Comparison of Trends in Self-reported Cigarette Consumption and Sales in England, 2011 to 2018

Sarah E. Jackson, PhD; Emma Beard, PhD; Bernard Kujawski, PhD; Ella Sunyer, BA; Susan Michie, DPhil; Lion Shahab, PhD; Robert West, PhD; Jamie Brown, PhD

Abstract

IMPORTANCE Population cigarette consumption is declining in many countries. Accurate estimates of long- and short-term changes are vital for policy evaluation and planning. Survey data and sales data that are used to make these estimates each have important potential biases, so triangulation using different methods is required for robust estimation.

OBJECTIVES To compare monthly estimates of cigarette consumption in England from a nationally representative survey and recorded cigarette sales and to triangulate an accurate estimate of changes in cigarette consumption since 2011.

DESIGN, SETTING, AND PARTICIPANTS This study used times series analyses based on survey data and recorded cigarette sales to estimate and compare trends in population cigarette consumption in England from 2011 to 2018. Survey participants were representative samples of 1700 people aged 16 years or older each month in England.

MAIN OUTCOMES AND MEASURES Monthly cigarette retail sales data from August 2011 through February 2018 were obtained from a data agency. Monthly self-reports of cigarette consumption were collected over the same period using the Smoking Toolkit Study.

RESULTS A total of 136 677 individuals (51.1% female; mean [SD] age, 46.7 [18.8] years) were surveyed. Over the study period, mean monthly cigarette consumption in England was 2.85 billion (95% CI, 2.78 billion to 2.93 billion) cigarettes based on survey data compared with 3.08 billion (95% CI, 3.03 billion to 3.13 billion) estimated from sales data. Over the whole period, cigarette consumption declined by 24.4% based on survey data and 24.1% based on sales data. This equated to 118.4 million and 117.4 million fewer cigarettes consumed per month (or approximately 1.4 billion cigarettes per year) based on survey data and sales data, respectively. After adjusting for underlying trends, month-by-month changes in cigarette consumption were closely aligned: a 1% change in survey-estimated cigarette consumption was associated with a 0.98% (95% CI, 0.53%-1.44%) change in sales estimates.

CONCLUSIONS AND RELEVANCE Survey data and sales data were closely aligned in showing that overall cigarette sales in England have declined by almost a quarter since 2011, amounting to more than 1 billion fewer cigarettes smoked each year. The alignment between the 2 methods provides increased confidence in the accuracy of parameters provided by the Smoking Toolkit Study and sales data. It indicates that estimated changes in cigarette consumption are robust and provide a meaningful basis for policy evaluation and planning.
Introduction

Cigarette smoking is one of the leading risk factors for morbidity and mortality worldwide and is associated with a large economic burden. It is important to have accurate estimates of changes in aggregate cigarette consumption at a national level to evaluate and plan policies aimed at reducing smoking. These estimates can be obtained from surveys or sales data, but each has potential biases. This study compared survey and sales data in England to assess their degree of alignment and triangulate a robust estimate of national changes in cigarette consumption.

Survey data and sales data show a steady decline in smoking in many countries, including England, since the 1970s. The decline has been greater in some years than others and varies from country to country and within sociodemographic groups within countries. This variation has been crucial to the evaluation of policies such as tax increases, smoking bans, and marketing restrictions and the rate of decline is being used in some countries to plan policies aimed at moving those countries to what has been termed the "end game" in tobacco control.

National surveys provide crucial information for evaluating and planning national tobacco control policies. They can assess long-term progress and short-term changes that may be used to assess the effects of policies or events. However, they are potentially subject to biases. When it comes to estimating national cigarette consumption, measurement bias may result from not wanting to admit to smoking, inaccurate recall of number of cigarettes consumed, motivation to underestimate cigarette consumption, a tendency to round daily consumption to the nearest 5 or 10 cigarettes, and inaccurate translation of weight of tobacco to number of cigarettes in the case of hand-rolled cigarettes. There is evidence that failure to admit to smoking is rare in national surveys, including countries such as the United Kingdom, but it is not negligible. Sample bias may result from failure to access representative samples. This could arise from limitations in sample identification methods, inability to access certain respondents, or refusal by respondents to participate. This last issue has become increasingly challenging in recent years.

Recorded sales data are potentially subject to different types of bias. These could in principle include failure of the recording method to capture the full national picture, failure to capture use of tobacco purchased illicitly or abroad, wastage in the sense of cigarettes bought but not consumed, and stockpiling by smokers. Recorded sales data also have the limitation that they cannot provide information on smoking prevalence, which is a crucial indicator in its own right. Also, they cannot provide a breakdown by sociodemographic characteristics.

There has been a high-profile call for science to renew its focus on triangulation. Robust estimation of changes in total national cigarette consumption requires confidence that the information being provided by survey data and sales data are in alignment. While several studies have compared survey and sales data for alcohol consumption and reported strong correlations between them, 1 study has used cigarette sales data in Scotland to estimate the sources of bias that might lead to overestimation or underestimation of consumption, and 1 study has combined sales data with smoking prevalence estimates to calculate cigarette consumption per adult smoker in Great Britain, none, to our knowledge, have examined alignment in monthly trends between cigarette survey data and sales data in England. The Smoking Toolkit Study (STS), undertaken in England, provides a unique opportunity to study this because it involves monthly surveys of representative samples in which smoking status and self-reported cigarette consumption are assessed. Previous research has shown close correspondence between major smoking indices in this survey when aggregated annually and annual national surveys.

Thus, this study aimed to compare survey measures of smoking in a large sample of English adults aged 16 years and older from the STS with sales-based cigarette consumption data in England from 2011 through 2018. Specifically, this study addressed the following research questions:

1. What is the degree of concordance between estimated total national cigarette consumption in England over the study period as assessed by the STS and sales data?
2. What is the degree of concordance between the long-term trend in total national cigarette consumption in England over the study period as assessed by the STS and sales data?
3. What is the degree of concordance between monthly changes in total national cigarette consumption in England after taking account of the long-term trend and seasonal variation?

Methods

The study used time series analyses to compare trends in population cigarette consumption in England from August 2011 through February 2018 based on survey data and recorded cigarette sales. Ethical approval for the STS was granted originally by the University College London Ethics Committee and participants provided full written informed consent. The data in this study were not collected by University College London and were anonymized before being received by University College London, thus were exempt from requiring study-specific ethical approval. This study followed the American Association for Public Opinion Research (AAPOR) reporting guideline.

Data Sources

Survey Data
The STS is an ongoing monthly survey designed to provide information about smoking prevalence and behavior in England at a population level. The study uses a form of random location sampling to select a new sample of approximately 1700 adults aged 16 years or older each month. The survey typically covers 200 to 300 census output areas each wave, which are sampled at random (after stratification by geodemographic analysis of the population) from more than 170 000. Interviewers travel to the selected areas and perform computer-assisted interviews with 1 participant older than 16 years per household until quotas based on factors influencing the probability of being at home (working status, age, and sex) are fulfilled. Random location sampling is considered superior to conventional quota sampling because the choice of properties approached is reduced by the random allocation of small output areas. However, interviewers can still choose which houses within these areas are most likely to fulfill their quotas, rather than being sent to specific households in advance. Response rates are therefore not appropriate to record, unlike random probability sampling, in which interviewers have no choice as to the properties sampled, and so response at each address can be recorded. Full details of the study methods are available elsewhere, and comparisons with national data indicate that key variables such as sociodemographic characteristics and smoking prevalence are nationally representative. During the study period, data were collected from a total of 136 677 participants ranging in age from 16 to 99 years.

Sales Data
We obtained cigarette retail sales data from a data agency that provides measurement and analytics for market research. Data were collected from electronic point of sale (EPOS) to estimate cigarette retail sales for the Great Britain (England, Scotland, and Wales) market. The company collects census sales data from large grocery retailers on a weekly basis, comprising scanned EPOS readings of the type and volume of each pack of cigarettes sold, and a net retail price. The sampling frame includes most of the large supermarket chains but excludes discount retailers Lidl and Aldi, although neither sells cigarettes. Sales for remaining retail outlets, which are generally smaller and used for top-up and impulse purchases, are estimated using EPOS data from a weighted stratified random sample. Some independent outlets do not collect EPOS data, so sales are estimated via manual audits undertaken by the data agency, which examine invoices, purchase data, and stock levels. Data inputs are designed to be representative of Great Britain as a whole and separately for England and Wales (combined) and Scotland, and capture approximately 90% of total grocery market sales.

Sales data are not collected separately for England and Wales, but are segmented using Broadcasters’ Audience Research Board regions, with data for Wales grouped with the West of England. To get as good a match as possible between the geographical regions used in both data collection methods, we rescaled the figures for the Wales and West region to obtain reliable figures just for England. We were unable to find exact information about the geographical coverage of the
West Broadcasters’ Audience Research Board region. Thus, we used the Broadcasters’ Audience Research Board map provided on the organization’s website and overlaid local authority district boundaries obtained from the UK Data Service. Based on the overlap, we identified 11 local authority districts within the West region. Next, we extracted from the Office for National Statistics (ONS) population estimate data for those identified local authority districts to estimate the number of adults (age ≥16 years, for comparability with the STS sample) in the West region (1 838 080 individuals). Finally, using the ONS 2016 figure for the Wales population aged 16 years and older (2 556 071 individuals), we calculated a scaling factor (0.42) for the West and Wales region to estimate West-only sales figures. All analyses reported here are based on our modeling of the agency-provided sales data to estimate cigarette sales in England.

Measures
The primary measures from the STS were (1) smoking prevalence and (2) cigarette consumption among those who reported smoking. Prevalence was recorded as the proportion of participants who reported smoking cigarettes (including hand-rolled) daily or occasionally at the time of the survey or during the preceding 12 months. Smokers were identified as those who reported smoking cigarettes daily or occasionally. Self-reported consumption was measured by the question, “How many cigarettes do you usually smoke?” with participants given the choice of reporting number per day or per week, scaled up to monthly consumption by multiplying by the number of days in a given month. The mean cigarette consumption was multiplied by the smoking prevalence to give a per capita cigarette consumption figure. This was then multiplied by the total population estimate of those aged 16 years or older from the ONS for the year in question to scale up monthly consumption figures to the population in England aged 16 years or older. Where ONS population size estimates were currently unavailable (2017 and 2018 data), population size was estimated using a linear fit for the latest 6 years available in the ONS data set (2011-2016).

The primary measure from the sales data was volume sales, calculated as the number of manufactured cigarettes plus hand-rolled cigarettes estimated at 0.48 g of tobacco per cigarette.

Statistical Analysis
The analysis plan was registered on the Open Science Framework before data analysis. Data were aggregated monthly or yearly, depending on the analysis. Complete case analysis was used for the STS such that only participants with responses contributed to the aggregated level data. We analyzed data using R studio statistical software (R Project for Statistical Computing) in 3 stages.

First, we calculated total national cigarette consumption estimates from survey and sales data with the mean taken across the entire study period. Then, the 2 estimates were compared by calculating the ratio of the STS estimate over the sales estimate.

Second, we examined annual trends for cigarette consumption separately for the survey and sales data using a range of regression models and compared the overall parameters from the best-fitting models. Total national cigarette consumption from the STS and sales data were regressed on to time in separate linear regression models. Then several additional models were assessed to determine whether they provided a better fit: (1) quadratic trend model; (2) logarithmic regression (level-log model or logarithmic trend model); (3) exponential regression (log-level model or exponential trend model); and (4) power regression (log-log model or power trend model). To identify the best overall model, we compared the Akaike information criterion as the primary measure of fit and the adjusted R² and Bayesian information criterion as secondary measures of fit. The simplest model within 2 Δ (2 Akaike information criterion units) was selected. The slopes of the 2 models were then compared by calculating the slope of the STS-time trend over the sales-time trend.

Third, we assessed the concordance between monthly estimates of cigarette consumption from the STS and sales data, using bootstrapping to calculate the confidence interval, and used autoregressive integrated moving average with exogenous input (ARIMAX) modeling. The ARIMAX
method of modeling is an extension of autoregressive integrated moving average analysis, which produces forecasts based on prior values in the time series (autoregressive terms) and the errors made by previous predictions (moving average terms).

We followed a standard ARIMAX modeling approach. For this analysis, the input variable was specified as the survey-based estimate and the output variable was the sales-based estimate. First, each time series was assessed for outlying values that could bias the results; none were identified. The series were log transformed to stabilize the variance and first differenced and seasonally differenced if required. First differencing involves calculating the change between 1 observation and the next, whereas seasonal differencing involves calculating the change between 1 year and the next. The autocorrelation and partial autocorrelation functions were then examined to determine the seasonal and nonseasonal moving average and autoregressive terms. To identify the most appropriate transfer function for the continuous explanatory variables, we checked the sample cross-correlation function. We also checked that autocorrelation terms included in the model were statistically significant, and that model residuals were normally distributed, random, and independent. The Ljung-Box test for white noise was used to statistically evaluate the degree to which the residuals were free of serial correlation.

Model coefficients can be interpreted as estimates of the percentage change in sales-based total monthly national cigarette consumption for a 1% change from the series mean in the survey-based figure. All model assumptions were met.

![Figure 1. Trend Models for the Sales and Smoking Toolkit Study Data: Annualized Means of Monthly Cigarette Consumption in England From 2011 Through 2018](https://jamanetwork.com/)

Table 1. Model Fit Indices Assessing the Association Between Year and the Sales and STS Data From 2011 Through 2018

<table>
<thead>
<tr>
<th>Model</th>
<th>Sales Data</th>
<th>STS Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>BIC</td>
</tr>
<tr>
<td>Linear</td>
<td>312.5</td>
<td>312.8</td>
</tr>
<tr>
<td>Quadratic</td>
<td>312.2</td>
<td>312.5</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>322.9</td>
<td>323.1</td>
</tr>
<tr>
<td>Exponential</td>
<td>315.4</td>
<td>315.6</td>
</tr>
<tr>
<td>Power association</td>
<td>324.8</td>
<td>325.0</td>
</tr>
</tbody>
</table>

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; STS, Smoking Toolkit Study.

$R^2$ is adjusted for the number of terms associated with each model fitted.
Results

A total of 136,677 individuals (51.1% female; mean [SD] age, 46.7 [18.8] years) were surveyed. The mean cigarette consumption per month across the study period was 2.85 billion (95% CI, 2.78 billion to 2.93 billion) for the STS data and 3.08 billion (95% CI, 3.03 billion to 3.13 billion) for the modeled sales data. Thus the survey-based estimate was 93% of the sales estimate.

Figure 1 and Table 1 show the annualized means of the monthly cigarette consumption over the study period. In 2011 this was 3.40 billion according to the STS and 3.41 billion according to the sales data. In 2018 this had declined to 2.57 billion according to the STS and 2.58 billion according to the sales data. This represented an overall decline of 24.4% and 24.1% for STS and sales data, respectively, over the period. The mean monthly decline was 118.4 million according to the STS and 117.4 million according to the sales data. This amounted to 1.42 billion and 1.41 billion fewer cigarettes smoked per year according to the STS and sales data, respectively. The best-fitting curve for the annualized data was linear for the sales data but logarithmic for the STS data (Table 2).

Monthly changes in cigarette consumption estimated by STS and sales data were highly correlated ($r = 0.72$; 95% CI, 0.57-0.82). Figure 2 shows the monthly trends. The best-fitting ARIMAX model included a nonseasonal moving average term and 1 order of differencing, and found that for every 1% decrease from the series mean in self-reported cigarette consumption, sales decreased by 0.98% (95% CI, 0.53%-1.44%). There was no evidence of outliers and all assumptions were met. However, there was some evidence of additional seasonal autocorrelation. As a sensitivity analysis, a seasonal ARIMAX model with 1 nonseasonal moving average term, a seasonal

<table>
<thead>
<tr>
<th>Data Source</th>
<th>$B$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STS data (log year)</td>
<td>−424 130 540 (−508 318 860 to −339 942 220)</td>
</tr>
<tr>
<td>Sales data (year)</td>
<td>−105 110 439 (−127 167 777 to −83 053 101)</td>
</tr>
</tbody>
</table>

Abbreviation: STS, Smoking Toolkit Study.

Figure 2. Monthly Estimates of Cigarette Consumption in England From 2011 Through 2018 Based on Self-reported Cigarette Consumption From the Smoking Toolkit Study and Recorded Sales
autoregressive term, and 1 order of differencing was also run. This model indicated that for every 1% decrease from the series mean in self-reported cigarette consumption, sales decreased by 0.89% (95% CI, 0.36%-1.42%). The additional seasonal term did not significantly improve the fit of the model.

**Discussion**

The results showed a high degree of alignment between Smoking Toolkit Study data and sales data on the decline in aggregate cigarette consumption in England from 2011 through 2018. Both estimated that more than 1.4 billion fewer cigarettes were sold each year over the period, amounting to a decrease in consumption of approximately 24%. There was also a high correlation between the monthly changes estimated by the 2 methods after accounting for the overall trend and seasonality.

The alignment between the 2 methods provides confidence that the estimated changes in consumption are robust and provides a meaningful basis for policy evaluation and planning. The alignment also testifies to the accuracy of the component figures going into each estimate. In the case of the STS data, these figures are smoking prevalence, daily cigarette consumption among smokers, and population size. In the case of the sales data, the results validate the modeled sales figures and the extrapolation from the sample frame to retail outlets across the country.

The decline in national cigarette consumption has been dramatic and exceeded the decline in smoking prevalence, which over the same period was approximately 15%. The remainder of the decline has been in mean daily consumption. It should be noted that a decline in cigarette consumption among people who continue to smoke may not translate into a decline in toxin exposure, as there is evidence that smokers increase the intensity with which they smoke each cigarette to maintain their customary nicotine intake. However, studies have indicated that reduction may have benefits for certain disease outcomes (albeit only if the reduction is substantial and sustained) and benefit specific population groups (eg, pregnant smokers), and promote later attempts at cessation. It is also important to note that expenditure by smokers on their cigarettes has not declined because of tax increases and increases in prices charged by the tobacco companies.

Although the decline in cigarette consumption was very similar in the STS and sales data, the overall level was approximately 7% lower for the STS than sales data. This may be because we overestimated the cigarette equivalence of the weight of tobacco in hand-rolled cigarettes to produce the sales-based figure. There could also have been underreporting of smoking in the survey. Third, the survey may have underrecruited smokers. Fourth, the survey data were...
restricted to people aged 16 years and older. While it is illegal in the United Kingdom for individuals younger than 18 years to purchase cigarettes, there is some smoking among those younger than 18 years.63

Limitations
The study had a number of limitations. In terms of the sampling, it is possible that either or both the survey and sales data records may have failed to capture representative samples, although comparison of key variables (sociodemographic characteristics, smoking prevalence, and cigarette consumption) collected in the STS with those collected in other national surveys (the Health Survey for England and the General Lifestyle Survey) indicate that the sample is broadly representative of the adult population in England, and cigarette consumption across the surveys has been shown to be virtually identical.43 Over the study period, estimates of cigarette consumption among smokers participating in the STS were similar to those in the Health Survey for England64 and the Opinions and Lifestyle Survey (the successor to the discontinued General Lifestyle Survey)65 (Table 3). Thus, while both survey and sales methods may overestimate or underestimate cigarette consumption, it is unlikely results would be substantively different if these other surveys were used. In terms of measurement, the limitations of both types of data were described in the Introduction. In terms of generalizability, different countries may experience different types and degrees of bias, so one cannot assume that a similar alignment will be present in those jurisdictions. In addition, the analyses focused exclusively on cigarettes and did not include other tobacco products (eg, cigars) or alternative nicotine products (eg, electronic cigarettes), so one cannot assume that there is similar alignment for other product categories.

Conclusions
This study’s findings provide the most robust quantification to date, to our knowledge, of the decline in total cigarette consumption in England at approximately 24% from 2011 through 2018. The findings also provide increased confidence in the accuracy of parameters provided by the STS methods and sales data.
Administrative, technical, or material support: Michie.

Supervision: Sunyer, Shahab, West, Brown.

Conflict of Interest Disclosures: Dr Beard reported grants from Pfizer outside the submitted work. Dr Shahab reported grants from Pfizer outside the submitted work. Dr West reported grants and personal fees from and consultancy for Pfizer, personal fees from and consultancy for GlaxoSmithKline, and personal fees from and consultancy for Johnson and Johnson outside the submitted work. Dr Brown reported grants from Pfizer outside the submitted work. No other disclosures were reported.

Funding/Support: Cancer Research UK funded the Smoking Toolkit Study data collection (grants C1417/A22962 and C44576/A19501) and the salaries of Drs. Jackson, Beard, and Brown (grant C1417/A22962). Dr Jackson’s salary was also supported by the Economic and Social Research Council (grant ES/R005990/1). Dr Beard’s salary is partly funded by the National Institute for Health Research School for Public Health Research.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The views are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

Additional Information: Data are available from the authors on reasonable request. The study analysis plan is registered on the Open Science Framework at https://osf.io/t9zzv/. The School for Public Health Research is a partnership between the University of Sheffield, University of Bristol, Cambridge University, Imperial College London, University College London, The London School for Hygiene and Tropical Medicine, the LiLaC collaboration between the Universities of Liverpool and Lancaster and Fuse, and The Centre for Translational Research in Public Health, a collaboration between Newcastle University, Durham University, Northumbria University, the University of Sunderland, and Teesside University.

REFERENCES


17. Malone RE. Tobacco endgames: what they are and are not, issues for tobacco control strategic planning and a possible US scenario. Tob Control. 2013;22(suppl 1):i42-i44. doi:10.1136/tobaccocontrol-2012-050820


