

**MIGRATION IN RETROSPECT:
REMEMBRANCES OF THINGS PAST**

James P. Smith
RAND

Duncan Thomas
RAND and UCLA

This research was supported by NICHD Grant P01 HD28372. We are grateful to Iva MacLennan, Margaret Hwang, Becky Mazel, and Eileen Miech for assistance in the preparation of the paper. Discussions with Christine Peterson have been very helpful.

1. Introduction

In recent years, there has been an explosion in the quality and quantity of household survey data from developing countries. Analyses of these data have yielded many important insights about the lives of families and individuals in low income settings (Srinivasan and Behrman, 1995; Deaton, 1997). However, the vast majority of these household surveys are cross-sectional, limiting how much can be learnt about the dynamics that have brought each society to its current status quo. Longitudinal surveys in developing countries are probably still years away, and panel surveys launched today can, at best, only tell us about the future and not the past. Yet, it is the demographic and economic transitions of the past few decades that are crucial for understanding the process of development. The only realistic option then is to rely on recalling past events. Not surprisingly, an important emerging trend in household surveys has been an increased reliance on retrospective information, especially in developing countries.

The obvious limitation of these surveys involves legitimate questions about the quality of retrospective data. Questions about events that occurred decades ago are quite unlikely to yield as reliable a response as would a query about what took place last week or last year. There is an extensive literature on response errors in the survey literature, dating back at least to the 1930s; Sudman and Bradburn (1974) provide a classic summary. In the last few years, survey researchers have paid an increasing amount of attention to the way respondents answer questions and the study of autobiographical memory has proved to be an area of fertile collaboration between survey researchers and cognitive psychologists, (c.f., Rubin, 1996; Schwarz and Sudman, 1994; Sudman, Bradburn and Schwarz, 1996).

Given the growing importance of retrospective surveys, researchers need to not only be sensitive to the quality of the underlying data but are likely to benefit from investments in better understanding the extent and nature of biases inherent in these sources. This paper is the product of such an investment. Focussing on residential location and migration, we undertake a detailed evaluation of retrospective life histories contained in a pair of household surveys. The matched migration histories cover the same period of the respondent's life and were reported independently by the respondent in the First and Second Malaysian Family Life Surveys (MFLS1 and MFLS2), which were conducted 12 years apart.

Migration events are examined for three main reasons. First, at the individual level, migrations are often important lifetime events and so there is an *a priori* expectation that there will be some concordance between histories reported 12 years apart. The fact that not all moves are equally well-remembered will be exploited to glean some understanding of the nature of recall bias in these data. Second, at a macro level, migration plays a critical role in the process of economic development and understanding how mobility changes with development is a central theme running through the social science literature. Disentangling behavioral changes from recall error will be key for understanding the relationship between development and migration. Third, besides the paucity of longitudinal data from developing countries, collecting panel surveys of migrants is very expensive since migrants are the most costly respondents to follow (and most likely to attrit from any sample) precisely because they have moved from their original location. It may well be argued that, for analyses of residential mobility over the life course, retrospective surveys will often be the only game in town.

The rest of this paper is divided into 5 parts. The next section describes the structure of the Malaysian surveys. Section 3 discusses hypotheses that will be tested and Section 4 presents a descriptive summary of the reliability of the data in the surveys. Multivariate analyses of correlates of the quality of retrospective migration histories are presented in Section 5, and the final section summarizes the results.

2. Sources of data

The first MFLS was a precursor for many of the demographic and socio-economic surveys that have been conducted in developing countries over the last two decades, including the Demographic Health Surveys and the World Bank's Living Standards Surveys. In conjunction with the follow-up, MFLS2, it remains one of the most ambitious long-term retrospective surveys. MFLS1 collected information from over 1,200 ever-married women aged 18 to 49 along with their spouses. The survey was conducted in 52 districts covering all of Peninsular Malaysia over three rounds, four months apart, during 1976 and 1977 (Butz and Da Vanzo, 1978). In addition to socio-demographic characteristics of the household, each

male and female respondent was asked to provide an extensive life history covering, *inter alia*, schooling and training, marriage, occupation, labor supply, earnings and residential location.

MFLS2, conducted in 1988-89, sought to re-interview the same household members surveyed in MFLS1. All respondents in MFLS2 were asked to provide an account of their life history covering the same substantive ground as in MFLS1. Since our main goal is to evaluate the quality of retrospective autobiographies collected in a survey setting, attention is restricted to those respondents who completed life histories in both MFLS1 and MFLS2. Of the 1,262 primary respondent females in MFLS1, almost three quarters were relocated. Of them, complete migration histories were recorded in both waves of the MFLS for 851 women: they are the female respondents in our matched sample. About 80% of the women were still married at MFLS2 and their spouses were eligible for the male interview: 663 men completed migration histories in MFLS1 and MFLS2: they are the males in our matched sample.

While some of the life history modules in MFLS1 and MFLS2 differ, the migration histories were designed to be identical. Each respondent and spouse gave their location (district and state) at birth and at age 15. They were then asked to report all subsequent migration events involving a change in location that lasted at least three months in duration. In the survey, the definition of a migration differs for men and women. Female respondents were asked to recall every time they moved house, while men (who tend to move more often than women) only reported moves across a district line. For both sexes, a move from Malaysia to a foreign country or from a foreign country to Malaysia is counted as a migration event (but moves between foreign countries are not reported).

The two waves of the MFLS offer a unique opportunity to evaluate the quality of demographic retrospective data. For pre-1977 moves, each respondent is asked to describe precisely the same migration events in two independent interviews that took place twelve years apart. The degree of correspondence, or discrepancy, between the migration histories for the pre-1977 time span forms the basis of our evaluation of the extent and nature of recall error in retrospective migration data.

One difficulty in recalling migration histories stems from changes in district boundaries in Malaysia which have been made throughout its post World War II history. Most adjustments involved small changes in boundaries, although in some cases, as economic development proceeded and population

expanded, single districts were subdivided into two districts. In the most complex case, boundaries have been reconfigured among districts around Kuala Lumpur (the capital city) as a consequence of rapid growth, with a new district emerging that draws from parts of many districts.

Respondents and coders were instructed in MFLS1 to use 1970 district boundaries to define their place of previous residence, and to use 1980 boundaries in the MFLS2. While not unusual for any country over time, these revisions in district boundaries raise some issues. First, the revisions could cause confusion in the minds of either respondents or coders concerning the relation between current district contours and those that were in place when the respondent lived there some time ago. To some extent, we can capture the consequences of any confusion that may have resulted by including controls in our analyses for districts whose boundaries were altered.

Even if this problem were unimportant, district boundary changes between 1970 and 1980 could introduce another complication that must be dealt with to make comparisons between MFLS1 and MFLS2 meaningful. Given the geographic subdivisions that took place between the two surveys, some events that are inter-district migrations in MFLS2 would not have been inter-district moves in MFLS1. To preserve comparability, MFLS2 inter-district migrations that would not have been inter-district migrations using the MFLS1 district boundaries were redefined as intra-district moves.

3. Recall error and the quality of retrospective data

Autobiographical memory is an active area of research in cognitive psychology (Conway, 1990) and survey design (Sudman, Bradburn and Schwarz, 1996). Much has been learnt about how memory is organized and several mechanisms underlying an individual's recall of facts have been distinguished. This has led to general agreement that exploiting our understanding of the hierarchical organization of memory will improve the design of retrospective surveys. There is less agreement on how to fully exploit the implications of these insights from cognitive psychology, suggesting that a good deal will be learnt from actual field experience. Our goal here is to provide some evidence from that experience.

Events which took place a longer time ago are harder to remember and so-called "forgetting curves" have been drawn at least since Ebbinghaus's pioneering study in 1894. While there is little doubt that migrations many year prior to the interview are likely to be less well recalled by respondents, the existing literature suggests that the relationship between memory and time since the event is far from simple. "Forgetting curves" may shift with attributes of the respondent, the migration event, and the temporal coincidence of other life events.

For example, as the frequency of events increases, they may blur together so that remembering each single event becomes more difficult. For most people, migration is a relatively infrequent and so this is unlikely to afflict many respondents in MFLS. But, there are some respondents who do report many moves in our data and they may provide sufficient power to test this hypothesis.

Second, laboratory studies indicate that more meaningful material is forgotten more slowly, (Sudman, Bradburn and Schwarz, 1996) suggesting that salient migrations will be recalled with greater frequency. Two distinct aspects of salience are potentially important. The first aspect relates to characteristics of the move itself. For example, the first migratory event, those of longer duration or involving greater distance or even the move that brought the respondent to the current location may be more salient and consequently more readily recalled. The second element of salience involves other landmark events linked with the migration. The co-occurrence of these events has been shown to be associated with improved recall (Loftus and Marburger, 1983; Auriat, 1993). This may be because some coincident events -- such as a move at the same time as a marriage -- are more salient themselves although studies in psychology suggest that it may also be observed because recall of one event is triggered by cues associated with another simultaneously linked event. The MFLS data are particularly valuable in examining the role of marker life events because each respondent reports a life history covering not just migration but also marriage, child-birth and work.

In addition to whether events are remembered at all, the quality of retrospective data could be judged by the accuracy of recall. For example, a small number of studies in economics has found that retrospective reports of employment and unemployment spells are accurate but that the precise timing is not as well remembered (Duncan and Hill, 1985; Mathiowetz and Duncan, 1988). Indeed, the focus of

a substantial fraction of the survey design and cognitive psychology literature has been on the accuracy of dating and, specifically, the extent of telescoping (reporting events as having occurred more recently than they actually did).

The empirical evidence on telescoping is mixed (c.f., Bradburn, Huttenlocher and Hedges, 1994). For example, Becker and Mahmud (1984) sought to validate retrospective birth history data from Matlab, Bangladesh by matching reported birth dates with vital statistics gathered prospectively from the same household on a biweekly basis over 25 years. While event reporting is good, there is a general tendency for women to place the event too far back. Other studies report systematic forward dating (Gray, 1955; Loftus and Marburger, 1983).

The evidence is even more mixed on what type of events are more subject to telescoping. Some have argued that more vividly remembered ("accessible") events will be telescoped (Brown, Rips and Shevell, 1987). But, not all studies support this "accessibility" hypothesis. Thompson, Skowronski and Lee (1988) fail to find any relationship between clarity of recall and the extent of telescoping whereas Kemp (1988) reports that telescoping is greater for less well-remembered events. Auriat (1993) suggests that telescoping may be an exclusively short-term recall phenomenon. She reports that in a study of retrospective migration histories, there is no evidence of telescoping but dating errors are reduced when moves are temporally linked to landmark events. We will present new evidence on this issue below.

The potential value of landmark events in aiding respondent recall suggests that other individual characteristics may be associated with the quality of a person's retrospective responses. Apart from studies comparing men with women (c.f., Skowronski and Thompson, 1990; Auriat, 1993), there is little evidence about what attributes might matter, perhaps because few studies have had sufficient observed heterogeneity to provide convincing tests. The MFLS, a study covering all of Peninsular Malaysia, may be helpful in this respect. The relevance of understanding links between recall error and respondent characteristics is clear if we think about education, for example. One of the best documented facts in the migration literatures is that mobility rises with education. However, if better educated people give more complete migration histories, part of that correlation will have nothing to do with behavior but will simply reflect differences in respondent error by education. Inferences regarding the link between human capital

investments and development may be seriously misleading. New evidence on the importance of individual characteristics in assessing the quality of retrospective reports is presented below.

Finally, question wording, instrument complexity and survey conditions have all been shown to be important in eliciting responses (retrospective or otherwise) (c.f., Sudman and Bradburn, 1974; Sudman, Bradburn and Schwarz, 1996). In the Belgian migration study, Auriat (1993) argues that the small advantages of interviewing husbands and wives together are outweighed by the additional costs. In addition to examining this issue, we will assess whether the interviewer burden (measured by time to complete the instrument) affects recall quality and also use independent evaluations of the respondent's interest and reliability as measured by the interviewer. These latter indicators are of special interest because, in empirical applications, data on survey practice will be extremely valuable if it contains information that is associated with the quality of the recalls but is unrelated to the underlying behaviors that are being examined -- as in, for example, the relationship between mobility and education. These survey quality measures thus provide an opportunity to disentangle the role of behavior from recall error.

4. Correspondence of retrospective responses in MFLS1 and MFLS2

Before comparing migration histories, we evaluate the correspondence between location at birth and at age 15 reported in MFLS1 and reported again twelve years later in MFLS2. Place of birth, a common demographic question, would seem a reasonably salient life event. Although there is little hard evidence on test-retest reliability for such questions, our intuition is that most people in the U.S. would know their state of birth, but would have more problems with their county of birth (more analogous to a Malaysian district). This latter point is especially pertinent in Malaysia because district boundaries may have changed since the respondent's birth. A mismatch would also arise if the interviewer misreported the answer or if coding was incorrect (a problem perhaps exacerbated by the boundary re-definitions). In addition to comparing MFLS1 and MFLS2 responses, a special set of test-retest questions in MFLS1 is used to disentangle some reasons for non-correspondence.

4.1 Residence at birth and at age fifteen

Table 1 lists the extent of correspondence in location at birth and at age 15 for males and females separately. The first two columns refer to states (of which there are 14 in Malaysia and all foreign countries are treated as a fifteenth state). The remaining columns refer to district matches.

Even with the twelve year hiatus, state mismatches are relatively infrequent, occurring in fewer than 5 percent of cases for births and only slightly more frequently for location at age 15. Discrepancies are more common at the district level with about 85 percent of respondents reporting consistent districts at birth and at age 15 in the two waves of the MFLS. With the issue of changing district boundaries in mind, the fraction of consistent responses if geographically adjacent districts are included as matches is listed in the fourth and sixth columns of the table. Whether due to faulty memory or the problems associated with changing district boundaries, respondents locate themselves at birth and at age 15 in the same place, or somewhere close by, in over 90 percent of cases.

MFLS1 was conducted over three rounds and women reported location at birth and at 15 in the first round and again, four months later, in the second round. When contrasted with the correspondence over the twelve year interval, this short-term test-retest feature of the survey provides a measure of the effect of time delay and response consistency. Some of the reasons for discrepancies (such as interviewer and coder error) have little to do with memory lapses associated with time. As shown in the third row of Table 1, the short-term reliability of responses is higher with only 2.5 percent of respondents reporting inconsistent states and 4-5 percent reporting inconsistent districts. We conclude that about one-half of the inconsistent states and a third of the inconsistent districts in MFLS1 and MFLS2 are not associated with memory lapses that occurred in the twelve year hiatus between the surveys.

4.2 Correspondence in reported number of moves

We turn next to migration histories reported by the same respondent and covering the same time period from age 15 through to the MFLS1 interview. Table 2 summarizes the correspondence between the histories reported in the two interviews.

A comparison of the relative number of moves reported in the two survey waves is one index of the rate of forgetting. In MFLS1, 968 moves are reported by men. But 84 fewer moves are reported in

MFLS2 suggesting that, on net, some moves were forgotten during the twelve year break between the two surveys. In contrast with men, women report slightly more inter-district moves over this period in MFLS2, relative to MFLS1, indicating that the MFLS1 reports are not a complete catalogue of all moves.

The biggest differences in recall, however, are not between men and women but lie in the stark contrast between inter-district and intra-district moves (the latter being reported only by women). Intra-district moves are much more common, and fully 20 percent of those reported in the first interview were not recalled in the second interview. In fact, for every woman who lists more intra-district moves in MFLS2, there are almost two women who report more of these migrations in MFLS1. This tilt in favor of the first survey is not surprising as the additional 12 year distance from each event in MFLS2 apparently took a greater toll on recall of shorter distance, and apparently less memorable, intra-district moves.

4.3 Matching migration events

While the simple comparison of counts of reported moves is mildly encouraging, it does not tell us whether the respondents reported the same moves -- that is, to the same place or at the same time. We have, therefore, attempted to find exact matches for all migration events reported by the same person in the two interviews. For example, for each MFLS1 male inter-district move, a search was made across all MFLS2 inter-district migrations reported by the same respondent to identify one with which the MFLS1 move could be paired. In this pairing, a MFLS2 move could only be used once.

There were two reasons our search for a match for an MFLS1 move would not be successful. First, and most commonly, there were not enough MFLS2 migrations to do the job. To illustrate, if there were three inter-district migrations reported in MFLS1, but only one in MFLS2, two of the MFLS1 moves could not have a match. The second reason for a non-match was that the districts were not the same in MFLS1 and MFLS2. An identical procedure was employed to search for matches for all female inter-district and intra-district moves.

The match rate for male inter-district moves (in row 2.2) is 58 percent. It is higher among women: 69 percent of their original inter-district migrations were also listed in the follow-up survey and

a remarkably similar fraction of reported intra-district moves match. By this metric, women give higher quality retrospective migration histories, a fact that has emerged twice already.

In row 2.3, the procedure is reversed so that all MFLS2 moves that are reported as having occurred before the MFLS1 survey serve as the reference and we search for their matches in MFLS1. If more salient events are less likely to be forgotten with time, a move that took place prior to MFLS1 and is *reported* in MFLS2 will be more salient than a move *reported* in MFLS1. Hence, match rates should be higher conditioning on MFLS2 reports and the gap should be greatest for intra-district moves. Both predictions are supported by the data. For example, 67 percent of the intra-district moves reported in MFLS1 were also reported in MFLS2, but over 80% of those reported in MFLS2 were remembered at the MFLS1 interview.

For some analytical purposes, the principal value of recall geographic data centers around knowing where respondents were in the past. For example, we may want to understand how local labor market forces affect economic outcomes or how the quality of local infrastructure, such as health facilities or family planning programs, alter health or fertility. A useful summary index for this purpose is a calculation, based on the migration histories, of the percentage of time respondents would have been located in the same district in both surveys. Row 2.4 indicates that this would be 86 percent for male and 89 percent for female respondents. This more positive evaluation of recall migration data stems in part from the fact that the "forgotten" migrations are relatively short duration so that their impact on time spent in a particular location over the life course is less consequential.

Another measure of recall quality concerns the dating of moves reported in both surveys. Some summary statistics of date discrepancy are presented in the third panel of Table 2. To assess the extent of telescoping, we first divided the dates into three groups; those that are the same, those in which MFLS1 precedes MFLS2 and those in which MFLS2 precedes MFLS1 (rows 3.1-3.3). If moves are telescoped, then it should be more common for the MFLS1 date to precede the MFLS2 date. While around 10 percent of reported dates are the same, there is a tendency for dates reported in MFLS1 to precede those in MFLS2 (especially for female inter-district moves). This ordering suggests that migration events are, on average, telescoped. Our second measure concerns the absolute size of the date discrepancy. Because

there is tremendous heterogeneity in these discrepancies, as is apparent in the very fat tailed distributions in Figure 1, our measure is the quartic root of the (absolute value) of the difference in dates. (This measure is robust to outliers; square roots and trimmed means provide similar results). There apparently is a non-significant amount of telescoping in these data: the difference between the two date discrepancies (in row 3.6) is significant although not very large. For example, for men, the mean difference for men is about 6 months ($=1.87^4 - 1.54^4$) compared to slightly less than 8 months for female inter-district migrations and 5 months for female intra-district moves.

5. Multivariate analysis of quality of migration retrospectives

To provide some insights into the relation between recall error and characteristics of moves and of respondents, we turn to a multivariate analysis of two dimensions of the quality of retrospective data. Table 3 lists probit estimates of the probability a move is reported in MFLS2 given it was reported in MFLS1. A parallel strategy examining whether moves reported in MFLS2 were also reported in MFLS1 was explored, but since the results provide no additional insights, they are not discussed here. Attention shifts in Table 4 to the second measure of quality, the discrepancy in dates for matched migration events, using the quartic root of the absolute value of the difference in migration dates (measured in months) to reduce the influence of extreme outliers. Since only moves reported in both surveys are included, the regression models in Table 4 are based on fewer observations than the match rate regressions.

Results for male inter-district moves are reported in the first pair of columns in each table; the remaining three pairs of columns report estimates for all moves by women, their inter-district moves and their intra-district moves. Co-variates in the models fall into three generic categories: personal attributes, survey characteristics and attributes of the moves themselves. Summary statistics of the covariates are given in Appendix Table 1.

5.1 Personal Attributes

Many studies have shown that recall decays with distance from the event, often in a non-linear fashion. To examine this "forgetting" function, the probability a move is reported in both MFLS

interviews is allowed to depend on age and (a linear spline in) years since migration (both measured at the time of MFLS1). Since date discrepancies are determined by a given month and year, any time-dimensioned co-variates would have a purely mechanical relation with that outcome. Consequently, covariates that are linear in time are eliminated from the regressions in Table 4.

Controlling age, years since migration isolates "forgetting" from aging. The results in Table 3 indicate a tendency for moves that took place during the 5 years prior to MFLS1 to not be recorded 12 years later in MFLS2. The forgetting function is particularly steep for female (inter-district) moves. However, for moves prior to 1970, the function becomes essentially flat for females: moves reported in MFLS1 that took place more than 5 years prior to the interview were no less likely to be remembered in the 1988 interview, no matter how far back the move occurred. This flattening of the forgetting function is also apparent for men, although their memories do continue to dim as they reach further back to moves that took place prior to 1960. Thus, several moves reported in 1976 that took place only a few years prior to the interview were apparently not remarkable and by 1988, they had been forgotten.

Holding years since migration constant, an increase in the respondent's age at the time of MFLS1 increases the age when the migration occurred. Otherwise identical events that take place during adolescence or early adulthood may have less salience than migrations that occur later in adulthood. The results for male inter-district moves support this hypothesis. While a woman's recall of migration events is not associated with when in the life course the move took place, her recall of dates is related to her age (in Table 4). Because this is our only control for time in the latter regressions, it is not possible to distinguish between an effect of aging and an effect of longer recall.

MFLS interviews are demanding, asking respondents to organize and remember complex sequences of their life histories. Education is an important correlate of the quality of recall in the migration histories suggesting that more educated respondents may have less difficulty with this task. For example, among men with no formal schooling, less than half of all MFLS1 male inter-district moves are also reported in the follow-up survey. In comparison, males with at least 6th Form schooling report moves in both interviews 90 percent of the time. Even after controlling for all observables in the regression, the gap in match rates between these two education groups is 30 percent.

To what extent does this reported increase in mobility with schooling reflect behavioral migration differences instead of differential reporting propensity across education groups? This generic issue, which applies to many research questions, highlights the value of having side-by-side comparisons of the reporting and behavioral migration functions. In our sample, post-secondary school male graduates report about 5 inter-district moves compared to about one for men with no schooling. Differential reporting propensities are small in comparison and so they certainly cannot fully explain the observed rise in the number of migrations across education levels. We note, also, that additional years of schooling are not only associated with an increased probability of reporting inter-district migration events, but with reduced inconsistencies in the dating of these migrations.

Finally, those personal attributes that are not the central interest in our research are briefly discussed. Conditional on own schooling, parents' education has consistent effects only in improving the quality of women's reported intra-district migration data. Although there are no consistent ethnic differences for men, Indian and Chinese women apparently supply more accurate migration data, with a higher likelihood of reporting the same migration event in both surveys (particularly intra-district moves) alongside a smaller time discrepancy in reported start dates.

5.2 Survey characteristics

The second set of co-variates capture several dimensions of survey operations that may be related to the quality of responses. The presence of other people during the interview is a controversial survey practice although it is often unavoidable in households in low income settings, like Malaysia. In MFLS2, we know whether other people were present and whether anyone helped the respondent during the interview. Two categorical variables are created from this information—whether the spouse was present, and whether any help was received during the interview. While we find no impact on any dimension of survey quality when a person other than the spouse assisted the respondent, the presence of a spouse does affect the quality of responses, albeit in an asymmetric fashion. A wife's attendance at the MFLS2 interview reduced the odds that a MFLS1 move was reported by a man in the follow-up survey -- a result that contrasts with Auriat's (1993) finding for Belgian couples. When this result is explored more deeply, we find that the wife's presence has no effect on match rates of post-marriage moves but the effect is

concentrated entirely in pre-marriage moves. A man is, apparently, liable to forget many of the moves he made when single if his wife is listening. On a more positive note, the husband's presence lowered the time discrepancy in move dates reported by women.

In MFLS2, an interviewer assessment was made about the quality of each life history module. Interviewers were asked their subjective evaluation of the reliability of the respondents' answers. Those interviews judged as more reliable either had higher quality migration modules or the quality was no different than other cases. In particular, those cases for which the interviewer judged the answers to be very reliable or of average reliability had consistently higher match rates for all female and female inter-district migrations and lower time discrepancy for male migrations. Survey characteristics such as evaluations of respondent reliability are quite valuable because they are part of an assessment of the quality of retrospective migration data, but do not belong in a model explaining the behaviors underlying migration decisions. Therefore, these variables are good candidates for statistical instruments to distinguish the behavioral migration function from the function describing the probability of reporting migration events.

Another characterization of the actual interview is the total length of the male or female life history modules (of which migration is only a small part). While long interviews are more complete ones, fatigue may become a factor in a particularly long interview tempting respondents to truncate. The average MFLS1 interview was about 1 hour for men and nearly 2 hours long for women; MFLS2 interviews were typically shorter. There is little evidence in these data that response quality declines as the time burden rises. Time discrepancies in intra-district moves rise with interview time, but match rates for inter-district moves improve as length of interview increases.

Given the potential problems created by changing district boundaries over time, it is reassuring that, once districts were made comparable in the two surveys, an indicator variable that identifies migration involving a post-1970 district is not associated with differences in the quality of male reporting. In fact, these adjusted districts were characterized by higher quality of reports for female inter-district moves.

5.3 Characteristics of moves and migration salience

The final set of covariates captures alternative dimensions of the move itself. Hands down, the most salient aspect of a move is its duration. Among both men and women, move duration is strongly related to the likelihood of reporting migrations, with short-term moves the most likely to be forgotten. For example, roughly a third of male moves in MFLS1 lasting less than six months are also reported in MFLS2. This match rate increases steadily with move duration until it more than doubles among moves that lasted at least 20 years (76 percent). While the details differ, this pattern of rising match rates with move duration is quite strong for all female and for female inter-district moves.

These results simultaneously illustrate the potential and limitation of recall migration histories. Long term retrospective migration histories do not seem to be a useful way of collecting information on the short-duration, often circular migration patterns that characterize many developing countries. Respondents do far better, however, with the more salient longer duration moves, of which up to 80 percent are remembered.

Our second set of salience measures concerns the distance of the move, which is captured by geographic distance between district centers as well as by indicators for whether moves involved crossing a district or state line. Match rates are significantly higher for male and female inter-state moves, and the twelve year gap apparently took a greater toll on female intra-district moves. After controlling for other covariates, female reporting rates are 19 percentage points higher if the move crossed a district boundary. Geographic (or straight-line) distance does not appear to be associated with the memorability of the move and this is not just because long distance moves are captured by the inter-state controls: distance is not significant even without those controls. Table 4 indicates that none of the measures of distance included in the regression has any detectible impact on the consistency of reported migration dates. Given the geographic impediments like mountains ranges, straight-line distance may not be a good proxy for actual distance or moves. The deterrent effect of distance may also change with development as roads and other infrastructure development that ease the burden of moving across a given distance.

Sequencing of Moves

How a particular move fits into the lifetime migration sequence may add or subtract from its meaning. The move that first brought the respondent into the district lived in at the time of the MFLS1 is particularly memorable (although dating is no better than any other move). In contrast, there is little evidence that the first move (after age 15) adds any special meaning in respondents' ability to recall.

Our models also distinguish moves made before and after marriage and, within post-marriage moves, those made jointly with a spouse are differentiated from those taken alone. Whether joint moves took place is only known for inter-district moves and is included in only those functions. Joint moves should be more memorable in part due to the greater disruption caused when the entire family pulls up roots. This intuition is correct: joint moves made while married are much more easily remembered. The accuracy of migration dating, however, is not affected by the jointness of a move.

Many migration sequences (especially in developing countries) are circular, with an individual moving from place A to place B, often for temporary employment or reasons associated with family care, and then subsequently returning again to place A. Given their temporary nature and the lack of commitment to the intermediate destination, migrations that are within a circular pattern may be less salient. While the statistical significance is often weak, there is some evidence that circular moves are more likely to be forgotten and are particularly difficult to date.

The cognitive psychology literature suggests that when there are several similar events in a history, they merge together and are not easily recalled as discrete events. In the extreme, a sole migration event is more salient than a move that is just one among many. This is not much of a problem for inter-district moves which are relatively few in number. However, among the more numerous female intra-district migrations, match rates are much lower among women who report more moves (although they do a better job on timing, conditional on reporting the move).

Coincident Events

Events may be more memorable when they are temporally coincident with another salient life event which may serve as a marker and help trigger memories associated with the migration. Since the emotional and cognitive power of coincident events may differ among people, the impact of three markers are explored: marriage of respondents, birth of a child to the wife, and the start of a significant new job by the husband. We select these three because marriage is an important event in all cultures and, in Malaysia, it is traditional for women to return home to their mothers around the time of child birth. In contrast, labor force participation rates are low among women, but labor market concerns are a primary motivation for the majority of male moves.

Family-oriented events -- marriage and birth -- have no detectible impact on the ability of men to recall either the migration itself or the date when it occurred. In contrast, the start of a new job significantly increases the probability a migration is reported in his interviews and reduces the time discrepancy in the dates that he reports.

Matters could hardly be more different for women where family-centered events serve as a strong memory aid. Marriage significantly increases the odds a woman reports a concurrent migration and reduces the difference in reported dates of this move. Although the results for the birth of a child are not as powerful or consistent, the weight of evidence indicates that it also improves the quality of reporting, especially for female intra-district moves. In contrast to men, career changes of their husbands do not help their recall of migrations. The memory enhancing value of other salient life events for recreating migration life histories suggests that they should be incorporated explicitly into the survey design.

Telescoping

Telescoping refers to remembering events as having taken place more recently than they did, so that for telescoped events the MFLS1 date on average will precede the MFLS2 date. In the absence of telescoping, knowing whether the MFLS1 reported date preceded or followed the MFLS2 reported date would provide no information about the size of the absolute difference in the time discrepancy. In the presence of telescoping, however, on average, reported MFLS1 dates will tend to precede those in MFLS2. With telescoping, the absolute value of the time discrepancy will be smaller when the MFLS2 date is

reported first. To test this implication, the regressions in Table 4 include a control for whether the MFLS1 reported date was the earlier of the two. In all cases, it is positive and significant implying that, after controlling personal, survey and move attributes, telescoping is present and is slightly larger for inter-district relative to intra-district moves.

Some have argued that telescoping is greater among more salient (or accessible) events (dubbed the "accessibility hypothesis"). If this is true then, if the date reported in MFLS1 precedes the date reported in MFLS2 for a migration that took place coincident with a salient life time marker event, the discrepancy in reported dates should be larger. To test this hypothesis, the models in Table 4 were re-estimated by adding interactions between the control for MFLS1 being reported earlier and the three life time marker events: a job change among men, marriage and birth of a child. The results are reported in Table 5 (with all other covariates suppressed from the table).

The evidence favors the accessibility hypothesis: telescoping is exacerbated by co-incident salient events. Consistent with our evidence on gender differences in what constitutes a salient event, men telescope their migrations that are coincident with the start of a new job, but show no such tendency for moves coincident with important family dates. Similarly, women do not telescope their moves coincident with their husbands' job changes, but do telescope moves taking place around the date of their marriage. Among women, if a move occurred around the marriage date and the MFLS1 date preceded the MFLS2 date, not only is the time discrepancy larger but the direct effect of reporting the MFLS1 date first is no longer significant. Apparently, marriage is doing a good job of identifying salient events that underlie telescoping by women.

6. Conclusion

While opinions are often strongly held, there has been little scientific evaluation of the quality of retrospective data based on fieldwork in a low income country setting. By asking the same questions twelve years apart, the two Malaysian Family Life Surveys provide an opportunity to quantitatively evaluate the quality of long term recall data in a rapidly changing developing country.

Until recently, such long-term retrospective data had a poor reputation in the empirical social sciences with doubts cast on the ability to recall events just a few months ago, let alone a decade or even further in the past. That verdict seems premature not because retrospective data are perfect but because these questions can elicit useful information if we are also able to isolate those events people remember from those they tend to forget. Using migration life histories, our research demonstrates that respondents tend to remember salient moves, those linked with other important life events such as the start of a marriage, the birth of a child, change in a job and moves that lasted for a long time. In contrast, migrations that dim in memory as time passes are typically short duration or local moves, often made while the respondent was young. This kind of mobility is apparently not successfully retrieved by simple retrospective interviewing. But for many purposes, these may be less important moves as they span relatively small segments of people's lives. Dating of events is not perfect, and there is a tendency for respondents to recall events as having taken place closer to the interview date than they actually did, particularly in the case of salient events.

We find significant differences in the quality of recall data across different socio-demographic groups -- particularly among different education groups -- and point out the need to interpret retrospectives with this fact in mind. The study has also demonstrated that there is scope for disentangling the influence of differences in reporting from differences in behaviors by exploiting characteristics of the survey operation and interviewer evaluations of the responsiveness and reliability of the respondents since these are factors that do not affect behaviors but are correlated with the quality of the retrospective data.

Long-term recall data are not a panacea, and are certainly not perfect. But, until recall data have been scientifically evaluated in a wider array of settings, the current presumption against the use of long-term recall questions in field surveys ignores a potentially rich source of data. Rather, it would seem more profitable to better understand the nature of the information contained in retrospective data and identify field practices that enhance their usefulness in empirical research. These practices include using other life events as markers and reporting interviewer evaluations of respondent engagement in the interview.

References

- Auriat, N. 1993. "My wife knows best: A comparison of event dating accuracy between the wife, the husband, the couple and the Belgium Population Register", *Public Opinion Quarterly*, 57:165-90.
- Becker, S. and S. Mahmud. 1984. "A validation study of backward and forward pregnancy histories in Matlab, Bangladesh", WFS Scientific Report 22, Institute for International Statistics, Voorburg.
- Bradburn, N. M., J. Huttenlocher and L. V. Hedges. 1994. "Telescoping and temporal memory" in N. Schwarz and S. Sudman (eds.) *Autobiographical memory and the validity of retrospective reports*, New York: Springer-Verlag.
- Brown, N. R., L. J. Rips and S. K. Shevell. 1986. "The subjective dates of natural events in very long-term memory", *Cognitive Psychology*, 17:139-77.
- Butz, William P. and Julie Da Vanzo, 1978. *The Malaysia Family Life Survey: Summary Report*. R-2351-AID. The RAND Corporation.
- Conway, M. A. 1990. *Autobiographical memory: An introduction*, Milton Keynes: Open University Press.
- Deaton, A. 1997. *Analysis of Survey Data*, Baltimore: Johns Hopkins University Press.
- Duncan, G. and D. Hill. 1985. "An investigation of the extent and consequences of measurement error in labor economic survey data", *Journal of Labor Economics*, 3.4:508-32.
- Ebbinghaus, H. 1894. *Memory: A contribution to experimental psychology*, New York: Dover.
- Gray, P. G. 1955. "The memory factor in social surveys", *Journal of the American Statistical Association*, 50:344-63.
- Kemp, S. 1988. "Dating recent and historical events", *Applied Cognitive Psychology*, 2.3:181-8.
- Loftus, E. and W. Marburger. 1983. "Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports of landmark events", *Memory and Cognition*, 11.3: 114-20.
- Mathiowetz, N. and G. Duncan. 1988. "Out of work, out of mind: Response errors in retrospective reports of unemployment", *Journal of Business and Economic Statistics*, 6.2:221-9.
- Rubin, D. C. 1996. *Remembering Our Past: Studies in Autobiographical Memory*, New York: Cambridge University Press.
- Schwarz, N. and S. Sudman. 1994. *Autobiographical Memory and the Validity of Retrospective Reports*, New York: Springer-Verlag.
- Schwarz, N. and S. Sudman. 1995. *Answering Questions*, San Francisco: Jossey-Bass.
- Skowronski, J. J. and C. P. Thompson. 1990. "Reconstructing the dates of personal events: Gender differences in accuracy", *Applied Cognitive Psychology*, 4.5:371-8.

- Srinivasan, T. N. and J. R. Behrman. 1995. *Handbook of Development Economics*, Volume 3, Amsterdam: North Holland.
- Sudman, S., and N. M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*, Chicago, IL: Aldine Publishing Company.
- Sudman, S., N. M. Bradburn and N. Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*, San Francisco: Jossey-Bass Press.
- Thompson, C. P., J. J. Skowronski and D. J. Lee. 1988. "Telescoping in dating naturally occurring events", *Memory and Cognition*, 16:461-8.

Figure 1: Empirical distribution of time discrepancies

MFLS1 date - MFLS2 date [measured in years]

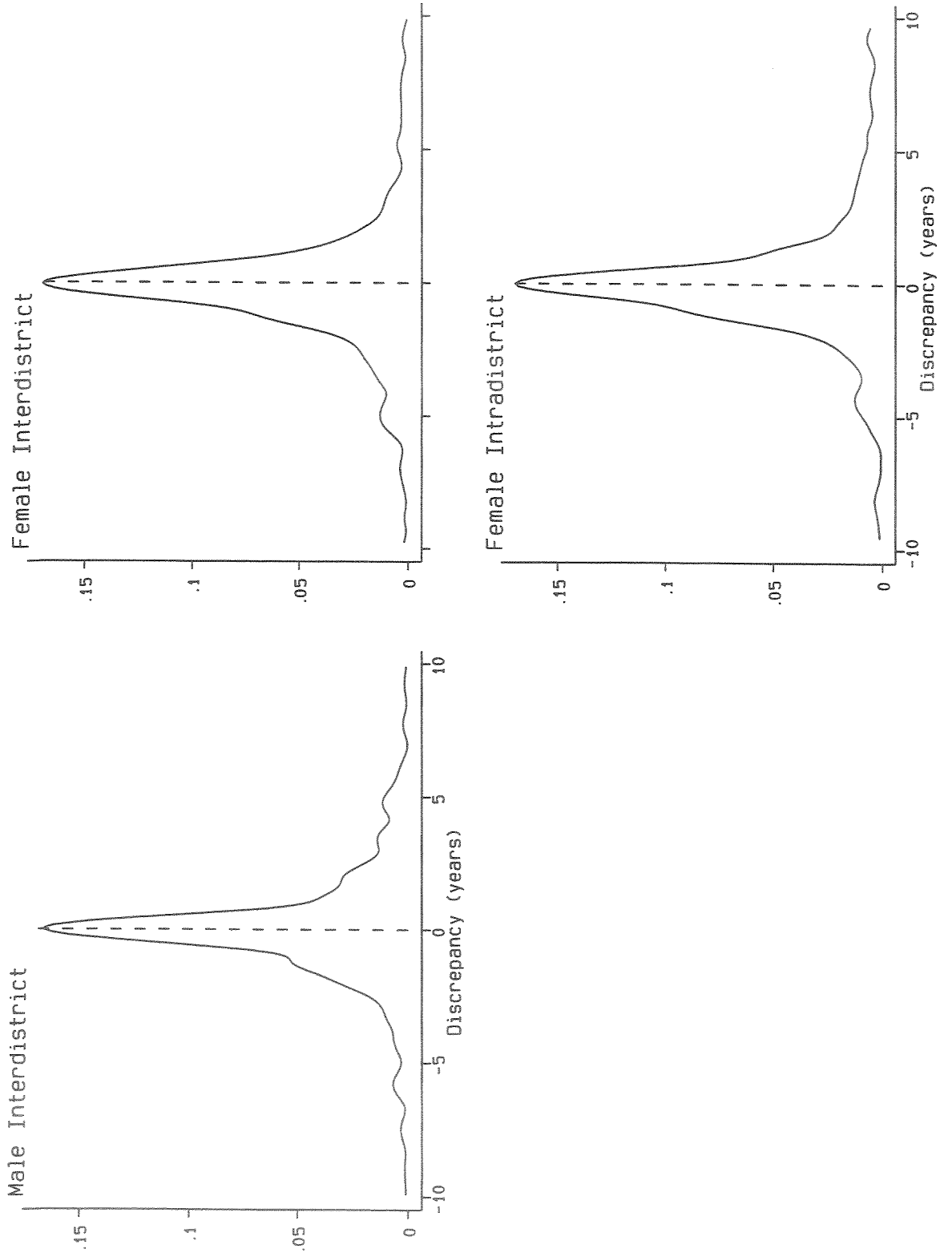


Table 1

Percent matches of reported residence at birth and residence at age 15

	STATE		DISTRICT			
	OF BIRTH	@ AGE 15	OF BIRTH	@ AGE 15		
	SAME	SAME	SAME ADJ.	SAME ADJ.		
	(1)	(2)	(3)	(4)	(5)	(6)
MALES						
MFLS2=MFLS1	95.2	91.9	85.7	93.1	83.4	90.6
FEMALES						
MFLS2=MFLS1	95.3	94.1	84.4	95.3	83.5	92.7
MFLS1						
Round1=Round2	97.4	97.7	96.0		94.6	

Notes: State of birth and at age 15 asked twice of females in MFLS1 (in Round 1 and Round 2). Males only asked in Round 3. MFLS2=MFLS1 Round 1 response for females. ADJ. denotes match with same district or geographically adjacent district.

Table 2
Summary Measures of Retrospective Responses

		MALES	All	F E M A L E S	
				Inter-district	Intra-district
1. NUMBER OF MOVES					
1.1	Reported in MFLS1	968	1878	690	1188
1.2	Reported in MFLS2	884	1652	700	952
1.3	Difference (MFLS1-MFLS2)	84	226	-10	236
2. MATCH RATES					
% respondents report same number of moves					
2.1	in MFLS1 and MFLS2	57.0	42.8	70.2	44.8
% moves reported in other survey given move					
2.2	reported in MFLS1	58.0	70.7	69.0	66.7
2.3	reported in MFLS2	63.5	75.9	67.9	81.8
% of time from age 15 to MFLS1 interview					
2.4	spent in same district	85.7	.	88.7	.
3. DATE DISCREPANCIES					
Percentage of reported dates for matched moves					
3.1	MFLS1 precedes MFLS2	45	48	53	47
3.2	Same	11	8	8	8
3.3	MFLS1 follows MFLS2	44	44	39	45
Discrepancy in date of reported move					
Mean quartic root of (absolute value of MFLS1 date - MFLS2 date) in months					
3.4	MFLS1 precedes MFLS2 (standard error)	1.87 (0.05)	1.89 (0.03)	1.91 (0.05)	1.92 (0.04)
3.5	MFLS2 precedes MFLS1 (standard error)	1.54 (0.05)	1.66 (0.03)	1.54 (0.05)	1.71 (0.04)
3.6	Difference (standard error)	0.33 (0.07)	0.23 (0.04)	0.37 (0.07)	0.21 (0.06)

Table 3
 Probit estimates of probability migration reported in MFLS2
 conditional on migration having been reported in MFLS1

	M A L E S		ALL MOVES		F E M A L E S		INTRA-DISTRICT	
	INTER-DISTRICT		β	t	β	t	β	t
PERSONAL ATTRIBUTES								
Age (years)	0.029	(2.89)	0.005	(0.70)	-0.007	(0.52)	0.009	(1.02)
Yrs since move: 0-5	-0.082	(1.39)	-0.078	(2.36)	-0.218	(2.90)	-0.043	(1.08)
6-15 years	0.025	(1.33)	0.017	(1.25)	0.041	(1.58)	-0.013	(0.76)
> 15 years	-0.029	(2.27)	0.005	(0.42)	0.020	(0.93)	-0.006	(0.40)
Education - Completed								
Standard 1-5	0.323	(2.19)	0.106	(1.30)	0.049	(0.33)	0.098	(0.95)
Standard 6	0.346	(2.20)	0.152	(1.43)	-0.022	(0.12)	0.271	(1.97)
Form 1-3	0.180	(0.70)	0.332	(2.11)	0.400	(1.21)	0.238	(1.25)
Form 4-5	0.660	(3.03)	0.485	(3.00)	0.442	(1.64)	0.162	(0.72)
Form 6/more	1.090	(4.49)						
Father-some educ	0.031	(0.29)	0.192	(2.73)	-0.122	(0.94)	0.376	(4.19)
Mother-some educ	0.391	(2.77)	-0.036	(0.35)	0.066	(0.37)	-0.132	(0.96)
Ethnicity								
Chinese	-0.130	(1.18)	0.225	(2.99)	0.052	(0.38)	0.385	(4.03)
Indian	0.035	(0.20)	0.290	(2.87)	0.138	(0.73)	0.296	(2.29)
SURVEY OPERATIONS								
At interview								
Someone helped	0.012	(0.07)	-0.025	(0.27)	0.046	(0.25)	-0.137	(1.19)
Spouse present	-0.299	(2.87)	-0.140	(1.82)	-0.149	(1.01)	-0.089	(0.92)
Reliability of responses								
Very reliable	0.035	(0.10)	0.526	(3.46)	0.978	(3.28)	0.227	(1.20)
Average	-0.081	(0.23)	0.452	(3.16)	0.709	(2.56)	0.293	(1.63)
Interview time (hours)								
in MFLS1	-0.023	(0.15)	0.011	(0.26)	0.227	(3.00)	-0.040	(0.72)
in MFLS2	0.370	(1.87)	0.051	(0.54)	-0.188	(1.00)	0.152	(1.28)
District boundary	0.088	(0.82)	0.162	(2.37)	0.347	(2.70)	0.055	(0.64)
MIGRATION ATTRIBUTES								
Duration of move								
(1) if 2-4 years	0.487	(4.06)	0.329	(3.62)	0.536	(3.27)	0.241	(2.07)
5-10 years	0.629	(4.15)	0.579	(5.01)	0.693	(3.24)	0.597	(4.07)
>10 years	0.518	(3.19)	0.448	(3.61)	0.636	(2.97)	0.481	(2.88)
Distance (kms)	-0.043	(0.99)			-0.067	(1.28)		
Inter-state move	0.310	(3.06)	0.264	(2.46)	0.272	(2.07)		
Inter-district move			-0.089	(0.94)				
First time in dist.	0.552	(4.43)	0.267	(3.01)	0.322	(2.15)		
First move	0.194	(1.62)	-0.091	(0.93)	0.155	(1.03)	0.150	(1.43)
Move while married	0.074	(0.54)	0.070	(0.91)	-0.176	(1.14)	0.036	(0.38)
Moved with spouse	0.877	(6.31)			1.286	(7.22)		
Circular move	-0.182	(1.55)	-0.144	(1.61)	-0.086	(0.45)	-0.012	(0.11)
# moves reported	0.020	(0.94)	-0.085	(4.62)	-0.058	(1.80)	-0.130	(5.24)
COINCIDENT EVENTS								
Husband changed job	0.258	(2.60)	0.121	(0.95)	0.121	(0.57)	0.095	(0.56)
Marriage	-0.141	(1.03)	0.240	(2.81)	0.062	(0.42)	0.429	(3.82)
Birth of child	-0.136	(0.80)	0.061	(0.72)	0.034	(0.23)	0.047	(0.42)
χ^2 (all covariates)	268.79		245.37		196.15		188.88	
\ln likelihood	-524.27		-1060.56		-328.01		-661.74	
Sample size	968		1878		690		1188	

Notes: Sample is all reported moves in MFLS1. Asymptotic t statistics in parentheses. Education Form 4-5 includes Form 6 and above for females. Year since move measured from MFLS1 interview.

Table 4

Time discrepancy between migration dates

OLS estimates: Dependent variable= $\sqrt[4]{|MFLS1 \text{ date} - MFLS2 \text{ date}|}$ measured in months)

	M A L E S		ALL MOVES		F E M A L E S		INTRA-DISTRICT	
	INTER-DISTRICT				INTER-DISTRICT			
	β	t	β	t	β	t	β	t
PERSONAL ATTRIBUTES								
Age (in years)	0.009	(1.58)	0.012	(3.01)	0.016	(2.20)	0.012	(2.32)
Education completed								
Standard 1-5	0.016	(0.14)	-0.014	(0.26)	-0.047	(0.58)	0.034	(0.49)
Standard 6	-0.135	(1.11)	-0.003	(0.05)	-0.040	(0.37)	0.009	(0.11)
Form 1-3	-0.091	(0.50)	-0.184	(1.95)	-0.354	(2.30)	-0.089	(0.74)
Form 4-5	-0.296	(1.93)	-0.295	(3.06)	-0.733	(5.51)	-0.010	(0.07)
Form 6/more	-0.599	(4.09)
Father-some educ	0.017	(0.23)	-0.122	(2.72)	-0.055	(0.79)	-0.146	(2.50)
Mother-some educ	0.085	(0.96)	-0.161	(2.43)	-0.058	(0.59)	-0.259	(2.91)
Ethnicity								
Chinese	0.282	(3.38)	-0.087	(1.77)	0.050	(0.65)	-0.148	(2.31)
Indian	-0.062	(0.58)	-0.167	(2.60)	-0.036	(0.39)	-0.206	(2.36)
SURVEY OPERATIONS								
At interview								
Someone helped	0.027	(0.22)	0.032	(0.50)	0.065	(0.66)	0.007	(0.09)
Spouse present	0.096	(1.30)	-0.082	(1.59)	0.046	(0.56)	-0.163	(2.47)
Reliability of responses								
Very reliable	-1.114	(3.39)	0.005	(0.05)	0.158	(0.77)	-0.078	(0.55)
Average	-1.066	(3.24)	0.136	(1.22)	0.267	(1.35)	0.089	(0.66)
Interview time								
in MFLS1	-0.151	(1.30)	0.036	(1.33)	-0.035	(0.85)	0.074	(1.99)
in MFLS2	0.224	(1.52)	0.050	(0.83)	-0.012	(0.10)	0.046	(0.64)
District boundary	-0.007	(0.10)	-0.009	(0.20)	-0.144	(2.17)	0.076	(1.31)
MIGRATION ATTRIBUTES								
(1) if moved during								
1960-1969	0.085	(0.96)	-0.011	(0.18)	0.036	(0.39)	-0.029	(0.35)
1970-1976	0.096	(0.74)	-0.130	(1.61)	0.063	(0.49)	-0.184	(1.73)
Distance (kms)	0.011	(0.38)	.	.	-0.023	(0.81)	.	.
Inter-state move	0.052	(0.70)	-0.021	(0.30)	0.013	(0.18)	.	.
Inter-district move	.	.	0.027	(0.43)
First time in dist.	0.117	(1.53)	-0.014	(0.24)	0.024	(0.30)	.	.
First move	0.078	(0.98)	-0.101	(1.62)	0.001	(0.01)	-0.138	(1.98)
Move while married	-0.143	(1.64)	0.109	(2.17)	0.180	(2.19)	0.074	(1.13)
Moved with spouse	-0.045	(0.56)	.	.	-0.112	(1.47)	.	.
Circular move	0.130	(1.50)	0.123	(2.10)	0.095	(0.78)	0.189	(2.56)
# moves reported	0.004	(0.33)	-0.041	(3.05)	0.029	(1.46)	-0.098	(4.66)
COINCIDENT EVENTS								
Husband changed job	-0.356	(5.18)	-0.053	(0.71)	0.055	(0.50)	-0.129	(1.26)
Marriage	0.035	(0.34)	-0.309	(5.58)	-0.344	(4.10)	-0.333	(4.50)
Birth of child	0.153	(1.26)	-0.165	(3.13)	-0.070	(0.87)	-0.223	(3.21)
TELESCOPING								
MFLS1 date earlier	0.282	(4.27)	0.287	(6.90)	0.301	(4.74)	0.272	(4.97)
Intercept	2.263	(5.20)	1.499	(6.16)	1.036	(2.52)	1.743	(5.65)
F(all covariates)	5.09		12.60		5.89		9.55	
R ²	0.23		0.23		0.29		0.25	
Sample size	561		1328		476		792	

Notes: Sample is all matched moves in MFLS1 and MFLS2. Education Form 4-5 includes Form 6 and above for females.

Table 5

Time discrepancy between migration dates: Telescoping and salient coincident events

OLS Estimates: Dependent variable= $\sqrt{(\text{MFLS1 date}-\text{MFLS2 date})}$ measured in months

	M A L E S		A L L M O V E S		F E M A L E S	
	INTER-DISTRICT β	t	β	t	INTER-DISTRICT β	INTRA-DISTRICT t
MFLS1 reported date earlier	0.203	(2.34)	0.079	(1.48)	0.131	(1.60)
* Husband changed job	0.323	(2.46)	0.020	(0.13)	-0.220	(1.01)
* Marriage	-0.157	(0.75)	0.524	(5.52)	0.635	(4.54)
* Birth of child	-0.319	(1.32)	0.057	(0.56)	-0.241	(1.54)
					0.048	(0.70)
					0.105	(0.52)
					0.475	(3.74)
					0.190	(1.39)

Notes: Sample is all matched moves in MFLS1 and MFLS2. Model includes all covariates in Table 5 in addition to interaction between indicator variable for MFLS1 reported earlier and the three coincident events.

APPENDIX TABLE 1
Means and (standard errors) of covariates

	M A L E S		F E M A L E S			
	INTER-DISTRICT		INTER-DISTRICT		INTRA-DISTRICT	
PERSONAL ATTRIBUTES						
Age (in years)	42.232	(0.305)	37.343	(0.290)	36.428	(0.233)
Years since move	16.88	(0.310)	13.96	(0.300)	11.95	(0.240)
Spline: 0-5	5.526	(0.035)	5.439	(0.044)	5.083	(0.045)
6-15	6.962	(0.128)	6.069	(0.156)	4.818	(0.124)
>15	4.397	(0.203)	2.456	(0.153)	2.052	(0.121)
Education: Fraction						
Standard 1-5	0.323	(0.015)	0.325	(0.018)	0.314	(0.013)
Standard 6	0.299	(0.015)	0.178	(0.015)	0.172	(0.011)
Form 1-3	0.060	(0.008)	0.052	(0.008)	0.060	(0.007)
Form 4-5	0.085	(0.009)	0.052	(0.008)	0.036	(0.005)
Form 6/more	0.092	(0.009)
Fraction						
Father had some educ	0.457	(0.016)	0.467	(0.019)	0.412	(0.014)
Mother had some educ	0.219	(0.013)	0.168	(0.014)	0.118	(0.009)
Chinese	0.277	(0.014)	0.326	(0.018)	0.375	(0.014)
Indian	0.123	(0.011)	0.161	(0.014)	0.142	(0.010)
SURVEY OPERATIONS						
Fraction						
Someone helped at interview	0.102	(0.010)	0.151	(0.014)	0.149	(0.010)
Spouse present at interview	0.409	(0.016)	0.254	(0.017)	0.254	(0.013)
Very reliable responses	0.414	(0.016)	0.277	(0.017)	0.303	(0.013)
Average reliability responses	0.568	(0.016)	0.670	(0.018)	0.642	(0.014)
Interview time (hours)						
MFLS1	0.918	(0.010)	2.001	(0.034)	1.903	(0.023)
MFLS2	0.526	(0.008)	0.878	(0.013)	0.883	(0.011)
Fraction involve						
district boundary change	0.286	(0.015)	0.333	(0.018)	0.399	(0.014)
MOVE CHARACTERISTICS						
Duration of move (years)	6.329	(0.246)	8.213	(0.293)	6.792	(0.199)
Fraction duration 2-5 years	0.253	(0.014)	0.261	(0.017)	0.286	(0.013)
5-10 years	0.151	(0.012)	0.190	(0.015)	0.243	(0.012)
>10 years	0.221	(0.013)	0.320	(0.018)	0.224	(0.012)
Fraction moves between						
1960-1969	0.206	(0.013)	0.270	(0.017)	0.247	(0.013)
1970-1976	0.105	(0.010)	0.180	(0.015)	0.243	(0.012)
Distance (kms)	1.251	(0.037)	1.449	(0.046)	0.002	(0.002)
Fraction moves involve						
interstate move	0.607	(0.016)	0.581	(0.019)	0.005	(0.002)
first time moved to district	0.320	(0.015)	0.536	(0.019)	0.382	(0.014)
first moves	0.320	(0.015)	0.536	(0.019)	0.415	(0.014)
circular moves	0.233	(0.014)	0.110	(0.012)	0.332	(0.014)
moves while married	0.238	(0.014)	0.233	(0.016)	0.378	(0.014)
moved with spouse	0.205	(0.013)	0.267	(0.017)	0.018	(0.004)
Number of moves reported	4.988	(0.102)	4.035	(0.098)	3.223	(0.065)
COINCIDENT EVENTS: Fraction (within 6 months of move)						
husband changed job	0.355	(0.015)	0.094	(0.011)	0.076	(0.008)
married	0.135	(0.011)	0.372	(0.018)	0.294	(0.013)
birth of child	0.095	(0.009)	0.223	(0.016)	0.220	(0.012)
Fraction MFLS1 reported date						
precedes MFLS2	0.447	(0.021)	0.532	(0.023)	0.470	(0.018)
* Husband changed job	0.171	(0.016)	0.059	(0.011)	0.044	(0.007)
* Marriage	0.050	(0.009)	0.210	(0.019)	0.184	(0.014)
* Birth of child	0.037	(0.008)	0.113	(0.015)	0.117	(0.011)

Number of moves in MFLS1
Number of matched moves in MFLS1 & 2

968
561

690
476

1188
792