

STATISTICS IN TRANSITION new series, Summer 2014
Vol. 15, No. 3, pp. 477–486

BOOK REVIEW

***Sampling Elusive Populations: Applications to studies of child labour*, by Vijay Verma.** International Labour Organization, Geneva. 2013. XIX + 821 pp. ISBN 978-92-2-128321-8.

Reviewed by Włodzimierz Okrasa

Surveying populations difficult to find or to reach (*hard-to-reach*), elusive or invisible, mobile or intentionally avoiding statistical observation for some reasons, presents not only additional problems to those being faced by a survey maker in the context of researching regular populations. The problems involved in the former are often qualitatively different from those in the latter, at practically each stage of the process, from sampling populations with non-existing or imperfect frame, through data collecting and processing, to estimating the looked-up parameters. Comparison of methodologies focused on researching basically the same units – such as child workers – in both types of contexts can serve as a laboratory case for demonstrating differences between the problems specific to surveying regular and irregular populations. We are now in a perfect position for making such type of methodological comparisons as Vijay Verma – the author of the widely acclaimed book (cf. Kordos' review in this journal, 2008) on *Sampling for household-based surveys of child labour* prepared within the framework of the ILO Statistical Information and Monitoring Programme on Child Labour (SIMPOC) – has elaborated a complementary volume on *Sampling Elusive Populations: Applications to studies of child labour*, also published by the International Labour Organization.

This book offers the comprehensive treatment of methods specifically designed to meet challenges posed by the irregular type of populations, in general, with extensive presentation of their applications to formal and practical issues in researching children's work. As such, the book fills a key gap in the survey methods literature, on the one hand, while promoting the approach especially suited to dealing with this complex research objects. In a consequence, its usefulness goes far beyond studying child labour within the public statistics, but contributes also to economic and social analysis of labour market and work activities. With respect to the former, the book addresses also several problems fundamental to the methodology of surveying hard-to-survey populations, a

"common but not universal characteristic of (which) is that they are rare populations for which no separate sampling frames exist" (Kalton, 2014, p. 401). Such populations are becoming of growing interest to statisticians inspired also by economists, demographers and other social scientists in view of the fact that in addition to the problems with forming a list of potential respondents – hard-to-sample populations – they are more and more frequently being faced with insurmountable sometimes difficulties to persuade members of such populations to take part, or to be interviewed. A good indication of the rising demand for operating knowledge, i.e., for manuals on dealing with such a type of challenges and problems – which are covered jointly by the term hard-to-survey populations – might be a book just under the title *Hard-to-survey populations*, edited by Roger Tourangeau et al., 2014, which was published recently. In addition to the first of the two ILO-issued books by Vijay Verma, this seems to complement its contents and the expected audience from the side of general survey methodology, including the post-sampling survey making procedures.

At a glance, the enormous scope and richness of the issues addressed in Vijay Verma's book (*Sampling Elusive Populations* – for convenience an abbreviation SEP will stand for it in this text) can best be characterized by indicating its size – 821 pages – and composition: fourteen chapters, written as self-contained thematic sub-manuals. With his extraordinary experience and knowledge of methods of surveying child labour – including awareness of the fact that "collecting comprehensive data on child labour is a challenging task, and no single survey method can satisfy all data needs" (SEP, p. 1) – the Author begins with useful introduction to the context of the questions covered by his manual. Starting with sketching main differences between household-based child labour survey – which misses working children not living at home – and the specificity of such an elusive population as presented by labouring children (ill-defined population of a great heterogeneity, rare and highly mobile, partly hidden and reclusive, staying away from participation in a survey) the Author is constructing – not simply applying – a methodology of sampling especially suited to accounting for these characteristics. His understanding of the basic concept – elusive population – is given explicitly: "By elusive populations we mean *populations for which – by virtue of their characteristics, or of the lack of suitable sampling frames, or difficulties in obtaining the required information – adequate samples cannot be defined, drawn or implemented using the normal procedures of general population sampling.*" (*ibid.*, p. 4). Since the problems of under-coverage or of non-response may occur in a survey of populations considered 'normal', they also may be added to the *most elusive* population groups given that they are the ones of the greatest policy concern. The typology of populations in terms of

characteristics important from sampling standpoint (following Kish, 1991), along with the specification of the nature of the methodological problem involved, organize the book contents of this book.

Accordingly, this well-organized (reader-friendly) manual – which is conceived as a complementary to the previous one – concentrates on various groups of working children not living at home. Their key characteristic is a degree of connection to their household as the starting point for being identified or traced in a way. Also, some categories of children living at home and working away are included due to insufficient information available to their parents/guardians on the location or type of work they are doing. All the possible to imagine situations which can appear in a variety of configurations, given different circumstances of work and relations to the household, make the survey (sample) designer open to any of the theoretically envisioned approaches, including the need to enumerate working children at the place of their work based on a sample of such places.

An overview of the book's contents

After providing the reader with preliminary ideas and concepts of working children as an example of elusive population – an object of chapter 1 – the Author concentrates on systematic presentation of available sampling methods, in a way facilitating also the proper choice of the one, most effective in the concrete problem context.

In chapter 2, *Child labour situations, data needs and sources*, diverse and specific to the particular situation child labour problems are discussed as the understanding of the situation to be studied is considered by the Author essential for choosing an appropriate survey methodology and sampling strategy. The chapter illustrates the variety of situations and types of child labour in order to provide the necessary background for the diverse sampling techniques discussed in subsequent chapters. The variety of forms of child labour covers diverse sectors: child domestic work; agriculture including commercial crops; fishing and aquaculture; mining and quarrying; manufacturing including handicrafts; construction; street work and the informal sector; and also various 'unconditionally worst forms of child labour' including child trafficking, commercial sexual exploitation, forced or bonded labour, engaging or living in armed conflict, and children's involvement in illicit activities, in particular in drug trafficking.

It is followed by an overview of different data sources for different types of child labour: household-based surveys; supplementary sources or surveys (school-

based surveys, community-level inquiries, general national household surveys, censuses, other secondary sources); employers' surveys; establishment surveys; baseline surveys and studies; and rapid assessments. Two major strategies of generating data on child labour – household-based surveys and rapid assessment studies – are characterized briefly. The former as a large probability sample of the general population; the latter as a small-scale but intensive survey. The two form two ends of the range of application of the various sampling techniques addressed in the book.

Chapter 3, *Basic sampling and estimation procedures*, is devoted to providing a reminder of some basic principles concerning sample design and selection which underlie the more specialised techniques discussed in this book, such as: principles of probability sampling, common departures from simple random sampling (stratification, clustering, unequal selection probabilities), probability proportional to size (PPS) sampling, and systematic sampling. Also, the chapter reviews basic principles concerning weighting of sample data and estimation from a sample, along with sources of information for weighting and presentation of a step-by-step procedure for weighting (computation of design weights, adjustment for non-response, calibration against external standards, and trimming and scaling of the weights).

The problem of sampling from imperfect frames is discussed in chapter 4, *The sampling frame*. It starts with reviewing shortcomings of sampling frames and basic concepts (the survey population, the sampling frame for single-stage and for multi-stage sampling). The surveys cover selected sectors; most of the establishments in the sectors are small and a high proportion employ child labour. Several common aspects concerning sampling frames are presented using an illustrative material – the problems such as surveying a population in the absence of an existing sampling frame; including the cost and quality implications of the quantity of information to be collected for each unit during the operation; economising research by sharing the costs between different surveys; using the listing operation for making substantive estimates; and special problems related to the type of units in the frame (e.g. establishments versus other locations where working children are found). Basic requirements and desirable quality, efficiency and cost-related properties of *area frames* are also discussed. The problems of *list frames* are considered from a practical perspective, focusing on the correspondence between listing, sampling and analysis units.

The second theme of the chapter is developing and explaining such important the concepts as: (i) correspondence between sampling and analysis units (any analysis unit is associated with *at most one* sampling unit in *direct sampling*, otherwise it would be 'indirect sampling'); (ii) *sampling with multiplicity*, as the

multiplicity estimator links many of the sampling techniques discussed in this book. Situations when the sample has to be obtained by exploiting links between analysis units themselves, *link trace sampling*, are also addressed (for being discussed in later chapters).

Chapter 5, *Sampling establishments employing children*, discusses sampling aspects which apply equally to both small and informal sector establishments and to larger establishments, as they differ significantly in terms of sampling considerations and procedures. The main difference concerns the selection of establishments – for large and medium-sized establishments samples are often selected directly from lists, and the chapter describes sampling procedures for this type of selection procedures. However, the second type of design concerns samples for surveys of small and informal sector establishments which, like households, are small-scale, numerous and widely dispersed in the population. The commonly used samples for small and informal sector establishments are area-based and involve two (or sometimes more) stages of sampling. The technical issues discussed include: (a) characteristic features of small and informal sector establishments and their consequences for survey design; (b) the choice between integrated multi-sectoral and separate single-sector surveys; (c) stand-alone versus surveys attached as modules to other surveys; (d) the construction and use of 'strata of concentration' of different types and sectors of establishments to control distribution of the sample; (e) procedures for selecting establishments within sample areas; and (f) issues in survey implementation.

The rare populations are discussed in chapter 6: *Sampling rare populations*, the characteristic feature of which is that sampling the whole population with normal procedures does not yield a representative sample of adequate size for the subpopulations of interest because of their small size. In surveying different types of child labour, the rare populations of interest – working children – are generally unevenly distributed among the general population of children. The Author discusses five aspects of the strategy: (1) locating concentrations of the rare population using existing large-scale sources; (2) partitioning the frame according to the degree of concentration of the rare population (using different techniques); (3) oversampling strata of concentration, making use of the patterns of concentration identified; (4) listing, screening and two-phase sampling, aimed at the identification and sampling of the final elements (households, children); (5) employing special procedures to increase selection probabilities of units in the rare population and thereby increase the achieved sample size. There are a number of other procedures discussed in subsequent chapters (such as multiplicity sampling, multiframe sampling and adaptive cluster sampling) and the common link between them is that they involve *sampling with multiplicity*.

Multiplicity sampling is discussed in chapter 7, focused on situations where the approach may be useful in surveying the rare populations of labouring children. Since the basis of multiplicity sampling is the relationship between sampling units and analysis units, sampling with multiplicity arises when an analysis unit is linked to *more than one sampling unit*. The Author discusses potential advantages and uses of multiplicity sampling, identifying situations where multiplicity sampling may be useful, but also addresses its limitations and the problems of the method. For instance, reporting biases are often larger for multiplicity counting rules than for ordinary unitary counting rules. Another concern is the increased complexity. There can also be serious ethical, confidentiality and privacy concerns in using the method. Procedures for estimation with multiplicity sampling are also explained in this chapter. The standard *multiplicity estimator* takes the weights as inversely proportional to the unit's multiplicity.

The next, 8th chapter, is devoted to **Multi-frame sampling**, discussed in the context of child labour surveys, generally, in order to reduce coverage errors when no single sampling frame can provide a complete representation of the target population. Typically, the multiple sampling frames overlap and procedures such as constructing a new single frame without duplicates or by accounting for the duplicates in the estimation procedure need to be used. Therefore, the Author presents the main methods of removing the duplicates and constructing non-overlapping frames, as well as the main procedures for accounting for duplications and estimation from overlapping frames. Both types of situation when multiple frames can involve multiplicity in the selection of units are considered: either a unit may appear in more than one frame, or within any of those frames the unit may appear more than once. The chapter considers practical aspects of implementing this procedure in the context of a child labour survey.

Adaptive cluster sampling discussed in chapter 9 is a technique designed to obtain more adequate and efficient samples for a population which is rare and very unevenly distributed. The technique specifically involves selecting an additional sample in the neighbourhood of points where a concentration of the population of interest is found during implementation of the initial sample. It is presented as being most effective when the population of interest tends to be concentrated in relatively few and large clusters, but little information is available on the extent, location and patterns of its concentration – e.g., such populations include street children, children engaged in street trades and child beggars. Several technical aspects are discussed as well, such as: unequal unit selection probabilities; stratification with adaptive sampling; multistage sampling; multivariate criteria for adaptive sampling; adaptive sampling using 'order

statistics'; arbitrary rules for stopping the adaptive process; problem of imperfect detectability; and aspects of the estimation procedures with adaptive sampling. In addition to discussing issues involved in its implementation, the procedure is illustrated in detail on the basis of an artificially constructed small population. The illustration demonstrates how adaptive sampling can help in locating large concentrations of the population of interest by increasing the chance of their appearance in the sample, and hence also in obtaining a larger number of elements of interest (such as children working in a particular sector).

Sampling mobile populations is described in chapter 10, including special problems and issues which arise in this approach, while stressing that the concept of 'mobile population' is more general than simply not having a fixed place of residence or work – sometimes it is *necessary* or *preferable* to sample and enumerate units through their *mobility* (movement).

The following questions are involved in difficulties of enumerating such populations: (i) who are the eligible respondents for the survey, and (ii) where and (iii) when to find them; also (iv) what information concerning their mobility to ask them for, and (v) how to obtain the information; (vi) how to use sample data to produce valid estimates for the population, and (vii) how to assess variances and biases to which those estimates are subject. The Author develops a framework to organise the variety of circumstances, problems and solutions encountered in sampling mobile populations – four important concepts in the framework are: sampling locations, observation points, time segments, and 'time-location primary sampling units'. Also, procedures for estimating the *probability of selection and sample weight* of a mobile individual are developed, along with quantitative expressions for variations in individual selection probabilities in a number of commonly encountered situations.

The approach discussed in chapter 11, *Capture-recapture sampling*, is devoted to sampling techniques which involve taking two (or more) independent samples from the same population and using the overlap found between the samples to estimate the selection probabilities applied to obtain those samples and the total population size. Capture-recapture applications in the social field are usually based on a combination of sample surveys and administrative sources. The Author provides instructive illustrations of application of this technique, stressing its usefulness (and robustness) even in the situation of departures from the assumed statistical model, and the fact that statistical procedures have been developed to control the effect of certain departures from the original simple model. In the Author's view, a major technical contribution of the chapter concerns the development of procedures for the estimation of sample weights in a

more general situation, along with explanation of procedures for putting together all these effects.

Controlled selection and balanced sampling discussed in Chapter 12 is a procedure to control the structure of the sample beyond what is possible with ordinary independent selection within strata. Surveys, in particular of mobile and other difficult-to-access populations, often have to be restricted to a limited area and to a small number of primary units. The Author provides several arguments for using controlled selection, especially when one has to select a small sample of primary units, but at the same time ensures that it is 'balanced' and 'representative' of the population in terms of many characteristics (or control variables). He also discusses this procedure in the context of the modern theory of balanced sampling, thus providing the possibility of dealing with a wider range of issues and more efficient sampling algorithms. These control variables may include one or more stratification variables, which correspond to controlled selection. The formal considerations are complemented by illustrations useful for a reader interested in practical applications.

The reclusive populations are discussed in chapter 13, **Snowball sampling**, which is meant as an approach to surveying reclusive populations of labouring children. In particular, the term snowball sampling refers to a convenience sampling mechanism in settings characterised by the lack of a serviceable sampling frame. A unit of the target population can enter the sample through direct selection into the initial sample, or by being identified ('named') for inclusion by someone already in the sample. There are a number of parameters which define the design of a snowball sample: the number of waves, number of contacts to request, and criteria for including a participant in the sample. The Author considers it especially useful in the context of an exploratory study, but accents its recent development and advantages within a more advanced analysis. The primary advantage of the method is its success in identifying individuals from unknown populations and from small, hidden groups dispersed within a large population; also, it provides a means of accessing social groupings which are vulnerable, etc.

Noting also its deficiencies – in particular selection bias which limits the validity of the sample – the Author outlines a simplified procedure for estimating size of a hidden population. A number of illustrations from surveys in diverse settings present both positive and negative experiences in applying the method.

Respondent-driven sampling (RDS), discussed in the last chapter (14th), is considered to be an improved variant of the usual snowball sampling, as both procedures are types of *chain-referral* sampling. As with snowball sampling, a

unit of the target population can enter the sample through direct selection into the initial sample, or by being identified for inclusion by someone already in the sample. In both cases, the process starts with a small number of peers, usually chosen non-randomly. However, as an improvement over ordinary snowball sampling, RDS is designed to produce a closer approximation to probability sampling. It incorporates features such as *the direct recruitment of peers by their peers*, a *dual system of incentives* (for participation and for recruiting), and *recruitment quotas* (e.g. a maximum of three recruits per respondent). As regards the assessment of the RDS procedure, the chapter analyses the experience with two sets of studies. The first set involves performance comparisons of the RDS and alternative sampling approaches. In the second set of examples, studies concerning assessment of RDS examine how well its procedures are implemented in terms of the theoretical assumptions of the model. For both of them numerous examples of studies undertaken for assessing comparative performance and validity of the RDS techniques are provided.

A set of three annexes and references – bibliography, author index and subject index – concludes impressive contents of this very professionally prepared book, both by the Author and the ILO editors. This book is for anyone interested in researching the labouring children, and also for all interested in surveying elusive populations, in general – students and academicians, as well as policy makers and practitioners. As mentioned earlier, it is a self-contained manual, providing the reader/user with a great piece of the subject-matter knowledge and an advanced methodology at work.

REFERENCES

- KALTON, G., (2014). Probability sampling methods for hard-to-sample populations. Chapt. 19 [in] Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates (Eds.) *Hard-to-Survey Populations*. Cambridge University Press, Cambridge, UK.
- http://www.amazon.com/gp/reader/1107031354/ref=sib_dp_pt#reader-link.
- KISH, L., (1991). Taxonomy of elusive populations, in *Journal of Official Statistics*, 7(3), 339–347.
- KORDOS, J., (2008). Book review. Sampling for household-based surveys of child labour, by Vijay Verma, *Statistics in Transition new series*, Vol. 9, No. 3, pp. 587–590.
- TOURANGEAU, R., EDWARDS, B., JOHNSON, T. P., WOLTER, K. M., BATES, N. (Eds.), (2014). *Hard-to-Survey Populations*. Cambridge University Press, Cambridge, UK.
- VERMA, V., (2008). Sampling for household-based surveys of child labour prepared within the framework of the ILO Statistical Information and Monitoring Programme on Child Labour (SIMPOC).