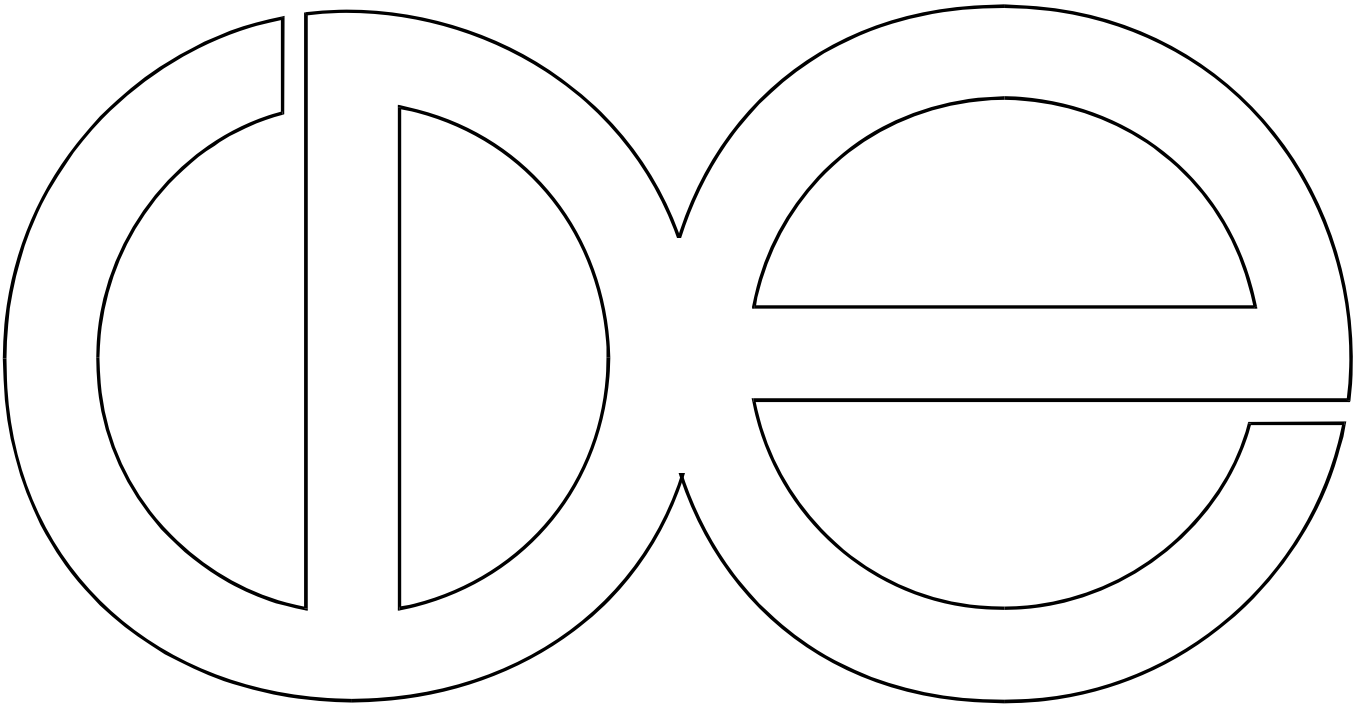
Center for Demography and Ecology

University of Wisconsin-Madison

"Public Use Samples of 1910 & 1920 Puerto Rican Censuses"
Grant Application to the
Department of Health and Human Services
Public Health Service

Alberto Palloni
Francisco Scarano
Halliman Winsborough

CDE Working Paper No. 2007-18



Note to readers: The grant proposal reproduced in this working paper was written in 2000. Please cite as:

Palloni, Alberto, Francisco Scarano, and Halliman Winsborough. 2000. "'Public Use Samples of 1910 & 1920 Puerto Rican Censuses.' Grant Application to the Department of Health and Human Services Public Health Service." CDE WP 2007-18. Center for Demography and Ecology, University of Wisconsin Madison.

I. SPECIFIC AIMS

The **first goal** of the proposed project is to produce micro samples of two Puerto Rican censuses, for 1910 and for 1920. Both of these censuses are comparable to those fielded in the US during the same years. The most important difference is that the 1910 Puerto Rican census contains information on duration of union, not to be found in US or, with a handful of exceptions, any other censuses for that matter. As we show below, information on union duration is useful to apply robust tests on the nature of fertility changes.

We propose to collect and organize the information in ways that are strictly comparable to those used to collect and organize the US based IPUMS data (Ruggles and Sobeck, 1995). The two micro-samples from the Puerto Rican census will be a very rich data set for social and economic historians and demographers alike, as it can be utilized to study social and economic conditions in several regions of the island, to inform the nature of social and economic processes that took place in the aftermath of Spanish domination, and to understand how these affected the composition of the labor force, the organization of the household, the spatial distribution of the population, and the differential growth of cities.

We highlight three potential uses of the census micro samples by social and economic historians. First, students of twentieth-century Puerto Rico will, for the first time, be able to answer questions that have dogged the historiography for decades: Beyond the generalities of migration displacements, in what specific ways did the economy of 'American Porto Rico' with its rapid shift away from coffee planting and toward sugar and tobacco, affect residential and family organization of working groups? How did families cope with the revamped labor market and, in particular, with the substantial entry of rural women into the waged work force? (Acosta-Belén, 1986). Second, using complementary landholding and farm tax data now being analyzed, who among the smallholders of the coastal plains took advantage of the American tariff subsidy to participate in the plantation economy as satellite cane farmers? (Scarano 1990). And what effects did this have on their demographic profile? These are two problems connected with much broader issues in the history of the emergence of newer working classes. The focus is on processes of proletarianization of the peasantry previously engaged in coffee production and with the concomitant transformation into laborers or small land holders.

Third, beyond the easy dismissal by some scholars of certain census categories as "social constructions," what realities of social interaction and family formation lie behind the sharp increase in the percentage of "whites" and the concomitant decline of "blacks" in the early twentieth century?

The answers to these questions will have an influence on new research on Caribbean and Latin American plantation spaces that were quite comparable to Puerto Rico. Economic and social historians of those areas that Wagley aptly called "plantation America" stand to benefit tangibly from the availability of the Puerto Rican materials. It has been a tenet of the comparative literature for some time that in this part of the American tropics the strictures of social and economic life were fundamentally determined by the human ecology associated with one or another of the "after-dinner crops" especially those that were dominant in Puerto Rico: sugar, coffee, and tobacco (Wagley 1957; Steward 1956). In the hands of comparativists, our samples will permit a reexamination of as yet untested postulates about the comparability ofplantation economies and especially of the working populations shaped by them.

In addition, the public data set we propose to produce will include an unusual component, namely, a sample of Puerto Ricans living in the US. Since there is information on length of residence in the US and place of birth, one can compare a number of processes among those living in Puerto Rico and in the US. As of 1910 there were 1,513 Puerto Ricans living in the US, a number that grows to 11,811 in 1920. This is enough to assess the effects of assimilation and to trace the origins of contemporary differentials.

As we discuss at length below, the data we propose to collect and organize will provide rich materials for the study of demographic transformations that took place at the turn of the century. It will be a precious source of information to test theories of fertility decline, the role of public health interventions, and the factors that trigger changes in residential preferences.

To summarize: the first goal of the proposed project is the production of a public good for a variety of scholars, from demographers interested in population dynamics to social and economic historians who study economic and social conditions of the population in developing areas in the aftermath of colonial domination. We envision a micro sample strictly comparable to those produced by the IPUMS project, and similar to those generated in other countries of the region.

The remaining goals of our project are analytic in nature as they relate to substantive

problems to be addressed through the use of the census micro sample. Thus, the **second goal** is to describe levels, patterns and trends of mortality, fertility, nuptiality during the first two decades of the century, and to compare these with those experienced both in the US and in the rest of the Latin American and Caribbean region. The main objective of these analyses is to identify differences and similarities in population dynamics in a society that, unlike those of the rest of Latin America and the Caribbean, experienced direct, first hand exposure to the US influence, but that, unlike the US and like the rest of Latin America, was greatly affected by conditions set forth during the Spanish colonial period. Although the bulk of the Puerto Rican fertility decline occurs after 1940, the period 1900-1940 is one of significant changes. It is during this period that important transformations begin to occur in mortality, fertility and nuptiality. It is also during this period that the traditional household organization begins to crumble, under stress from the combined pressure of new demographic constraints and revamped individual residential preferences, the latter emerging as a result of new ideas and values, the ideological debris of heavy internal and external migration flows. We know of no other detailed source of information at the national level that would enable us to study mortality and fertility processes during a time of dramatic changes, just at the onset of an irreversible demographic transition. There are only a handful of other examples of micro-level data of a population at the moment they begin to experience large scale demographic changes. As mentioned before, the 1910 Puerto Rican census, unlike the majority of censuses ever taken contains information of marriage duration. This permits analyses of fertility that are unique and of unusual robustness (see Methods and Analysis).

The **third goal** is to evaluate the nature and determinants of changes in household composition that took place during the period 1910-1920. During this period Puerto Rico experiences the brunt of the impact that results from incorporation into a US dominated territory. First, the foundations of the economy are overhauled and the transition from a coffee-based to a sugar-based economy is completed in less than twenty years. Second, although still heavily dependent on one export, the Puerto Rican economy begins to develop light manufacture and industry which also lead to important changes in the nature of labor markets. Third, and more importantly, the old social order and cultural suprastructure are weakened partly through the direct influence of new ideas and values, partly through the influence of migrants, and partly because the material conditions change very rapidly. Modernization theory would have us believe that, as a consequence of these transformations, household and family composition must change as well, and they experience the passage from an extended form to a nuclear form. But, did such transition ever take place at all? Was the household composition at the outset of the process identical to what would be expected according to modernization theory? And, if a change did indeed occur, can we identify the influence of changing demographic constraints and separate it from the influence of changing values?

Our **fourth** and final goal is to perform comparative analyses of conditions and behaviors of Puerto Ricans living in Puerto Rico and those residing in the US. Because the 1910 and 1920 US censuses contain information about Puerto Ricans who resided in US, we also propose to extract a sample of US census records pertaining to Puerto Ricans living in the five boroughs of New York city, where about 75 percent of all Puerto Rican migrants resided during 1910-1920. The two micro-samples from the 1910 and 1920 Puerto Rican censuses and the household records for Puerto Ricans living in New York in 1910 and 1920 will provide a unique source for the study of characteristics of individuals and households in a comparative perspective, and supply the basis for testing hypotheses about assimilation and acculturation during the very early stages of what turned out to be a massive process of Puerto Rican migration to the US.

II. BACKGROUND AND SIGNIFICANCE

We now spell out the value added of a Public Use sample of the Puerto Rican censuses and the research problems that it is uniquely suited to address.

II.1. The richness of census micro samples

As of this date the IPUMS project has produced a series of micro samples from the decennial US censuses during the period 1850-1950 (Ruggles and Menard, 1994; 1995). The availability of these samples has given a formidable impulse to socioeconomic history and demographic research that requires detailed individual information and associated community and geographic location identifiers and characteristics. Judging by the number of research projects and publications that rest on data description and analyses based on the IPUMS samples, the project has been a huge success (Ruggles, 1998).

The IPUMS has led to a flurry of research in historical demography in the US, with

particular emphasis on studies of changes in infant and child mortality (Preston and Haines, 1991), marriage and divorce(Ruggles, 1997), household organization (Ruggles, 1996; 1988; Shoemaker, 1992; Kramarow, 1995; Mancunovich et al., 1999), name giving patterns (Watkins and London, 1994) and the occupational structure (Sobeck, 1997). There are countless other scientific endeavors that can be tackled with the data created..

Projects similar to IPUMS have led to the collection and organization of micro-census samples for countries in other regions of the world, including Latin America (Palloni and DeVos,1994) Western and Eastern Europe (Sandefur and De Vos, 1999), Asia (Hirschman, 1990) and Sub-Saharan Africa (Zuberi, 1996). Although the procedures followed are not always the same, the rate of utilization of the data sets is very different, and the main objectives of the associated projects are quite diverse, there is little doubt that research on past trends and more recent developments in the socioeconomic and demographic history of these areas has been enriched as much as has in the case of the US. The census micro data from Latin America have made possibly uniquely rich analyses of fertility, mortality and family composition during a period of two and in some cases three decades. The micro census samples from Africa are of more recent origins but have already allowed interesting analyses of mortality, fertility, household composition, and residential pattern of children (Van de Walle, 1999). In sum, the impetus given to this type of data collection by the experience of a few but well undertaken projects is opening opportunities in other areas of the world where there are unused samples of censuses. Examination of these data sets is a promising research avenue that needs to be undertaken

II.2. Puerto Rico as a unique historical case

The case of Puerto Rico is unique. It is the result of a historical experiment that brought together a blend of contrasting economic, social and cultural influences. Up until 1898 the history of Puerto Rico was intimately connected to Spanish colonial domination, and its most important historical features replicated quite closely characteristics encountered in other parts of the Latin American and Caribbean regions under Spanish colonial rule. Starting in 1898, however, Puerto Rico became a territory legally and administratively dependent of the US. The economic infrastructure was completely revamped, with formidable effects on the social class composition and the regional distribution of the population. Furthermore, the ideological and cultural suprastructure experienced an overhaul that led to a contrast between the cultural and social experience of the island and that of the rest of Latin America. More than figuratively, the political incorporation of Puerto Rico into the US at the turn of the century constituted an extraordinary historical experiment with a number of consequences that historians, sociologists, economists and demographers cannot afford to ignore. The changes experienced and their impact on a number of conditions led also to a social, economic and demographic regime that is in stark contrast to those created in the rest of Latin America. Yet, simultaneously, Puerto Rico was the site of a historical experience quite dissimilar from the one in the US. Simple comparative statistics should enable us to address intriguing questions and problems, related to issues pertaining to social and economic history as well as to demographic theory and analyses. In what follows we elaborate on some of the most provocative of these issues.

II.3. Puerto Rico as a historical singularity: social, economic and ideological transformations in the aftermath of the US invasion

Like its sister colonies of Cuba and Hispaniola (the eastern third of the island of Hispaniola), Puerto Rico's early colonial history under Spain was one of relative isolation from its European metropole and from the circuits of trade developed by France, England, and Holland in the Atlantic. Until the late seventeenth century, and after a period of marked depopulation of its Amerindian inhabitants, this fertile island of some 3,500 mi^2 remained sparsely cultivated and was populated by only a few thousand colonials, and a majority population of African slaves. As economic activity sprouted all around it in the eighteenth century, with the rise to prominence of the French and English "sugar islands," Puerto Rico slowly began to develop along the "developmentalist" lines that the new Bourbon monarchy lay down for it, through policies meant to augment commercial production of sugar, coffee, and tobacco, and to increase the supply of African slaves. Although it remained, primarily, a strategic colony for the defense of Spain's sprawling empire in the Americas, by the second half of the eighteenth century Puerto Rico's turn toward commercial agriculture seemed assured by Spain's determination to emulate its European rivals in the creation of export-oriented colonies. In parallel to the economic changes that began to unfold in the latter 1700s, the population began an upward ascent that would not subside until the latter decades of the twentieth century. By 1800, the island's population, already near the 150,000 mark, was increasing at an annual rate of more

than 3.2 per cent per year.

The nineteenth century brought a profound transformation in the nature of Spanish colonialism and, concomitantly, in local economic and social structures. As the Spanish empire crumbled and the neighboring English and French "sugar colonies," for various reasons, declined--Haiti, in particular, was destroyed by a slave revolt and the war that followed it--Puerto Rico and Cuba gained a position of leadership that they had never before enjoyed. Although overshadowed by Cuba's "sugar revolution" of the 1830s through the 1850s, the Puerto Rican sugar plantation sector boomed starting around 1815 (Scarano 1984). By the middle of the nineteenth century, Puerto Rico stood as the second largest sugar producer in the Caribbean (after Cuba) and the second largest supplier of the United States market. Gains in the export sector had exacted a large price in terms of human freedom, however. Between the 1790's and the 1840's a large number of Africans (60-80,000, by our calculations) were imported as slaves to labor on the sugar *haciendas*. Then, as the import of slaves slowed to a trickle and finally stopped around mid-century, many more nominally free peasants, called *jíbaros* by the island elites, were forced under penalty of law to obtain employment in the export sector and to observe a strict labor discipline. Of little consequence to the Spanish colonial state was the fact that the main beneficiaries of these policies--the sugar hacendados--were a group composed largely of foreigners, many of them recent immigrants from Europe and North America.

One aspect of this history that informs our work with the 1910 and 1920 censuses is the presence of a robust, if increasingly threatened, island peasantry. Unlike the classic Caribbean "sugar colonies" of the eighteenth century, in Puerto Rico a large peasant sector remained viable throughout the process of island integration into the world market, which spanned the entire nineteenth century. A rugged mountainous interior, a benign climate and rich soils that allowed for subsistence yields from relatively small labor inputs, and a history of easy access to land, mostly through squatting rights, all contributed to the strength of this peasant group. This *jíbaro* peasantry constituted, at all times before the turn of the twentieth century, a majority of the island population. The crucial weight of this peasantry in demographic terms, as well as in class relations and colonial politics, clearly stood the Puerto Rican case apart from that of Cuba, which was then a much more urbanized society and where slaves constituted a significantly higher proportion of the population. Hence, even while the percentage of land devoted to export crops grew steadily throughout the 1800s, at the end of this period the vast majority of Puerto Ricans were still rural dwellers, and most of these had access to a small portion of land for subsistence activities. As in similar situations throughout the world, many of the latter mixed independent farming with seasonal employment in haciendas (Steward 1956; Bergad 1983).

The peasantry would not survive the forces unleashed by the U.S. invasion, however. The rise of coffee cultivation in the mountainous interior after ca. 1870, followed by a vigorous development of the coastal sugar plantations after the US invasion of 1898, accelerated the proletarianization of Puerto Rico's population. The proportion of arable land devoted to food crops sharply declined, especially after the island was incorporated into US economic space via the enactment of free trade between metropole and colony in 1900. Pastures along the coastal plains were plowed under, and the ensuing cane fields incorporated into the huge landownings of "corporate land-and-factory combines," as Wolf and Mintz have described them (1957). A large portion of the economically active population found employment in sugar (and in its handmaiden in the economy of "American Puerto Rico," tobacco). By 1930, the sugar industry alone employed nearly 100,000 people, our of a total population of 1.3 million. As two American critics graphically put it in 1928, Puerto Rico had become "a community of agricultural laborers" (Diffie and Diffie, 1931).

The sugar boom prompted rapid shifts in residential and employment patterns. During the aegis of coffee (ca. 1870-1898), an era of bustling activity for highland *municipios*, many coastal Puerto Ricans migrated, whether seasonally or permanently, to the interior regions, where jobs were more abundant. As the coffee industry declined steeply in the first few years of US rule and as the coastal plains came back to life, this trend reversed. The coffee regions were the first to bleed migrants, not only to the new centers of the sugar economy along the coast, but also to locations beyond the island: Hawaii, Cuba, Panama, Ecuador, and, of course, New York City and its environs. At this time, young men often took leave of their households for long periods, living in barracks provided by local sugar corporations during the harvest. Others, such as the 5,000 Puerto Ricans who left the island for Hawaii in 1900-1901, did so at the prompting of employment agents sent there specifically to recruit the young and unemployed (Rosario Natal 1983).

II.4. Puerto Rico as a demographic singularity: mortality, nuptiality, and fertility

It is hard to imagine that such a historical setting could evolve without undermining the prevailing demographic regime. Prior to 1898 Puerto Rico shared many features typical of other countries in the region exposed to Spanish colonial domination. First, its economy was primarily extractive and almost completely dependent on exports of a single agricultural commodity (coffee) produced mostly in areas on the western part of the island. Second, political institutions were stunted and dominated by a wasteful, inefficient, and suffocating bureaucracy that kept political participation to a minimum but that constituted an important source of employment for a minority of privileged individuals. Third, infrastructure was scarce and mainly concentrated in the urban areas. Transportation routes were poor, internal communications difficult, and the availability of services to the population virtually nonexistent or entirely dependent on voluntary organizations associated with the Catholic Church. The dominance of a conservative, obscurantist and Catholic-inspired ideology was virtually complete.

The prevailing demographic regime was not unlike others in the rest of Latin America during the same period of time. Mortality fluctuated sharply in response to epidemics (particularly outbreaks of cholera) and to natural disasters associated mainly with hurricanes. For most of the XIXth century, background levels of mortality sustained life expectancies of around 32 years and infant mortality levels exceeding 200 deaths per 1,000 births (Vazquez-Calzada, 1988). Although there was a slight increase in life expectancy during the second half of the century, little progress was made until the 1900's and crude death rates remained around 28-30 per 1,000, a very high level even for a population where more than half of its members were younger than 15 years of age.

Parish records and a number of indirect estimates suggest that the crude birth rates for the period 1800 to 1900 were heading downwards, from levels around 55 per 1,000 to about 45 per thousand. As of 1900 the General Fertility Rate was of the order of 200 (per 1,000 women of childbearing age), a level that translates into the equivalent of a TFR of about 6.8 children per woman approximately. As in the case of mortality, fertility rates began to decline early on, during the second half of the XIXth century but the rate of decline remained sluggish until the first two decades of the XXth century.

Finally, the nuptiality regime was characteristic of what has been called the 'consensual belt' of Latin America, a region stretching North of Ecuador and Paraguay and spreading over the Caribbean, where consensual unions represent more than one third of all unions, and where the prevalence of unmarried young women (and young men) is higher than 10 percent (Palloni and DeVos, 1999).

Puerto Rico becomes a singularity in the continent soon after the US invasion and the establishment of US rule in 1898. Although the changes in political institutions and national administration were momentous, nothing parallels the traumatic transformation of the economy and the subsequent changes of values, norms and individual outlook that accompanied the uprooting of old social classes and the establishment of a new social and economic order. The American influence was to be felt in every sphere of life, including those domains that had a direct impact on health and mortality, fertility decision making, and residential preferences. The most important change was the radical transformation from a coffee-based to a sugar-based economy, one controlled and regulated by large American corporations catering to international markets serviced by these companies. This process led to the transformation of a large fraction of coffee producers into dispossessed agricultural laborers and later, industrial workers. This process is analogous to the development of a proletariat out of the debris left by land enclosures in Western and Northern Europe. And, we argue, its demographic effects may have been not too dissimilar. The massive effects on the structure of the labor force were paralleled by important technological implications for it led to the introduction of new cultivation and processing techniques and plants, as well as to large public works to improve transportation and communications. Despite these transformations the overall nature of the economy changed very little. Indeed it continued to be dependent on exports and tax revenues associated with a single commodity whose price, as those of coffee before, experienced sharp fluctuations in international markets. A consequences of the change was the decay of coffee producing regions, massive migration from the Western interior to the North Eastern interior and coastal regions, and rapid urbanization of key centers of commerce in the North East.

The US intervention changed the mode of production without transforming it into something qualitatively different from or superior to those in place in other peripheral economies in Latin America. As late as 1940 the economic order could not have been too different from that

prevailing in Dominican Republic or Panama. Yet the social order had been completely transformed and its traditional ideological foundation replaced. The declining domination of the Church is the first salient feature. A second is the pervasive influence and diffusion of a forward ideology, tolerant and encouraging of new and modern ideas, and assigning liberties and responsibilities to individual as opposed to corporate groups. Thus, the experience of Puerto Rico is one where the material basis for a shift in reproductive regime is at best weak whereas its ideological foundation is very strong. How is this manifested in the demographic regime of the epoch? What is the nature of the social differentials across classes and regions?

The conditions experienced by Puerto Rico in 1900-1920 are very different from those that explain the early transition in countries such as Argentina and Uruguay, where the demolition of the old order occurs during the same time interval but its roots are deep in economic not ideological or social transformations. Paradoxically, the situation lived by Puerto Rico in that period of time is more akin to that being lived today by peripheral countries such as Bolivia, Brazil, Peru, Ecuador where, despite economic backwardness, the high fertility regime is no longer sustainable.

i. The collapse of the old demographic order

How did demographic regime change after these transformations? We argue that under the pressure exerted by US economic and ideological influence, the main pillars of the demographic order collapse and a demographic transition gets under way.

☛Migration: To begin with, new migration flows emerged. Above all, there was increased migration away from the coffee producing areas of the South West toward the North-East and the newly opened sugar estates. Second, the beginning of what turned out to be a tide of migrants toward the US takes place soon after the US invasion of the island. According to historical records, in 1900 about 1000 Puerto Ricans left the island bounded for the US (Rosario Natal 1983) Over the next forty four years the total number of migrants grew to about 90,000 with little return migration before 1930 (Vazquez-Calzada,1988). There is no question that those who migrated early in the century must have been a very select group, perhaps very different from the average migrant after World War II. They are forerunners and even a descriptive study of their characteristics should help us shed light on the mechanisms that generated the migratory flows. This is an important reason to propose the comparison of characteristics between Puerto Ricans residing in the US in 1910 and 1920 and those who remained in the island.

☛Mortality: Admittedly, mortality had already declined somewhat before 1900 but it did so at a sluggish pace, with frequent setbacks due to natural catastrophes (hurricanes) and periodic reintroduction of epidemics, particularly cholera and typhoid fever. The pace of decline accelerates in the period 1900-1940 as mortality levels tumble and the CDR drops from average levels close to 28 per 1,000 to about 22 per 1,000. The real decline is probably higher since the estimate of the initial level is contaminated by the effects of the 1918-19 influenza pandemic. Infant mortality declined from levels close to .200 in 1900 to about .140 in 1920, a 30 percent decline. Life expectancy is estimated to have increased from 32 years in 1900 to about 40 in 1930. This marks the initiation of what turned out to be an irreversible process. Why did mortality decline during this period? How can infant (and child) mortality have been reduced by 30 percent after years of little if any progress? Historians suspect that the US invasion led to transformations that worked in favor of health levels. To begin with, large-scale interventions to eradicate infectious illnesses began just during this time. It is known that, in addition to their direct impact, these interventions have spill-over effects as they are vehicles to diffuse ideas about what infectious diseases are and how best to protect against them. If so, mortality could decline faster in areas directly affected by the interventions and_ less so in others. There is, of course, the matter of medical innovations and the utilization of new knowledge, mostly developed after the introduction of germ theory. But antibiotics were not yet generally known and only a few vaccines were administered to the general populations. Changes, then, might have come from other sources. Some suspect that social and economic changes must be responsible: the economy takes a turn for the better and there are important improvements in families' standards of living. Literacy alone grows from a low of about 20 percent in 1900 to about 70 percent in 1940. Judging by figures on the percentage of the population in school age that actually attended school, it appears that about 70 percent of this improvement took place before 1920. The bulk of the increase in elementary school attendance therefore occurred between 1900 and 1920, and should be better understood with the proposed census samples. If this is so, we should certainly see the association across subgroups and regions: Which subgroups fare better? What regions are the main forerunners of this mortality decline?

Which are the followers and which remained stagnant?

With the exception of Cuba, the impact of truly exogenous factors occurs much later in other countries of Latin America. And, with the exception of Argentina and Uruguay, the effects of improvements in standards of living in the continent are not felt before exogenous interventions have a chance to operate. Thus, Puerto Rico's mortality transition, at the very least its initial phases, is unique in the region: propelled by eradication campaigns to reduce mortality associated with malaria, uncinariasis (hookworm disease), yellow fever, typhoid, and TB, the country also experiences significant changes in standards of living. Both forces work jointly to reduce mortality to levels that the rest of the region attains after World War II.

Because we can time precisely the exogenous interventions, identify the geographic space where their direct effects must be the largest and, simultaneously, assess individual characteristics that serve as control for tastes, preferences, information and access, we have a unique opportunity. Indeed, there is the tantalizing possibility that through examination of child mortality over two decades covered by the censuses, with all the detail retrievable from census records (see below), we will actually identify the magnitude of the impact associated with exogenous interventions, an elusive accomplishment we managed to realize only at the expense of obtaining highly fragile estimates (Preston, 1976; 1980; Palloni and Wyrick; 1981).

☛Fertility and nuptiality: as in the case of mortality, patterns of change in fertility and nuptiality in Puerto Rico could not have been more different from those prevailing in the rest of the region. We suspect that, as it happens in the case of mortality, changes in fertility and nuptiality take place as an outcome of changes induced by exogenous factors. First, in theory at least, the changes in child mortality could have had a short-run, albeit muted, impact on fertility. One would expect this to be reflected in changed levels of fertility during the period 1910-1920 precisely in those areas with the largest decline in child mortality. For a number of reasons it is unlikely that this effect was too large (Palloni and Rafalimanana, 1999).

Second, three series of events signal the early onset of strong secularizing influences which could have played an important role in massive fertility decline that begins after World War II . The first event is the legalization of divorce right after the US invasion. Although there are no divorce statistics until 1932 --the micro samples from the census themselves will be useful to estimate proportions divorced by age-- Puerto Rico is the first country in Latin America to legalize divorce, an event that signals the weakening effects of the Catholic church in affairs of the family.

The second set of events are also related to nuptiality. The fraction remaining single increases after 1910 and so does the mean age at marriage. Could these changes be due to the large-scale migration from the Western interior to the coastal areas prompted by the shift from coffee-based agriculture to sugar-based agriculture? Are they generalized to only some social groups? And if so, which ones? The fact that marriage patterns changed steadily until about 1950 is not at all common in the other countries of the region. This presents us with a conundrum: why? Exactly what led to nuptiality changes of these magnitude? And what effects did they have on general fertility? We can use the data from the micro census samples to identify the social groups where the changes are sharpest, and establish whether they are related to increased urbanization as well as the massive transformation of small proprietors into landless laborers.

The third set of events is related to the creation of a movement in support of family planning and practice of contraception. According to historians, the public debate may have posted headlines starting as early as 1920 after which it gained formidable -momentum. But it is suspected that even before then there was a fairly widespread subterranean movement supporting and promoting the practice of contraception in various forms, including withdrawal, condoms and vaginal wash, the most commonly known methods at the time. There is a startling synchrony between these events and the pace of fertility decline. In fact, as noted before, measures of general fertility indicate that a slow decline started right after 1900. This decline continues gradually until 1945 approximately, a year when the TFR begins a free fall that will take it from 4.5 children per women to replacement levels in a very short period of time. The gradual decline between 1900 and 1945, however, masks a much sharper decline in marital fertility rates for, although the general fertility decline during 1900-1920 was about 20 percent of the total decline experienced between 1900 and 1950, this takes place while the nuptiality regime contracting. The aggregate evidence available suggest that during this period the age at marriage was increasing and the proportion of women marrying was decreasing. This was partly a result of reductions in consensual unions. Simultaneously, divorce rates must have been increasing as well. These three forces, later

marriage, lower proportion marrying, and decreased duration of first marriages due to divorce, combined with gradually declining general fertility measures, strongly suggest that marital fertility rates must have been declining much faster. This means that control of family size could not have been confined to a few pockets and may have been fairly widespread even during the early part of the XXth century.

There are a number of related questions that, if clarified, could shed light on the mechanisms that lead to the onset of fertility decline and give to the latter an overwhelming momentum: Was it indeed the case that marital fertility was coming under stricter control as early as 1910 and 1920? And whose fertility was declining? Were all social groups and regions affected or only some forerunners, perhaps the same whose mortality had been declining fast during the same period of time? How similar where these innovators to those who migrated to the US? Were their fertility levels and patterns similar to the levels and patterns of those who resided in the US? What are the main factors that account for these changes? Superficial indications retrieved from age patterns of fertility in the forties and fifties suggest that fertility limitation was decidedly parity-dependent. Was the marital fertility decline between 1900 and 1920 also parity dependent? Or. alternatively, was it an event that affected all women alike, regardless of birth order? How much of the observed fertility decline between 1900 and 1920 was due to changes in marriage and how much to changes in marital fertility? Were the contribution of marriage and marital fertility the same across regions and social classes?

II.5. What can we learn from PURIPUS?

i. Material conditions, ideologies and demographic patterns. There is a very sophisticated anthropological literature that explores the connection between different crop systems and their attendant work patterns, on one hand, and the workers' cultural adaptations, ideologies or 'mentalities' in Puerto Rico. This literature, based on historical reconstruction and ethnographic observation in Puerto Rico, lays out a neat model of how, in a poor, agrarian, dependent context, different crops and technologies of production came to shape a variety of cultural adaptations. But this literature says nothing about demographic correlates. In this project we intend to perform the first systematic investigation into the demographic correlates of these ecological variables. The usual presumptions are that sugar cane "devoured" people, turning them into landless rural proletarians with loose family relations (usually via consensual unions), open to revivalist Protestantism and spiritualism, and politically quite on edge, as they were usually responsive to Socialist preachings. These folk, living on coastal plains, descended from Black slaves and exhibited fairly flexible ideas about race and its significance in social interaction. On the other hand, highland coffee workers and their families, usually of a lighter phenotype and demonstrably descended from Spaniards, and Catholic to the bone, held on to some land in a system permeated by relations of clientelism, had larger families (and usually greater stability in unions), held very strong and prejudicial views on racial relations, and voted along the clientelist lines that defined many of their other social relations. Finally, the tobacco workers, located in an intermediate geographic zone between the coffee and sugar areas, were the most prosperous and "modern" of the three--their land holdings were larger and they usually had title to them--and exhibited cultural patterns and behaviors that mixed some of those of the coffee workers with some of those from the sugar areas. Can one verify the existence of demographic diversity that agrees with the above mentioned ideological and material differences across these groups? This project presents us with the first opportunity to analyze fertility and mortality correlates of workers' ecological and cultural adaptations. These censuses capture the "moment" at which this population was being subjected to the greatest pressures for change, launched by a huge amount of U.S. investment into export agriculture. Furthermore, the Puerto Rican findings could serve as a baseline from which to assess similar relations worldwide.

ii. Adaptations among Hispanic migrants to the U.S. A significant number of Puerto Ricans have been counted in the U.S. censuses since 1900, but beginning in 1920 the number rose to very significant levels, especially in and around the five boroughs of New York City. By 1930 the number of Puerto Ricans in the City had grown immensely, since the valve of European immigration had been largely shut off by legislation. The PURIPUS for Puerto Rico would be of immense value to students of U.S. immigration to understand how population characteristics and social behaviors are modified, even as in many ways they remain the same, between the origin society and the receiving society.

The following is a sample of questions we can address:

✦To what an extent did the proletarianization of the labor force induced by the destruction of the coffee plantation lead to different mortality and fertility regimes? Was the demographic response

similar to the one experienced during the protoindustrialization period in Western Europe?

✦Did the beginning of the fertility and mortality decline occur in several social strata or social groups simultaneously, or are there identifiable forerunners as hypothesized for Western Europe (Livi-Bacci, 1986)? If there are forerunners, micro census data will allow us to get as close to them as sample sizes permit. We can thus learn a great deal about the onset of fertility and mortality decline and about the mechanisms involved +For many industrialized countries we have lost forever the opportunity to understand how societies bring fertility and mortality under control. For some developing countries, such as those with late transitions, we are now witnessing the process as it occurs. If there was a diffusion process early on In Puerto Rico, is this similar to what we are observing today in countries such as Brazil or Mexico? Can we at least find some support for the idea that there are universals governing processes of fertility decline, transcending time and space?

✦Puerto Rican who migrated to the US afford us an opportunity to study two different process. The first is the degree to which they become assimilated to US culture. The availability of length of residence in the US as key datum to detect differences in the patterns of behavior by duration of residence. But we can also investigate the reverse side of the migratory process: are recent migrants selected from a particular social class with distinctive and particular behaviors, or are they simply a random sample of the labor force in Puerto Rico? And, if they are different, who are they? How different are they from forerunners among whom morality and fertility began to decline first?

III. Preliminary Studies

III.1. Alberto Palloni is completing a manuscript on the history of the population in about 20 Latin American countries during the period 1850-1995. This manuscript employs multiple data sets, including vital registration, nationally representative surveys, and micro-census samples drawn from decennial censuses for all but a handful of countries for the period 1970-1990. The analyses cover several themes including the secular decline of fertility and mortality, changes in nuptiality, and the relation between fertility and mortality changes. In addition, there are extensive tests of competing explanations for the decline of fertility and mortality, such as those invoking the importance of diffusion of ideas, and those that emphasize changes in socioeconomic conditions. Finally, the project includes extensive description of socioeconomic differentials in fertility and child mortality throughout the period 1920-1995. This manuscript is the product of work that spreads over fifteen years in which Palloni, supported by NICHD, NIA and NSF, produced many papers on the subject and developed a number of techniques to estimate fertility and child and adult mortality levels and patterns. In addition to systematizing vital registration and published censuses information that stretches back to 1850, Palloni (in collaboration with two former students, Glen Ulmer and John Marcotte) produced standardized files from the microcensus samples of decennial censuses fielded in the region during the period 1970-1990. These samples were processed, cleaned, and standardized and are available on line at the web site of the Center for Demography and Ecology. Although the actual extraction of the samples was coordinated by CELADE -and followed procedures that are not always comparable to those applied in the IPUMS proj ect-the cleaned microdata have proven to be invaluable sources of information to estimate child mortality (using Brass's techniques), nuptiality (using Coale-McNeil models), age-specific fertility rates (using own-children procedures) and fertility parameters (using Coale and Trussell, and David-Sanderson CPA models). Numerous verification studies and consistency tests confirm the high quality of the data contained in the sample.

Palloni and Susan De Vos are preparing a second manuscript on the transformations of families and households in Latin American and the Caribbean. This work utilizes sources such as the micro-census samples referred to above, the household files from the World Fertility Survey (WFS), and household files from the Demographic and Health Survey, as well as a number of surveys carried out separately by each country. Although this work rests on the studies carried out by Palloni, the study of families and household adds new components. In particular, an important domain is the test of hypotheses about the role played by changes in preferences and changes of demographic conditions in the determination of living arrangements. In order to falsify hypotheses, the work utilizes microsimulation models of the family such as SOCSIM. (Wachter et al., 1977) and MOMSIM (Ruggles, 1987)

Some of the work referred to above has already. appeared in print. We cite here the most important references. Palloni, Massagh and Marcotte (, 1984); Palloni and Kominsky (1984); Palloni and Rafalimanana (1999); Palloni and Tienda (1986); Palloni and Tienda (1992);

Palloni, 1990 (Annals); Palloni and Devos (1989) Palloni and De Vos (1999); Palloni and Arias (1999); Palloni (2000); Palloni and Hill (1997); Palloni, Hill and Pinto (1996); Palloni and Lu (1999); Palloni et al 2000.

III.2. Francisco Scarano: Francisco A. Scarano's is a historian with considerable experience in the history of Puerto Rico in its Caribbean and Latin American contexts. His research interests and past work cover a broad set of issues, from demographic to political, and cultural history, as well as historiography and historical synthesis (Scarano 1993, Scarano 1996, Scarano 1999). Some of his earliest work, leading to graduate degrees at Columbia University, hinged on his discovery at the Archive of the Indies in Seville, Spain, of a series of 23 annual population *padrones* or "censuses" from the late eighteenth century. He used those simple but valuable compilations to explore the underlying causes of the island's population "explosion" of the second half of the eighteenth century and the early years of the nineteenth, when annual growth rates averaged more than 3 per cent (Scarano 1974). His analyses of the *padrones* for 1776-1803 suggested that stepped-up immigration, a decline in mortality, and a rise in fertility jointly determined the population upturn. He further suggested that rapid changes in landholding practices and structures were ultimate explanatory factors of those changes, as the mostly landless, roving peasant population obtained greater access to land and began to turn to more intense cultivation practices, which replaced an inefficient system of land exploitation (see also Scarano 1996).

Scarano has also worked on Puerto Rican historical demography in the nineteenth and twentieth centuries. In a monograph on the sugar economy in the nineteenth century he advanced the argument that slave labor (supported by an active slave trade) was much more important to the rise of that crop to island-wide prominence than previously thought (Scarano 1984). This research also yielded an estimation of the number of slaves imported into Puerto Rico in the nineteenth century. A parallel study suggests the possibility that, unlike most slaveholding societies at the time, this Spanish Caribbean colony probably experienced a positive rate of natural growth of its slave population before the 1855-56 cholera epidemic, which killed about 30,000 enslaved persons, wiping out most of the gains of the evolving reproductive regimen (Scarano 1986).

In preparation for this project, Scarano has already coded data from three territorial barrios (subdivisions of municipalities) of the 1910 and 1920 Puerto Rican censuses. The main purpose is to test some widely held tenets of Puerto Rican historiography regarding the nature and speed of peasant proletarianization, and the effects of economic changes (the rise of sugar and tobacco and the decline of coffee) on family structures, migration patterns, residential tendencies, fertility and nuptiality The three selected barrios are those where anthropologists Eric Wolf, Sidney Mintz, and Robert Manners conducted ethnographic and historical research in the 1940s (Steward 1956); hence, considerable observations exist on all of the above processes, which should now be refined on the basis of census data.

III.3. Halliman H. Winsborough is Emma Welsch Conway-Bascom Professor Emeritus in Sociology. Winsborough retired from the University of Wisconsin, Madison in January of 2000 and became Interim Director of the Inter-university Consortium for Political and Social Research at the University of Michigan, Ann Arbor. Winsborough's appointment was through August of 2000. He has returned to residence at U.W. Madison.

From the point of view of this project the most important thing about Winsborough's forty-year career in Sociology and Demography is his experience as a builder and developer of capital goods for social science research and on the development and use of Public Use Sample Files from the U.S. Census. In the mid-1960's, Winsborough, in collaboration with Joseph Spengler, Reynolds Farley, and Kurt Back established the Population Studies Center at Duke University. The computing facilities available to the Center made it possible for Winsborough to present one of the first papers to PAA using the initial 1960 1-in-a- Thousand sample from the 1960 Census of Population. In 1967 Winsborough moved to the University of Wisconsin, Madison to become Director of the small Center for Demography and Ecology that Norman Ryder had begun there. During Winsborough's tenure as Director, CDE grew to very nearly its present size and structure. Winsborough was PI for one of the earliest Center Grants from NICHD, one that acquired a small mainframe computer for the Center among other things. That machine was probably the second earliest computer under the control of an academic social science department in the United States, the first being at Johns Hopkins under the control of James Coleman and Peter Rossi. Its purpose was primarily to process the rather ample . Public Use and Summary Tape

Files made available for the 1970 U.S. Census.

In the mid 1970's Winsborough, in collaboration with Karl Taeuber and Robert M. Hauser, began the process of creating public use sample files from the 1940 and 1950 Censuses of Population. This project, funded by the National Science Foundation, is the only example of non-current PUMS files being created from manuscript censuses records which where not in the public domain. The files were constructed in collaboration with the Bureau of the Census and was completed in the early 1980's. The files have been in continuous use ever since, serving as a backbone of time series and repeated cross sectional analysis of the growth, composition, and distribution of the U.S, population. In the later 1980's Winsborough was chair in Sociology and in that role was able to help create the Wisconsin Survey Center with James Sweet and Charles Palit. A notable feature of this survey center is its operation of a continuous national sample of the U.S. population, one in which each day's telephone interviews are a random sample of the population.

In the early to mid 1990, Winsborough established the Social Science Computing Cooperative at the University of Wisconsin. The organization is a consumer's coop comprised of the Center for Demography, the Institute on Aging and Adult Live, the Department of Sociology, the Department of Rural Sociology, the Institute for Research on Poverty, the Social Systems Research Institute (the research component of the Department of Economics,) and the Department of Political Science.

IV. METHODS AND ANALYTIC PROCEDURES

This section of the proposal consists of three parts: the first describes the nature of the PURIPUS samples, the second discusses procedures and methods to create the samples, and the third defines methods of estimation of demographic parameters and formulates models for testing conjectures.

IV.1. Nature of microsamples

IV.1.1 General strategy

Our goal is to generate files that are as similar as possible to those in the larger IPUMS data file. There are three reasons for wanting to conform to the IPUMS standard as closely as possible. First is that we plan for the Puerto Rican samples to become a part of the IPUMS data file in order that they may be used for comparative purposes by other scholars studying the United States. The second reason for using the IPUMS standard is that an important motivation for our project is to compare the demographic characteristics of Puerto Ricans living in the island with those living in New York State. The latter group will be sampled as a subcontract to the IPUMS group at the Minnesota Population Center using their standard facilities. Finally, the funding of IPUMS-International through a National Science Foundation Enhancing Infrastructure for the Social and Behavioral Sciences grant forecasts the availability of a much larger and more diverse collection of census micro samples from around the world. Although their construction must necessarily be more heterogeneous than are the samples for the U.S., we believe that the IPUMS standard will be the major point of reference. Maintaining the IPUMS standard, then, makes it most likely that in the future the Puerto Rican samples can be compared with those of a large number of other countries at comparable times or comparable levels of economic development.

The best way to insure comparability of the Puerto Rican sample with the IPUMS standard would be for the Minnesota Population Center people to prepare it. However, they are currently working at capacity on the U.S. 1910 sample and, although willing to draw the over sample of Puerto Ricans in New York for 1910 en passant and return to their 1920 programs to prepare the 1920 over sample, they cannot undertake the additional data processing load of the Puerto Rican samples themselves. Dr. Ruggles has, however, agreed to provide us with the programs used for the U.S. samples for 1910 and 1920. These are all standard code for Microsoft Access. We will make the necessary modifications to this code to fit the minor variations in the schedule for Puerto Rico from that for the U.S., for the change from English to Spanish, and for our larger sampling fraction. Ruggles has also agreed to undertake the missing data allocation procedure to insure that it is handled in a way consistent with other IPUMS files.

Rather than develop a new "shop"of our own to undertake the clerical activities, we have decided to subcontract the task to the Wisconsin Survey Laboratory, http://www.wisc.edu/uwsc/. Constructing a public use census sample is very much like processing a survey. The data entry phase is like the interviewing phase of a survey but with less opportunity to query the respondent about confusions. The development of dictionaries for coding responses and undertaking the coding is very similar between PUMS and survey procedures. We will be translating and adapting the dictionaries used for the U.S. comparable census so that some of the work is actually less for the PUMS than for a survey. Evaluation of quality, editing and cleaning, and the construction of new variables are also tasks

similar to survey activities. Thus, by subcontracting the work to an ongoing survey organization we gain a practiced professionalism similar to that extant at the Minnesota Population Center without having to create it *de novo*. Information on the subcontract is included in the budget and its explanation.

IV.1.2. Extraction of samples

The overall sampling rate will be one in ten. The more sparsely populated coffee growing regions will be sampled at a slightly higher rate. Overall in the two censuses we expect about 240,000 cases in the main sample and an additional 20,000 in the coffee growing region over sample. The sampling procedure will be the procedure used for IPUMS adapted to raise the sampling fraction from one in a hundred to one in ten. The plan is to obtain a sample of households and all persons in the selected households will be included in the sample. In the conventional scheme used by IPUMS five person-lines are chosen from every 50th page of the fifty-line enumeration schedule pages. If the person selected is a household head, all persons in the household are included in the sample. If the person selected is not a household head, no persons in that household are included. Group quarters - units with more than 30 residents in the IPUMS procedure - are dealt with as if they were single individual households.

Our modification of this plan is to draw five sample lines **from each schedule page**. This provides each line with a one in ten chance of being selected. This plan is slightly more costly than an exact analogy to the IPUMS scheme but has the advantage of yielding much less geographic clustering. If the sampling unit were the individuals, we could stop there. We want samples of households, however, thus we need an additional step. The chance that a selected line is from a household of a given size is proportional to that size. If we took all the household members from the selected line's household, we would wind up over representing large households. In the 1940 PUMS project we arrived at the same point in sampling and then counted the line's household size and selected the whole household randomly with a probability inversely proportional to the size. The IPUMS solution is a simpler way of accomplishing the same thing. The chance that the sample line falls on the household head is simply one over the household size. Thus, the sampling rule includes the correcting experiment

Data entry will follow the same IPUMS techniques and, indeed, use its programs for controlling the entire process. Data will be entered as text. Data dictionaries will be constructed to transform text to numeric codes following entry. The dictionaries will be based on translations of IPUMS dictionaries but modified and expanded as seems empirically necessary. Over all, we expect that about a tenth of the cases will be double entered and reconciled for verification. Double entry will be more likely in the beginning of the project and for a new employee entering in the course of the project.

Coding will be also consistent with the IPUMS practice. Household and family structure variables may take some additional attention to conform to the Puerto Rican reality. Occupation will be coded to both the 1900 and the 1950 classification schemes. We expect to redefine existing or introduce new categories to make them more consistent with the realities of Puerto Rican labor markets at the time.

The Wisconsin Survey Center estimates that their part of the project on the two decennial censuses can be completed in 30 months. Data entry is estimated to take 24 months. Start-up and final data file preparation and documentation add an additional three months each.

IV. 1.3 Problems with Traditional Public Use Sample Procedures.

Public Use Sample files, and especially those included in the IPUMS collection, suffer a number of problems that have developed traditional ameliorations over the years. A first problem is one of cross-temporal comparability in coding. Occupational titles, for example, do not necessarily mean the same thing over time. Stevedores in 1910 were laborers who carried materials from ships on their backs. In 2000 they are operatives who drive fork lift trucks. If one were interested in comparability in function, then the word stevedore means much the same thing over time. However, if one is interested in comparability with respect to relative pay, riskiness, or social standing, then the word means something quite different in the two time periods. Similar problems occur with respect to industry, family relationship, and even geographic coding schemes. To be a member of the "urban" population differs over time. The traditional solution to this problem, which we will employ, is to provide the original text as well as the coded value and, perhaps, to code to more than one standard scheme.

A second problem of some concern with PUMS files made from long-past censuses is undercount. In our particular instance, we know of no work estimating the extent of undercount in Puerto Rico in 1910 or 1920. Even if such work were available, it is not clear what one should do

about it in compiling a file for public use. Undoubtedly the proper adjustment is closely related to the nature of the analysis being done. Thus, adjustment for undercount is probably best left to the analysts. In truth, however, adjustments are rarely made in the analysis of these files.

A third problem is one of the allocation of missing data. Again, how this process is optimally accomplished is probably conditioned by the analysis to be undertaken. For the mainland data for the two decades, a hot deck like procedure has been used as a kind of average solution. We will provide the same solution for Puerto Rico. Of course allocated data are indicated and so the analyst has the freedom to choose another method that better fits their needs if he or she chooses.

IV.2. Methods and models

IV.2.1. General Mortality

As a background for and to complement the study of child mortality we will estimate intercensal adult mortality levels and changes. In particular, we will estimate the magnitude of mortality decline between 1900 and 1920. This can be done by using basic equalities from generalized stable populations (Preston and Coale, 1982; Preston et al., 1989). We will use the 1910 and 1920 census age distributions, exogenous knowledge about migration, and births in the intercensal period, to derive estimates of life expectancy for the three year period centered around the censuses as well as estimates of changes in mortality in the intercensal period. Due to stringent information requirements about intercensal migration flows, the technique cannot be used to estimate adult mortality at less aggregate levels.

IV.2.2. Child mortality

i. Purpose and strategies: Information on children ever born and surviving provides the raw materials for the estimation of child mortality. To do this we will utilize a whole array of techniques, most of which are based on the seminal work of W. Brass ("Brass multipliers") and associated developments. These techniques have been used widely not just in the study of child mortality in the developing world but also in the investigation of mortality conditions in the US (see Preston and Haines, 1991. We divide these tasks into three groups described below in increasing order of analytic strength.

ii. Important clarification: we will estimate mortality indices for aggregates of variable nature. First, we will characterize Puerto Rico as a whole as well as major regions. Second, we will distinguish smaller spatial units within each regions, "municipios," which we will call residential areas. Third, we will distinguish aggregates defined by levels of education, nature of occupation and other characteristics that proxy for poverty levels (home ownership, access to sewage, access to piped water, access to electricity). We will refer to these as social groups or strata. Finally, in some cases, we focus on individuals and their own characteristics as well as those pertaining to places where they reside.

☛Estimation for large aggregates: The basic idea is to transform the observed proportion of children born to women in quinquennial age groups, (D(x); x=15, 20,...45) into estimates of the conditional probabilities of dying before ages i, (q(i), i=1,2,3,5,10,15,20)), where i depends on x. The procedures to do so are well established and will not be reviewed here (Brass, 1975; Trussell, 1975; Sullivan, 1972; Palloni and Heligman, 1985) . Brass' types of estimates are known to be quite robust but somewhat sensitive to violations of assumptions about constant fertility, constant mortality, and prevailing pattern of mortality. Through the use of modifications proposed by Preston and Palloni (Preston and Palloni, 1978) we will generate estimates that do not depend on assumptions about fertility. Instead of indexing an assumed fertility pattern we employ the observed age distribution of children surviving in the household. Biases due to changes in mortality will be eliminated by applying well-established corrections procedures proposed by Palloni and Feeney (Palloni, 1981; Palloni and Heligman, 1985; Feeney, 1978). Finally, we will use the new UN model mortality patterns (instead of the Coale-Demeny models) as they are more suitable to capture age patterns of mortality in Latin American countries (Palloni and Heligman, 1985).

The resulting estimates-and associated standard errors-- will characterize the country, entire regions, major cities, and large agglomerations of rural villages. Since we will obtain two sets of estimates, one from the 1910 census and the other from the 1920 census, and since the techniques enable us to retrieve trends of child mortality for a period of 5 to 10 years before the date of the census, we will be able to assess with some precision the prevailing levels of mortality during the entire period 1900-1920 in the country as a whole and in the various other aggregates. This accomplishment in itself is worthwhile since we will be able to generate concise and precise estimates of mortality for a developing country prior to the revolution that brought down the old mortality regime. Historians and demographers have been able to do this only in a handful of cases. However, this is not enough to identify the determinants affecting mortality levels. In order to produce robust estimates of effects of covariates we need other procedures

☛Estimation for subgroups: It is in our best interest to produce estimates of child mortality at levels of aggregation smaller that regions, cities or rural agglomerations. In fact, we are particularly interested in testing hypotheses about differentials by maternal education, head's occupation, and poverty levels as assessed by characteristics of the dwelling (availability of water, sewage and electricity). We are particularly interested in testing hypotheses about the existence of differentials early in the century between mothers of different educational levels, regardless of disparities in heads occupation or even poverty levels. Is it in indeed true that maternal education had then, as it has now, a strong impact? If so, the hypotheses that diffusion of information about cleanliness and hygiene was behind the early mortality decline would gain additional strength.

Estimates of mortality levels by subgroups can be obtained by simply reversing the operations suggested by W. Brass. For each well-defined subgroup we estimate the expected proportion of children dead to mothers of different age groups. This quantity is calculated as the fraction of children dead that would be observed in the subgroup under a standard pattern of mortality given the observed fertility levels of the subgroup. Since what we observe are not the details of the fertility pattern but only indicators of cumulated performance (children born by age of mothers) we assume that women in each subgroup have a fertility pattern identical to the average of the subgroup. We can then use the Brass multipliers in combination with the indicators of cumulated fertility performance to calculate the expected proportion of children dead. The ratio of the observed proportion dead to the expected one yields a positive valued, continuous mortality indicator. The behavior of this indicator and its association with subgroups characteristics can then be studied using standard regression techniques. This procedure was first suggested by Trussell, and Preston (1981) and employed by Preston and Haines in their investigation of mortality with the US 1910 census. We propose to use a similar procedure with added fine-tuning. The first adjustment is to estimate the expected number of children dead using a pattern of mortality more suitable for Puerto Rico that can be extracted from the UN models of mortality. The second adjustment is to estimate the expected proportion of children dead **after adjusting** for past mortality changes. This can be done in a straightforward manner using the methodology proposed by Palloni and Heligman (1985). Finally, rather than using linear models for the analysis of the indicator, we propose the use of non linear models of the following type. Let $\delta(x,Z)$ and $d(x,Z)$ be the expected and observed fraction of children dead for women aged x, with a vector of characteristics Z. Let $B(x,Z)$ be the observed number of children ever born for the same subgroup of women. The products of $B(x,Z)$ and $\delta(x,Z)$ and $d(x,Z)$ yield, respectively, the expected and observed number of children dead, say $\Delta(x,Z)$ and $D(x,Z)$. $D(x,Z)$ can then be modeled as a Poisson variable, a function of an offset, $\Delta(x,Z)$, and selected covariates.

☛Estimation for individuals: The same procedure described above can be applied to individual women. Two new problems emerge. The first is that we cannot compute robustly the expected number of children dead for each woman knowing only her age and the total number of children ever born. To circumvent this problem we need to assume that a women's life time fertility performance -at least up until the time of the census-- can be properly identified through indices of cumulated fertility for some subgroup to which she belongs. The smaller the subgroup we choose, the more unstable the fertility index becomes. In some applications, researchers have used the average pattern of fertility observed in the population. Rather than adopting this solution we propose to utilize estimates of fertility patterns by subgroups obtained from own children procedures (see below) and the cumulated indices associated with them.

The second problem is that for the majority of women the fraction of children born who have died will be 0. Therefore, the dependent variable-the ratio of observed to expected fraction children dead-will be truncated at 0. The proper procedure to study patterns of association with individual woman's characteristics is not OLS but Tobit regression. Palloni and Pinto (2000) have used this procedures extensively in their application to data from micro census samples for about 15 countries.

IV.2.3. Model building and hypotheses testing

i. Preamble on estimation of parameters: The foregoing procedures provide the basic materials for elaborating models of child mortality. With one exception, the indicator of mortality level among children born to a mother will be assessed at some level of aggregation higher than individual. This means that the indicator is computed for a subgroup within which there will be some within group variance. These subgroups will vary (see above) as they can be regions, regions of residence or social strata. It is only when we focus on the ratio of expected to observed deaths per woman that we take within group heterogeneity into account.

ii. Model building

We begin assuming that the mortality indicator is a function of basic demographic characteristics summarizing mother's childbearing experience. Most importantly, we pose the hypothesis that individual's resources, such as maternal levels of education, and household endowments, such as region of residence, will exert independent effects. If indeed mortality decline proceeds in the same way it did in the US during a similar period of time, maternal education should have very high returns, higher than those associated with proxies for income (head's occupation) (Preston and Haines, 1991). We also postulate that the direct or indirect impact of health interventions will have an independent effect, regardless of individual endowments. Household endowment includes not just access to water, sewage or electricity, but also distance to sites where public health interventions took place. Interventions for the eradication of malaria, dengue and yellow fever, typhoid and TB can be precisely timed and located in space. For obvious reasons, distance to such sites is a proxy for the direct impact of the intervention. But it is also a proxy for the indirect impact, that is, the influence that derives from knowledge about hygiene and protective measures that altered the population's levels of exposure to diseases. If this household endowment has similar effects regardless of individual conditions, we can actually evaluate the 'pure' direct and indirect effects of the intervention. There are a number of ways of estimating these effects but here we focus on only one of them, namely, analysis of temporal shifts. The proper model is one where the indicator of mortality level at time t for mother i in household j and residential area k, $L_{ijk}(t)$, is a function of characteristics associated with the mother, the household and the area of residence. Among these we include maternal education, characteristics of the household and head. The model can be written as

$$T(L_{ij}(t))=G(M_i, H_j, R_k)) \qquad (1)$$

where T stands for Tobit transform, G(.) is a linear linear function of inputs, including maternal characteristics, M_i, household characteristics, H_j, and characteristics of the residential area, R_k. This last set of variables includes aggregate properties, such as proportion of households with connection to sources of water and sewage and access to electricity. It will also include one (or more) dummy indicators signaling the proximity to an area of intervention. Expression (1) can be estimated separately for each year and then the behavior of all estimates over time compared. In particular, one can determine whether the relation shifted significantly during the intercensal period, as could be suspected if exogenous interventions had effects not captured by (1).

One of the hypotheses is that, once the mortality decline begins as a result of interventions (some of which may involve works of infrastructure whereas others result from either focalized interventions or public health campaigns), the effects will be stronger among those who are better 'prepared' to accept the associated changes. Thus, the health returns to maternal education should increase between 1910 and 1920 and do so more in residential areas favored by interventions.

A second application will take advantage of the fact that we the information on child mortality is for units of observation (subgroups of individuals in social classes, residential areas within cities, entire regions) that form a pseudo-panel since for each unit we have information over two periods of time. This provides us with an opportunity to estimate fixed effects models to yield more robust estimates of effects for relevant covariates. If units are sufficiently large, there is no need to resort to Tobit transforms and a linear model will do. The fixed-effect model for subgroups j is expressed as follows:

$$L_j(t)=H(Z_j(t))+\delta_j+\varepsilon_{jt} \qquad (2)$$

where $L_j(t)$ is the mortality indicator for subgroup j at time t, $Z_j(t)$ is a vector of covariates for subgrouop j evaluated at time t, H is a linear transform, δ_j is a fixed (over time) unmeasured factor associated with subgroup j, and ε_{jt} is a random error satisfying the usual assumptions.

IV.2.4. Problems in the estimation of child mortality

Needless to say, the procedures described before are not problem-free. In addition to the difficulties we noted and for which some solutions were proposed, there are other remaining shortcomings.

☹Age reporting: it is well known that age declaration is faulty. Most of these errors, however, can be handled by grouping ages into sets likely to include both the true and the reported age. For example, digit preference that consists of rounding up to 0 or 5 is tractable by defining non conventional groups, such as 12-17 and 18-23 instead of 15-19 and 20-24 which, except in degenerate cases, will contain most of both the reported and the true age of individuals who rounded up their ages. Systematic age

misreporting is more difficult both to detect and correct. To deal with these we can only propose to apply consistency checks in the age distribution of subgroups before we estimate parameters.

☹Reporting of children ever born and dead: one of the better known problems with reports on children ever born and children dead is that very young women tend to over report 0 children and older women tend to selectively forget children who died far into the past. The first of these is generally attributed to a coding error and can be easily adjusted (Palloni, 1981). The second problem is more delicate and can only be detected rather than solved. Detection involves fitting a curve to the reported fraction of children born by subgroups and then verifying the slope of the top part of the curve behaves properly according to a standard. If irregularities are noted one must only use a sample of women aged Z or below where Z is an age above which irregularities become salient.

☹Differential undercount: this problem is the most difficult to deal with. Assuming that one has acceptable estimates of general undercount and possibly of differential undercount by large aggregates, it would be possible to reweight estimates of mortality and of estimated effects so that they represent better the true population. A better procedure is to divide the country into areas classified according to probable levels of undercount and to estimate our models within those areas, rather than grouping them irrespective of suspected levels of undercount. Unfortunately neither solution works if undercount selects out individuals or families who are more likely to experience higher (lower) levels of mortality. If that is the case, our estimates of differentials will be biased downwards and should be take as conservative bounds.

IV.2.5 Nuptiality and fertility

i. Purpose and strategies: Puerto Rico's fertility decline is a singularity in the continent because of the unusual historical circumstances characterizing its social and economic development. We expect that fertility decline will have a different timing and pace as well as a different profile of social and economic differentials. The data included in the Puerto Rican censuses is not only very rich but also unique. It is rich because it contains information normally not included in other censuses (duration of marriage,_ number of children ever born and surviving) thus enabling us to deploy a number of tools that are impossible to apply except in very rare situations. It is unique because such information is available neither in censuses of Latin American after 1950 (and certainly not before) an in very few European censuses.

Our analytic plan includes a number of strategies, some of them purely descriptive, whereas others are for hypotheses verification and model building. We describe them in increasing order of complexity.

ii. Important clarification: As we did in the case of mortality, we will estimate fertility indices for aggregates of variable nature: nation and region, residential places. social groups or strata, and individual. For a definition of these, see Section IV. 3.3.ii above.

☛Coale Indices: combining information on number of births (at the national and regional level) from vital statistics and on age distributions from two censuses, we will compute all three indices utilized in the Princeton Fertility Survey, namely, I_m, I_g and I_f (Coale, 1972). It is, therefore, possible to produce basic comparisons of Puerto Rican's conditions during the first two decade of this century with that of Western European countries or of countries in Latin America. Since there is a multiplicative relation between I_f, on the one hand, and I_g and I_m on the other, changes in general fertility over time can be easily decomposed into those attributable to changes in marriage and those associated with marital fertility. Furthermore, we will decompose I_m into two parts, one associated with consensual unions, I_{cm}, and the other associated with marriages, I_{mm}, so that $I_m=I_{cm}*I_{mm}$. This will enable us to further decompose changes attributable to nuptiality into those associated with the regime of consensual unions and those associated with official marriages.

☛Own children estimates: the workhorse of the fertility section in this project will be estimates of age and time-period specific fertility rates produced with own-children methods. Own children procedures rely on the observed age distribution of children surviving in the household, total number of children ever born to a woman in a household, her age and her relation to a head (and to other members of the household). With this information one transforms the age distribution of children surviving in a household at the time of the census into estimates of the time distribution of births in the past and of fertility rates by age and by year before the survey. There are a number of feasible methods to produce these estimates (Cho and Retherford, 1986). John Marcotte (with the help of Alberto Palloni and Glen Ulmer) generated a modified version of the general method proposed by Retherford and Cho. This modification enables the user to adjust mortality levels and patterns, adjust for differential undercount of children and mothers and takes into account the possibility of

differential age misreporting among children and mothers. This procedure has been successfully applied to estimate fertility in about 20 Latin American countries from 1950 until 1990 (Palloni and Pinto, 2000). We propose to employ the same methodology with the Puerto Rico census micro samples.

Estimates of age and year specific rates, F(x,t), for ages x=15 to x=49, for the ten years preceding each census, t=1,...,10, are generally quite reliable. They can be estimated for subgroups of women according to characteristics of the household or of the household head, including region of residence, education and occupation, etc... Thus, we will have detailed information on F(x,t) for a number of subgroups and social classes for a period of time stretching from 1900 until 1920.

Surprisingly, the 1910 Puerto Rican census also includes information on union duration for all women in a household who reports to be in a union. Thus, all estimates associated with the 1910 census (all the fertility rates from 1900 until 1910) can be simultaneously estimated by age and duration in union, a most unusual condition in historical populations. This will enable us to produce estimates of age and duration specific fertility rates and to estimate robustly parameters of birth control.

☛Estimates of birth control (non-duration dependent): we will first adopt the Coale-Trussell fertility model to retrieve estimates of parity-dependent birth control from the estimated age-specific fertility. The model assumes that the observed fertility pattern departs from a natural fertility pattern via practice of birth control as a stopping behavior (Coale, 1969). A birth control parameter, m, captures the extent to which the observed fertility schedule deviates from one where there is complete absence of parity-specific control. A nuisance ('level') parameter, M, reflects conditions that affect fertility through spacing and that are unrelated to parity-specific control, such as lactation or marital disruption due to migration. The expression for fertility rate at age x and time t is

$$F(x,t)=M(t)* \eta(x)* \exp(v(x)*m(t)) \quad (3)$$

where $\eta(x)$ is a standard schedule of natural fertility, $v(x)$ is a standard schedule of deviations from natural fertility, and $M(t)$ and $m(t)$ are the two (time-varying) parameters reflecting 'degree' of birth control and 'level' of fertility. Estimates of M&m can be obtained for relevant subgroups of the population, particularly those defined by indicators of social strata (education, occupation) and region of residence. They, as much as the fertility rates and summary measures of fertility curves (such as TFR) on which they are based, can be considered dependent variables and their profile over time studied as a function of other characteristics. This will be one of a handful of opportunities demographers have had to assess the degree of birth control in a historical population who begins to legitimize the practice.

In theory, we could estimate m and M for all years t for which we have available own children estimates, F(x,t), namely, for ten years before each census, a period stretching from 1900 to 1920. In order to increase robustness, however, we suggest to group estimates in five-year clusters thus generating a time series with four point estimates of m and M during the twenty year period. This clustering of estimates does not eliminate but minimizes the noise due to age heaping and misreporting. It preserves, however, a time series of estimates (m&M) which is highly recommended for proper interpretation.

As we have done before we will define subgroups (by social class via education, occupation and indexes of dwelling characteristics; by place of residence, by region) and calculate estimates of m&M. For descriptive purposes, the estimates can be treated as separate time series for subgroups. For model building we can treat it as a pooled cross section and time series (see below).

☛Estimates of birth control parameters (duration dependent: a modification of the above described model makes use of the information on duration of union. This extension first proposed by Page (1997) can only be used with the 1910 own children estimates by age and duration of union. The expression for the fertility rate at age x and union duration d is as follows:

$$F(x,d)=K* \eta(x)* \exp(-\mu*d) \quad (4)$$

where K is a level parameter for the age-duration dependent fertility schedule, $\eta(x)$ is as before, and μ is a parity dependent measure of birth control within unions. The estimate of μ in expression (4) does not depend on an assumed pattern of deviation from natural fertility, only on reported union duration. In theory at least, it should be a more robust estimate of the birth control parameter. In general, however, m and μ should not differ by much unless union duration is vastly misreported or birth control in the population is highly *sui generis*.

As before, estimates of K and μ will be obtained for different social groups, for each year before the 1910 census and, to secure robustness, for two quinquennial period, 1900-05 and

1905-10.

☛Cohort Parity Analysis (CPA)

The availability of information on union duration in the 1910 census, enable us to pursue a line of research that is rarely possible with historical data. In fact, to avoid restrictive assumptions associated with the model of parity dependent birth control, we can apply Cohort Parity Analysis (CPA) to the age-duration information of the 1910 census. In a nutshell CPA is a procedure that enables us to estimate lower and upper bounds of the true proportion of the population who are birth controllers at any given parity. When the resulting range is reasonable narrow, it provides a gauge for the presence of birth control practices used for either stopping or for spacing behavior. Because CPA models **fertility control intentions** instead of focusing exclusively on outcomes, it will enable us to detect even imperfect birth control practices, a feat that is simply not possible with the Coale-Trussell model. Because the description of the details of CPA lengthy and complex, we do not present it here. Suffice it to say that our goal is to use its final product, namely, estimates of the lower and upper bounds of the proportion controlling per parity j, $C_L(j)$ and $C_U(j)$. These quantities can be estimated for social subgroups (social classes, places of residence, regions) and be treated as dependent variables with conventional models for discrete dependent variables (logit or probit models). The only qualification required here is that it will be important to simultaneously model both quantities, the upper and lower bound, rather than only one of them. This is because the width of the range conveys information about the precision of the estimate. In order to do so we can use likelihood procedures for the range of a proportion rather than for the proportion itself. If the range for unit(group of parity j) j is $\{C_L(j),C_U(j)\}$ then the corresponding likelihood for unit j requires integration of the appropriate density (normal or logistic) between the two boundaries. A STATA routine to optimize the such likelihood is available in house.

Although it will not be possible to repeat the exercise with the 1920 census, we will be able to analyze the association between CPA and M&m estimates in 1910. It is therefore feasible, albeit not optimal, to produce indirect CPA estimates with information from M&m in 1920 and the estimated relation between CPA and M&m estimates.

☛*Nuptiality: the Coale-McNeil function*

Nuptiality can be analyzed via three different procedures. The first is to use the index I_m mentioned before. The second is to calculate SMAM from the observed proportions never marrying in each census. Since the conventional estimate of SMAM assumes that the nuptiality regime is invariant and we argued that nuptiality was changing during this period, we suggest to use a modified procedure that relies on the difference between the proportions ever married by age in the two censuses. This procedure is a straightforward modification of the intercensal growth approach and has been already applied successfully in similar contexts.

The third, and perhaps more fruitful approach, is to estimate the parameters of the Coale-McNeil nuptiality model and the effects of covariates on each one of them. The Coale-McNeil model is an accelerated time failure model postulating that the proportion of ever married women at age x can be described by the following expression:

$$G(x)=T*G_s((x-\alpha)/\lambda) \qquad (5)$$

where T represents the proportion who will ever marry in the observed population, $G_s(y)$ is the proportion ever married in a standard nuptiality schedule, α is the age at which a significant fraction of women begin to enter into marriage in the observed population, and λ is a measure of the pace at which marriage proceeds in observed population relative to the standard. Rodriguez and Trussell (1980) re-parameterized the model and suggested a Maximum Likelihood procedure to estimate the parameters. Palloni and Fussell (2000) derived a Maximum Likelihood procedure (implemented in STATA) to estimate the effects of covariates on the re-parameterized version of the model. We propose to use this procedure to investigate the effects of basic characteristics, such as education, place of residence, region, on the marriage process during the interval 1900-1920.

An important consideration in the case of Puerto Rico is the separate estimation of parameters for women who are in union and those who are married, As shown by Palloni and Fussell (2000), this separation is feasible and can be achieved efficiently with a simple three-state multistate representation modelling simultaneously the hazard of both a marriage and union with the same re-parametrized nuptiality function. The final result of the estimation process will be two sets of parameters, one for unions and one for marriages, for the period 1900-1920. This should provide sufficient information to describe marriage changes during this period.

☛*Simultaneous estimation offertility and nuptiality*

So far we treated nuptiality and marital fertility separately. This is not imperative. Indeed, assuming the Coale-McNeil marriage function and the parity dependent control model of fertility one could jointly model both components as any fertility rate F(x,t) can be described as

$$F(x,t)=\{T(t)*G_s((x-\alpha(t))/\lambda(t))\}*$$

$$\{\eta(x)* \exp(\ln(M(t))+v(x)*m(t))\} \qquad (6)$$

The idea for this joint model and suggested quasi-likelihood procedures for estimation were first proposed by Rodriguez and Phillipov (1995). We propose to generalize (6) to include effects of covariates on at least some of the parameters. To do this we need to define appropriate social subgroups (regions, areas of residence, social strata) and to design a numerical routine for maximization of likelihood. Note that results from the joint model need not be identical to those from a separate estimation of the nuptiality and marital fertility model.

☛*Cumulated fertility*

The index B(x,t), the number of children ever born to a woman aged x at time t, is the simplest and also the only one that can be calculated for individual women with the data available to us. Clearly, one could calculate B(x,t) for different social subgroups and perform analyses of differentials. However, its main advantage lies precisely in the fact that it'-enables us to focus on individual outcomes. Admittedly it continues to be an 'aggregate' indicator as it reflects the fertility experience since the onset of the childbearing period. In order to make good use of this indicator and test conjectures and theories, such as the diffusion hypotheses, it is necessary to introduce some basic parameterization. We will follow Brass's suggestion and adopt a two-parameter Gompertz fertility model. Once standardized for lifetime fertility, B(x,t), follows the profile of a Gompertz distribution function. The first parameter of the transformation controls the spread (variance) of the fertility function whereas the second one reflects its timing. The introduction of effects of covariates (region, area of residence, social strata) is unproblematic, and so is the analysis of intercensal changes.

IV.2.6. Model building and hypotheses testing

i. Preamble on estimation of parameters: with the exception of the information on children ever born by age (and duration in union), B(x,t), and its Gompertz transformation, our fertility estimates refer to units involving some aggregation of individuals. In the case of CPA the target quantities are two proportions (upper and lower bound of fraction contracepting) whereas in all other cases the target quantities are either marital fertility rates or general fertility rates (by age and/or duration of union). The CPA-associated estimates can be handled with modified probit or logit models. Fertility rates for a subgroup with characteristics Z can be modeled assuming that the counts of births are Poisson variables with mean and variance equal to the product of exposure time, O(x,d,t;Z), and a model fertility rate F(x,d, t;Z). In general, the models will have the following form:

$$E(B(x, d, t; Z))= O(x, d, t; Z)*F(x, d, t; Z) \qquad (7)$$

where Z is the vector of covariates. The function O(.) is the number (or transformation) of persons-years lived by women in the corresponding age-duration-time period-Z characteristic cell, whereas the function M(.) represents the model fertility parameterization (a duration dependent or independent model).

ii. Model building

As we did in the case of mortality, we now illustrate how we propose to carry out model-building using the assessments of fertility and nuptiality levels and patterns described before. Due to space constraints we discuss only one generic example regarding the role of diffusion in birth control prevalence.

☛Were birth control practices diffusing?

Although this question is paramount for demographic theories of fertility decline (Palloni, 2000; Casterline and Cohen, 2000) it has proven to be a hard nut to crack. A whole volume is devoted to conceptual and methodological issues raised by the conjecture(Casterline and Cohen, 2000).

The model we propose to use derives from a more general model suggested by the PI's (Palloni 2000). It consists of including an indicator of weighted proximity to social actors who exhibit the new behavior (controlled fertility), and to determine if such an indicator is associated with important effects net of other characteristic of the unit under observation. The model may include a lagged formulation or simply focus on a cross section. Thus, suppose we have estimates of marital fertility rates F8(x,t) for social groups g=1,...,G. Suppose also we hypothesize that the practice of birth control within the group is encouraged by imitative behavior using as reference social groups of similar standing within a well-defined neighborhood. Then the following model is appropriate for counts of births at age x

$$E(x,t;g)=O(x,t;g)*\{M(t)*\eta(x)*\exp(v(x)*m_g(t)\}$$

where all functions are as before. To avoid cluttering, we make the assumption that imitative behavior affects parity-related birth control only. Thus, M is constant across g whereas m_g is defined as:

$$\exp(m_o + \Sigma_h (\beta_n \omega_{gh} m_h))$$

a function of the weighted average of the m's in areas of residence located in a predefined neighborhood of g. The factor ω is a weight that could depend on physical distance or other measures of social interactions or proximity, and β is an associated effect. Assuming that β is constant leads to a simple linear form where m_g is a function of the weighted average of m's in social groups that are alike and located within predefined neighborhoods. A similar model for differences in counts of births between any two quinquennial periods before 1910 and between 1910 and 1920 can also be posed. Its parameters can be identified using a fixed- effects estimator.

IV.2.7. Estimation problems

Since estimates from own-children constitute the core of our estimation of fertility, we briefly discuss the three central problems associated with own children estimates and the solutions we propose. We neglect problems that are of secondary importance.

☹Differential Undercount: census undercounts per se are not threatening to own children estimates. This is because if entire households are missed, errors associated with the reconstructed fertility rates contain omissions in both numerator and denominator. The problem arises when undercount is associated with factors that are also determinants of fertility. We have no ready-made solutions for this shortcoming since there are no extant estimates of regional census undercounts for the period.

☹Age misreporting and selective omission

Age misreporting among women is subject to the usual pattern that leads to digit preferences (ended in 5 or 0) and to mild systematic downward biases among those in middle ages. Application of own children can bypass these errors since one can estimate fertility rates in unconventional age groups, that contain the attractive ages as well as the true ages. This type of groupings leads to remarkably smooth age patterns of fertility (Palloni and Pinto, 2000).

Age misreporting in the population of children is less regular than among adult women but has one important invariant feature: the tendency to omit children age 0 and/or to systematically report their ages as 1 or 2. The effect of this is to artificially depress fertility rates for the year before the census and to artificially increase it for the period of two to three years before the census. The obvious solution is to group years in such a way as to capture the true date of the events misreported

Omission of recently born children can be detected in a straightforward way by examining the age distribution of children surviving. There are two solutions to the problem. The first is to simply neglect the information on children age 0 and abandon any pretense to estimate fertility rates in the year before the census. This is our preferred strategy since it involves no extraneous assumptions. The second solution is to utilize estimates of infant and child mortality, survive backwards the (adjusted for-age-misstatement) population in the 1-4 age group, and then adjust once more for changes in the age-specific rates of growth. This is a more convoluted procedure that relies on a number of assumptions not all of which can be justified with ease..

☹Allocation of non-own children The average fraction of children living in a household who could not be allocated to a mother in the Latin American censuses is of the order of .08 (Palloni and Pinto, 2000). This is a low value, justifying the proportional allocation of non-own children as a fair solution likely to lead to consistent estimates.

Because the categories of relations to head in the Puerto Rican censuses are much richer than in the average Latin American census and because we suspect that the practice of fostering out children, even for short periods of time, is less of a concern in Puerto Rico than elsewhere in the region, we believe that the average fraction of non-own children will be even lower than 8 percent. If it were higher that ten per cent we propose to use two alternative procedures: (a) proportional allocation and (b) imputation within a household so that a match between a potential child and a potential mother is made only when certain conditions regarding children and mother's age, mother's marital status, and characteristics associated with other **own children** of a potential mother are established.. It is our experience (Palloni and Pinto, 2000), however, that alternative imputation algorithms to match mothers and non own children do not lead to results that are starkly different from those that one obtains with simple proportional assignment.

IV.2.8. Household composition

What was the typical household

organization in this period? How did it change between 1900 and 1920? The transformation of the cultural landscape of Puerto Rican society, and the massive changes in the fabric of economic and social relations that took place after Spanish domination, strongly suggest that patterns of residential arrangements and household configurations may have changed as well. This is in keeping with classic modernization theory that sees transformations like those experienced during this period as a potential root cause of a transition from a family regime dominated by extended family arrangements to a regime dominated by the nuclear household. But, what mode prevailed around 1910? How prevalent was the extended household and how large where regional and social class differentials? And if there was a change between 1910 and 1920, what was its root cause? Could it have been that the mortality, fertility and migration changes in the al period stifled individual preferences for extended household arrangements and forced an' increase in nuclear residential patterns? Or, was the opposite true, namely, the extended form increased propelled by demographic constraints, not by changes in individual preferences?

i. Purposes and strategies: The questions posed at the outset of this section (in bold) are the core of a conundrum which has not been satisfactorily resolved for Western European (Wachter et al., 1978), North American (Ruggles, 1987) and Latin American societies (Palloni et al., 2000). We do not intend to resolve them in this project. But, at the very least, we can add a case, an example illustrating the existence of a particular pattern. For starters we can answer the first question: what were the patterns prevailing during 1910 and 1920? We can do this in a rather straightforward manner, by simply determining the frequency of different types of household organizations. The categories or types we propose to use include very categories and types included in simple typologies (lonely individual, nuclear, extended) and more complex ones (Hammel and Laslett's and UN categorization) that enhance our ability to compare the situation in Puerto Rico with that in other societies and periods of time.

We can similarly dispose of the second issue pertaining to changes between 1910 and 1920 in an equally straightforward manner. We may need to utilize elementary statistical procedures to determine if changes between distributions of households across several types deserve to be taken seriously. But, asides from this minor hurdle, the required empirical work is remarkably unproblematic.

A third task is less straightforward. This is that we need to establish and assess the magnitude of differentials in household configurations. Because of its relation to processes of production and distribution, we expect large differences in household configurations across regions and social strata. Similarly, to the extent that changes in preferences are culturally bounded and sometimes directed by imitation, we also expect changes if transformations of values in the period-from secularization to adoption of a more 'western' type of approach-was sufficiently pronounced and advanced. In order to evaluate differentials we need to undertake some minimum modeling where the outcomes of interest, namely, the distribution of households by type, is made a function of characteristics (region, area of residence and social strata).

The most difficult endeavor is to answer the key question: if important changes were observed between 1910 and 1920, were they due to changes in individual (family) preferences or were they the result of adaptation to constraints imposed by a changing demographic regime? The problem is well-known and so is the solution (Wachter et al., 19xx; Ruggles, 19xx). In the sections that follow we provide a minimalist description of this solution

ii. Important clarification: we will estimate distribution of households for aggregates of variable nature. As in the case of mortality and fertility, these will include regions, areas of residence (municipios) and empirically defined social strata.

IV.3.9. Model buildin and nd hypotheses testing

Model building for household organization has two components. The first is one related to establishing social differentials. The second regards the model for disentangling the effects of preferences and demographic constraints on changes in household configuration. We deal with these in turn.

i. Assessing differentials: We will adopt conventional tools (multinomial logistic models) to describe characteristics of residential arrangements. Because there are key issues related to direction of causality that remain unresolved, these are not satisfactory tools to test important hypotheses about factors affecting household configuration. However, in some cases we will be able to infer causal direction. for example, because changes in agriculture occur in well defined regions-the influence of coffee producing regions dwindles and that of sugar growing areas increases-- we can study how these affected residential arrangements in each. Since we know when these changes began to take place and when they were more or less completed, we

can provide some time order to the sequence of events. Thus, sharp alterations in residential arrangements between 1910 and 1920 in coffee producing regions (but not in other regions) can in no uncertain terms be related to the overhaul of the mode of production.

In general, however, interpretation will not be straightforward. Yet, estimates from these models will provide us a first glance to the relation between relevant social and economic factors (region, social classes) and particular household arrangements. The typical model will be as follows:

$$P(Hj(t))=F(Zj(t))$$

where P is a probability transform dictated by the model being adopted (multinomial logit or probit) Hj is an observed household category or type for unit j, Zj refers to a vector of characteristics associated with the unit, and F is an operator (usually linear).

ii. The question of primacy: the logic and tools for deciding the validity of the proposition that changes in household configurations owes little to constraints associated with the demographic regime, are well-established and will not be repeated here. Due to space constraints we can only summarize the steps involved in testing the fundamental hypothesis of primacy of demographic factors. Admittedly, ours is a highly simplified description of a model-building task that is extraordinarily complex. We hope to identify the most basic elements to understand the strategy.

First, we will use a microsimulation model for households and families that reflects well the prevalence of a known demographic regime. This can be done using a number of possible models already tested: SOCSIM was originally used by Wachter and colleagues; CAMSIM is being used at Cambridge by Smith, Oeppen and colleagues; MOMSIM was used by Steven Ruggles in his study of the US census; FAMSIM is being used by Palloni and DeVos in Latin American countries. The differences between these models are in flexibility of use and degree of complexity they can represent. Our approach will be to use at least two of them to make sure that differences in modeling approach do not yield to sharply different results. If they do, these differences ought to be at least documented.

The second stage consists in determining the feasible distribution of household configurations and to assess whether there are statistically significant differences between these and those observed in the empirical case. If there are, then individual and family preferences may also have an influence on observed configurations.

IV.3.10. Fertility and mortality among Puerto Rican migrants to the US

The last objective of the proposed project is to systematically compare child mortality, fertility, nuptiality and residential arrangements between Puerto Ricans who migrate to the main land and those who remain in the island. We choose to focus on those residing in the five boroughs of New York, where more than 70 percent of the population of Puerto Rican migrants to the US resided in 1920.

The scope of the proposed analysis is ambitious for we seek to establish whether those who migrated to the US between 1900 and 1910 were indeed members of a selected population and whether or not, with increasing duration of residence in the US, their demographic patterns resemble more those of the mainland than those of the island.

Measurement of child mortality, fertility, nuptiality and residential arrangements among migrants living in the US presents no intrinsic difficulties as we will follow identical guidelines to those developed before for the census micro sample of Puerto Rico. There are, however, three important considerations to keep in mind.

i. Waves of migration: migrants belong to a first wave (in the first five years of the XXth century) are likely to be quite different from those that followed. At the time social networks were not established and migration was mostly of single workers, not families. This changed gradually over time. It is therefore important to distinguish migration cohorts as they are likely to display different traits and behaviors. This requires poses a problem for we also would like to assess the effects of duration of residence in the US. With only two censuses and the migration experience spread only over twenty years, it will be more difficult to adjudicate effects between migration cohort of membership and duration of residence in the US.

ii. Information available: the census information retrievable fro Puerto Ricans in the mainland is not identical to the information retrievable from the census in the island. The most important differences are two: (a) there is no information on duration of marriage for those in the mainland in 1910 and (b) there is information on residence in the US for individuals who live in the US and Puerto Rican born. Whereas the lack of information on duration of marriage or union prevent us from applying models such as CPA for assessing birth control, the added information on duration of residence is a saving feature since it will enable use to control for selection (see below) and to assess the strength of assimilation.

iii. *Modeling difficulties*: in principle, the investigation of similarities and contrasts between the migrant and the sending population is straightforward. Models appropriate for the demographic phenomenon of interest were reviewed before. These are also appropriate for establishing the contrast we seek. It suffices to account for whether or not a unit of observation belongs to one population or the other. Whether or not there is an effect associated with migration per se is, however, a more difficult nut to crack. In order to make sure that we do not misinterpret results, we must control for duration of residence in the US. Indeed, if migrants are drawn from selected groups in Puerto Rico, there will be a contrast at the time migration takes place. Consequently, a control for duration of residence may lead to consistent estimates of the effects of migration. With one exception. This occurs when the effects on demographic behavior of factors on which the migrant population is selected influences those behaviors with a time lag. For example, assume that migrants are more predisposed to take more risks and being more adaptive. Then their demographic behavior may not be **initially** different from the rest of the Puerto Rican populations (so we will not detect a difference at the outset), but that does not mean they are not selected. Insofar as the features involved are latent traits, we have no protection against this contingency. The only way out is to estimate models with unmeasured heterogeneity even though, as is well known, these are not robust to misspecification of the distribution of the latent trait.

A second problem is that we are interested in assessing effects on demographic behavior of area of residence within New York. There is evidence suggesting that decisions about residential location in the US led to differences in behavior. However, this presents us with another selection problem: how to distinguish the direct effects of residing in a particularly area of the city from those of factors that affect both decision about residential location and demographic behavior?

Although not resolved we identify these modeling problems as they are inherent in the product we propose to distribute, the PURIPUS.

References

Acosta-Belén, Edna, 1986. *The Puerto Rican woman : perspectives on culture, history, and society*, 2nd. ed. New York: Praeger.

Arias, Elizabeth and Alberto Palloni. 1999. "Prevalence and Patterns of Female Headed Households in Latin American: 1970-1990." *Journal of Comparative Family Studies* 30(2):257-79.

Bergad, Laird W. 1983. *Coffee and the Growth ofAgrarian Capitalism in Nineteenth-Century Puerto Rico*. Princeton, NJ: Princeton University Press.

Brass, W. 1975. "Methods for Estimating Fertility and Mortality from Limited and Defective Data." Chapel Hill, NC: Laboratories for Population Statistics, Carolina Population Center.

Casterline, J. and B. Cohen. 2000. *Social Processes Underlying Fertility Changes in Developing Countries*. (forthcoming) National Research Council.

Cho, Lee-Jay, Robert D. Retherford, and Minja Kim Choe. 1986. *The Own-Children Method of Fertility Estimation*. Honolulu, HI: Population Institute.

Coale, A. J. 1972. *The Growth and Structure of Human Populations: A Mathematical Investigation*. Princeton, N.J.: Princeton University Press.

Coale, Ansly. 1969. "Decline of Fertility in Europe from French Revolution to World War R." In *Fertility and Family Planning*, edited by S. J. Behrman, L. Cousa and R. Freedman. Ann Arbor, MI: University of Michigan.

Devos, Susan and Gary Sandefur. 1999. "The Living Arrangements of Elders in Five European Countries Circa 1990." Center for Demography and Ecology working paper No. 99-14. University of Wisconsin-Madison.

Diffie, Bailey and Justine Diffie. 1931. *Porto Rico: A Broken Pledge*. New York, NY: Vanguard Press.

Feeney, G. 1978. "Estimating Infant Mortality Trends from Child Survivorship Data." East-West Population Center, Honolulu, HI.

Hirschman, C. and P. Guest. 1990. "Multilevel Models of Fertility Decline in Southeast Asia." Population Association of America. Toronto, Canada, May 2-6.

Kramarow, Ellen. 1995. "The Elderly Who Live Alone in the United States: Historical Perspectives on Household Change." *Demography* 32(3):335-52.

Livvi-Bacci, Massimo. 1986. "Social-Group Forerunners of Fertility Control in Europe." In *The Decline offertility in Europe*, edited by Ansley J. Coale and Susan Cotts Watkins. Princeton, NJ: Princeton University Press.

Macunovich, Diane J., et al. 1995. "Echoes of the Baby Boom and Bust: Recent and Prospective Changes in Living Alone Among Elderly Widows in the United States." *Demography* 32(1):17-28.

Natal, Rosario Carmelo. 1983. "Exode Puertorriqueno: Las Emigraciones al Caribe y Hawaii." San Juan.

Page, H. J. 1997. "Patterns Underlying Fertility Schedules: A Decomposition by Both Age and Marriage Duration." *Population Studies* 31(1):85-106.

Palloni, A., K. Hill, and G. Pinto. 1996. "Economic Swings and Demographic Changes in the History of Latin America." *Population Studies* 50(1):105-32.

Palloni, A. and R. Wyrick. 1981. "Mortality Decline in Latin America: Changes in the Structures of Causes of Deaths, 1950-1975." *Social Biology* 28(3-4):187-216.

Palloni, Alberto. 1981. "Adjusting Data on Children-Ever-Born for Nonresponse." *Social Biology* 28:308-14.

_____. 1990. "Fertility and Mortality Decline in Latin America." *Annals of the American Academy of Political and Social Sciences* special issue: 126-44.

_____. 1999. "Increment-Decrement Life Tables." In *Elements of Demography*, edited by Samuel H. Preston, Patrick Heuveline and Michel Guillot. Oxford University Press.

_____. 2000a. "Living Arrangements of the Elderly." United Nations Technical Meeting on Population and Aging. New York, NY, February 8-10, 2000.
_____. 2000b. "Diffusion in Sociological Analyses: How Useful is It for the Study of Fertility?" In *Social Processes Underlying Fertility Changes in Developing Countries*. (forthcoming) National Research Council.
Palloni, Alberto, Hill K. 1997. "The Effects of Economic Changes on Mortality by Age and Cause: Latin America, 1950-1990." Pp. 75-128 in *Demographic Responses to Economic Adjustment in Latin America*, edited by G. Tapinos, A. Mason and J. Bravo. Oxford: Clarendon Press.
Palloni, Alberto and Susan Devos. 1989. "Formal Models and Methods for Analyzing Kinship and Household Organization." *Population Index* 55(2):174-98.
Palloni, Alberto, et al. 1994. "Changes in Families and Households in Latin America, 1970-1990." Population Association of American. Miami.
Palloni, Alberto, Susan Devos, and Martha Pelaez. 1999. "Aging in Latin American and the Caribbean." Center for Demography & Ecology Working Paper 99-02. University of Wisconsin-Madison.
Palloni, Alberto and Beth Fussell. 2000. "Nuptiality in Latin America."
Palloni, Alberto and Larry Heligman. 1985. "Re-Estimation of Structural Parameters to Obtain Estimates of Mortality in Developing Countries."
Palloni, Alberto, Kenneth Hill, and Guido Pinto. 1996. "Economic Swings and Demographic Changes in the History of Latin America." *Population Studies* 50(1):105-32. Palloni, Alberto and R. Kominski. 1984. "Estimation of Adult Mortality Using Forward and Backward Projection." *Population Studies* 38(2):479-93.
Palloni, Alberto and Hsien-Hen Lu. 1999. "Adult Health and Adult Mortality: Recent Trends." University of Wisconsin-Madison.
Palloni, Alberto, Michael Massagli, and John Marcotte. 1984. "Estimating Adult Mortality with Maternal Orphanhood Data: Analysis of Sensitivity of the Techniques." *Population Studies* 38(2):255-79.
Palloni, Alberto and Guido Pinto. 2000. "Fertility Patterns and Change in Latin America." Palloni, Alberto and Hanta Rafalimanana. 1999. "The Effects of Infant Mortality on Fertility Revisited: New Evidence from Latin America." *Demography* 36(1):41-58.
Palloni, Alberto and Marta Tienda. 1986. "The Effects of Breastfeeding and Pace of Childbearing on Mortality at Early Ages." *Demography* 23(1):31-52.
_____. 1992. "Demographic Responses to Economic Recessions in Latin America Since 1900." *Sociological Inquiry* 62(2):246-70.
Palloni, Alberto and R. Wyrick. 1981. "Mortality Decline in Latin America: Changes in the Structure of Causes of Death, 1950-1975." Social Biology 28:187-216.
Preston, S. H. 1976. *Mortality Patterns in National Populations*. New York: Academic Press. Preston, S. H. and A. J. Coale. 1982. "Age Structure, Growth, Attrition and Accession: A New Synthesis:" *Population Index* 48:217-59.
Preston, Samuel. 1980. "Causes and Consequences of Mortality Declines in Less Developed Countries During the Twentieth Century." In *Population and Economic Change in Developing Countries*, edited by R. Easterlin. Chicago, IL: University of Chicago Press.
Preston, Samuel H. and Michael R. Haines. 1991. *Fatal Years: Child Mortality in Late Nineteenth-Century America*. Princeton, NJ: Princeton University Press.
Preston, Samuel H., Christine Himes, and Mitchell Eggers. 1989. "Demographic Conditions Responsible for Population Aging." *Demography* 26(4):691-704.
Preston, Samuel and Alberto Palloni. 1978. "Fine-Tuning Brass-Type Mortality Estimates with Data on Ages of Children Surviving." *Population Bulletin of the United Nations* 10:72-91.
Rodriguez, German and Dimiter Philipov. 1995. "Fitting the Coal-Trussell Model by Maximum Quasi-Likelihood." Office of Population Research, Princeton University.

Rodriguez, German and James Trussell. 1980. *Maximum Likelihood Estimation of the Parameters of Coale's Model Nuptiality Schedule from Survey Data*. World Fertility Survey Technical Bulletins, vol. 7.

Ruggles, S. 1998. *Public Use Microdata of the 1910 Census*. Proposed project, NICHD. Ruggles, Steven. 1987. *Prolonged Connections: The Rise of the Extended Family in Nineteenth Century England and America*. Madison: University of Wisconsin Press.

_____. 1988. "The Demography of the Unrelated Individuals: 1900-1950." *Demography* 25:521-36.

_____. 1996b. "Living Arrangements of the Elderly in America, 1881-1980." Pp. 254-71 in *Aging and Generational Relations Over the Life Course: A Historical and Cross-Cultural Perspective*, edited by Tamara K. Hareven. New York: Aldine de Gruyter.

_____. 1997a. "The Rise of Divorce and Separation in the United States, 1880-1980." *Demography*.

Ruggles, Steven and Russell R. Menard. 1994. *Public Use Microdata Sample of the 1880 United States Census of Population: User's Guide and Technical Documentation*. Inter-University Consortium for Political and Social Research.

_____. 1995. "The Minnesota Historical Census Projects." *Historical Methods* 28:6-10.

Ruggles, Steven and Matthew Sobek. 1995. "Integrated Public Use Microdata Series: User's Guide." Social History Research Laboratory.

Scarano, Francisco A. 1974. "The Puerto Rican Population, 1765-1815: A Statistical Analysis." Master's thesis. Columbia University.

_____. 1984. *Sugar and Slavery in Puerto Rico: The Plantation Economy of Ponce, 1800-1850.* Madison, WI: University of Wisconsin Press.

_____. 1986. "Poblacion esclava y fuerza de trabajo: problemas del andlisis demogrdfico de la esclavitud en Puerto Rico, 1820-1873," *Anuario de Estudios Americanos* XLIII:3-22.

_____. 1990. "Estructuras de la Plantacion Azucarera Esclavista: El Modelo Clasico y Sus Variaciones." *Del Caribe (Santiago de Cuba)* VI(16-17):6-14.

_____. 1993. *Puerto Rico: Cino Siglos de Historia*. Bogota: McGraw-Hill Interamericana.

_____. 1996. "The Jibaro Masquerade and the Subaltern Politics of Creole Identity Formation in Puerto Rico, 1745-1823." *American Historical Review* 101(5):1398-431.

_____. 1999. "Slavery and Emancipation in Caribbean History." In *General History of the Caribbean, Vol VI, Historiography*. Kingston: UNESCO.

Shoemaker, Nancy. 1992. "The Census as Civilizer: American Indian Household Structure in the 1900 and 1920 Censuses." *Historical Methods* 25:4-11.

Sobek, Matthew. 1997. "Occuptional Structure and the Labor Force in the United States, 1881-1990." Ph. D. Diss. University of Minnesota.

Steward, Julian H. 1956. *The People of Puerto Rico: A Study in Social Anthropology*. Urbana, IL: University of Illinois Press.

Sullivan, J. 1972. "Models for the Estimation of the Probability of Dying Between Birth and Exact Ages of Early Childhood." *Population Studies* 26(4):79-98.

Trussell, J. 1975. "A Re-Estimation of the Multiplying Factors for the Brass Technique for Determining Childhood Survivorship Rates." *Population Studies* 29(1):97-107. Trussell, James and Samuel Preston. 1981. "Estimating the Covariants of Childhood Mortality from Retrospective Reports of Mothers." In *Methodologies for the Collection and Analysis of Mortality Data*, edited by Jacques Vallin, John H. Pollard and Larry Heligman.

VandeWalle, Etienne. 1999. "Where Are the Children?" Population Studies Center, University of Pennsylvania.

Vazquez-Calzada, Jose L. 1988. *La Poblacion de Puerto Rico y Su Trayectoria Historica*. Universidad de Puerto Rico.

Wachter, Kenneth W., Eugene A. Hammel, and Peter Laslett. 1978. *Statistical Studies of Historical*

Social Structure. New York: Academic Press.
Wagley, Charles, 1957. "Plantation America: A Culture Sphere." In *Caribbean studies. A symposium*. ed. Vera Rubin. Mona, Jamaica, University College of the West Indies, Institute of Social and Economic Research, in association with Columbia University Program for the Study of Man in the Tropics [1957].
Watkins, S. C. and A. S. London. 1994. "Personal Names and Cultural Change: A Study of the Naming Patterns of Italians and Jews in the United States in 1910." *Social Science History* 18:169-209.
Wolf, Eric R. and Sidney W. Mintz. 1957. "Haciendas and Plantations in Middle American and the Antilles." *Social and Economic Studies* 6(3):380-412.
Zuberi, (Mcdaniel), Antonio and E. M. Zulu. 1996. "Fathers, Mothers and Children: Patterns in Child-Parent Living Arrangements in Sub-Saharan Africa." African Population Studies, December 1-30.

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: palloni@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu