

THE ESTABLISHMENT SURVEY

TECHNICAL REPORT

March 2017



1. Introduction

The collection of data for the Establishment Survey commenced on 1 July 2016. The aim is for media owners, marketers, and advertisers to use the survey as a strategic inter-media planning tool, as the ES surveys media usage across all media types by all major demographics.

It provides a demographic, geographic, lifestyle, product and multi-media landscape of South Africa. In addition, ES allows for cross-platform and cross-device media consumption analysis.

The survey consists of a face-to-face interview designed to collect information about households and individuals aged 15 years and older, representative of the adult South African population.

ES will be released bi-annually based on twelve-month rolling data. A “soft launch” will take place after the initial six months of data collection.

2. Definition of terms

Area classification

Areas are classified into one of three area types based on a fixed set of criteria and using the municipal code and geography type as per Stats SA Census 2011, namely:

- **Metro** – Small Areas (SAs) falling within the boundaries of the eight ‘Metropolitan Municipalities’ as defined by the National Demarcation Board (that being Ekurhuleni, City of Johannesburg, City of Tshwane, Mangaung, Buffalo City, Nelson Mandela Bay, eThekweni and City of Cape Town) AND classified as ‘Urban’ according to the geo type variable in the Census 2011 data.
- **Urban** – Small Areas that are within the borders of one of the ‘Metropolitan Municipalities’ (municipal code) and defined as ‘Farm’ or ‘Traditional’ (geography type) or are within the borders of an ‘Other Municipality’ (municipal code) and defined as ‘Urban’ (geography type).
- **Rural** – Small Areas that are classified as ‘Traditional’ or ‘Farm’ in the Census 2011 data (based on the geo type variable) and are located within ‘Other Municipality’.

CAPI

CAPI (Computer Assisted Personal Interview) interviewing is conducted face-to-face using a questionnaire deployed on a tablet. An interviewer conducts the interview as per the CAPI script, and captures the answers on the tablet. Data is uploaded to a central database once the interview is complete.

Geography type

The geography type is based on the official Statistics South Africa classification which uses input from the Surveyor General. The three geography types are Urban, Farm and Traditional.

Household

A group of people who eat and sleep in the house at least 4 days out of a week. If a house has a backroom, granny flat, cottage, shack or other separate dwelling, the first step is to determine if the person (or persons) living in that dwelling forms (or form) a separate household, or if they are part of the household living in the main dwelling according to the following:

If a person sleeps in a separate dwelling on the same property, but shares meals with the household in the main dwelling at least 4 days out of a week, they form part of the same household.

However, if the person (or persons) in the backyard dwelling eats (or eat) separately, they form a household on their own and must be counted as such. If there is more than one household on a stand then the KISH grid will be used to determine which household is to be interviewed.

Household income

The total claimed income of all members in the household before tax or any other deductions.

Household roster

Upon recruiting the household, the interviewer lists all the household members, from oldest to youngest, in the household roster. Those who are away during the interviewing period, are listed separately as non-qualifying members.

IHS

IHS are expert demographers, providing comprehensive global market, industry and technical expertise. For Establishment Survey, IHS provides up-to-date population estimates and includes the latest boundary information, ensuring ongoing representation.

Kish grid

A method developed by statistician Leslie Kish in 1949, used to randomly select the member of a household to be interviewed by way of a pre-assigned table of random numbers.

Language

Own language

Language spoken most often by the interviewed respondent

Other own languages

Other languages spoken by the respondent

Household language

Language spoken most often in the household

Languages read and understand

All languages read and understood by the interviewed respondent

Municipal code

Identifies the official municipality of the Small Area. Municipalities and their boundaries are officially recognised areas as defined by the Municipal Demarcation Board. Eight municipalities have been classified as 'Metropolitan Municipalities', namely Ekurhuleni, City of Johannesburg, City of Tshwane, Mangaung, Buffalo City, Nelson Mandela Bay, eThekweni and City of Cape Town.

Personal income

The claimed personal income before tax or any other deductions.

PPS sampling

Probability Proportionate to Size (PPS) sampling is an approach in which the selection probability for each element is set to be proportional to its size. For ES, SAs are listed along with their associated populations (age 15+). These numbers are cumulated – effectively numbering the population. Random numbers are used to select these notionally numbered people. This results in a list of SAs in which these people live, the probability of a SA being selected then being proportional to its population size.

Proportionate sampling

The ES sample is drawn in direct proportions to the South African population in terms of province and area type, in order to preserve the purity and stability of the population data. No disproportionate sampling will be implemented, to ensure that the sample in every six-month data period is proportionate to major conurbations, as well as province.

Province

The nine provincial boundaries in the ES sample align with those of Statistics South Africa.

Respondent

The person in the household (aged 15 years or older) randomly selected using the household roster and Kish grid. The respondent completes the interview.

Sampling frame

The TNS master sampling frame lists all Small Areas (SAs) together with a number of variables at SA level, such as population size (used for SA selection based on PPS sampling), province, main place name and code, sub-place name and code, dominant race group and TNS Field Manager allocation.

Sampling Interval

The sampling interval is the number of houses to be counted between each visiting point. The interval is calculated within each SA, using the number of households per SA as a crude measure of size.

Small Areas (SA)

The Small Area (SA) code is based on the 2011 Census and is the officially-recognised classification of areas with fixed boundaries. The SAs are a very low level geographical classification of areas and in 90% of cases, the SA is the same as the Enumeration Area (EA), while the balance of SAs are made up of two or more EAs.

Starting point

The starting point is a random point, marked as an 'X' on the map supplied to the interviewer. The random starting point is generated through Geographic Information System (GIS) technology, using a random number generator to allocate a random value for the x- and y- coordinates (longitude & latitude) for a point located in the SA map.

Substitution

Substitution means the purposeful replacement of an originally selected household, in accordance with agreed rules.

Total population

Total population refers to the total number of people living in the Small Area.

Total population \geq 15 years of age refers to the total number of people aged 15 years or older living in the Small Area. In other words, this is the total number of people who could be interviewed.

Total population $<$ 15 years of age refers to the total number of people below the age of 15 years living in the Small Area. In other words, this is the total number of people who are not eligible for interviewing but who still reside in the Small Area.

3. The ES sample

The annual sample size is 25,000 interviews.

Interviewing is spread over 49 weeks of the year, and evenly across the country. This ensures that all periods, area types and provinces are consistently covered resulting in stable data.

Four (4) household interviews are completed per Small Area; therefore, 6,250 Small Areas per annum are drawn in order to achieve the annual sample of 25,000 households.

The sampling approach

The sample is drawn through strata created using province and area type as explicit variables.

The dataset used to develop the ES master sampling frame is delivered by IHS on a bi-annual basis.

TNS conducts quality assurance checks on the IHS dataset. Checks are done on the IHS dataset focusing on:

- Google Earth enabled verification of low population SAs
- Duplicates
- Range of data (e.g. no negative values for population figures)
- Current dataset vs previous IHS dataset (e.g. consistency in geography type and municipality)

The quality-assured IHS dataset forms the basis of the TNS master sampling frame which lists a number of variables at SA level, such as population size, province, main place name and code, sub-place name and code, dominant race group and TNS Field Manager allocation.

SAs identified as Industrial, Institutional, Parks and Recreation are excluded from the master sampling frame used for PPS sampling.

Area selection

Using the master sampling frame, a stratified multistage sample design is used to draw a sample of 12,500 households per wave, from 3,125 Small Areas (SAs).

Area splits follow actual population distribution as provided bi-annually by IHS.

A sampling frame of SAs at sub-sample level (per area type per province) has been developed.

A Probability Proportional to Size (PPS) approach is applied to ensure the correct and random selection of starting points per area type per province. The PPS is

generally accepted to be the most appropriate approach to sampling for programmes such as the Establishment Survey.

Small Areas (SAs) are then randomly selected from the sampling frame according to the area type and province, and used by the TNS Field team to select households to interview.

Household selection

Once the SAs for interviewing have been selected, the households to interview in that area must be identified.

Using the sampling interval, and referring to the random starting point, visiting points and substitute points are identified before continuing on to identify the person to be interviewed in each household. One substitute point is allowed for each visiting point, that is, if four interviews are to be completed per SA, a total of eight visiting points is identified per SA (four original points and four substitute points).

To identify the visiting points, interviewers will be required to walk in a clockwise direction (in a spiral formation) and count the number of houses on both sides of the road until the sampling interval number is reached.

Once the visiting points and possible substitute visiting points have been established, the interviewer will go to the first visiting point (starting point) and begin interviewing.

Respondent selection

To identify the respondent in the home to be interviewed for the placement survey, a Kish grid is used. This approach requires the interviewer to record all the individuals at that household who are 15 years or older and who reside and eat there four or more nights of the week.

By using the Kish grid, the interviewer will select the respondents at random and regardless of gender, resulting in a broad range of responses to the questions included in the survey. In other words, not just heads of household or 'gatekeepers' will be chosen for the interview.

Following the outlined process produces a randomly selected area, randomly selected households within that area, and randomly selected persons to interview for the survey within the selected household.

Substitution

There are two types of substitution which may be applied: area substitution and visiting point or household substitution. Strict controls are in place to manage substitutions and there is a strong focus on keeping substitutions to a minimum.

a) Area substitution

In terms of area substitution, an SA may be substituted for another similar SA in the same area only if that area is inaccessible for a specific reason. Reasons may include that there are no homes in that area or it is unsafe due to excessive crime. In these instances, the entire SA would need to be replaced with a similar SA in the same area. Where possible, we use our in-house Geographic Information System (GIS) and Google Earth to check the validity of claims.

It may also happen that the interviewer in an SA selects visiting points and substitute visiting points but cannot complete any interviews in that area for valid reasons. For this scenario, the SA would need to be substituted in its entirety. In some instances, the interviewer may only be able to complete some of the required number of interviews for any given area, in which case, the area may be 'closed' with fewer completed interviews. A substitute SA will then be selected for the completion of the outstanding balance.

b) Household substitution

An interviewer may use a substitute visiting point within the SA only if they have been back to the originally selected home on four separate occasions at different times of the day and on different days of the week (for Metro and Urban only) and have been unable to complete an interview with the randomly chosen respondent for a valid reason, such as refusal or in the case of the visiting point being a child-headed household.

Up to a total of eight visiting points are allowed per SA (four original points and four substitutes).

No substitution of individuals within a household is allowed. If the selected individual is not present at the time of the original placement interview, the interviewer will make an appointment to return to conduct the interview.

4. Interviewer training

Before the start of fieldwork, face-to-face briefings across the provinces were conducted with all interviewers. Any changes to the questionnaire made were then briefed with interviewers again.

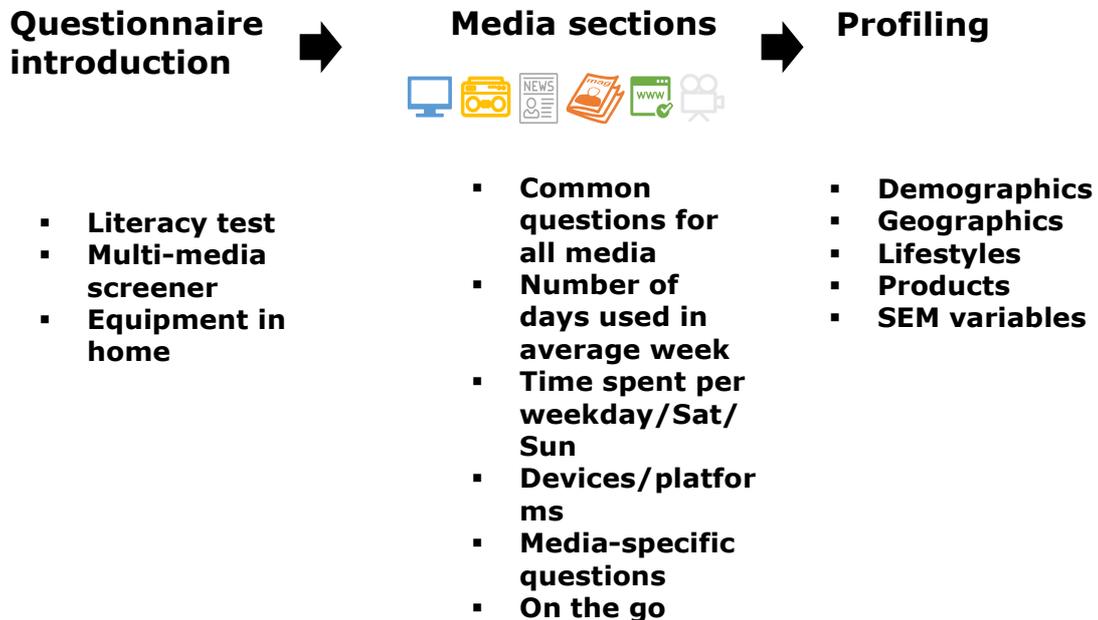
All interviewers are supplied with interview instructions and a mock questionnaire to refer to in field.

The interviewers are assigned areas they are familiar with and will therefore be able to properly engage with respondents in their preferred language and relate to them well.

5. Survey technique

The ES Questionnaire

The questionnaire uses a “top-down” structure (not brand led). The graphic below illustrates the main questionnaire components.



Interview equipment and software

The interview is conducted via CAPI on a tablet, with the use of paper showcards.

Interview day

The interview days are spread across all days of the week.

6. Quality control

Interviews are back-checked independently of the field teams. Back-checks are done telephonically or face-to-face. A minimum of 20% of all questionnaires (interviews) and a minimum of 10% of each interviewer’s work is back-checked.

7. Sample Weighting

After data collection, the data will be appropriately weighted, to align the sample with the target population as closely as possible.

The ES data is weighted to match the data released by IHS. Weights are applied within each province/area type combination, of which there are twenty-three. The ES data is weighted on both an individual and a household level.

The following variables are used to weight the data on an individual level:

- Province
- Area type
- Race
- Gender
- Age

The following variables are used to weight the data on a household level:

- Province
- Area type
- Race

Random Iterative Method (RIM) weighting is applied. RIM weighting uses a mathematical algorithm to help provide an even distribution of results across the province and area type, while balancing the required weighting variables with predetermined totals.

It weights the specified variables simultaneously and disturbs each variable as little as possible. This in turn protects the integrity of the data and supports the precision of research estimates.

A weighting efficiency report is produced on a bi-annual basis.

8. SEM (Socio-Economic Measure)

The SEM has been developed using a correspondence analysis as the basis. Fourteen variables are used to calculate the SEM score. Each individual is given a score from 0 (low socio-economic living) to 100 (high socio-economic living). The continuum is currently divided into ten equal groups (SEM 1 to SEM 10), but users have the flexibility to spread the scores in any way they wish.

9. Reporting

A six-monthly presentation is produced on data from the six-months preceding the report.

In a specific year the six-monthly data refers to the following waves:

Wave 1: July – December

Wave 2: January – June

The data is released to the industry through Third Party Software Providers, approved by the BRC and PRC.