A national sample of US paternity tests: do demographics predict test outcomes?

David Bishai, Nan Astone, Laura Argys, Robert Gutendorf, and Chris Filidoro

BACKGROUND: Nearly 300,000 paternity tests are performed in the United States annually to include or exclude a male as the biological father of a child. Little is known about how well the test outcomes could be predicted simply on the basis of the subjects' social background. Our objective is to document the demographic composition of test subjects and to determine how well ethnic background and age predict paternity inclusion.

STUDY DESIGN AND METHODS: A national database of 9999 paternity test results was formed weighted to represent the population of paternity establishment cases for the United States. Multiple regression determined the odds ratio of a paternity inclusion based on demographic variables such as race and age.

RESULTS: The test results indicated paternity inclusions in 72 percent of cases, overall. The percent included varied little across racial and/or ethnic categories with a low of 67 percent in African Americans and a high of 82 percent among Native Americans, but these differences were not significant in multivariate analysis. The likelihood of inclusion showed a modest but significant correlation with increased maternal age. Mothers of European origin who nominated non-Europeans as fathers were less likely to identify the correct man as father.

CONCLUSION: For any subject receiving a genetic test in a child support office, predicting a 72 percent probability that the test will show paternity inclusion offers as good an estimate of the test outcome as a subgroup estimate based on a subject's age, race, and child characteristics.

Roughly 280,000 DNA based paternity tests are performed in the United States annually.[1] Most of these tests are performed at the request of the mother or state child support agency (IV-D office) to facilitate the establishment of a child support order. Over the past two decades, state IV-D offices have begun to require mothers to assist in paternity establishment as a condition of public assistance, which has increased the use of genetic testing to determine paternity. Not all unmarried mothers seek government assistance in obtaining child support, however, and only a fraction of women filing for child support obtain genetic testing. In 2000, there were 7,571,000 female-headed households with children under age 18.[2] On an annual basis, state offices of child support enforcement established 1,555,581 paternities in 2000. It was able to establish nearly half of these (688,510) through voluntary in hospital acknowledgment programs.[3]

Very little is known about the population seeking genetic verification of paternity or whether characteristics of the mother, child, or alleged father are correlated with the test results. What we know about children for whom paternity needs to be established comes primarily from the demography of unmarried mothers.[4] There have been no published studies correlating paternity test outcomes with the background of the clients. There are two reasons why it is worthwhile to look for such correlations. First,

From the Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; the University of Colorado at Denver, Denver, Colorado; Long Beach Genetics, Long Beach, California, and Orchid Biosciences-GeneScreen, Princeton, New Jersey.

Address reprint requests to: David Bishai, Department of Population and Family Health Sciences, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Suite E 4005, Baltimore, MD 21205-2103; e-mail: dbishai@jhu.edu.

The authors acknowledge the support of the National Institute for Child Health and Development (1 RO3 HD042545-01).

Received for publication June 20, 2005; revision received September 26, 2005, and accepted September 26, 2005.

doi: 10.1111/j.1537-2995.2006.00806.x

TRANSFUSION 2006;46:849-853.

policy makers can benefit from having a benchmark for the probability of paternity in population of child support enforcement agency clients. Suppose a policy maker wanted to estimate potential child support collections that might be obtained from a population of 10,000 individuals who had not yet been tested. A natural question would be whether the estimate of the inclusion rate would be greatly improved if administrative data (from birth certificates) on the race, mother's age, and child's sex were used as the basis for the estimate. A second reason to assess the predictability of paternity tests is to search for potential demographic factors that might confound the relationship between a paternity inclusion and later child outcomes or paternal behaviors. Suppose there were to be a claim that in a population of children whose mothers sought child support that, controlling for household income, children of fathers who were included in a paternity test had different school performance from children where the man was not included. If paternity inclusion rates were significantly lower by mother's age or by race, one could not judge whether the paternity inclusion or these background factors were the principal mediators of the relationship. There is a nascent literature aspiring to measure causal effects of paternity inclusion on behaviors (other than the legally mandated payment of child support) that will benefit from assessing demographic background of genetic test cases by whether paternity is established.[4-6]

It has been suggested that the legal determination of paternity may alter interactions between a father and his child and that the impact of paternity may vary by the method by which paternity is established. For instance, children for whom paternity has been established by any means are much more likely to receive child support and to have contact with their nonresident fathers than their counterparts for whom paternity is uncertain.[8] There is also some evidence that genetic testing, and other nonvoluntary methods of paternity establishment, result in a much higher incidence of child support awards, but a slightly lower incidence of father-child contact than children whose fathers' names appeared on the birth certificate or who otherwise voluntarily acknowledged paternity. These results suggest that, in the absence of paternity establishment, there is some uncertainty regarding the identity of the child's father. If we want to credit the test outcome with changing the behavior of men, we must first establish the degree to which both parties anticipate test results.

The objective of this article is to describe the racial and ethnic background as well as the mean age of mothers and alleged fathers who obtain DNA-based paternity tests in the United States. In addition, we identify any systematic patterns by race, age, and child sex in the frequency with which the test yields a paternity inclusion for an alleged father in both interracial and intraracial unions. Since the physical features of a child may make a child's paternity more obvious, a plausible hypothesis is that couples with dissimilar racial backgrounds would be more likely to yield tests that prove that a man is the father.

MATERIALS AND METHODS

Cases were obtained from a company that performs roughly half of all DNA-based paternity tests in the United States and has contracts with state child support enforcement offices in more than 33 states; it offers a promising foundation for a nationally representative sample of genetic tests from the population attending child support offices. From an initial sample of 152,737 mother-father-child trios (i.e., mother, child, and putative father) submitting to DNA tests in 2000, we excluded 5711 cases because the father's or child's age was missing or miscoded. From the remaining data set, we prepared a nationally representative random sample of 9999 trios with sampling weights proportional to paternity establishment caseloads for each of the following regions; Northwest, Northeast, North Central, Central Midwest, East Central, South Central, Mid-Atlantic, Southwest. (States that were aggregated into regions were as follows: Northwest—AK, *WA*; Northeast—CT, *RI*, *NY*, *ME*, MA, NH, VT; North Central—ID, *ND*, *OK*; Central Central—*NE*, KS; Eastern Central—*KY*, *OH*, *TN*; South Central—*AR*, LA; Mid-Atlantic: *MD*, VA, WV; and Southwest—*CO*, NV, NM, *TX*, UT, WY. States that had sufficient samples to represent themselves were as follows: *AL, AZ, CA, DE, FL, GA, HI, IA, IL, IN, MI, MN, MS, MO, MT, NJ, OR, PA, SD, VT,* and *WI*. Only the 33 italicized states had data available to contribute. Nonitalicized states were represented by setting the sample weights from the states with available data in that region equal to the population weight for the entire region.) The purpose of the regional aggregates was to obtain representation for the 17 states. In addition to the regional aggregates, weighted entrants to the sample were also drawn from California and 21 other states that were not aggregated because they had sufficient sample sizes. Based on the weighted sample, we tabulated mother and alleged father's race and age and child sex. We also tabulated inclusion rates for the overall sample and by age and race. The data on caseloads used to form the sample weights came from state caseload estimates for 2002, whereas the data for the genetic tests came from 2000. We are assuming negligible proportional shifts between states in paternity establishment caseloads.

Genetic tests were performed using routine polymerase chain reaction analysis with the Promega Powerplex system, (MiraiBio Inc., Hitachi Software Engineering America Ltd., Alameda, CA). Fourteen STR loci were routinely utilized during the time frame for the cases of this study. The power of exclusion for this battery of tests was greater than 0.99999 for all races analyzed. Inclusion of paternity required that the tested man have all the neces-

TABLE 1. Race and/or ethnicity and paternity inclusion rates by race and/or ethnicity

Race and/or ethnicity	Father's race	Mother's race	Percent with positive match by father's race	Percent with positive match by mother's race	p Value for null that match rates are same
European American	40.07	46.43	74.69	75.18	0.61
African American	37.50	32.70	67.39	67.98	0.61
Latino	14.66	13.44	73.33	74.92	0.35
Asian American/Middle Eastern/North African	1.01	0.96	80.2	79.07	0.84
Native American	0.66	0.78	81.82	85.71	0.54
Other unspecified	6.09	5.69	76.35	73.23	0.23
Number	9999	8930	9999	8930	

sary obligatory alleles for each system examined to contribute to the child. Minimum probabilities of paternity required for any inclusionary report exceeded 99 percent and a paternity index of 100. The study was approved by the Committee on Human Research of the Johns Hopkins Bloomberg School of Public Health.

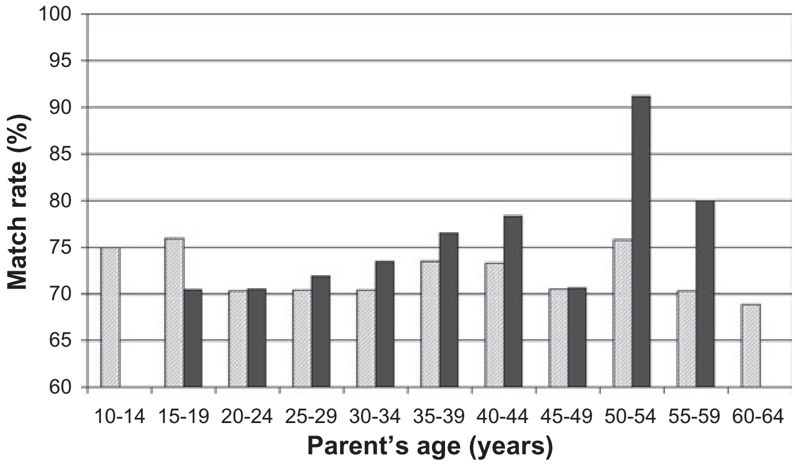

Fig. 1. Paternity inclusion rates by parent's age. (▨) Percent with positive match by father's age; (■) percent with positive by mother's age.

RESULTS

Mothers did not submit genetic samples in 10.15 percent of cases, so in our proportionally weighted sample of 9999 cases there are only 8930 cases for which mothers' data are available. For this sample, the overall paternity inclusion rate was 72 percent. Male children composed 51 percent of the sample. There was no significant difference in inclusion rates by the sex of the child.

Table 1 shows the racial composition of our weighted sample and the proportions yielding a paternity inclusion by the racial group of the parents. Inclusion rates range from a low of 68 percent for African American mothers to a high of 86 percent for mothers of Native American origin. The last two columns show that paternity inclusion rates are similar whether race and ethnic groups are defined by mother's self-definition or father's self-definition. Figure 1 shows the proportion of cases and the proportion yielding a paternity inclusion by mother's age and by father's age. Inclusion is more likely for women at older ages, but there is no clear increase in the inclusion rate as men get older.

The children in our sample, for whom paternity is contested, range in age from infants to 19 years, although the majority (51%) are below the age of 4. There was no systematic relationship between a child's age and the of proportion of subjects showing a paternity inclusion.

Table 2 shows the caseloads by the various combinations of parental ethnicity. Parents of matched ethnicity appear along the diagonal and account for 6900 of the 8965 for whom complete ethnicity data were available. There were more women from European backgrounds (4047) than men from European backgrounds (3541) in the sample.

Table 3 shows the proportion with paternity inclusion by various racial and ethnic combinations of the couples and can be used to assess the hypothesis that couples of matched ethnic backgrounds would have more uncertainty of paternity because the father's features and mother's features are so similar. If paternity were more uncertain in ethnically matched couples, then father inclusion rates along the diagonal of Table 3 would be lower then elsewhere. The mean paternity inclusion rate for ethnically matched couples and for unmatched couples is 74, suggesting no difference in paternity inclusion by ethnic matching.

The results of logistic regression models shown in Table 4 indicate that the child's sex has no correlation with the likelihood of a paternity inclusion in the DNA test set. Maternal age increases the likelihood of paternity inclusion, and African American ethnic group decreases the likelihood of inclusion. Overall Table 4 demonstrates that very few characteristics of the mother, child, or alleged father explain the likelihood of a paternity inclusion. One

TABLE 2. Case volume by ethnicity of parents

	Mother's ethnicity						
Father's ethnicity	European	African American	Latino	Asian/ Mideast	Native	Other	Total
European	3114	74	136	8	17	192	3541
African American	428	2725	96	3	5	152	3409
Latino	259	68	873	3	17	94	1314
Asian/Mideast	14	1	8	64	24	10	121
Native	28	1	7	0	0	5	41
Other	204	98	100	8	5	124	539
Total	4047	2967	1220	86	68	577	8965*

* Sums to 8965 because mothers did not submit samples in 998 samples and did not report ethnicity in an additional 36 samples.

TABLE 3. Paternity inclusion rates (percent of cases where fatherhood is proven)

	Mother's ethnicity					
Father's ethnicity	European	African American	Latino	Asian/Mideast	Native	Other
European	75.63	71.62	71.32	*	*	76.04
African American	70.33	69.98	78.13	*	*	73.03
Latino	72.2	67.65	74.91	*	*	74.47
Asian Mideast	*	*	*	81.25	*	*
Native	*	*	*	*	*	*
Other	78.92	67.35	85	*	*	66.94

* Fewer than 50 couples.

TABLE 4. ORs predicting paternity inclusion*

Mother's age	1.018	1.018	1.018
	(3.36)†	(3.30)†	(3.35)†
Father's age	0.995	0.995	0.995
	(–1.490)	(–1.510)	(–1.500)
Child is a boy	1.017	1.016	1.013
	(–0.350)	(–0.310)	(–0.260)
Father is African American	0.903		0.78
	(–1.040)		(3.95)†
Father is Latino	0.871		0.926
	(–1.700)		(–1.120)
Father is Asian	1.054		1.198
	(–0.230)		(2.31)‡
Father is Native American	0.791		0.986
	(–1.160)		(–0.080)
Father is other/unspecified	1.052		1.025
	(–0.410)		(–0.220)
Mother is black	0.832	0.771	
	(–1.930)	(4.61)†	
Mother is Latino	1.109	1.018	
	(–1.060)	(–0.230)	
Mother is Asian	1.219	1.285	
	(–0.650)	(–1.510)	
Mother is other/unspecified	0.923	0.903	
	(–0.790)	(–1.240)	
Mother is Native American	1.926	1.735	
	(–1.930)	(–1.870)	
Observations	8942	8942	8942

* Robust z statistics in parentheses.
† Significant at 1 percent.
‡ Significant at 5 percent.

way to assess the quality of the statistical model is to compare the predicted probability that a test set would show an inclusion from a statistical model to the actual inclusion outcome. To do this we use the model in the first column of Table 4 to generate estimates of a predicted inclusion rate for each child, based solely on the demographic characteristics included in the regression. These predicted inclusion rates turn out to range from 0.62 to 0.85. To have clinical utility, one could consider setting thresholds based on the statistical prediction, for example, predict paternity inclusion if regression score is greater than 0.65, greater than 0.70, etc. Setting the threshold low would generate many false inclusions and have low specificity. Setting the threshold high would generate many false exclusions and have low sensitivity. Table 5 shows the sensitivity, specificity, and number of correct classifications at various thresholds. The table shows that if one simply "guesses" that all men would be included in paternity tests, this guess performs as well or better than any possible cutoff value of the demographic index at correctly classifying test subjects.

DISCUSSION

DNA-based paternity tests identify the alleged father as the actual father in only 72 percent of cases in a sample of genetic tests that is weighted to be equally representative of paternity establishment caseloads in the United States for 2002. There is almost no evidence that couples with

TABLE 5. Performance of demographic variables at predicting paternity inclusion at various cutoff values*

Cutoff value for predicted inclusion†	Specificity (%)	Sensitivity	Percent correctly classified
0, all assumed included	0	100	72
0.65	0.25	99.91	72
0.70	22.26	81.88	66
0.75	72.24	33.78	45
0.80	98.66	2.47	28
100, all assumed excluded	100	0	28

* Statistical prediction of paternity inclusion based on first column of Table 4. Predicted odds of paternity inclusion ranged from 0.62 to 0.8.

† All values above this threshold considered included.

dissimilar ethnic backgrounds have a higher or lower rate of paternity inclusion than couples of similar backgrounds.

Only a few weak associations between parental race and age and the likelihood of a paternity inclusion in a paternity test were found. Although it might have been the case that collinearity between maternal age and race could lead to a loss of precision in estimating independent race and age effects, Table 4 showed that as long as only one parent was included both age and race effects could be estimated as significant. Collinearity between mother's and alleged father's race appears to be limiting the precision of estimates of racial effects on inclusion in the multivariate models in column 1 of Table 4. We speculate that older mothers involved in disputed paternity may potentially have less exposure to true uncertainty due to multiple sexual partners during the month of conception. This may account for the higher rates of inclusion in older mothers, but we point out that the odds ratio (OR), although significant, is only 1.02. Consequently these variables make almost no contribution to one's ability to predict a man's paternity in the absence of a genetic test.

We conclude that knowing parental race, ages, and sex of the child is of little assistance in predicting which samples will produce a paternity inclusion. It remains an open research question how many of the test results come as a complete surprise to the parties involved. A woman's own knowledge of her sexual activity during the month of conception would give her a tremendous advantage in predicting who the father of her child is likely to be. Cases of contested paternity rest on a man's uncertainty that he was the sole partner during the month of conception. Our statistical analysis rests on considerably less information than would be available to the parties involved, so it is possible that both the man and the woman are able to predict their inclusion rates with a lot more accuracy.

REFERENCES

1. Annual report summary for testing in 2000. Prepared by the Parentage Testing Standards Program Unit. Bethesda: American Association of Blood Banks; 2000.
2. Statistical abstract of the United States: 2000. 111th ed. Washington (DC): U.S. Bureau of the Census; 2000.
3. FY2000 preliminary data preview report. Washington (DC): Office of Child Support Enforcement; 2001.
4. Aizer A, Mclanahan S. The impact of child support enforcement on fertility, parental investment and child well-being. Working Paper 11522. Cambridge (MA): National Bureau of Economic Research; 2005.
5. Seltzer J, McLanahan S, Hanson T. Will child support enforcement increase father-child contact and parental conflict after separation? In: Garfinkel I, McLanahan S, Meyer D, Seltzer JA, editors. Fathers under fire. New York: Russell Sage; 1998. p. 157-90.
6. Graham JD, Beller AH, Hernandez PM. Effects of child support on educational attainment. In: Garfinkel I, McLanahan S, Robins PK, editors. Child support and child well-being. Washington (DC): Urban Institute Press; 1994.