How to Sample Drug Users and Other ‘Hidden’ Populations:
Methodological Implications for Qualitative Marketing Researchers

Abstract

Social marketing has been employed to change the behaviour of consumers for non-commercial reasons connected to both the individual and the public good, for example in protecting women against domestic violence, and in facilitating government rehabilitation programmes for drug users and others. Marketing researchers are often tasked with studying the effectiveness of government rehabilitation programmes among drug users by using survey questionnaires and interviews. However, there are accessibility and methodological problems in finding respondents among this hidden population. This paper therefore proposes the use of stratification techniques for both probabilistic and non-probabilistic snowball sampling to investigate hidden populations. Stratification is defined as the division of the target population into strata or groups that are linked by some demographic characteristic, such as age or income, in order to enhance the research findings. In reality, since most respondents (or research topics) are heterogeneous and knowledge of a particular topic is randomly distributed in the population, the use of stratification leads positively to sampling diversity and efficiency, i.e. higher precision, leading in turn to enhanced quality of information and more robust findings, which can then be used to design strategic marketing policy applicable to the population under investigation as a whole.
1. Social Marketing and Hidden Populations

Traditionally, marketing used to be employed by profit-making companies to sell products or services; however, marketing is now also used widely in the public and voluntary sectors (Wood, 2012; Hastings and Angus, 2011; Weinreich, 2010; Kotler et al., 2009). This public and voluntary sector marketing, which is also known as ‘social marketing’ (non-profit), has been employed to change the behaviour of consumers for non-commercial reasons connected to both the individual and the public good. It has been employed to promote the concept of ‘Corporate Social Responsibility’ in the commercial sector (Grigore, 2011), to protect women against domestic violence (Vanya, 2006), or to facilitate professional services in assisting government rehabilitation programmes for drug users (Rahtz and Sirgy, 2000; Bickel and DeGrandpre, 1995).

In the application of social marketing, marketing researchers are often tasked with studying the effectiveness of government rehabilitation programmes among drug users and drug dealers by using survey questionnaires and interviews. However, there are accessibility and methodological problems for marketing researchers in finding respondents among this hard-to-reach group or hidden population. A hard-to-reach group is defined here as a group which is difficult to approach because no immediate sampling frame is available. This means the data is not publicly available, for example, in public directories or on electoral records. Such data (respondents’ details) may be available from official institutions such as police authorities, probation offices or courts. However, under the Data Protection Act 1998 this data cannot be disclosed to anyone. This creates difficulties for marketing researchers wishing to study such populations. The purpose of this paper is therefore to propose a sampling strategy (both non-probabilistic and probabilistic) to access hard-to-reach groups. It is intended that this paper will be helpful for marketing researchers to design a robust sampling methodology, in turn leading to the production of more relevant marketing policy implications. This paper employs conceptual and numerical discussions to achieve a diversity sampling strategy to reach hidden populations.

In this study, there are three types of hidden population or hard-to-reach group (Illustration 1, Appendix). The first group is in the white market, the second in the grey market and the third in black markets. A hard-to-reach group in the black market could be, for example, economic agents associated with illegal or criminal activities such as prostitutes, drug users, arms dealers, benefit cheats, shoplifters, tax evaders, etc. The group from white markets includes physically disabled or diabetic consumers, violin players, etc. Hard-to-reach groups in grey markets include tax avoiders, in situations where no clear legislation or precedent exists. Here, a tax avoider (grey markets) is defined as a person who attempts unethically to reduce the amount of tax that is payable according to his or her means, but still remains within the law by establishing an offshore company in a tax haven. While, tax evaders (black markets) are individuals who evade the payment of taxes by illegal means, for example by under-declaring income, overstating deductions, etc (Malkawi and Haloush, 2008; Bridges and Green, 1998; Simmons and Cheng, 1996).

2. Snowball Non-Random Sampling

If marketing researchers wish to write a commissioned research paper to evaluate the effectiveness of drug prevention and rehabilitation programmes implemented by a Local Government Authority, they need to investigate the buying behaviour of all prospective buyers, including drug users and dealers etc. It may be extremely difficult to locate the respondents in these situations since the population can be regarded as hidden and rare. Also, there is no exhaustive public list of population members for these groups, so in practice researchers tend to use ‘traditional’ snowball sampling technique (Franks et al., 2009; Kalton and Anderson, 1986; Berg, 1983; Goodman, 1961). In addition, from a legal
standpoint, drug dealers or drug users are unlikely to participate in a survey owing to the risk of imprisonment, public embarrassment, etc. In practice, marketing researchers normally use monetary incentives to enable the use of snowball or referral sampling without stratification (Illustration 2, Appendix). Illustration 2 depicts a snowballing process in which only six drug users are needed to participate in the study.

Snowball or referral sampling is a non-probabilistic sampling technique which is widely used when the population is hidden and rare, and is therefore difficult to locate. In this technique, marketing researchers first recruit a few initial respondents, who are then asked to refer other respondents within the population (Franks et al., 2009; Kalton and Anderson, 1986; Berg, 1983; Goodman, 1961). A snowball or referral sampling technique is practical because when marketing researchers are able to identify one or two respondents willing to participate in the research, it is likely that they will know other individuals in their area and can help to locate them. The same process is used for other hidden populations wishing to keep their identity secret, such as benefit cheats, illegal immigrants, ex-offenders, etc.

However, one problem with this approach is that, whilst the real population is heterogeneous, the snowballing technique tends to attract more homogeneous respondents. Snowballing respondents tend to recruit those whom they resemble in age, education, income, ethnic origin, nationality and other non-heterogeneous characteristics (Binsardi and Green, 2012; Binsardi and Mclean, 2008). Therefore, although this technique can dramatically lower search costs, this comes at the expense of introducing bias because the technique itself reduces the likelihood that the sample will represent a cross section of the population. Accordingly, only well-connected homogeneous respondents tend to be over-sampled. Correspondingly, because of its homogeneity, a snowball sample is lacking in diversity. Sampling diversity is a methodological prerequisite for yielding robust and refreshed findings. As a result, marketing researchers tend to obtain ‘static or monotonic’ findings and no new refreshed ‘insights variation’ is discovered.

In order to increase sampling diversity, stratification is encouraged. Stratification is a sampling technique wherein the researcher divides the entire population into different subgroups or strata, then selects the final subjects proportionally from the different strata (Binsardi and Green, 2012; Thompson, 1991; Golder and Yeomans, 1973). For example, if social marketing researchers wish to investigate the effectiveness of government rehabilitation programmes, they will use snowball sampling to recruit respondents (drug users). Understandably, snowball sampling is an attractive strategy to obtain more drug users. However, this paper proposes that marketing researchers should first stratify initial respondents for example by gender (male and female) or by income (low, middle, upper) (Illustration 3, Appendix). By undertaking this stratification, more variation and sampling diversity will be achieved, which will generate new insights in relation to drug users’ behaviour – it is to be expected theoretically that the behaviour of drug users varies with gender and income levels.

Illustration 3 shows that (based on a behavioural theory) there are two possible stratifications affecting drug users’ behaviour, namely income and gender. If income stratification is chosen, then two low-income drug users, two drug users of middle income and other two drug users of high income can be selected. This income stratification yields a total of six drug users. Alternatively, if gender stratification is selected, three male and three female drug users can be recruited to produce total six respondents. Comparing snowball sampling without stratification (Illustration 2) and with stratification (Illustration 3), snowball sampling with stratification provides a superior methodology because sampling diversity will be achieved. This is because the snowballing process, which starts after stratification, makes the population of drug users become heterogeneous. Conversely, sampling without stratification (Illustration 2) leads to over-sampling of homogenous respondents.
Therefore, the benefit of snowballing with stratification is to reduce the variance \( (s^2) \) of a population. Reduced variance means smaller errors; hence, the findings will be more robust, as smaller errors means more precise explanations of samples to explain the whole population of drug users. To illustrate this, let us assume a finite population of nine drug users \((N=9)\) has the following population values \((X)\) measuring population characteristics of drug users’ behaviour:

\[
1, 3, 2, 2, 3, 1, 3, 2, 1
\]

Accordingly, the population mean and variance can be computed as follows:

\[
\mu = \frac{\sum y}{N} \quad \text{and} \quad s^2 = \frac{\sum (y - \bar{y})^2}{N-1}
\]

Applying these formulae to the above hypothetical data, the population mean \((\mu)\) and population variance \((\sigma^2)\) equal 2 and 0.75, respectively. However, if snowballing with stratification is selected, then the population can be divided more efficiently into three strata (stratum 1 for low-income drug users, 2 for middle income, and 3 for high income). Correspondingly, the population can then be divided up into three homogeneous strata. The stratum sizes are \(N_1, N_2\) and \(N_3\), where \(N=N_1+N_2+N_3\). Within each stratum, respective samples of sizes \(n_1, n_2\) and \(n_3\) can be taken. Note: the sampling is independent in the different strata \((N, \mu \text{ and } \sigma)\) refer to the population data, while \(n, \bar{y} \text{ and } s^2\) refer to the sample data.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y)</td>
<td>1, 1, 1</td>
<td>2, 2, 2</td>
<td>3, 3, 3</td>
</tr>
</tbody>
</table>

While a snowball sampling technique without stratification, several samples are possible. Let us assume the following samples:

\[
\begin{align*}
1, 1 & \quad 1, 2 & \quad 1, 3 \\
1, 3 & \quad 2, 3 & \quad 3, 3 \\
\end{align*}
\]

The sample mean \((\bar{y})\) can range from 1 to 3, while the sample variance \((s^2)\) ranges from 0 to 1. However, if marketing researchers employ the above stratification (strata I, II and III) and select one, two or three samples from each stratum, then possible samples can also be provided below.

\[
\begin{align*}
n=1 & \quad 1, 2, 3 \\
n=2 & \quad 1, 2, 3, 1, 2, 3 \\
n=3 & \quad 1, 2, 3, 1, 2, 3, 1, 2, 3 \\
\end{align*}
\]

In the case of stratification, the sample mean \((\bar{y})\) should always be 2, which is always the same as the population mean \((\mu)\). Correspondingly, the sampling variation would be zero. That is to say, stratification improves the precision of the findings. Following the above numerical illustration, a stratum is useful when the population is heterogeneous and it is possible to create homogenous strata within the population. That is, the units within each stratum should be close to each other but the means of the strata should differ as much as possible to enhance sample diversity.

Accordingly, the benefits of stratification are, firstly, that snowballing with stratified samples can provide greater precision in the findings than snowballing without stratified samples of the same size. From the first point, it can be deduced that because snowballing with stratified samples provides greater precision, secondly, it often requires a smaller sample, which is more cost efficient. Referring to Illustration 3, it would be more robust to recruit three respondents from snowballing with stratification than six homogenous respondents from snowballing without stratification, in this case one respondent of low income, one of middle and one of high income. Thirdly, snowballing with stratification can guard against an ‘unrepresentative’ sample (such as all-male or all-female respondents from a
mixed-gender population). Fourthly, it makes marketing researchers’ tasks more organized by focusing on individual strata. For example, drug users in the UK can be re-stratified against four locations, namely Wales, England, Scotland and Northern Ireland. Hence, in this case, the organization and administration of the research will be easier and more manageable for marketing researchers.

It is not only demographic variables that can be used to stratify respondents; time and contextual variables can also be used (Bryman, 2012; Binsardi and Green, 2012). For example, it is important that people or events are stratified at different times of the day and week so that inferences about what they do are not limited to their habitual behaviour during one particular time frame. In addition, people behave differently in different settings so the research may require that a variety of stratified settings be selected.

The spirit of snowballing with stratification is similar to the concept of the purposive sampling technique with maximum variation. The relevant sampling literature (Bryman, 2012; Denzin and Lincoln, 2011; Morse, 2004; Rao, 2000; Raj, 1968) describes several purposive sampling techniques, such as homogeneous sampling, typical case sampling, extreme case sampling, critical case sampling, total population sampling, maximum variation sampling, and expert sampling. Purposive and maximum variation sampling is similar to snowballing with stratification in that it aims to achieve as much as variation in the data as possible in order to obtain more precise and robust findings. Otherwise, the findings will be less precise and more anecdotal, meaning that any strategic social marketing policy implications of the findings with regard to the drug users will not be optimal because they cannot be applied to the whole population of drug users. The disadvantage of stratification is that it may require more administrative effort (i.e. time and other resources); this is surely justified, however, if more refreshed findings are obtained owing to stronger methodology.

Unfortunately, another drawback of this ‘traditional’, ‘non-random’ snowball sampling technique is that the findings can be generalized and sampling errors cannot be estimated (Rao 2000, Stuart 1984, Cochran 1946, 1977, Raj 1968). To answer these problems, Respondent-Driven Sampling (RDS), a breakthrough in sampling methodology, was developed by Heckathorn (2007, 2002) and Salganick and Heckathorn (2004). This will be conceptually discussed in the next section.

3. Snowball Random Sampling

Respondent-Driven Sampling (RDS) is also known as snowball random sampling, and solves the classic dilemma of ‘traditional’ snowball non-random sampling without stratification (Illustration 2). In traditional snowball sampling with stratification, respondents are recruited in a non-random way within the strata. While in RDS, however, respondents are selected randomly within the strata. While the traditional snowball sampling method uses a direct estimate from sample to population (Illustration 4, Figure 1), a respondent-driven sample employs an indirect method (Illustration 4, Figure 2) by initially using a sample to make estimates about the social network connecting the population of respondents.

This predicament was initially without a solution, since snowball sampling precipitates biases of unknown size and unknown direction (Salganick and Heckathorn 2004). Accordingly, any inference made based on snowball sampling would be merely homogeneous; subjective findings do not reflect the diversity of the population. Using RDS, marketing researchers ask initial respondents to refer their next snowballing respondents, then forming a network-based sample. The networked samples can then be stratified according to demographic, cultural and other characteristics. The sampling process (randomization) begins when the marketing researchers select a small number of seeds, i.e. the first respondents to be included in the sample. This RDS methodology is based on a synthesis and extension of two areas of mathematics, namely Markov chain theory and biased network theory (Heckathorn
2007, Salganick 2006, Salganick and Heckathorn 2004). These theories indicate that the estimates of RDS are asymptotically unbiased, no matter how the seeds are selected. In addition, the RDS sample can be efficient if the total population is identifiable (Sudman and Kalton 1986). There is also some evidence that RDS is consistent with the idea that respondents recruit randomly from their networked friends by combining the breadth of coverage of network-based methods with the statistical validity of standard probability sampling methods. This makes it possible to draw more robust, diverse and valid samples from previously unreachable, thinly spread groups.

Further stratification procedures can be found inter alia in Ekman (1959) and Dalenius and Hodges (1959). When a theoretical population consists of G strata of sizes \(N\), the theoretical observation of the strata can be represented by \(x_{gi}\), \(g = 1, 2, 3 \ldots G\) and \(i = 1, 2, 3 \ldots \Ng\). The total, mean and variance of the observations of the \(g\)th stratum can be given by:

\[
X_g = \sum_1^{N_g} x_{gi} \quad \text{and} \quad \bar{X}_g = \frac{X_g}{N_g}
\]

\[
S_g^2 = \frac{1}{N_g - 1} \sum_1^{N_g} (x_{gi} - \bar{X}_g)^2
\]

From this expression, the advantages of random stratification in GT are, inter alia, estimates for each stratum that can be obtained individually and the total and mean of the population that can be estimated efficiently with higher precision. This leads to an enhanced quality of data diversity compared to that derived from snowball sampling without stratification.

3. Conclusion and Direction for Further Research

It has been noted in the mainstream qualitative literature (Denzin and Lincoln, 2011; Lewis-Beck, Bryman, Liao, 2004; Auerbach and Silverstein, 2003; Strauss and Corbin, 1990) that qualitative analysis is related to non-probabilistic sampling (‘interpretivist’ paradigm) while quantitative analysis is more positively linked with probabilistic sampling (‘positivist’ paradigm). Samples in qualitative analysis are usually non-random (non-probability) samples because qualitative research is more concerned with generalising findings to theory development (‘interpretivist’) rather than to populations (‘positivist’). Due to this ‘traditional divergence’ in the positivist versus interpretivist debate, most scholars have been reluctant to apply a more robust or innovative sampling technique. Instead, they have merely used purposive sampling for qualitative research, which blocks the development of ‘methodological innovation’ in qualitative marketing research.

This paper indicates that the use of snowball (random and non-random) sampling with stratification increases efficiency and convenience for marketing researchers, enabling them to focus on their research efforts regarding subpopulations of special interest. In addition, using a snowball sampling technique with stratified samples yields more accurate and robust findings. Such research will produce more relevant strategic marketing implications for the population of interest. Recent research has yielded a more random snowballing technique, the so-called ‘respondents-driven’ technique. Further research should be directed towards investigating the efficiency property of stratification in other types of ‘qualitative’ or non-random sampling, such as quota, adaptive, judgmental, and purposive sampling. This research is intended to fill a methodical gap in qualitative marketing research to guide researchers to select samples for designing social marketing policy.
References


______ and Strauss, A. 1967. The Discovery of Grounded Theory, Aldine, Chicago, IL, USA.


9


Morse, J and Richard, L. 2002. Read Me First for a Qualitative Guide to Qualitative Methods, California, Thousands Oaks, USA, Publisher: Sage Publications.


Appendix

Illustration 1
Hard-to-reach groups classified by its legal status

Illustration 2
Snowballing Process without Stratification
Illustration 3
Snowballing Process with Stratification

- Drug Users Samples=6
  - Possible Strata Income
    - Low Income
      - Snowballing Process
        - 2 Drug users
    - Middle Income
      - Snowballing Process
        - 2 Drug Users
    - High Income
      - Snowballing Process
        - 2 Drug Users
  - Possible Strata Gender
    - Male
      - Snowballing Process
        - 3 Drug Users
    - Female
      - Snowballing Process
        - 3 Drug Users

Illustration 4

Figure 1
Traditional Sampling Estimation

Figure 2
Respondent Driven Sampling

Adapted From