Descriptive evaluation of the leading methodologies to estimate the scale of illicit tobacco trade

Indrek Saar and Petr Janský

Abstract

The illicit tobacco trade is a large-scale worldwide phenomenon that can be associated with many negative consequences, such as criminal activity, lost tax revenues and higher prevalence of tobacco-related diseases. Customs authorities responsible for tackling illicit tobacco trade face the difficulties of monitoring and measuring the illicit tobacco market and mostly rely on the seizure data or studies funded by tobacco industry. To provide recommendations about an appropriate methodology to measure the scale of illicit trade, we performed qualitative descriptive evaluations of the two leading approaches: survey-based and trade discrepancy methods. We used the existing literature and basic economic identities to develop a straightforward economic accounting framework. Using this framework, we assess both methods and evaluate them against the criteria of reliability, validity, feasibility and country coverage. While the trade discrepancy approach has its strengths, we find that across these four criteria the survey-based approach performs better and should be preferred.

1. Introduction

Illicit trade in tobacco products (ITTP) is a worldwide phenomenon that can be associated with negative consequences that include criminal activity, lost tax revenues and a higher prevalence of tobacco-related diseases. Large amounts of cigarettes and other tobacco products are seized both at the borders and inland by customs authorities in many countries. One of the key observations from our participation in the illicit trade in tobacco products in the European Union project, carried out in 2018–2019 by the Secretariat of the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC), was that seizure data is the main source of information for customs authorities to monitor the scale of the illicit market, combined with estimates from studies funded by the tobacco industry. This became clear from the interviews with customs officials during the study visits (World Health Organization [WHO] Framework Convention on Tobacco Control [FCTC], 2020). The lack of reliable estimates complicates decision-making at both the political and operational levels. There are, however, several analytical, independent studies that have been carried out that attempted to provide an approximation of the importance of ITTP through various methods. Not surprisingly, given its illicit nature, the studies vary in the estimated scale of ITTP, but as much as 12 per cent of the global cigarette market is thought to be illicit, with the associated loss of at least US$40 billion in tax revenues (Joossens et al., 2009). Such estimates, if reliable and feasible, assist with understanding the size of the problem and with monitoring the effects of policy interventions.

There are various methods that can be used by customs authorities or others to measure the scale of illicit tobacco trade. Relying on some review studies (Gallagher et al., 2018; National Research Council, 2015; Yurekli et al., 2001) as well as more specific recent analyses (Barrera et al., 2019; Aldridge & Décary-Hétu, 2015), we can identify five broad methodological categories: expert surveys, web-crawling approaches, modelling and econometric approaches, survey-based approaches and trade discrepancy methods.

Expert surveys can generate quite accurate estimations; however, their quality varies significantly from year to year due to the changing knowledge and experience of experts (Yurekli et al., 2001). In addition, expert surveys tend to provide qualitative observations rather than quantitative estimates (Agency for Research on Cancer, 2008, p. 140; see also Kupka & Tvrdá, 2016) and are typically used in combination with other methods (see Ajmal & Veng, 2015 for one example).

Web-crawling approaches involve the use of software to index all content in a website, then extract relevant information such as exact product names, vendors, prices, destinations and sources of the product, and even customer feedback, storing them in databases that can be searched and analysed (Barrera et al., 2019; Aldridge & Décary-Hétu, 2015; Thelwall, 2001). Since this approach can measure only the trade taking place or initiated through online platforms, it is not a valid method to address the entire market unless the internet becomes the main platform for ITTP.

There are also various modelling or econometric approaches. For example, the impact of incentives for smuggling could be modelled by examining the effects of price differences between neighbouring countries on sales and/or consumption of tobacco products (see Goel & Saunoris, 2019; Prieger & Kulick, 2018; Yürekli & Sayginsoy, 2010; Goel, 2008; Yurekli et al., 2001). Due to its complexity, the uncertain quality of data and analytical proficiency (Gallagher et al., 2018), the feasibility of modelling or econometric approaches is rather low.

The remaining approaches can be grouped into two categories: those based on consumer surveys and those relying on discrepancies in international trade statistics. In terms of ease of implementation and feasibility, these two approaches are often the most practical and have been frequently adopted in the literature. For this reason, the remainder of the current paper will focus on the two latter methodological categories.

Methodological approaches to estimating ITTP vary in reliability, validity and feasibility, as well as in which countries they are available. We used these four criteria to evaluate consumer survey-based and trade discrepancy approaches to estimate the scale of ITTP. It is widely recognised that no currently available method for assessing ITTP is flawless (Gallagher et al., 2018, p. 9), and our evaluation suggests which methods might most reliably track ITTP over time in multiple countries.

The rest of the paper is structured in the following way: section 2 outlines the evaluation methods used, section 3 develops a simple framework for economic accounting of illicit tobacco trade that helps to capture the strengths and weaknesses of the various measurement approaches, while sections 4 and 5 analytically describe these two leading methodological approaches to estimate the scale of illicit tobacco trade. Section 6 provides a discussion and concludes the paper.

2. Methods

To understand the merit of consumer survey-based and trade discrepancy methods, we performed a qualitative descriptive evaluation. Specifically, we described each method based on the information we collected from prior studies that have employed or reviewed these methods. Our descriptions compare the characteristics, weaknesses and strengths of each method against four criteria: reliability, validity, feasibility and country coverage. The criteria were selected in relation to their suitability for evaluating the impact of initiatives such as the WHO Protocol to Eliminate Illicit Trade in Tobacco Products (WHO, 2013) and on target 16.4 of the United Nations Sustainable Development Goals (SDGs) to reduce illicit financial flows, which includes illicit tobacco trade. For such policy purposes, reliability and validity are not sufficient criteria, and feasibility and country coverage should also play an important role. Such criteria should also work well for estimating the scale of illicit tobacco trade more generally, for example by customs authorities for operational purposes.

In this paper the four criteria used are defined and interpreted as follows:

Reliability refers to the consistency of resulting estimates. A reliable method should produce similar results when applied more than once. In other words, reliability refers to the accuracy of the method, making estimates more comparable across settings/countries.

Validity, on the other hand, expresses the expectation of being precise with respect to the actual value that is measured. In the current context, estimates obtained via a valid methodology are close to the actual scale of illicit trade. Similarly, invalid methods might capture only part of the illicit trade and therefore cannot produce the true scale of illicit tobacco trade.

Feasibility implies that the data collection process should be possible in practical terms, economically reasonable and methodologically straightforward, and the estimates should be unambiguously interpretable.

Country coverage means that the application should be available for use by as many countries as possible. In a sense, this criterion complements feasibility: even if the method is feasible in some countries, this might not be true for others. This is an especially important criterion for evaluating methods that rely on secondary data because availability of data is an essential part of implementation. For methods that need primary data collection, the existence of prior applications of the method is one indicator of country coverage.

We combined information from various data sources to perform the evaluation. However, we began by developing a theoretical framework to structure the subsequent discussion and evaluation. Building on a few basic economic identities, we generated a simple model to distinguish between various components and types of economic activity in tobacco markets.

The characteristics of both methods under evaluation are described in the literature. We first identified review or methodology papers that could be used to map the main logic and features, as well as the overall pros and cons of each methods. To understand how each model works in practice, we also examined empirical studies utilising the methods. Rather than presenting a systematic review of related literature, our evaluation relies on studies specifically addressing illicit tobacco trade and/or its measurement. The studies were identified using various bibliographic platforms, including Scopus, Web of Science, ScienceDirect, Wiley Online Library, SpringerLink, and Taylor & Francis. Additionally, a Google web search was used to find other analytical papers on the issue. We used the following search terms: 'illicit tobacco', 'tobacco trade', 'illicit cigarettes', 'measurement of illicit trade' and 'cigarette smuggling'. In addition to studies specifically related to illicit tobacco or its measurement, we looked for studies and information addressing general aspects of both methods such as limitation of consumer surveys or trade statistics.

3. Accounting the illicit tobacco trade

In this section we present an accounting framework for ITTP. Implicitly, we focus on one country in one year and its interactions with the rest of the world. Assuming that total production of tobacco products Q consists of licit and illicit production, Q^L and Q^I, respectively, therefore:

$$Q = Q^L + Q^I \qquad (1)$$

The products can be sold legally in domestic markets, and sales are denoted as S^L. In most instances, $S^L = Q^L$. However, a certain amount of locally manufactured products might be exported legally before any sale (sold abroad), denoted as X^L. Considering that imported licit products M^L are also sold in the domestic market, officially recorded sales of tobacco products in the country can be expressed as follows:

$$S^L = Q^L - X^L + M^L \qquad (2)$$

Note that the entire amount of the products sold is not consumed by domestic consumers. After sale, they might be re-sold informally to consumers in other countries, denoted as X^{LI}. In the case of tobacco, this might happen in the form of 'bootlegging', for example. In addition, if there is a considerable cross-border trade taking place between neighbouring countries, additional export and import components appear, denoted as X^T and M^T, respectively. Therefore, to get the legal consumption of domestic consumer C^L, the last formula must be modified as follows:

$$C^L = S^L - X^{LI} - X^T + M^T \qquad (3)$$

Assuming that total consumption C of tobacco consists of licit consumption C^L and illicit consumption C^I, therefore:

$$C = C^L + C^I \qquad (4)$$

Equation (3) shows the components of licit consumption. For illicit consumption there are three sources. The first one is the legal export of trade partners that is diverted into an illegal form when entering the country. This can be expressed as the discrepancy between partners' export X^P and officially recorded import of the country from these partners M^L. The second source is imports that are not recorded as exports in partner countries nor as imports in the importing country. Again, a typical example is bootlegging. This is denoted as M^I. The third source is domestic illicit production Q^I. We must also consider that part of this illicit production might be exported, and that we denote as X^I. Accordingly, illicit consumption can be formulated as follows:

$$C^I = (X^P - M^L) + M^I + Q^I - X^I \qquad (5)$$

From these equations it is easy to derive simple formulas to estimate illicit consumption C^I. Survey-based studies manipulate equation (4) and illicit consumption is measured as follows:

$$C^I = C - C^L \qquad (6)$$

Note that the equation (6) captures all the components shown in equation (5). If the data about total consumption and tax paid consumption is available (or possible to collect), then this formula can be utilised. However, in practice it is usually assumed that $C^L \approx S^L$ and therefore tax paid on sales is used instead of consumption. While sales might be a good approximation for legal consumption under certain conditions, this approach ignores the variables X^{LI}, X^T and M^T (see equation (3)). For example, if a considerable amount of sold products is moving out of the country in the form of cross-border trade or bootlegging, this approach underestimates actual illicit consumption. One way to overcome these problems would be to measure C^I directly, and that is often done in practice.

The trade discrepancy approach employs equation (5) by including only the first two terms, as follows:

$$C^I = X^P - M^L \tag{7}$$

Again, the three variables M^I, Q^I and X^I are ignored since they are not included in the official trade statistics. This means that if there is considerable informal production or illegal importation occurring, the trade discrepancy approach will underestimate the actual illicit consumption. The survey-based approach is, therefore, potentially more exact compared to the trade discrepancy method if the overall aim is to reliably measure the consumption of illicit tobacco.

Having introduced both methods and having highlighted the pitfalls of these in a systematic manner using the accounting framework, we now move to more detailed descriptions of each method.

4. Survey-based approaches

Survey-based approaches rely on the data collected directly from market participants, usually from consumers. Although there have been studies carried out that gather information from the producers' side of the market (for examples see Skinnari & Korsell, 2016), it is not reasonable to expect that producers or sellers of illicit products are willing to reveal information that could threaten their business, especially if the data collection takes place on a regular basis.

Two general approaches have been used to arrive at the estimate of the scale of ITTP. First, it is measured on the information collected from consumers about their purchases and consumption of illicit products (for example, proportion of illicit packs consumers have bought). The second, and less straightforward approach, is based on data on overall consumption of tobacco products, including both licit and illicit, avoiding the need to separate them. In this case the total consumption figures can be combined with tax paid on sales to derive the illicit part of the consumption, as expressed in equation (6). The data for tax paid on sales should be available in most countries, since it can be easily calculated based on the excise tax revenues collected from the tobacco products. Examples of studies that have utilised this approach are Nguyen et al. (2014), Ahsan et al. (2014) and Blecher (2010). It is also employed by tax and customs authorities, for example, in the United Kingdom (HM Revenue & Customs, 2020).

Comparing these two survey-based approaches, the second indirect one is certainly more feasible in terms of country coverage with respect to secondary data sources. Although conducting surveys is typically quite expensive, the data on total tobacco use/smoking prevalence are collected regularly in many countries, which cannot be said about illicit trade or consumption. For example, 76 countries in the world effectively monitor tobacco use through representative and periodic surveys of adults and young people (World Health Organization, 2017) and 146 countries regularly monitor smoking behaviour among the population aged greater than or equal to 15 years (WHO, 2018). Considering that regular monitoring of tobacco use is a requirement of the 181 parties to the WHO Framework Convention on Tobacco Control (FCTC), there is a potential to increase the number of countries even more.

Regarding data on illicit tobacco products, there are at least two regular surveys conducted but both fail to meet the criteria of validity and reliability or are not considered sufficiently transparent for us to evaluate. The Euromonitor database is quite widely used, but the methodology of this data remains unclear (see Schafferer et al., 2018 and Prieger & Kulick, 2018). Another option would be to rely on estimates provided by KPMG, an international services network funded by the tobacco industry. Several authors have employed this data (Transcrime, 2016; Calderoni, 2014), but there is relatively strong evidence indicating its non-reliability (Gallagher et al., 2018; Stoklosa, 2016).

Various methodological options exist for survey-based approaches to acquire the required information from consumers. Smokers might be interviewed face-to-face on the street or at home, via telephone or asked to mail their empty cigarette packs for examination. The latter means that data collection could involve examination of the cigarette packs by the interviewer to clarify if it is smuggled and/or taxes are paid (Yurekli et al., 2001). Respondents do not necessarily have to be aware of the legality of products they have purchased, and it might require knowledgeable interviewers to detect illicit packs. Curti et al. (2019), on the other hand, used an indirect approach to separate illegal and legal products by analysing brand information and prices.

If the sample is representative and large enough, inferential statistics can enable estimates for illicit tobacco consumption to be derived for the entire population. While there is evidence that venue-based (Muhib et al., 2001; Thomas & Freisthler, 2016) or street intercept (Graham et al., 2014) sampling can work relatively well, it is less likely to draw a representative sample with known statistical properties. Superior statistical confidence and higher reliability can be achieved if nationally representative random samples are available, whether the sampling was constructed specifically for the tobacco study, as in Joossens et al. (2014), or is included as part of another study as in Ciecierski (2007).

Regardless of the choice between different approaches or sampling procedures, it must be acknowledged that survey-based studies rely heavily on self-reported data. Therefore, they suffer from social desirability bias, according to which self-reported assessment of substance use such as alcohol, tobacco or drugs tends to underestimate the actual consumption or heavy users might remain out of the reach of researchers (Liber & Warner, 2017; Johnson, 2014; Tourangeau & Yan, 2007; Yurekli et al., 2001; Pérez-Stable, 1990). Johnson (2014) discusses various reasons behind misreporting tendency, such as the need to avoid social threat, or feelings of shame, embarrassment or weakness. In some cases, however, researchers must also account for the possibility of over-reporting that is motivated from trying to impress peers, for instance (Johnson, 2014).

There are certain measures researchers have tried to use to reduce reporting bias. For example, examination of purchased packs by interviewers might be preferred rather than allowing respondents to identify the legality of the products. Other options include the collection of additional data from the respondent's significant other, providing the respondent with greater privacy when speaking with the interviewer or exerting some sort of psychological pressure on respondents to answer more truthfully (Johnson, 2014). For example, Joossens et al. (2014) point to the at-home setting of an interview as a limitation if the interview takes place in front of the family members. Ciecierski (2007) has rather thoroughly addressed and tried to minimise under-reporting bias by providing private and at-home settings for respondents that aim to offer anonymity and a feeling of comfort to answer what might be uncomfortable questions. In addition, authors may structure interviews so that they encourage respondents to feel as comfortable as possible. Specifically, authors may intentionally ask additional questions so that tobacco-related themes do not seem overly important.

Regardless of these attempts to tackle the problem, misreporting remains an unsolved issue (Liber & Warner, 2017), comprising the main weakness of survey-based methods. Some studies have tried to deal with the under-reporting through sensitivity analysis, presenting estimates under various assumptions about the level of under-reporting. For example, Ahsan et al. (2014) and Nguyen et al. (2014) both assume under-reporting levels of 10–30 per cent. Results from these studies show that the change in the level of under-reporting by 10 per cent changes the estimated scale of illicit trade up to twofold, or even more in some years. Therefore, under-reporting bias causes substantial uncertainty in estimating the scale of illicit trade.

Due to potential misreporting, estimates derived from self-reported assessments might suffer from both low validity and reliability and should be interpreted cautiously. Specifically, in terms of validity and due to social desirability bias, estimates might resemble preferred levels of consumption rather

than actual levels. Regarding reliability, although respondents might give consistent answers over a short period, consumption estimates are based on actual consumption as well as factors affecting respondents' individual motivations to misreport. For example, in Indonesia, where smoking might be socially desirable (Ahsan et al., 2014), consumers could be inclined to over-reporting. In the United States, with decreasing smoking prevalence, the stigmatisation of smoking is probably increasing (Liber & Warner, 2017) and an under-reporting tendency might prevail. This makes it a rather complicated task to determine the actual change in consumption.

Other typical limitations that all surveys face are also important issues for tobacco surveys. According to survey error models, one must account for possible errors like sampling, coverage, nonresponse and inferential errors (Lavrakas, 2013). In this context, several errors could be decreased by focusing surveys on total consumption rather than illicit consumption. For example, it is probably less embarrassing or inconvenient for a consumer to report frequent smoking than frequent purchase of illicit tobacco, potentially reducing nonresponse errors (Tourangeau & Yan, 2007). On the other hand, focusing directly on illicit products, a researcher can decrease misreporting by personally examining tobacco products to identify illegal ones. Depending on circumstances, a direct approach to measuring illicit consumption might still be preferable.

In addition, based on the accounting framework above, the indirect approach is also inclined to under-estimate the size of the illicit market if a considerable proportion of legal sales are exported via bootlegging or via cross-border trade (that is, because too large a value is subtracted from the total consumption figure). The indirect approach is also inclined to over-estimate the size of the illicit market if tobacco products are imported via cross-border trade into the country. If the scale of bootlegging and cross-border trade can be detected and assessed using other methods (that is, expert opinions and/or other studies), the resulting error might not be a substantial problem.

5. Trade discrepancy approach

The trade discrepancy method uses international trade statistics to compare reported tobacco exports destined for a country against reported tobacco imports. More specifically, comparing reported export from country A to country (or region) B to reported imports to country (or region) B, the difference between official exports and imports is largely composed of illicit trade. One important assumption underlying this logic is that since there are usually no taxes imposed on exports, the exporting country most likely reports correct figures; but as the imports are taxed, importers have incentives to hide the true values and under-report or divert the goods to the illicit market (Yurekli et al., 2001). As Yurekli et al. (2001) have noted, 'persistent discrepancies between these amounts – discrepancies that cannot be explained by other factors – provide an estimate of the amount of wholesale smuggled tobacco.'

This method has been applied to estimate the scale of ITTP, but it has its origins in estimating a more general phenomenon, illicit financial flows, where it is usually referred to as the 'trade mirror statistics method'. As described in more detail by Cobham and Janský (2020), the approach contrasts what a country claims to import from (or export to) the rest of the world with what the rest of the world states it exported to (or imported from) that given country. The progressive development of this method has taken place since Morgenstern (1950, 1974) and Bhagwati (1964, 1974), the latter being used as an example by Yurekli et al (2001). These studies usually use publicly available international trade data sets. Frequently used data sources include the United Nations' (UN) Comtrade data, used recently by Kellenberg and Levinson (2016) and Ndikumana (2016), while prominent research by Global Financial Integrity's Spanjers and Salomon (2017) and Ndikumana and Boyce (2010) use International Monetary Fund's (IMF) Direction of Trade Statistics. The two data sets are similar, but the coverage

of countries and the level of aggregation (data are available at a so-called 'product level' and, for recent years, monthly) make UN Comtrade the preferable source. In addition, only the UN Comtrade provides detailed, separate information on tobacco trade. For example, the UN Comtrade data has been used recently as one of the sources for studying ITTP in Paraguay (Gomis et al., 2018). Using data on commodities '240220 – Cigarettes; containing tobacco', authors argued that the methodology can be useful despite many caveats.

One of the strengths of this method is good country coverage, as estimations are based on data that are available for most countries. More specifically, the UN Comtrade database stores standardised official annual trade statistics reported by more than 170 countries and reflects international merchandise flows with coverage reaching up to 99 per cent of world merchandise trade (United Nations [UN], n.d). In addition, since the data is accessible online, the method is quite practical. Results are also readily understood and interpreted.

Although practices to collect the data used for the trade discrepancy approach vary across countries, in developing countries it is based on officially reported figures drawn primarily from customs declarations (UN, 2013). These figures are reasonably reliable and substantial errors are rare. Developed countries, however, rely heavily on other sources such as administrative records associated with taxation or enterprise surveys, partly due to abolition of customs control, for example within the European Union (UN, 2013).

Nevertheless, it should be noted that several potential errors can affect the accuracy of this method. Errors in commodity classifications, for example, or errors from intentional over- or under-invoicing – for the purposes of money laundering (see Altinkaya & Yucel, 2013) – can produce biased estimates. Varying valuation methods are also used to compile export and import data, sometimes requiring application of a threshold level above which illicit activity can be detected (Nitsch, 2016).

The validity of the trade discrepancy approach seems to be the most serious concern. Specifically, it is only possible to measure trade flows based on legal exports. If a considerable part of the export from country A to country B is taking place in the form of bootlegging or other forms of illegal trade, then the estimate only expresses part of the illegal flow of tobacco products. The same applies to illegal exports from country B: its magnitude affects the validity of this approach, as indicated by equation (5) above. For example, if we were to observe the export of tobacco products from Belarus to Estonia in the 2010s, we cannot detect any official trade according to the UN Comtrade database, although Belarusian illicit products are the most common ones in the Estonian illegal market, as we found out from interviews with customs officials. There is similar evidence from Latin America. Specifically, while Paraguay's tobacco company Tabesa does not have an official export to Brazil, customs authorities in the latter country seized millions of cigarettes in the 2010s (Gomis et al., 2018). Therefore, based on trade statistics, the illicit trade is not always identified or may be considerably distorted.

One simple explanation could be that these products are mostly 'cheap white' cigarettes that are produced legally but then diverted to illicit markets without being reported as exports. Therefore, the entity measured through the trade discrepancy method is narrower than defined above: international flows of illicit tobacco products, net of unreported export of partner countries (X^1 in equation (5)). The possibility also exists that the goods can be diverted to illicit markets of other countries (Yurekli et al., 2001). Therefore, the trade discrepancy approach might not produce valid results in country comparisons. Including an entire region in the analysis might give more valid estimates. However, the precision of the estimate could turn out lower if too many countries are compared because illicit trade flows are most likely destined to a limited number of countries (Nitsch, 2016).

Another potential limitation of the trade discrepancy approach is that it does not capture domestic illicit production without payment of taxes. This is also evident from equation (7) above. Yurekli et al. (2001) argue that because of this feature, the method is not recommended for use in countries that are significant producers of tobacco. One option to overcome this challenge would be to utilise complementary data and materials to improve the interpretability of trade statistics. For example, Gomis et al. (2018) searched the websites of transnational tobacco companies and Paraguay's tobacco company (Tabesa) for activities since the mid-2000s to understand the company's global business strategy. In addition, Gomis et al. (2018) utilise data such as imports of cigarette paper (HS 4813) and cellulose acetate (HS 391211 and 391212), two key cigarette components, to account for domestic illicit tobacco production.

There are evidently several essential assumptions and related limitations for the trade discrepancy approach. For example, the method assumes that all goods reported by an exporting country but not by the corresponding importing country are ultimately smuggled into the same importing country (and no other country). Furthermore, bootlegging is not covered by this method (Yurekli et al., 2001). More generally, Nitsch (2016) observes that knowledge about trade misreporting practices is scarce and more detailed analysis about the reasons of discrepancies in trade statistics is needed to improve the validity of this approach. For example, transaction-level data with information from both countries could enable further validation of the method. Neither the UN Comtrade nor other similar international databases contain such data, and transaction-level data has been used only in a limited number of single-country case studies like Wier (2020) that generally do not focus on tobacco. Above all, it is hard to find evidence that the method fulfils the validity criteria that the estimates obtained are close to the actual scale of ITTP.

When it comes to comparing trade discrepancy and survey-based methods, Nguyen et al. (2014) have provided estimates for Vietnam using both approaches. The trade discrepancy method produced comparable estimates with the survey-based method when an assumption of roughly 30 per cent under-reporting of tobacco use was used, although some clear differences were seen in the respective results. Specifically, while the survey-based approach revealed a decreasing trend in illicit trade over time, the trade discrepancy approach showed an opposite trend (after 2006). Nguyen et al. (2014) highlighted how the two methods measure and reflect different aspects of ITTP: 'The survey-based method measures the magnitude of illicit tobacco present in the market, whereas the trade discrepancy method measures international flows of illicit tobacco products.'

6. Discussion

We have reviewed and evaluated two approaches that are often used to assess the scale of illicit trade in tobacco products. The key findings are presented in Table 1. Both survey-based and trade discrepancy approaches have specific strengths and weaknesses that must be considered when used for measurement purposes. According to our descriptive evaluation, the trade discrepancy method slightly outperforms the survey-based method on country coverage. While secondary data is available for both methods, a larger range of countries is covered by trade statistics. In terms of feasibility, the methods are equally practical under the availability of secondary data. However, the survey-based approach might be somewhat easier to interpret as the main data is collected directly from consumers while the trade discrepancy method derives the estimate indirectly based on certain assumptions.

Table 1: Results of qualitative descriptive evaluation

Criteria	Survey-based method	Trade discrepancy method	Recommendation
Validity	Misreporting	Ignores domestic production and flows that are illicit along the entire supply chain	Survey-based method
Reliability	Consumer surveys sensitive to various errors and biases	Customs declarations and enterprise surveys more reliable	Trade discrepancy
Feasibility	Practical, straightforward interpretation	Practical, less straightforward interpretation	Survey-based method
Country coverage	146 countries currently, potentially 181	More than 170 countries, 99% of world merchandise	Trade discrepancy
Overall	While the country coverage, feasibility and reliability of both methods are comparable, due to the low validity of the trade discrepancy method we evaluate that a survey-based approach would generally outperform the trade discrepancy method.		

Regarding the reliability of methods, both methods rely at least partly on data collected via surveys. However, trade statistics, which in turn rely on enterprise surveys, are probably less sensitive to various errors as businesses (unlike consumers) routinely keep records of their transactions. In addition, trade statistics are largely based on officially reported customs declarations, which we take to be more reliable than consumers' self-reported data. This leads us to favour the trade discrepancy approach in relation to the reliability criterion.

Both methods have limitations that challenge the validity of the obtained estimates. For the trade discrepancy approach, the main limitations concern the lack of information with respect to domestic production, and illicit flows along the entire supply chain. If these forms of illicit activities are substantial, the trade discrepancy method alone is not suitable. This and other limitations, including the poor availability of transaction-level trade data, imply that an acceptable level of validity remains out of reach.

In the case of survey-based methods, one of the key challenges to acknowledge and deal with is the possibility of misreporting, which threatens the validity of estimates. Unfortunately, there is no good

solution for that problem as yet. However, there are several methods that researchers can employ to decrease the misreporting error. One is to turn to an indirect approach by measuring the scale of illicit trade through discrepancy between sales and survey-based total consumption, since the latter is probably a less sensitive survey topic than addressing illicit tobacco through direct measurement tools. One of the strengths of this approach is the fact that tobacco consumption or smoking prevalence is measured in many countries worldwide, providing good country coverage and practicality.

The misreporting challenge does not mean that survey-based methods are inapplicable. In comparison to the trade discrepancy approach – that is completely incapable of capturing certain illicit transactions – achieving a satisfactory validity is more probable with survey-based approaches. However, if the objective is to estimate the absolute scale of ITTP, studies based on self-reported data might need some form of cross-validation through less subjective statistics such as national and international trade statistics or taxes on paid sales. To obtain estimates of high quality these options should be reviewed and implemented with great care. Several contextual factors, such as existence of domestic production, overall trend in smoking prevalence and existence of bootlegging, should be accounted for.

We conclude with proposals for uses and refinements of the existing methodologies. Overall, we find that for policy purposes, such as FCTC evaluation or the SDGs target, survey-based approaches are preferred. There are aspects of the survey-based method to be improved, but across the four criteria it performs better. In the case of the trade discrepancy method, it is hard to find evidence that the method, as applied to the UN Comtrade international trade statistics, fulfils the validity criteria. As the method stands currently, we discourage its use as the main method for policy evaluations in the international context. Future research should continue to examine misreporting issues surrounding substance use, to clarify further optimal use of existing measures.

Acknowledgements

Disclaimer

The present paper was produced as part of the 'Pilot project on illicit trade in tobacco products in the European Union 2017–2019', implemented by the Secretariat of the WHO Framework Convention on Tobacco Control, with the financial assistance of the European Union. The opinions expressed in this document are those of the authors. They do not purport to reflect the opinions or views of the Secretariat of the WHO Framework Convention on Tobacco Control and the Protocol to Eliminate Illicit Trade in Tobacco Products.

References

Ajmal, A., & Veng, I. U. (2015). Tobacco tax and the illicit trade in tobacco products in New Zealand. *Australian and New Zealand Journal of Public Health*, *39*(2), 116–120.

Aldridge, J., & Décary-Hétu, D. (2015, January 4). *Cryptomarkets: The darknet as an online drug market innovation* (Final report to NESTA). University of Manchester/University of Montreal. https://daviddhetu.openum.ca/files/sites/39/2017/04/Nesta-Final-Report.pdf

Ahsan, A., Wiyono, N. H., Setyonaluri, D., Denniston, R., & So, A. D. (2014). Illicit cigarette consumption and government revenue loss in Indonesia. *Globalization and Health*, *10*(75).

Altinkaya, Z., & Yucel, O. (2013). The effects of international trade on international money laundering from the perspectives of international law and international trade in Turkey. *European Scientific Journal*, *1*, 116–124.

Barrera, V., Malm, A., Decary-Hetu, D., & Munksgaard, R. (2019). Size and scope of the tobacco trade on the darkweb. *Global Crime*, *20*(1), 26–44.

Bhagwati, J. (1964). On the underinvoicing of imports. *Oxford Bulletin of Economics and Statistics*, *27*(4), 389–397.

Bhagwati, J. N. (1974). On the underinvoicing of imports. In J. N. Bhagwati (Ed.), *Illegal transactions in international trade* (1st ed., pp. 138–147). Elsevier.

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

Calderoni, F. (2014). A new method for estimating the illicit cigarette market at the subnational level and its application to Italy. *Global Crime*, *15*, 51–76.

Ciecierski, C. (2007, September 4). *The market for legal and illegal cigarettes in Poland: A closer look at demand and supply-side characteristics* (IDRC Working Paper Series/ITEN Working Paper Series No.1). International Development Research Centre.

Cobham, A., & Janský, P. (2020). Estimating illicit financial flows: A critical guide to the data, methodologies and findings. Oxford University Press.

Curti, D., Shang, C., Chaloupka, F. J., & Tong, G. T. (2019). Tobacco taxation, illegal cigarette supply and geography: Finding from the ITC Uruguay Surveys. *Tobacco Control*, *28,* s53–s60.

Gallagher, A. W., Evans-Reeves, K. A., Hatchard, J. L., & Gilmore, A. B. (2018). Tobacco industry data on illicit tobacco trade: A systematic review of existing assessments. *Tobacco Control*, *28*, 334–345.

Goel, R. K. (2008). Cigarette smuggling: Price vs non-price incentives. *Applied Economic Letters*, *15*, 587–592.

Goel, R. K., & Saunoris, J. W. (2019). Cigarette smuggling: Using the shadow economy or creating its own? *Journal of Economics and Finance*, *43*, 582–593.

Gomis, B., Lee, K., Botero, N. C., Shepherd, P., & Iglesias, R. M. (2018). 'We think globally': The rise of Paraguay's Tabacalera del Este as a threat to global tobacco control. *Globalization and Health*, *14*(110).

Graham, K., Bernards, S., Clapp, J. D., Dumas, T. M., Kelley-Baker, T., Miller, P. G., & Wells, S. (2014). Street intercept method: An innovative approach to recruiting young adult high-risk drinkers. *Drug and Alcohol Review*, *33*, 449–455.

HM Revenue & Customs. (2020, July 9). *Measuring tax gaps 2020 edition. Methodological Annex*. HM Revenue & Customs Corporate Communications. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/943842/Measuring_tax_gaps_2020_edition_-_methodological_annex.pdf

Johnson, T. P. (2014). Sources of error in substance use prevalence surveys. *International Scholarly Research Notices*. Article ID 923290. https://doi.org/10.1155/2014/923290

Joossens, L., Merriman, D., Ross, H., & Raw, M. (2009, January 22). *How eliminating the global illicit cigarette trade would increase tax revenue and save lives*. International Union Against Tuberculosis and Lung Disease. https://theunion.org/sites/default/files/2020-08/How%20Eliminating%20the%20Global%20Illicit%20Cigarette%20Trade%20Would%20Increase%20Tax%20Revenue%20and%20Save%20Lives%20-%20EN.pdf

Joossens, L., Lugo, A., La Vecchia, C., Gilmore, A. B., Clancy, L., & Gallus, S. (2014). Illicit cigarettes and hand-rolled tobacco in 18 European countries: A cross-sectional survey. *Tobacco Control*, *23*, e17–e23.

Kellenberg, D., & Levinson, A. (2016). *Misreporting trade: Tariff evasion, corruption, and auditing standards*. Cambridge: National Bureau of Economic Research. http://www.nber.org/papers/w22593

Kupka, P., & Tvrdá, K. (2016). Cigarette smuggling for greenhorns: Recent patterns of illegal tobacco trade in Czech Republic. *Trends in Organized Crime*, *19*, 218–235.

Lavrakas, P. J. (2013). Applying a total error perspective for improving research quality in the social, behavioral, and marketing sciences. *Public Opinion Quarterly*, *77*(3), 831–850.

Liber, A. C., & Warner, K. E. (2017). 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), 113–119.

Morgenstern, O. (1950). *On the accuracy of economic observations*. Princeton University Press.

Morgenstern, O. (1974). On the accuracy of economic observations: Foreign trade statistics. In J. N. Bhagwati (Ed.), *Illegal transactions in international trade* (pp. 87–122). North Holland Publishing Company.

Muhib, F. B., Lin, L. S., Stueve, A., Miller, R. L., Ford, W. L., Johnson, W. D., & Smith, P. J. (2001). Venue-based method for sampling hard-to-reach populations. *Public Health Reports*, *116*, 216–222.

National Research Council. (2015). *Understanding the US illicit tobacco market: Characteristics, policy context, and lessons from international experiences*. The National Academies Press. https://doi.org/10.17226/19016

Ndikumana, L. (2016). Trade misinvoicing in primary commodities in developing countries: The cases of Chile, Côte d'Ivoire, Nigeria, South Africa and Zambia. United Nations Conference on Trade and Development (UNCTAD). United Nations. https://unctad.org/system/files/official-document/suc2016d2_en.pdf

Ndikumana, L., & Boyce, J. K. (2010). Measurement of capital flight: Methodology and results for sub-Saharan African countries. *African Development Review*, *22*(4), 471–481.

Nguyen, M. T., Denniston, R., Nguyen, H. T., Hoang, T. A., Ross, H., & So, A. D. (2014). The empirical analysis of cigarette tax avoidance and illicit trade in Vietnam, 1998–2010. *PLoS ONE*, *9*(1), e87272. https://doi.org/10.1371/journal.pone.0087272

Nitsch, V. (2016). *Trillion dollar estimate: Illicit financial flows from developing countries* (Darmstadt Discussion Papers in Economics No. 227). Darmstadt University of Technology. http://econpapers.repec.org/paper/zbwdarddp/227.htm

Pérez-Stable, E. J., Marin, B. V., Marin, G., Brody, D. J., & Benowitz, N. L. (1990). Apparent underreporting of cigarette consumption among Mexican American smokers. *American Journal of Public Health*, *80*(9), 1057–1061.

Prieger, J. E., & Kulick, J. (2018). Cigarette taxes and illicit trade in Europe. *Economic Enquiry*, *56*(3), 1706–1723.

Schafferer, C., Yeh, C-Y., Chen, S-H., Lee, J. M., & Hsieh, C-J. (2018). A simulation impact evaluation of cigarette excise tax increase on licit and illicit cigarette consumption and tax revenue in 36 European countries. *Public Health*, *162*, 48–57.

Skinnari, J., & Korsell, L. (2016). The illicit tobacco market in Sweden – from smuggling to warehousing. *Trends in Organized Crime*, *19*, 273–299.

Spanjers, J., & Salomon, M. (2017, April). *Illicit financial flows to and from developing countries: 2005–2014*. Global Financial Integrity®. https://secureservercdn.net/45.40.149.159/34n.8bd. myftpupload.com/wp-content/uploads/2017/05/GFI-IFF-Report-2017_final.pdf

Stoklosa, M. (2016). Is the illicit cigarette market really growing? The tobacco industry's misleading math trick. *Tobacco Control*, *25*, 360–361.

Thelwall, M. (2001). A web crawler approach for data mining. *Journal of Information Science*, *27*(5), 319–325.

Thomas. C., & Freisthler, B. (2016). Assessing sample bias among venue-based respondents at medical marijuana dispensaries. *Journal of Psychoactive Drugs*, *48*, 56–62.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.

Transcrime – Research Centre on Transnational Crime. (2016). An European outlook on the illicit trade in tobacco products. *Trends in Organized Crime*, *19*, 300–328.

United Nations (UN). (2013). *International merchandise trade statistics: Compilers Manual, Revision 1 (IMTS 2010-CM)*. https://unstats.un.org/unsd/trade/EG-IMTS/IMTS2010-CM%20-%20white%20 cover%20version.pdf

United Nations (UN). (2015). *UN Comtrade*. https://unstats.un.org/unsd/comtrade_announcement.htm

United Nations (UN). (n.d). *What is UN comtrade?* https://unstats.un.org/unsd/tradekb/ Knowledgebase/50075/What-is-UN-Comtrade

Wier, L. (2020). Tax-motivated transfer mispricing in South Africa: Direct evidence using transaction data. *Journal of Public Economics*, *184*, 104153.

World Health Organization (WHO). (2013). *Protocol to eliminate illicit trade in tobacco products*. WHO Framework Convention on Tobacco Control. https://apps.who.int/iris/bitstream/ handle/10665/80873/9789241505246_eng.pdf;jsessionid=C8B8CE5CF4775E6B9BB7B2BBCC1F BBD3

World Health Organization (WHO). (2017, July 19). *WHO report on the global tobacco epidemic, 2017: monitoring tobacco use and prevention policies*. https://www.who.int/tobacco/global_ report/2017/en/

World Health Organization (WHO). (2018). *WHO Global Report on Trends in Prevalence of Tobacco Smoking 2000–2025*. (2nd ed.).

World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC). (2020). *Illicit tobacco trade in the European Union 2017-2019: Report on study visits of academic researchers to national customs authorities*. https://www.who.int/fctc/protocol/activities/pilot-project-eu-protocol/ en/

Yurekli, A., de Beyer, J., & Merriman, D. (2001). *Understand, measure, and combat tobacco smuggling* (World Bank economics of tobacco toolkit no. 7. Smuggling). World Bank Group. http://documents.worldbank.org/curated/en/418961468163740317/Understand-measure-and-combat-tobacco-smuggling

Yurekli, A., & Sayginsoy, Ö. (2010). Worldwide organized cigarette smuggling: An empirical analysis. *Applied Economics*, *42*, 545–561.

Indrek Saar

Indrek Saar is a professor at the Financial College of the Estonian Academy of Security Sciences and an associate professor at the School of Governance, Law and Society of Tallinn University. He received his PhD in economics from the University of Tartu. His main teaching areas include economics, taxation, and quantitative methods. His main research interests are related to the fields of public finance and policy analysis, particularly taxation and customs policies, economics of law enforcement and cost-benefit analysis.

Petr Janský

Petr Janský is an economist with an empirical research interest in public finance, taxation, tax havens and international development. With academic degrees from the University of Oxford and Charles University in Prague, Petr Janský is currently an associate professor and head of department of European economic integration and economic policy at Charles University. He publishes in scholarly journals such as Economic Geography, International Tax and Public Finance or World Economy. Petr Janský pursues impact through collaboration with international organisations such as the European Parliament and the United Nations.