

Winter 1991

Estimates of Census Underenumeration Based on Genealogies

John W. Adams

University of South Carolina - Columbia, jwadams0@mailbox.sc.edu

Alice Bee Kasakoff

Follow this and additional works at: https://scholarcommons.sc.edu/anth_facpub



Part of the [Anthropology Commons](#)

Publication Info

Social Science History, Volume 15, Issue 4, Winter 1991, pages 527-543.

© [Social Science History](#) 1991, Cambridge University Press.

This Article is brought to you by the Anthropology, Department of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

Estimates of Census Underenumeration Based on Genealogies

JOHN W. ADAMS & ALICE BEE KASAKOFF

WE HAVE BEEN STUDYING the migrations of the descendants of nine men who came to Massachusetts before 1650 and have compiled a computerized database that includes all the people born before 1860 in the patriline. Thus we have what the nine genealogists who studied these families thought was close to a complete list of family members alive in 1850. Here we focus on our attempts to find these individuals on the 1850 federal census.

To facilitate our task, we made up a search list that contained all males alive in 1850, but we omitted females known to have married by 1850. The search list included both the last known

John W. Adams and Alice Bee Kasakoff are professors of anthropology at the University of South Carolina. Their current research, which uses the database described in this article, focuses on changes in migration and family patterns as the American North filled (1620–1880). An earlier version of the article was presented at the November 1989 meeting of the Social Science History Association in Washington, DC. The research has been supported by a fellowship from the National Endowment for the Humanities, administered by the Newberry Library; a separate fellowship from the Newberry Library; grants from the National Science Foundation, its Geography and Regional Sciences Program (SES nos. 8016384 and 8501821), its Anthropology Program (BNS no. 8305214), and its EPSCOR Program, administered through the University of South Carolina; and a Faculty Research and Development Grant from the University of South Carolina. The authors are also grateful to the National Endowment for the Humanities for fellowships to attend a summer seminar in economic history with Stanley Engerman, and to the many students who have collected, coded, and keyed in the census data. Special mention is given to Dawn Deaton, Dan DuPre, Andrea Kao, Kimberley Kelly, A. and R. Rajendran, Conrad Reining, Jr., and Jean Russo.

Social Science History 15:4 (Winter 1991). Copyright © 1991 by the Social Science History Association. CCC 0145-5532/91/\$1.50.

place in each individual's record before 1850 and the next known place after 1850.¹ The list was deliberately over-inclusive. In the 30% or so of the cases in which we did not have death dates for individuals, we carried them on the list until they were 100 years old. Again, the census is supposed to have been taken as of 1 June, but we included anyone who was born or died in 1850, whether it was before or after that date. All told, we searched for 7,627 individuals, about two-thirds men and one-third unmarried women.

Our first step was to use the *Accelerated Indexes* of the census, which were available for all states.² The indexes list every different surname within a household, enabling us to find individuals who had left home and were living with strangers.³ Relying on the indexes alone, we were able to find 72% of the men aged 20 or over.

For people who could not be found through the indexes, we then hand-searched towns where the genealogies placed them. Our assistants also looked at the agricultural censuses and sometimes found additional people.⁴ Altogether, this more intensive phase yielded an additional 7%, for a total of 79% in all age groups.

THE EXTENT OF UNDERENUMERATION

Tables 1 and 2 present the proportion of each five-year age group that we were able to recover. They treat as "eligible" only those individuals with an event on record after 1850. The genealogist had lost track of some of these people, but if we found them on the 1860 census, they were included as eligible.⁵ Thus, we restrict ourselves to the people who we know were alive after the census. The third column is the number of people found, divided by the number eligible. However, these rates required some minor corrections.

We removed the people who were born in 1850 but too late to be included in the census or who died in 1850 before it was taken. We simply assumed that the census was accurate as of 1 June, as it was supposed to be, and removed people according to the exact date of the event on the genealogy.⁶

We also added people to the "found" list who we were sure had been enumerated but whom for various reasons we could not link conclusively to the people on the eligible list. Most of these were living away from their families, so the usual clues for linkage

Table 1 Recovery rates of males alive in 1850

Age in 1850	Eligible	Found	Raw percentage found	1850 event	Linkage	Corrected percentage found
0	100	42	42	-47		79
1-4	336	258	77	-5		78
5-9	425	349	82			82
10-14	431	345	80		3	81
15-19	409	335	82		5	83
20-24	335	254	76	-1		76
25-29	297	216	73		6	75
30-34	264	195	74	-2	5	76
35-39	249	195	78	-1		79
40-44	212	173	82	-1		82
45-49	186	163	88	-1		88
50-54	157	130	83			83
55-59	104	91	88			88
60-64	90	78	87	-1		88
65-69	75	60	80	-2		82
70-74	59	50	85	-1		86
75-79	30	24	80	-1		83
Over 79	30	22	73			73
All	3,789	2,980	79	-63	19	81

Source: Database taken from nine printed genealogies (Bisbee 1956; Chaffee 1909; Farwell 1929; Faunce 1973; Greely 1905; Holman 1928; Pelton 1892; Shedd 1921; Wellman 1918) linked to U.S. Census Office 1850.

Note: Eligibles are all men with an event recorded after 1850. Corrected rates are obtained by subtracting column 4 from the eligibles and adding column 5 to those found.

(such as names of wives and children) are missing.⁷ The adjusted recovery rate for males is 81% overall. It is somewhat lower during the first five years of life and hits a plateau in the low 80s in the age group 15-19. It then declines to a low point of 75% for those in their late 20s, rises again to the high 80s for some ages, and falls for men in their late 70s or older.

For women, the overall rate is only 77%, but the age patterns are quite similar. The unmarried women over 30 were less apt to be recovered than men of the same age. Some of these older women had probably married by the time of the census, but they were carried along as unmarried because the genealogist did not

Table 2 Recovery rates of unmarried females alive in 1850

Age in 1850	Eligible	Found	Raw percentage found	1850 event	Linkage	Corrected percentage found
0	70	31	44	-31		79
1-4	302	227	75			75
5-9	369	285	77	-1		77
10-14	323	270	84		6	85
15-19	234	199	85	-1		85
20-24	130	100	77			77
25-29	70	42	60	-1	2	64
30-34	51	25	49			49
35-39	26	13	50	-1		52
40-44	23	12	52			52
45-49	18	14	78			78
50-54	10	4	40			40
55-59	11	7	64			64
60-64	9	5	56			56
65-69	4	3	75			75
70-74	4	4	100			100
75-79	0	0	0			0
Over 79	0	0	0			0
All	1,654	1,241	75	-35	8	77

Source: See Table 1.

Note: Eligibles include all women not married by 1850 who had an event recorded after 1850. Corrected rates are obtained by subtracting column 4 from the eligibles and adding column 5 to those found.

record their marriages. In the early years, between the ages of 1 and 9, however, the recovery rate for girls was a few percentage points below that for boys, probably an indication of the girls' greater underenumeration.⁸

Had we hand-searched the cities where we thought some of the unrecovered people were living, we could certainly have found more, so the true rate of underenumeration is probably a bit lower. But even if all the unrecovered people on our list who the genealogist thought were living in cities were actually found there, the recovery rate would be raised by only 2%.

Another estimate of underenumeration can be obtained by asking what proportion of the people we found on the 1860 census

Table 3 1850 recovery rates for people found in 1860

Age in 1850	Males		Females	
	Percentage	(No. eligible)	Percentage	(No. eligible)
0	80	(40)	85	(26)
1-4	88	(200)	85	(213)
5-9	91	(263)	86	(235)
10-14	87	(226)	86	(139)
15-19	89	(191)	98	(47)
20-24	83	(191)	85	(27)
25-29	83	(57)	76	(17)
30-34	86	(144)	93	(14)
35-39	87	(141)	44	(9)
40-44	93	(114)	50	(6)
45-49	95	(109)	83	(6)
50-54	91	(85)	0	(1)
55-59	93	(59)	60	(5)
60-64	95	(42)	80	(5)
65-69	85	(27)	100	(1)
70-74	95	(21)	100	(1)
75-79	100	(7)	0	(0)
Over 79	100	(1)	0	(0)
All	88	(1,918)	85	(752)

Source: See Table 1; database also linked to U.S. Census Office 1860.

Note: Eligibles are people found on the census of 1860. Rates have been corrected in the same way as they were for 1850 (see Tables 1 and 2).

were found on the 1850 one (Table 3). Overall, the rates of recovery are some 7% higher than the ones we reported above. The age patterns are essentially the same.⁹

Why the difference? There are several reasons. First, a person found on one census is likely to be found again. Thus, a retrospective search is always more successful. Second, we know that the group we found for 1860 was not representative of the general population. Not every state had been indexed for that year, so we used rather different search procedures and recovered more older than younger men, more farmers than nonfarmers, and more people living in rural areas. As we show below, and as Ginsberg has suggested, these features make people significantly more likely to be enumerated on a census.¹⁰

COMPARISON WITH OTHERS' FINDINGS

Our estimates of underenumeration are higher than some others, which have converged on a figure of "7 to 10%" (see, for example, Eblen 1965; Knights 1971). Of course, genealogists are not without their own foibles; like the enumerators, they also have difficulty following the most mobile segments of the population. This certainly explains why at least some of our hand searches failed to turn up any "finds." On the other hand, some 90% of those we found were living in the state where the genealogist had placed them. Furthermore, most migration was over short distances, and many of the families were geographically clustered, so we searched by hand most of the places where the lost people should have been living, especially if they were living with other family members. We were quite likely to find most of these short-distance migrants if they had been enumerated.

In a previous study in which we used the census to check on the accuracy of the genealogist (Adams and Kasakoff 1988), we estimated that they lost some 9% of the children, but the losses were concentrated in the newly settled areas. Only 3% of the children censused in southern New England, for example, were missing from the genealogy, though 23% of the children censused in the Midwest were not mentioned by the genealogist.

All of this suggests that the proportion missed because the genealogist lost track of people occurred largely in the areas most recently settled. But few people from our families were living there, at least according to the genealogies (4% in 1850; see Table 5). Granted that some of the people we were unable to find in the East were really in the West, it is still improbable that many of them were, given the barrier the Plains posed to further migration.

Moreover, our estimates agree rather well with Ginsberg's (1988) study comparing a map of households in a small area of Massachusetts with those enumerated in the census of 1870. She found 80% of the mapped households overall and 85% of those individuals living in sparsely populated areas. Her analysis focused on households, however, not on individuals. We find slightly more men in their 40s than she found household heads. This may mean that our population was living in more sparsely populated areas or that enumeration for our group was better

overall than for the one she was studying, which included several foreign-born. In any case, our figures are remarkably close to hers.

We suggest that the lower underenumeration estimates of other researchers result from the higher recovery rates of their retrospective searches and from their counting household heads rather than dependents.

UNDERENUMERATION OF CHILDREN

As far as we know, this is the first study to include women and children in an estimate of undercount. Including them allows us to determine whether there was underenumeration within households as well as skipping of entire households.

The majority of the missing children—66% of the children aged 1–14 whom we did not find on the census—however, were not counted because their fathers, while living, had not been enumerated. Another 13% were children whose fathers were dead.¹¹ These children, though they accounted for only 6% of the group we were looking for, were particularly hard to find on the census; we were only able to locate slightly more than half of them, much less than the 79% recovery rate for all children. The rest of the unrecovered children, some 20%, were children who were not found even though their fathers were.

We examined the cases of missing children aged 1–4 whose fathers were found and discovered that about a quarter of them were in fact listed on the census but were missed by our own assistant researchers; another quarter probably should not have been on the eligible list because their birth or death dates were quite approximate. All in all, we doubt that the increase in recovery rates with age is nearly as steep as it seems from Tables 1 and 2. Our revised recovery rates for ages 1–4 are 81% for males and 78% for females.

The number of children under age 5 is often used to estimate fertility in child-woman ratios. If the missing children whose fathers were found were all true cases of underenumeration, the child-woman ratio would be underestimated by about 5%. Of course, some of the children we did not find may have been enumerated elsewhere. Perhaps, if their mothers were dead, they were listed under other surnames. On the other hand, these should be offset

Table 4 Logistic regression predicting enumeration in 1850 (adult males only)

Variable	Mean	Beta	Chi-square	<i>p</i>	<i>R</i>
Logd20	2.40	-0.11	8.08	0.0045	-0.063
Farm	0.69	0.47	11.54	0.0007	0.079
Fert	5.10	0.01	0.27	0.6038	0
Age group					
Elder	0.10	-0.11	0.21	0.6455	0
Young adult	0.38	-0.47	8.43	0.0037	-0.065
Married	0.84	0.53	8.19	0.0042	0.064
Model chi-square (6 d.f.) = 67.41					
<i>p</i> = 0.00000 <i>R</i> = 0.185					

Source: See Table 1.

Note: Number of males aged 20 or older: 1,513; number of those men found on the 1850 census: 1,209 (found = 1). There are five independent variables: (1) Logd20, a measure of mobility, the natural logarithm of the distance traveled before age 20; (2) Farm, whether or not the man was a farmer (these are lifetime occupations, and the farm group includes men who pursued both farm and non-farm occupations in their lives; farmer = 1); (3) Fert, the person's total lifetime fertility; (4) Age group, in three age categories devised to reveal the patterns of underenumeration discussed: Elders, men over age 64; Young adults, men aged 20-34; Middles, men aged 35-64 (more likely to be enumerated than the others and not listed in the regression); (5) Married (= 1).

by children whose fathers had died, who might not be enumerated, either. Our best guess would place the child-woman ratio as perhaps 4% too low.

UNDERENUMERATION OF ADULT MALES

As for missing adults (males aged 20 or older), we have done logistic regression to explain who was or was not censused (Table 4). This regression, as expected, demonstrated that the more mobile individuals were less likely to be found on the census. Nonfarmers were also less likely to be found than farmers. Among the age groups, the younger adults were significantly less likely to be found than the middles, but there was a smaller difference between the elders and the middles, and it was not significant. Married persons were more likely than unmarried to be found, even when age had been controlled for. Fertility did not affect the probability of being found.

Table 5 Proportion recovered by last known region

Region in 1850	Males		Females	
	Percentage	(<i>N</i>)	Percentage	(<i>N</i>)
Northern New England	86	(1,077)	82	(497)
Southern New England	82	(919)	80	(444)
NY-NJ-PA	78	(832)	79	(347)
Old Midwest	78	(425)	74	(179)
New Midwest	72	(140)	58	(55)
Other	48	(85)	52	(23)
All	80	(3,478)	78	(1,545)

Source: See Table 1.

Note: *N* is the number of individuals known to be in the region according to the last known place before the 1850 census. Figures are adjusted for births and deaths in 1850. New Midwest: Illinois, Kansas, Minnesota, Nebraska, Missouri, Wisconsin. Old Midwest: Ohio, Indiana, Michigan.

We have also tabulated recovery rates according to the regions where the genealogist expected the people to be found (Table 5). The highest recovery rates were for people whose last known place of residence was northern New England. The rates for people whom the genealogist expected to be living in southern New England were also high. The lowest rates were for people whom we expected to find outside the North altogether and for those in the New Midwest. It is interesting that this same pattern was found in overenumeration. Overenumeration was highest in northern New England, second highest in southern New England, and lowest in the Old Midwest.

It is probably no surprise, then, that we were able to find so many people in the areas of the country most densely populated by our families. Dense population would also provide the greatest opportunity for double residence with relatives. Since newer regions also experienced more migration, these findings are consonant with the idea that the most mobile individuals were the most likely to be missed.¹²

EVIDENCE OF OVERENUMERATION

There was a small but quite interesting group of individuals who were censused more than once. In fact, two individuals were actu-

ally censused three times—one, Sally Chaffee, aged 12, probably on the same day in the same town, Amity, Pennsylvania, within the space of some 50 households! Her case exemplifies one of the reasons for overenumeration: circulation of children. She was listed with her natal family, where she came fifth on a list of nine children (there were three older sisters and a younger one at home); she was also listed with each set of grandparents, whose households were much smaller. Her mother's parents would have been living alone were it not for Sally; her father's father was dead, but his widow was living with her youngest son, who had just married.

Our second thrice-found was Peter Shedd, aged 25. He exemplifies another type of overenumeration, that of the young family just starting out. Peter was found in Milford, New Hampshire, in his father's household, listed as the oldest child, along with five brothers and four sisters. Two days before, he had been censused in Wilmington, Massachusetts (Middlesex County), as a laborer with no property and two young children, the eldest born in New Hampshire.¹³ The same household was also listed in Danvers, Massachusetts, again showing Peter as a laborer.¹⁴

Bona fide examples of the same family enumerated twice as independent households are remarkably rare; in addition to Peter Shedd's, we found only three, all young couples. Several other families were enumerated twice but not as independent households. Usually, these are young couples found once living independently and a second time living with the husband's or wife's parents. The youngest children found twice were children of these couples, except for one three-year-old boy whose mother had died, probably giving birth to him. He was listed both with his mother's parents and with his father, who had remarried.

Table 6 gives the age distribution of people enumerated more than once. They amount to 1% of the males we have found and 1.4% of the females. However, 3% of the girls aged 15–19 were censused more than once, as were 4% of the boys in the next age group. Finally, there is also evidence of overenumeration among the elderly. Overenumeration, then, offsets some of the underenumeration in the same age groups, but it does not do so completely.¹⁵

The girls censused twice were, like Sally Chaffee, often from neighboring farm households. Usually in these cases, a younger

Table 6 Percentage censused two or three times, by age and sex

Age in 1850	Males		Females	
	Percentage	(N)	Percentage	(N)
0	0		2.1	(1)
1-4	0.3	(1)	0.9	(3)
5-9	0		0	
10-14	1.0	(4)	1.6	(6) ^a
15-19	1.6	(6)	3.1	(9)
20-24	4.2	(12)	1.4	(2)
25-29	1.7	(4) ^a	2.7	(2)
30-34	1.0	(2)	1.9	(1)
35-39	0.5	(1)	0	
40-44	0		4.3	(1)
45-49	0		0	
50-54	0		0	
55-59	0		0	
60-64	0		0	
65-69	0		0	
70-74	0		0	
75-79	3.2	(1)	0	
Over 79	3.8	(1)	0	
All	1.0	(32)	1.4	(25)

Source: See Table 1.

Note: Percentages are those of people found at all who were found more than once.

^aIncludes one person found three times.

married brother or sister took in a middle daughter of an older sibling who had several other children at home. The younger sibling was at a much earlier stage in family formation, when there were several very young children, at a stage during which researchers have hypothesized a severe economic squeeze on the family (Smith 1984). There are also examples of poor families with several girls who sent one to a wealthier but unrelated family; the youngest such girl we know of was seven. These live-in helpers (if that is what they actually were) remained quite near their natal families.

This sort of labor exchange occurred less often with boys. As a result, the overall rate of overenumeration was lower for them than for girls. Probably, if a family needed the services of a boy

Table 7 Percentage living at home, by age and sex

Age in 1850	Males		Females	
	Percentage	(N)	Percentage	(N)
0	98	(60)	98	(48)
1-4	99	(314)	97	(319)
5-9	99	(388)	99	(392)
10-14	98	(369)	97	(352)
15-19	92	(340)	91	(298)
15	97		95	
16	97		95	
17	90		93	
18	87		86	
19	89		91	
20-24	65	(229)	88	(136)
20	79		90	
21	69		89	
22	69		89	
23	54		93	
24	50		81	
25-29	41	(170)	73	(62)
30-34	26	(123)	79	(34)
35-39	11	(91)	76	(13)
40-44	19	(64)	70	(10)
45-49	7	(41)	67	(6)
Over 49	24	(42)	75	(4)
All	77	(2,231)	94	(1,674)

Source: See Table 1.

Note: Denominator is the number of children whose fathers were also found on the 1850 census. Numerator is the number of times they were both found in the same household.

who also lived nearby, he continued to live at home. But a girl may have been more useful if she lived with the family she was helping.

One would expect a "bathtub-shaped" distribution of overenumeration: a large number enumerated twice within a few days in neighboring households, like Sally; then the number falling off but slowly rising again as the time between entries increased. Cases of overenumeration whose dates we have examined bear this out: those found within a few days of each other seem to involve tem-

porary or dual residence; those found farther apart in time, people who moved before the census was completed.

These cases of overenumeration allow us to glimpse some of the short-term movement that has usually been invisible both in censuses and (particularly) in genealogies. They alert us to four periods of residential fluidity during the life cycle. The first occurred when a person left home; the second, early in marriage, when some married couples, especially poor ones, lived with either the wife's or the husband's family; the third, when widowed women or their children were taken in by in-laws; and the fourth, when poor elderly widowers were boarded with strangers or in almshouses but alternated between this arrangement and residence with married children or grandchildren.

CONCLUSION

Most of the literature on enumeration has focused on either adult males or missing households, but not on dependents. Our analysis suggests that the underenumeration of this elusive segment of the population was considerable; for each unenumerated parent there were several missing children, and to them we must add the children who either worked away from home or were orphaned. The undercounting of children is particularly important because the population was so young.

Our focus on dependents has also brought to light the phenomenon of overenumeration. Though the group prone to overenumeration is the same as that prone to underenumeration, and the two offset each other somewhat, the effects are far from equal in magnitude. A large proportion in the youngest age groups escapes enumeration altogether. Claims of underenumeration result from a comparison of census numbers with (better and worse) "guesstimates" of the size of the actual population. So if the census numbers are padded with double counts, the real population is even more greatly undercounted than previously supposed. The problem may be quite severe for many minority groups in the U.S. today, especially those that move frequently and among which child fosterage is common, for they often cause overenumeration. The census totals may contain several people who have been counted twice. This would mean that the group was even more underenumerated than it would appear from the published figures.

(Margo Anderson [pers. com., 1989], then of the Census Bureau, informed us that double counting was being studied in connection with the undercounting of minorities in the 1990 census. But it is well to remember that it is not entirely absent from any census of population.)

Using both the census and genealogies allows us to observe more than we could by using either source alone. The census uncovers forms of circulation that are impossible to observe through genealogies, but genealogies allow us to link individuals living away from home to their natal families. They also allow us to examine the constellation of relatives who were available to help out during critical periods. Together, the two sources furnish unique and unexpected information on how families coped with dependency changes inherent in their growth and decay.

NOTES

- 1 As we progressed, we eliminated people we had already found from the list and updated it from our "finds" on the 1860 census, which we were also searching.
- 2 We quickly discovered a set of alternate spellings and misspellings of the surnames we were seeking and looked at the indexes under several listings in an effort to find the people we wanted.
- 3 In each state searched, we copied all the households to which the indexes led us. It was only later that we tried to link them to the genealogies. So a person who was not where we expected him or her to be on the basis of the search list would still have been found. We searched the entire North. (In 1850 some 2% of the people alive were living in Canada. We do not discuss them further.)
- 4 We are grateful to Peter Knights for sending us his copy of the agricultural census of Maine, which was not otherwise available to us outside the state.
- 5 Some 30% of all the people we found on the census were people for whom we had no record of their having lived beyond 1850; they are not included in Tables 1 and 2.
- 6 When the exact date of a birth or death in 1850 was not known, we assumed that five-twelfths had occurred before the census.
- 7 For example, we collected more than one 20-year-old John Chaffee living in Connecticut away from home, and we also have records of the natal households from which they are missing; but which is which? These Johns continue in our records as having not been found. This problem especially plagues the very large genealogies and usually occurs only in areas where the family is clustered. Since most of the unlinked individuals had common given names, we made the adjustment on that basis. There would seem to be no reason why males with common given names should be enumerated

more or less often on the census; they should be found at the same rate as people with rare given names. But in our data the recovery rate of men with rare given names is 5% higher than that of men with the most common given names. Inexplicably, the difference is most pronounced in the age group 5–9, where we had expected no difficulty, since the children were at home and we had had very little problem linking them to the genealogy. But there were also pronounced differences in recovery rates at the ages when men were apt to be living away from their families, especially in the late 20s, the age for which underenumeration appears to be greatest. The difference in recovery rates at this age is 6.7% (in favor of the rare names). Our solution was to add men to the “found” column at the ages when men lived away from home, so that the rate for common names reached that for rare names. (John was the most common given name, accounting for some 6.4% of all males on the search list, followed closely by William, 6.3%, and then George, Charles, and James. Some 33% of all males had the eight most common given names.)

- 8 The recovery rate of women with rare given names was actually lower overall than that of women with the more common names, except for the two age categories that have been corrected. In determining the list of common names, we used just the first five letters. For men this procedure caused no problem, but for women the most common given name, Mary, was often followed by another name; Eliza, the next most common, could also have been Elizabeth.
- 9 The greatest discrepancy occurs in middle adulthood: starting with men in their 30s, the recovery rates are usually 10% to 12% higher than those we reported before.
- 10 The lack of indexes led us to hand searches, and, since we paid our student assistants on the basis of their productivity, they quite naturally searched the towns where they would find the most with the least effort. Many used the agricultural censuses as guides. When we realized the difficulty of finding people in 1860, we decided to concentrate on 1850 instead. In the states for which there were indexes, our 1860 group is more diverse, but in the other states, unfortunately a majority of those we were looking at, it contains a disproportionate number from rural areas where our families had established several farms.
- 11 These children were not often living away from home (see Table 7), so it appears that some children were not enumerated even when their natal households were found. Such underenumeration appears to decline with age. The highest rate occurs in the group aged 1–4, where about 9% of the children whose fathers were found were missed; it falls to 4.6% at ages 10–14. There also seems to be a tendency on the part of either the parent or the census taker to forget children at the end of a long list. Contrary to what we expected, there was no bias against females in this “forgetfulness.” In other cases, the children’s mothers had died and the father no longer lived with them.
- 12 We were also interested in finding out whether poor people were less likely to be found. We were able to test this possibility using the information on the subset of men found in 1860 (some 985 individuals, 56% of the larger

group described above). Did the wealth of these men in 1860 help explain whether or not they had been found in 1850? We ran the same regression, adding the logarithm of the sum of their real and personal property in 1860. The sign was positive, but the magnitude was low and the variable not significant. We also tried a dichotomous variable indicating whether or not the individual held any property at all, real or personal. This result also was not significant. We looked at the three age groups separately, hoping that wealth would predict the likelihood of being found for at least some age groups, but the best result we obtained was $p = .09$ for the dichotomous variable for the younger group (between ages 20 and 35), but for this group the entire model was not significant. Since the subset of men found in 1860—a more stable group, biased towards farmers who had remained in older towns—was probably wealthier than the entire set of eligibles in 1850, there may in fact be a difference, but we were unable to detect it due to the special nature of the group we located on the census of 1860.

- 13 He is listed some 40 households after his father's brother and wife, who had no occupation but were living in a household headed by their son, who was, like Peter, also newly married and a laborer. It is tempting to see the two cousins moving there together.
- 14 In the two listings some of the ages and birthplaces are different. The Danvers listing says that all the people were born in Massachusetts; according to the other listing, the wife was born in Vermont. The youngest child in Danvers is listed as having been born three months later than the same child in the Wilmington listing—or were they censused three months later? The genealogy confirms the Wilmington information. The Danvers enumerator was a frequent dittoer of birthplaces and not very careful. In the Milford listing, both Peter and his father were farmers, but his father's real property was worth only \$700, far below the average value of farms in New Hampshire at the time. Peter had no real property on any of the listings.
- 15 The recovery rates in Tables 1 and 2 count each "find" only once. Also, it is quite probable that some of the linkage problems we mentioned above for these same age groups are actually cases of double listings. If so, the number in those age groups would be even higher, and if we had not stopped looking for people whom we had already found in the more intensive phase, we certainly would have found more.

REFERENCES

- Adams, John W., and Alice Bee Kasakoff (1988) "The effect of sex ratios of children on the occupations of their fathers: Evidence from the American North, 1750 to 1860." Paper presented at the annual meeting of the Social Science History Association, Chicago, 5 November.
- Anderson, Margo (1989) Conversation with authors, November.
- Bisbee, Frank J. (1956) *Genealogy of the Bisbee Family*. East Sullivan, NH: Privately printed.
- Chaffee, William (1909) *The Chaffee Genealogy*. New York: Privately printed.

- Eblen, J. E. (1965) "An analysis of nineteenth-century frontier populations." *Demography* 2: 399–413.
- Farwell, J. D. (1929) *The Farwell Family*. Orange, TX: Privately printed.
- Faunce, James Freer (1973) *The Faunce Family: History and Genealogy*. Akron, OH: Privately printed.
- Ginsberg, Caren A. (1988) "Estimates and correlates of enumeration completeness: Censuses and maps in nineteenth-century Massachusetts." *Social Science History* 12: 71–86.
- Greely, G. H. (1905) *Genealogy of the Greely-Greeley Family*. Boston: Privately printed.
- Holman, Mary Lovering (1928) *Ancestors and Descendants of John Coney of Boston, England, and Boston, Massachusetts*. Concord, NH: Privately printed.
- Knights, Peter R. (1971) *The Plain People of Boston, 1830–1860: A Study in City Growth*. New York: Oxford University Press.
- Pelton, J. M. (1892) *Genealogy of the Pelton Family in America*. Albany, NY: Privately printed.
- Shedd, Frank E. (1921) *Daniel Shed [sic] Genealogy*. Boston: Privately printed.
- Smith, Richard M. (1984) "Some issues concerning families and their property in rural England, 1250–1800," in Richard M. Smith (ed.) *Land, Kinship, and Life-Cycle*. Cambridge: Cambridge University Press: 1–86.
- U.S. Census Office (1850) *Seventh Census of the United States*. National Archives, Washington, DC.
- (1860) *Eighth Census of the United States*. National Archives, Washington, DC.
- Wellman, Joshua W. (1918) *Descendants of Thomas Wellman of Lynn, Massachusetts*. Boston: Privately printed.