Understanding the Unobservable Population in Call Detail Records through Analysis of Mobile Phone User Calling Behavior
A Case Study of Greater Dhaka in Bangladesh

Ayumi Arai
Center for Spatial Information Science
University of Tokyo
Tokyo, Japan

Apichon Witayangkurn
School of Engineering and Technology
Asian Institute of Technology
Pathumthani, Thailand

Teerayut Horanont
Sirindhorn International Institute of Technology
Thammasat University
Pathumthani, Thailand

Xiaowei Shao
Ryosuke Shibasaki
Center for Spatial Information Science
University of Tokyo
Tokyo, Japan

Abstract—The understanding of mass population movements has greatly advanced with the rapid spread of ubiquitous devices. Anonymized call detail records (CDRs) for mobile phones have enabled us to not only trace individual trajectories but also approximate activity patterns, including significant locations such as homes and workplaces. The majority of studies analyzing CDRs attempt to utilize the mobility patterns of anonymized crowds to improve transportation and public health. This is quite reasonable because CDRs can capture the movements of people at given times and places, whereas general statistics usually account for a population based on their locations of residence. However, it has also been pointed out that there are discrepancies between the movements of people as observed through CDRs and those of an entire population in a given area. This is because CDRs only represent device users. In fact, we can never learn about the population that is unobservable through CDRs only by analyzing CDRs. Therefore, this study attempts to provide clues to help us understand the whereabouts of the unobservable population by analyzing two months of the CDRs for 58 volunteers with mobile device service from a major telecommunications company in combination with field survey data from Dhaka. We surveyed the personal and household attributes of mobile users in relation to their calling behavior. The analysis results show that per mobile user observed in CDRs, there is an average of roughly 2.4 to 2.8 unobservable people. Their age groups and gender composition are also provided. We find that male and female users exhibit opposite trends in call locations according to the presence of children within the household. In addition, based on field observations, we find that the location and time distributions of small children follow some specific routines. Our findings contribute to the understanding of the whereabouts of the unobservable population, the majority of whom are children and are considered to be vulnerable to disasters or infectious diseases but are difficult to locate through CDRs alone.

Keywords—mobile phone; CDRs; demographic attributes; sample bias

I. INTRODUCTION

With the increasing trend in the growth of global mobile subscriptions, the mobile penetration rate is expected to reach 96% by the end of 2014. The number of subscriptions in the developing world is estimated to represent more than three-quarters of the total, as the speed of growth these areas is almost twice that in the developed world [8]. In fact, in some developing countries, it is not uncommon for people to have access to mobile phones even when they do not have bank accounts, electricity, or access to clean water [19]. The mobile phone is arguably one of the most ubiquitous platforms and prominent infrastructures that will allow us to understand and address various issues in the developing world. This is because mobile infrastructure is already more developed than other basic infrastructures and can link large numbers of people with devices to information and technologies distant from them.

Studies of human travel patterns have greatly advanced owing to the rapid diffusion of ubiquitous devices, which generate large-scale spatiotemporal datasets such as GPS logs and call detail records (CDRs). With the capability to trace mass individual trajectories, the majority of these studies focus on quantitating properties of human mobility patterns. The features of repeatedly visited locations are estimated, taking social norms into consideration. Utilizing the bounded nature of human mobility patterns [7], significant locations such as homes and workplaces can be estimated using spatiotemporal data [9][15]. Apart from the features of trajectories and visited locations, some research attempts to link location histories—derived from GPS logs—to user attributes by assuming that people who have similar location histories share similar interests and preferences. These studies measure user similarity based on the sequence properties of people’s trajectories and the hierarchical properties of their location histories [11].

However, the movement patterns described in these studies show the trajectories of crowds because the data are anonymized. Therefore, their major fields of application are
concentrated in transportation, where quantitating the volume and speed of mobility can contribute to improving transportation planning and policy interventions. Further, it is increasingly being noted that the population captured by ubiquitous device data is not representative and therefore does not provide an accurate depiction of the general population, because such data can only capture device users. In fact, recent research examines heterogeneous mobile phone owners according to user attributes, such as gender and socioeconomic status, comparing mobile users to the general population [5][17]. This allows us to understand that such heterogeneity impacts the estimation results of human mobility analyses using CDRs [18]. This means that the interpretation and analysis of results may be misleading if there is no clear understanding of which parts of society the data represent. This constraint significantly limits the application of data when the population composition of the data matters. Although such limitations have been revealed to some extent, thus far, there are no established methods to help us address this issue in the analysis of biased data. Therefore, in this study, we attempt to answer following research questions:

- How dose the population composition of mobile users and non-users differ? How is the difference taken into account when analysis results from CDRs are used to address issues in the society?
- Are there any ways to help understand non-mobile users, who are not included in CDRs, through analysis results of CDRs?

First, we describe the population composition disparities between those who are captured by CDRs and those who are unobservable. To do so, we conducted a field survey to collect information on the demographic attributes of mobile phone users with service from a telecommunications company in Bangladesh and the members of their households. Based on an analysis of the data, we provide sets of descriptive statistics for the sex, role within the household, and age group of those captured by the CDRs and those who are unobservable. Second, we attempt to find clues in the calling behavior of mobile users that indicate the presence of the unobservable population. We suggest that key features of calling behavior enable us to estimate hidden properties in the CDRs even when the data are anonymized.

The contributions of our work are described below:

- Descriptive analyses of mobile users and the members of their households, who are not mobile phone users and are therefore unobservable in CDRs, are provided.
- Trends in calling behavior according to the presence of those unobservable in CDRs are provided. These traits can serve as clues to understand the unobservable population in CDRs based on the calling behavior of mobile users.

The remainder of this paper is structured as follows. Section 2 describes the data used in this study. Section 3 explains the demographic attributes of mobile users and those unobservable in CDRs as well as the typical activity patterns of the unobservable. Section 4 explains the types of calling behavior that can indicate the presence of those unobservable in CDRs. The final section includes our conclusions.

II. DATA

In this study, we use two datasets. One includes data collected through a questionnaire survey. We surveyed 810 households that included at least one mobile phone user with service from one of the leading telecommunications companies in Bangladesh (hereinafter called the operator). The other contains CDRs from 58 volunteers, who are also users of the same company. We targeted users of one company so that we could compare users in both datasets.

A. Social Background Considered for Data Collection

Bangladesh is no exception among many developing countries where access to mobile infrastructure is much easier than access to other basic infrastructures, such as electricity or clean water. This is probably because the construction of mobile infrastructure has been quick, allowing people to instantly access mobile networks once they obtain a mobile device. In addition, mobile device costs are not always very high. In a city like Dhaka, where huge income disparity can be found, people enjoy mobile services in various manners, depending on the trade-offs between their needs and financial constraints. As a result, the costs of obtaining a mobile phone and the amounts spent for daily use vary widely.

In Dhaka, with just 10 US dollars, you can buy a very simple feature phone on the street, which is common among lower-income populations. Most of them use the phone only for calls rather than for messaging because most of them do not know how to type text. Those who cannot read and therefore cannot fill out application forms to buy SIM cards from authorized shops can purchase them from street vendors for around two dollars. They normally load the cards with between five and 50 cents per week and sometimes use missed calls to communicate with family members for free. In contrast, smartphones are becoming popular among higher- or middle-income populations. Smartphones cost between 100 and 1,000 dollars and allow for not only calling and texting but also various internet services. According to our field observations, those who have stable jobs spend a minimum of one dollar per week. Although internet services are becoming popular, calling seems to be the primary method of mobile communication for the majority of people because internet services are relatively more expensive than calling. Thus, we consider the analysis of CDRs to still be relevant to understanding the calling behavior of diverse populations at this moment. In addition, it is obvious that there are differences in mobile phone usage according to income level that need to be taken into account for data collection.

B. Field Survey Data

We conducted a field survey—the Survey on Patterns of Activity for Comprehensive Explorations of Mobile Phone Users in Dhaka (SPACE)—to understand the personal attributes and calling behaviors of mobile users and the members of their households. We interviewed 3,288 respondents, including 922 mobile users. The survey period
was from November 2013 to January 2014, and the survey sites were selected areas of Greater Dhaka. SPACE collected one-day call records for mobile users in addition to their demographic attributes and activity, as well as information on the mobile ownership of all household members.

To capture the attributes of this diverse population, we employ two-stage stratified sampling based on land use and household income levels. We first classify all primary sampling units (PSUs) into three groups based on their dominant type of land use: residential, commercial, or industrial. Of these, 15 PSUs are selected, taking population numbers into account. From each PSU, 18 households are selected for three income groups: high, middle, and low income. This means that we have 270 households for each income group, and therefore 810 households in total. From the 18 households in the low-income group in each PSU, we sample the slum population as part of the low-income group if the ratio of the slum population to the total population in the PSU is more than 25%; otherwise, we do not sample the slum population for the low-income group. As a result, the low-income group includes 189 households from the slum population. For the following analysis, the 810 households are re-classified into four categories: high, middle, low, and slum income levels. Since there are different numbers of households sampled for each income level, the analysis is based on the ratio and distribution within each income level. We consider it crucial to include the slum population in the target population because this population is increasing. In fact, the slum population in Dhaka more than doubled, to an estimated 3.4 million, between 1995 and 2005, and is still increasing, while the total population increased from nine million to 13 million during this 10-year period [16]. Furthermore, this population group is not well captured by official statistics in general. One of the significant aspects of CDRs is that they can link anyone who uses a mobile phone regardless of status or living conditions. This enables us to understand certain populations that are not listed in official records but do exist de facto.

We need to note that the SPACE data do not represent the general population of each income group due to the sampling condition that households without mobile phone users of the operator are rejected from the sample. The ratios of rejection in the high-, middle-, and low-income groups are 23%, 31%, and 28%, respectively. It indicates that the operator is relatively popular among higher-income populations. However, the SPACE data represent the population of households with mobile users of the operator for each income group, which almost fulfills our study’s purpose. There were different rates of participation in our survey according to income level. The rates were 30%, 40%, and 48% of households we approached in high-, middle-, and low-income groups, respectively.

C. Call Detail Records (CDRs)

We use two months of CDRs for 58 mobile users of the operator from November to December 2014. We interviewed users of operators for four income levels who accept our survey during the same period of SPACE. They allowed us to analyze their calling patterns, which are determined based on the time and location distribution of calls, in combination with their attribute information. In addition to the call record information, we obtained information on household structures, types of activity, and significant locations from the users. Their household incomes skew to higher levels: three-quarters are from households in the higher- and middle-income populations. Forty-seven percent of them are male and 40% are engaged in income-earning activity. Roles within the household are distributed as follows: 47% are household heads; 43% are the spouses of heads; and the remainder includes the children, parents, and extended family members of heads.

III. DESCRIPTIVE STATISTICS FOR MOBILE USERS AND THE UNOBSERVABLE POPULATION IN CALL DETAIL RECORDS

In this section, we describe the demographic attributes of 922 mobile users and the members of their households—that is, of 2,366 non-mobile users—from the SPACE data. Sex, role within the household, and age group are examined separately for mobile users and non-users. This allows us to show which parts of the population can be captured by CDRs. Descriptive statistics are provided for four income levels.

A. Characteristics of Mobile User Households

Table I describes the basic characteristics of households from the SPACE data according to income level. As shown in row (A), the average household size is almost four for all income levels. This means that there are four persons per household, on average, across all income levels. The ratio of males is 51% for all income levels, according to row (B), which is close to the 54%, the ratio of males in the general population in Dhaka [2]. The higher the income level, the more highly educated the users are, as described in row (C). Row (D) shows that the higher the income level, the greater the ratio of mobile users when we disregard the telecommunications provider. According to [3], the subscription ratio in Bangladesh is 69%, but the figure obtained from our field survey is much lower. This is probably because in developing countries, it is common for users to have more than one SIM card so they can benefit from lower tariffs offered by several operators. As shown in row (F), we can confirm a certain ratio of multiple SIM card holders across all income levels. Interestingly, average number of mobile users per household does not differ according to income level, as described in row (E). Given that the average household size is between 4.0 and 4.1 for all income levels, we can assume that, on average, each person identified on CDRs represents 2.4, 2.6, 2.6, and 2.8 persons from the high-, middle-, low-, and slum income levels, respectively.

<table>
<thead>
<tr>
<th>TABLE I. BASIC TRENDS IN HOUSEHOLD ATTRIBUTES BY INCOME LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income level</td>
</tr>
<tr>
<td>--------------</td>
</tr>
<tr>
<td><strong>(A)</strong> Average household size</td>
</tr>
<tr>
<td><strong>(B)</strong> Male ratio</td>
</tr>
<tr>
<td><strong>(C)</strong> Average years of education of users</td>
</tr>
<tr>
<td><strong>(D)</strong> Average number of users per household (regardless of provider)</td>
</tr>
<tr>
<td><strong>(E)</strong> Average number of users per household (who specified the operator as their provider)</td>
</tr>
<tr>
<td><strong>(F)</strong> Ratio of multi-SIM holders</td>
</tr>
</tbody>
</table>
B. Attributes of Mobile Users and the Unobservable in CDRs

In this subsection, we examine user attributes to understand who is represented in the operator’s CDRs. Descriptive statistics for the sex, role within the household, and age group are provided by income level. Because the number of households and mobile users differ by income level, we employ proportions to compare the distributions among the four income levels. First, we examine the composition of sex among mobile phone users. Table II provides descriptive statistics for the sex of the users. As shown in row (A), males are the predominant users overall. Comparing the user ratio among males in row (B) and that among females in row (C), the ratio is higher overall for males. Table III provides the number of males and females who are assumed to exist but not included in CDRs, given the presence of one male or female in the CDRs, by income level. As discussed in Section A, we assume that there are roughly 2.4 to 2.8 unobservable persons per male or female in CDRs. For instance, when we find 10 male users in slum areas, they are assumed to represent 27 persons, 12 males and 15 females.

Table II: Descriptive Statistics: Mobile Users’ Sex by Income Level

<table>
<thead>
<tr>
<th>Income level</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>68%</td>
<td>32%</td>
</tr>
<tr>
<td>Middle</td>
<td>63%</td>
<td>37%</td>
</tr>
<tr>
<td>Low</td>
<td>52%</td>
<td>48%</td>
</tr>
<tr>
<td>Slum</td>
<td>62%</td>
<td>38%</td>
</tr>
</tbody>
</table>

Table III: Sex Composition of the Unobservable per Mobile Phone User by Income Level

<table>
<thead>
<tr>
<th>Per user in CDRs</th>
<th>Male (M)</th>
<th>Female (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Middle</td>
<td>1.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Low</td>
<td>1.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Slum</td>
<td>1.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Second, we examine the roles of mobile users within their households. The roles are classified into four categories: household head, spouse, child of the head, and others. People classified as others include the parents of the household head, relatives, servants, or workers living within the household. As can be seen, across all income levels, the majority of unobservable males are students and children. This indicates that most unobservable males are still dependent on their family members. Further, almost half of the unobservable females do household tasks, meaning that they are most likely housewives. This trend is consistent across all income levels.

Then, we examine the composition of the unobservable by role within the household against the four sets of 100 users. The number of unobservable differs by income level because the ratio of unobservable people to mobile users differs by income level. For instance, if the total populations surveyed for an income level consist of 250 mobile users and 750 non-users, the number of unobservable individuals per 100 users is 300. If another income level consists of 200 mobile users and 400 non-users, the number of unobservable people per 100 users is 200. Fig. 2 shows the composition of the unobservable population for each income level per 100 users. As mentioned, the total number of unobservable people varies according to income level because the ratio of non-mobile users to the total population differs according to income level. The greater the number of unobservable in Fig. 2, the lower the ratio of mobile users for that income level. As can be seen, the majority of the unobservable are the children of household heads, followed by their spouses. This trend is consistent across all income levels.

Fig. 3 describes the main activity of the unobservable for each income level. “Child” indicates a child below school enrollment age, and “HH tasks” indicates a person who does household tasks. As can be seen, across all income levels, the majority of unobservable males are students and children. This means that most unobservable males are still dependent on their family members. Further, almost half of the unobservable females do household tasks, meaning that they are most likely housewives. The remainder of unobservable females consists of students and children. The results demonstrate that CDRs are not highly capable of capturing the whereabouts of children, who are considered vulnerable.
Third, we compare the age group distributions between mobile users and those who do not use mobile phones within the users’ households to extract the population groups that are not captured by CDRs. The age group distribution among users is illustrated in Fig. 4 and that of the unobservable is shown in Fig. 5. As can be seen in Fig. 4, the dominant age group among users is that between 30 and 49. In almost all age groups, there are more male users than female users. The overall trends in the distribution are similar across all income levels.

Fig. 5 describes the age group distribution among the unobservable per 100 mobile users and by income level. The number of unobservable persons differs according to the income level, since the ratio of users varies. Overall trends are similar across all income levels, while the ratio of younger age groups is higher in lower income levels. Based on the descriptive statistics for the unobservable, we find that vulnerable people, who are often either young or dependent on their family members, such as the elderly, tend not to be included in CDRs. The unobservable, explained as children, tend to mostly comprise a very young population considering the form of the population pyramid in Bangladesh. It is very wide at the base with a median age below 25, indicating a half of the population are age below 25 and skew to younger ages.

Thus, we need to note that the CDRs do not capture the whereabouts of the vulnerable, particularly in the case of children. Based on the age distribution in Fig.4, it is fair to say that CDRs seldom capture people below the age of 20. These findings describe exactly what we observed in our fieldwork. Typical households in Dhaka tend to comprise nuclear families, which are composed of a household head, a spouse, and some children; it is quite common for household heads to own a mobile phone. Spouses, who are predominantly female because of the very high proportion of male-headed households in Dhaka, sometimes own mobile phones. Most of females do household tasks at home and some females work outside the home. This partially reflects social norms in Muslim societies, where males are supposed to be engaged in income-earning activities outside the home while females stay at home. The remaining household members seem to have limited opportunities to own mobile phones. They are often the children of the head and the ratio of extended family members, e.g., parents, brothers, sisters, or relatives of the head or spouse, seems to be higher among lower-income levels.

C. Whereabouts of the Unobservable

We briefly discuss trends in the whereabouts of children, who predominantly compose the unobservable population in CDRs based on our field observations. We found that children in Dhaka are primarily engaged in educational activities regardless of their household income levels. Even though many households cannot afford to send their children to public school, many NGO-run schools and educational institutions still provide educational opportunities for children from such households. Except for colleges and universities, school hours are basically only in the mornings or afternoons. Most schools operate from Saturday to Thursday, and Friday is generally a weekend. Some children go to coaching after school hours and others stay at home or around their homes. This means that on weekdays, the majority of children spend half of their days at school and the rest of the day around their homes. It is common to choose schools and coaches that are close to home due to heavy traffic in Dhaka. In other words, we can assume that children generally have similar routines such that their locations and activity patterns are very specific. Therefore, we expect that understanding the whereabouts of children at given times and locations is possible if we can identify the locations of mobile users from households with children.

IV. CLUES TO FINDING THE UNOBSERVABLE BASED ON MOBILE USERS’ CALLING BEHAVIORS

This section explores clues to finding the unobservable in CDRs based on the calling behaviors of mobile users. The results in previous sections show that the majority of the
unobservable are children. Hence, we first examine past studies on the link between travel patterns and personal or household attributes, which are related to the presence of children. By doing so, we attempt to identify the features of the activities of people with children within their households. Then, we analyze the calling behavior data obtained from SPACE and two months of CDRs from volunteers to investigate whether the features of the activities extracted from the literature review are reflected in the calling behavior characteristics.

A. Impacts of the Presence of Children within the Household on Travel-activity Behavior

Conventional trip diary data collected through field surveys have long contributed to enhancing the understanding of human activity and travel patterns for research on urban planning and transportation. Part of human mobility is well explained as travel demand in association with activity patterns, because urban travel is considered to be driven by the demand of people who have the need or desire to participate in activities [12]. In travel demand modeling, travel activities are explicitly combined with the presence of children within households. The presence of children is an important factor affecting the time allocated to out-of-home and non-work activities [10]. Furthermore, the time allocated to non-work or out-of-home activities differs among males and females when there are children in the household because children are dependent on their primary caregivers, which are predominantly their mothers. This implies that in terms of locations, calling behavior characteristics may reflect such behavioral differences according to the presence of children [14]. For non-employed people, the presence of children significantly affects travel-activity behavior [13]. Given that CDRs enable us to identify significant locations such as home and work locations [9], they can provide a partial view of users’ time and location distributions. This implies that time allocated to out-of-home activities, as extracted from calling behavior, can be a clue to finding people who belong to households with children.

B. Trends in Call Locations among Males and Females

First, we compare trends in call locations between males and females by analyzing one-day call records of 922 mobile users, collected through SPACE. Call locations are classified as home, primary out-of-home location, and other. We underline that this classification only specifies the home as a particular location, and that the primary out-of-home location is defined based on the number of call records per location. This definition is based only on frequency because this information can be extracted by counting the number of calls in CDRs. This means that the method applied to the one-day call records from the SPACE data is also applicable to CDRs. We calculate the ratio of calls at home and at the primary out-of-home location for each user as (1) and (2), and take the average by income level. By doing so, we attempt to find differences in call location trends between males and females.

![Fig. 6. Ratios of calls from (A) home and (B) the primary out-of-home location](image)

**Fig. 6.** Ratios of calls from (A) home and (B) the primary out-of-home location

C. Trends in Call Locations According to the Presence of Children

Second, we examine whether there are any differences in call location trends according to the presence of children within a user’s household. We analyze two months of CDRs for 58 volunteers: 27 males and 31 females. Considering the variety of lifestyles among mobile users, we classify the seven days of the week into two groups: primary routine days and non-routine days. We define primary routine days as days in which a mobile user is engaged in her/his primary routine. The primary routine could include any activities on which users spend the majority of their time. For instance, the primary routine for students is to go to school, and that of salary workers is to work at the office. If working days are from Sunday to Thursday, these five days of the week are the primary routine days, making Friday and Saturday the non-primary routine days. In the following, we first examine trends in call locations for each type of day according to the presence of children. Then, we analyze trends in call locations on non-primary routine days versus those on primary routine days.

![Fig. 7. Ratio of calls by location type on primary routine days among mobile users with and without children within their households](image)

**Fig. 7.** Ratio of calls by location type on primary routine days among mobile users with and without children within their households
tend to make a similar number of calls from home and from the primary out-of-home location. Similarly, (B) shows the trends among females. Females tend to call from home regardless of the presence of children within their households. Although we can to some extent see different trends in call locations according to the presence of children, the difference is not sufficient to distinguish between those with and without children based only on call locations.

**NON-PRIMARY ROUTINE DAYS**

Fig. 8 compares the ratios of calls at home and calls at the primary out-of-home location on non-primary routine days between the two groups and by sex. As can be seen, males without children within their households tend to call more from the primary out-of-home location. Conversely, those with children tend to make similar ratios of calls at home and at the primary out-of-home location. Among males, the trends on non-primary routine days are the opposite of those on primary routine days. As for females, again, their trends are opposite from those of males but the difference in females is not as significant as that in males. This indicates that males without children within their households have more flexibility to spend time out of the home. However, this difference appears to be insufficient to identify males with children based on calling behavior alone.

**NON-PRIMARY ROUTINE DAYS VERSUS PRIMARY ROUTINE DAYS**

Finally, we analyze trends in call locations on non-primary routine days versus those on primary routine days according to the presence of children within users’ households. To do so, we calculate the ratios, expressed as (3) and (4), and compare traits between males and females.

\[
\begin{align*}
\text{RH} &= \frac{\text{Ratio of calls from home on non-primary routine day}}{\text{Ratio of calls from home on primary routine day}} \quad (3) \\
\text{RO} &= \frac{\text{Ratio of calls from primary out-of-home location on non-primary routine day}}{\text{Ratio of calls from primary out-of-home location on primary routine day}} \quad (4)
\end{align*}
\]

Fig. 9 compares \( \text{RH} \) and \( \text{RO} \) according to the presence of children within the household and by sex. As can be seen from (A), the values of \( \text{RH} \) and \( \text{RO} \) among males with children within their households are close to unity. On the other hand, among males without children, \( \text{RO} \) is much greater than \( \text{RH} \), which means that they tend to make more calls at the primary out-of-home location on non-primary routine days. This indicates that the trends in call locations for males with children are almost unvarying across all days of the week. In contrast to males, \( \text{RH} \) and \( \text{RO} \) values among females are close to unity for those without children within their households. That is, males and females exhibit opposite directions of association between the presence of children within the household and trends in call location. The results also suggest that weekly trends in call locations are the key to identifying mobile phone users who have children within their households through their calling behavior.

**D. Application of Our Findings**

In this section, we find that males and females exhibit opposite trends in call locations on non-primary routine days versus primary routine days. This indicates that we can identify mobile users with children in their households by analyzing trends in call locations when we identify gender from anonymized CDRs. In fact, some studies attempt to predict the demographic attributes and socioeconomic status of mobile users by analyzing sensor data from smartphones [6] as well as calling behavior [4]. The analysis of call records collected through a field survey also shows that calling behavior exhibits gender-specific characteristics [1]. We will explore the application of such calling behavior characteristics to estimate gender and the presence of children among mobile users through the analysis of anonymized CDRs in future studies.

**V. CONCLUSIONS**

CDRs are receiving increased attention due to their ability to capture the mobility patterns of large-scale populations. Research on a variety of topics such as transportation, urban planning, disaster management, and public health is flourishing, utilizing the movement of people to address societal issues. However, the application of such data may be misleading if the population groups captured by CDRs are not relevant to the issue at hand. In fact, discrepancies between the populations captured by CDRs and the general population are noted in several studies. In this study, we demonstrate gaps between those who are captured by CDRs and those who are unobservable. We find that CDRs do not fully capture the movements of entire population groups in Dhaka. In terms of population size, there are roughly 2.4 to 2.8 unobservable people per mobile user identified in CDRs. Considering that for three income levels out of four, males represent more than 60% of mobile users, we can say that the majority of the operator’s users are males. In addition, more than 70% of the users are married, and their ages are mostly within the range of late twenties to late fifties. Our findings show that CDRs do not capture specific population groups such as students or...
children below school enrollment age. This implies that the application of CDRs needs to take such biases into account.

We also provide clues to identifying households with children from the calling behavior of mobile users. We first examine trends in call locations between males and females, and then compare their calling behaviors on primary routine days and non-primary routine days according to the presence of children. The results show that male users with children in their households exhibit consistent trends with regard to calling locations regardless of the type of day. Interestingly, female users exhibit the opposite trend. Although we demonstrated trends in call locations according to sex and the presence of children within the household, we discussed little on the whereabouts of the children. We will further examine time and location distribution of the unobservable including children in future studies.

Finally, we would like to emphasize that our findings imply that there is considerable potential to utilize CDRs to address issues related to part of the vulnerable population in the developing world. In many capitals of developing countries, urbanization is a common issue that accompanies rapid economic growth. This is because rural people migrate to urban areas expecting to find more income-earning opportunities in cities. Most of these people, who are poorly educated and lack assets, have to reside in vulnerable areas where the risks of disasters and infectious disease are potentially high. To improve such conditions, it is crucial to understand how many and what kinds of people reside in which places. However, official statistics cannot provide such information because most of these people are not registered in the places in which they are living but rather in their hometowns. We believe that this study can shed light on such areas by providing clues to understanding the whereabouts of the vulnerable, including those unobservable from CDRs.

ACKNOWLEDGMENTS

We thank the telecommunications company and volunteers who provided the data for this research. Part of this work was supported by GRENE-ei, funded by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT).

REFERENCES


