control*, 26, 29–33.

Lakhdar, C.B. (2008). Quantitative and qualitative estimates of cross-border tobacco shopping and tobacco smuggling in France. *Tobacco Control*, 17, 12–6.

Liber, A.C., & Warner, K.E. (2018). Has Underreporting of Cigarette Consumption Changed Over Time? Estimates Derived From US National Health Surveillance Systems Between 1965 and 2015. *American Journal of Epidemiology*, 187(1), 113–9.

Little, M., Ross, H., Bakhturidze, G., et al. (2019). Illicit tobacco trade in Georgia: prevalence and perceptions. *Tobacco Control.*

Maas, C.J., & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86–92.

Maldonado, N., Llorente, B.A., Iglesias, R.M., et al. (2018). Measuring illicit cigarette trade in Colombia. *Tobacco Control*.

Merriman, D. (2010). The Micro-Geography of Tax Avoidance: Evidence from Littered Cigarette Packs in Chicago. *Am Econ J Econ Policy*, 2, 61–84.

Ministry of Finance in Poland. (accessed 2019, Aug 4). *Wzory banderol podatkowych i legalizacyjnych.* https://www.podatki.gov.pl/media/2993/fsr-win-003-mf-gov-pl-folders_citrix-fqgv-pulpit-wzory-banderol.pdf

Observatorio Chileno de Drogas. (2017). *Décimo Segundo Estudio Nacional de Drogas en Población General.* Santiago de Chile, Chile.

Observatorio Chileno de Drogas. (2015). *Estudios nacionales de drogas en población escolar de Chile 2015.* Instituto Nacional de Estadísticas y Censos. Tercera Encuesta Nacional de Factores de Riesgo para Enfermedades No Transmisible. Buenos Aires, Argentina. http://www.senda.gob.cl/observatorio/estudios/poblacion-escolar/

Paraje, G., Araya, D., & Drope, J. (2018, Dec 15). Illicit cigarette trade in Metropolitan Santiago de Chile. *Tobacco Control.*

Paraje, G. (2018). Illicit Cigarette Trade in Five South American Countries: A Gap Analysis for Argentina, Brazil, Chile, Colombia and Peru. *Nicotine & Tobacco Research.*

Patrick, D.L., Cheadle, A., Thompson, D.C., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: a review and meta-analysis. *American Journal of Public Health*, 84(7), 1086–93.

Philip Morris International. (2019) *Investor information*, July 2019.

Pizarro, M.E., Giacobone, G., Shammah, C., & Stoklosa, M. (In review). *Empty Pack Survey to Estimate Illicit Tobacco Trade in the City of Buenos Aires, Argentina.*

Ross, H. & Blecher, E. (2019). *Illicit Trade in Tobacco Products Need Not Hinder Tobacco Tax Policy Reforms and Increases. Tobacconomics White Paper.* Chicago, IL: *Tobacconomics, Health Policy Center, Institute for Health Research and Policy, University of Illinois at Chicago.*

Ross, H. (2015). *Understanding and measuring tax avoidance and evasion: A methodological guide.* Washington, DC. http://www.tobaccoecon.uct.ac.za/sites/default/files/image_tool/images/405/Publications/reports/Understanding-and-measuring-tax-avoidance-and-evasion-A-methodological-guide1.pdf.

Ross, H., Tesche, J., & Vellios, N. (2017) Undermining Government Tax Policies: Common strategies employed by the tobacco industry in response to tobacco tax increases. *Preventive Medicine*, 105, S19–S22.

Ross,H., Vellios, N., Batmunkh, T., et al. (2019). Impact of tax increases on illicit cigarette trade in Mongolia. *Tobacco Control*.

Rowell, A., Evans-Reeves, K., Gilmore, A.B. (2014). Tobacco industry manipulation of data on and press coverage of the illicit tobacco trade in the UK. *Tobacco Control*, 23, e35–43.

Sejm of the Republic of Poland. (2019). *Ustawa z dnia 9 listopada 1995 r. o ochronie zdrowia przed następstwami używania tytoniu i wyrobów tytoniowych.*

Sejm of the Republic of Poland. (accessed 2019, Jan 20). *Ustawa z dnia 6 grudnia 2008 r. o podatku akcyzowym. Dz. U. z 2009 r. nr 3, poz. 11. 2008.* http://prawo.sejm.gov.pl/isap.nsf/download.xsp/WDU20090030011/O/D20090011.pdf

Sáenz de Miera Juárez, B., Reynales Shigematsu, L., Stoklosa, M., et al. (2020, March 5). Measuring the illicit cigarette market in Mexico: a cross-validation of two methodologies. *Tobacco Control.*

StatsSA. (2018). *Mid-year population estimates.* Pretoria: Statistics South Africa.

Stelmach, R., Fernandes, F.L.A., Carvalho-Pinto, R.M., Athanazio, R.A., Rached, S.Z., Prado, G.F., et al. (2015). Comparison between objective measures of smoking and self-reported smoking status in patients with asthma or COPD: are our patients telling us the truth? *Jornal brasileiro de pneumologia: publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia*, 41(2), 124–32.

Stoklosa, M., & Ross, H. (2014). Contrasting academic and tobacco industry estimates of illicit cigarette trade: evidence from Warsaw, Poland. *Tobacco Control*, 23, e30–4.

Sudman, S. (1983). *Handbook of Survey Research: Applied Sampling.* New York: Academic Press, Inc.

Szklo, A., Iglesias, R.M., Carvalho de Souza, M., Szklo, M., & Maria de Almeida, L. (2017). Trends in Illicit Cigarette Use in Brazil Estimated From Legal Sales, 2012–2016. *American Journal of Public Health*, 108(2), 265–9.

United Nations. (accessed 2019, Aug 4). *Protocol to Eliminate Illicit Trade in Tobacco Products.* https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IX-4-a&chapter=9&lang=en

U.S. National Cancer Institute and World Health Organization. (2016). *The Economics of Tobacco and Tobacco Control. National Cancer Institute Tobacco Control Monograph 21*. NIH Publication No. 16-CA-8029A. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, and World Health Organization.

Vellios, N., Van Walbeek, C., & Ross, H. (2019) Illicit cigarette trade in South Africa: 2002–2017. *Tobacco Control.*

World Health Organization (WHO). (2013) *Protocol to Eliminate Illicit Trade in Tobacco Products.* Geneva: World Health Organization.

World Health Organization (WHO). (2010) *WHO Technical Manual on Tobacco Tax Administration.* Geneva: World Health Organization.

World Bank. (2018) *Economics of Tobacco Taxation Toolkit.* Washington: World Bank.

Zubovic, J., Jovanovic, O., Djukic, M., Jolovic, N., Vladisavljevic, M. (forthcoming). *Study on Tobacco Consumption in Serbia 2019*, Institute of Economic Sciences, Belgrade, Serbia.

Methodology and Code Appendices

6

The Methodology and Code Appendices build upon the Stata background discussed in the previous toolkit's appendix. The individual Stata commands are placed in angle brackets < > and are italicized. This is for illustrative purposes only. The command itself must be used without brackets for the code to run in Stata. The variable names used in the examples are italicized.

6.1 Selection of high-traffic congregation points in Santiago de Chile

The Santiago de Chile study selected high-traffic congregation points using data on foot traffic in 767 high-traffic areas. These high-traffic areas were obtained from the Survey of Origin-Destination in the Greater Santiago Area, while the data on the built surface (in square meters) by infrastructure type (for example, commercial buildings, sport facilities, schools, universities, office buildings, health care facilities, parks, etc.) was obtained from the registry of built surfaces at the Inland Revenue Service. These two data sets were merged to estimate the relationship between infrastructure type and foot traffic in the city. The following ordinary least squares (OLS) model was utilized:

$$Ni = \beta 1Ci + \beta 2Si + \beta 3Ei + \beta 4Oi + \beta 5Qi + \beta 6Hi + \beta 7Ti + \beta 8Gi + \varepsilon i$$

where Ni is the number of visits to area **i** (data from the Survey of Origin-Destination), while the independent variables are the square meter areas of commercial buildings (Ci), sport facilities (Si), educational and cultural facilities (Ei), office buildings (Oi), churches and cultural places (Qi), health care facilities (Hi), transportation and telecommunications (Ti), and parks (Gi) (data from the registry of built surfaces). This model's coefficient estimates were used to predict the foot traffic in all 66,467 blocks represented in the registry of built surfaces in Santiago de Chile. This predicted foot traffic volume allowed the researchers to identify 424 congregation points—blocks with exceptionally high foot traffic. Forty of 424 congregation points were randomly selected with weighted probabilities proportional to foot traffic in the area. Congregation points with higher traffic were more likely to be selected into the sample.

6.2 Sample survey instrument

INDIVIDUAL QUESTIONNAIRE

A1 Location:

A2 Census sector number:

A3 Street Name:

A4 Code of the place:

A5 Start time of the interview:

A6 Start time of the interview in the census sector:

A7 Full address:

Approach the home and ask the first person contacted.

Hello, my name is _____ and I am making this contact on behalf of an international team of health researchers from _____ Institute. This study is sponsored by the American Cancer Society. We are conducting a survey on smoking. I would like to ask some questions to a person aged 18 or over to see if someone from that household would be eligible for our research, would that be possible? I will only take a few minutes of your time and all information will be strictly confidential.

B1 How many smokers 18 years of age or older live in your household?

|__| |__|

88 Refused to respond

99 Do not know

Are any of these smokers at home at the moment? (If yes) Can I talk to him / her? (If it's the person himself) Can I talk to you?

When identifying the smoker who completes the quota, submit the free and informed consent form and start the interview afterwards.

Personal characteristics

B2 Sex:

1 |__| Male 2 |__| Female

B3 What year were you born? Register in 4 digit

|__| |__| |__| |__|

8888 Refused to respond
9999 Do not know

B4 What month were you born?

|__|__|

01 January
02 February
03 March
04 April
05 May
06 June
07 July
08 August
09 September
10 October
11 November
12 December
888 Refused to respond
999 Do not know

B5 What day were you born?

|__|__|

88 Refused to respond

99 Do not know

B6 What is the last grade and level of education that you completed?

|__|__|

1.0 Never attended school regularly

Elementary School

2.0 Kindergarten
2.1 1st grade
2.2 2nd grade
2.3 2rd grade
2.4 4th grade
2.5 5th grade
2.6 6th grade

Middle School

2.7 7th grade
2.8 8th grade

High School

3.1 9th grade, (Freshman)
3.2 10th grade, (Sophomore)
3.3 11th grade, (Junior)
3.4 12th grade, (Senior)
4.1 Technical or scientific high school, 1 year
4.2 Technical or scientific high school, 2 years
4.3 Technical or scientific high school, 3 years

Higher Education

5.1 Bachelor's degree, 1 year
5.2 Bachelor's degree, 2 years
5.3 Bachelor's degree, 3 years
5.4 Bachelor's degree, 4 years
5.5 Bachelor's degree, 5 years
5.6 Bachelor's degree, 6 years
5.7 Bachelor's degree, 7 years
5.8 Bachelor's degree, 8 years
6.0 Post-graduation (specialization, master's, doctorate)
88 Refused to respond

SMOKING

Now I'm going to ask you about your smoking of factory-made cigarettes. Please do not include marijuana cigarettes, hand-rolled cigarettes, and electronic cigarettes in your answers.

B7 Currently, do you smoke any tobacco product?

1. |__| Yes, daily
2. |__| Yes, less than daily
8. |__| Refusal
9. |__| Do not know

If 8 or 9, complete the questionnaire

If 1 or 2, move on

B8 On average, how many cigarettes do you smoke per day or per week?

1. |__| One or more, daily. How many per day (B9)? |___|___|
2. |__| One or more, weekly. How many per week (B10)? |___|___|
3. |__| Less than once a week. How many per month (B11)? |___|___|
4. |__| Less than one a month. How many per year (B12)? |___|___|
8. |__| Refusal
9. |__| Do not know

CIGARETTES OF THE LAST PURCHASE

The next questions are the last time you bought ***factory-made cigarettes*** for your own consumption.

B9 The last time you bought cigarettes for your own use, how many cigarettes did you buy?

(Enumerator: Record quantity and, when necessary, record unit details)

UNIT (20)	QUANTITY	UNIT DETAIL	
1 |__| Cigarettes	|__|__|__| (B10) units		
2 |__| Packs	|__|__|__| (B11) units	How many cigarettes were there in the pack? cigarettes	|__|__|__| (B12)
3 |__| Cartons	|__|__|__| (B13) units	How many packs were there, inside the carton? packs	|__|__|__| (B14)

4 |__| I never bought cigarettes

B10 In total, how much did you pay for it?

$ |__|__|__|__|__|__|__|__|__|__|__|__|__|

B11 The last time you bought cigarettes for your own use, which brand of cigarettes did you buy?

_____ |__|__|__| (brand code) (B17)

B12 Is that your regular brand?

1. |__| Yes
2. |__| No

If 2, continue to B19, otherwise move on to B21.

B13 What is your regular brand?

_____ |__|__|__| (brand code) (B20)

B14 The last time you bought cigarettes for yourself, where did you buy them?

(Enumerator: expect spontaneous response)

1. |__| Grocery store
2. |__| Supermarket
3. |__| Street vendor
4. |__| Pharmacy
5. |__| Newsstand
6. |__| Duty free (Duty-free shop)
7. |__| Out of the country
8. |__| Internet
9. |__| Vending machine
10. |__| Other _____

99. |__| No responding

FEATURES OF CIGARETTE PACK

It is very important for the study to know and describe some characteristics of cigarette packs consumed by all smokers participating in the study. I would like to see your pack of cigarettes, write down its characteristics, and photograph it. Do you have a pack of cigarettes with you? Can you show it to me?

(When shown the pack) Is this the cigarette pack from your last purchase? (If it is not the last-purchased pack) Would you have the last-purchased pack with you? Can you show it to me? Can I take a picture of this pack?

C1 The informant allowed:

1. |__| View and photograph of the pack
2. |__| Pack view only
3. |__| Only pack's photo

C2 Is the packet viewed / photographed the one of the last purchase?

1. |__| Yes
2. |__| No

C3 Brand Coding:

|__|__|__|

999: Brand of cigarettes not registered with national authorities

If 999 continue to C4, otherwise move on to C5

C4 Brand Name: _____

C5 Is there an official tax stamp visible on the packet?

1. |__| Yes, the pack has an official stamp visible
2. |__| The pack has a stamp visible, but this stamp is of another country
3. |__| It has fragments of the tax stamp, and the origin of the stamp is hard to define
4. |__| No, there is no stamp and no fragment of the stamp visible

C6 Are there any health warning images on the pack?

1. |__| Yes

2. |__| No

If 1, continue to C7
If 2, move to C10

C7 What is the language of the text message that accompanies the image warning?

1. |__| English

2. |__| Spanish

3. |__| Portuguese

4. |__| Other

C8 Code of the image warning:

|__|__|

99 A different image.

C9 Size of the warning (image + text) in relation to the face in which it is exposed?

1. |__| Less than 30%

2. |__| 30%

3. |__| Greater than 30%

Mark if any of the following items have present in the pack:

C10 Light, ultralite

1. |__| Yes

2. |__| No

C11 Name of the country the pack is intended for:

1. |__| Yes If yes, what country _____

2. |__| No

6.3 Sample exemption application submitted to the Institutional Review Board

Exemption Screening Questions

If you answer 'Yes' to any of the questions A through D below, then STOP and use the application form for initial IRB review.

If you answer 'No' to all questions A-D below, continue to complete this claim of exemption application.

Important: Please include a completed screening form with your application

A. For research involving special populations, interventions or manipulations

1. Does your research involve pregnant women, fetuses, or prisoners? ❑ Yes ❑ No
2. Does your research involve using survey or interview procedures with children? ❑ Yes ❑ No
3. Does your research involve the observation of children in settings where the investigator(s) will participate in the activities being observed? ❑ Yes ❑ No

B. For research using survey procedures, interview procedures, observational procedures and questionnaires

1. If data are to be recorded by audiotape or videotape is there potential harm[1] to subjects if the information is revealed or disclosed? ❑ Yes ❑ No
2. If the subjects are to be identifiable either by name or through demographic data, is there potential harm to participants if the information is revealed? ❑ Yes ❑ No
3. Will data collection include sensitive information (for example, illegal activities, or sensitive themes such as sexual orientation, sexual behavior, undesirable work behavior, or other data that may be painful or very embarrassing to reveal, such as death of a family member, memories of physical abuse?) ❑ Yes ❑ No

C. For research using existing[2] or archived data, documents, records, or specimens only

1. Will any data, documents, records or specimens be collected from subjects after the submission of this application? ❑ Yes ❑ No
2. If the data, documents, records, or specimens are originally labeled in such a manner that subjects can be identified, directly or indirectly through identifying links, is the investigator recording the data for the purposes of this research in such a manner that subjects can be identified, directly or indirectly through identifying links (that is, demographic information that might reasonably lead to the identification of individual subjects – name, phone number; or any code number that can be used to link the investigator's data to the source record – medical record number or hospital admission number? ❑ Yes ❑ No

D. For research using protected health information

1. Will the research involve the use or disclosure of individually identifiable health information including: names, dates (other than years), telephone numbers, fax numbers, electronic e-mail addresses, social security numbers, medical records numbers, health plan beneficiary numbers, account numbers, certificate/license numbers, device identifiers and serial numbers, web URLs, internet addresses, biometric identifiers, full face or comparable images, or any unique identifying number, characteristic, or code? ❑ Yes ❑ No

[1] Harm to subjects means that any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or can be damaging to the subjects' financial standing, employability, or reputation.
[2] Existing means the items exist before the research was proposed or was collected prior to the research for a purpose other than the proposed research.

6.4 Sampling of geographic regions using predicted smoking prevalence probability proportional to size (PPS) method in Brazil

The researchers in the Brazilian study began their geographical region sampling by investigating the number of smokers that live in each census tract. Their littered pack collection was implemented based on these results. The researchers used data from the National Health Survey to estimate smoking probability by socioeconomic characteristics. These results gave the researchers the approximate number of smokers in each census tract.

The Stata code used is below. The procedure starts by declaring the National Health Survey design. It identifies the survey's primary sampling unit variable and weight variable. Here, the variable names were psu and weight, respectively.

```
svyset psu [pweight=weight]
```

The variables included in the model are:

- fumprodind: manufactured cigarette use (0: no, 1: yes); this is the dependent variable in the model
- idadetest: age group (0: 18–24yo, 1: 25–34, 2: 35–44, 3: 45–54, 4: 55–64, 5: 65+)
- sexo: sex (0: male, 1: female)
- renda_total: average monthly income of the household owner

The following logistics regression was used to estimate the relationship between smoking and socioeconomic characteristics:

```
xi:svy:logit fumprodind i.sexo i.idadetest renda_total
```

where <*xi*> allows for the inclusion of categorical variables (for example, age) in the model, while <*svy*> adjusts the results of the previously defined survey weights. The model uses categorized age variables. This is good practice, because the relationship between smoking and age is

usually not linear. If income is not an available variable in either the survey or in the individual PSU data, other variables such as education can serve as a proxy for income.

The model's estimated results can be used to predict the smoking prevalence in each census tract using the Stata <*margins*> option and the adult population proportions with each characteristic within each census tract:

```
margins, at(_Isexo_1="VALUE=PROPORTION OF FEMALES"
_Iidadetest_1="VALUE=PROPORTION OF 25-34"
_Iidadetest_2="VALUE=PROPORTION OF 35-44"
_Iidadetest_3="VALUE=PROPORTION OF 45-54"
_Iidadetest_4="VALUE=PROPORTION OF 55-64"
_Iidadetest_5="VALUE=PROPORTION OF 65+"
renda_total ="VALUE= AVERAGE MONTHLY INCOME OF THE HOUSEHOLD
OWNER")
```

Assume that 60 percent of the population is female, based on the census tract distributed such that 32 percent are 25 to 34 years old, 23 percent are 35 to 44 years old, 10 percent are 45 to 54 years old, 5 percent are 55 to 64 years old and 4 percent are age 65 or older. The household owner's average monthly income is 3800 Brazilian Real. The command to predict the smoking prevalence in that census tract is:

```
margins, at(_Isexo_1=0.6 _Iidadetest_1=0.32 _Iidadetest_2=0.23 _Iidadetest_3=0.1
_Iidadetest_4=0.05 _Iidadetest_5=0.04 renda_total=3800)
```

The number of smokers in each census tract is the product of the estimated adult smoking prevalence in each PSU and the total number of adults in the PSU.

The final stage in the sampling process requires the random drawing of the PSUs. The probability of being selected is proportional to the number of smokers in each PSU relative to the total. The <*gsample*> Stata function is used for sampling with probabilities proportional to size (PPS). If the function is not already installed, the following command can be used:

```
ssc install gsample
```

The <*gsample*> function will require <*moremata*> package to run. The following command will install the package:

```
ssc install moremata
```

The PSU dataset needs to be arranged in long format; each row is a different PSU. The pop variable represents the number of smokers in each PSU. The sampling weight variable is created with the following commands:

```
egen popsum=sum(pop)
gen sweight= (pop/popsum)
```

The <*gsample*> function and the previously calculated weights are used to draw a sample of n PSUs without replacement. In the example below, n = 50.

```
gsample 50 [aw=sweight], wor
```

6.5 Simple random sampling

Unlike the PPS presented in the example above, Stata does not require a new procedure to execute a simple random sampling. Assume that a full list of PSUs in a given city is provided. If researchers need to select 5 percent of the PSUs for the sample, the Stata command below will give 5 percent of the sample:

```
sample 5
```

If the list contains PSUs in different cities, with the cities identified by city variable, then researchers can draw a simple random sample of 5 percent of PSUs per city using the following command:

```
sort city
by city: sample 5
```

6.6 Multivariate weighting – Chicago example

Assume that, similar to Merriman (2010), researchers want to draw a weighted random sample of geographical regions. Areas with a larger population (100 percent weight) and higher employment levels (50 percent weight) have a higher chance of being selected into the sample. The idea is to mimic the average smoker's smoking behavior, such that two-thirds of cigarettes are discarded at home and one-third at work.

First, a database needs to be created containing the entire universe of geographical regions. The database needs to contain the following variables:

- id: id number of a given geographical region
- pop: population of smokers living in that geographical region
- emp: population of smokers working in that geographical region

The following commands can be used to generate weights for the sampling:

```
egen popsum=sum(pop)
egen empsum=sum(emp)
gen weight= (pop/popsum)+0.5*(emp/empsum)
```

The <*gsample*> function and the survey weights can be used to draw a sample of n geographical regions without replacement. In the example below, n=23:

```
gsample 23 [aw=weight], wor
```

6.7 Imputing compliance status for smokers who do not show packs to enumerators: logistic regression approach

In the created dataset, variable compl indicates whether the pack shown by the smoker was tax-paid or not. The values of the variable are missing for smokers who did not show their packs. The researchers determined that the following variables are good predictors of the compl variable. Logistic regression of compl on those variables yields high Pseudo-R2 and low values of AIC and BIC.

- compl: compliant (tax-paid) cigarettes (0: no, 1: yes)
- age: age group (0: 18–24yo, 1: 25–34, 2: 35–44, 3: 45–54, 4: 55–64, 5: 65+)
- sex: sex (0: male, 1: female)
- brand: brand of cigarettes (values 0 to 10)

The tax compliance probability for smokers who did not show their packs can be predicted by:

```
xi: logit compl i.age i.sex i.brand
predict pred, p
```

To assign either smoker compliance or noncompliance with missing compl data, a cutoff point for the pred variable should be chosen. The standard cutoff point is 0.5:

```
gen compl2 = 0
replace compl2=1 if pred>0.5
```

The actual and imputed compliance for the observations with non-missing values for compl can be compared:

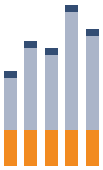
```
tab compl compl2
```

If needed, other cutoff points can be chosen. Finally, the imputed values of compliance should be assigned to the observations with missing compl values:

```
replace compl=compl2 if compl==.
```

INSTITUTE FOR
HEALTH RESEARCH
AND POLICY

UIC



tobacconomics

Economic Research Informing
Tobacco Control Policy

www.tobacconomics.org
@tobacconomics