# Creating a New Database for the Study of Southern Lynchings: Public Use Microdata, *The Historical United States County Boundary Files 1790 - 1999* and Forensic Demography

**Stewart E. Tolnay ***
Sociology Department
University of Washington
Seattle, WA 98195-3340

and

**Amy Kate Bailey**
Sociology Department
Center for Studies in Demography and Ecology
University of Washington
Seattle, WA 98195-3340

* Submitted for consideration to the Population Association of America for presentation at the 2006 Annual Meeting. This research was supported by a grant to Tolnay from the National Science Foundation (SES0521339). We are grateful to Patty Glynn for programming support. Direct correspondence to: Stewart Tolnay, Department of Sociology, Box 353340, University of Washington, Seattle, WA 98185; email: tolnay@u.washington.edu.

During the late 19th and early 20th centuries, thousands of African Americans were lynched in the southern region of the United States. Previous research on lynching has been based largely on existing inventories of the victims of this form of extralegal punishment. While extremely valuable, these inventories suffer from important limitations in how much they can tell us about those victims, about the local conditions that affected the risk of an individual becoming the victim of a lynch mob, or about how the relationship between individual characteristics and the likelihood of being lynched may have varied across time and place. We are creating a new source of data that will introduce the individual victim into research on lynching, and provide information about the social relations and environment within which each victim was embedded. Our procedure merges information from an inventory of nearly 2,800 individuals who were lynched between 1882 and 1930 (Beck and Tolnay 2005 – See Table 1) with: 1) the original enumerators' manuscripts for the Census immediately preceding the lynching (e.g., the 1880 Census for an 1885 lynching); and 2) geocode information on the county of lynching and county of residence for identified victims. The main contribution of this work is twofold: we develop new spatially-oriented methods of matching individuals to their census records; and, we create an historical data source that will allow researchers studying lynching, hate crimes, and ethnic conflict to link information for individuals and households to characteristics of the local context within which the victim resided or was lynched.

**Review of the Literature**

A variety of approaches have been used to document the "tragedy of lynching" and to gain a better understanding of the social, cultural, economic, and demographic factors that promoted it. Social activists and anti-lynching advocates made important contributions during the lynching era, itself, by documenting the phenomenon and attempting to link lynch mob activity to underlying social and economic forces (Ames 1942; Raper 1933; Wells-Barnett 1892 [1969], White 1929 [1969]). Case studies have also made important contributions, marshalling extensive evidence to provide rich detail on single lynching incidents and the circumstances surrounding them (Dinnerstein 1968; Downey and Hyser 1991; Griffin 1993; Smead 1986; Wexler 2003). However, while telling thorough and nuanced stories, the extent to which the facts of a single event can be generalized to a larger set of lynchings is questionable.

The most common sociological strategy for studying lynching has been comparative, using inventories that include limited information about a large number of lynchings to examine the distribution of incidents or victims over time and across space, and covariation between temporal or spatial rates of lynching with corresponding variation in social, economic, political or demographic factors. Most comparative analyses of lynching have been specifically designed to test theories about the antecedents of lynching, for example, regarding the influence of economic conditions, political climate, or racial composition (see, e.g., Beck, Massey and Tolnay 1989; Beck and Tolnay 1990; Brundage 1993; Corzine, Creech, and Corzine 1983; Corzine, Huff-Corzine and Creech 1988; Olzak 1990; Soule 1992; Tolnay and Beck 1992a, 1995; Tolnay, Beck and Massey 1989; Tolnay, Deane and Beck 1996). While complete consensus might not, yet, exist, previous comparative studies have indicated that southern lynchings (primarily of African Americans) were more likely to occur where or when (1) the black population was proportionately larger, (2) the white population suffered economic hardship, (3) the Democratic party was stronger, (4) the white population felt economically or socially threatened by the black population, and (5) the level of black out-migration was lower. Although the comparative approach yields conclusions that are more generalizable, it offers much less detail than the case study approach.

One distinct limitation of comparative analyses of lynching is their heavy reliance on cross-sectional, generally county-level, analyses to draw conclusions about the conditions that created a climate favorable to mob violence; the nature of existing lynching inventories has precluded analyses in which individuals are the units of analysis and the dependent variable describes whether or not the individual was lynched within a given time period. A second limitation to prior work is the extremely limited amount of information about the lynching victims that is included in existing inventories; we know that black males were more likely than other individuals to be lynched, but have little additional information beyond that. This may be particularly important given what we know about the context of lynchings, as the kinds of individuals targeted for lynching may have varied with community characteristics. For example, in some contexts "drifters" with few ties to the local community may have been chosen as lynch targets, while in others, prosperous blacks

may have been "made an example of" through lynching. The data source we are creating will enable researchers to explore the interaction between spatial, social, and individual factors and to test hierarchical models that locate individual lynch victims within their communities.

An illustration of the information to be gained by linking lynch victims to their census records, and then to data about their community of residence and the county in which they were lynched may be useful. From the Beck-Tolnay inventory of lynch victims, we know that Theo Calloway, a black male, was lynched in Lowndes County, Alabama, on 28 March 1888, after he was accused of committing murder. Prior to linking this victim to his census records for 1880, that is the extent of our knowledge about Mr. Calloway. However, after the successful linkage, we also know that Mr. Calloway: (1) was 45 years of age in 1880, and therefore 53 years old when he was lynched; (2) was born in Alabama; (3) was a farmer; (4) lived with his 43-year-old wife, who reported the occupation of farm laborer; (5) had eleven children ranging in age from 4 months to 18 years, with all children 11 years and older reporting an occupation of farm laborer; and (6) was illiterate, as were all other household members. Furthermore, by using geocodes to link the individual and household records for Mr. Calloway to county-level characteristics, we know that Lowndes County's population was 82% black, and heavily agricultural with one farm for every nine residents. The average farm was 90 acres, and roughly four out of five farms were operated by a tenant farmer or a sharecropper.

**Methodology for Matching Census Records, Lynching Inventory, and Geospatial Information**

Using the information about each lynching victim that is contained in the Beck-Tolnay inventory, we will search for each victim in the decennial census prior to the year in which they were lynched. We begin by looking for a Soundex match by last name in the state and county in which the individual was lynched. All records with a matching last name are then visually inspected to determine whether the name on the census record is a reasonable approximation of the name reported as that of the individual who was lynched. Our initial criteria for a successful link also includes race, sex, and an age range that focuses on individuals who would have been between 10 and 75 years old at the time of lynching.

Because not all lynch victims would have been lynched in the county in which they resided or were enumerated, spatial distance between counties is a key criteria for identifying likely matches. If an acceptable match is not found in the county of lynching, we search each contiguous county as well. For counties bordering a state line, the spatial criterion will include counties from both states. We use historical county boundaries, newly available from *The Historical United States County Boundary Files 1790-1999*, to identify adjacent and proximate counties.

This sequential strategy for restricting the geographic scope of our search is based on the reasonable (though certainly not unchallengeable) assumption that most migrants moved short or modest distances during this historical period, and uses this logic to restrict the number of potential "matches" without excluding nearby areas due to the arbitrary nature of political boundaries. Counterbalancing these strengths are a variety of weaknesses, including that these restrictions may selectively eliminate more transient victims from being linked with their census records, and that some "true links" will be missed by the restricted geographic scope. A discussion of these and of steps to remedy some of these potential pitfalls follows. Additionally, because county-level political boundaries may have changed throughout the half-century under investigation, meaning that a small minority of lynchings occurred in counties that either did not exist in the year of census enumeration, or for which the counties encompassed by the adjacency criterion may have changed between the time of census enumeration and the date of the lynching, we use census-date boundaries for all searches (Horan and Hargis 1989).

**Potential Obstacles to Successful Linkages**

As with any attempt to link records across two unique sources, our efforts to identify lynch victims face a variety of challenges. These include:

- Census Under-Enumeration: Under-enumeration, particularly of black males, is a well-established historic fact (Coale and Rives 1973; Rosenwaike, et. al., 1998). For any given lynch victim who is not included in the census manuscripts, this undercount reduces the likelihood of a successful match to zero. Unfortunately, there is no easy remedy to this problem.

- Geographic mobility: There is no guarantee that lynchings occurred in the county in which the victim resided during census enumeration. Individuals who migrated across county lines will be more difficult to link to their census records than those who are residentially stable. However, while geographic mobility was common during the period under investigation, most moves involved exchanging residences within the same county rather than migrating between counties (Tolnay 1999).[1] We believe that the use of contiguous counties as potential locations for lynch victims will help overcome the problem of geographic mobility. Clearly, the approach we have chosen eliminates the possibility of successfully linking victims who migrated long distances between the time of census enumeration and the date on which they were lynched. However, as much of the existing long-distance migration – particularly among African Americans – during this time period likely involved moves *out of* rather than *within* the South, influence on our results should be limited.
- Destruction of 1890 census records: The original enumerator's manuscripts for the 1890 census were destroyed in a fire. This limits our ability to successfully match individuals lynched in the 1890s – the bloodiest decade in the lynching era – to their 1890 census records. However, the bulk of lynchings in the 1890s were committed in the early years of the decade, and lynching rates declined over time. By using the 1880 census records and restricting our searches to the early years of the 1890s, we substantially increase the number of potential matches with only a small decrement to our confidence in each individual match. For example, there are 372 victims in the lynching inventory who were lynched between July 1 1890 and December 31 1893.
- Name irregularities and common names: The historical records we rely on – both the census enumerators' manuscripts and the newspaper accounts used to construct the Beck-Tolnay inventory of lynch victims – were created in an era when a large proportion of adult blacks in the South were illiterate. The names of county residents and individual lynch victims were recorded not by the individual him/herself, but by a census worker or a reporter. Census workers often obtained names for all household members from one member, who may or may not have been the eventual lynch victim, a relative or even a permanent member of the household. Reporters may have received victims' names from members of the lynch mob, the local black community, or local authorities. There is no guarantee that any of these sources knew the victim's true name, or its correct spelling. The incorrect recording of a name for Person A, either in the inventory, the census, or both, will make a successful match more difficult. Additionally, in many lynchings, different newspaper reports provided more than one name – particularly variations on the first name – for each victim. For example, for a black man lynched in Campbell County, Tennessee on 26 October 1926, three different newspaper accounts agreed on his last name – Bell – but variously reported his first name as "Nip," "Rip," and "Herbert." Common names and nicknames pose a similar problem. For example, a search for Wilbur Smith, lynched in Montgomery County, Alabama in 1920, yielded no Wilbur Smiths, but 23 black men named "Will," "Willie," or "William" Smith. We discuss our strategy for estimating our level of confidence in potential matches with multiple names reported, or multiple likely "matches" identified below.

**Estimating Confidence: Single and Multiple Potential Matches**

It is difficult to be completely confident in the accuracy of a "successful" match, even when there is only one candidate. For lynch victims with more than one potential match located in the census manuscripts, three strategies for selecting the best possible match – or for winnowing the number of potential matches -- are employed. First, members of our research team will consult the original newspaper articles used to construct the lynch victim inventory for additional information (such as characteristics of the victims and/or names of family members). Second, optimal matching techniques (see, e.g., Abbott and Tsay 2000; Stovel 2001) will be used to compare the names of the potential matches with the name of the victim as recorded in the Beck-Tolnay inventory. Adopted from the field of genetics, optimal matching compares two strings of characters by identifying the minimum number of changes required to make the two strings identical – a quantitative, although not definitive, measure of the degree of similarity between all recorded names. Third,

[1] See, however, Ferrie 2004 for contradictory evidence with surprisingly high rates of mobility.

we will search for the names of all potential matches in the focal census (e.g., 1910) in the subsequent census (e.g., 1920). Not finding a potential match in the later census does not necessarily mean that the individual was the lynch victim, because they could have migrated or not been enumerated. However, if a potential match for the focal census *is found*, with reasonable certainty, in the later census, then s/he could not have been the lynch victim. For these census-to-census matches, we will have more information at our disposal, including age, state of birth, parents' state of birth, literacy, marital status, and even spouses and children's characteristics for married men.

In some cases, these expanded efforts will identify the "true" match. It is more likely, however, that it will reduce the number of potential matches. When it is possible to reduce the number of potential matches for each name recorded in the lynching inventory to five or fewer, we will record information for all possible matches and assign a score that represents our best (albeit subjective) estimate that each potential match is the lynch victim. The sum of the probabilities for each name reported for any lynch victim will be 1. Subjective probabilities will be independently assigned by two members of our research team, and discrepancies discussed and resolved during team meetings. When there is no reason to believe that one individual is more likely than another to be the "true" match, then all potential matches will receive the same estimated probability – for example, 0.5 in a case with two possible matches.

When it is impossible to narrow the number of potential matches for each recorded name to five, then the case is recorded as unsuccessfully linked. For those cases, any additional information obtained from the original newspaper accounts of the lynching will be recorded and used as supplementary evidence to assess whether the successfully linked cases differ systematically from the unsuccessful cases.

**Pilot Testing: Preliminary Evidence**

Our preliminary work has focused on linking black males who were lynched in Alabama, Georgia, and Tennessee between 1882 and 1890 and between 1920 and 1930 with their 1880 or 1920 census records (See Table 2). In roughly 40% of the cases, a single individual was identified in the county in which the lynching occurred or in a spatially contiguous county; in roughly one in three cases, multiple likely matches were identified; in one-quarter of the cases, no match was found. We are encouraged by the results from these efforts, and feel that they are in line with "success" rates from prior attempts to link individuals to their census records (Ferrie 1996; Guest 1987; Rosenwaike, et. al. 1998; Steckel 1988). These results suggest that the methodology we have developed to link individual and contextual information about known lynch victims will result in a resource sufficiently large to support both statistical and more qualitative studies of southern lynching. They also suggest that black male lynch victims may differ in meaningful ways from a randomly-selected sample of black men. In both 1880 and 1920, lynch victims appear to be statistically more likely than members of the random sample to have been literate, and in 1920, they are also significantly more likely to have been married at the time of the census enumeration. We have not yet assessed whether spatial variation in these differences exists.

**Conclusions**

This project's innovations include the use of newly-available historical spatial resources to facilitate Census linkages, vastly expanded information on the individual factors correlated with risk of lynching, and the ability to link community contextual information with characteristics of individual lynch victims. As with any historical data collection process, the resulting product will be limited by both the completeness and quality of the data to which we have access, and by the temporal separation between researchers and the phenomenon under examination. However, we believe that this new data source will facilitate a new arena of research on "the tragedy of lynching," and provide unique insights into racially-motivated hate crimes in our nation's past. Specifically, it will introduce the victim and his/her socioeconomic, demographic, family structural, and community's contextual characteristics into the study of lynching.

**BIBLIOGRAPHY**


Abbott, Andrew and Angela Tsay. 2000. "Sequence analysis and optimal matching methods in Sociology: Review and prospect." *Sociological Methods & Research* 29:3-33.

Ames, Jesse Daniel. 1942. *The Changing Character of Lynching: Review of Lynching, 1931-1941, With a Discussion of Recent Developments in This Field*. Atlanta: Commission on Interracial Cooperation.

Beck, E. M., James L. Massey and Stewart E. Tolnay. 1989. "The gallows, the mob, the vote: Lethal sanctioning of Blacks in North Carolina and Georgia, 1882 to 1930." *Law & Society Review*, 23:317-331.

Beck, E. M. and Stewart E. Tolnay. 1990. "The killing fields of the Deep South: The market for cotton and the lynching of Blacks, 1882-1930." *American Sociological Review*, 55:526-539.

Brundage, W. Fitzhugh. 1993. *Lynching in the New South: Georgia and Virginia, 1880-1930.* Urbana: University of Illinois Press.

Coale, Ansley J. And Norfleet W. Rives Jr. 1973. "A statistical reconstruction of the Black population of the United States1880-1970: Estimates of true numbers by age and sex, birth rates, and total fertility." *Population Index* 39:3-36.

Corzine, Jay, James Creech, and Lin Corzine. 1983. "Black concentration and lynchings in the South: Testing Blalock's power-threat hypothesis." *Social Forces* 61:774-96.

Corzine, Jay, Lin Huff-Corzine, and James Creech. 1988. "The tenant labor market and lynching in the South: A test of the split labor market theory." *Sociological Quarterly* 58:261-78.

Dinnerstein, Leonard. 1968. *The Leo Frank Case*. New York: Columbia University Press.

Downey, Dennis B. and Raymond M. Hyser. 1991. *No Crooked Death: Coatesville, Pennsylvania, and the Lynching of Zachariah Walker*. Urbana: University of Illinois Press.

Ferrie, Joseph P. 1996. "A new sample of males linked from the public use microdata sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census manuscript schedules." *Historical Methods* 29:141-156.

Griffin, Larry J. 1993. "Narrative, event-structure analysis, and causal interpretation in historical Sociology." *American Journal of Sociology* 98:1094-1133.

Guest, Avery M. 1987. "Notes from the National Panel Study: Linkage and migration in the late nineteenth century." *Historical Methods* 20:63-77.

Horan, Patrick M. and Peggy G. Hargis. 1989. "The County Longitudinal Template." Paper presented at the Annual Meeting of the Social Science History Association, Chicago.

Olzak, Susan. 1990. "The political context of competition: Lynching and urban racial violence." *Social Forces* 69:395-421.

Raper, Arthur. 1933. *The Tragedy of Lynching*. Chapel Hill: University of North Carolina Press.

Rosenwaike, Mark E., Samuel H. Preston, and Irma T. Elo. 1998. "Linking death certificates to early census records: The African American matched records sample." *Historical Methods* 31:65-74.

Smead, Howard. 1986. *Blood Justice: The Lynching of Charles Mack Parker*. New York: Oxford University Press.

Soule, Sarah A. 1992. "Populism and black lynching in Georgia, 1890-1900." *Social Forces* 71:431-49.

Steckel, Richard H. 1988. "Census matching and migration: A research strategy." *Historical Methods* 21:52-60.

Stovel, Katherine. 2001. "Local sequential patterns: The structure of lynching in the Deep South, 1882-1930." *Social Forces* 79:843-880.

Tolnay, Stewart E. 1999. *The Bottom Rung: African American Family Life on Southern Farms*. Urbana: University of Illinois Press.

Tolnay, Stewart E. and E. M. Beck. 1992a. "Racial violence and Black migration in the American South, 1910 to 1930." *American Sociological Review*, 57:103-116.

Tolnay, Stewart E. and E. M. Beck. 1995. *A Festival of Violence: An Analysis of Southern Lynchings, 1882-1930*. Urbana: University of Illinois Press.

Tolnay, Stewart E., E. M. Beck and James L. Massey. 1989. "Black lynchings: The power threat hypothesis revisited." *Social Forces*, 67:605-623.

Tolnay, Stewart E., Glenn Deane and E.M. Beck. 1996. "Vicarious violence: Spatial effects on southern lynchings, 1890-1919." *American Journal of Sociology*, 102:788-815.
Wells-Barnett, Ida B. 1892[1969]. *Southern Horrors: Lynch Law in All Its Phases*. New York: Arno Press.
Wexler, Laura. 2003. *Fire in a Canebreak: The Last Mass Lynching in America*. New York: Scribner.
White, Walter. 1929[1969]. *Rope and Faggot: A Biography of Judge Lynch*. New York: Arno Press.

**Table 1. Number of victims in Beck-Tolnay Inventory, 1882 – 1930, by state, decade, & victim's race [a]**

| | 1882-89 | | 1890-99 | | 1900-09 | | 1910-19 | | 1920-30 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Black* | *White* | *Black* | *White* | *Black* | *White* | *Black* | *White* | *Black* | *White* |
| Alabama | 37 | 4 | 112 | 15 | 63 | 0 | 52 | 1 | 9 | 4 |
| Arkansas | 16 | 20 | 65 | 16 | 41 | 6 | 35 | 4 | 21 | 1 |
| Florida | 18 | 4 | 62 | 3 | 50 | 4 | 49 | 4 | 43 | 3 |
| Georgia | 54 | 4 | 117 | 9 | 99 | 2 | 126 | 3 | 43 | 2 |
| Kentucky | 30 | 19 | 47 | 18 | 35 | 3 | 12 | 2 | 4 | 1 |
| Louisiana | 45 | 9 | 126 | 32 | 83 | 7 | 50 | 1 | 10 | 4 |
| Mississippi | 106 | 11 | 137 | 5 | 140 | 3 | 66 | 1 | 60 | 2 |
| North Carolina | 27 | 6 | 13 | 6 | 16 | 2 | 14 | 0 | 13 | 1 |
| South Carolina | 28 | 2 | 53 | 3 | 33 | 1 | 21 | 0 | 13 | 0 |
| Tennessee | 32 | 13 | 73 | 16 | 42 | 5 | 22 | 0 | 6 | 3 |
| **Total** | **393** | **92** | **805** | **123** | **602** | **33** | **447** | **16** | **222** | **21** |

[a] Victims whose race was unknown are excluded from the table.

**Table 2. Preliminary results, comparing characteristics of lynch victims with samples of non-victims for Alabama, Georgia, and Tennessee in 1880 and 1920.**

| | 1880 | | 1920 | |
|---|---|---|---|---|
| | Comparison Sample | Lynch Victims | Comparison Sample | Lynch Victims |
| N of Cases | 996 | 52 | 741 | 30 |
| Mean Age | 33.0 | 32.6 | 32.7 | 28.9 |
| % Can read and write | 19.4 | 30.8* | 62.5 | 87.5** |
| % Mulatto | 10.6 | 11.5 | 10.3 | 9.4 |
| % Head of Household | 65.8 | 73.1 | 57.5 | 62.5 |
| % Married | 61.4 | 71.4 | 55.6 | 76.7* |
| % Owns Home | NA | NA | 17.5 | 10.5 |

* $p < .05$; ** $p < .01$ (two-tailed tests)