

Chapter 2

Methodology

2.1 Introduction

This chapter presents the methodology used to carry out the study including research design, studied area, sampling design, sample size, selection of respondents, techniques of data collection and types of tools, data collection and field experience, data entry and processing, analysis and statistical techniques applied and indices used and chapter plan for the thesis. It also portrays the field experience and approvals obtained for the study.

2.2 Research and sampling design

Research design

While the study requires describing characteristics of the refugee population, it also tries to analyse relationship existing between two phenomena of the population. Hence, descriptive research design is found appropriate for it. The study uses both quantitative and qualitative capsules of data collection. Quantitative technique includes semi-structured questionnaire, while the qualitative capsule uses In-Depth Interview (IDI) method based on few guiding questions with necessary probing into it.

Sampling design

For a sound sampling design, a good sampling framework is essential. In order to obtain the universe of the refugee population in Tamil Nadu, the Department of Rehabilitation, Government of Tamil Nadu was approached for details about refugee population. The Department periodically compiles information about the status of refugee camps in Tamil Nadu. In 2007, it published an operational manual for government officials who are associated in any way with the refugees to lay out a clear pathway to the service of refugees. As a part of the manual, detailed information is provided on the number of camps spread across districts of Tamil Nadu along with basic demographic information such as family size and population of camps in each district. This information has been used in the study as a universe for total refugee population and camps as primary sampling units. The same information has been used for selecting required number of camps and respondents from each camp.

Selection of sample size

Given the settlement of conflict in Sri Lanka, it was assumed that at least one-fourth of the refugee population would prefer repatriating to Sri Lanka. Hence, proportion who would prefer repatriation is arbitrarily assumed as 60%. With the above assumption, the following method was used to determine the sample size for quantitative component.

$$n = \frac{t^2 \times p(1-p)}{m^2}$$

where,

n - required sample size

t - confidence level at 95% (standard value of 1.96)

p - proportion who would favour repatriation (0.25)

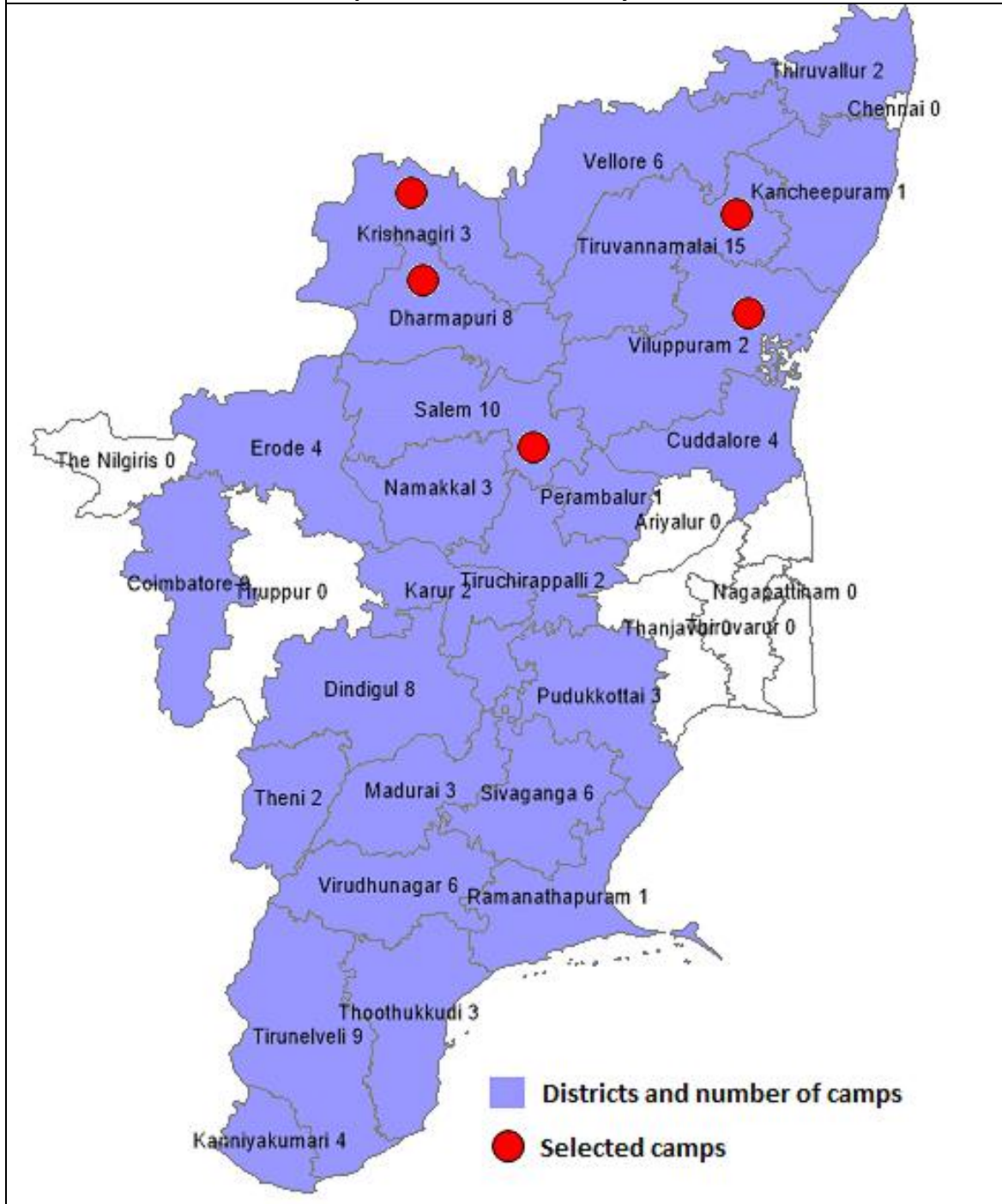
m - margin of error at 5% (standard value of 0.05)

With the above method, the sample size works out to be 288 refugee households. For the study, the same number of refugee households were targeted for data collection.

Selection of districts

In order to have representation of geographical areas, it was found ideal to select five districts, one each from east, west, north, south and central part of Tamil Nadu. While such selection would have given a fair representation, the researcher had operational challenges for this selection. Hence, based on the availability of permission from district and camp authorities, the selection of camps was confined only to the northern, north-eastern and central parts of Tamil Nadu. Also, care was taken that each district will have only one camp. There were few districts where permission was denied to do studies of refugee population in camps and hence those districts were not selected. The details on geographies of selected camps are presented in figure 2.1. One commonality about all these camps and most of the refugee camps are that they are mostly situated in the rural areas.

Figure 2.1 District-wise distribution of Sri Lankan refugee camps in Tamil Nadu and camps selected for the study, 2007



Selection of households and respondents

Initially, total population size of each camp was obtained. The required number of households was proportionally distributed per camp based on the camp population size. However, due to some other operational challenges encountered in Villupuram camp, data collection camp was stopped in the middle. Also, there was a shortfall in

targeted sample size in other camps. Finally, the total sample contacted for the study were 204 households. The breakup of sample size of five selected refugee camp is given in Table 2.1.

From each selected household, one adult member, preferably from economically productive age group of 18-59 years was selected. In few households, respondents above 59 years were also selected. In order to have representation of both males and females, care was taken to select alternative households for male and female members. The profile of respondents in terms of age and sex is presented in the next chapter. A total of 17 respondents were selected for qualitative in-depth interviews. Of these, a majority of them were from Dharmapuri and Krishnagiri camps (Table 2.1).

Table 2.1 Selected refugee camps and sample size			
Sl. No.	District	Sample size	
		Quantitative	Qualitative
1	Krishnagiri	75	5
2	Thiruvannamalai	22	2
3	Villupuram	16	2
4	Salem	29	2
5	Dharmapuri	62	6
	Total interviews	204	17

2.3 Data collection and methods

As mentioned earlier, data have been collected using both qualitative and quantitative techniques. In addition, camp level information about infrastructure such as availability of health services, transport, communication, environment and sanitation was collected. Due to the war against militant groups, outsiders not were allowed to visit refugee camps or interact with refugees. There were stringent rules in refugee camps which prohibited any refugee to visit other camps. Due to these reasons, data collection was not initiated for a couple of years after the civil war ended in May 2009. With intermittent follow-up with local networks associated with camps, it was learnt in 2011 that refugees are allowed to move outside the camps and outsiders are allowed to visit the camps. Hence, data collection was carried out during May-July 2011.

Tools used for data collection

A semi-structured quantitative questionnaire was used to collect data using face to face interview method. The respondents were primarily adult members of the households, preferably in the economically productive age group. The questionnaire had both close-ended and open-ended questions. Close-ended questions were primarily of outcome questions that are discreet and quantitative in nature. They followed a unique coding pattern that is generally applied in surveys at large-scale. In order to capture qualitative information, the questionnaire had also open-ended questions which helped in capturing aspects that had limitations to explain with any close-ended options.

In addition to a quantitative semi-structured questionnaire, qualitative tool was used to conduct IDIs. This tool had general questions with sufficient guiding questions. Adequate and relevant probing was done to explore more possible answers related to the primary questions. The qualitative interviews were conducted with the refugees who had returned at least once to Sri Lanka, either voluntarily or otherwise, and again returned back to India as refugees. The sample for qualitative component is selected from the quantitative survey which elicited information through a question on “number of moves as refugees”. Those respondents who moved as refugees more than once were selected for qualitative interviews. Responses for qualitative interview were handwritten by the researcher and later transcribed into a soft copy to use as anecdotes. Copies of both quantitative and qualitative tools are given in the Annexure.

Additional information that was not planned while designing stage of the study but collected at the time of data collection was on the camp infrastructure. Any knowledgeable person in the camp or camp head was the primary respondent of this information. Certain indicators like cleanliness and greeneries of the camp surrounding were obtained using observation method of data collection where the researcher had his own judgment to assess and rank the parameters.

Pilot study

In order to understand the studied population, a pilot study was conducted in early 2008. As a part of this exercise, informal interaction was done with 12 refugees which

focused on various aspects of the population such as livelihood, employment, employment opportunities, possession of assets and amenities, opinion about returning to Sri Lanka, environmental conditions, etc. The study helped in forming and conceptualizing the proposal. It also helped in framing open ended and close ended questions.

Field experience

Aftermath of the war against militants made the data collection more complex while approaching the refugees in the camps. As explained in earlier sections, no outsiders or civilians were allowed to visit camps without going through the proper channel which is through camp authorities. Any outsider who visited the camps is looked with suspicion. It was felt a valid identity was required to visit the camps and seek permission for interview. Hence, besides carrying the student identity card, a formal letter was obtained from the Institute stating the identity of the student/researcher and his motive to visit camps.

Prior to interview with the refugees, permission was sought from Village Administrative Officers (VAO) and respective Camp Heads explaining the purpose of the visit and the intended research work. District level officials were also approached to seek formal permission. After a detailed narration about the research work, the concerned authorities were convinced about the genuineness of the intended study and granted permission and support to conduct interviews. Besides the authorities, the camp residents raised queries about the purpose of such a study and wanted to know as to that how it would benefit them. Respondents in almost every selected alternative household enquired about the purpose of the study. Thereafter, the refugees cooperated for the study.

Before the interview, discussions were initiated with respondents or groups of people, introducing the study and the use of findings. Though there were differences of opinion among the camp people about such a study, they were finally convinced about the ongoing steps taken by the government for their benefit, and the interest taken by international community to support their well-being. The researcher being a Tamil

speaking native, it was possible to secure cooperation of the officials and convince people for the interview. The Tamil use of language impressed people and they accepted the researcher as one among them.

Besides the aforementioned practical challenges, certain rules of research methodology were. The basic problem as followed in any other studies was maintaining the sampling interval. Respondents, despite informing them about the purpose of the survey, were under the impression that they might get some assistance from the Government through this study. So every household was keen to be interviewed. Since houses in the camps are very close to each other, sometimes it was difficult to exactly locate the selected household. Further, refugees tended to increase the number of family members, especially by including separated sons, or married daughters. So they had to be probed well to get the exact number of family members residing in the camp. The smallness of in the sample size is mainly due to refusal of the permission by the authorities.

2.4 Plotting of maps

Maps were used in several instances in the study to plot the geographies such as selection of districts, Tamil diaspora across the world and place of origin of the respondents. A database management technology called DevInfo has been used to customize the study data and plot maps. In 2004, DevInfo was endorsed by the United Nations Development Group (UNDG) to assist countries in monitoring the Millennium Development Goals (MDGs). This technology has been extensively used by UN allies including UNDG and United Nations Children Education Fund (UNICEF) (DevInfo, 2005).

2.5 Various indices used

Composite indicators are widely used to assess the state of development and disparities among various regions/groups. One of the appropriate methods to arrive at relative measure of various units is by using multiple dimensions that represent the development parameter. Normalization technique is widely used in social science and in various studies when the indicators are of multiple dimension (Nanda et al, 2007;

Pasquale et al, 2010). The normalization technique helps in explaining different indicator dimensions to unique measurement unit from which composite index can be devolved.

While different technique is used to construct relative index at the regional level, a different approach needs to be applied for constructing index for unit such as household or individual. One of the strongest methodologies applied calculating household level index is articulated in National Family Health Survey (IIPS and ORC Macro, 2000). This kind of index has certain limitations as it does not differentiate between necessities and luxury parameters which are region specific. However, the advantage of using such index is it superimposes the mentioned limitations where the studies are at large scale. Same technique has been applied to construct an index that represent social network among refugees. In small surveys like present study, there is a scope to be more practical and closer to reality when applying such methods to construct index. Hence, variables are selected with caution to construct “social network index”.

Infrastructure Index

To assess the status of camps in terms of infrastructure availability, an “Infrastructure Index” was generated using the information of infrastructure availability in the camps. The identified indicators were standardized in terms of units such as kilometre, availability and non-availability, per person, etc. The indicators which have no variation across camps were dropped. Further, to explain the indicators in unique direction, the normalization procedure was applied as given below. Application of the normalisation technique uses a formula to unify positive and negative indicators. Positive indicators are those that have direct relation with the development status. The camp with lowest value will have a score of zero and the camp with highest value will have a score of one. For instance, higher data value of the indicator denotes better status and one good example is cleanliness surrounding the camp. Similarly, a negative indicator has indirect relation with the development status where lower data value denotes better status. The camp with lowest value will have a score of one and the camp with highest value will have a score of zero. For example, an indicator like the number of person using one

toilet denotes the status of a particular camp where lesser number of people sharing a toilet denotes better status. Following are the formula applied for positive and negative indicators for the composite index.

For positive indicators that would have direct relation with infrastructure availability, the procedure given below was used for normalisation. In this case, the index value is the actual value minus the minimum value in the series, divided by maximum value minus minimum value in the series.

$$(V_i - V_{\min}) / (V_{\max} - V_{\min})$$

For negative indicators that would have inverse relation with infrastructure availability, the following procedure was used for normalization. The index value is the maximum value minus the actual value in the series, divided by maximum value minus minimum value in the series.

$$(V_{\max} - V_i) / (V_{\max} - V_{\min})$$

where

‘V’ is the selected infrastructure variable of interest

‘i’ is the index value of the selected infrastructure variable

‘max’ is the maximum value among the series of values for a variable

‘min’ is the minimum value among the series of values for a variable

After normalization, the results of different indicators are summed to arrive at the Infrastructure Index. From the summed up index value, the poor status of camp infrastructure is defined when the index data value is close to ‘0’ and good infrastructure is defined when it is close to ‘1’. Thus, all five camps are rated using this technique based on different infrastructure available in and around the camp. The results of this index is presented in chapter 3.

Table 2.2 Variables used to construct Infrastructure Index and their units	
Variable used	Unit
Distance from the nearest government hospital	Kilometre
Private hospitals/clinics in the vicinity of camp	Binary <i>(1-available and 0-not available)</i>
Number of bathrooms	Per person
Number of toilets	Per person
Availability of waste water outlet	Binary <i>(1-available and 0-not available)</i>
Cleanliness surrounding the camp	Five point scale <i>(1-very bad and 5-very good)</i>
Greeneries surrounding the camp	Five point scale <i>(1-very bad and 5-very good)</i>
Connectivity to metallic road	Binary <i>(1-available and 0-not available)</i>
Frequency of public transport to camp	Minutes
Vicinity of nearest bus stand	Kilometre

Social network index (SNI)

In the present study, an attempt is made to understand social network of respondents with their social connections. Hence, connections at three levels have been taken into account while constructing the index – connections within refugee circle, connections in the place of origin and connections with the place of origin. Along with these three variables, other variables used include the mutual contact with one another and connection established with natives after coming to India.

Table 4.4 in chapter 4 presents the indicators selected for developing SNI. In this section, a composite indicator called SNI is developed to further measure and establish relationship with other factors. Considering that these variables have their own importance in explaining the social network among the refugees, all the ten variables are given equal weight. The index value for social network would range from 0 to 10 where '0' would mean absence of any social network while '10' would mean strong social network. Detailed analysis of SNI is presented in chapter 4.

Table 2.3 Variables used to construct Social Network Index and their weights	
Variables used	Weights
Do any of your friends/ relatives/ family members from Sri Lanka live in Tamil Nadu?	Yes = 1; No = 0
Do you contact them?	Yes = 1; No = 0
Do they contact you?	Yes = 1; No = 0
Do you have any friends/ relatives/ family members in Sri Lanka?	Yes = 1; No = 0
Do you contact them?	Yes = 1; No = 0
Do they contact you?	Yes = 1; No = 0
Do you have any friends/ relatives/ family members in Tamil Nadu who are Indian citizen?	Yes = 1; No = 0
You know them from Sri Lanka or know them after coming to Tamil Nadu?	Yes = 1; No = 0
Do you used to contact them?	Yes = 1; No = 0
Do they contact you?	Yes = 1; No = 0

Refugee adjustment index

Using the same method which was adopted in NFHS (1998-99), an index called “Refugee Adjustment Index” (RAI) has been developed. This is a composite measure that would differentiate the refugees by their level of adjustment on social, cultural and economic variables. Table 2.4 shows the variables used in the index and their weights. The index value for these variables ranges from the minimum of ‘0’ to a maximum of respondent with highest monthly income. The actual data shows the range of 1.4 to 22, with a mean value of 7.7. For the convenience of binary logistic regression analysis, the RAI has been made into a dichotomous variable considering mean value as the central cut-off point.

Table 2.4 Variables used to construct Refugee Adjustment Index and their weights	
Variables used	Weights
Attend any festivals in the locality	Yes = 1; No = 0
Attend any marriage in the locality	Yes = 1; No = 0
Assent used while speaking Tamil	Purely Tamil assent = 2 Mixed assent = 1 Purely Sinhala assent = 0
Household monthly income (continuous variable)	Income in rupee / 1000

2.6 Statistical techniques applied

In the present study univariate, bivariate, and multivariate techniques have been applied. Univariate analysis shows only the percentage and frequency of distribution. On the other hand, bivariate analysis provides only the association between two variables under investigation and chi square test provides the statistical significance of association between the two variables. Along with the univariate analysis, mean and standard deviation have been calculated wherever required. Multivariate techniques like binary logistic regression have been employed to find association between dependent variables and their explanatory variables mentioned in research hypothesis.

Chi square

Most commonly used test of significance for finding out the association between two nominal variables is chi square (χ^2) test. If two variables have a non-zero relation, then to test whether the relation is statistically significant, chi square test is applied. The chi square equation is written as:

$$\chi^2 = \frac{\sum(O - E)^2}{E}$$

where

O = observed frequency

E = Expected frequency

This test compares the observed and expected (assuming the relation is zero) frequencies in each category. The larger the summed difference between observed and expected values as a ratio of expected values, the larger is the value of chi square and the greater is the likelihood that the relationship is statistically different.

Fisher's exact test

While chi-square test is the most commonly used test of significance where it assumes that the expected frequency of each cell has a value of five or more. This test is not appropriate when the expected frequency of any particular cell is less than five. In such a case, Fisher's exact test is recommended as it does not have any such assumption and

can be used regardless of how small the expected frequency is. Hence, Fisher's exact test has been used in the analysis of significance wherever chi-square assumption fail. Fisher's exact test equation is written as:

$$N = \sum_i R_i = \sum_j C_j$$

where

R_i = Row sum

C_i = Column sum

This test calculates the difference between the observed and expected data, considering the given marginal and assumptions of the model of independence. It works in the same way as the chi-square test for independence. However, the chi-square gives only an estimate of the true probability value, an estimate which might not be very accurate if the marginal is very uneven or if the value is less than five in one of the cells. In such cases, Fisher's exact test is a better choice than the chi-square. This test procedure calculates an exact probability value for the relationship between two dichotomous variables, as found in a two by two cross-Table.

Logistic regression

When the response variable is dichotomous or binary (0, 1), application of binary logistic regression is appropriate. Any regression other than binary logistic regression would violate the basic underlying assumptions that the line tends to be very poor because the response variable can assume only two values that are 0 or 1. Hence, binary logistic regression technique is quite suitable in which independent variable may be continuous or categorical, or mixture of both (Retherford and Choe, 1993).

In the logistic regression model, outcome of an event reduces to occurring or not occurring. Hence, the concept of probability comes in. If the event occurs, the probability (p) is 1, otherwise 0. In the logit function the probability of occurring of an event is explained as:

$$p = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}$$

where, z is a linear combination of independent (predictor) variable in a multivariate logistic function and e is the base of natural logarithm. Thus,

$$Z = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

In other words, the probability of not occurring of an event is

$$1 - p = 1 - \frac{1}{1 + e^{-z}} = \frac{e^{-z}}{1 + e^{-z}}$$

Hence, odds of occurring of an event are $\frac{p}{1 - p} = e^z$

Clearly p/1-p equals the odds of occurring of an event. The logistic model for log odds of given predictor variable is

$$\text{logit } p = \log \frac{p}{1 - p} = \log \text{odds}$$

However, it is easier to interpret the results in terms of odds ratio, instead of log odds. Odds ratio is defined as the ratio of odds of an event when there is an increase of one unit in independent variable holding other independent variables constant. However, the choice of using either log odds or odds ratio is a matter of convenience. When the model is thought to be in additive form, then log odds are used but when it is of multiplicative form, odds ratio is used. In other words, when the response variable is log odds, the logit model is an additive model but if it is odds ratio, it is multiplicative model (Retherford and Choe, 1993).

Thus odds ratio model can be written as

$$\frac{P}{1 - P} = e^{a + b_1 X_1 + B_2 X_2 + \dots + B_k X_k}$$

Exponential power 'b' is the factor by which the odds changes when the independent variable increase by one unit. If 'b' is positive, this factor will be greater than 1, which means that odds are increased. If 'b' is negative, the factor will be less than 1, which shows that odds are decreased. When 'b' is 0, the factor equals 1 and the odds remain unchanged. In the current study, odds ratio has been used to interpret the results.