December 22, 2003

Instructions for creating the circular migration file

Strategy:
The goal is to create a file that characterizes the migration experience of migrants in terms of departures and returns from the village of origin. A departure and a return constitutes a circular migration. A person can have numerous such circular migrations. The first one is "order 1;" and the order for subsequent circular migrations increase one at a time.

We will build relatively few variables for this first file. But it would be nice to program it in such a manner that additional variables can be built as research needs arise.

For our initial analysis using these circular migration files we will be analyzing a subset of people for whom we have life history data. But given that that selection might change, or that different folks might have reasons to select different samples, our feeling is to go ahead and create these circular migration records for all for whom we have life history data. But if there is a good reason to create the circular migration records just for our initial analysis sample, let's discuss it.

These are variables that will be made available to folks here at UNC, at IPSR, and relatively soon to anyone who wants them.

We refer to data collection in 1994 and 2000, but it needs to be remembered that the migrant follow-up data collection spilled over into 1995 and 2001.

Variables:

ID#	Person ID#
ORDER	Order of the circular migration (1,2,...)
TYPEi	type of circular migration, order i 1= returned to village 2= still gone 3= never left village 4= not in origin village since age 13
RESNUMi	a count of stops (e.g.specific residences) before the circle is completed by the individual returning to the origin village

RESij — the jth residence in circular migration order i; coded using the actual residence codes; j will go as high as necessary. (Note that when finished, j will equal RESNUMi for the ith circular migration interval.)

AGERESij — age when the person was first in RESij

BEGAGEi — age at the start of circular migration order i

BEGYRi — year of the start of circular migration order i. Convert from the BE system to the AD system. This is YEAR in 1994 and BEYEAR in 2000.

ENDAGEi — age at the end of circular migration order i; use SAS missing if circular migration order i is an open interval (TYPEi = 2)

DURi — duration of circular migration order i
= (ENDAGEi - BEGAGEi)

STUDENTi — whether the person was a student at any time during circular migration order i
1= yes 0= no

CONSTRi — whether the person was a construction worker for the entire duration of circular migration order i
1= yes 0= no

PROXYi — Whether the life history data was a proxy report
1= yes 0= no

SOURCEBi — source of the data at the beginning of the circular migration order i
1= village 94
2= migration FU 94
3= village 00
4= migrant R–>U 00
5= migrant R–>R 00

SOURCEEi — source of the data at the end of the circular migration order i
1= village 94
2= migration FU 94
3= village 00
4= migrant R–>U 00
5= migrant R–>R 00

TWOSOUi whether 1994 and 2000 are available for this person.
1=both 1994 and 2000 data are available
2=just 1994 data are available
3=just 2000 data are available

LEFTCEN Left censored. This is a variable that is a characteristic of a person and not a characteristic of any given interval, and hence no "i" is part of the name. Yet lets make it a variable in each of the circular migration records. It should be 1 for those for whom we do not have a 1994 life history and the 2000 life history began at some age other than 13. Code 0 otherwise.

VILAT13 This is a variable that indicates whether the person was in the village at age 13 for the start of the life history calendar. This is also a variable that is a characteristic of a person and not a characteristic of any given interval, and hence no "i" is part of the name. Code 1 if the person was in the village at age 13 and 0 otherwise (except code SAS missing if LEFTCEN=1).

Overview:
For each person, the life history data are structured such that there is one record for each year/age for that person. In 1994, the records begin at age 13 and move forward to the person's current age. In 2000, the records begin either at age 13 or 1990 (if there was life history data on the person from the 1994 data collection). For each year/age there are up to 6 possible residence locations. We make no assumption that the residence at the end of one year is the same as the residence at the beginning of the next year. Even though there is one record for each year/age, conceptually it might be easier to visualize the data as one long string, beginning with the first residence at age 13.

Hand drawn figure 1 illustrates the process for the first circular migration interval. Start with the first residence at age 13. If it is not in the origin village, search the residence data until you find when the person was in the village. If never in the village, TYPE1=4 (never in the village). This is a terminal state; there are no more circular migration intervals for this person.

The first time the person is in the village (top of Figure 1) marks the start of looking for the first departure from the origin village (that is code for residence not equal to origin village code). If the person never left the village after the first time in the village, then TYPE1=3 (never left the village). This is a terminal state; there are no more circular migration intervals for this person.

Once you find that the person has left the village, center of Figure 1, the next step is to see the next destination. If there is no further residence, then TYPE1=2 (open interval). This is a terminal state; there are no more circular migration intervals for this person. If the next residence is the origin village, then TYPE1=1 (circular migration).

If the next residence is some place other than the origin village, then there is a repeatable loop, which we will call the "next destination loop." While in this next destination loop, the counter,

RESNUMi is accumulating, and RESij and AFERESij are being created. The only way out of the next destination loop is by returning to origin, in which case TYPE1=1, or by running out of destinations, in which case TYPE1=2.

Anytime a circular migration interval ends back in the village, TYPEi=1, this signals the beginning of the next circular migration interval as illustrated in the handwritten Figure 2. The person is finished whenever the last circular interval is TYPEi= 2, 3, or 4.

We are envisioning that each circular migration interval order i will constitute a record. Thus some will have multiple records. But if there is a better way to arrange things, lets talk.

Once there is a beginning and end point for the circular migration interval, then other variables will be created. In the instructions below, there is a "student loop" described to create STUDENTi, and a "construction loop" to create CONSTRi.

Which data to use if more than one available?

Based on some numbers from a memo that Rick wrote Dec 4, 2002, it looks like we have almost 8,000 cases where we have data from both 1994 and 2000. In some cases there will be an overlap of 4 years (1990-1993). In other cases there will be an overlap that is longer.

In all cases where there is overlap for calendar year1994, we will want to use the data from 2000. The reasoning is that the calendar year 1994 data from the1994 fieldwork are based on only half or less of 1994 because of the timing of the fieldwork. Similarly, if we have data for calendar year 1995 from form 10 in 1994 and from either form 5 or form 11 in 2000, we will want to use the 2000 data.

When there is overlap on calendar years 1990-1993 (and perhaps 1994 for migrants interviewed in 1995), if one of the sources is a proxy report and the other is not, always use the one where the person for whom we have life history data is also the person who was the respondent. A definition of proxy reports is provided below.

For all others for calendar years 1990-1993 when there are two sources, use the 1994/5 data. The reasoning is that the data collection occurred closer in time to when the events were occurring.

If a person has data from both the 1994 and the 2000 fieldwork, the life history data will have to be stitched together. There will be cases where the circular migration interval begins with 1994 data and then ends with 2000 data. When this happens, use the 1994 instructions as appropriate and then switch to 2000 instructions. This can be accomplished by first creating a unified file. But it can also be accomplished by programming such that once the 1994 data should end the program automatically moves to the 2000 data.

SUSANA WILL CHECK THE 1994 RESIDENCE CODES AGAINST THE 2000 RESIDENCE

CODES. IF THEY ARE DIFFERENT THEN WE WILL NEED TO RECONCILE THEM. THIS INCLUDES DEALING WITH PROVINCES THAT MIGHT HAVE SUBDIVIDED. WE WILL WANT THE SAME RESij CODING THROUGHOUT THE FILE IRRESPECTIVE OF WHETHER THE DATA CAME FROM 1994 OR 2000.

The detailed instructions below are based on the household survey data, that is the data collected in the village based fieldwork. It will be necessary to check if the migrant life history data is coded exactly as the household survey data.

A NOV 19 EMAIL FROM RICK SAYS THAT ALL THE 1994 MIGRANT LIFE HISTORIES END AT AGE 35. SOUNDS LIKE SOMETHING IS WRONG AND WE NEED TO FIND OUT WHY

Detailed instructions for 1994 village household survey.
Here and throughout, we are assuming that there are 51 villages. Each has an ID number. For a person from one of those villages, his or her origin village residence code is the ID number for that village. (It is logically possible that someone was residing in one of our origin villages in 1994 and a different one of the 51 villages in 2000. This would create ambiguity as to which was the true origin village. However, in practice, we do not have links if people move from one of the 51 villages to another; and hence the potential ambiguity re which is the origin village is invisible to us.)

Unlike the 2000 data, the 1994 data have something called “sequence order of the places of residence.” These programming instructions do not use the sequence order of place of residence. There is not a specific reason for not using the sequence number, rather it was not clear that it would be helpful. Hence, the 1994 and 2000 instructions are more similar than they might otherwise be.

We are using the detailed information on place of residence and not the condensed information made available on the public use files.

These instructions will be in a series of steps, beginning with step A.

A. Start at age 13 and see if the person for whom the life history data is being recorded was present. CEP94 is the id for the person for whom the life history data is being recorded. CEPINF is the id for the person providing the life history information. If CEP94=CEPINF, then PROXY1=0; otherwise, PROXY1=1. Also create SOURCEB1, which will be coded 1 or 2 depending on whether the data come from the village fieldwork or the migrant follow-up. Also create TWOSOU1 indicating whether 2000 and 1994 data are available for this person. Then go to step B.

B. Check their first listed residence. If Q5_5PL1 is coded 10000, 20000, or 2xxxx, then the person's first listed residence is in their origin village. Code VILAT13=1. Go to step D. If Q5_5PL1 is NOT coded 10000, 20000, or 2xxxx, then the person did not start out in the origin village at 13. VILAT13=0 unless LEFTCEN=1 in which case code VILAT13=SAS missing. Go to step C. (Note that we are treating missing data as if the residence was not in the origin village.)

C. Check the next residence. This might be Q5_5PL2 if there was a second residence or it might be Q5_5PL1 on the age 14 record. The general idea is to look for the first instance where Q5_5PLi is coded 10000, 20000, or 2xxxx, indicating that the person is in the village for the first time since age 13. Once this is found, go to step D. If no place (Q5_5PLi) is found that equals 10000, 20000, or 2xxxx (using both the 1994 and 2000 data), then this person has not lived in the village since age 13. Code TYPE1=4. and code remaining variables SAS missing. That is it for this person.

D. At this point, we have people who have started in the origin village at some point at or since age 13. The next step is to see when the person first left the origin village. Check the records to find the first instance where Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999. If such a place is never found before a person's current age, that is the last record for 1994 and 2000, then it means that the person never left the village. Code TYPE1=3. BEGAGE1= age (Q5_1_1) represented by the row where the first Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999 was found. YEAR1=YEAR represented by the row where the first Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999 was found. Convert to the AD system of years. ENDAGE1 equals the age on the last row of the person's life history. Calculate DUR1 and SOURCEE1 as described in step G below. Then that is it for this person. Code any other variables SAS missing.

For the rest, the first place found where Q5_5PLi NOT equal to 10000, 20000, 2xxxx, 99998 or 99999 represents the first non-origin destination. This is the first destination, so set RESNUM1=1. RES11=the first Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999. BEGAGE1= age (Q5_1_1) represented by the row where the first Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999 was found. YEAR1=YEAR represented by the row where the first Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999 was found. Convert to the AD system of years. AGERES11=age (Q5_1_1) represented by the row where the first Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999 was found. (Note that BEGAGE1 and AGERES11 are the same age.) Go to step E.

E. At this point the first circular migration interval has begun. The next step is to look for the next residence after RES11. There are three possibilities: there is no next residence, the next residence is the origin village, or the next residence is a different destination. We will discuss each in turn.

If there is no next residence (that is you never find a subsequent residence that differs from RES11, using the 1994 and 2000 data), then it means that there are no other places. Code

TYPE1=2. ENDAGE1 equals the age on the last row of the person's life history. Calculate DUR1 and SOURCEE1 as described in step G below. Calculate STUDENT1 by doing the student loop described in step H. Calculate CONSTR1 by doing the construction loop described in step I. Then that is it for this person.

If the next residence is the origin village (that is the next Q5_PLi NOT equal to RES11 or 99998 or 99999 AND equal to 10000, 20000 or 2xxxx) this is the terminating stop for this circular migration order 1. ENDAGE1= the age (Q5_1_1) represented by the row where the terminating stop for this circular migration was found. Go to step G

If the next residence is a new stop that is not equal to the origin village or to the current destination (that is the Q5_PLi NOT equal to RES11 or 99998 or 99999 AND NOT equal to 10000, 20000 or 2xxxx) then this is the second destination. (Note that this is a conservative view of next destination. The next listed destination could be such that Q5_PLi equal to RES11. This might indicate that the person moved within the destination. If the destination was a province, this could be a move of considerable distance and considerable change in type of place.) Set RES12=the first Q5_PLi NOT equal to RES11 or 99998 or 99999 AND NOT equal to 10000, 20000 or 2xxxx. AGERES12=age (Q5_1_1) represented by the row where RES12 was found. Go to step F.

F. We are now in the next destination loop. We will go though the steps for the third non-origin stop in the first circular migration interval, but the idea is to through this loop until the person either goes back to origin or there is no other place where the person goes.

Again, like in step E, there are three possibilities: there is no next residence, the next residence is the origin village, or the next residence is a different destination.

If there is no next village (that is you never find a subsequent residence that differs from RES12), then it means that there are no other places. (Here and elsewhere, the assumption is being made that if the 1994 life history ends, then the 2000 life history is checked.) Code TYPE1=2. ENDAGE1 equals the age on the last row of the person's life history. Calculate DUR1 and SOURCEE1 as described in step G below. Calculate STUDENT1 by doing the student loop described in step H. Calculate CONSTR1 by doing the construction loop described in step I. Then that is it for this person.

If the next residence is the origin village (that is the next Q5_PLi NOT equal to RES12 or 99998 or 99999 AND equal to 10000, 20000 or 2xxxx) this is the terminating stop for this circular migration order 1. ENDAGE1= the age (Q5_1_1) represented by the row where the terminating stop for this circular migration was found. Go to step G.

If the next residence is a new stop that is not equal to the origin village or to the current destination (that is the Q5_PLi NOT equal to RES12 or 99998 or 99999 AND NOT equal to 10000, 20000 or 2xxxx) then this is the third destination. Set RES13=the first Q5_PLi NOT equal to RES12 or 99998 or 99999 AND NOT equal to10000, 20000 or 2xxxx. AGERES13=age

(Q5_1_1) represented by the row where RES13 was found. Go back through the next destination loop looking for the fourth destination, a return to origin or the fact that the person did not go to any other places.

G. At this point we have circular migration order 1 where the person returned to the village.

Set TYPE1=1.

DUR1= (ENDAGE1-BEGAGE1) Note: if this is a negative number, that suggests we have a problem.

SOURCEE1= the source of the data being used when the person returned to the village, thus ending the circular migration order 1.

Calculate STUDENT1 by doing the student loop described in step H

Calculate CONSTR1 by doing the construction loop described in step I

Once STUDENT1 and CONSTR1 are finished, circular migration order 1 is finished for this person. Then go to step J to begin work on circular migration order 2.

H. Student loop.
The goal of the student loop is to determine whether the person was a student at any point while gone from the origin village. Once it is determined that the person has ever been a student while gone from the village, STUDENTi gets coded 1, and that is it.

Figure 3 might help provide an overview. It is for a person whose migration began at age k+1 (v) and ended at age k+6 (z). For each age, we know from question 5.6 whether the person was a student at that age. Then there are up to six residence locations (sequence order) where the person might have been a student. If there is an age where the person was not in the village, in this case ages k+2, k+3, k+4, and k+5, if the person was a student (question 5.6) we know he/she was a student outside the origin village, and hence STUDENTi gets coded 1, and that is it. However, if the person was a student in the age beginning or ending the circular interval, in this case k+1 and k+6, then it is necessary to go to the sequence of student variables to see if the person was a student while outside the village.

What is considered a student? Codes 1-20, and 96, are considered to be student. Code 0 indicates the person never studied. In the life history data there should not be any code 0's. If there are we will need to decide what to do with them. 95, every level of non formal education, is NOT a student. 97 stands for kindergarten, and there should not be any. If there are we will need to decide what to do with them. 98, NA, and 99, missing/don’t know, are to be treated as NOT being a student. For codes 95, 96, 97, 98 and 99 we need a sense of how frequently they occur. We are assuming that they are relatively rare.

If TYPEi= 3 or 4, code STUDENTi= SAS missing

Starting with BEGAGEi and continuing until ENDAGEi, check Q5_6ED. If Q5_6ED= any of {1 through 20 or 96} then check to see if Q5_1_1(the age in that row) = either BEGAGEi or ENDAGEi. If it does NOT, that means it is an age for which the person was gone the entire age, hence STUDENTi=1. If it is either BEGAGEi or ENDAGEi, then need to check whether the schooling was in the origin village or not, as follows.

For BEGAGEi, need to determine whether the schooling occurred after the person left the village (and if a short move, before the person returned to the village). There are up to six sequence orders indicating where the person was while being a student, Q5_6OR1 through Q5_6OR6. This needs to be checked, but it appears that 98 is used if the person was not a student while residing in residence"i". Also, this needs to be checked, but it appears that Q5_6OR1 through Q5_6OR6 line up with Q5_PL1 through Q5_PL6. Recall from step D, that the first place found where Q5_5PLi NOT equal to10000, 20000, 2xxxx, 99998 or 99999 represents the first non-origin destination. The "i" in this Q5_5PLi is the indicator for the "i" to be checked in Q5_6O1 through Q5_6OR6. (For the purposes of the next paragraph, let this "i' be denoted as v.) If the code for this Q5_6ORi begins with 3 through 9 or the person was abroad, then code STUDENTi=1. If not continue checking the Q5_ORi+1, 2, 3, 4, 5, but make sure to stop if the person moved back to the village during this age. See steps E and F above.

For ENDAGEi need to determine whether schooling occurred before the person returned to the village. Note that this is only a relevant concern if TYPEi=1. The location of the return to the village in the Q5_6PLi series was found in step E or F above. Let this be location z. If ENDAGEi not equal to BEGAGEi, then check all the Q5_6OR1 through Q5_6ORi that precede the return to the village, that is for all i less than z. If the code for any of these Q5_6ORi begins with 3 through 9 or the person was abroad, then code STUDENTi=1. For those cases where ENDAGEi=BEGAGEi, check all the Q5_6OR1 through Q5_6ORi for i greater than v and less than z. If the code for any of these Q5_6ORi begins with 3 through 9 or the person was abroad, then code STUDENTi=1.

I. Construction loop

The goal of the construction loop is to determine whether, while away from the origin village, the person was always in construction. If always in construction, then CONSTRi= 1. Otherwise, CONSTRi=0. The first time a non-construction job is found, CONSTRi=0, and that is it for this person/circular migration interval.

Occupation information is found in Q5_7OC1 through Q5_7OC6. Construction is code 4; all other codes, including 9 (missing data) are considered not construction. Note that a legitimate not applicable does not count as "not construction."

As described in the student loop above, location v is the beginning of the circular migration order i (put differently, it is the first stop outside the village), and location z is the end (the return

to the origin village) if it is a TYPEi=1 circular migration. If it is a TYPEi=2 circular migration, then there is no z, just the end of the life history.

Figure 4 provides an overview. It is for a person whose migration began at age k+1 and ended at k+4. There are up to 6 occupations for any given age. For each occupation there is a variable, Q5_7OC1 through 6 that provides the code for the occupation, and one variable, Q5_7OR1 through 6 that provides the sequence order for that location. It appears (but this will have to be checked) that a person could have more than one occupation in a destination, and these multiple occupations could be sequential or simultaneous. We can not distinguish between the sequential and simultaneous possibilities.

Start with the Q5_7OC variable associated with location v, the beginning of the circular migration. If this Q5_7OC variable is coded 1,2, 3, 5,6, or 9, then this is not construction, and CONSTRi should be coded 0 and that is the end of the construction loop. If this Q5_7OC variable is coded 4, then the person is in construction and it is necessary to check the next occupation in location v. If the next Q5_7OC variable in location v is coded 1,2, 3, 5,6, or 9, then this is not construction, and CONSTRi should be coded 0 and that is the end of the construction loop. If this next Q5_7OC variable is coded 4, then the person is in construction and it is necessary to check the next occupation in location v. Continue like this until there are no more construction variables in location v. Then go to location v+1 checking for construction vs non-construction. Then continue to location v+2, and so forth until location z is reached. Location z is the return to the origin village for TYPEi=1. Check all the occupations in location z-1. If they are still all construction, code 4, then CONSTRi is to be coded 1. (If TYPEi=2 {open interval} then check until you run out of places to check.)

J. The work on circular migration order 2 is identical to that on circular migration order 1 with two exceptions. First, i will be 2 instead of 1. Second, there is no need to check whether they are in the village. That fact that circular migration order 1 ended in the village means that circular migration order 2 will start in the village.

TWOSOU2=TWOSOU1

If we are still in the 1994 data, then PROXY2=PROXY1

BEGAGE2=ENDAGE1

SOURCEB2=SOURCEE1.

Then follow the logic at step D and beyond. Proceed with circular migration order 3, and so forth as illustrated in the hand drawn figure 2.

<u>Detailed instructions for 2000 village household survey</u>.
Conceptually, the steps for using the 2000 are quite similar to using the 1994 data.

Again, the assumption is that we are using the 51 1984 villages. Splits since 1984 are to be ignored.

For people for whom we have life history data from both 1994 and 2000 it is most likely they will be past age 13 by the time the 2000 data is used. In this case, step B does not apply, but rather the exact point where the 2000 data will be used is determined by the point where the 1994 data ended. Hence it could be Step C, D, E or F. When attempting to stitch together the 1994 and 2000 life histories, it is possible, indeed likely, that ages will not match for matched calendar years. How we handle such situations will depend on how frequently this occurs, and the patterns of not matching. Hence, we cannot describe procedures ahead of time. The first thing to do is obtain a count of how often this occurs, and whether the 1994 and/or 2000 reports are proxy reports.

A. Start at age 13 and see if the person for whom the life history data is being recorded was present. CEP00 is the id for the person for whom the life history data is being recorded. X5_05 is the id for the person providing the life history information. If CEP00=X5_05, then PROXY1=0; otherwise, PROXY1=1. Also create SOURCEB1, which will be coded 3,4, or 5 depending on whether the data come from the village fieldwork or the migrant follow-up. Also create TWOSOU1 indicating whether 2000 and 1994 data are available for this person. Then go to step B.

Note that there are some people, about 1200, for whom we do not have 1994 life history data and the 2000 life history data start at some age greater than 13. These are the folks for whom LEFTCEN=1. For these people, start the instructions at the first age for which we have data.

B. Check their first listed residence. If X5_5R1 is coded 2xxxx, then the person's first listed residence is in their origin village. Go to step D. If X5_5R1 is NOT coded 2xxxx, then the person did not start out in the origin village at 13. Go to step C. (Note that we are treating missing data as if the residence was not in the origin village.)

C. Check the next residence. This might be X5_5R2 if there was a second residence or it might be X5_5X1 on the age 14 record. The general idea is to look for the first instance where X5_5Ri is coded 2xxxx, indicating that the person is in the village for the first time since age 13. Once this is found, go to step D. If no place (X5_5Ri) is found that equals 2xxxx, then this person has not lived in the village since age 13. Code TYPE1=4, and code remaining variables SAS missing. That is it for this person.

D. At this point, we have people who have started in the origin village at some point at or since age 13. The next step is to see when the person first left the origin village. Check the records to find the first instance where X5_5Ri NOT equal to 2xxxx, SAS missing or 99999999. If such a place is never found before a person's current age, that is the last record for 1994 and 2000, then it means that the person never left the village. Code TYPE1=3. BEGAGE1= age (X5_1AGE) represented by the row where the first X5_5Ri NOT equal to2xxxx, SAS missing or 99999999 was found. YEAR1=BEYEAR represented by the row where the first X5_5Ri NOT equal

to2xxxx, SAS missing or 99999999 was found. Convert to the AD year system. ENDAGE1 equals the age on the last row of the person's life history. Calculate DUR1 and SOURCEE1 as described in step G below. Then that is it for this person and code the remaining variables SAS missing.

For the rest, the first place found where X5_5Ri NOT equal to 2xxxx, SAS missing or 99999999 represents the first non-origin destination. This is the first destination, so set RESNUM1=1. RES11=the first X5_5Ri NOT equal to 2xxxx, SAS missing or 99999999. BEGAGE1= age (X5_1AGE) represented by the row where the first X5_5Ri NOT equal to2xxxx, SAS missing or 99999999 was found. YEAR1=BEYEAR represented by the row where the first X5_5Ri NOT equal to2xxxx, SAS missing or 99999999 was found. Convert to the AD year system. AGERES11=age (X5_06AGE) represented by the row where the first X5_5Ri NOT equal to 2xxxx, SAS missing or 99999 was found. (Note that BEGAGE1 and AGERES11 are the same age.) Go to step E.

E. At this point the first circular migration interval has begun. The next step is to look for the next residence after RES11. There are three possibilities: there is no next residence, the next residence is the origin village, or the next residence is a different destination. We will discuss each in turn.

If there is no next residence (that is you never find a subsequent residence that differs from RES11), then it means that there are no other places. Code TYPE1=2. ENDAGE1 equals the age on the last row of the person's life history. Calculate DUR1 and SOURCEE1 as described in step G below. Calculate STUDENT1 by doing the student loop described in step H. Calculate CONSTR1 by doing the construction loop described in step I. Then that is it for this person.

If the next residence is the origin village (that is the next X5_5Ri NOT equal to RES11 or SAS missing or 99999999 AND equal to 2xxxx) this is the terminating stop for this circular migration order 1. ENDAGE1= the age (X5_1AGE) represented by the row where the terminating stop for this circular migration was found. Go to step G

If the next residence is a new stop that is not equal to the origin village or to the current destination (that is the X5_5Ri NOT equal to RES11 or SAS missing or 99999 AND NOT equal to 2xxxx) then this is the second destination. Set RES12=the first X5_5Ri NOT equal to RES11 or SAS missing or 99999 AND NOT equal to 2xxxx. AGERES12=age (X5_1AGE) represented by the row where RES12 was found. Go to step F.

F. We are now in the next destination loop. We will go though the steps for the third non-origin stop in the first circular migration interval, but the idea is to through this loop until the person either goes back to origin or there is no other place where the person goes.

Again, like in step E there three possibilities: there is no next residence, the next residence is the origin village, or the next residence is a different destination.

If there is no next village (that is you never find a subsequent residence that differs from RES12), then it means that there are no other places. Code TYPE1=2. ENDAGE1 equals the age on the last row of the person's life history. Calculate DUR1 and SOURCEE1 as described in step G below. Calculate STUDENT1 by doing the student loop described in step H. Calculate CONSTR1 by doing the construction loop described in step I. Then that is it for this person.

If the next residence is the origin village (that is the next X5_5Ri NOT equal to RES11 or SAS missing or 99999999 AND equal to 2xxxx) this is the terminating stop for this circular migration order 1. ENDAGE1= the age (X5_1AGE) represented by the row where the terminating stop for this circular migration was found. Go to step G

If the next residence is a new stop that is not equal to the origin village or to the current destination (that is the X5_5Ri NOT equal to RES12 or SAS missing or 99999 AND NOT equal to 2xxxx) then this is the third destination. Set RES13=the first X5_5Ri NOT equal to RES12 or SAS missing or 99999 AND NOT equal to 2xxxx. AGERES13=age (X5_1AGE) represented by the row where RES12 was found. Go back through the next destination loop looking for the fourth destination, a return to origin or the fact that the person did not go to any other places.

G. At this point we have circular migrations order 1 where the person returned to the village.

Set TYPE1=1.

DUR1= (ENDAGE1-BEGAGE1) Note: if this is a negative number, that suggests we have a problem.

SOURCEE1= the source of the data being used when the person returned to the village, thus ending the circular migration order 1.

Calculate STUDENT1 by doing the student loop described in step H

Calculate CONSTR1 by doing the construction loop described in step I

Once STUDENT1 and CONSTR1 are finished, that is, it is finished for circular migration order 1 for this person. Then go to step J to begin work on circular migration order 2.

H. Student loop.
The goal of the student loop is to determine whether the person was a student at any point while gone from origin village. Once it is determined that the person has ever been a student while gone from the village, STUDENTi gets coded 1, and that is it.

Figure 3 might help provide an overview. It is for a person whose migration began at age k+1 and ended at age k+6. For each age, we know from question 5.6 whether the person was a student at that age. Then there are up to six residence locations where the person might have been a student. If there is an age where the person was not in the village, in this case ages k+2,

k+3, k+4, and k+5, if the person was a student (question 5.6) we know he/she was a student outside the origin village, and hence STUDENTi gets coded 1, and that is it. However, if the person was a student in the age beginning or ending the circular interval, in this case k+1 and k+6, then it is necessary to go to the residence of student variables to see if the person was a student while outside the village.

What is considered a student? Codes 1-20 are considered to be student. Code 0 indicates the person never studied. In the life history data there should not be any code 0's. If there are we will need to decide what to do with them. 94, religious studies, is NOT a student. 97, non formal education, is NOT a student. 97 stands for kindergarten, and there should not be any. If there are we will need to decide what to do with them. SAS missing and 99, missing/don't know, are to be treated as NOT being a student. For codes 94, 97 and 99 can we get a sense of how common they occur. Our expectation is that they are relatively rare.

If TYPEi= 3 or 4, code STUDENTi= SAS missing

Starting with BEGAGEi and continuing until ENDAGEi, check X5_6ED. If X5_6ED= any of {1 through 20} then check to see if X5_1AGE (the age in that row) = either BEGAGEi or ENDAGEi. If it does NOT, that means it is an age for which the person was gone the entire age, hence STUDENTi=1. If it is either BEGAGEi or ENDAGEi, then need to check whether the schooling was in the origin village or not, as follows.

For BEGAGEi, need to determine whether the schooling occurred after the person left the village (and if a short move, before the person returned to the village). There are up to six residences indicating where the person was while being a student, X5_6EDR1 through X5_6EDR6. This needs to be checked, but it appears that SAS missing is used if the person was not a student while residing in residence"i". Also, this needs to be checked, but it appears that X5_6EDR1 through X5_6EDR6 line up with X5_R1 through X5_R6. Recall from step D, that the first instance where X5_5Ri NOT equal to 2xxxx, SAS missing or 99999999 was used to define the beginning of this interval. The "i" in this X5_5Ri is the indicator for the "i" to be checked in X5_6EDR1 through X5_6EDR6. (For the purposes of the next paragraph, let this "i' be denoted as v.) If the code for this X5_6EDRi begins with 3 through 10, then code STUDENTi=1. If not continue checking the X5_EDRi+1, 2, 3, 4, 5, but make sure to stop if the person moved back to the village during this age. See steps E and F above.

For ENDAGEi need to determine whether schooling occurred before the person returned to the village. Note that this is only a relevant concern if TYPEi=1. The location of the return to the village in the X5_6Ri series was found in step E or F above. Let this be location z. If ENDAGEi not equal to BEGAGEi, then check all the X5_6EDR1 through X5_6EDRi that precede the return to the village, that is for all i less than z. If the code for any of these X5_6EDRi begins with 3 through 10, then code STUDENTi=1. For those cases where ENDAGEi=BEGAGEi, check all the X5_6EDR1 through X5_6EDRi for i greater than v and less than z. If the code for any of these X5_6EDRi begins with 3 through 10, then code STUDENTi=1.

I. Construction loop

The goal of the construction loop is to determine whether, while away from the origin village, the person was always in construction. If always in construction, then CONSTRi= 1. Otherwise, CONSTRi=0. The first time a non-construction job is found, CONSTRi=0, and that is it for this person/circular migration interval.

For each row (year/age) in the life history, a person can have up to three occupations in each residence and up to six residences. Thus for each age/year there are 18 possible occupations. For the three possible occupations in any given residence we can not tell whether they held the occupations simultaneously or sequentially.

Occupation information is found in X5_7Jp_q, where p indicates the job in any given residence running from 1 to 3 and q indicates the residence in that year running from 1 to 6. Construction is code 5. Codes 1,2,3,4,6,7, and 9 are considered not construction. Note that a legitimate not applicable (SAS missing) does not count as "not construction." There is a code 8, "Monk, student, military service or job unchanged." Before we can finalize instructions for the 2000 construction loop, we need to better understand what this code 8 means. Clearly, monk, student or military service would not count as construction. But "job unchanged" could mean that they continued in construction work. The first step will be to see how common code 8 is in the data. Then we need to look at the pattern of code 8 relative to other information in the life history. From X5_2 we know if they were a monk that year. From X5_3 we know if they were in the military that year. From X5_6ED we know if they were a student that year. Perhaps the best thing to do is dump some records and see what patterns emerge.

As described in the student loop above, location v is the beginning of the circular migration order i (put differently, it is the first stop outside the village), and location z is the end (the last stop outside the village) if it is a TYPEi=1 circular migration. If it is a TYPEi=2 circular migration, then z is the end of the life history that was censored by the interview. Note that if the circular migration has but one non-origin village stop, then v = z.

The following will need to be modified once we understand what code 8 means.

Start with the X5_7J1_q variable associated with location v, the beginning of the circular migration. If this X5_7J1_q variable is coded 1,2, 3, 4, 6, 7 or 9, then this is not construction, and CONSTRi should be coded 0 and that is the end of the construction loop. If this X5_7J1_q variable is coded 5, then the person is in construction and it is necessary to check the next occupation in location v. If the next X5_7J1_q variable in location v is coded 1,2, 3, 4, 6, 7 or 9, then this is not construction, and CONSTRi should be coded 0 and that is the end of the construction loop. If this next X5_7J1_q variable is coded 5, then the person is in construction and it is necessary to check the third occupation in location v. If still construction, then go to location v+1 checking for construction vs non-construction. Then continue to location v+2, and so forth until location z is reached. Location z is the return to the origin village. Check all the occupations in location z-1. If they are still all construction, code 5, then CONSTRi is to be

coded 1. (If TYPEi=2 (open interval) then check until you run out of places to check.)

J. The work on circular migration order 2 is identical to that on circular migration order 1 with two exceptions. First, i will be 2 instead of 1. Second, there is no need to check whether they are in the village. The fact that circular migration order 1 ended in the village means that circular migration order 2 will start in the village.

TWOSOU2=TWOSOU1

If we are still in the 2000 data, then PROXY2=PROXY1

BEGAGE2=ENDAGE1

SOURCEB2=SOURCEE1.

Then follow the logic at step D and beyond. Proceed with circular migration order 3, and so forth as illustrated in the hand drawn figure 2.

SINCE WE HAVE AGE AT EACH STOP AND DETAILS AT THE STOP, I'M ASSUMING THAT IT WOULD BE RELATIVELY "EASY" FOR ANY MULTIPLE STOP INTERVAL TO SEE (AT SOME LATER TIME IF SOMEONE'S SUBSTANTIVE INTEREST REQUIRED IT) IF THEY CAME BACK TO NANG RONG DISTRICT RATHER THAN TO THE ORIGIN VILLAGE. IS THIS RIGHT? RICK, IF THIS QUESTION IS UNCLEAR, DO NOT SPEND MUCH TIME ON IT. I CAN DRAW A PICTURE THAT MAKES IT VERY CLEAR. NOTE HOWEVER, THAT THE STUDENT VARIABLE WOULD NOT BE CORRECT BECAUSE IT IS AN "EVER" VARIABLE. THE CONSTRUCTION VARIABLE WOULD BE FINE BECAUSE IT IS AN "ALWAYS" VARIABLE.

<u>Frequencies</u>

For TYPEi, produce a frequency distribution separately for each "i"

For AGERESij, produce a frequency distribution separately for each "i" and "j" combination.

For DURi, STUDENTi, CONSTRi, PROXYi, SOURCEBi, TWOSOUi produce a frequency distribution for each "i"

Finally frequency distributions for LEFTCEN and VILAT13