Racial and Educational Differentiation in Extended Kinship Structures in the United States

Jonathan Daw[a]

Ashton Verdery[b]

September 21, 2012

[a]: Institute of Behavioral Science and Institute for Behavioral Genetics, University of Colorado-Boulder

[b]: Department of Sociology and Carolina Population Center, University of North Carolina-Chapel Hill

This research was supported by NIH grants T32 HD007289, and R25 DA026401.

ABSTRACT. Although social differences in kin contact, co-residence, support, and exchange have attracted much demographic research, studies of kinship structures themselves have received substantially less attention. We describe the extended kin structures of the contemporary American population and how these differ by race and education. We employ nationally representative, intergenerational data on sibling and parental ties collected over 41 years to create extended kinship networks through social network methods. Results show that, adjusting for age and the opportunity for a tie in the data, there are no statistically significant differences by race in the presence of any type of kinship tie. However, statistically and substantively significant differences in the distribution of many tie types are observed by education. These results suggest that widely documented racial differences in frequencies and types of interaction with kin may be attributable to cultural or geographic differences in kinship processes rather than the kinship structure itself.

Introduction

Understanding group-level differences in demographic outcomes and influences, especially by race and ethnicity, has long been a core project of American demography. While researchers have looked at inter-group differences in health, sexual activity and contraceptive use (Kirby and Kaneda 2010; Browning, Leventhal, and Brookes-Gunn 2004; Stephen, Rindfuss and Bean 1988), income and homeownership (Bianchi 1980; Krivo and Kaufman 2004), infant mortality and birthweight (Eberstein and Parker 1984; Cramer 1995; Van den Oord and Rowe 2000; Frank 2001; Zuberi 2001; Van den Oord and Rowe 2001), and even census and vital statistic error rates (Coale and Kisker 1986; Fein 1990; Preston, Rosenwaike and Hill 1996), work on group differences in social context seems to have captured the most attention. This line of work looks at the lived experiences of individuals by group, documenting differences in, e.g., neighborhood and social network density of same group ties (Farley 1977; Bruch and Mare 2006; McPherson, Smith Lovin and Cook 2001; Moody 2001; Berry 2006), knowing someone in prison (Zheng, Salganik and Gelman 2006), and household and family structure (White 1981; London and Elman 2001; Page and Stevens 2005).

This article extends work on inter-group differences in social context by focusing on differences in kinship. Surprisingly, demographers have not paid sufficient attention to kinship, despite being well poised to do so (Entwisle 2007; Mare 2011). Kinship is one of the most important variables across a broad range of social science topics – e.g., it has been demonstrated important for economic factors (e.g., Grieco 1987; Zimmer and Aldrich 1987), demographic factors (e.g., Tilly and Brown 1967; Sandberg 2005), political factors (e.g., Padgett and Ansell 1993; Hammel 1968), health and psychological factors (Christakis and Fowler 2008; Fowler and Christakis 2008), and sociological factors (e.g., Vaisey and Lizardo 2010; Mills 1958). In terms of analytical scale, kinship as explored in this paper falls between neighborhoods and acquaintance networks at the macro level, and household and family structures at the more micro-level.

Prior work on social differences in kinship has been quite varied, but, we argue, has not proceeded with a demographic perspective. A core component of the demographic perspective is a focus on opportunity structures, often referred to as the population at risk of an event occurring. The most familiar examples of this are the various differences between types of rates; e.g., the total fertility rate vs. the crude birth rate. A number of demographic analyses have looked at group level differences in various activities people might engage in with kin – co-residence, contact, discussing important matters, providing assistance, etc. – but we know of no studies to date which have looked at group differences in the opportunity structures – the number and nature of various types of kin or friendship ties – which condition the possibility for these activities to occur (though some papers look at age or cross-national differences in kinship). Without this basic descriptive understanding, researchers may be misattributing the sources of group differences in kinship activities. For instance, they might argue that cultural, psychological or socio-economic differences between groups is the source of differing levels of, for instance, discussing important matters (McPherson et al. 2006), when in reality there are no differences in such activities conditional on the opportunities that individuals have to engage in them.

In this study, we extend demographic understandings of racial and socioeconomic differentiation in kinship structure through two primary contributions. First, we estimate the typical distribution of broadly-defined kinship ties in the U.S. as well as variation therein. Second, we describe racial and educational differences in the central tendency these distributions and discuss the implications of these findings for understanding racial and educational differences in household coresidence and different forms of social support[1].

Background

In our exploration of kinship, we draw heavily on the theories and methods of social network analysis. Conceptually, we follow House et al. (1988) in distinguishing between three different

[1] Before beginning, we wish to offer a key caveat. Our intent in this paper is to describe group differences in kinship, not to attribute those differences to membership in said groups. Thus, when describing educational differences in kinship, we are explicitly not asserting that educational attainment leads people to have differing numbers of kin as it is highly probable that the relationship goes in the other direction (differing kinship structures constrains individuals opportunities for educational attainment). Instead, the purpose of this paper is to provide a – to the best of our knowledge – first glimpse at how extensive kinship patterns are by groups.

approaches to studying social networks: studies of dyadic relational content such as social support and contact, ego-network studies of social integration and isolation, and studies of network-level structure. The first approach, which is the most common type of study in this literature, focuses not on the structure of the network itself, but some characteristic of different types of dyadic ties such as the exchange of monetary assistance (e.g., Hofferth 1984; Hogan et al. 1993; Cohen & Casper 2002). While important, these studies only incompletely capture the kinship structure because they measure only those kinship ties across which financial assistance flows. In addition, their observation is conditional on the ability of individuals to provide assistance and others' need for it. The second approach involves measuring, for instance, whether one has contacts with whom to discuss important matters, many of whom are kin (e.g., Fischer 1982, 2009; Lin 2001; McPherson et al. 2006, 2009), or else the frequency of contact with kin (e.g., Murphy 2008; Raley 1995). Such studies are primarily studies of relational content because the set of persons with whom one could have such a relationship is much broader than the typically low number of ties that this 'core discussion' network includes. Finally, the third type of study closely approximates our goal for the present study. These typically attempt to measure, for instance, the density, triadic structure, and connectivity of ties between nodes within a population of interest. In this section, we review these studies to situate the importance of kinship within a broader context, to demonstrate the extent to which prior research on group differences in kin relations has not accounted for opportunity structures, and to see whether that literature offers hypotheses about the extent and direction of inter-group differentiation.

Previous Investigations of Social Differentiation in Family and Kinship

The literature on racial differences in kinship network structure is surprisingly sparse. Most previous investigations of racial, ethnic, and (less commonly) educational differentiation in family and kinship do not examine the kinship structure itself, or else does so in a limited fashion. Rather, most looks at differences in household co-residence (e.g., Elmon and London 2011; Ruggles 1994, 1996; Angel and Tienda 1982; Hofferth 1984), contact (e.g., Raley 1995), and support (e.g., Mazelis and Mykyta 2011; Sarkisian and Gerstel 2004). This research typically finds that African Americans have larger households on average (Choi 1991; Kamo 2000; Peek et al. 2004), and higher fertility rates (Census 2011), from which one could hypothesize that African Americans have larger kinship networks on average than whites. Although all of these variables may be related to kinship size and are important in their own right, in absence of a detailed literature on these factors this connection cannot be assumed. Furthermore, it is uncertain whether these descriptions describe racial differences in kinship structures in a representative and complete manner. For instance, it is difficult to generalize from qualitative investigations of these attributes. Social support studies focus exclusively on activated ties rather than the availability of genealogical ties to living relatives. Similarly, in the cross-section, household co-residence ties focus only on a particular sort – those currently residing in the same household. Although all of these lines of research usefully contribute to demographic understanding of family and kinship in the U.S., they are no substitutes for studying the structure of kinship itself. By investigating the structure of kinship networks by race in the U.S., we have the potential to understand how these relational measures arise, which ones don't become activated, and how multiple generations influence demographic processes for the focal one (Mare 2011).

Studies of Household Co-Residence

Household structures will be related to kinship structures because most people in the U.S. live with biolegal kin. However, household co-residence is primarily a form of relational content - co-residence is one type of relationship one can have with one's kin. Most recent research on patterns of household co-residence have been concerned with the emergence of non-nuclear households, also called 'extended' household structures. While the 'traditional' nuclear family consists of two parents and their young biological children living together in a common household, this actually was a fairly recently-burgeoning family form itself, thought to have evolved in response to the rise of economic industrialism (Burgess 1916; Parsons 1949). This form included particular focus on emotional support on those within the nuclear family, often to the exclusion of those without (Bengtson 2001; Hogan et al. 1993; Swartz 2009) as well as the homemaker/breadwinner gendered division of labor (Bengtson 2001; Swartz 2009).

The emergent extended kinship household, in contrast, is thought to have arisen as a result of a combination of demographic and socioeconomic changes (Swartz 2009). First, these family structures are typically three-generational, which requires that grandparents and grandchildren be alive together for substantial periods of time. Until recently in human history this has been comparatively uncommon - for instance, life expectancy at birth in the U.S. was 49.2 in 1900, compared to 77.9 in 2009. This large change in human longevity means that concurrently living grandparent-grandchild ties are far more common than in the past, and three-generational household structures are therefore more commonly possible (Crosnoe & Elder 2002; Logan & Spitze 1996; Mills et al 2001; Silverstein & Marenco 2001; Uhlenberg 1996; Setterson 2007). Furthermore, changing patterns of fertility over this same time period have resulted in decreasing birth parity per generation, resulting in what is frequently called a 'beanpole' kinship structure (Bengtson 2001; Uhlenberg 1996) consisting of multiple generations of married pairs with two or fewer children apiece.

These patterns are strongly related to social position, as intergenerational co-residence among adults parent-child pairs is far more common among African Americans, Latinos, and lower-educated persons compared to whites and higher-educated persons in the U.S. (Angel & Tienda 1982; Hofferth 1984; Mollborn 2011). It is frequently argued that this household structure is a practical response to the demographic conditions of social disadvantage in socioeconomic resources and characteristics of the marriage market (Swartz 2009; Crimmins & Ingegneri 1990; Glick 1999; Kamo 2000; Sarkisian & Gerstel 2004b; Sarkisian et al. 2007; Van Hook & Glick 2007).

Studies of Kin Contact

The frequency of contact with kin among Americans is typically high, especially intergenerationally. About 80% of parent-adult child pairs speak together at least once a week, although contact with mothers is more common than with fathers (Lawton et al. 1994; Connidis 2001; Swartz 2009). Furthermore, parents and children who are not co-resident typically live fairly nearby - more than 50% of adult children with living parents live within an hour's drive thereof, and this is even more common for frail, aging parents (Lawton et al. 1994; Lin & Rogerson 1995; Connidis 2001; Lye et al.

1995). Because grandparent-grandchild contact is typically mediated through the middle generation, this also means that grandparents often have frequent opportunities for contact with their grandchildren (Silverstein & Marenco 2001).

The frequency and type of kin contact is strongly related to social position by race, ethnicity, gender, and socioeconomic status. Women, in line with their frequent role as 'kin keepers,' are in more frequent contact with their kin than men (Lee et al. 2003; Eggebeen & Hogan 1990). Race and socioeconomic status are also strongly related to these processes, as both African Americans and working class persons are more likely to live near their kin and have more frequent contact with them than whites and middle class persons respectively (Sarkisian 2007; Connidis 2001). Finally, these racial differences in the frequency of contact with non-co-resident kin are shaped by gender, as well - African American are in more frequent contact with their kin than white women, whereas the reverse is true when comparing men by race (Raley 1995).

Studies of Relationship Satisfaction

Marital quality and satisfaction are important predictors of marital disruption and therefore family structure (Amato & Rogers 1997; Veroff et al. 1995). SPELL OUT WHY. Previous research shows that relationship satisfaction is shaped by race and socioeconomic status in important ways. For instance, African Americans report lower relationship satisfaction and experience separation and divorce at higher rates than whites (Adelmann et al. 1996; Broman 1993, 2005; Cherlin 1998; Kposowa 1998; Landale & Oporesa 2007; Lichter et al. 1992; Phillips & Sweeney 2006; Tucker & Mitchell-Kernan 1995), and differences in the financial situation of whites and blacks explain an important proportion of these differences (Broman 1993, 2005; Clark-Nicolas & Gray-Little 1991). Accordingly, higher SES is related to higher marital stability and relationship quality (Conger & Elder 1994; Karney & Bradbury 2005; Heaton 2002; Martin 2006; Orbuch et al 2002; Popenoe 2007; South 2001; Stanley et al 2006; Amato et al. 2007; Conger et al 2010; Duncan et al. 2007). This association appears to result from a combination of both causal and selective mechanisms (Schoon et al. 2002; Wickrama et al 2008; Haas 2006; Conger et al

2010).

Studies of Kin Support

Kin support plays an important role in the lives of most Americans, and is linked to physical and mental health trajectories and general well-being. Kin support may be divided into two primary types: instrumental support such as financial exchange and child care, and affective support including such as being there for kin during trying times (Swartz 2009). Both types of ties have a large literature linking them to social position by race and socioeconomic status. As such this review will be necessarily brief due to space considerations.

A long-lived literature investigates the link between race, ethnicity, socioeconomic status, and kin orientation (e.g., Aschenbrenner 1973; Stack 1974; Moynihan 1965). Countering the Moynihan report (Moynihan 1965), a large literature arose arguing that, although the ‘traditional’ nuclear family among African Americans faced more instability, this was compensated for by stronger extended kinship ties (Aschenbrenner 1975; Martin & Martin 1978; Stack 1974). As such the frequent assumption in this line of research was that African Americans should exchange more assistance with their extended kin than do whites. However, the empirical reality has turned out to be more complicated, such that whites tend to exchange more financial and emotional support with their extended kin whereas African Americans on average provide more practical and housing assistance (Swartz 2009; Berry 2006; Cohen & Casper 2002; Eggebeen & Hogan 1990;Hofferth 1984; Hogan et al 1990, 1993; Kamo 2000; Lee & Aytac 1998; Sarkisian & Gerstel 2004; Sarkisian et al. 2007; Mazelis & Mykyta 2011; Cooney & Uhlenberg 1992; Eggebeen 1992; Hogan et al 1993; Hoyert 1991; Goldscheider & Goldscheider 1991; Roschellle 1997).

It is typically thought that these patterns are explainable as a joint function of racial differences in ability to assist and need for assistance (e.g., Jayakody 1998; Berry 2006). Furthermore, kinships of all races show evidence of strong reciprocity norms; since African Americans have far less wealth on average than whites (e.g., Oliver & Shapiro 1995; Conley 1999), they may be less likely to accept financial support than whites because they are less able to reciprocate in kind (Antonucci et al 1990; Malson 1983; Martin & Martin 1978; McAdoo 1982; Testa & Stack 2002). Additionally, this process works in reverse as well - because of the degree of racial stratification in financial resources and the association of kin resources, those with the greatest need for assistance frequently have the least access to it through kinship networks (Harknett & Hartnett 2011; Heflin & Patillo 2006). It is also frequently argued that core demographic processes such as intergenerational spacing plays a role in creating racial disparities in kin-based resource exchange, as tighter spacing frequently results in greater dependencies on middle aged persons, limiting the potential for wealth accumulation (Burton 1996; Jarrett & Burton 1999; Burton et al. 1995; Swartz 2009). Finally, the race-exchange relationship is somewhat self-perpetuating, as being kin with poor persons is independently associated with low rates of resource accumulation due to one's kin frequent need for financial assistance (Stack 1974; Heflin & Patillo 2002).

Contributions of the Present Study

The literature just discussed has shed a great deal of light on the facts of kinship in the U.S. However, it has only dimly illuminated the distribution of characteristics of kinship networks themselves. A number of basic questions about kinship remain surprisingly unanswered: how many living grandparents, parents, children, aunts/uncles, cousins, and nieces/nephews do persons of different ages, racial background, and socioeconomic status typically have? How much dispersion is there in these characteristics? At the network level, what is the distribution of different types of dyadic ties by race and socioeconomic status, how large is the typical kinship network, and how much dispersion is there in this characteristic?

Although the questions are demographically important in themselves, answering them well will also contribute to demographic understanding of the functions of kinship, as measured in the previously discussed studies of contact, co-residence, relationship quality, and support. Crucially, it is difficult to truly understand how racial or socioeconomic differences in financial support by kin tie type arise if one does not know the number of different sorts of kin one is related to. For instance, suppose that African Americans are more likely to receive financial support from cousins than are whites - this could be because each African American's cousin is more likely to provide support, or it could be because each such kin has an equal chance of providing support regardless of race, but African Americans have more cousins on average. Until the distribution of kin of different types is measured, it is impossible to distinguish the two. Similar claims could be made about studies of kin contact, affective support, and co-residence. Just as it is essential to measure whether one is in a romantic relationship before one can measure relationship quality, it is equally important, though less obvious, to investigate what kin are available to an individual before studying the nature of their relational ties. This argument relies on the bedrock demographic emphasis on distinguishing between persons at-risk for a phenomenon and those who experience it. In this case, no social exchange can occur between a kin pair which does not exist. Indeed, these arguments are familiar to demographers as they form the essence of the demographic perspective/comprise the thrust of American demographic research over the past 60 years. I feel like we should relate them to demographic ideas of at risk and what not.

As a first step toward this goal, this study employs data from the Panel Study of Income Dynamics Family Information Mapping System file to study the distribution of characteristics of white and African American kinship networks in the U.S. Our analysis will: describe the distribution of characteristics of this dataset; descriptively analyze the distribution of biolegal kinship ties from the network and ego perspective; describe racial and socioeconomic differences in these characteristics; and then validate our findings using more limited GSS data and more assumption-laden simulation techniques.

Data Source

Data for this study are drawn from the Panel Study of Income Dynamics Family Information Mapping System (PSID FIMS). The PSID began in 1968 by following a nationally representative, household-based sample of over 18,000 respondents. It has been collected since 1968, with data available through 2009, the PSID follows the original, nationally representative sample of U.S. households using a genealogical design - as members of the original households left home, the study followed these new households, as well. Samples were collected annually from 1968 to 1997; thereafter they were collected biannually. Since most households consist of biolegal kin, this means that this study tracks the evolution of biological and marital lineages over a 41 year span. The FIMS dataset provides linkage variables delineating parent-child (biological and adoptive) and sibling (distinguishing full-, half-, and step-siblings) ties among observations; marital ties are determined using the primary dataset. Because these are the three elementary ties in a kinship network (Batagelj & Mrvar 2007), this means that all existing biolegal ties among sample members may be calculated using matrix multiplication of these linkages, as described below.

Over time, 71,285 individuals have been interviewed as part of the PSID. However, not all of these observations are employed in the present analysis, for several reasons. First, not all persons interviewed were part of the biolegal networks of the original households, as many were roommates and non-married significant others of members of these lineages. These persons are excluded from the present analysis. Second, refreshed households were added to the study over time to maintain the population representativeness of the sample. The largest supplementary sample (added in 1990) consisted of Latino households to account for the substantial increase in the population share of Latinos since 1968. However, because of the much shorter timeline available for this family which would result in far less fully characterize kinship networks, these data are not analyzed in this paper. Third, many persons interviewed in this dataset died and/or permanently left the sample over time. Persons who died are included in the kinship network calculations (described below), but are not counted as ties themselves. (So, for instance, if one's parent died but is survived by a sister, that person will still count as one's aunt, although the deceased parent will not be counted.) Finally, many persons who are part of the biolegal networks of the original households joined by marrying into the lineage; while these persons are important parts of the lineage, their own lineages are necessarily incomplete because there is no chance that their parents, siblings, cousins, etc. were interviewed due to the study design. Thus these persons are counted as ties for focal observations but not as observations in themselves.

Measures

Dependent variables: Kinship Ties

In contrast to other work on kinship which has primarily focused on a small set of ties specific to the question being explored – e.g., studies of intergenerational support looking at child and parent ties –

our analysis describes a broad range of different kinship ties. Specifically, we focus on parent, child, spouse, full sibling, half sibling, grandparent, grandchild, aunt and uncle, niece and nephew, and cousin ties. These ties are all potentially important conduits of social support, information and influence and represent both inter- and intra-generational ties. Of course, other ties could also be examined – e.g., in-laws, second-cousins, step-siblings – but the data are less well suited to examine such types of ties, as the next section makes clear.

We characterize kinship ties using a modification of the methods described by Batagelj and Mrvar (2007), incorporating information on biological, adoptive, and marital ties (henceforth, biolegal ties) to characterize the full kinship networks. The key intuition is that all biolegal kinship ties can be defined as a function of three elementary matrices - parent matrices (**P**, a non-reciprocal matrix in which person *j* is person *i*'s parent if $\mathbf{P}_{\mathbf{i,j}}$=1 and =0 otherwise), sibling matrices (**S**, a reiprocal matrix in which *j* is *i*'s sibling if $\mathbf{S}_{\mathbf{i,j}}$ =1), and spousal matrices (**M**, a reciprocal matrix in which *j* is *i*'s spouse if $\mathbf{M}_{\mathbf{i,j}}$=1). For instance, one's grandparent is one's parent's parent, one's aunt is one's parent's sibling or the spouse of one's parent's sibling, and one's cousins are one's parent's parents' children's children who are not one's siblings or alternatively, one's parents' siblings' children (we did not differentiate half cousins).

These measures are operationalized in two different ways. We begin with a structural analysis by describing the size distributions of kinsets of several of the kin-relation types we focus on. Doing so lets us examine the entire distribution, which is important because it helps situate our second set of analyses – conducted at the individual level – within a group context. Because the notion of a kinset is most comprehensible for reciprocated ties, we limit our analyses to the tie types which are reciprocal (e.g., sib-sets and cousin-sets). We then turn to an individual level analysis. Here, we look at the individual distribution of counts of each type of tie, i.e., how many ties of a given type we observe for the average person. We explore racial and ethnic group differences in these counts for each tie type. *Key confounders: Opportunities for ties*

Features of the PSID-FIMS study design present potential problems for our analysis. The danger is that aspects of how individuals come into the study might affect our likelihood of observing certain types of ties from or to them. This arises from the twin facts that (a) we can only define kinship ties between individuals who co-resided at some point between 1968 and 2009, and (b) that there are three means of entering the subset of the PSID survey we use: being in the original wave in 1968, being born to a member of the original wave or one of his or her descendants, or marrying someone already in the study[2]. Generally, this means that kinship data will be more complete for individuals who are younger in the most recent wave of the survey as we will have had time to observe the generations before them. It also means that those who marry into the survey will have incomplete data as we will not observe the structure of their prenuptial households. Additionally, these effects will interact and we will have limited ability to see complete counts of, for instance, grandparents for younger generations because we will only observe one parents' parents.

To cope with these issues and ascertain the extent to which they bias our measures of kinship and racial and ethnic differences therein, we define a measure of an individual's *opportunity for a tie* (OFT). Figure X describes the tests which we conducted for each person.

Generation within a lineage is defined by the number of parent-child ties preceding oneself in a lineage. So, first generation (G1) respondents are those who have no parents, grandparents, great-grandparents, etc. in the PSID dataset; second generation (G2) respondents have a parent but no grandparent or higher; third generation (G3) respondents have a grandparent but not great-grandparent or higher; and so on. These measures are defined without regard for whether the kin in question is still alive in the 2009 wave. For those who are not direct biological or adoptive descendants of the 1968 householders, for certain purposes the generation of one's spouse who is such a descendant is assigned to the person in question. Generations are not defined for those who are neither descendants nor other biolegal relatives of the original householders.

The type-specific measure that we call the '*opportunity for tie*' (OFT) is a measure of whether an individual's place in the PSID data collection design means that it is very unlikely that a tie of a specific

[2] One may also be a roommate with an original member or his or her descendants, but we exclude those people never linked by kinship as described in the data section above.

sort will be observed for that individual. For instance, under the sampling design it is very unlikely that we will observe the grandparent of a G1, because these ties are only observed when one is G3 or higher, because observing a grandparent requires observing one's parent's parents. Whether one has the opportunity for a specific type of tie, however, is defined differently by the type of tie of interest, and depends on one's generation in the lineage and age, as follows:

OFT(C): =1 for any generation, age $\geq$20

OFT(HS,FS,P,NN): G2 or higher

OFT(GP,AU,CO): G3 or higher

OFT(GC): Any generation, age $\geq$40

OFT(M): Any generation, age $\geq$18,

assuming a minimum marriage age of 18 and minimum intergenerational interval of 20 years. Although these assumptions are surely violated in a few cases, specifying a reasonable cutoff ensures that intergroup comparisons of tie distributions are subsets to demographically reasonable subsets in each group so that the comparison is valid.

Independent variables: Race, Education, and Age-category

One of the key contributions of this paper is to describe group differences in kinship distributions. While we used all individuals ever in the subset of the PSID dataset described above to define kin ties (but only counted ties to kin who are currently living) and the opportunity for such ties, we define the following independent variables only for each individual who is living in 2009. *Race* is defined at a individual level for those alive in our analytic sample in 2009, following the household-based collection design of the PSID, and for analytical purposes is defined as having only two categories - white and black, defined as the race of the householder in the original wave of data collection. All others are dropped from the analysis for the sake of comparability. *Education* is individually-defined, and is measured as the most recent valid response (in 2009) to the question, "What is the highest grade or year of school that (he/she) has completed?" which is recoded into four categories: less than high school, high school, some college, and a 4-year degree or higher. *Age* is measured as either one's PSID-calculated age in 2009 or, if one is still alive but a non-respondent in 2009, the last valid reported age plus the differences in years since that report. *Living status* is tracked by PSID between survey waves, and is measured as whether the respondent died since the last valid interview. If death is reported for the respondent in between any survey waves, they are counted as dead in 2009.

Analysis

The analysis proceeds as follows. We begin with descriptive measures that focus on the distribution of individuals' kinship ties and the distribution of the sizes of kin groups – sib-ship, cousin-ship, and all kin-ship sets – in the population. After this we turn to inferential methods that are designed to cope with the complex nature of our kinship data. Specifically, we use a series of zero-inflated Poisson regression models predicting counts of each type of kin. Poisson regression models are appropriate for analyzing the distribution of count dependent variables, as is the case in the present analysis. However, often count dependent variables are dominated by zeroes disproportionate to the expectation in different Poisson distributions. A common reason for there to be more zeroes than would be expected from a Poisson process is that the extent of zeroes might be influenced by a separate process; diagnostic tests of our data indicate that this is the case. To cope with these separate processes – one influencing the likelihood of zeroes and one influencing the likelihood of other counts –zero-inflated Poisson regression models consist of two models – a logistic (or probit) model predicting zero counts, and a Poisson regression model predicting non-zero counts (Long 1997). As our theoretical interest in kinship focuses on racial and educational differences in numbers of available kin, we run two models for each type of kin tie. Both models use age (specified cubically) as a critical control because kinship opportunities are directly related to age (e.g., it is more likely that a thirty year old has a child than a 20 year old). The first model then interacts age with race (model A), while the second interacts age with education level (model B). Hereafter, we refer to the race models as A models and to the education models as B models. Within each model, we must also adjust for the over-inflation of zeroes. To do this, each model – following the standard procedures for zero-inflated Poisson models – also includes a separate equation predicting the likelihood of zeros as a function of cubic age. In all our regression models, standard errors are adjusted for the non-independence of observations within lineages using the sandwich estimator (Rogers 1993). We subset the analysis to those who have the opportunity for that tie (as defined above), to age ranges for which at least 100 observations with the opportunity for that tie are observed in the analytical dataset, and to individuals who are direct descendants of the heads of households of the original PSID household. The first of these is done because of structural inadequacies of the PSID-FIMS data we use. The second because there are age-based demographic processes which eliminate the potential for certain ties in certain age groups (i.e., ten year olds cannot be grandparents). And the third is done to ensure that we use only the highest quality data available; it has the effect of excluding unrelated roommates, for example.

The third and fourth portions of our analyses focus on further understanding racial and educational discrepancies in kinship counts. To do this, for both our race and education models, we run two complementary models that control for counts of how many deceased relatives of each type one has by 2009 (models A2 and B2), and how many 'intervening' ties one has for a count of that type (models A3 and B3). Counts of deceased ties are calculated based on how many alters occupied the ego-network position in question but who are not alive in 2009. Controlling for this and comparing the race or education effect between models controlling and not controlling for this will indicate the role of kin mortality in creating group differentials in counts of living ties in the 2009 data. We define 'intervening' ties as those ties required to observe the tie type in question. For instance, one must have children before one can have grandchildren. So, controlling for one's count of children in a model predicting counts of grandchildren controls for the effect of intervening ties in group differences in grandchild counts. Similarly, one must have aunts/uncles to have cousins, and full or half siblings to have nieces or nephews. The count of ever-living ties is used to measure intervening tie counts.

Results

Descriptive Statistics

Figure 1 shows the distribution of all measured kin across individuals, encompassing all spouses, full and half siblings, parents, grandparents, children, grandchildren, aunts, uncles, nephews, nieces, and cousins, as well as step-children, step-parents, sisters- and brothers-in-law, and step-grandparents, -

grandchildren, -aunts, -uncles, -nieces, -nephews, and other living members of one's kinship network. As can be seen, kinship networks come in a wide variety of sizes, and some observed individuals have 50 or more living relatives whose relation is within the logical bounds of closeness. The distribution is skewed right, and has a median of 12, a mean of 14.8, and a standard deviation of 10.8, and a range of 1 to 122. Only those with any observed relatives are included in this and subsequent analyses, resulting in an analytical sample of 35,484.

Figure 2 depicts the logged component size distribution for all relatives, full siblings, half-siblings, and cousins. A component size distribution is a common descriptor in social network analysis (e.g., Newman et al. 2001), and is used to characterize the frequency with which we observe sets containing a given number of persons of a given tie type. As can be seen, we most frequently see small sets of half-sibships , followed by full siblings and cousin sets, which are commonly fairly large. Obviously, the distribution of all relatives is substantially larger. All such sets show that larger sets are less common than smaller sets, and that there is considerable diversity in the size of these sets.

Age Patterns of Full and Half Siblings by Race and Education

Figures 3A and 3B show patterns of full sibling counts by age, race, and education as predicted from the zero inflated Poison models described above. Both show that, by current age patterns, there is a curvilinear relationship between age and the number of living full siblings one has. Figure 3A shows that, for both whites and blacks, Americans under 30 average approximately one living sibling; for older Americans, however, this figure is substantially higher. This pattern is to be expected from the decline in fertility over the last half century. Similarly, patterns of full sibling counts by age and education (Figure 3B), show similar shapes for all educational groups. However, whereas we found no significant racial differentials in sibling counts, we see in figure 3B that this curvilinear relationship is substantially higher for those who did not graduate high school; persons 50 and older average 2-2.5 siblings whereas more educated groups average approximately 1.5[3].

[3] The race differences in the regression model are not statistically significant, the educational differences are – there is a statistically significant (p<.10) negative main effect of attending some college or having a bachelor's degree, a

Figures 4A and 4B show patterns of half sibling counts by age, race, and education as well. As seen in Figure 4A, there is a negative relationship between age and counts of half siblings for both races, especially for those 40 and older; this likely reflects the trend toward increased prevalence of divorce and remarriage in American society. As with siblings, though, we see no statistically significant differences in half sibling counts by race. There are, however, large and statistically significant differences in predicted half-sibling counts by education. Across the age range, those who did not graduate high school have the highest expected counts, followed by high school graduates, college dropouts, and college graduates. The college graduate-high school dropout differences are always statistically significant from one another throughout the age distribution. (However, the regression coefficients on which this model is based do not achieve conventional significance.) A notable feature of this graph is the difference in age trends amongst those under 35 for the BA+ group vis-à-vis the less educated groups. While we see a positive association with age for the BA+ group, we see negative associations for all other groups.

Age Patterns of Parent and Child Counts by Race and Education

Figures 5A and 5B depict age patterns of parental tie counts by race and education.[4] Figure 5A shows that the expected count of one's living parent ties declines monotonically with age, a process that accelerates somewhat in middle age. Notably, however, there are no race differences in these patterns. In contrast, Figure 5B depicts large, statistically significant differences in expected parental ties by education, but, like figure 5A, shows a general decline amongst older age groups. We suspect that mortality differentials between the parents of those with greater and lesser amounts of education are driving these distinctions, differentials which may be greater than those between races. These outcomes are roughly ordered by educational attainment, although there are two crossovers in predicted values between high school graduates and college dropouts. Statistically significant regression coefficients are observed for the main and interactive effects of high school graduation compared to high school dropouts.

---

statistically significant positive interaction of age and the same two categories, a negative, statistically significant interaction of those categories and age-squared, and a positive interaction of those categories and age-cubed. The age interactions for attending some college are uniformly borderline statistically significant (p<.10) while the interactions with college graduation are uniformly fully statistically significant (p<.05).

[4] These models have not yet fully converged according to our criteria.

Figures 6A and 6B portray age patterns of child tie counts by race and education. As seen in Figure 6A, there is a curvilinear cross-sectional relationship between age and child tie counts, such that the derivative of counts with respect to education is higher in the late 20s and early 30s than at other points in the distribution. Overall there is a positive relationship between age and child tie counts throughout the distribution, and no statistically significant differences by race. However, there are large and statistically significant differences in expected child counts by education, with counts generally inversely related to educational attainment until the oldest ages examined. Among those 40 and younger in 2009, college graduates have statistically significantly lower counts of children than all other groups, as do college dropouts compared to high school graduates and high school dropouts. There are no statistically significant comparisons between high school dropouts and high school graduates in the regression coefficients, although the predicted values are statistically significantly different at older ages in Figure 6B. Finally, the education pattern is somewhat upended among those approximately 45 and older, as college graduates have higher expected child counts than college dropouts among these older groups. The underlying regressions show statistically significant main and interactive effects of education for college dropouts and college graduates compared to high school dropouts.

<u>Age Patterns of Grandparent and Grandchild Counts by Race and Education</u>

Figures 7A and 7B depict patterns of grandparental ties by race and education.[5] It should be noted that, with the data collection design of the PSID, no more than two grandparental ties can be observed except through remarriage. Nonetheless, differences in the age patterns of these ties should still be telling. Figure 7A shows that the expected count of living grandparents is sharply and approximately linearly age-graded, as those in their mid-30s in 2009 have an expected count of less than 0.5 living grandparents (out of a possible 2 given the structure of the dataset). Furthermore, Figure 7B shows that this pattern is strongly influenced by the grandchild's educational attainment; although there are crossovers in predicted values, at older ages in the model age range high school dropouts and graduates have significantly lower expected counts of grandparental ties than college dropouts and graduates. In the underlying regression

[5] These models have not yet fully converged according to our criteria.

model statistically significant interactions in the effect of age and college graduation are observed compared to high school dropouts. Statistically significant predicted counts by education are observed in older ages only.

Figures 8A and 8B illustrate patterns of grandchild ties by race and education. As seen in Figure 8A, there is a curvilinear upward relationship between age and grandchild counts that is statistically identical by race. Few under 50 have grandchildren, yet expected counts accelerate rapidly above that age. A similar pattern is observed in Figure 8B,[6] except that there are statistically significant differences in the regression coefficients by education and age, such that college educated persons have significantly different age patterns and counts of grandchildren than high school dropouts.

Age Patterns of Aunt, Uncle, Niece, and Nephew Counts by Race and Education

Figures 9A and 9B portray age patterns of aunt/uncle counts by race and education. Results show a curvilinear relationship between age and aunt/uncle counts, such that persons between approximately 18 and 30 have higher expected counts than those older or younger. However, no differences are statistically significant by race or education.

Figures 10A and 10B portray age patterns of niece/nephew counts by race and education. Predictably, there is a curvilinear age relationship with this count, such that persons 20 and younger have few nieces and nephews while older adults have much higher counts. Although there are no statistically significant differences in these counts by race, large differences are observed by education at older ages. For those approximately 45 and older, those with less than a high school education are expected to have statistically significantly more nieces and nephews than are there more highly educated peers. However, the main and interactive effects of educational attainment are not statistically significantly different in the underlying regression models.

Age Patterns of Cousin Counts by Race and Education

[6] This figure contains an error that we have not yet been able to resolve, as the age patterns of grandchild counts by education are presently incomplete.

Figures 11A and 11B depict age patterns of cousin counts by race and education. There is a concave down relationship between age and cousin counts for both whites and blacks, such that persons in their mid-20s are expected to have the most cousins, and those older and younger are expected to have fewer. Furthermore, Figure 11B shows that these age patterns are statistically significantly shaped by educational attainment, such that higher educated persons are expected to have lower cousin counts throughout the age range examined, with an especially large gap between those with less than a high school education and all others. The main and interactive effects of high school graduation are statistically significantly different compared to high school dropouts in the underlying regression model.

Age Patterns of Spouse Counts by Race and Education

Figures 12A and 12B depict age patterns of spouse counts (including current and living former spouses) by race and education. The age pattern, depicted in Figure 12A, shows a steep upward curve in spousal counts between ages 21 and 26, followed by a more gradual increase throughout the range. Furthermore, this is the only case we have observed in which a substantial number of statistically significant contrasts in expected counts are observed by race; here, between ages 35 and 49, statistically significant differences in the predicted counts are observed such that whites have higher spousal counts than blacks. However, these differences are not statistically significantly different in the underlying regression models. Differences in age patterns of spouse counts are even starker by education, as seen in Figure 12B. Between ages 30 and 60, high school dropouts consistently have lower counts of spouses than more highly educated persons. At younger ages, college graduates have among the lowest expected spouse counts, but by the top end of the age range they have higher expected spouse counts than any other group. At the youngest ages of the examined range, college dropouts have the highest expected counts. These predicted differences are statistically significant in the underlying regression model, as well, as the main and interactive effects of college graduation (compared to high school dropout) are statistically significant.

The Role of Kin Mortality in Educational Differentials in Kin Counts

Table 1 presents selected results from nested models comparing education effects in models with and without controls for counts of deceased ties of the indicated type (i.e., contrasting the results of the A models with the A1 models). Differences in the educational coefficients between paired columns will indicate the effect of kin mortality in group differentials in ties of the indicated type. In the interests of space only pairs of models with substantively interesting results are presented. Although these models were fit by race as well as by education (not shown), controls for deceased ties did not affect the substantive conclusions of any models by race.

Column (A) of Table 1 compares the effects of education on full sibling counts with and without controls for counts of deceased siblings. The results show that controlling for counts of deceased siblings completely eliminates the statistically significant influences of education on sibling counts. This suggests that sibling death is responsible for much of the observed educational differences in full sibling counts observed.

Similarly, controls for the number of deceased spouses one has had reduces the magnitude of most of the education coefficients on spouse counts. In particular, the main effect of BA+, and its interactions with age and age^2 are substantively reduced when counts of deceased spouses are controlled for. This finding suggests that spousal mortality partly underlies educational differences in living spouse counts.

Finally, controlling for one's number of deceased cousins introduces statistically significant effects of being a high school graduate (compared to a high school dropout) on living cousin counts. The main effect of being a high school graduate is fully statistically significant in the model with this control, and the interactive effects of HS with age and age2 are borderline statistically significant in this model. In contrast, no statistically significant coefficients were observed in the model without this control, suggesting that patterns of cousin death partially suppress a HS-<HS difference in cousin counts.

The Role of Intervening Ties in Educational Differentials in Kin Counts

We conducted a similar exercise controlling for counts of intervening ties when modeling counts of ties which required one (i.e., contrasting the A models with the A2 models). We found one interesting pattern (not shown). There is a statistically significant difference between the age patterns of grandchild counts of high school dropouts and college graduates. However, controlling for counts of the intervening tie (children) eliminates these statistical differences completely. This suggests that the educational difference in grandchild ties is entirely due to the educational difference in child ties.

Conclusion

To our knowledge, this is the first paper to describe full kinship structures in a population-based study in the United States. Previous research on kinship in the U.S. relies on household co-residence, contact, and exchange. These are very valuable investigations, but they do not measure the kinship structure itself – the distribution of the set of social ties bound through biological, marital, or adoptive ties that constitute the extended family. Although our data are somewhat limited, this investigation goes further in its ability to investigate race and socioeconomic differences in this structure than any previous investigation of which we are aware.

Our investigation is subject to a number of limitations. First, kinship tie data are only available when chains of household co-residence and parental, sibling, marital, or adoptive ties connect individuals. It could be that some kinship ties occur without the chain of household co-residence fully linking the individuals in question. Similarly, the PSID data is limited to a single lineage, and the kinship ties of those who marry or co-reside into the lineage in question are not measured. We partially address this limitation by only analyzing the structure of ties of those who are direct biological or adoptive descendants of the original households; however, offspring of these unions will be missing their grandparents, aunts, uncles, cousins, and so on from the side of the family of the person who ‘married in’ to the lineage in our dataset. However, it is likely that this limitation does not bias comparisons by race and education since the lineages followed are population representative and should therefore not be substantially different from the lineages of those who marry in. A more troubling potential limitation of the analysis is that race and socioeconomic groups may differ in the likelihood that parents of children never co-reside with them despite being part of their lives and a valid kin tie. Future research should take pains to capture relatives who are not connected to the focal individuals through chains of household co-residence and determine whether their omission biases these comparisons.

Despite these limitations, we believe that the present analysis substantially expands demographic understanding of kinship structures. In particular, our finding that educational differences, not racial ones, are the key determinant of kinship structure is an important contribution to demography and family studies. Future research should seek to determine how kinship structures and different forms of kin relations overlap and influence family life.

Figure 1: Distribution of Relative Counts

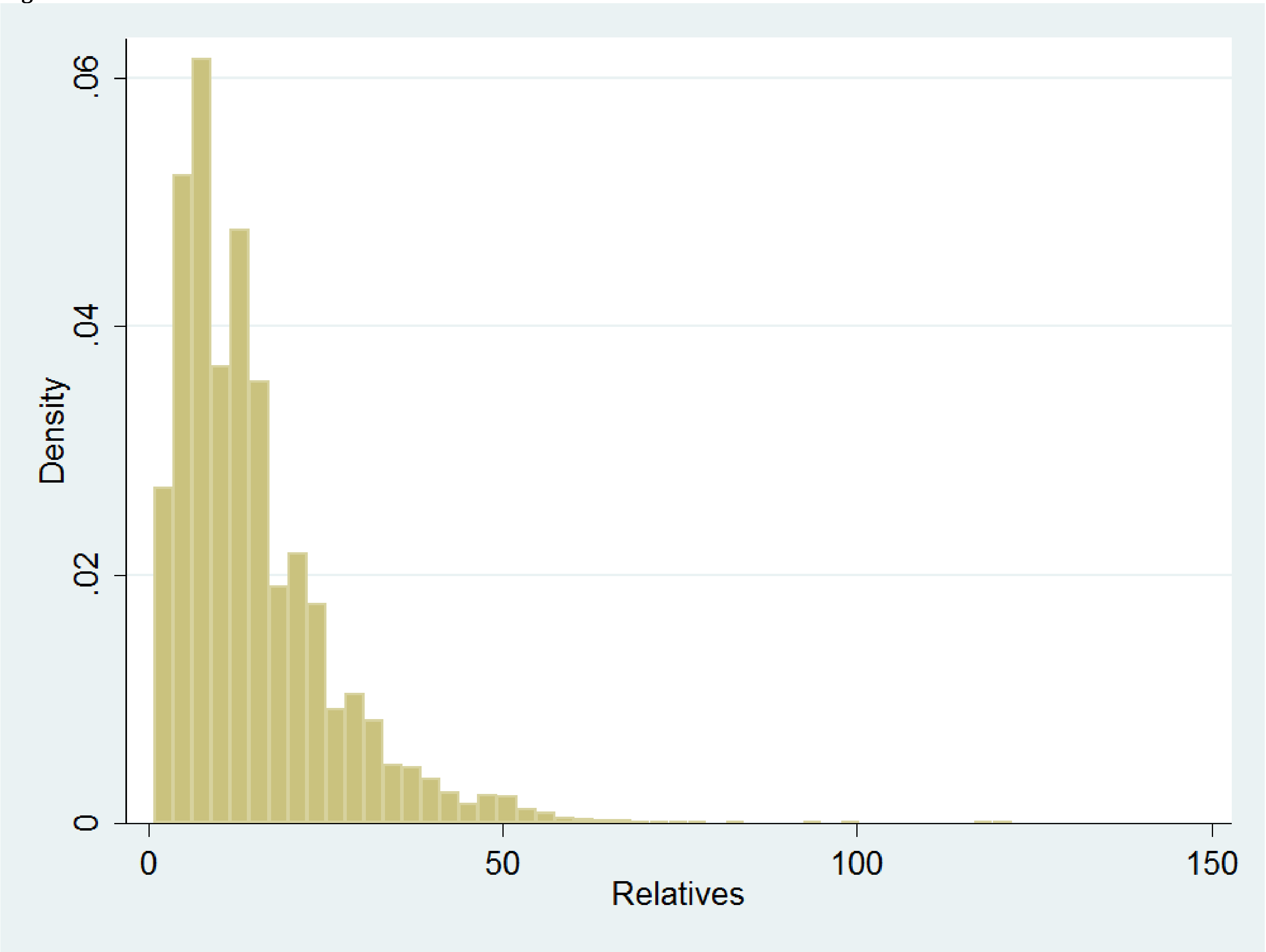

Figure 2: Component Size Distribution

Figure 3A: Predicted Full Sibling Counts, by Race and Age

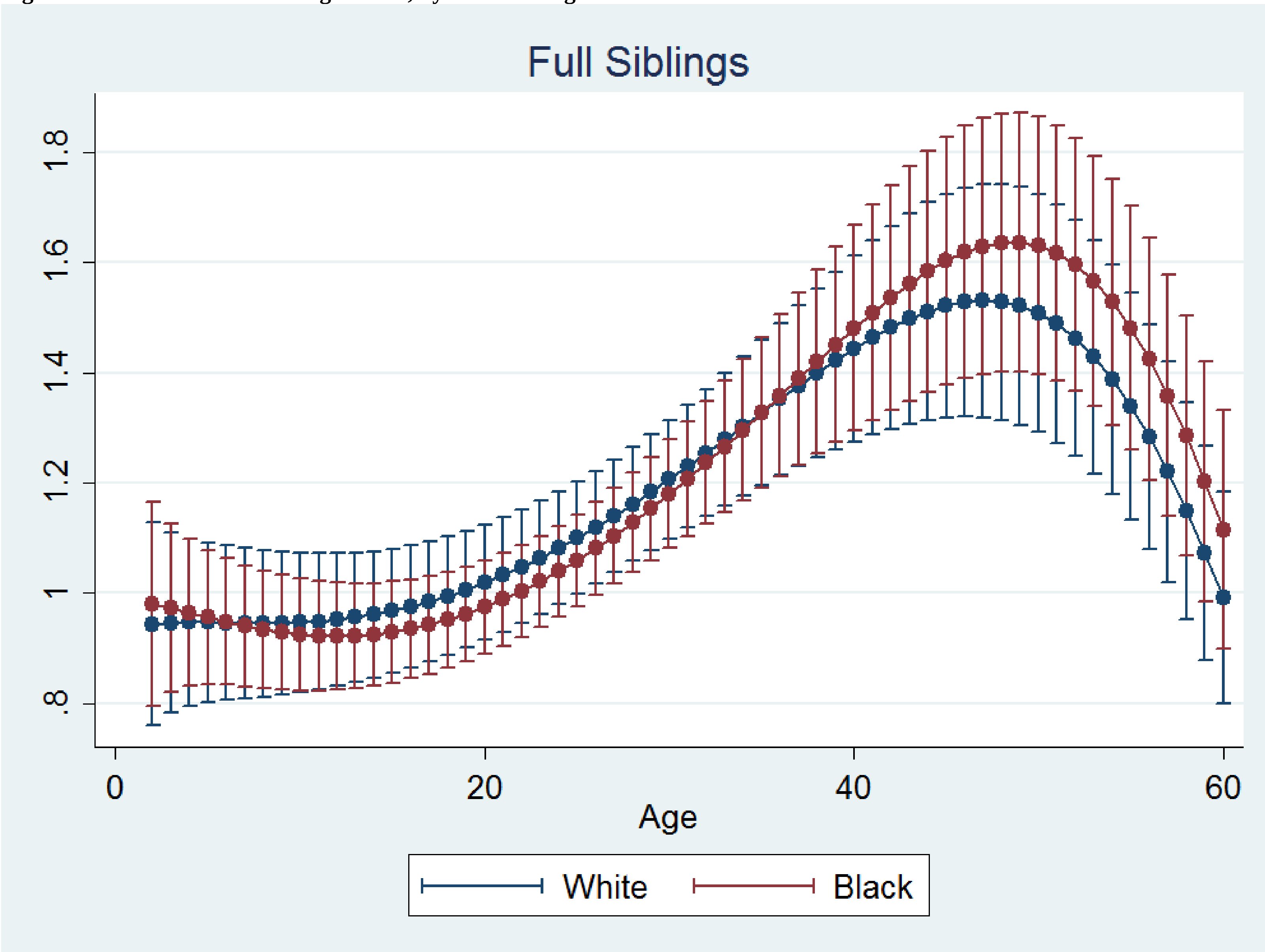

Figure 3B: Predicted Full Sibling Counts, by Education and Age

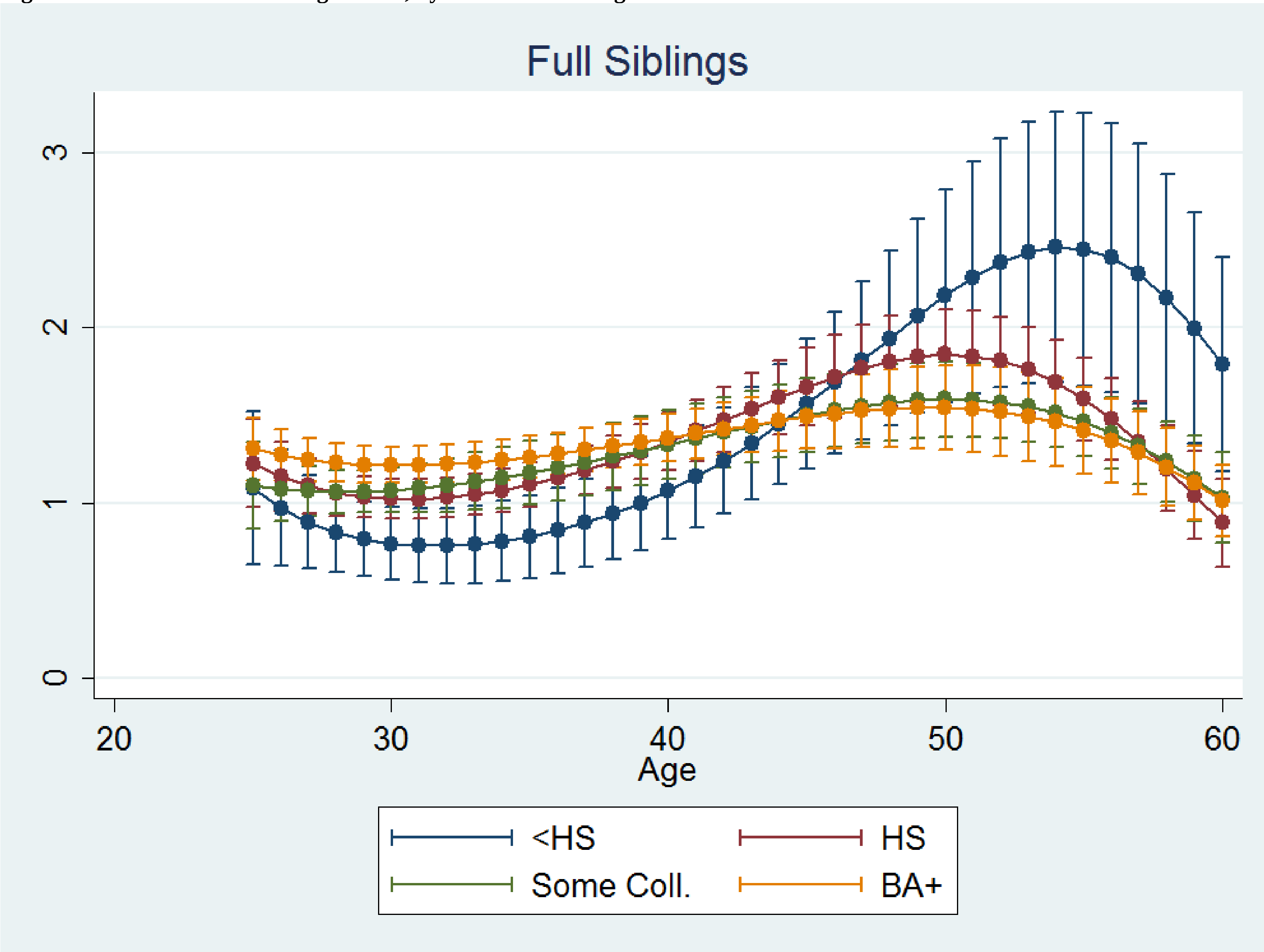

Figure 4A: Predicted Half Sibling Counts, by Race and Age

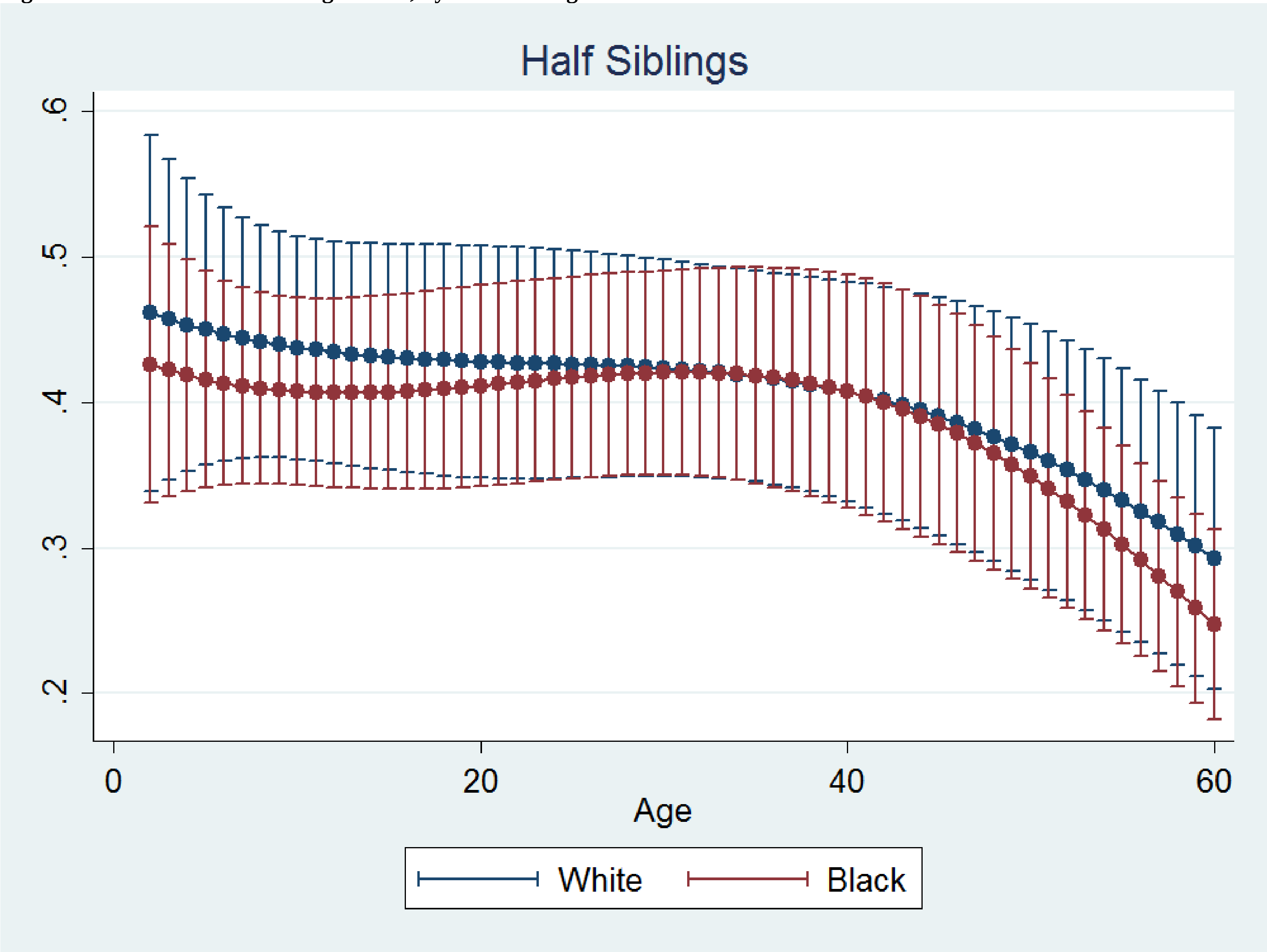

Figure 4B: Predicted Half-Sibling Counts, by Education and Age

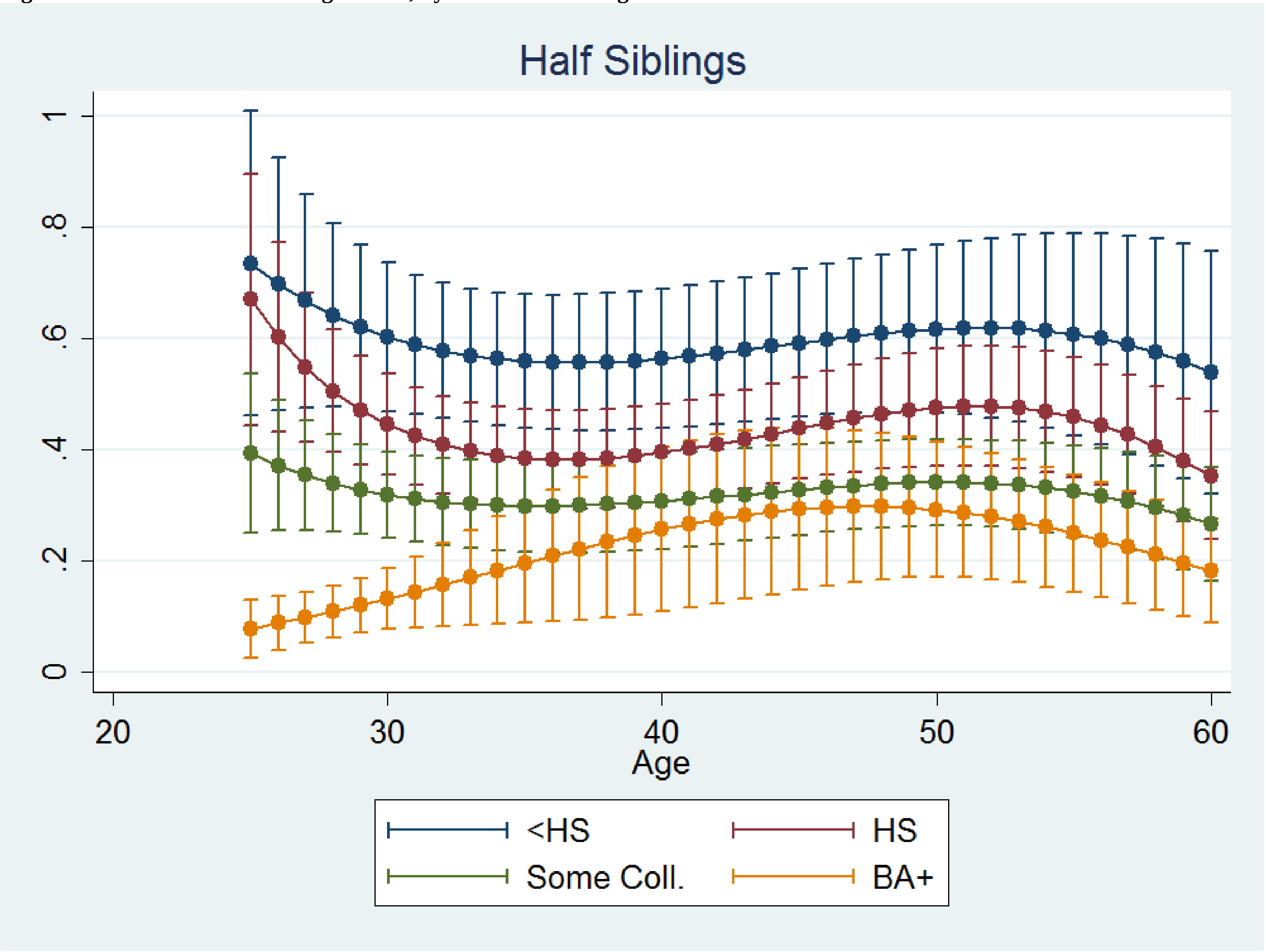

Figure 5A: Predicted Parent Counts by Race and Age

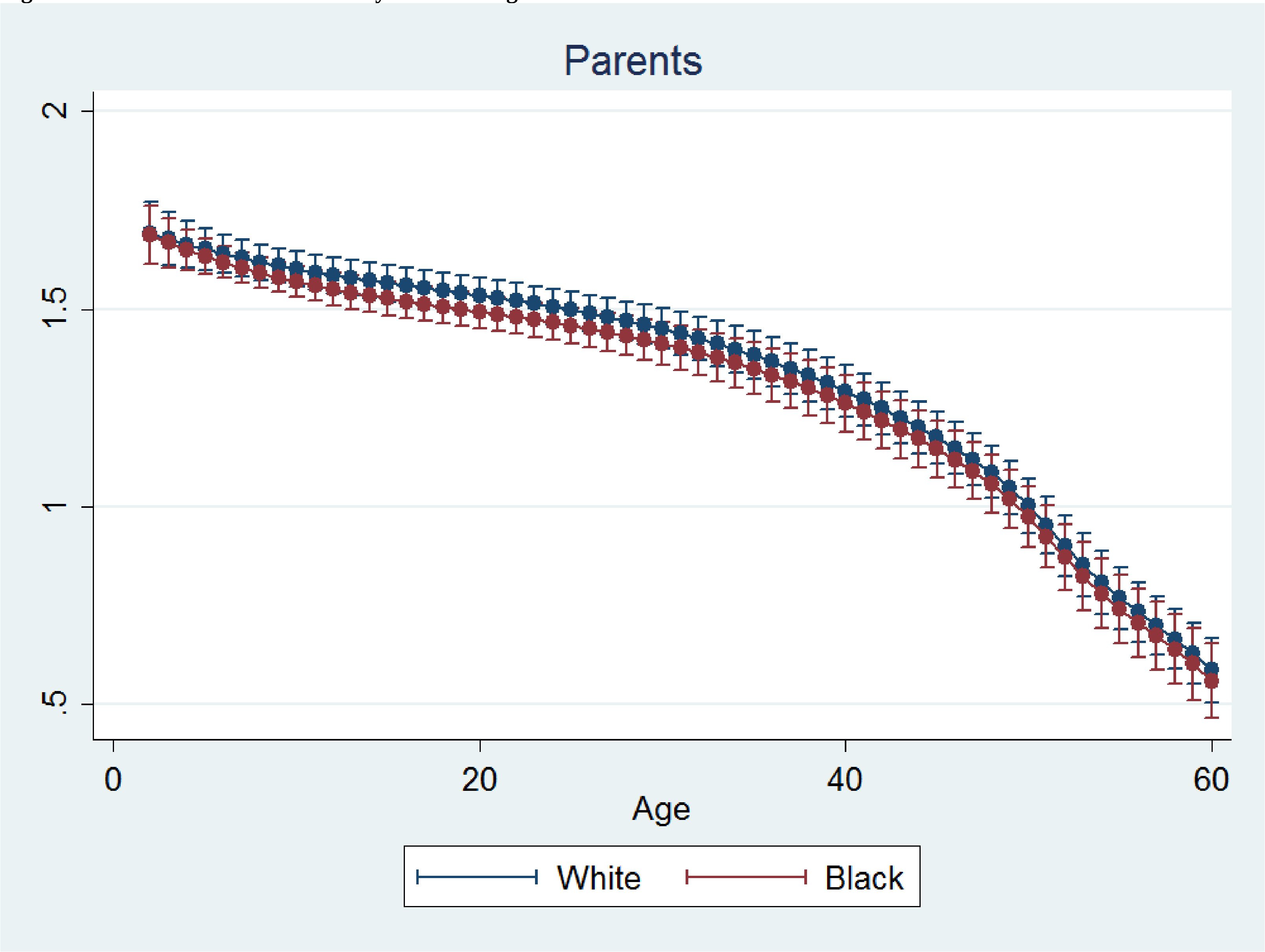

Figure 5B: Predicted Parent Counts, by Education and Age

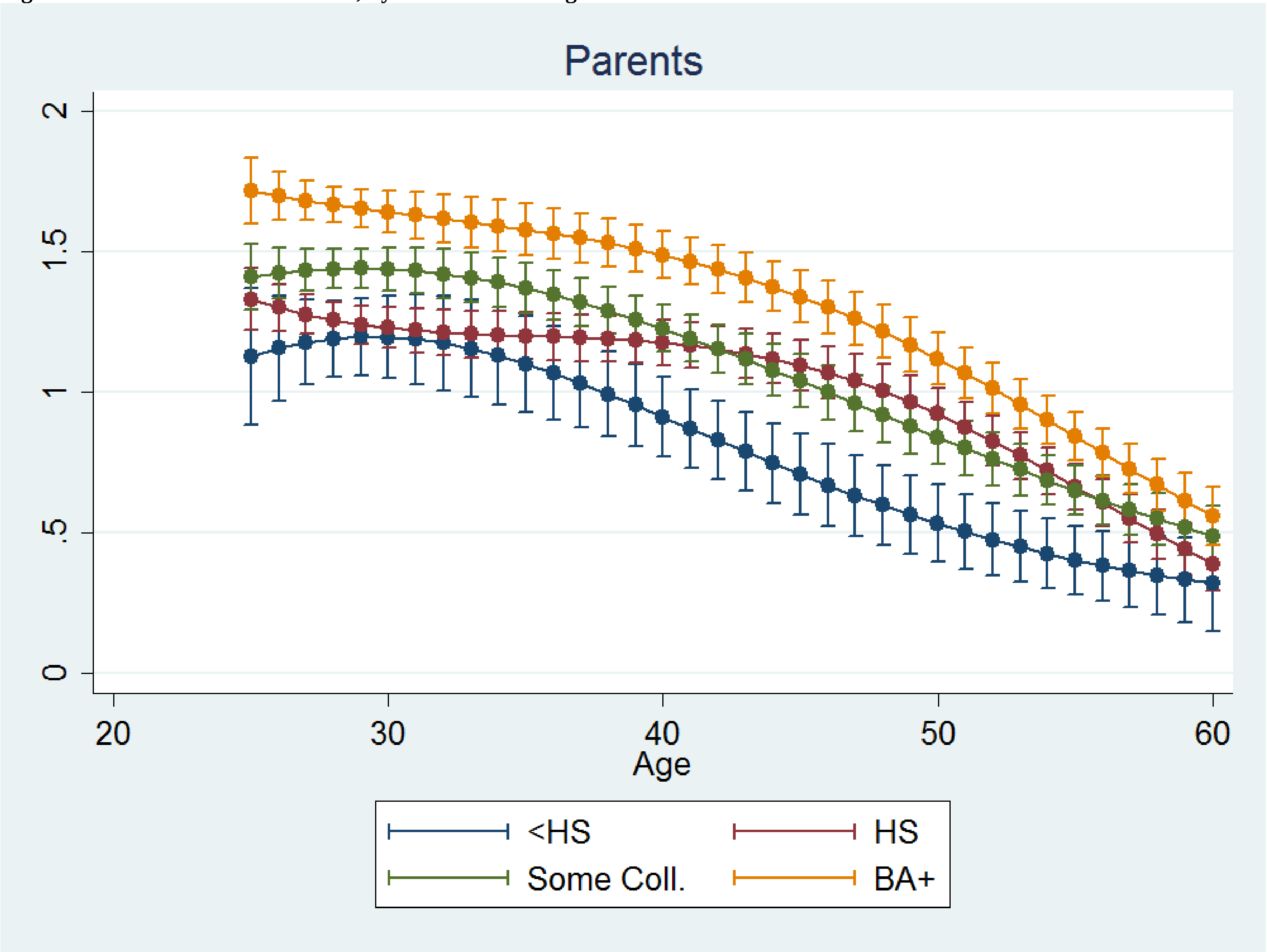

Figure 6A: Predicted Child Counts, by Race and Age

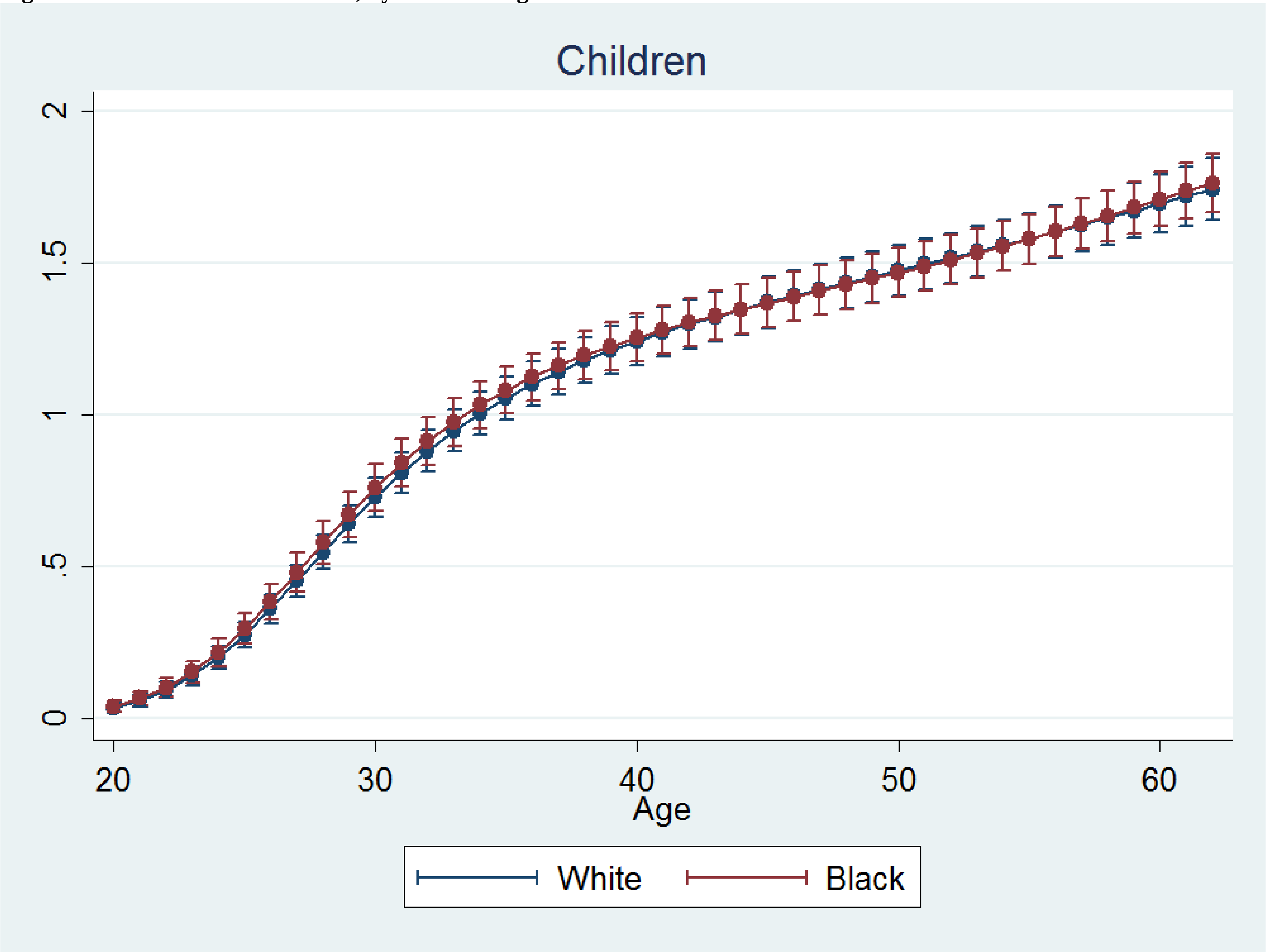

Figure 6B: Predicted Child Counts, by Education and Age

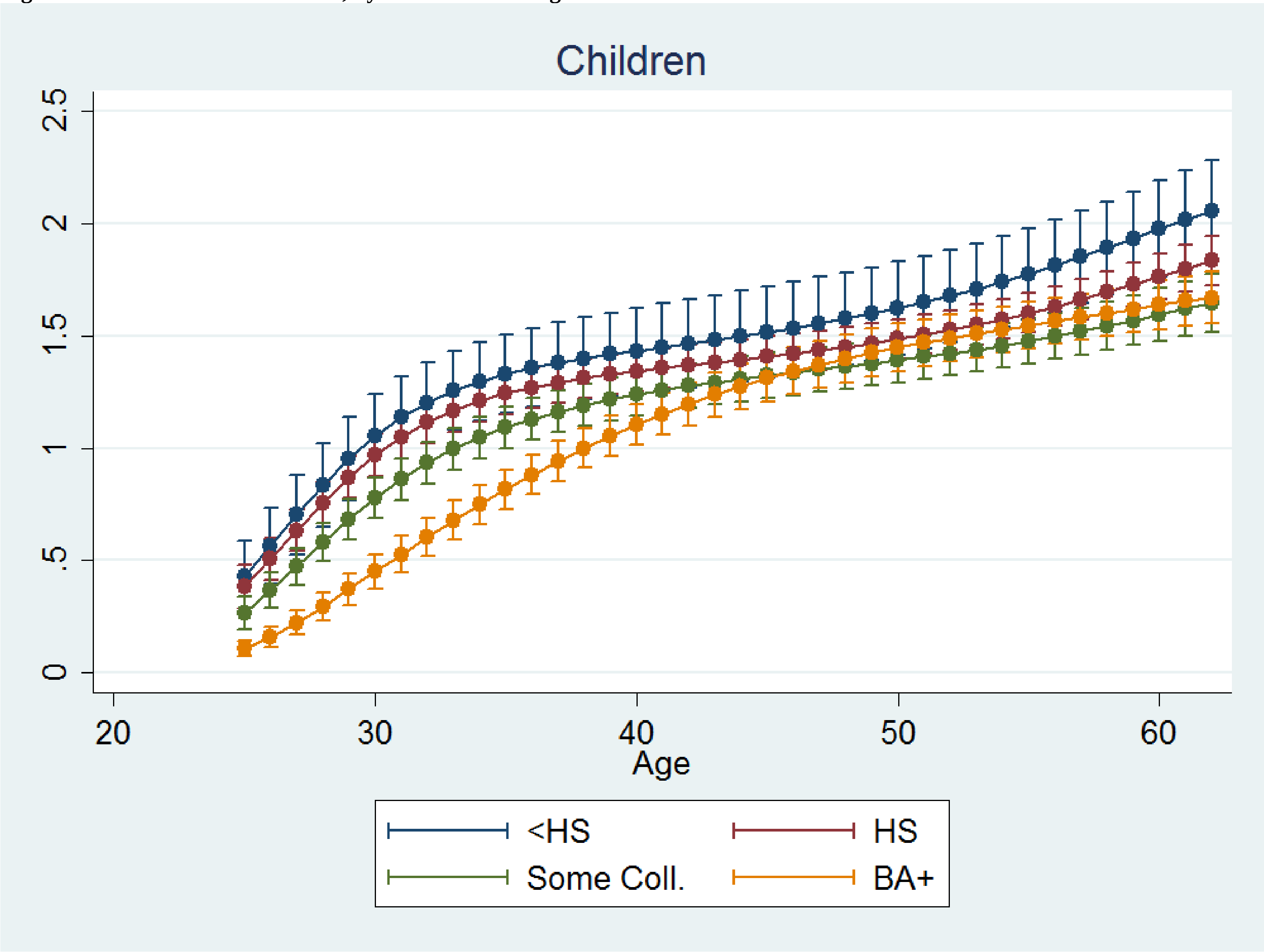

Figure 7A: Predicted Grandparent Counts, by Race and Age

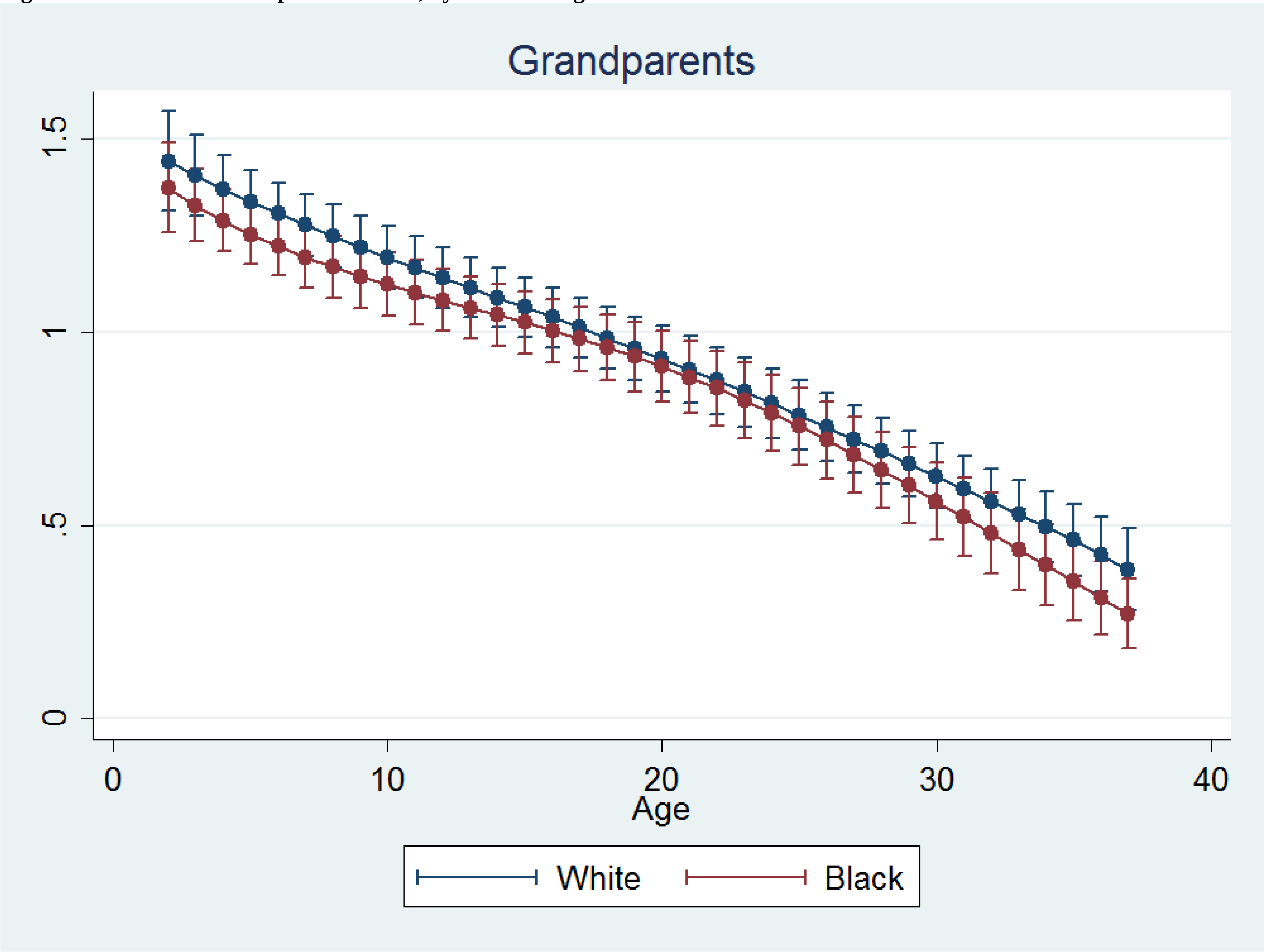

Figure 7B: Predicted Grandparent Counts, by Education and Age

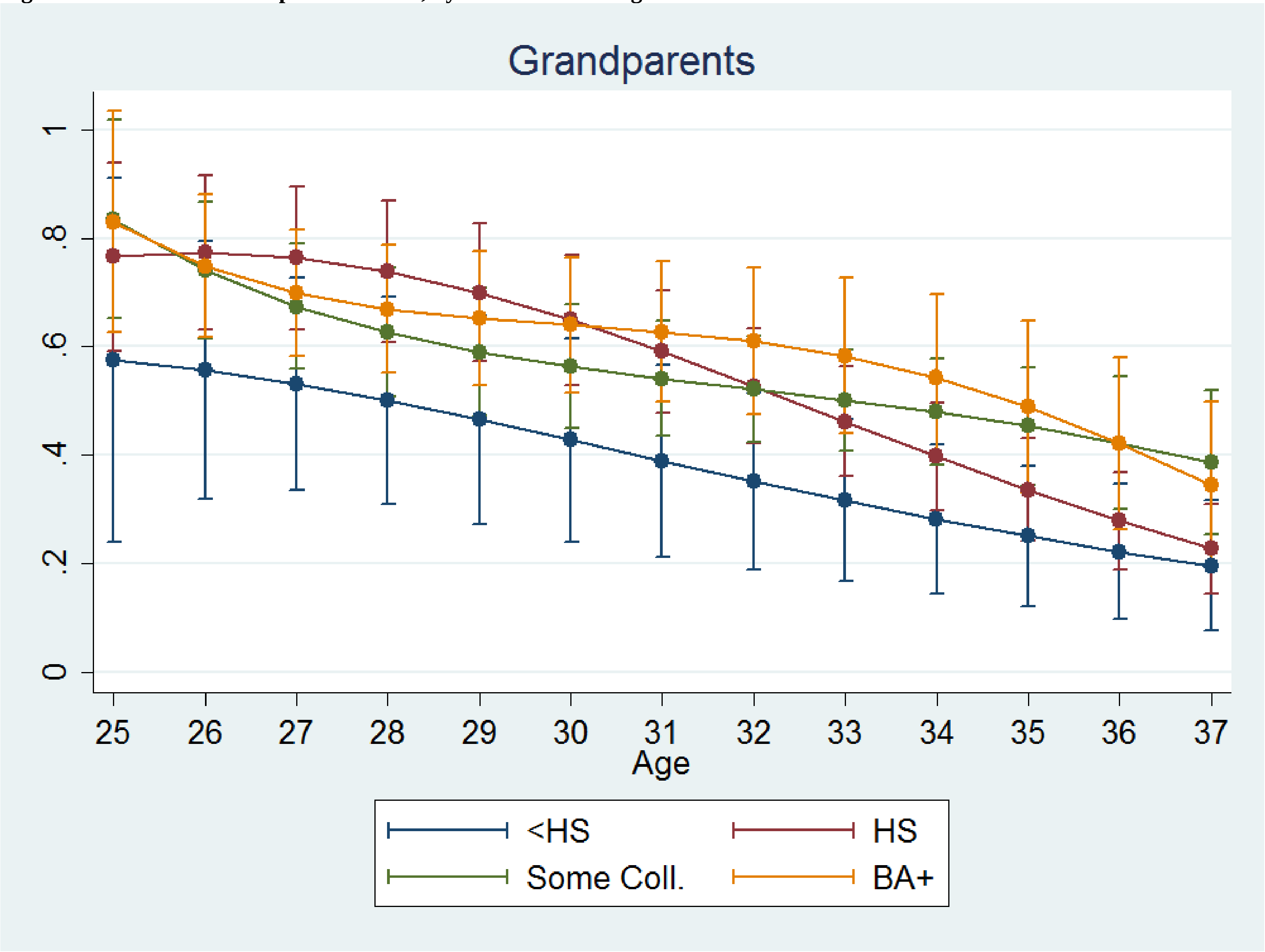

Figure 8A: Predicted Grandchild Counts, by Race and Age

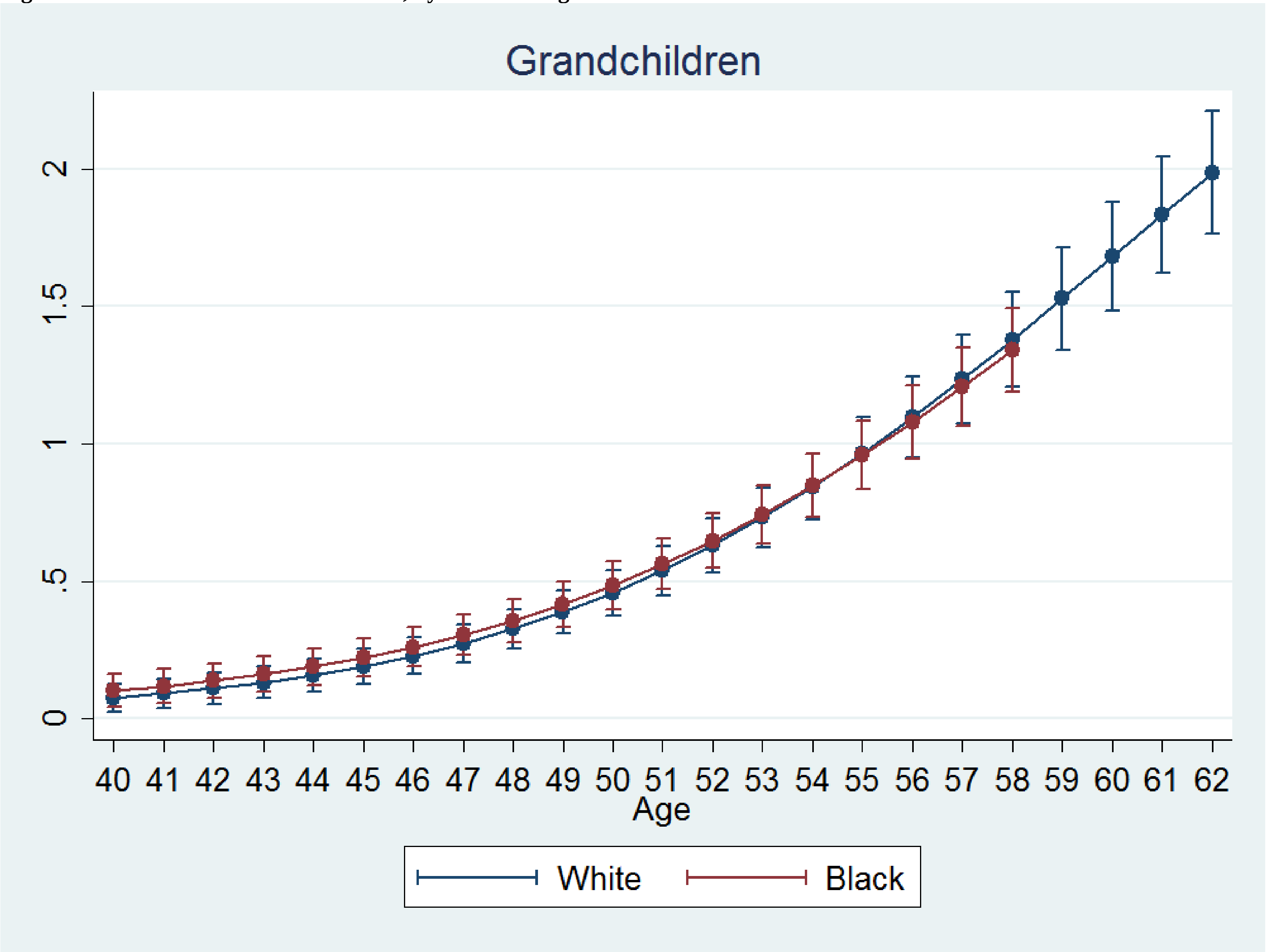

Figure 8B: Predicted Grandchild Counts, by Education and Age

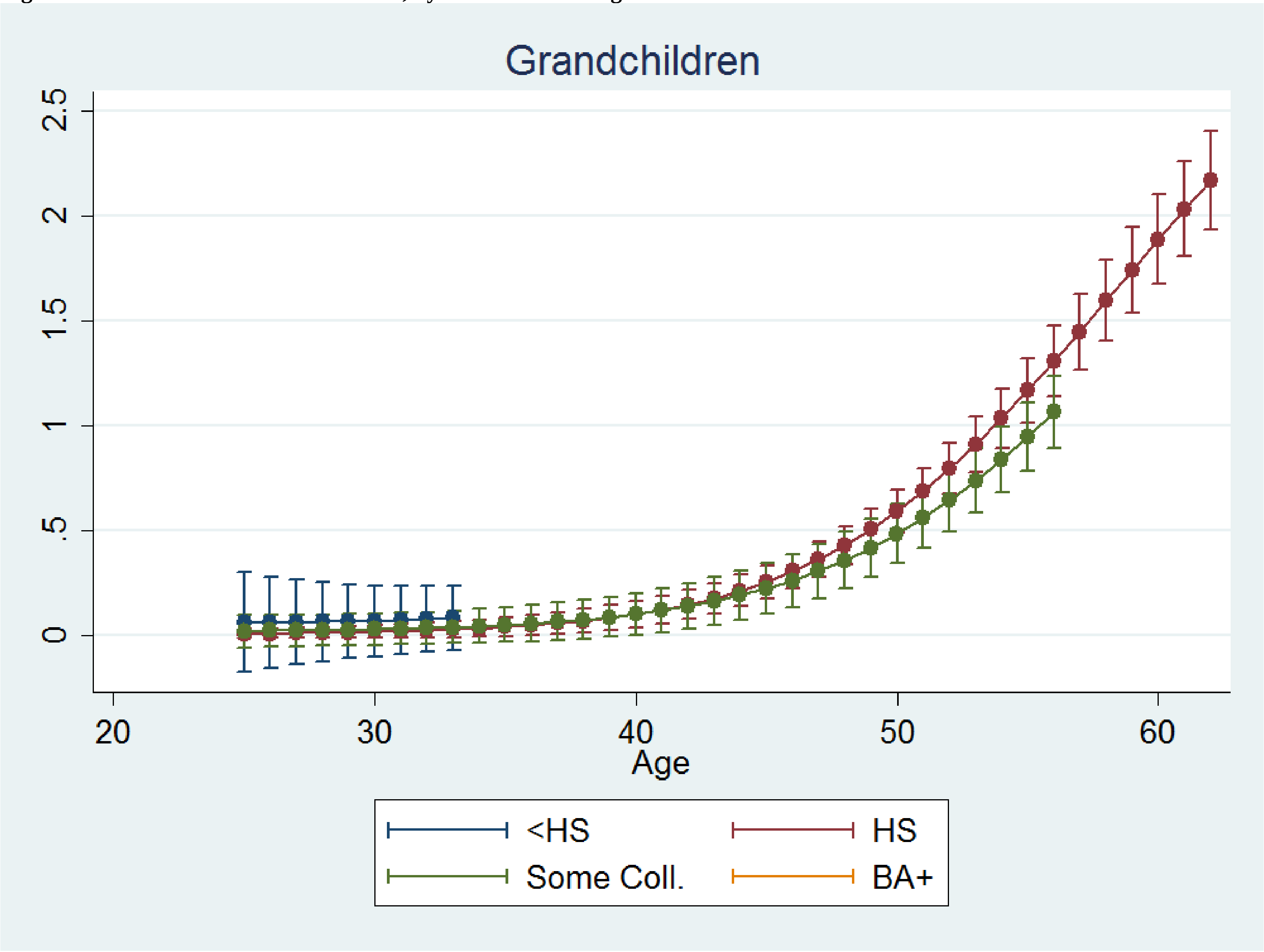

Figure 9A: Predicted Aunt/Uncle Counts, by Race and Age

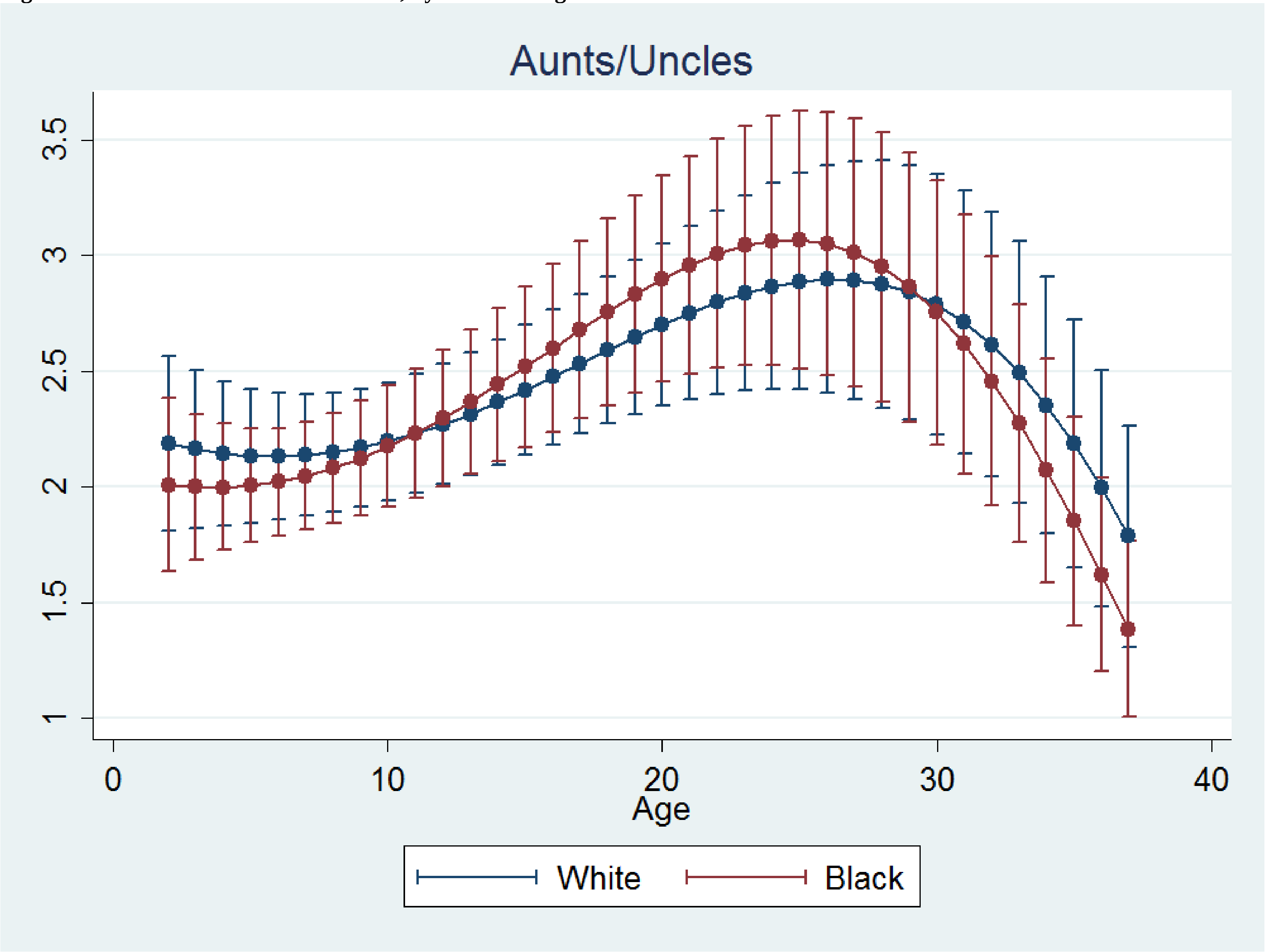

Figure 9B: Predicted Aunt/Uncle Counts, by Education and Age

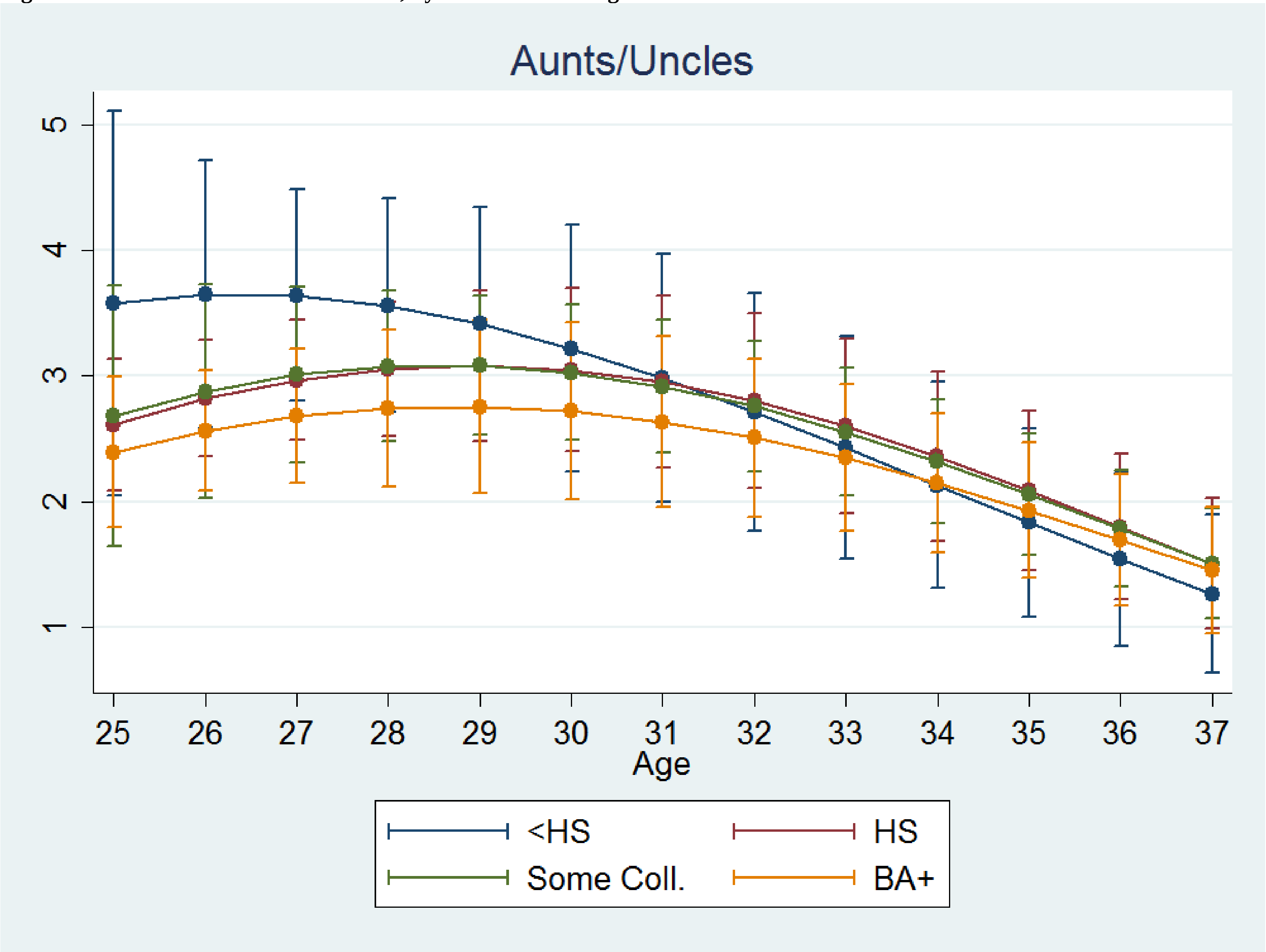

Figure 10A: Predicted Niece/Nephew Counts, by Race and Age

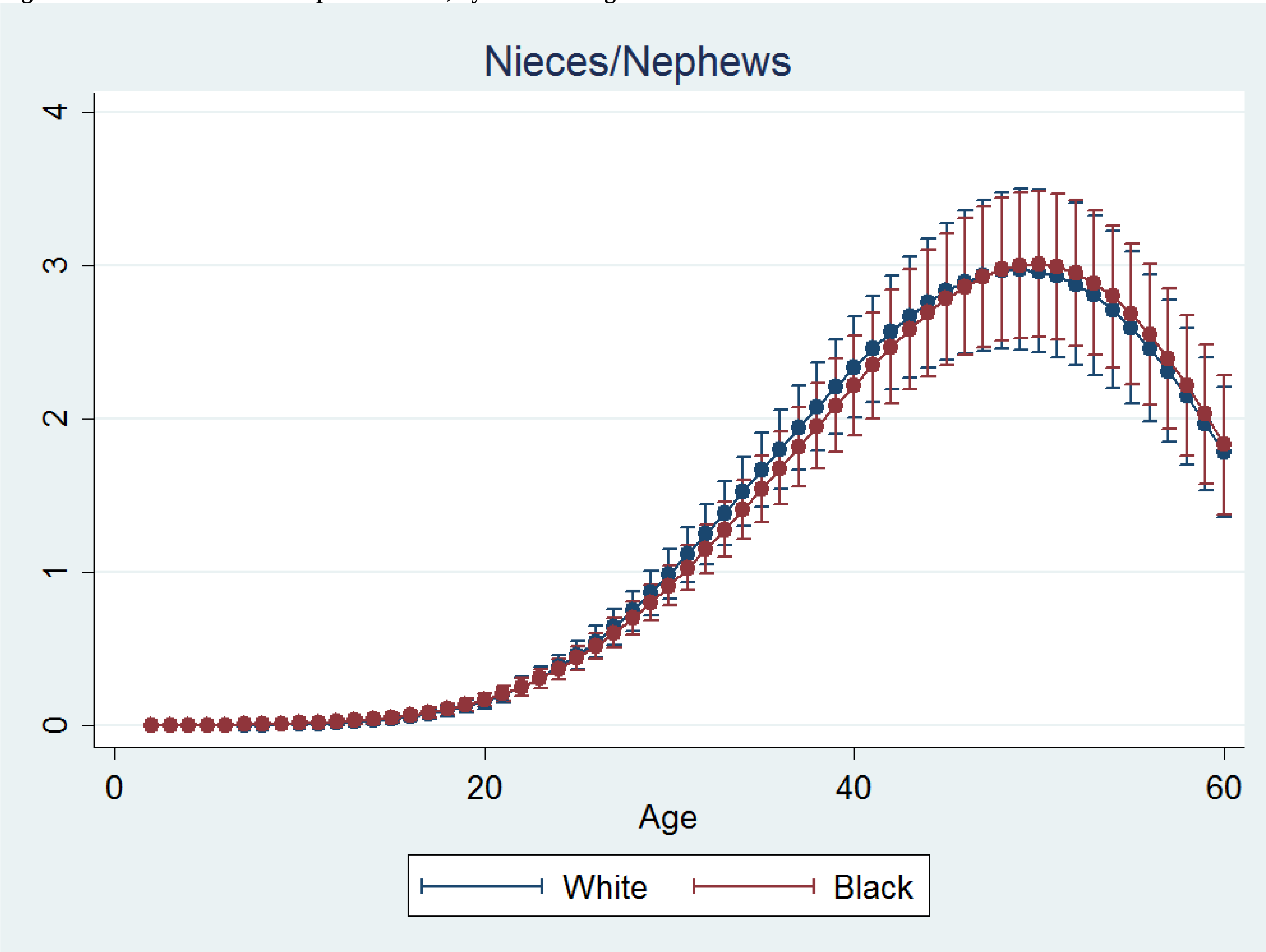

Figure 10B: Predicted Niece/Nephew Counts, by Education and Age

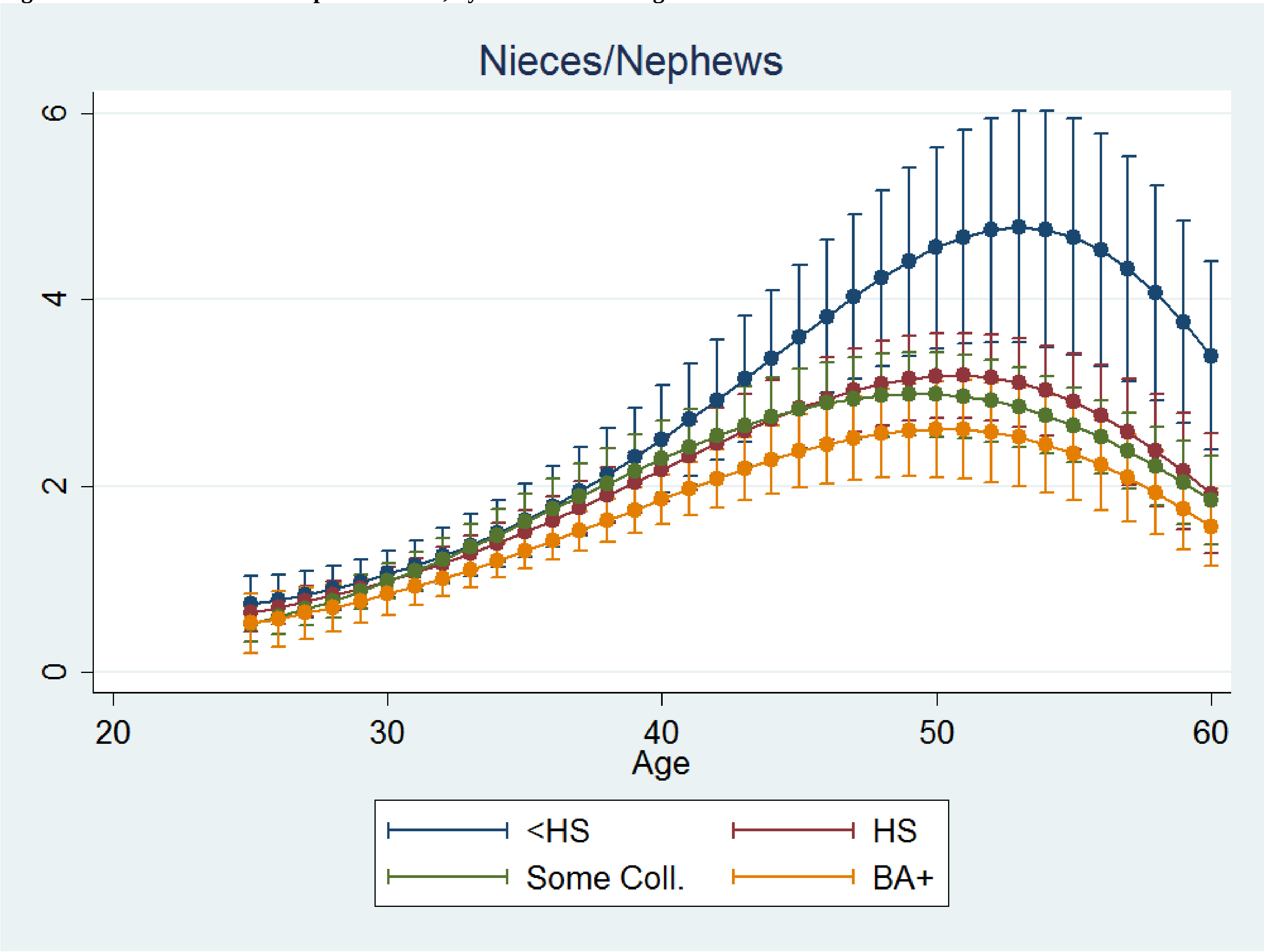

Figure 11A: Predicted Cousin Counts, by Race and Age

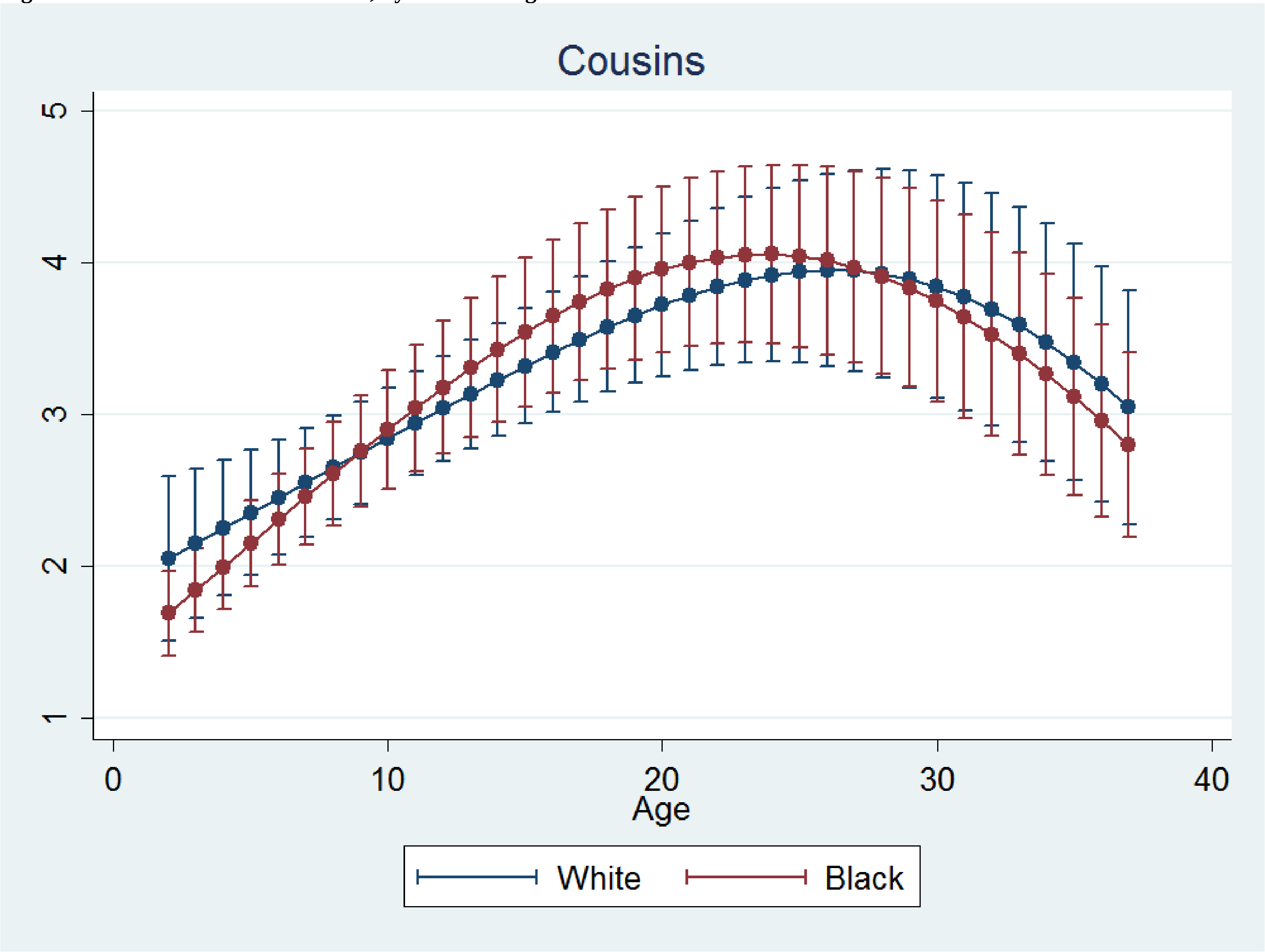

Figure 11B: Predicted Cousin Counts, by Education and Age

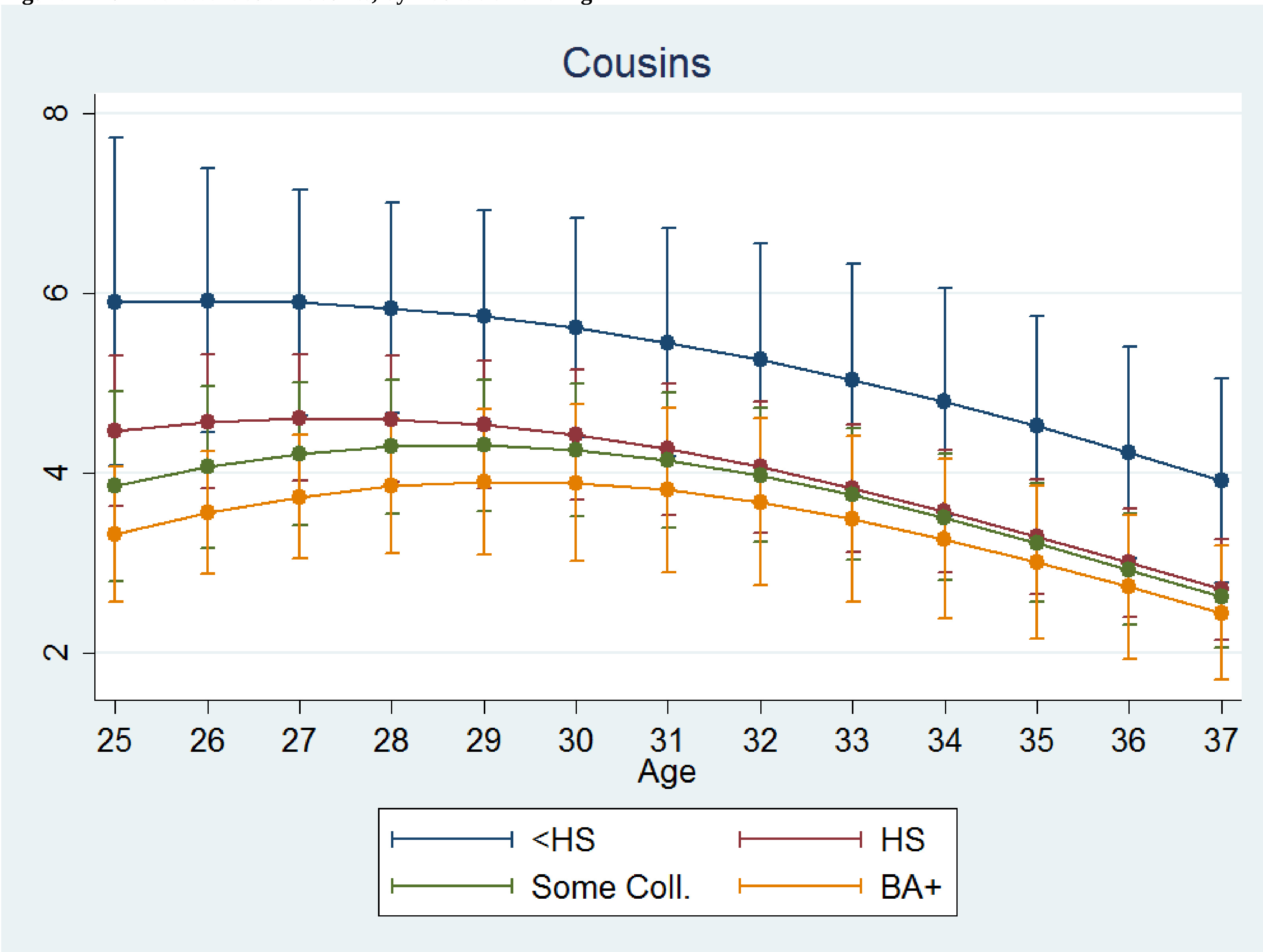

Figure 12A: Predicted Spouse Counts, by Race and Age

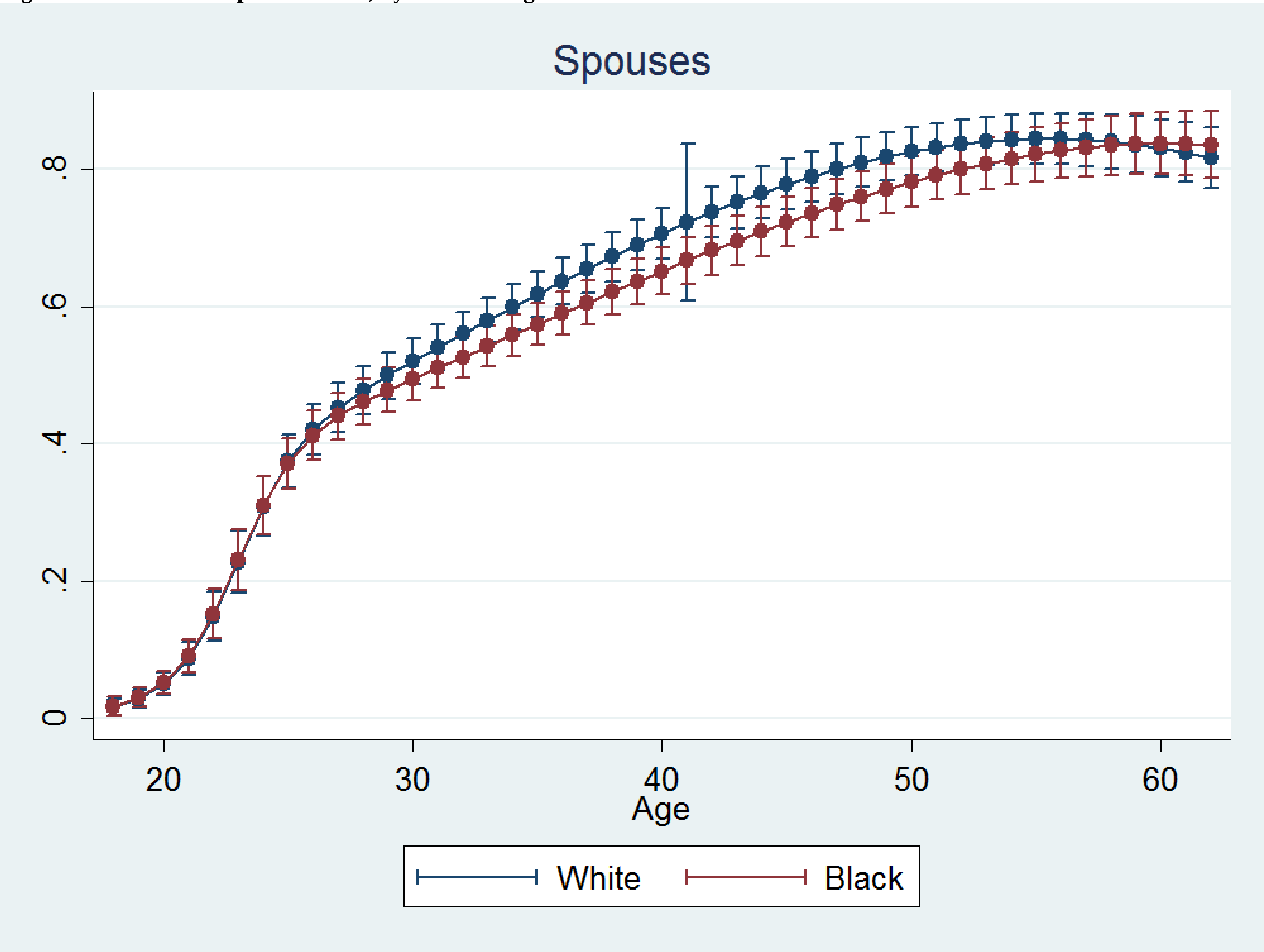

Figure 12B: Predicted Spouse Counts, by Education and Age

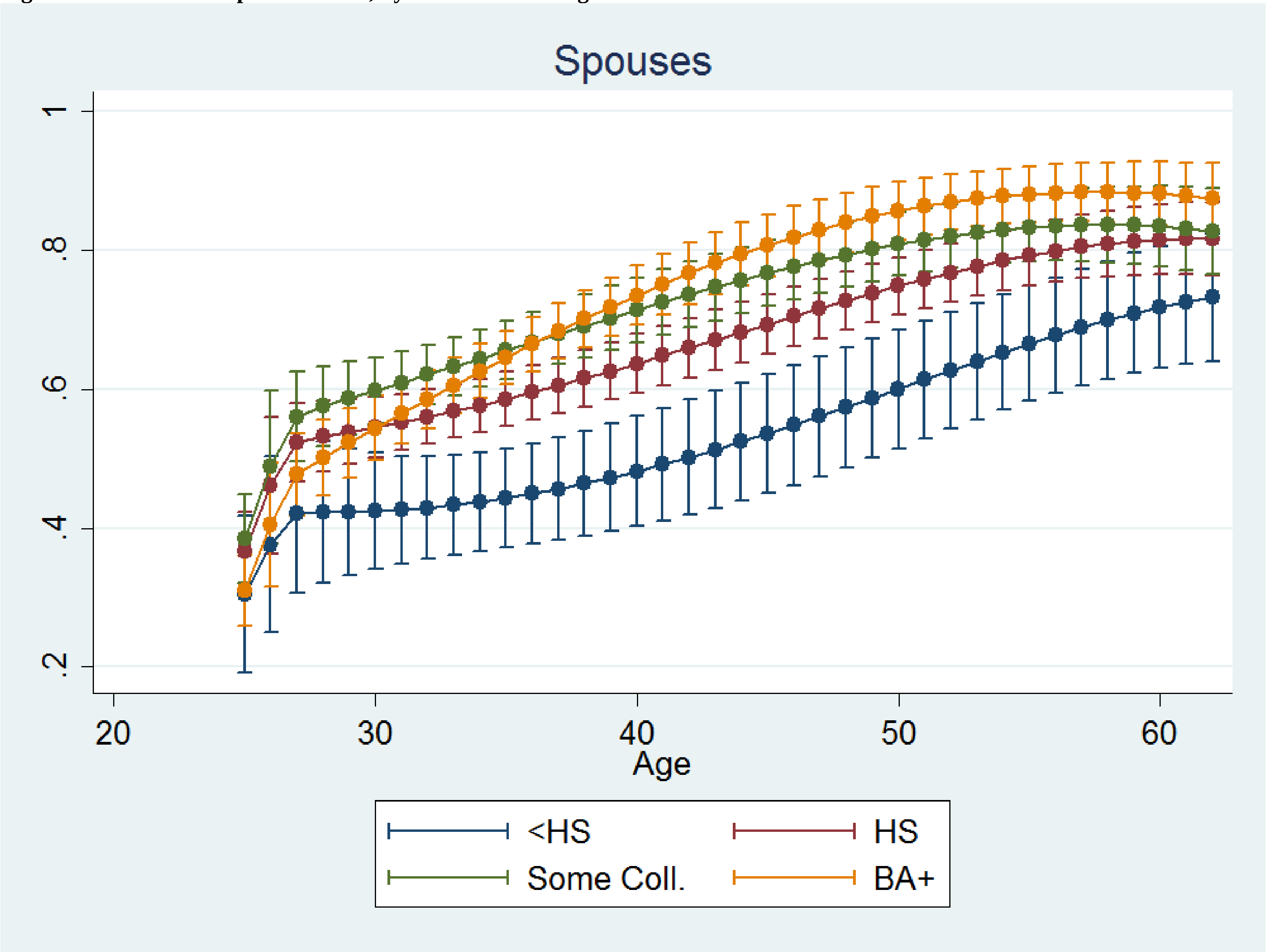

Table 1: Select Kin Mortality and Intervening Tie Effects on Educational Differences in Kin Counts

	(A) Full Siblings		(B) Grandchild		(C) Cousin		(D) Spouse	
	Baseline	+Deceased	Baseline	+Intervening	Baseline	+Deceased	Baseline	+Deceased
<HS (Ref.)								
HS	-3.624	0.273	-9.272	0.000	-3.819	-2.338*	-0.894	0.121
SC	-8.965+	-1.106	-3.761	-0.001	-7.331	0.328	-1.627	-1.847+
BA+	-8.700+	1.486	-32.013*	-0.002	-8.395	-3.684	-3.338*	-2.106*
HS*Age	0.273	-0.027	0.428	-0.002	0.298	0.223+	0.075	0.023
SC*Age	0.694+	0.113	0.158	-0.013	0.567	-0.037	0.130	0.151*
BA+*Age	0.700*	-0.065	1.256+	-0.052	0.626	0.296	0.228*	0.162**
HS*Age^2	-0.006	0.001	-0.007	0.000	-0.008	-0.007+	-0.002	-0.001
SC*Age^2	-0.016+	-0.003	-0.002	0.000	-0.014	0.001	-0.003	-0.003*
BA+*Age^2	-0.017*	0.001	-0.017+	0.001	-0.015	-0.008	-0.004*	-0.003**
HS*Age^3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SC*Age^3	0.000+	0.000	0.000	0.000	0.000	0.000	0.000	0.000*
BA+*Age^3	0.000*	0.000	0.000+	0.000	0.000	0.000	0.000	0.000*

NOTE: This table depicts the educational coefficients from zero-inflated Poisson regression models in two nested models. The ‘Baseline’ models are the models in which the indicated counts are modeled as an interactive function of education and age, cubically specified. The ‘+Deceased’ models add a control for the counts of known ties of that type who died since the beginning of the study. The ‘+Intervening’ model for grandchildren controls for the number of children the person has, which is required to have any grandchildren. See text for details.