

WHY TIME DEFICITS MATTER: APPENDICES

Ajit Zacharias, Rania Antonopoulos, and Thomas Masterson

July 2012



*Empowered lives.
Resilient nations.*





Appendix A Statistical Matching

Introduction

This appendix describes the construction of synthetic datasets created for use in estimation of the Levy Institute Measure of Time and Income Poverty (LIMTIP) for Argentina, Chile, and Mexico. This work was carried out for a project supported by the United Nations Development Programme and the International Labour Organization to develop alternatives to conventional income poverty thresholds. Construction of LIMTIP estimates requires a variety of information for households. In addition to basic demographics, the estimation process requires information about income and time use. No single data set has all the required data for any of the three countries. Thus, in order to produce LIMTIP estimates, a synthetic datafile is created for each country by combining two source data sets with statistical matching.¹ For Argentina, we use the Dirección General de Estadística y Censos' Encuesta Annual de Hogares (EAH) as the base data set, since it contains good information on demographics, income, transfers and taxes for a representative sample of households in Buenos Aires. Time use data comes from the DGEC's time use module of the same survey (UT). For Chile, we use the Encuesta de Caracterización Socioeconómica Nacional (CASEN) as the base data set since it contains good information on demographics, income, transfers and taxes for a representative sample of households in Chile. For the time use data, we use the Encuesta Experimental sobre Uso del Tiempo en el Gran Santiago (EUT), which covers only Greater Santiago de Chile. For Mexico, we use the Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) as the base data set, matching it with the Encuesta Nacional sobre Uso del Tiempo (ENUT), which includes rich time use data. Both of the Mexican data sets cover the entire country.

This appendix is organized as follows. Each section of the appendix details the statistical match for each country in turn. The source datasets are described and their demographic characteristics are compared. Then the quality of the match is reviewed for each.

Argentina

Data and Alignment

The source data sets for the time use match for the LIMTIP estimates for Argentina are the 2005 EAH and the 2005 UT. Note that our study is confined to the city of Buenos Aires, since that is the coverage of the EAH and UT. We use individual records from the 2005 EAH file, excluding those living in group quarters or in the Armed Forces. The EAH has a number of missing values, which we replaced by the

¹ See Kum and Masterson (2010) for details of the statistical matching procedure that we use.

method of multiple imputation with hot-decking.² This results in five replicates for each original record, for a total of 76,260, representing 2,696,930 individuals in Buenos Aires. Since the UT covers individuals aged 15 to 74 years old, we discard younger and older individuals from the EAH file. An additional detail concerns the sampling done on the UT: only one individual in each household is surveyed. Therefore, we will have now good way to compare household total household production, other than at the aggregate level. This leaves 57,335 records, which represents 2,044,405 individuals when weighted.

In order to create the estimates of the time-income poverty measure, we had to construct thresholds for the time spent on household production. The thresholds are defined for the household. The reference group in constructing the thresholds consists of households with at least one nonemployed adult and income around the official income poverty line. We divided the reference group into 12 subgroups based on the number of children (0, 1, 2 and 3 or more) and number of adults (1, 2 and 3 or more) for calculating the thresholds. The thresholds are simply the average values of the time spent on household production by households, differentiated by the number of adults and children. In principle, they represent the average amount of household production that is required to subsist at the poverty level of income.

For practical purposes, we defined the reference group as households with household incomes between 75% and 150% of the poverty line (this range is referred to as the poverty band hereafter), and with at least one non-employed adult. In order to transfer the hours spent by individuals on household production in the reference group as closely as possible, we used the following strata variables in the match: indicators for being within the poverty band, for having one or more non-employed adults in the household, the number of children, the number of adults, sex, employment status, and household income category. Table 1 compares the distribution of individuals by these variables in the two data sets. Since one survey (time use) was carried out as a subset of the other (income and demographic), we can expect them to be well aligned. We see that there are 3% fewer individuals in households without children in the time use sample than in the overall survey. Individuals in three-adult households make up a greater share of those in the time use survey, while those in one-adult and five-or-more adult households are more common in the EAH. About 3% fewer individuals are in households with at least one non-employed adult in the time use survey, compared to the overall survey, though the difference in the number of individuals in households within the poverty band in the two surveys is less than 1%. The distribution by household income is very close between the two surveys, with all categories differing by 1.7% or less. The nonemployed are under-represented in the UT relative to the EAH (3%). The distribution of individuals by sex is close in the two surveys, with females slightly less common (1.4%) in the EAH than in the UT. So, as expected, we have a very close alignment between the two surveys along all seven strata variables.

Match QC

Turning to the results of the match, we first look to the distribution of matched records by matching round in Table 2. The bulk of the matches, 68.9%, occur in the first round. This is lower than in other time use matches (see, for example, [Masterson 2010](#)), due to the higher than usual number of strata variables used in this match.³ The rest of the records are matched over an additional fourteen rounds,

² The variables with missing values were industry, occupation, educational attainment, as well as all of the income variables. 3,888 of 15,252 records had missing values for one or more of these variables.

³ In a typical time use match (as in [Masterson 2010](#)), five variables are used, yielding a total of 32 matching cells. In this match, using seven strata variables, the number of matching cells in the first round was 575.

with 5.8% receiving no match at all.⁴ Table 3 provides a comparison of the distribution of weekly hours of household production in the UT and the matched file. The tenth percentile is zero, so those ratios are undefined. The remaining percentile ratios are all relatively close, with the ratio of the median to the 25th percentile being exactly equivalent. The Gini coefficient is extremely close, 0.5555 in the matched file, compared to 0.5545 in the UT. Table 4 breaks down the mean and median of the three categories of household production and the total in the matched file and the UT.⁵ We can see that for all four variables the difference in the matched and the source file's mean is small, with the largest difference in procurement, which is 7% (or twenty minutes) lower in the matched file than in the UT. Median total household production is 9% higher in the matched file, but this amounts to only 35 minutes per week.

Examination of the quality of the match within population sub-groups shows generally good results. Figure 1 displays ratios of mean weekly hours of household production between the matched file and the UT for the seven strata variables. For almost all the categories, the average weekly hours in the matched file are within 10% of the UT. The one exception is for males, who have 26% higher weekly hours in the matched file, while females have 10% lower weekly hours. Table 5 has the actual numbers, and we can see that these large percentage differences represent relatively small differences in hours per week. In the case of sex, females have 2.8 fewer hours per week on average in the matched file, while males have 2.7 hours more. Notice that the ratios by category are well reproduced in the matched file. The largest deviation is by sex, as we would expect given the differences in the averages for females and males. The extent to which the match file reproduces the distribution of weekly hours of household production within matching cells is demonstrated in Figure 2.⁶ We can see very little difference between the matched file and the UT. Thus the distribution of household production is well preserved in the matching process, even at this level of detail.

Overall, the quality of the match is very good. It has its limitations, especially in terms of sex of individuals. But the overall distribution is transferred with reasonable accuracy, and the distributions within even small sub-groups, such as one adult with two children, is transferred with good precision.

Chile

Data and Alignment

The source data sets for the time use match for the LIMTIP estimates for Chile are the 2006 CASEN and the 2007 EUT. We use individual records from the 2006 CASEN file, excluding those living in group quarters or in the Armed Forces. Since the EUT covers individuals aged 12 to 98 years old in Gran Santiago only, we discard younger and older individuals, as well as all individuals in areas outside of Gran Santiago from the CASEN file. This leaves a total of 28,039 individual records, representing 4,792,489 individuals when weighted. The EUT, unlike the UT in Argentina, covers all individuals in each household surveyed. Unfortunately, the EUT's only income variable is household income and that is categorical. There are also a small number of non-respondents: 162 of 3,561 households (i.e., under 5 percent nonresponse). Since household income category is a strata variable for this match, we multiply imputed the household income categories, creating five replicates for each record, resulting in 14,410 records representing 4,857,143 individuals.

⁴ The unmatched records are assigned the median values of hours of household production for their original matching cells.

⁵ The three categories are care (child care, elder care, etc.), procurement (shopping, etc.), and core (cooking, cleaning, laundry, etc.).

⁶ For the sake of clarity of the plot, only the number of children and number of adults is used.

The strata variables for this match are identical to those used in the Argentina match described above. Table 6 compares the distribution of individuals by these variables in the two data sets. We see that there are 1.9% fewer individuals in households without children in the time use sample than in the CASEN. Individuals in one- and two-adult households make up a greater share of those in the EUT (2% and 7.2%, respectively), while those in three-, four- and five-or-more adult households are more common in the CASEN. About 2.3% fewer individuals are in households within the poverty band in the time use survey, compared to the overall survey, though the difference in the number of individuals in households with at least one non-employed adult in the two surveys is less than 1%. The distribution of individuals by household income is very different in the two surveys, with 13.6% more individuals in households in the middle category in the EUT than in the CASEN. Clearly this is due to the relatively poor income data in the EUT.⁷ The nonemployed are under-represented in the UT relative to the EAH (3%). The distribution of individuals by sex and employment status is very close in the two surveys, with less than 1% difference between the CASEN and the EUT for both variables. So, we have a very close alignment between the two surveys along five of the seven strata variables. We can expect that the match will be off by number of adults and household income, though only in the former can we readily predict the direction of the bias: with fewer individuals in one and two adult households represented in the CASEN, the household production weekly hours of the matched data set are likely to be lower for those individuals than in the EUT, since many will be matched with individuals in the EUT from households with more adults, who will tend to have lower household production hours.

Match QC

Turning to the results of the match, we first look at the distribution of matched records by matching round in Table 7. The bulk of the matches, 66.8%, occur in the first round. Again, this is lower than in other time use matches, but similar to the result in the Argentina match. The rest of the records are matched over an additional twenty-three rounds, with 2.7% receiving no match at all.⁸ Table 8 provides a comparison of the distribution of weekly hours of household production in the EUT and the matched file. In the EUT, and the matched file, both the 25th and the 10th percentile are zero, so most of the ratios are undefined. The remaining percentile ratios are exactly the same in the matched file as in the EUT. The Gini coefficient is extremely close, 0.6152 in the matched file, compared to 0.6162 in the EUT. Note that the Gini coefficient is higher and the percentile values are lower in the case of Chile than in Argentina. This is due to the inclusion of 12 to 14 year olds, who tend to do less household production work, in the Chilean time use survey. Table 9 breaks down the mean and median of the three categories of household production and the total in the matched file and the EUT. We can see that for all four variables the difference in the matched and the source file's mean is small, less than 2% in all cases (with differences ranging from three to eight minutes per week). The median values are all carried over exactly.

Examination of the quality of the match within population sub-groups shows generally good results. Figure 3 displays ratios of mean weekly hours of household production between the matched file and the EUT for the seven strata variables. For almost all the categories, the average weekly hours in the matched file are within 5% of those in the EUT. The exceptions are for individuals in households with three or more children, who have 5.8% lower weekly hours of household production in the matched file, individuals in households outside of the poverty band, who have 5.3% higher hours in the matched file, and males, who have 11.1% higher weekly hours in the matched file, while females have 5.1% lower weekly hours. Table 10 has the actual numbers, and we can see that these large percentage differences

⁷ Multiply imputing the missing values did not significantly alter the distribution of household income in the EUT.

⁸ As in the case of Argentina, the unmatched records are assigned the median values of hours of household production for their original matching cells.

represent relatively small differences in hours per week. Individuals in households with three or more children have 1.4 hours fewer hours in the matched file than in the EUT, while those in households outside of the poverty band have just less than one hour higher weekly hours. In the case of sex, females have about 90 fewer minutes per week on average in the matched file, while males have 70 minutes more. Notice that the ratios by category are very well reproduced in the matched file. The largest deviation is by sex, as we would expect given the differences in the averages for females and males. The medians are for the most part exactly carried over, but where they are off it is by 3.5 hours. This is due to the nature of the time use data, which was collected for one day and recorded in half hour chunks of time.

The extent to which the match file reproduces the distribution of weekly hours of household production within the reference group is demonstrated in Figure 4.⁹ We can see very little difference between the matched file and the UT. In Table 11, we see the ratios of the average and median household total weekly hours of household production in the matched file to the EUT, for households in the reference group, broken down by number of children and number of adults. We can see that household production is well preserved in the matching process, more so for the more populated cells (two-adult households, especially).

Overall, the quality of the match is very good. The overall distribution is transferred with good accuracy, and the distributions within even small sub-groups, such as one adult with two children, is transferred with good precision.

Mexico

Data and Alignment

The source data sets for the time use match for the LIMTIP estimates for Mexico are the 2008 ENIGH and the 2009 ENUT. We use individual records from the 2008 ENIGH file, excluding those living in group quarters or in the Armed Forces. The ENIGH has a number of missing values, which we replaced by the method of multiple imputation with hot-decking.¹⁰ This results in five replicates for each original record, for a total of 594,635, representing 106,866,209 individuals in Mexico. Since the time use data in the ENUT covers individuals aged 12 years of age and older, we discard younger individuals from the ENIGH file. This leaves 454,910 records, which represents 81,941,246 individuals when weighted. The ENUT file, from a nationally representative sample survey, contains records for all individuals, but time use data only for those twelve years of age and older, like Chile but unlike Argentina. The entire data set contains 57,918 individual records, representing 109,047,562 individuals. Missing values in the base data set were multiply imputed, resulting in five replicates for each individual record.¹¹ This produces 289,590 records. Once those individuals aged 11 or less are dropped, 222,524 records remain, representing 84,679,672 individuals.

⁹ For the sake of clarity of the plot, only the number of children and number of adults is used.

¹⁰ The variables with missing values at the household level were indicators for homeownership, availability of water, type of fuel, number of rooms in the house, type of sewage removal, electricity, type of garbage removal, laundry, sink, water tank, gas tank, and heater. 4,815 of 29,468 household records had one or more missing values. The variables with missing values at the individual level were occupation and usual hours of work. 579 of 118,927 individual records had one or more missing values.

¹¹ The variables with missing values at the household level were indicators for availability of water, type of fuel, number of rooms in the house, type of sewage removal. 1,896 of 30,958 household records had one or more missing values. The variables with missing values at the individual level were marital status, age, wage income, transfer income, and other income. 1,681 of 57,918 individual records had one or more missing values.

The strata variables for this match are identical to those used in the Argentina and Chile matches, with one addition: a geographical indicator for rural households. The latter is never dropped in the matching process, so all matches are within rural and urban segments of the recipient and donor data sets.¹² Table 13 compares the distribution of individuals by these variables in the two data sets. Since both surveys are nationally representative and carried out within one year of each other, we can expect them to be well aligned. About 2.6% more individuals are in households with at least one non-employed adult in the time use survey, than in the income survey, while the difference in the number of individuals in households within the poverty band in the two surveys is 2.3%. The distribution by household income is less close between the two surveys, with households in the ENIGH 4% more likely to be in the second income category. The distribution of individuals by sex is very close in the two surveys, with females only slightly less common (0.3%) in the ENIGH than in the ENUT. There are 3.5% more employed individuals and 4.3% fewer inactive individuals in the ENIGH than in the ENUT. Finally, there is a small difference in the distribution of individuals by rural/urban status (0.7%). So, as expected, we have a very close alignment between the two surveys along almost all eight strata variables.

Match QC

Turning to the results of the match, we first look to the distribution of matched records by matching round in Table 14. While the bulk of the matches, 80.3%, occur in the first round, the remainder of the records required an additional 61 rounds of matching to completely exhaust individual records. Less than 1% of records received no match at all.¹³ This high number of matching rounds is unusual in our experience. The large number is due to the high number of records and strata variables. Table 15 provides a comparison of the distribution of weekly hours of household production in urban and rural areas in the ENUT and the matched file. The percentile ratios are all relatively close, with the ratios for urban areas being slightly higher, and those for rural areas slightly lower in each case. The Gini coefficients are also extremely close, with the urban Gini 0.13 points lower and the rural Gini 0.24 points higher than in the ENUT. Table 16 breaks down the mean and median of the three categories of household production and the total in the matched file and the ENUT.¹⁴ We can see that for all four variables the difference in the matched and the source file's mean and medians are small, with the largest differences in median rural care and total weekly hours of just over 1.5% (about 15 and twenty minutes, respectively) higher in the matched file than in the ENUT.

Examination of the quality of the match within population sub-groups shows generally good results. Figure 5a displays ratios of mean weekly hours of household production between the matched file and the ENUT for the seven strata variables in urban areas. In all cases, the average weekly hours in the matched file are within 5% of the ENUT. In fact, with the exception of the lowest household income category which has 4.8% lower weekly hours in the matched file than in the ENUT, the averages for all sub-groups are within 3% of the ENUT average. The rural results (in Figure 5b) are slightly worse. Individuals in households without a non-employed adult have 7.5% higher weekly hours of household production in the matched file than in the ENUT, employed individuals have 15% higher weekly hours and males have 8.1% higher weekly hours. Table 17 has the actual numbers, and we can see that even the larger percentage differences represent relatively small differences in hours per week. In the case of sex, rural females have 1.4 fewer hours per week on average in the matched file, while rural males have 1 hour more. The rural employed have three hours more household production per week in the

¹² Rural is defined in the Mexican national poverty statistics as in a municipality with fewer than 2,500 residents.

¹³ The unmatched records are assigned the median values of hours of household production for their original matching cells.

¹⁴ The three categories are care (child care, elder care, etc.), procurement (shopping, etc.), and core (cooking, cleaning, laundry, etc.).

matched file, the largest deviation by averages. Notice that the ratios by category are well reproduced in the matched file, even for the categories with the largest average deviation. The largest difference is by sex in rural areas, as we would expect given the differences in the averages for rural females. The medians closely follow the patterns of the averages with similar differences in the rural areas by employment and sex. The extent to which the match file reproduces the distribution of total weekly hours of household production in households in the reference group is demonstrated in Figure 6. We can see there are some differences between the matched file and the ENUT. The upper tails are fatter for each of the cells in the reference groups. The ratios of the average and median household total weekly hours of production for the reference group in the matched file to the ENUT are presented in Table 18. In both rural and urban areas the largest differences are in one-adult households with children. These are the cells in the reference group which are the smallest, meaning they are likeliest to be off substantially, but also make less difference in terms of the target measure. Other cells have somewhat large differences, but most of these are households with three or more adults, which, again, are smaller cells. The largest difference among two-adult households is the average weekly hours for rural households with two adults and two children, among which households have 12% higher weekly hours of household production in the matched file than in the ENUT. Overall, the distribution of household production is well preserved in the matching process, even at this level of detail.

Overall, the quality of the match is very good. It has its limitations, especially in terms of sex of individuals. But the overall distribution is transferred with very good accuracy. And the distributions within even small sub-groups, such as one adult with two children, is transferred with good precision, compared with the other two countries.

References

Kum, Hyunsub, and Thomas Neal Masterson. 2010. Statistical matching using propensity scores: Theory and application to the analysis of the distribution of income and wealth. *Journal of Economic and Social Measurement* 35, no. 3 (January 1): 177-196. doi:[10.3233/JEM-2010-0332](https://doi.org/10.3233/JEM-2010-0332).

Masterson, Thomas. 2010. "Quality of Match for Statistical Matches Used in the 1992 and 2007 LIMEW Estimates for the United States." Levy Economics Institute Working Paper 618 (September). http://www.levyinstitute.org/pubs/wp_618.pdf

Tables

Table 1 Alignment of Strata Variables, Argentina

	EAH 2005	UT 2005	Diff
Number	2,044,405	2,128,750	4.1%
Number of children in household			
0	59.89%	56.79%	-3.1%
1	21.15%	21.21%	0.1%
2	12.86%	14.51%	1.7%
3+	6.10%	7.49%	1.4%
Number of adults in HH			
1	13.00%	12.05%	-1.0%
2	45.59%	45.09%	-0.5%
3	23.29%	26.48%	3.2%
4	13.13%	13.24%	0.1%
5+	4.99%	3.14%	-1.9%
Presence of non-employed persons 15-74 in HH?			
No	25.82%	22.98%	-2.8%
Yes	74.18%	77.02%	2.8%
Within Poverty band			
No	87.44%	86.80%	-0.6%
Yes	12.56%	13.20%	0.6%
Household income category			
<i>LT 1,000</i>	16.92%	18.45%	1.5%
<i>1,000-1,599</i>	22.25%	21.46%	-0.8%
<i>1,600-2,299</i>	20.70%	22.14%	1.4%
<i>2,300-3,499</i>	20.16%	18.45%	-1.7%
<i>GE 3,500</i>	19.97%	19.49%	-0.5%
Employed?			
No	32.51%	35.49%	3.0%
Yes	67.49%	64.51%	-3.0%
Sex of respondent, EAH 2005			
<i>Male</i>	46.32%	44.89%	-1.4%
<i>Female</i>	53.68%	55.11%	1.4%

Table 2 Distribution of Matched Records by Matching Round, Argentina

Round	Number	Percentage
1	1,408,402	68.9%
2	196,206	9.6%
3	57,115	2.8%
4	108,307	5.3%
5	39,850	1.9%
6	13,499	0.7%
7	24,833	1.2%
8	7,668	0.4%
9	4,992	0.2%
10	4,819	0.2%
11	46,477	2.3%
12	5,131	0.3%
13	3,366	0.2%
14	2,160	0.1%
15	2,872	0.1%
16	118,708	5.8%
Total	2,044,405	100.0%

Table 3 Distribution of Weekly Hours of Household Production in UT 2005 and Matched File

	p90/p10	p90/p50	p50/p10	p75/p25	p75/p50	p50/p25	Gini
UT05		4.092		9.000	2.455	3.666	0.555
MATCH		3.952		8.666	2.362	3.669	0.556
MATCH2		3.952		8.834	2.408	3.669	0.562

Table 4 Comparison of Mean and Median Time Use Variables in Matched File to UT 2005

	Mean HH Prod.	Mean Care	Mean Proc.	Mean Core	Median HH Prod.	Median Care	Median Proc.	Median Core
UT05	11.90	3.44	4.91	20.25	6.42	0.00	0.00	12.83
MATCH	11.71	3.32	4.57	19.79	7.00	0.00	0.00	12.84
Ratio	98.40%	96.51%	93.08%	97.73%	109.03%			100.08%

Table 5 Mean and Median Weekly Hours of Household Production by Strata Variable, UT 2005 and Matched File

Mean values of HH Production (Weekly Hours)			
	UT05	Match	Ratio
HH Production	20.25	19.83	97.9%
Core	11.90	11.71	98.4%
Procurement	3.44	3.31	96.2%
Care	4.91	4.66	94.9%
Distribution among population subgroups			
Ratio of Mean Values			
	UT05	Match	
Number of Children			
0	15.62	16.11	103.1%
1	24.14	22.70	94.0%
2	26.00	25.68	98.8%
3+	33.20	33.35	100.5%
Number of Adults			
1	16.54	15.63	94.5%
2	22.44	21.99	98.0%
3+	18.99	18.66	98.3%
Within Poverty Band?			
No	18.92	18.73	99.0%
Yes	29.01	27.12	93.5%
Presence of Non-Employed Adult?			
No	11.46	12.36	107.9%
Yes	22.87	22.37	97.8%
Household Income			
			Over All
LT 1,000	24.05	23.08	96.0%
1,000-1,599	22.22	21.18	95.3%
1,600-2,299	20.62	20.13	97.6%
2,300-3,499	16.91	18.47	109.2%
GE 3,500	17.22	16.41	95.3%
Employed?			
No	27.97	25.96	92.8%
Yes	16.01	16.81	105.0%
Sex			
Female	28.35	25.62	90.4%
Male	10.31	13.03	126.4%

Median values of HH Production (Weekly Hours)						
	UT05	Match	Ratio			
HH Production	12.83	12.84	100.1%			
Core	6.42	7.00	109.0%			
Procurement	-	-				
Care	-	-				
Distribution among population subgroups				Ratio of Median Values		
					UT05	Match
Number of Children						
0	8.75	10.50	120.0%	0/3+	0.32	0.38
1	18.09	16.33	90.3%	1/3+	0.66	0.58
2	19.83	19.25	97.1%	2/3+	0.72	0.69
3+	27.42	28.00	102.1%			
Number of Adults						
1	11.67	11.67	100.0%	2/1	1.20	1.20
2	14.00	14.00	100.0%	3+/1	0.90	1.00
3+	10.50	11.67	111.1%			
Within Poverty Band?						
No	12.25	12.25	100.0%	Yes/No	1.71	1.38
Yes	21.00	16.92	80.6%			
Presence of Non-Employed Adult?						
No	7.00	7.00	100.0%	Yes/No	2.25	2.25
Yes	15.75	15.75	100.0%			
Household Income					Over All	
LT 1,000	15.75	15.75	100.0%	LT 1,000	1.23	1.23
1,000-1,599	15.75	14.00	88.9%	1,000-1,599	1.23	1.09
1,600-2,299	14.58	13.42	92.0%	1,600-2,299	1.14	1.05
2,300-3,499	10.50	11.67	111.1%	2,300-3,499	0.82	0.91
GE 3,500	9.92	10.50	105.8%	GE 3,500	0.77	0.82
Employed?						
No	23.92	18.67	78.1%	Yes/No	0.44	0.56
Yes	10.50	10.50	100.0%			
Sex						
Female	24.50	19.25	78.6%	Female/Male	5.25	2.75
Male	4.67	7.00	149.9%			

Table 6 Alignment of Strata Variables, Chile

	CASEN	EUT	Diff
<i>individuals</i>	4,792,489	4,857,105	1.3%
Number of children in household			
<i>0</i>	33.70%	35.55%	1.9%
<i>1</i>	27.75%	28.27%	0.5%
<i>2</i>	22.75%	22.46%	-0.3%
<i>3</i>	11.18%	9.55%	-1.6%
<i>4+</i>	4.63%	4.17%	-0.5%
Number of adults in household			
<i>1</i>	4.89%	6.87%	2.0%
<i>2</i>	32.80%	40.01%	7.2%
<i>3</i>	28.58%	26.27%	-2.3%
<i>4</i>	19.81%	18.01%	-1.8%
<i>5+</i>	13.92%	8.85%	-5.1%
Is the household within the poverty band?			
<i>No</i>	82.01%	84.27%	2.3%
<i>Yes</i>	17.99%	15.73%	-2.3%
Is there a non-working adult in HH?			
<i>No</i>	26.65%	26.19%	-0.5%
<i>Yes</i>	73.35%	73.81%	0.5%
Household income category			
<i>LE 144,000</i>	6.86%	5.99%	-0.9%
<i>144,001-250,000</i>	11.66%	4.55%	-7.1%
<i>250,001-500,000</i>	28.33%	41.94%	13.6%
<i>500,001-1,000,000</i>	28.33%	27.62%	-0.7%
<i>GT 1,000,000</i>	24.81%	19.90%	-4.9%
Labor force status			
<i>Employed</i>	53.69%	53.85%	0.2%
<i>Unemployed</i>	4.07%	4.46%	0.4%
<i>Inactive</i>	42.24%	41.69%	-0.6%
Sex of respondent			
<i>Female</i>	52.36%	51.94%	-0.4%
<i>Male</i>	47.64%	48.06%	0.4%

Table 7 Distribution of Matched Records by Matching Round, Chile

Matching Round	Records Matched	Percentage	Cumulative Percentage
1	3,199,722	66.8%	66.8
2	232,383	4.8%	71.6
3	76,303	1.6%	73.2
4	442,051	9.2%	82.4
5	12,859	0.3%	82.7
6	64,310	1.3%	84.0
7	12,723	0.3%	84.3
8	20,621	0.4%	84.7
9	20,338	0.4%	85.2
10	144,445	3.0%	88.2
11	40,367	0.8%	89.0
12	20,173	0.4%	89.4
13	2,421	0.1%	89.5
14	114,634	2.4%	91.9
15	16,085	0.3%	92.2
16	102,113	2.1%	94.3
17	13,413	0.3%	94.6
18	7,688	0.2%	94.8
19	13,843	0.3%	95.1
20	6,867	0.1%	95.2
21	46,606	1.0%	96.2
22	20,017	0.4%	96.6
23	7,944	0.2%	96.8
24	26,661	0.6%	97.3
25	127,902	2.7%	100.0
Total	4,792,489		

Table 8 Distribution of Weekly Hours of Household Production in EUT and Matched File

	p90/p10	p90/p50	p50/p10	p75/p25	p75/p50	p50/p25	Gini
EUT07		5.33		.	3.00	.	0.6162
MATCH		5.33		.	3.00	.	0.6152

Table 9 Comparison of Mean and Median Time Use Variables in 2007 EUT and Matched File

	Mean Care	Mean Proc.	Mean Core	Mean HH Prod.	Median Care	Median Proc.	Median Core	Median HH Prod.
EUT	13.16	3.02	3.70	19.88	7.00	0.00	0.00	10.50
MATCH	13.04	2.97	3.64	19.74	7.00	0.00	0.00	10.50
Ratio	99.09%	98.34%	98.38%	99.30%	100.00%			100.00%

Table 10 Mean and Median Household Production Weekly Hours, 2007 EUT and Matched File

Mean values of HH Production (Weekly Hours)						
	UT05	Match	Ratio			
HH Production	19.88	19.74	99.3%			
Core	13.16	13.04	99.1%			
Procurement	3.02	2.97	98.3%			
Care	3.70	3.64	98.4%			
Distribution among population subgroups				Ratio of Mean Values		
					UT05	Match
Number of Children						
0	17.30	17.17	99.2%	0/3+	0.73	0.76
1	19.37	19.54	100.9%	1/3+	0.81	0.87
2	22.18	21.91	98.8%	2/3+	0.93	0.98
3+	23.82	22.45	94.2%			
Number of Adults						
1	21.35	21.90	102.6%	2/1	1.00	0.96
2	21.45	21.03	98.0%	3+/1	0.87	0.84
3+	18.50	18.48	99.9%			
Within Poverty Band?						
No	17.47	18.40	105.3%	Yes/No	1.19	1.10
Yes	20.73	20.22	97.5%			
Presence of Non-Employed Adult?						
No	19.42	19.33	99.5%	Yes/No	1.15	1.12
Yes	22.32	21.59	96.7%			
Household Income					Over All	
LT 1,000	22.67	21.78	96.1%	LT 1,000	1.14	1.10
1,000-1,599	21.99	21.57	98.1%	1,000-1,599	1.11	1.09
1,600-2,299	21.25	21.60	101.6%	1,600-2,299	1.07	1.09
2,300-3,499	18.77	18.64	99.3%	2,300-3,499	0.94	0.94
GE 3,500	17.21	17.45	101.4%	GE 3,500	0.87	0.88
Employed?						
No	25.67	25.02	97.5%	Yes/No	0.58	0.61
Yes	14.92	15.19	101.8%			
Sex						
Male	9.97	11.14	111.7%	Female/Male	0.34	0.40
Female	29.04	27.56	94.9%			

Median values of HH Production (Weekly Hours)						
	UT05	Match	Ratio			
HH Production	10.50	10.50	100.0%			
Core	7.00	7.00	100.0%			
Procurement	-	-				
Care	-	-				
Distribution among population subgroups				Ratio of Median Values		
					UT05	Match
Number of Children						
0	10.50	7.00	66.7%	0/3+	1.00	0.67
1	10.50	10.50	100.0%	1/3+	1.00	1.00
2	14.00	14.00	100.0%	2/3+	1.33	1.33
3+	10.50	10.50	100.0%			
Number of Adults						
1	14.00	17.50	125.0%	2/1	0.75	0.60
2	10.50	10.50	100.0%	3+/1	0.50	0.40
3+	7.00	7.00	100.0%			
Within Poverty Band?						
No	10.50	10.50	100.0%	Yes/No	1.00	1.00
Yes	10.50	10.50	100.0%			
Presence of Non-Employed Adult?						
No	10.50	10.50	100.0%	Yes/No	1.33	1.33
Yes	14.00	14.00	100.0%			
Household Income					Over All	
LT 1,000	17.50	17.50	100.0%	LT 1,000	1.67	1.67
1,000-1,599	17.50	14.00	80.0%	1,000-1,599	1.67	1.33
1,600-2,299	14.00	14.00	100.0%	1,600-2,299	1.33	1.33
2,300-3,499	7.00	7.00	100.0%	2,300-3,499	0.67	0.67
GE 3,500	7.00	10.50	150.0%	GE 3,500	0.67	1.00
Employed?						
No	17.50	17.50	100.0%	Yes/No	0.20	0.40
Yes	3.50	7.00	200.0%			
Sex						
Male	0.00	0.00		Female/Male	0.00	0.00
Female	24.50	21.00	85.7%			

**Table 11 Household Production Weekly Hours for Households in Reference Group, Ratio of Matched File to EUT
2007**

Number of Adults		Number of Children				
		0	1	2	3+	All
1	Mean	82.8%	57.1%	81.0%	107.8%	75.5%
	Median	90.9%	77.8%	0.0%	90.2%	80.0%
2	Mean	87.8%	75.0%	83.7%	97.8%	84.4%
	Median	80.0%	94.1%	70.0%	73.9%	78.9%
3+	Mean	95.5%	94.3%	105.3%	86.1%	94.4%
	Median	94.4%	95.7%	126.1%	78.6%	100.0%

Table 12 Alignment of Strata Variables, Mexico

	ENUT 2009	ENIGH 2008	diff
<i>individuals</i>	84,676,629	81,945,375	-3.2%
Number of children in household			
0	45.2%	43.2%	-1.9%
1	19.5%	18.8%	-0.7%
2	19.5%	19.1%	-0.5%
3	10.1%	11.8%	1.7%
4+	5.8%	7.1%	1.4%
Number of adults in household			
1	5.9%	5.6%	-0.3%
2	35.4%	36.5%	1.1%
3	19.7%	19.3%	-0.4%
4	14.5%	14.6%	0.1%
5	9.6%	8.4%	-1.3%
6	6.0%	6.4%	0.4%
7	3.7%	3.7%	0.0%
8	2.4%	2.8%	0.4%
9	1.4%	1.3%	-0.1%
10+	1.4%	1.5%	0.0%
Presence of non-employed adult in household?			
No	22.1%	24.7%	2.6%
Yes	77.9%	75.3%	-2.6%
Within poverty band for time use estimation?			
No	68.9%	66.7%	-2.3%
Yes	31.1%	33.3%	2.3%
Household income category			
<i>Less than 2500</i>	14.7%	12.1%	-2.6%
<i>2500 to 4999</i>	16.4%	20.4%	4.0%
<i>5000 to 7499</i>	16.2%	17.2%	1.1%
<i>7500 to 14999</i>	28.4%	27.6%	-0.8%
<i>15000 or more</i>	24.4%	22.7%	-1.7%
Sex of respondent			
<i>Female</i>	52.6%	52.3%	-0.3%
<i>Male</i>	47.5%	47.7%	0.3%
Labor force status			
<i>Employed</i>	49.6%	53.1%	3.5%
<i>Unemployed</i>	1.8%	2.6%	0.8%
<i>Inactive</i>	48.6%	44.3%	-4.3%
Rural household?			
No	79.5%	78.8%	-0.7%
Yes	20.5%	21.2%	0.7%

Table 13 Distribution of Matched Records by Matching Round, Mexico Match

Matching Round	Records Matched	Percentage	Cumulative Percentage	Matching Round	Records Matched	Percentage	Cumulative Percentage
1	65,795,628	80.3%	80.3%	33	6,944	0.0%	97.0%
2	2,983,564	3.6%	83.9%	34	9,931	0.0%	97.0%
3	1,023,700	1.2%	85.2%	35	28,308	0.0%	97.0%
4	4,590,113	5.6%	90.8%	36	228,473	0.3%	97.3%
5	156,005	0.2%	91.0%	37	31,713	0.0%	97.4%
6	496,904	0.6%	91.6%	38	34,885	0.0%	97.4%
7	63,297	0.1%	91.7%	39	12,074	0.0%	97.4%
8	241,519	0.3%	92.0%	40	103,770	0.1%	97.5%
9	675,121	0.8%	92.8%	41	75,976	0.1%	97.6%
10	342,831	0.4%	93.2%	42	19,539	0.0%	97.7%
11	90,381	0.1%	93.3%	43	12,484	0.0%	97.7%
12	233,351	0.3%	93.6%	44	30,291	0.0%	97.7%
13	34,068	0.0%	93.6%	45	55,343	0.1%	97.8%
14	22,281	0.0%	93.7%	46	832	0.0%	97.8%
15	10,257	0.0%	93.7%	47	33	0.0%	97.8%
16	10,854	0.0%	93.7%	48	3,152	0.0%	97.8%
17	1,307,221	1.6%	95.3%	49	2,768	0.0%	97.8%
18	82,304	0.1%	95.4%	50	79,576	0.1%	97.9%
19	52,877	0.1%	95.4%	51	42,172	0.1%	97.9%
20	1,758	0.0%	95.4%	52	42,826	0.1%	98.0%
21	25,433	0.0%	95.5%	53	171,961	0.2%	98.2%
22	32,047	0.0%	95.5%	54	6,968	0.0%	98.2%
23	3,765	0.0%	95.5%	55	41,902	0.1%	98.3%
24	109,684	0.1%	95.7%	56	25,595	0.0%	98.3%
25	4,643	0.0%	95.7%	57	38,067	0.0%	98.3%
26	36,383	0.0%	95.7%	58	93,886	0.1%	98.4%
27	50,289	0.1%	95.8%	59	5,103	0.0%	98.5%
28	210,461	0.3%	96.0%	60	114,463	0.1%	98.6%
29	4,629	0.0%	96.0%	61	473,449	0.6%	99.2%
30	177,398	0.2%	96.2%	62	358,417	0.4%	99.6%
31	136,605	0.2%	96.4%	63	322,389	0.4%	100.0%
32	466,714	0.6%	97.0%	Total	81,945,375		

Table 14 Distribution of Household Production in 2009 ENUT and Matched File

		p90/p10	p90/p50	p50/p10	p75/p25	p75/p50	p50/p25	Gini
ENUT09	Urban	21.848	3.538	6.175	5.959	2.296	2.596	0.4935
	Rural	22.933	3.161	7.255	6.276	2.250	2.790	0.4762
Match	Urban	22.108	3.557	6.215	6.017	2.307	2.608	0.4948
	Rural	22.444	3.135	7.159	6.061	2.201	2.754	0.4739

Table 15 Comparison of Mean and Median Time Use Variables in 2009 ENUT and Matched File

		Mean Care	Mean Proc.	Mean Core	Mean Personal	Mean HH Prod.	Median Care	Median Proc.	Median Core	Median Personal	Median HH Prod.
ENUT09	Urban	17.93	7.59	1.50	75.31	27.02	11.21	0.00	0.89	73.11	17.26
ENUT09	Rural	23.18	8.17	1.19	81.67	32.54	15.20	0.00	0.00	78.84	21.58
MATCH	Urban	17.88	7.65	1.49	75.15	27.02	11.13	0.00	0.89	73.02	17.22
MATCH	Rural	23.08	7.99	1.20	81.80	32.34	15.44	0.00	0.00	79.00	21.91
Ratio	Urban	99.72%	100.79%	99.33%	99.79%	100.00%	99.29%		100.00%	99.88%	99.77%
Ratio	Rural	99.57%	97.80%	100.84%	100.16%	99.39%	101.58%			100.20%	101.53%

Table 16 Urban Mean and Median Weekly Hours of Household Production, 2009 ENUT and Matched File

Mean values of HH Production						
Urban	ENUT09	Match	Ratio			
HH Production	27.0	27.0	100.0%			
Core	17.9	17.9	99.7%			
Care	7.6	7.7	100.8%			
Procurement	1.5	1.5	99.3%			
Personal	75.3	75.2	99.8%			
Distribution among population subgroups				Ratio of Mean Values		
					ENUT09	Match
Number of Children						
0	24.33	24.34	100.0%	0/3+	0.77	0.79
1	27.30	27.21	99.7%	1/3+	0.87	0.88
2	30.13	29.92	99.3%	2/3+	0.96	0.97
3+	31.48	30.90	98.2%			
Number of Adults						
1	26.76	27.34	102.2%	2/1	1.15	1.12
2	30.72	30.68	99.9%	3+/1	0.93	0.91
3+	24.95	24.75	99.2%			
Presence of Non-Employed Adult?						
No	25.00	25.43	101.7%	Yes/No	1.11	1.08
Yes	27.67	27.59	99.7%			
Within Poverty Band?						
No	26.39	26.47	100.3%	Yes/No	1.08	1.06
Yes	28.40	28.08	98.9%			
Household Income					Over All	
LT 1,000	29.98	28.54	95.2%	LT 1,000	1.11	1.06
1,000-1,599	30.04	29.84	99.3%	1,000-1,599	1.11	1.10
1,600-2,299	29.90	29.62	99.1%	1,600-2,299	1.11	1.10
2,300-3,499	26.61	26.39	99.2%	2,300-3,499	0.98	0.98
GE 3,500	23.26	23.83	102.5%	GE 3,500	0.86	0.88
Employed?						
No	33.35	33.78	101.3%	Yes/No	0.63	0.63
Yes	21.08	21.33	101.2%			
Sex						
Female	39.81	39.67	99.6%	Female/Male	3.13	3.03
Male	12.72	13.10	103.0%			

Table 17 Rural Mean and Median Weekly Hours of Household Production, 2009 ENUT and Matched File

Mean values of HH Production						
Rural	ENUT09	Match	Ratio			
HH Production	32.5	32.3	99.4%			
Core	23.2	23.1	99.6%			
Care	8.2	8.0	97.8%			
Procurement	1.2	1.2	100.8%			
Personal	81.7	81.8	100.2%			
Distribution among population subgroups				Ratio of Mean Values		
					ENUT09	Match
Number of Children						
0	30.53	30.59	100.2%	0/3+	0.88	0.90
1	32.39	32.13	99.2%	1/3+	0.94	0.95
2	33.88	33.35	98.4%	2/3+	0.98	0.98
3+	34.63	33.90	97.9%			
Number of Adults						
1	34.50	35.03	101.5%	2/1	1.05	1.03
2	36.09	36.00	99.8%	3+/1	0.86	0.84
3+	29.59	29.51	99.7%			
Presence of Non-Employed Adult?						
No	28.93	31.09	107.5%	Yes/No	1.14	1.05
Yes	33.09	32.63	98.6%			
Within Poverty Band?						
No	32.61	32.34	99.2%	Yes/No	0.99	1.00
Yes	32.40	32.34	99.8%			
Household Income					Over All	
LT 1,000	35.48	34.31	96.7%	LT 1,000	1.09	1.06
1,000-1,599	33.86	33.60	99.2%	1,000-1,599	1.04	1.04
1,600-2,299	31.82	31.37	98.6%	1,600-2,299	0.98	0.97
2,300-3,499	29.14	28.48	97.7%	2,300-3,499	0.90	0.88
GE 3,500	27.48	28.81	104.8%	GE 3,500	0.84	0.89
Employed?						
No	41.93	41.99	100.1%	Yes/No	0.46	0.53
Yes	19.37	22.30	115.1%			
Sex						
Female	50.87	49.44	97.2%	Female/Male	3.94	3.54
Male	12.90	13.95	108.1%			

Median values of HH Production			
Rural	ENUT09	Match	Ratio
HH Production	21.6	21.9	101.5%
Core	15.2	15.4	101.6%
Care	0.0	0.0	#DIV/0!
Procurement	0.0	0.0	#DIV/0!
Personal	78.8	79.0	100.2%
Distribution among population subgroups			
Ratio of Mean Values			
	ENUT09	Match	
Number of Children			
0	23.19	23.48	101.3%
1	20.98	21.04	100.3%
2	20.70	20.31	98.1%
3+	20.54	21.44	104.4%
Number of Adults			
1	28.12	28.77	102.3%
2	25.46	25.46	100.0%
3+	18.40	19.10	103.8%
Presence of Non-Employed Adult?			
No	22.03	23.98	108.9%
Yes	21.48	21.21	98.7%
Within Poverty Band?			
No	22.30	22.25	99.8%
Yes	20.27	21.21	104.6%
Household Income			
LT 1,000	27.83	26.30	94.5%
1,000-1,599	22.54	22.72	100.8%
1,600-2,299	19.14	19.67	102.8%
2,300-3,499	17.37	16.65	95.9%
GE 3,500	16.27	17.47	107.4%
Employed?			
No	39.72	39.86	100.4%
Yes	11.64	13.26	113.9%
Sex			
Female	51.39	49.34	96.0%
Male	8.44	9.05	107.2%

**Table 18 Household Production Weekly Hours for Households in Reference Group, Ratio of Matched File to 2009
ENUT**

			Number of Children				
	Number of Adults		0	1	2	3+	All
Urban	1	Mean	100.4%	69.7%	74.0%	84.9%	100.9%
		Median	100.0%	25.7%	59.2%	70.2%	94.6%
	2	Mean	102.5%	100.8%	102.0%	97.6%	98.9%
		Median	101.2%	98.6%	100.2%	93.0%	93.9%
	3+	Mean	104.1%	109.1%	105.0%	107.0%	104.7%
		Median	107.5%	116.0%	99.1%	106.2%	106.7%
	All	Mean	103.9%	104.3%	100.6%	100.0%	101.9%
		Median	103.4%	103.8%	98.1%	95.8%	98.7%
Rural	1	Mean	90.8%	92.8%	66.2%	91.0%	85.8%
		Median	94.8%	106.1%	73.9%	83.7%	92.2%
	2	Mean	95.7%	99.3%	112.1%	110.2%	102.3%
		Median	99.9%	98.8%	101.8%	96.2%	92.1%
	3+	Mean	105.4%	110.3%	109.4%	103.3%	107.0%
		Median	105.2%	118.8%	105.7%	104.4%	106.4%
	All	Mean	99.4%	103.3%	107.7%	108.4%	103.3%
		Median	99.1%	100.3%	99.8%	101.9%	96.9%

Figures

Figure 1 Ratio of Mean HH Production by Category (Match/UT 2005)

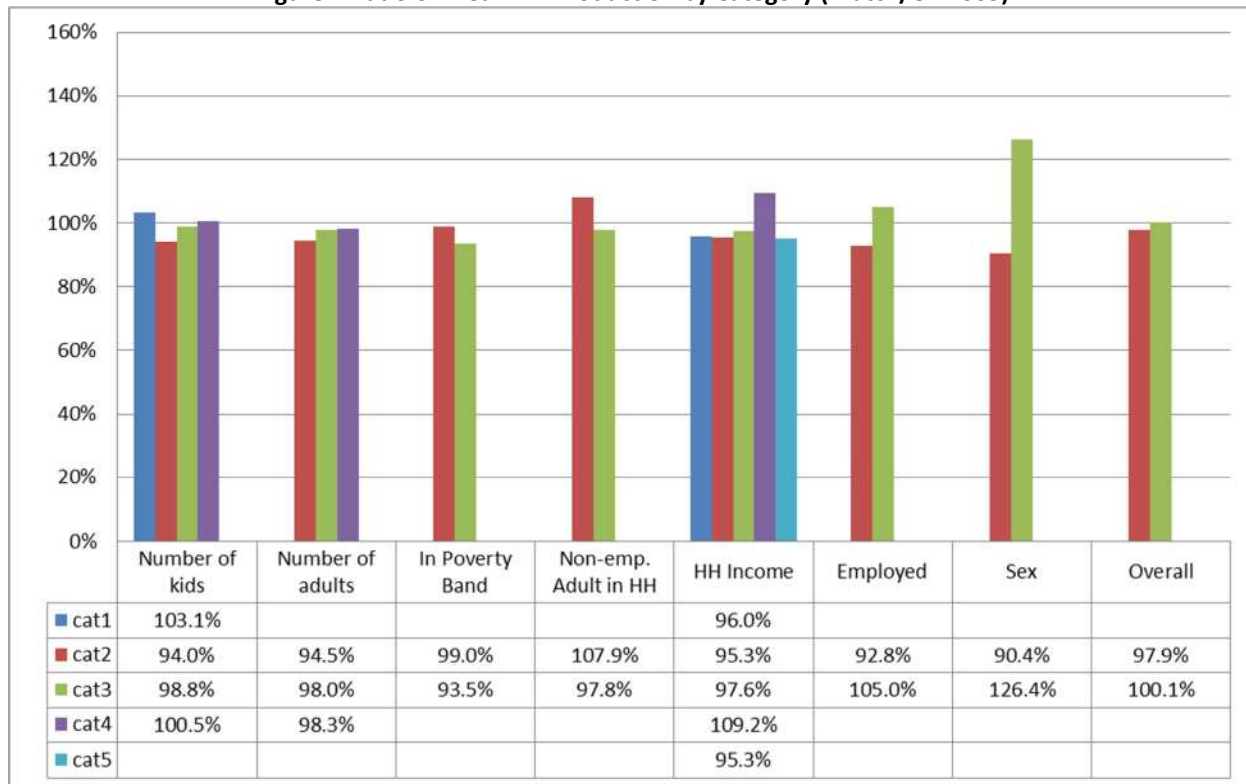


Figure 2 Household Production by Matching Cells, UT 2005 and Matched File

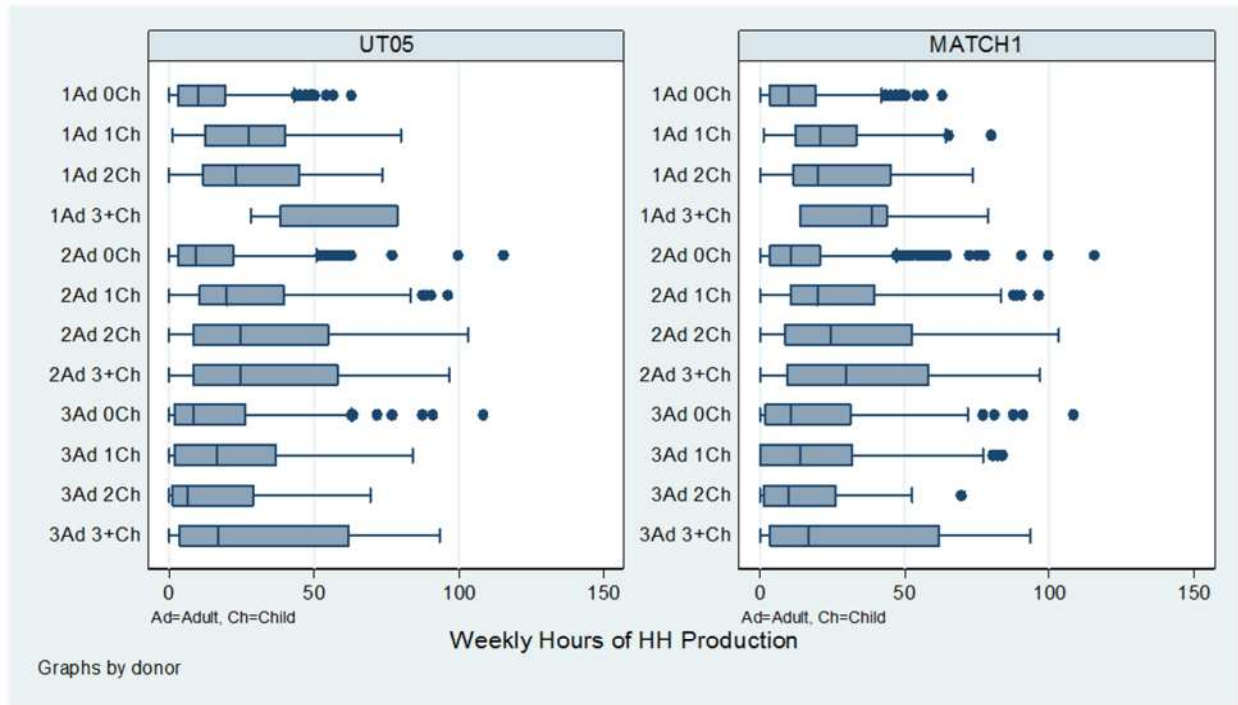


Figure 3 Ratio of Mean HH Production by Category (Match/EUT 2007)

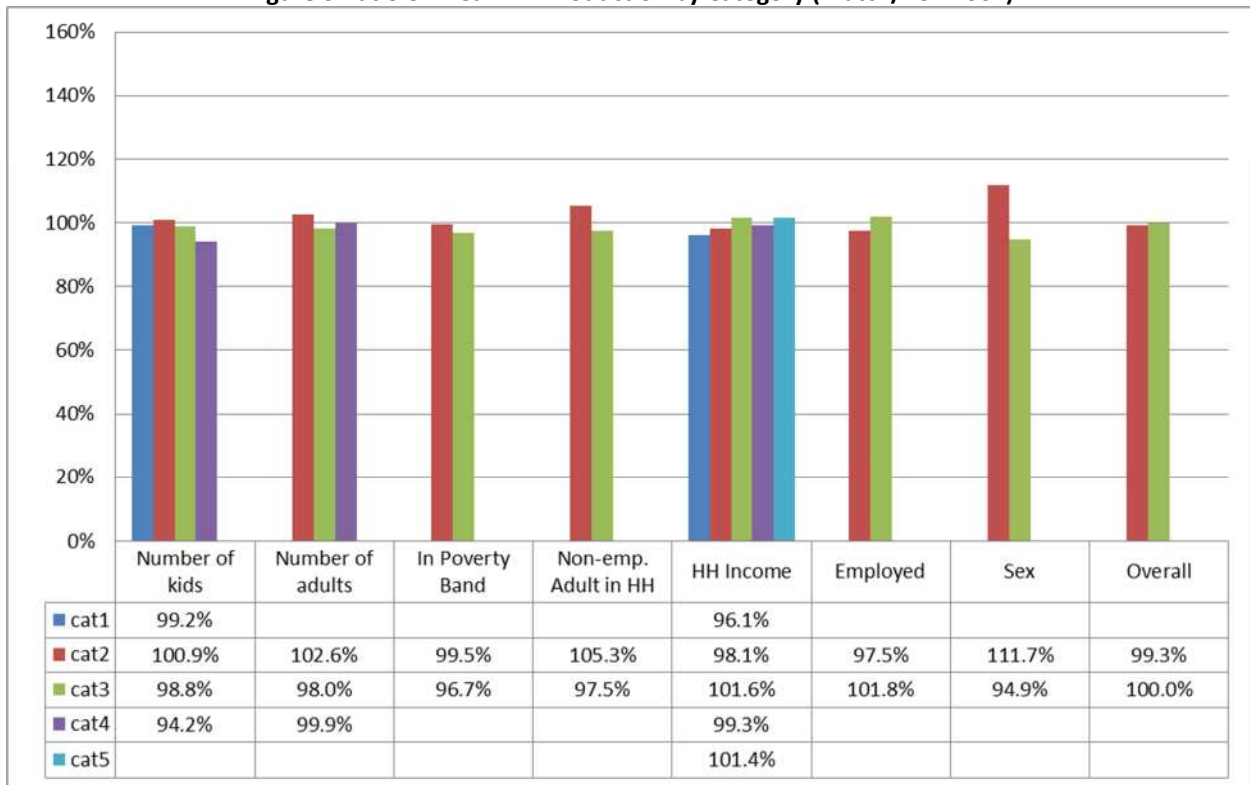


Figure 4 Household Production by Matching Cells, EUT 2007 and Matched File

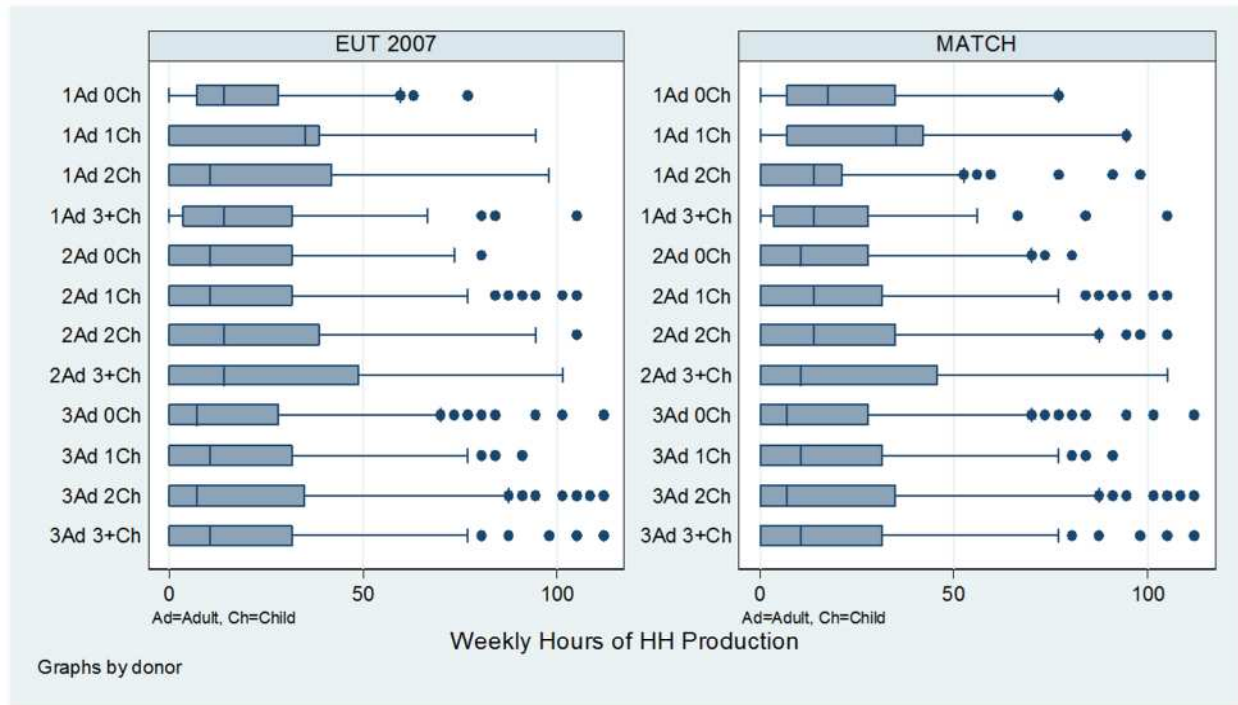
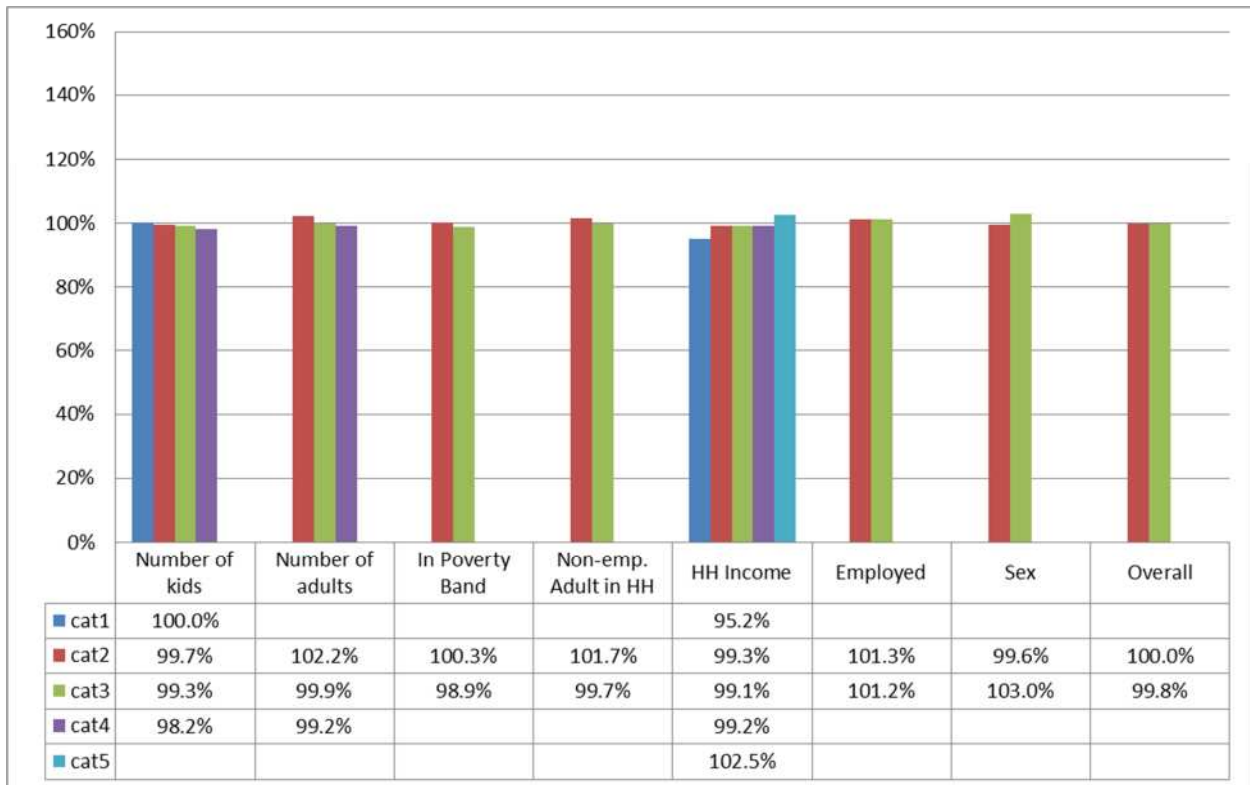


Figure 5 Ratio of Mean HH Production by Category (Match/ENUT 2009)

a. Urban



b. Rural

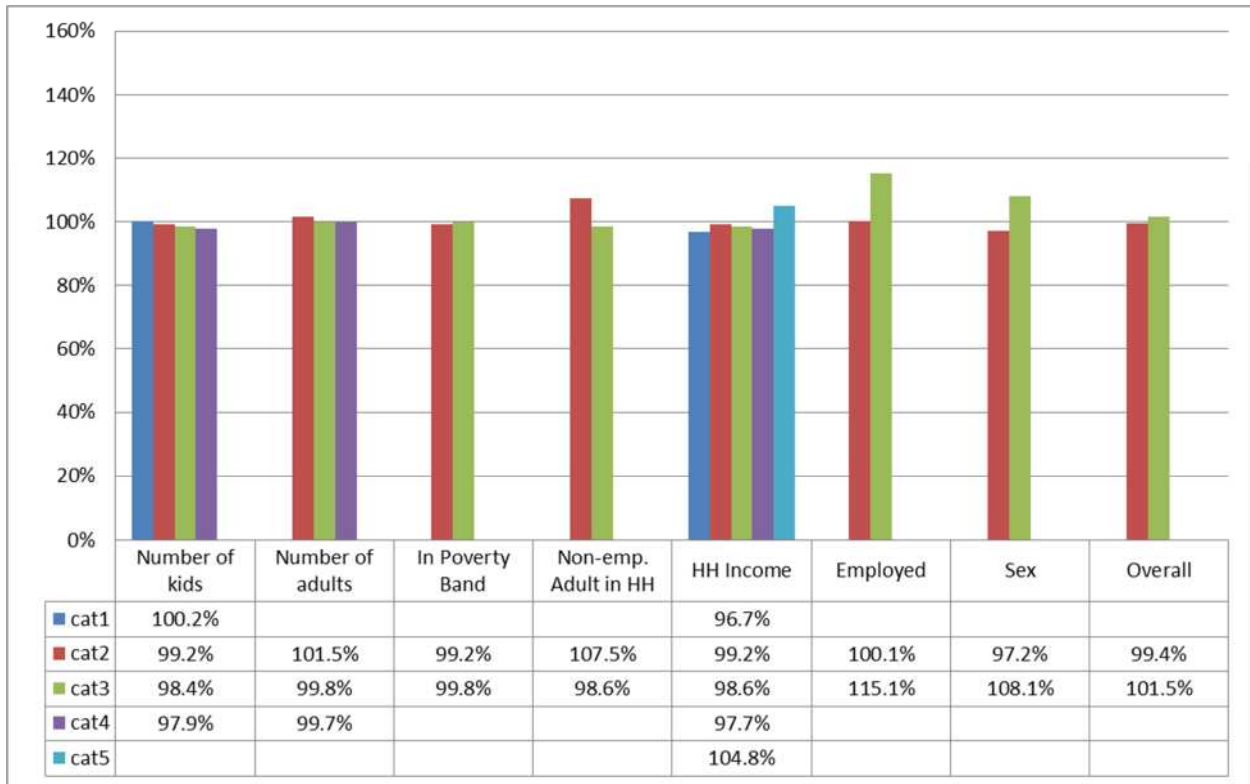
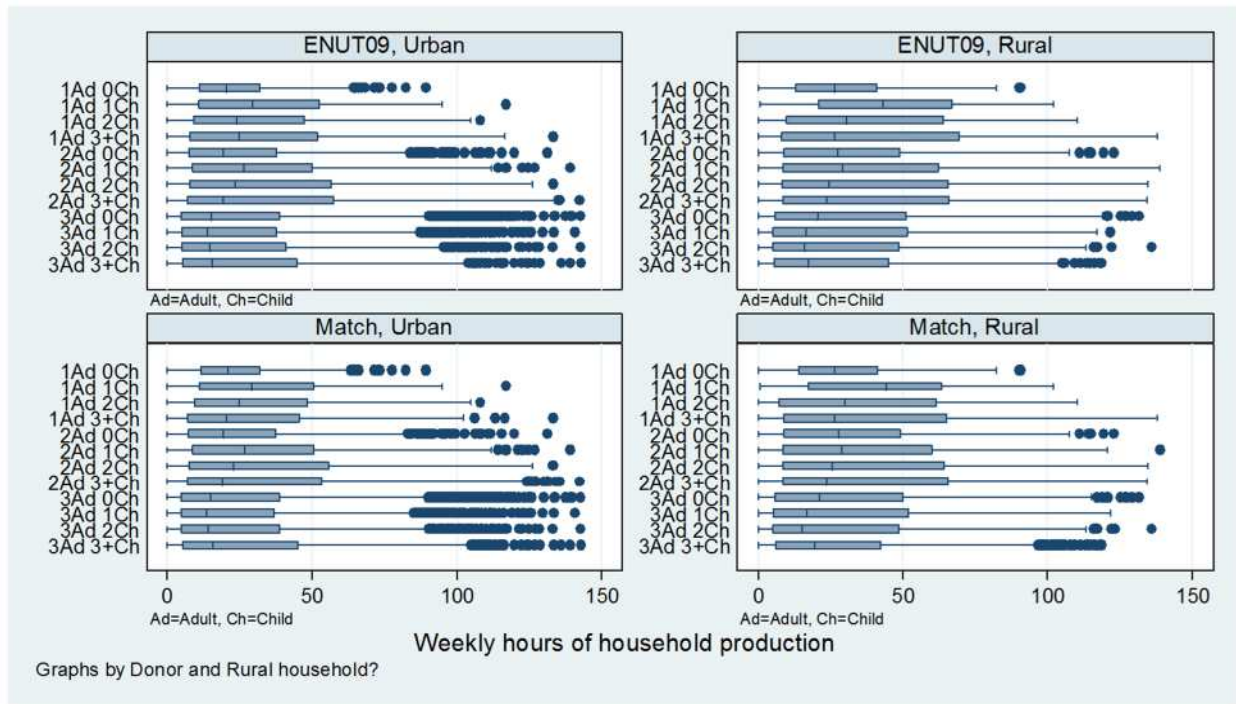


Figure 6 Household Production by Matching Cells, ENUT 2009 and Matched File



Appendix B Simulation of full-time employment and household work

The purpose of this appendix is to describe the microsimulation of full-time employment that was described in Section 1.2.4. We also present an assessment of the simulation results. We draw on and extend our work simulating the results of the American Recovery and Reinvestment Act ([Zacharias, Masterson, and Kim 2009](#)). The nature of this project is quite different than the prior study, which estimated the impact of a fiscal stimulus plan that aimed to increase employment generally. In this case, we assume that some unspecified way is found to employ those who are not employed full-time. We then assess the impact this change has on income and time poverty.

This paper is organized in the following way. First we describe the methodology involved in the imputation of occupation and industry, hours of employment and earnings, household income, and household production hours. Next we consider the possibilities for assessing the result of the imputation. Finally we assess the results, given the limitations outlined in the prior section.

Methodology

The data used in the simulations were created for each country via statistical matching of time use survey and income survey. We have described the procedures followed in the creation of the synthetic datafiles elsewhere ([Masterson 2011](#); [Masterson 2012](#)). The purpose of the simulations in the LIMTIP project is to assess the first order impacts of policies aimed at alleviating poverty via jobs policies, for example an employer of last resort (ELR) policy. As such, the simulations are a two-step procedure. The first step is imputing the earnings and the hours of employment of those to be assigned jobs, and household income of households with members who have been assigned jobs. The second step is to impute the new hours of household production of individuals in households affected by job assignments. With these steps completed, we can estimate the impact of a given policy on time and income poverty, both overall and on individual households. We first discuss the policy scenario in detail, then the steps involved in constructing the estimated outcome of the policy.

Policy Scenario

A very simplified job assignment scenario is envisioned in the LIMTIP project: that all eligible adults¹⁵ not working full-time¹⁶ receive full-time employment.¹⁷ This means that we need to draw donors from households in which all eligible adults currently work full-time. After eligible adults are assigned job, hours and earnings, the household income of households with eligible adult(s) is recalculated by adding the imputed amount of household earnings to the previous amount of household income. We assume that none of the other components (i.e. other than earnings) of household income undergo any change, i.e., we incorporate the maximum income effect of additional employment in our simulation. This assumption is, obviously, unrealistic for households that receive means-tested income transfers or receive income transfers that depend on employment status. While the effect of this assumption on biasing the results of our simulation is likely to vary from country to country, our belief is that such effects are likely to be small.

Once the employment and income simulation is complete, the hours of household production of individuals needs to be estimated in all households that contain job recipients. This occurs in a second

¹⁵ Eligible adults are defined as all individuals between the ages of 18 and 74 who are not disabled, retired, in school or in the military. These restrictions, other than age, could not all be applied for each country. The age restriction is simply the broadest age categorization for which all three countries had time use data.

¹⁶ Full-time is defined as working twenty-five hours per week or more.

¹⁷ An exception will be noted in the discussion of the labor force simulation.

round of hot-decking, with a second set of recipient and donor pools. The recipient pool contains all individuals included in the time use survey and living in households that contain at least one job recipient. The donor pool contains all individuals included in the time use survey and living in households in which all eligible adults engage in full-time employment.

Labour Force Simulation

This simulation builds on research done at the Levy Institute to estimate the impact of the American Recovery and Reinvestment Act of 2009 on U.S. income inequality. The procedure in this case is more straightforward, since we have no intermediate step of estimating the jobs created by given amounts of fiscal stimulus. The problem here is to assign hours and earnings to individuals receiving full-time paid employment. The method for assigning hours and earnings is a hot-decking procedure (for a review of hot-decking see Andridge and Little 2010). We use a nearest-neighbor method called affinity scoring to get a pool of records from which to match each record within matching cells determined by age, sex, and education. Before the hot-decking, we assign an industry and occupation to each job recipient. We also generate imputed wages and hours of work using a three-stage Heckit procedure. These four variables are used in the hot-decking assignment of hours and earnings.

Industry and Occupation Assignment

The first step in assigning jobs to recipients is to determine what are the likeliest industry and occupation for each of the recipients. This is done using a multinomial logit procedure. Both industry and occupation are regressed on age, sex, marital status, education and relationship to household head in the donor pool. The likelihood for each industry and occupation is then predicted in the recipient pool, using the results of the multinomial logit. Then each recipient, except those actually working part-time, is assigned the likeliest industry and occupation using those predicted likelihoods. This was also done for class of worker in the case of Mexico.¹⁸ Throughout the rest of the procedure, wherever industry and occupation are used, class of worker was also used for Mexico. For those recipients who were already working in the full-time scenario, industry and occupation were not changed.

Imputed Hours and Earnings

The imputations for the earnings and usual weekly hours of paid work were performed using a three-stage Heckit procedure (Berndt 1991, p. 627). The model, described below, was run separately for each combination of six age categories and sex. The first stage is a probit estimation of labour force participation:

$$lf_i = \alpha_1 + \beta X + \varepsilon_i \quad (1)$$

The explanatory variables, X , are indicators for the presence of male and female children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, number of children in the household, education, marital status, and spouse's age and education. The Mills ratio is calculated using the results of the first stage regression:

$$\lambda = f\left(\frac{\hat{lf}}{\hat{\sigma}_{lf}}\right) \bigg/ \left(1 - F\left(\frac{\hat{lf}}{\hat{\sigma}_{lf}}\right)\right) \quad (2)$$

¹⁸ This variable included the categories salaried, independent, employer, unpaid household, and other. In the simulation, we assigned recipients to only the first two categories.

Where f is the normal density function, F is the normal distribution function, $\hat{l}f$ is the estimated probability of labour force participation, and $\sigma_{\hat{l}f}$ is the standard deviation of $\hat{l}f$. The second stage is an OLS estimate of the log of hourly wage:

$$\ln w_i = \alpha_2 + \gamma_2 Z + \theta_2 \lambda + \mu_i \quad (3)$$

The regression is run only on those that are employed. The explanatory variables, Z , in this stage are the individual's education, age, marital status, industry and occupation (as well as state, in the case of Mexico), and finally, λ , the Mills Ratio from the first stage. Inclusion of the Mills Ratio corrects for the selection bias induced by limiting the regression to those in paid employment. The imputed log of wage is predicted for donors and recipients from the results of the regression, with industry and occupation replaced for the latter by the assigned industries and occupations from each scenario. The third stage is a regression of hours per week:

$$h_i = \alpha_3 + \gamma_3 Z + \omega \ln w_i + \theta_3 \lambda + \eta_i \quad (4)$$

The regression is run only on those in paid employment. The explanatory variables in this stage are the same as the previous stage, with the addition of the imputed wage for each scenario. Imputed hours per week are predicted for donors and recipients using the results of the regression, replacing the industry and occupation of the latter with the assignments from each scenario. The results of the last two stages give us the variables with which we perform the hot-decking procedure to assign earnings.

Hours and Earnings Assignment

Once we have assigned jobs, we can assign earnings and usual hours of work to those individuals who received a new job.¹⁹ The matches are performed within cells formed from combinations of age, sex and educational attainment. The assignment method is hot-decking. The variables used to assess nearness of match are household type, marital status, spouse's labour force status, assigned industry and occupation, indicators for the presence of male and female children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, number of children and the two imputed variables: log of wage and hours worked. Industry and occupation are the most heavily weighted variables. Next are imputed hours and wage, followed by household type and then the variables relating to children in the household. The selection of matches is done using affinity scoring.

Once the hot-decking is finished, we compare new earnings to previous earnings. In the full-time employment simulation, there were a small number of individuals who were actually employed part-time and who ended up with simulated earnings that were lower than their actual earnings. We did not change the hours of employment or earnings of these individuals and removed them from the pool of recipients.²⁰ For the other recipients, we revised their household income by adding the imputed amount of household earnings (the sum of earnings of all recipients in the household) to the pre-simulation amount of household income.

¹⁹ In Mexico, we also assign type of employment (own-account, unpaid family worker or wage/salary worker) and formality (formal/informal).

²⁰ In Mexico, we also kept the original type of employment and formality for those whose assigned earnings fell below their actual earnings.

Time Use Reallocation

We assume that as a result of the job assignment, the time use pattern of each eligible individual in the households that contain one or more job recipients from the first stage will change. An individual in the recipient household is considered “eligible” for time use reallocation if he or she belongs to the sampling universe of the time use survey. While the precise definition of the sampling universe varies from country to country, essentially this restriction amounts to excluding the very young and the old.²¹ We use a second round of hot-decking to assign new weekly hours of household production to each of the “eligible” individuals, based on updated labour force participation variables for the recipients of jobs in the first stage. The method is the same as the first stage, with the exception of the matching variables used and their relative weighting in the procedure. In this stage, the variables used to assess nearness of match are household type, marital status, spouse’s labour force status, indicators for the presence of male and female children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, number of children in the household, number of adults in the household, household income, the income share of each individual,²² and the two imputed variables from the first stage: earnings and usual weekly hours of employment. Household income and labour force status are updated to reflect the increased earnings and the new job assignments received in the previous stage. The number of children and number of adults in the household, household income, and income share are the most heavily weighted variables. Next are household type, updated earnings and usual weekly hours of employment, and labour force status, followed by marital status and spouse’s labour force status, then the variables relating to children in the household.

Assessment

The question of how to assess the results of the procedure outlined above is a difficult one to answer. Since we are creating a counter-factual distribution, we have nothing with which to make a valid comparison. The implication is that we have to compare the distribution of the imputed earnings, hours of employment and hours of household production within subgroups of the donor and recipient population. The comparisons we present here are between individuals in the donor pool and the recipients living in income poor households prior to the job assignment.²³ We choose the latter group because they are the target group of policies to alleviate poverty.

Argentina

The recipient and donor pool for the labour force simulation are presented in Figure 1, broken down by sex, age, and education.²⁴ The total number of males and females in the recipient pool was 30,525 and 60,925, respectively, while the donor pool contained 573,380 males and 414,428 females, indicating the sizeable gender mismatch between the recipient and donor population. The most striking pattern for

²¹ In Argentina, only individuals between the ages of 15 and 74 years were included in the time use survey. The Chilean and Mexican time use surveys included individuals of age 12 years and more (see [Masterson \(2011\)](#) for a discussion of the time use surveys of the individual countries and the adjustments made to them for use in the LIMTIP).

²² Income share is included to reflect changes in bargaining power within the household and its impact on the distribution of household production work.

²³ It should be noted that we are not using the official threshold of income poverty here. Our threshold is the standard poverty threshold plus the monetized value of time deficit. Time deficit is the amount of time required for the household to maintain itself minus the amount of time available to household members. For a comparison of our approach with some of the earlier approaches to measurement of time and income poverty, see Zacharias (2011).

²⁴ Percentages in this and similar figures to follow are of males and females in the recipient or donor pools.

both men and women is that the individuals in the recipient pool are less well-educated than those in the donor pool. A somewhat less pronounced feature is that more of the recipients are at the younger end of the age distribution. Figures 2 and 3 show the ratios of mean and median earned income and usual hours of work, respectively, of the recipients in adjusted income poverty to donors by matching cell. The shaded areas represent $\pm 20\%$ of parity. As we can see, for most of the cells, the imputed hours and earnings from the simulation are similarly distributed within the cells. Notable exceptions are males aged 45 to 54, especially college graduates. However, although there are a fair number of these males in the donor pool, the recipient pool has fewer than 250. The female comparisons look much better, as is to be expected, given the larger recipient pool. The imputation of usual hours of employment looks much better than earnings. This is simply because there is much smaller variation in usual hours worked, especially for full-time workers.

The recipient and donor pools in the hot-decking of hours of household production are presented in Figures 4 and 5. Figure 4 breaks down the pools by matching cells. There are 89,508 males and 587,076 females in the recipient pool, while there are 99,180 males and 494,486 females in the donor pool. Thus, unlike the sizeable gender mismatch in the hot-decking of hours of employment, there is much less mismatch in the hot-decking of hours of household production. We can see that, although the pools for men and women are slightly different than for the earnings match, the patterns are fairly similar: more of the recipients are less-educated. The recipients are concentrated in two or three or more adult households, while the donors are more evenly distributed.

The results of the hot-decking are presented in Figures 6 and 7. Again the ratios of individuals' mean and median weekly hours in the recipient pool to the donor pool are displayed by matching cell and reference group cells, respectively. The shaded area represents between 80% and 120%. For most of the cases, the ratios are within the twenty-percent interval. The worst variations are for males aged 45 to 54, especially those with less than a high school diploma and with college degrees. These categories include less than 10% of all men in the recipient pool, but the differences are substantial: roughly 10 hours per week at the median for both groups. Perhaps more concerning are the differences in the reference group cells. Women in the recipient pool in three or more adult, three or more children households have twenty-eight hours at the median while in the donor pool the median is less than five hours.

Chile

The recipient and donor pools for the first stage of the simulation for Chile are shown in Figure 8. The total number of males and females in the recipient pool was 70,281 and 201,132, respectively, while the donor pool contained 1,368,669 males and 901,541 females. As in the case of Argentina, we can see the gender mismatch between the donor and recipient pools, reflecting the greater proportion of women among the recipients and the greater proportion of men among the donors. Also, just as in Argentina, the recipient pool of both men and women is generally less educated than the donor pool, although for males, high school graduates are the majority in both pools (61% of recipients and 51% of donors), while for females, the ratios are slightly lower with only a plurality of donors in this category (55% of recipients and 49% of donors). Pluralities of both pools are less than 35 years old, though a greater share of both female and male recipients are in the youngest age group. Compared to Argentina, the recipient and donor pools for men and women look more similar for the Chilean full-time simulation.

The results of the first hot-decking imputation are shown in Figures 9 and 10. Looking at the imputed earnings in Figure 9 first, we can see that for the most part, the results look similar to the donor pool. There are some exceptions. The most different result was for median earnings for elderly males with

some college: median earnings of recipients were five and a half times that of donors, while mean earnings were over twice as high. However, this was one of the smallest cells in size, so the results are both understandable, and not worrisome. For females, there were no recipients in the latter group. The group with the largest divergence was therefore female college graduates aged 55 to 64, whose mean earnings were nearly twice as high as in the donor pool. The usual hours imputation results in Figure 10 show relatively small differences between the recipients and the donor pool.

The recipient and donor pools for the second stage of the full-time simulation are presented in Figure 11, broken down by matching cell. In the recipient pool there were 70,281 males and 201,162 females, while in the donor pool there were 1,368,669 males and 901,541 females, following the pattern for Argentina. We can see that again, while both donor and recipient pools are heavily skewed to the younger and less-educated cells, this pattern is more prevalent among the recipients than the donors. 56% of the male and 51% of the female recipients are less than 35 years of age, while only 48% and 42% of the corresponding donors are. And while 93% of the male and 94% of the female recipients have high school diplomas or less, 67% and 69% of the corresponding donors fall into this category. Figure 12 presents the breakdown of the recipient and donor pools by reference group cells. While the donors are concentrated in households with no children, the recipients are more evenly distributed, but among households with children.

Figure 13 presents the results of the hot-decking of hours of household production broken down by matching cells. There were 310,605 males and 341,721 females in the recipient pool and 835,612 males and 935,324 females in the donor pool, a much more gender-balanced set of pools than for the employment imputation. We can see that the median weekly hours of household production for males in most cells is zero. The mean values are similar to those of the donor pool. For females, we see that both mean and median weekly hours are quite similar. The exceptions among both males and females are older and more-educated. These are also the less-populous cells. Figure 14 shows the breakdown by reference group cells. The results for females are quite similar to the donor pool. Males look more divergent, though this impression is amplified by the fact that in many of the cells median weekly hours of household production for males is zero.

Mexico

Figure 15 shows the distribution of the recipient and donor pools by matching cell for the simulation for Mexico.²⁵ The total number of males and females in the recipient pool was 2,736,772 and 11,647,816, respectively, while the donor pool contained 22,927,775 males and 11,607,180 females. These figures suggest that the gender mismatch between the recipient and donor pools is much more severe for Mexico than for the other two countries. This is partly a reflection of the differences in geographical coverage. The Mexican sample is representative of the entire nation while the samples for Argentina and Chile cover, respectively, the city of Buenos Aires and Greater Santiago.

Similar patterns are apparent for the donor and recipient pools of men and women in Mexico as for Chile and Argentina. The recipient pool has less education than the donor pool: 55% of males and 59% of females in the recipient pool have less than a high school diploma, compared to 35% and 29% respectively in the donor pool. The distribution by age is more similar than we have seen in the other two countries. The results of the employment hot-decking are presented in Figures 16 and 17. It is clear

²⁵ Although each step of the process was done using rural/urban status as an additional strata variable to define matching cells, the results we present are for the country as a whole. Differences in earnings and hours of work in the household and the market were much greater by sex, age and education than by rural/urban status. In addition, the class of worker was used everywhere in the process that industry and occupation were used.

that compared to the other two countries in this study the simulation has produced results that much more closely match the distribution of earnings and usual weekly hours in the donor pool, especially the latter. This is most likely due to both the higher quality of the data itself, and the fact that there are so many more observations. The latter is important because the more records there are to choose from, the more likely it will be that a record will be matched with another that is quite similar to it during the hot-decking procedure. The most divergent result for earnings is for females with some college education aged 55 to 64, the mean and median earnings of which were 38% those of the donor pool. This group accounts for less than 0.1% of the female recipient pool.

Turning to the hot-decking of hours of household production, we see in Figures 18 and 19 the distribution of the recipient and donor pools by matching cell and reference group cells, respectively. As we have seen in the above simulations, the donor and recipient pools for the household production hot-decking were much more gender balanced: there were 16,587,175 males and 18,372,055 females in the recipient pool and 8,990,380 males and 9,690,580 females in the donor pool. Again, those in the recipient pool tend to have a lower level of education than those in the donor pool. The distribution of recipients and donors by reference group cells is quite different than in Argentina or Chile, however. In Mexico, fewer of the recipients are in households without children (17% of males and 16% of females) compared to the donor pool (37% of males and 32% of females). Also, almost none of the recipients are in one-adult households (1% of males and 3% of females) while many of the donors are (16% of males and 17% of females).

Figures 20 and 21 present the imputed hours of household production, broken down by matching cell and reference group cell, respectively. As we can see, the distribution of household production weekly hours among recipients does not differ greatly from that of the donors, by matching or reference group cell. In only four of the matching cells does the difference exceed 20%, while only two of the reference group cells show that great a difference.

Conclusion

The challenge in assessing the quality of simulations such as we have produced here is that there is no real-world situation to which to compare them. For each country, the imputed earnings, usual weekly hours of market work and household work tend to look reassuringly similar to the distribution in the donor pools. Intuition tells us that they should look similar, but the composition of the donor pools and the recipient pools is quite different. If we assume that the results *should* match up with the current situation, at least on an individual level, then we can use the results presented here as a guide to the quality of the simulations completed for this project. This is the best we can do in terms of quality assessment. Can we then conclude that the simulations that we ran for Mexico exceed in quality those for Chile and Argentina? Given the assumption that the results should match the current situation, we would answer yes. That assumption is not convincing, but absent a better metric, it will have to do.

References

- Andridge, Rebecca R., and Roderick J. A. Little. 2010. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review* 78 (1): 40-64. doi:10.1111/j.1751-5823.2010.00103.x.
- Berndt, Ernst R. 1996. *The Practice of Econometrics: Classic and Contemporary*. Har/Dis. Addison Wesley, March 14.
- Cranmer, Skyler J., and Jeff Gill. 2007. Multiple Hot Deck Imputation: A Non-Parametric Alternative to Multiple Imputation. September.
- Masterson, Thomas N. 2011. "Quality of match for statistical matches used in the development of the Levy Institute Measure of Time and Income Poverty (LIMTIP) for Argentina, Chile, and Mexico." Levy Economics Institute Working Paper 692 (October). Annandale-on-Hudson, NY: Levy Economics Institute of Bard College. http://www.levyinstitute.org/pubs/wp_692.pdf.
- Masterson, Thomas N. 2012. "Simulations of full-time employment and household work in the Levy Institute Measure of Time and Income Poverty (LIMTIP) for Argentina, Chile, and Mexico." Levy Economics Institute Working Paper 727 (July). Annandale-on-Hudson, NY: Levy Economics Institute of Bard College. http://www.levyinstitute.org/pubs/wp_727.pdf.
- Zacharias, Ajit. 2011. "The measurement of time and income poverty." Levy Economics Institute Working Paper 690 (October). Annandale-on-Hudson, NY: Levy Economics Institute of Bard College. http://www.levyinstitute.org/pubs/wp_690.pdf.
- Zacharias, Ajit, Thomas Masterson, and Kijong Kim. 2009. "Distributional Impact of the American Recovery and Reinvestment Act: A Microsimulation Approach." Levy Economics Institute Working Paper 568 (June). Annandale-on-Hudson, NY: Levy Economics Institute of Bard College. http://www.levyinstitute.org/pubs/wp_568.pdf



Figures

Figure 1 Donor and Recipient Pools for Argentina Labour Force Simulation by Sex, Age and Education

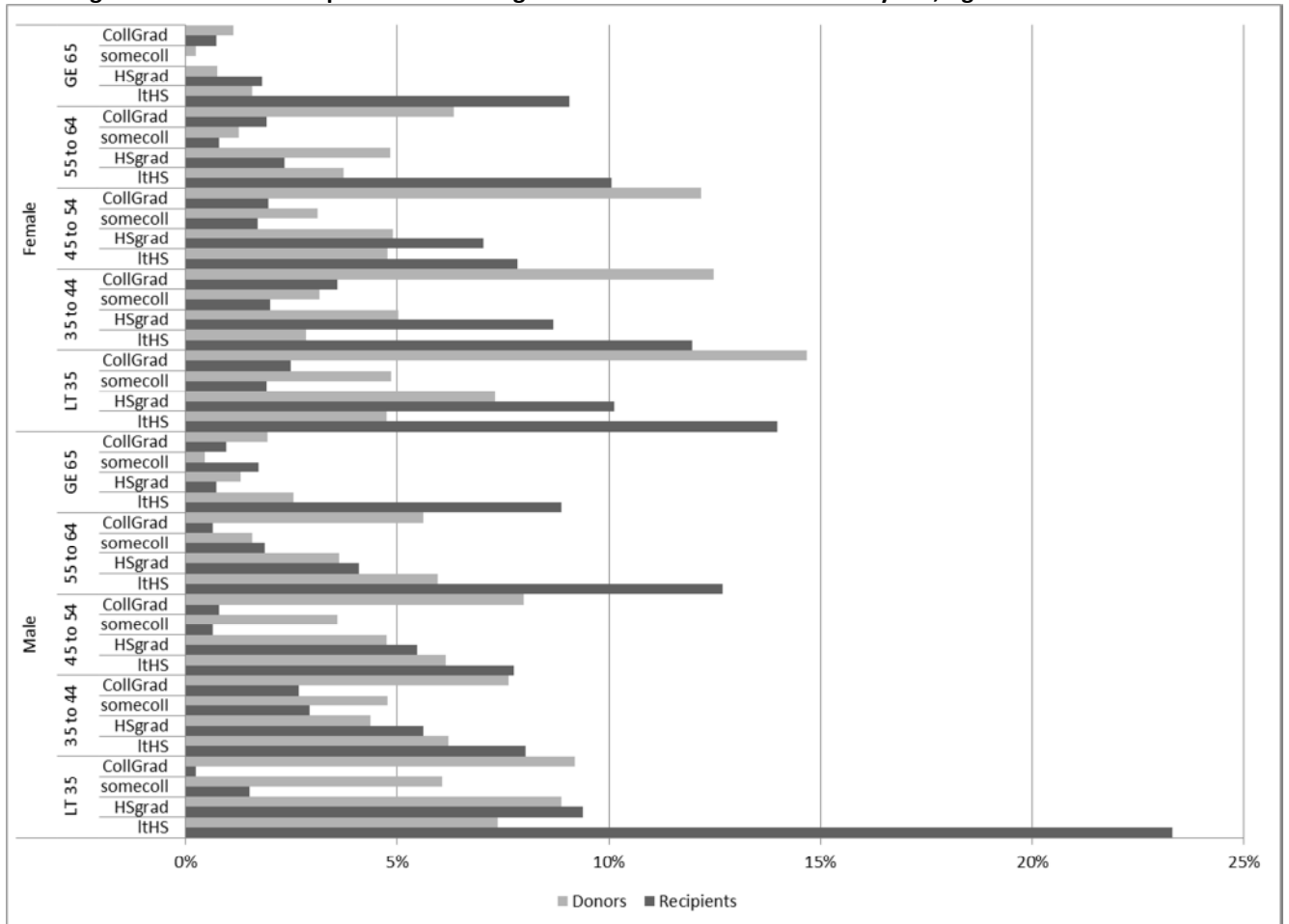


Figure 2 Mean and Median Earned Income for Donors and Recipients for Argentina Simulation by Sex, Age and Education

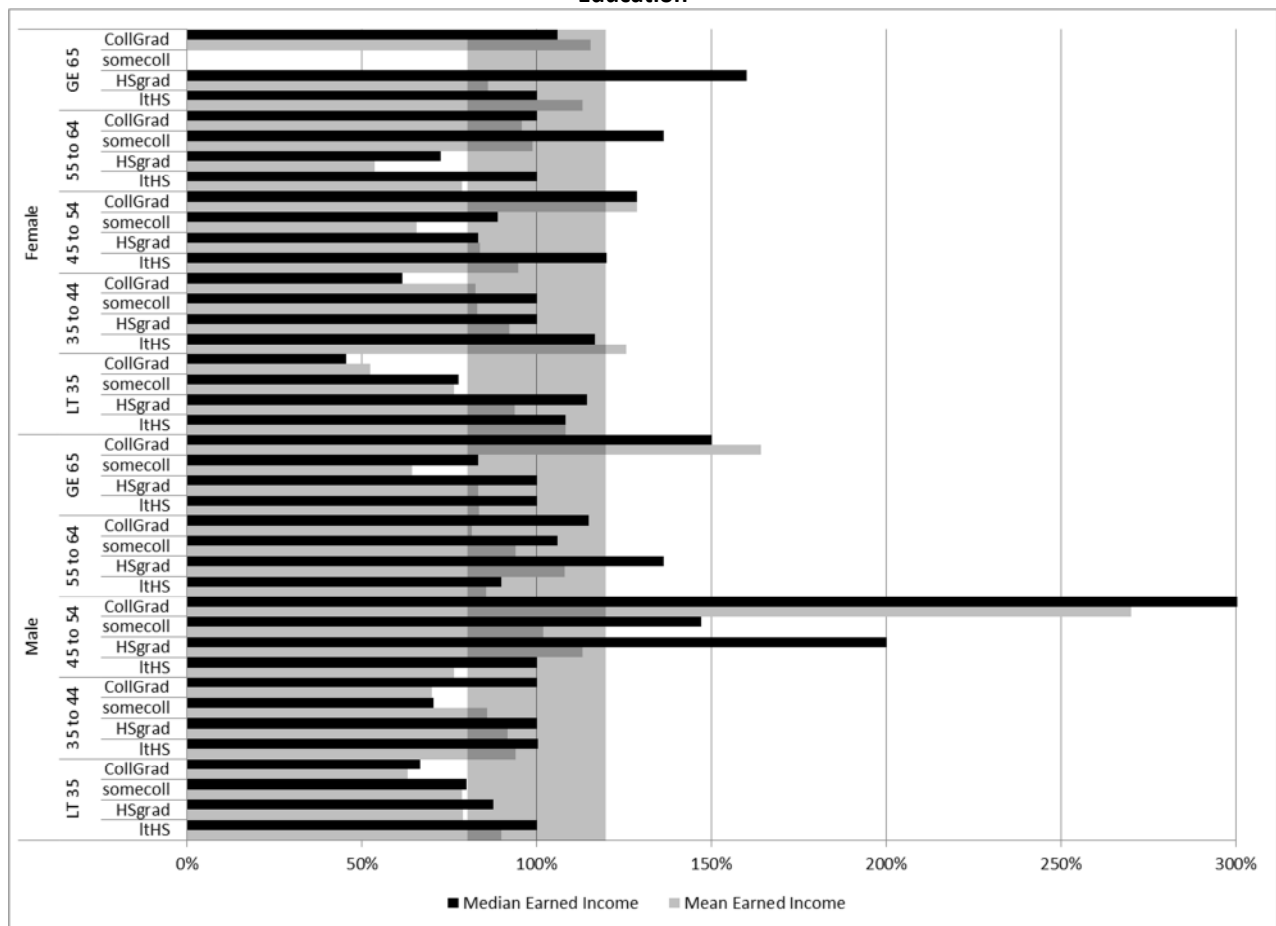


Figure 3 Mean and Median Usual Hours of Work for Donors and Recipients for Argentina Simulation by Sex, Age and Education

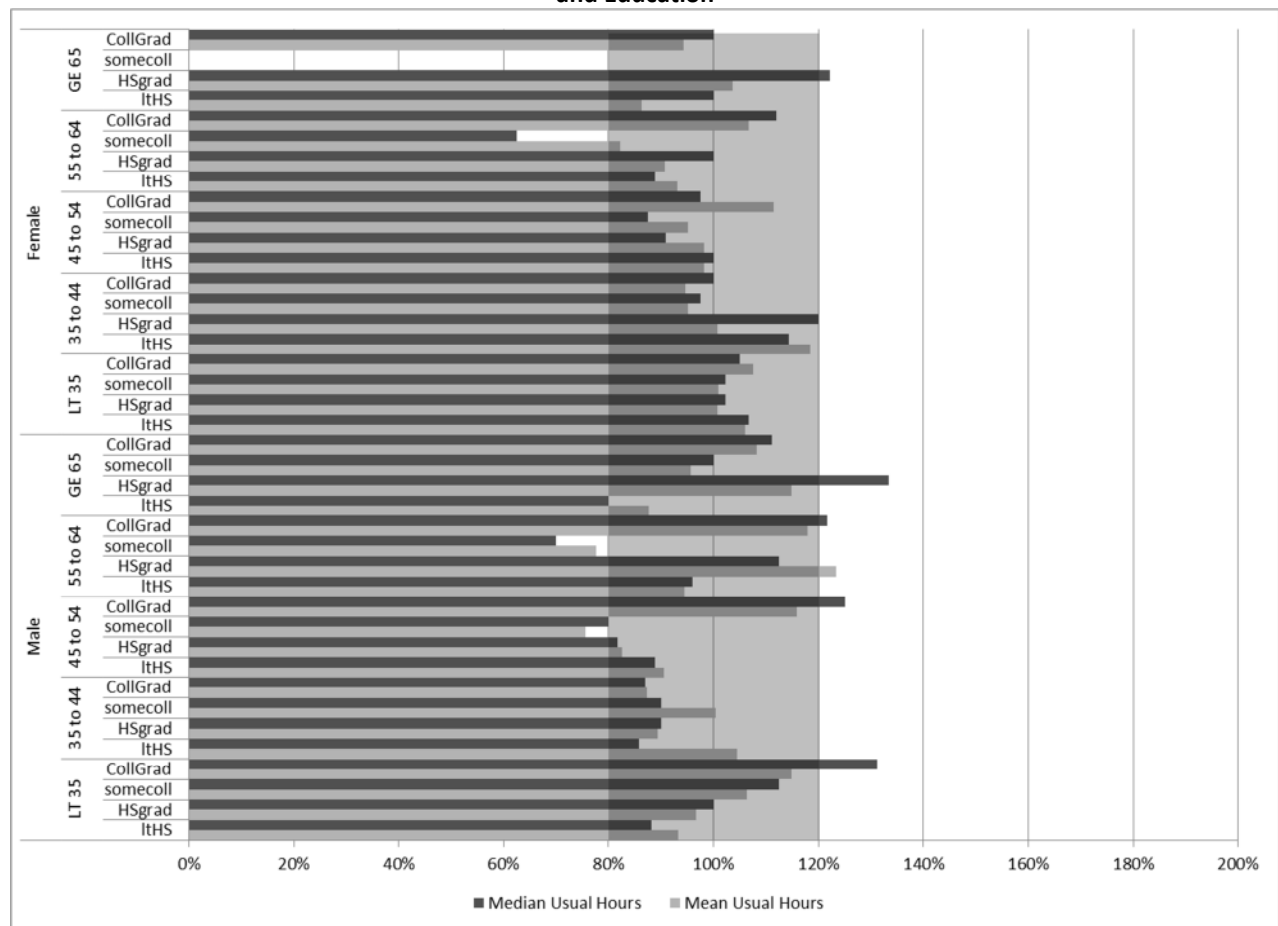


Figure 4 Donor and Recipient Pools for Argentina Time Use Simulation by Sex, Age and Education

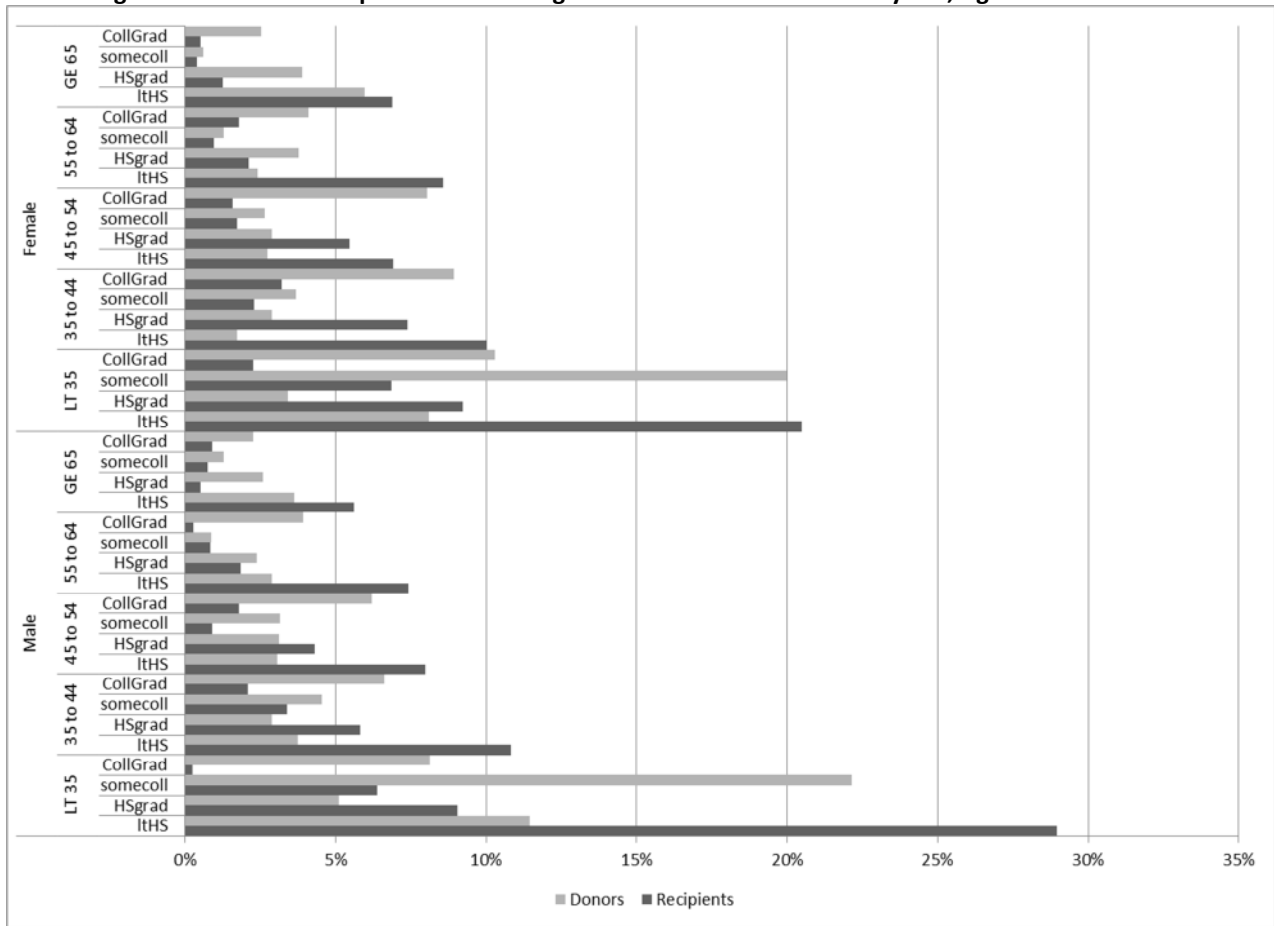
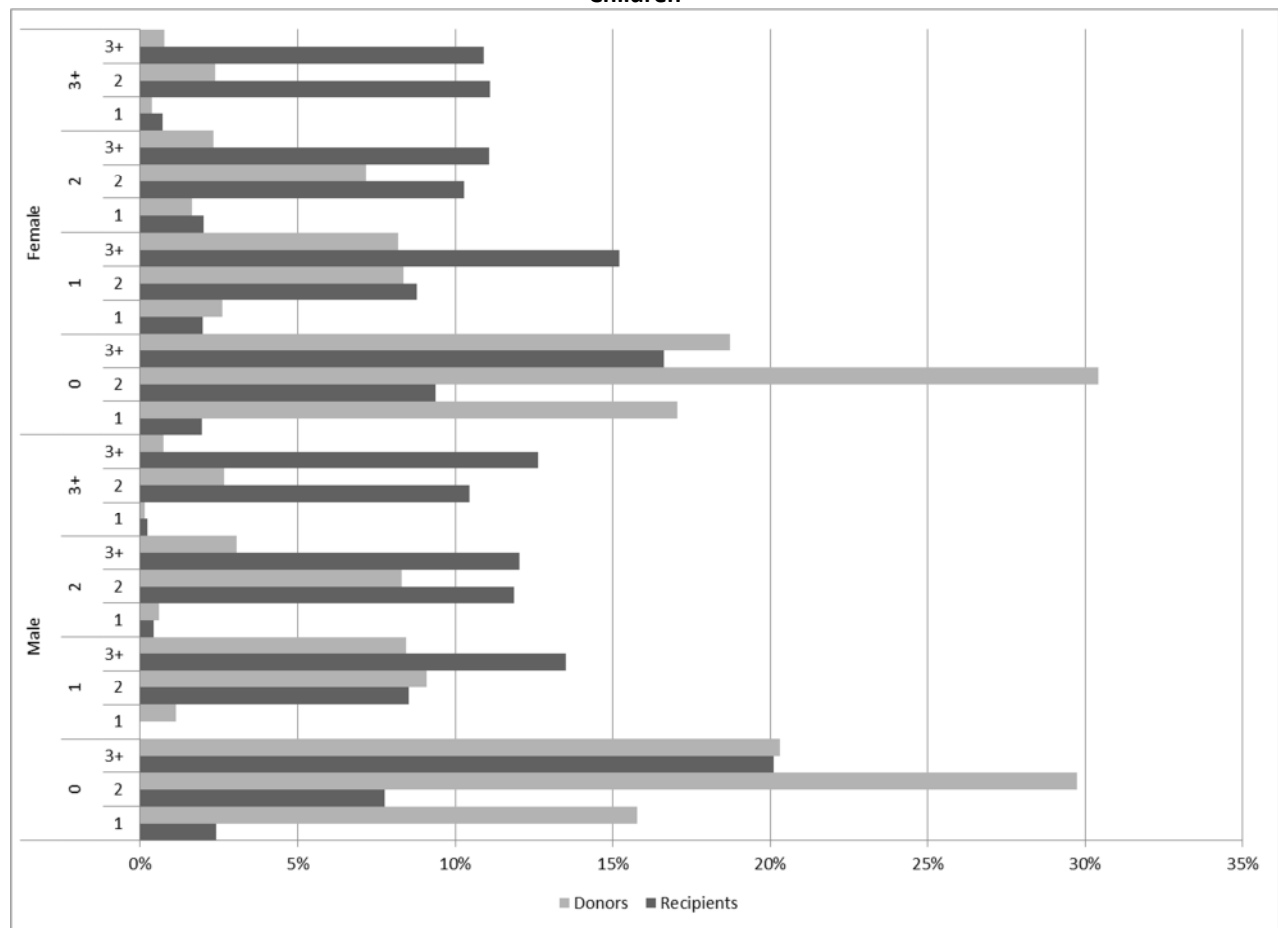


Figure 5 Donor and Recipient Pools for Argentina Time Use Simulation by Sex, Number of Adults and Number of Children



**Figure 6 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Argentina
Simulation by Sex, Age and Education**

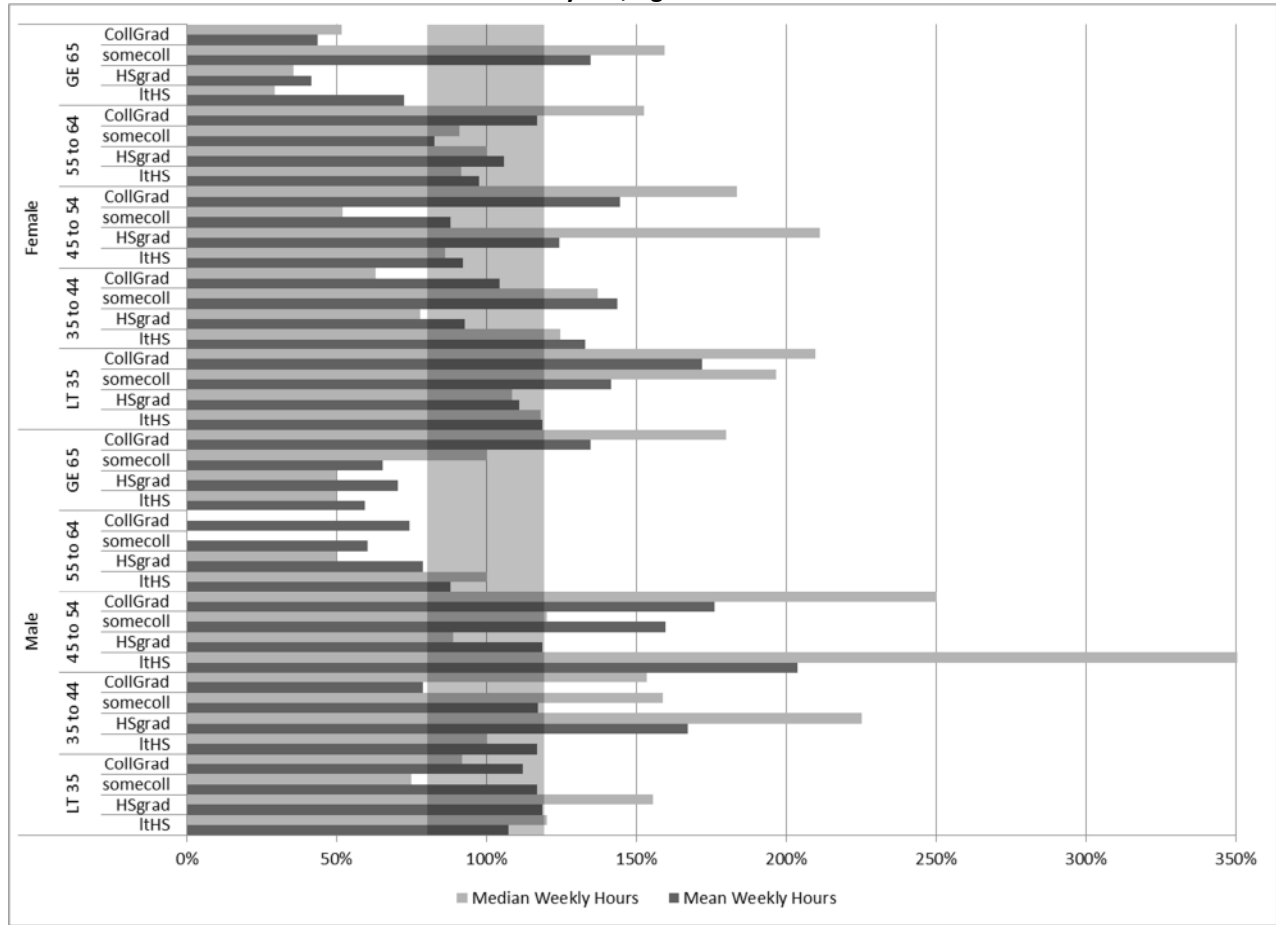


Figure 7 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Argentina Time Use Simulation by Sex, Number of Adults and Number of Children

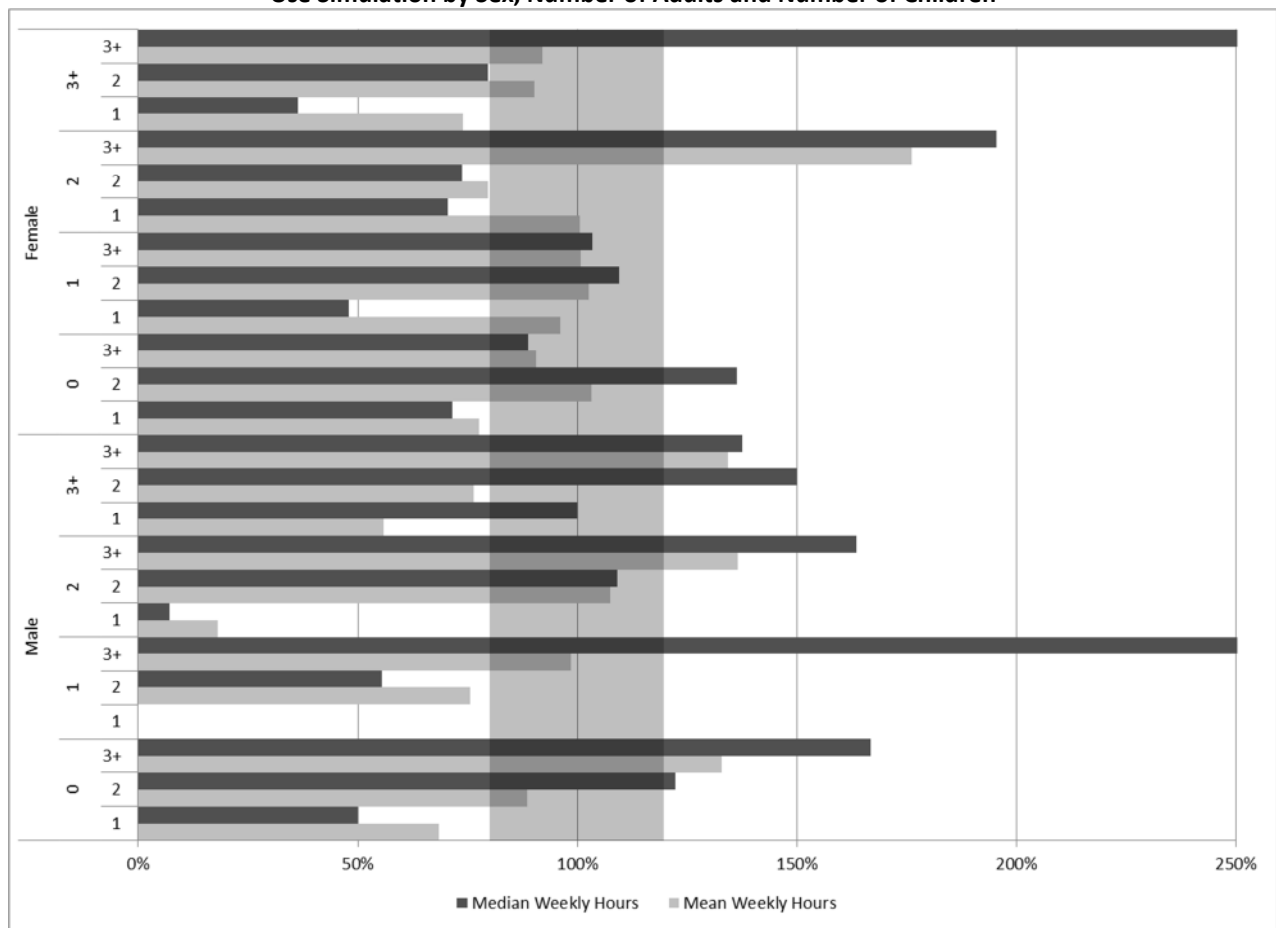


Figure 8 Donor and Recipient Pools for Chile Labour Force Simulation by Sex, Age and Education

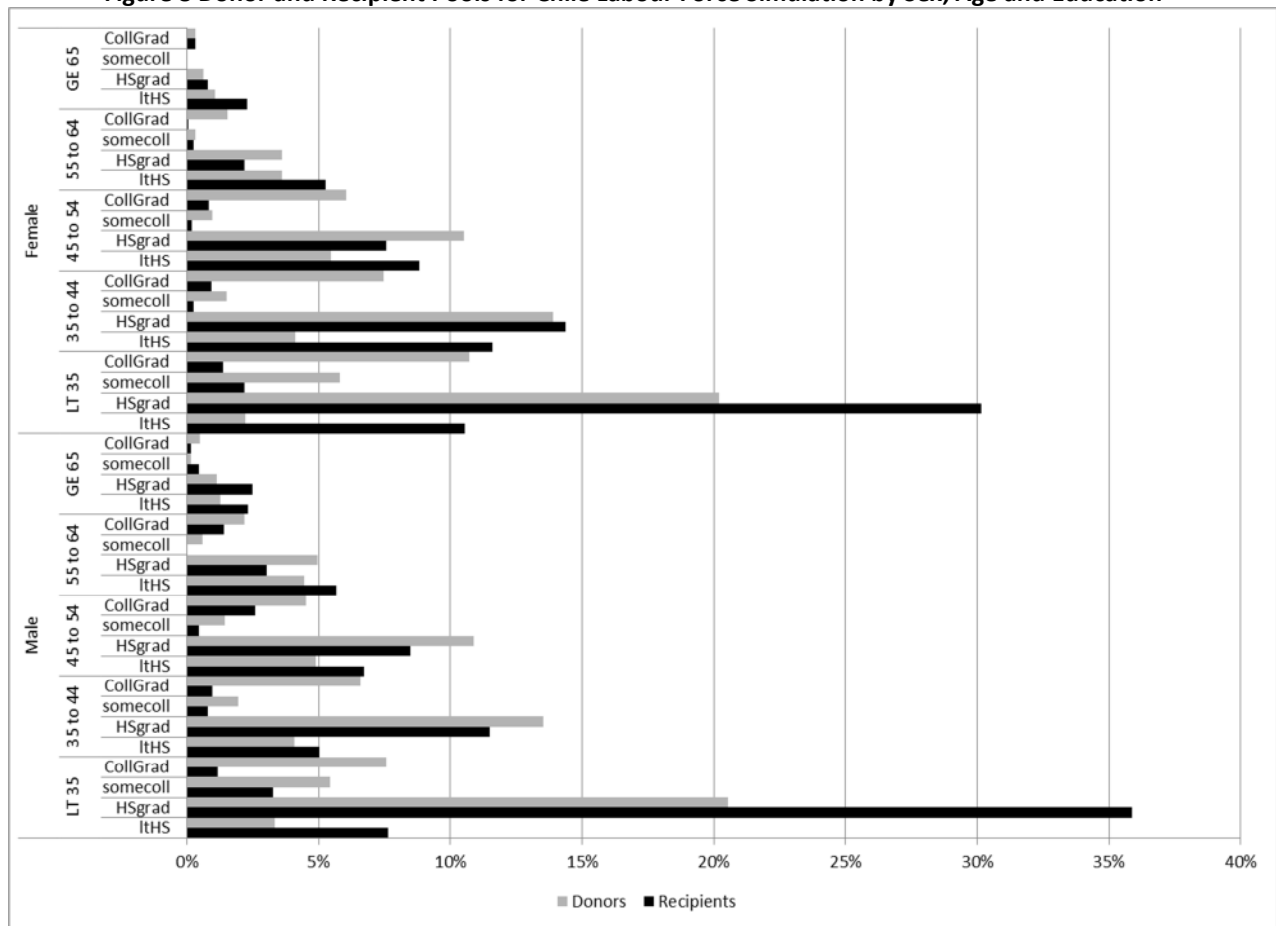


Figure 9 Mean and Median Earned Income for Donors and Recipients for Chile Simulation by Sex, Age and Education

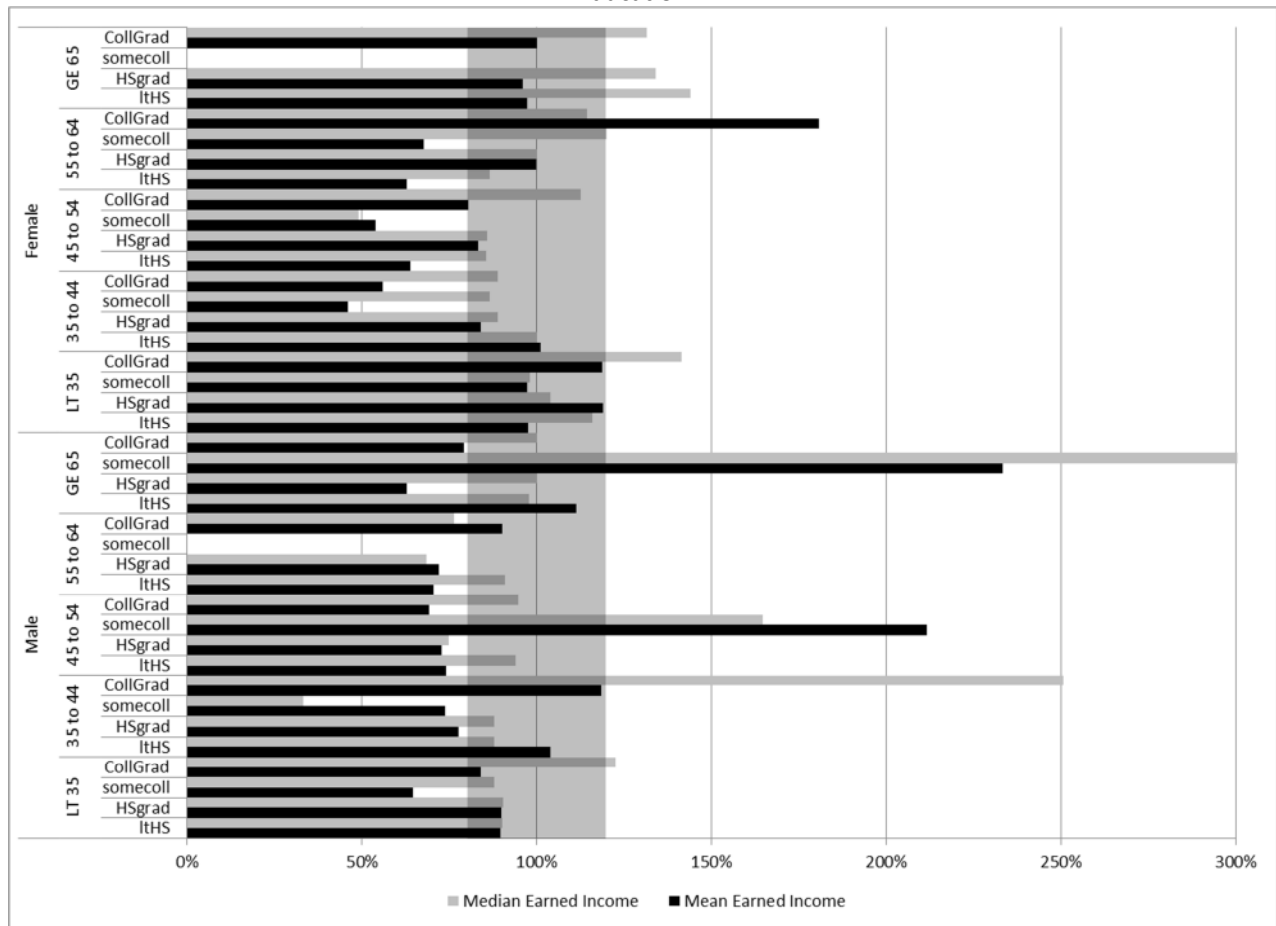


Figure 10 Mean and Median Usual Hours of Work for Donors and Recipients for Chile Simulation by Sex, Age and Education

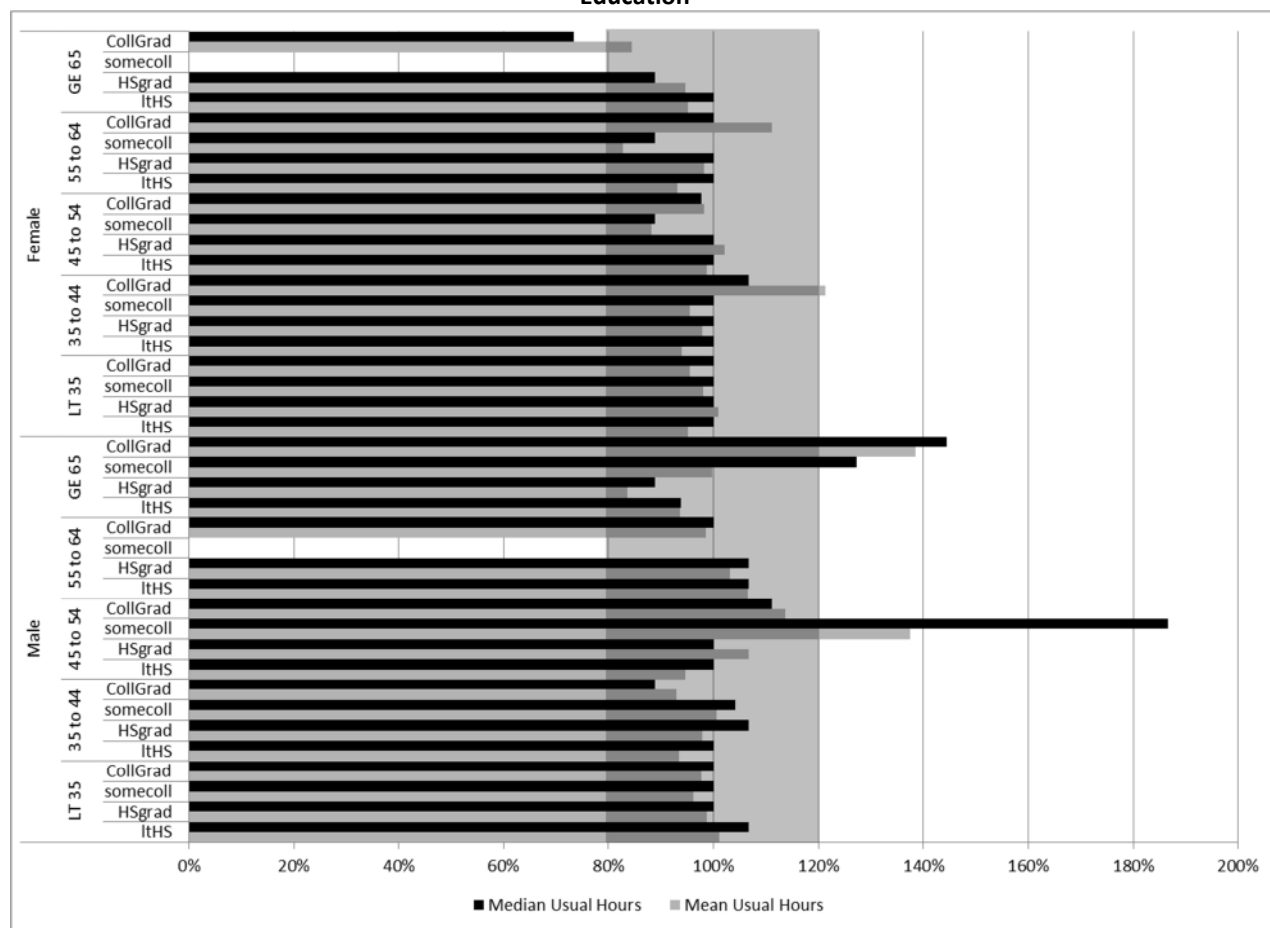


Figure 11 Donor and Recipient Pools for Chile Time Use Simulation by Sex, Age and Education

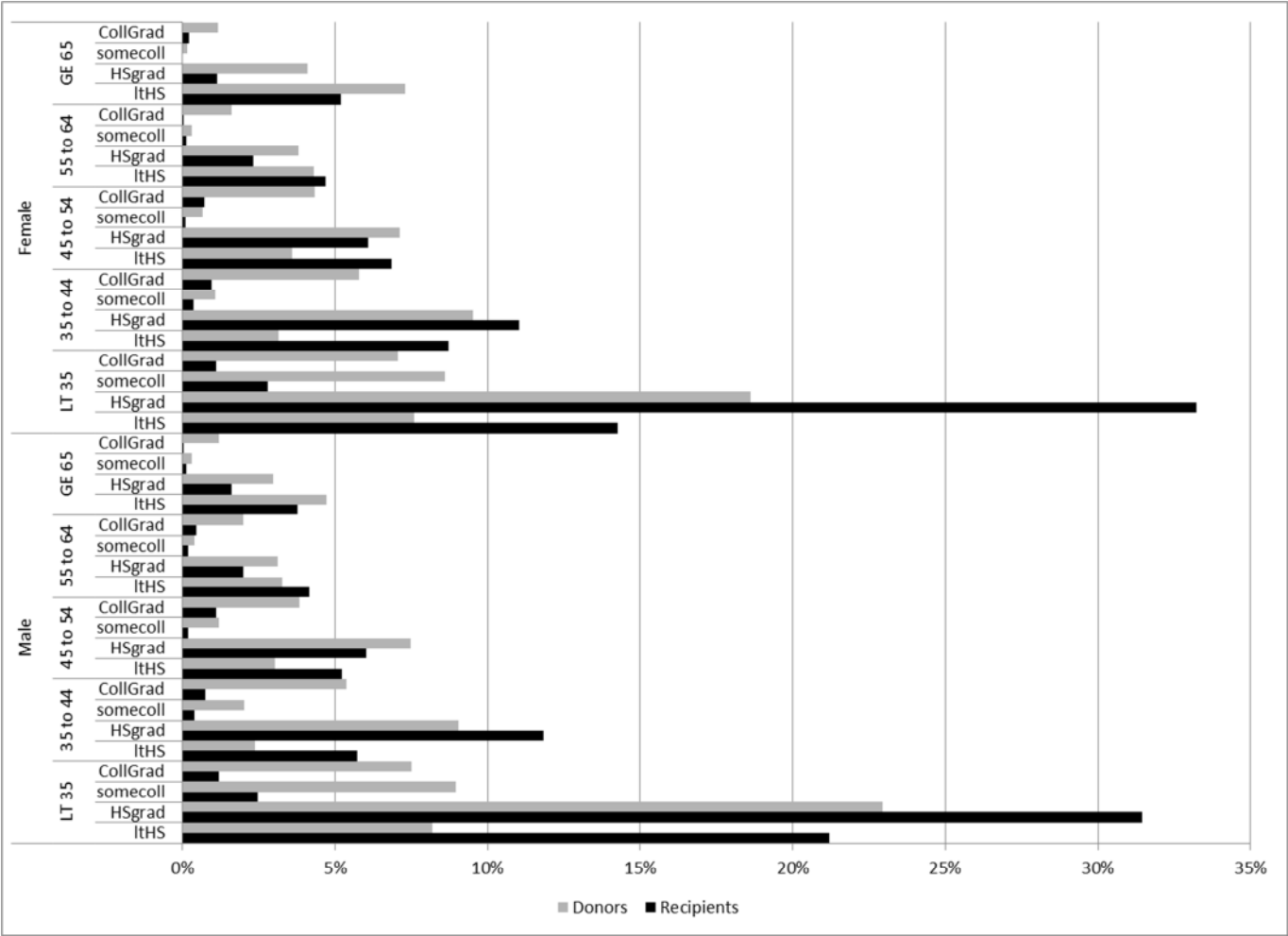


Figure 12 Donor and Recipient Pools for Chile Time Use Simulation by Sex, Number of Adults and Number of Children

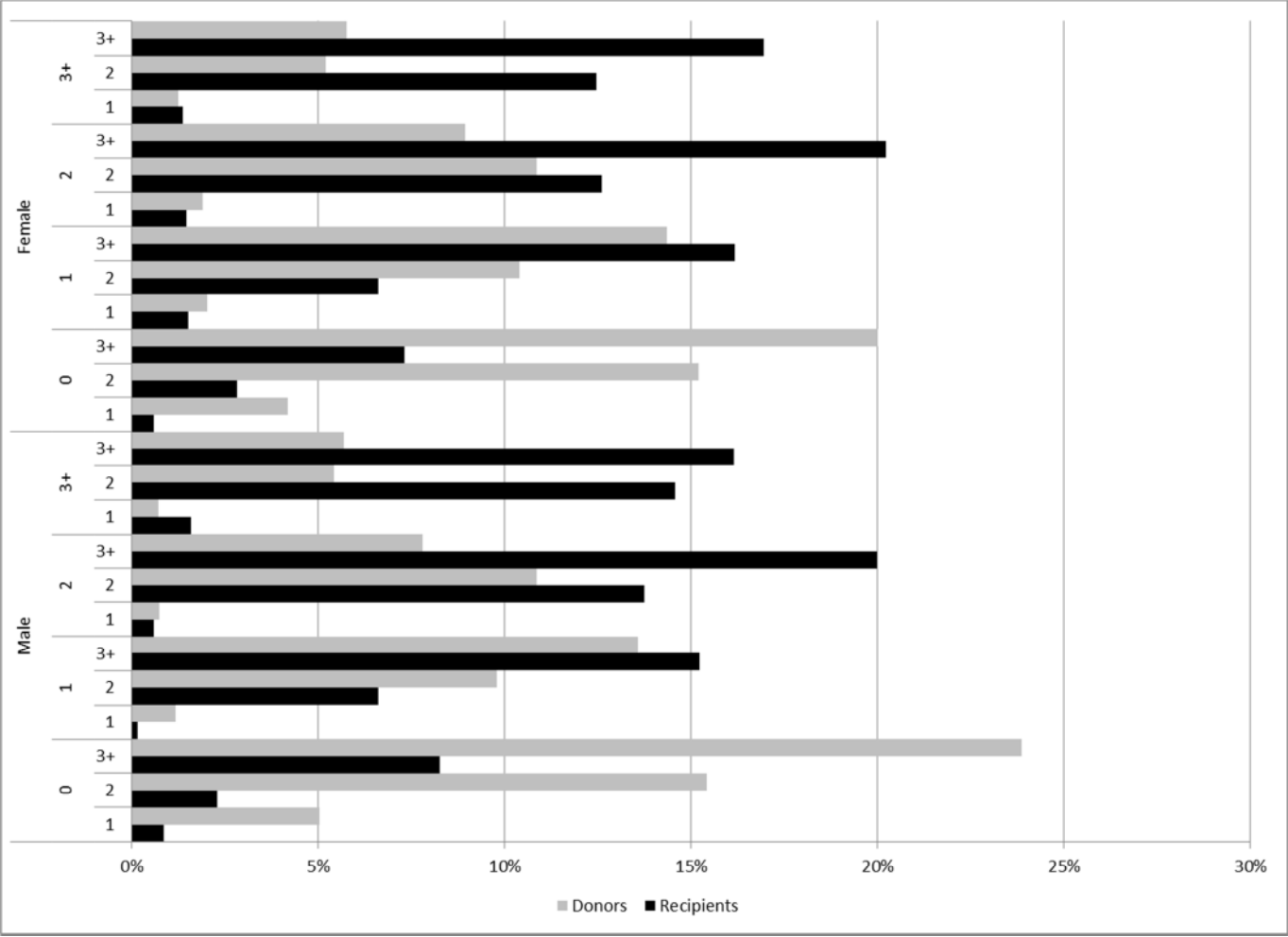


Figure 13 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Chile
Simulation by Sex, Age and Education

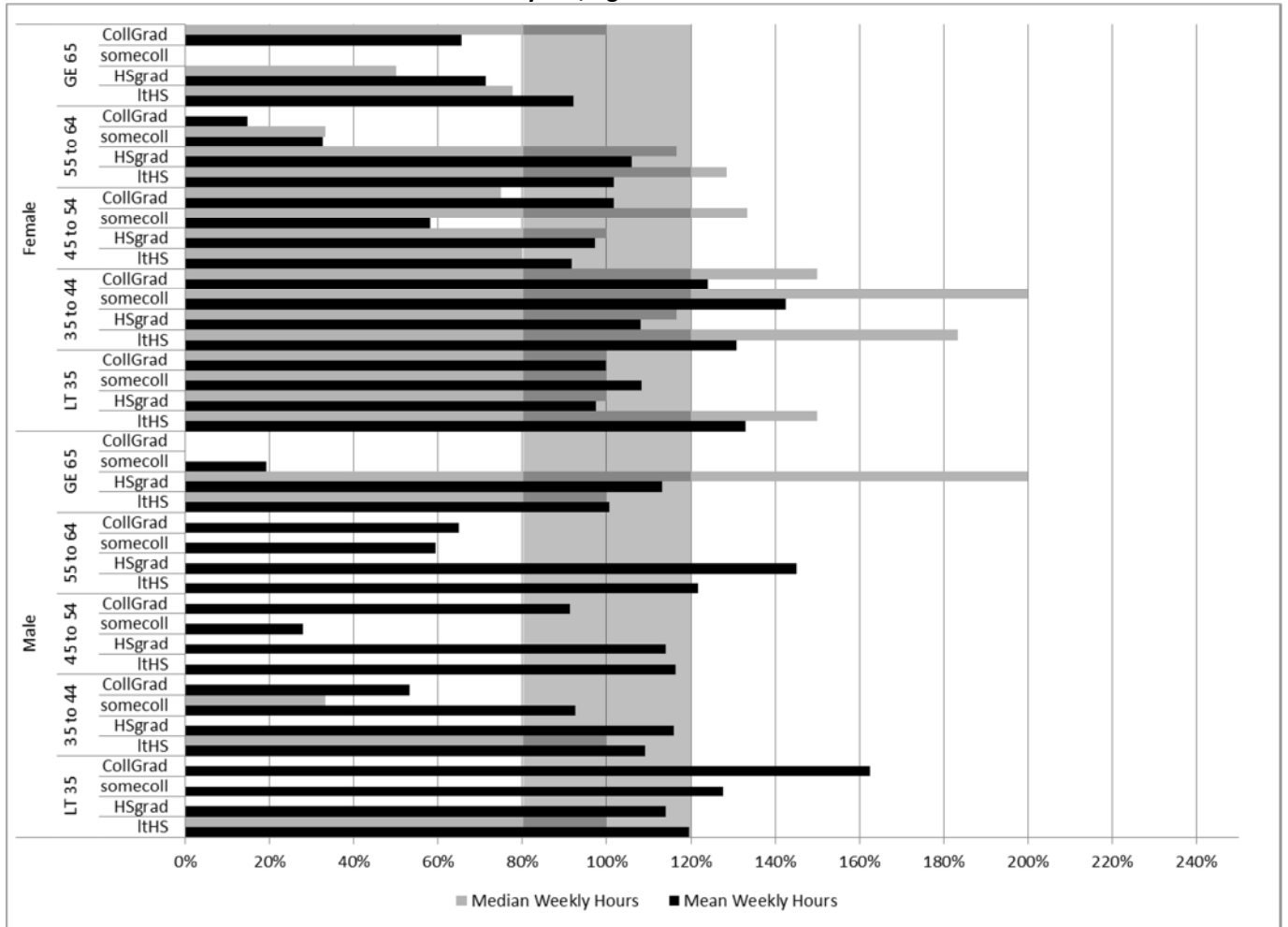


Figure 14 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Chile Time Use Simulation by Sex, Number of Adults and Number of Children

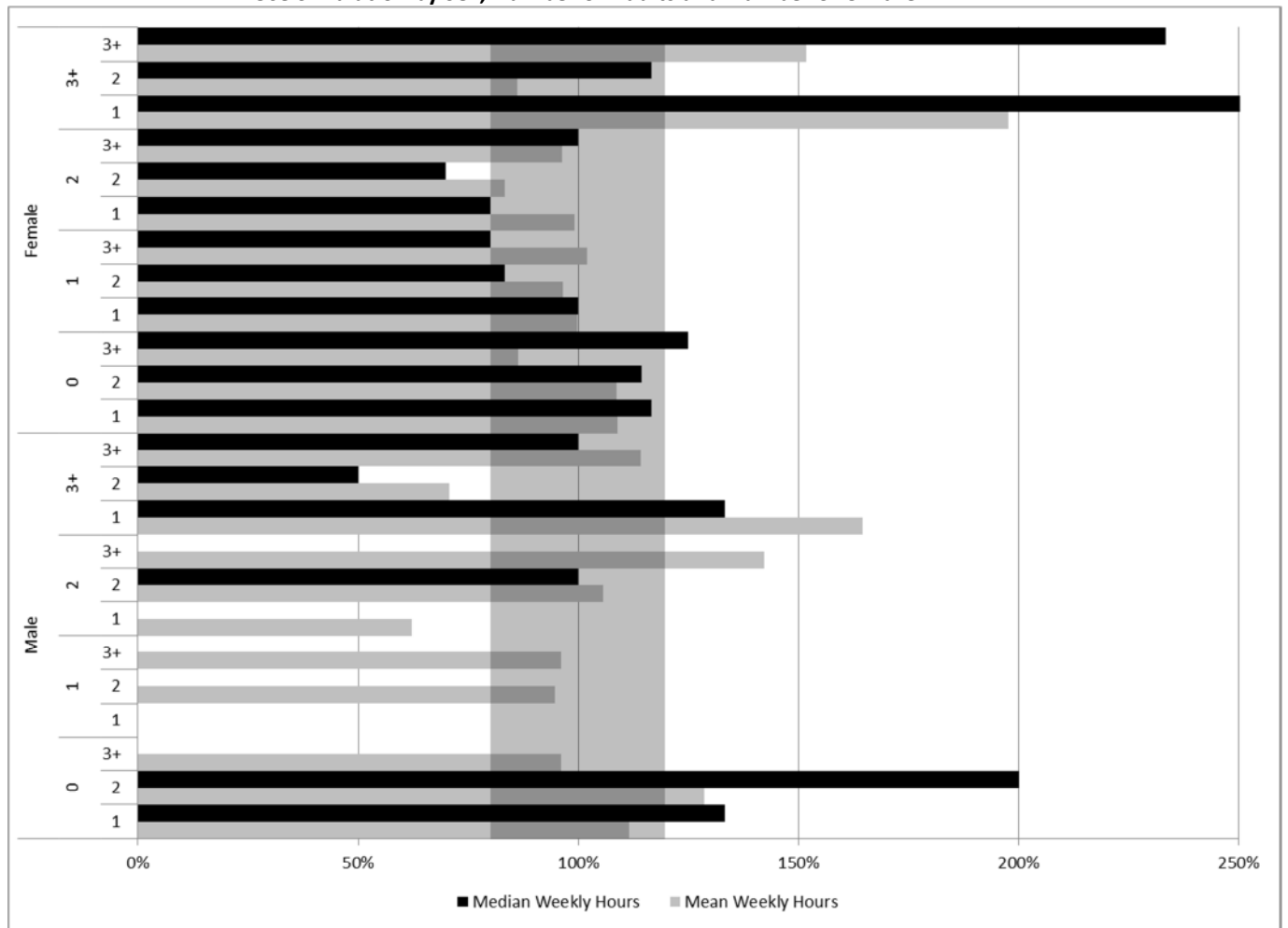


Figure 15 Donor and Recipient Pools for Mexico Labour Force Simulation by Sex, Age and Education

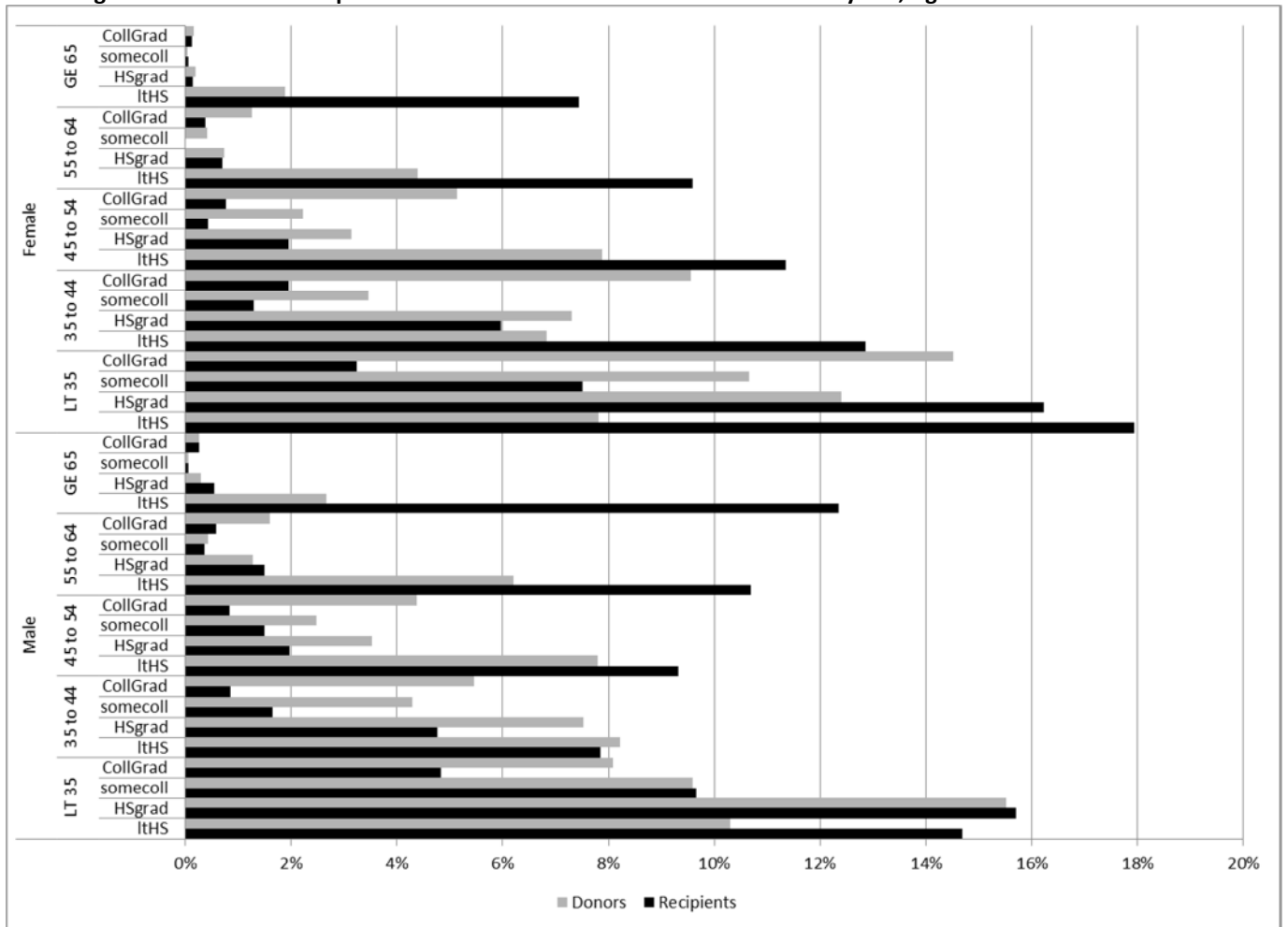


Figure 16 Mean and Median Earned Income for Donors and Recipients for Mexico Simulation by Sex, Age and Education

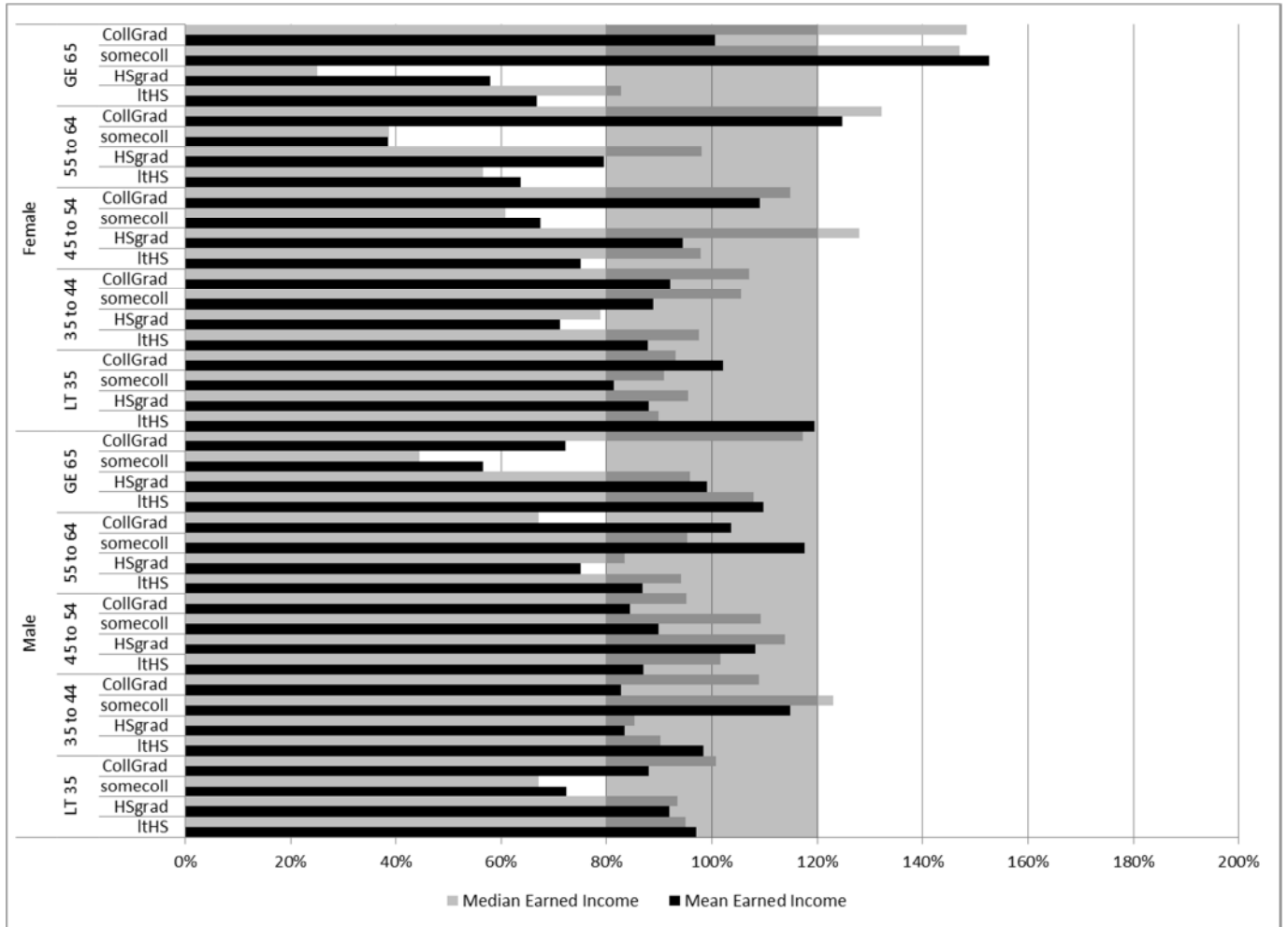


Figure 17 Mean and Median Usual Hours of Work for Donors and Recipients for Mexico Simulation by Sex, Age and Education

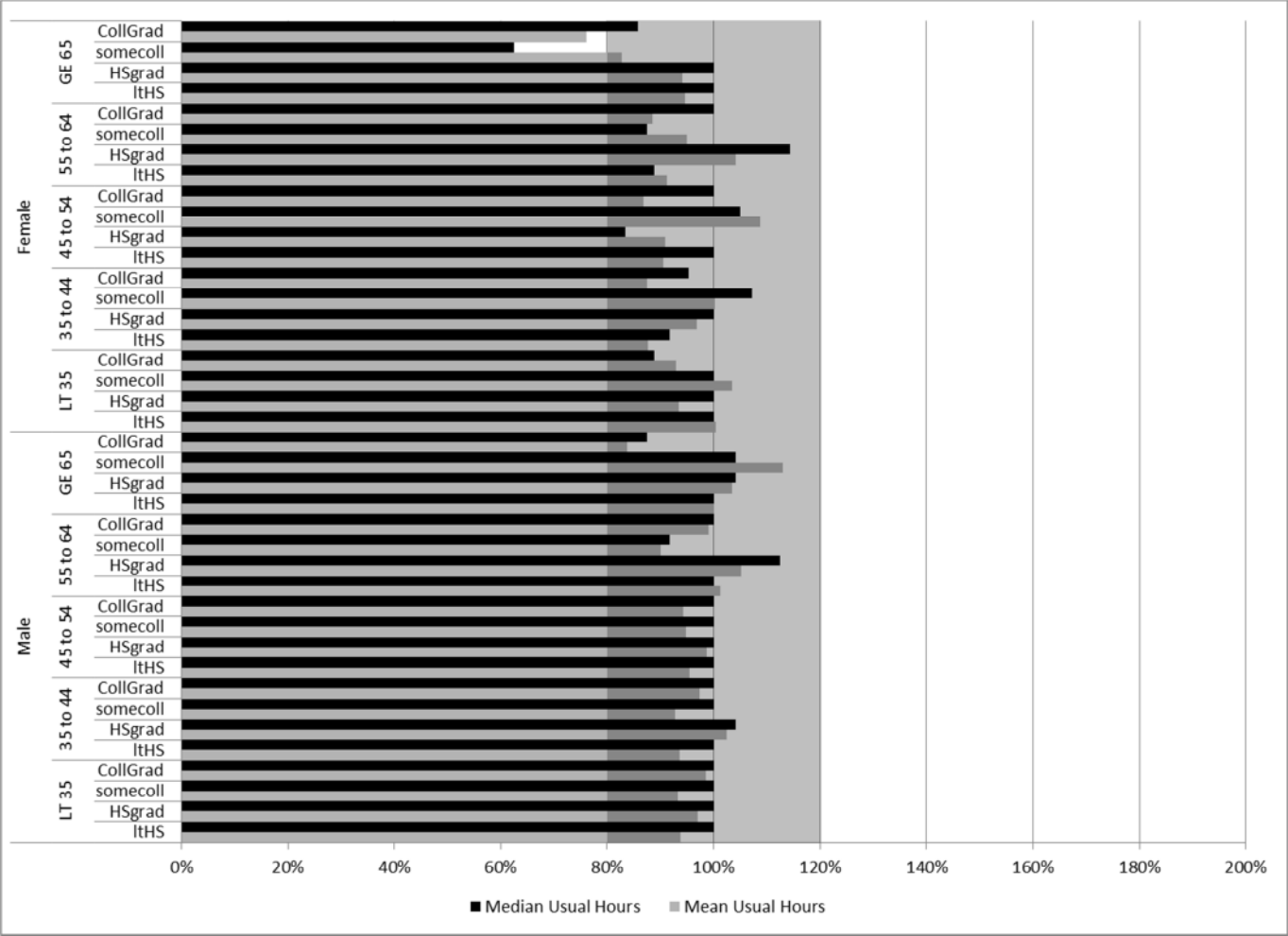


Figure 18 Donor and Recipient Pools for Mexico Time Use Simulation by Sex, Age and Education

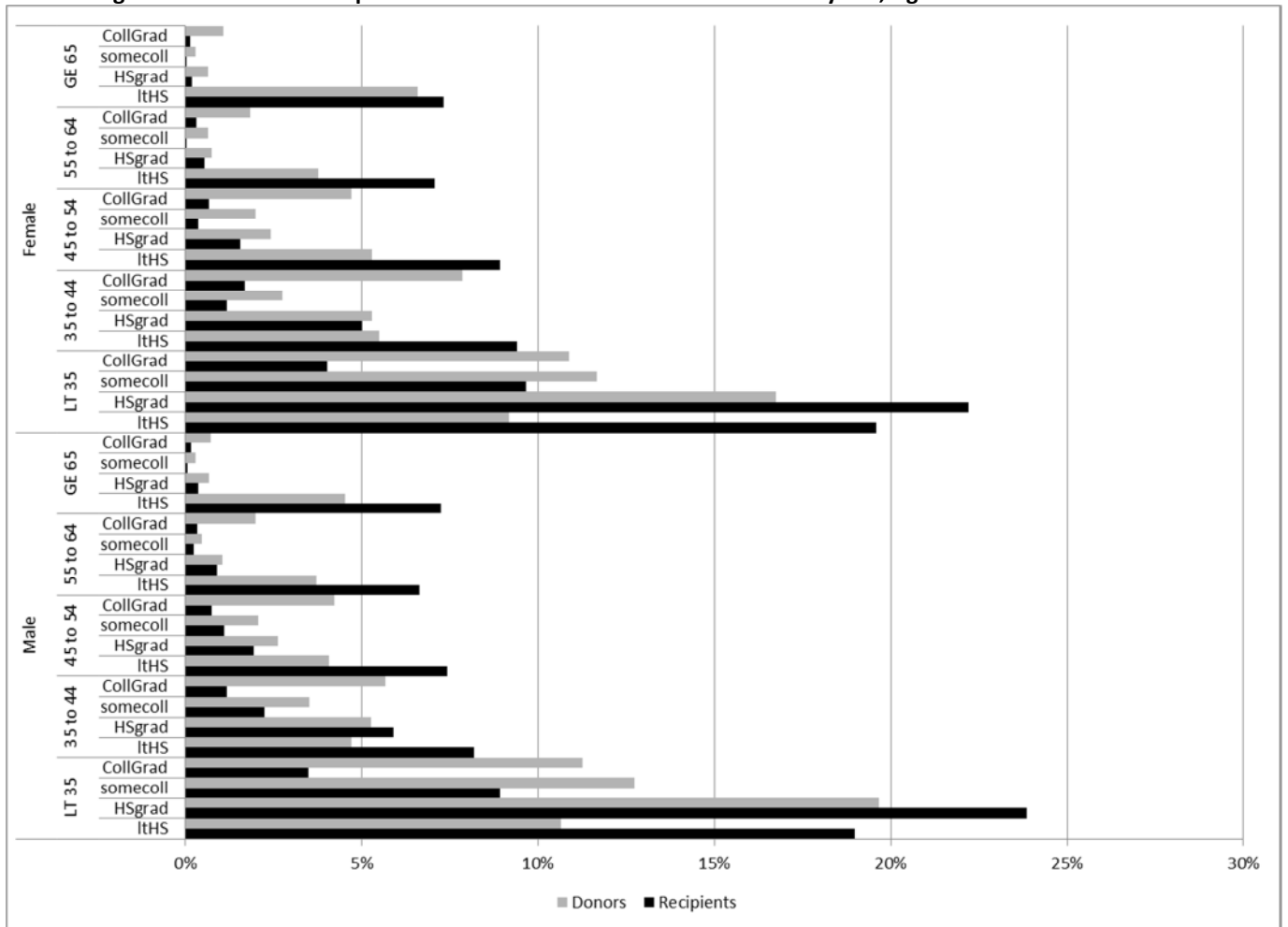


Figure 19 Donor and Recipient Pools for Mexico Time Use Simulation by Sex, Number of Adults and Number of Children

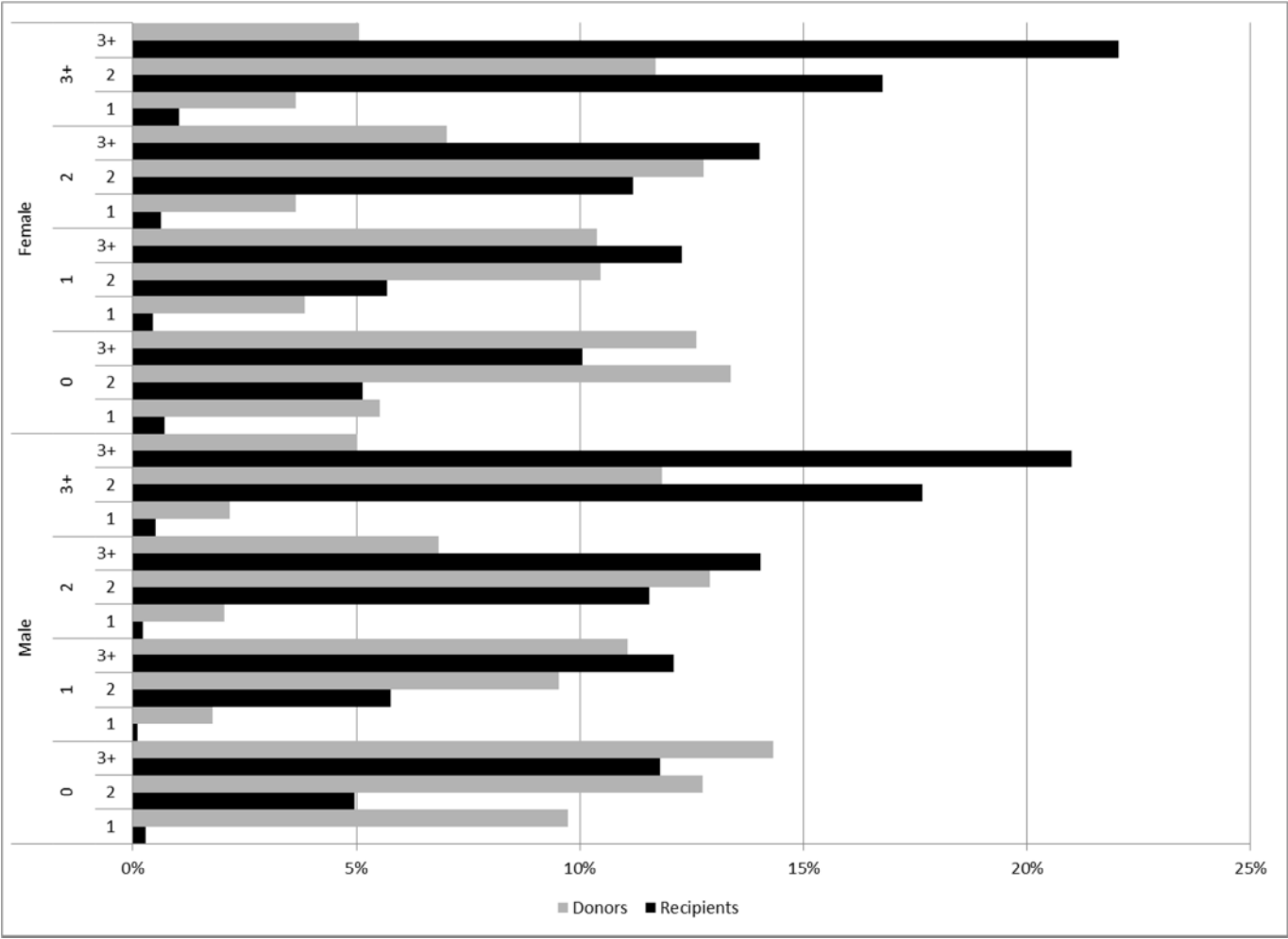


Figure 20 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Mexico
Simulation by Sex, Age and Education

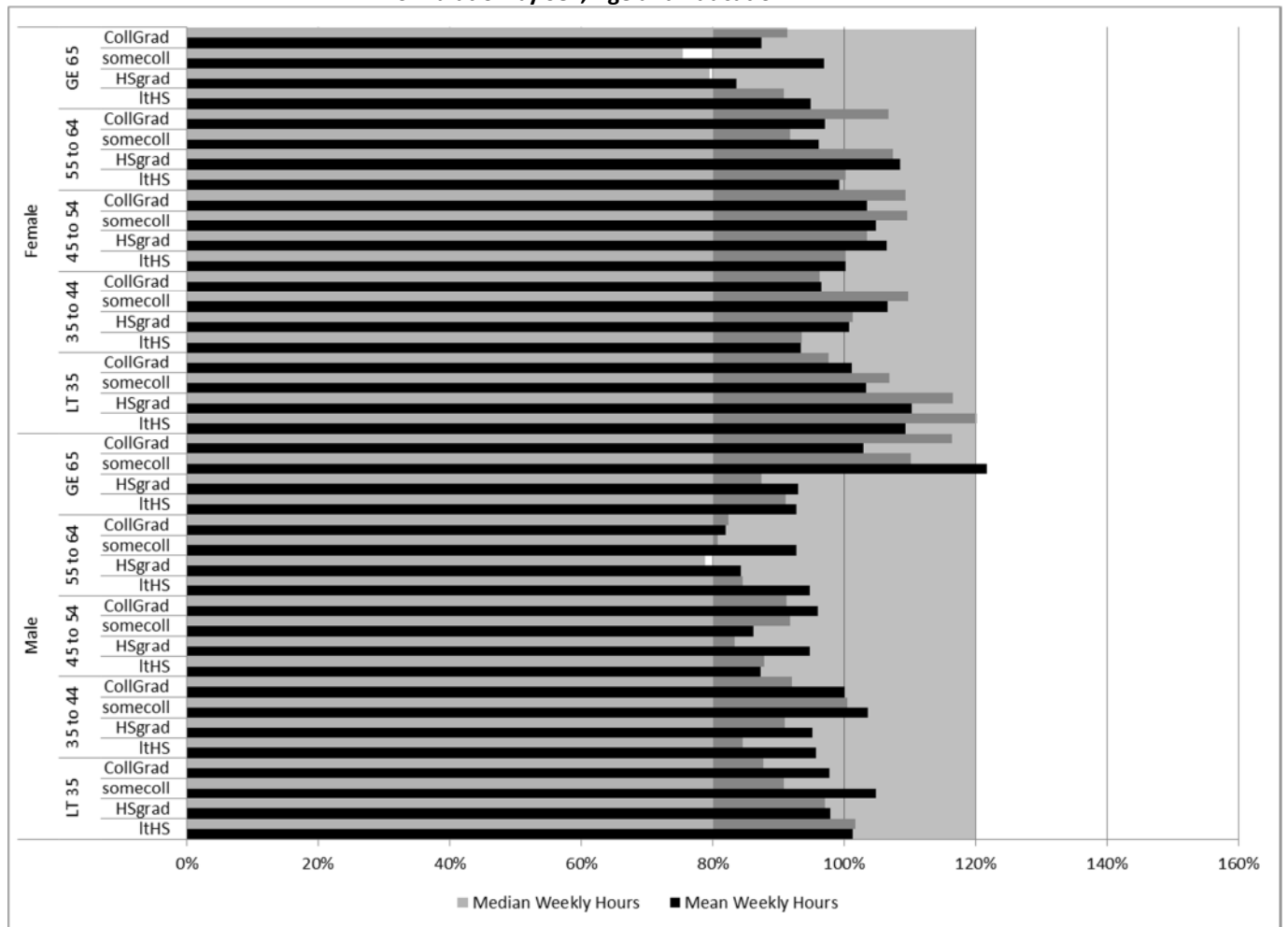


Figure 21 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Mexico Time Use Simulation by Sex, Number of Adults and Number of Children

