pagepress

Geospatial Health 2022; volume 17:1020

Geocoding precision of birth records from 2008 to 2017 in Kentucky, USA

Courtney J. Walker,[1] Steven R. Browning,[1] Jeffrey E. Levy,[2] W. Jay Christian[1]

[1]Department of Epidemiology, University of Kentucky, College of Public Health; [2]Department of Geography, Lexington, KY, USA

Abstract

Maternal address information captured on birth records is increasingly used to estimate residential environmental exposures during pregnancy. However, there has been limited assessment of the geocoding precision of birth records, particularly since the adoption of the 2003 standard birth certificate in 2015. To address this gap, this study evaluated the geocoding precision of live and stillbirth records of Kentucky residents over ten years, from 2008 through 2017. This study summarized the demographic characteristics of imprecisely geocoded records and, using a bivariate logistic regression, identified covariates associated with poor geocoding precision among three population density designations-metro, non-metro, and rural. We found that in metro areas, after adjusting for area deprivation, education, and the race, age and education of both parents, records for Black mothers had 48% lower odds of imprecise geocoding (aOR=0.52, 95% CI: 0.48, 0.56), while Black women in rural areas had 96% higher odds of imprecise geocoding (aOr=1.96, 95% CI: 1.68, 2.28). This study also found that over the study period, rural and non-metro areas began with a high proportion of imprecisely geocoded records (38% in rural areas, 19% in non-metro), but both experienced an 8% decline in imprecisely geocoded records over the study period (aOr=0.92, 95% CI: 0.92, 0.94). This study shows that, while geocoding precision has improved in Kentucky, further work is needed to improve geocoding in rural areas and address racial and ethnic disparities.

Correspondence: Courtney Walker, Department of Epidemiology, University of Kentucky, College of Public Health, 111 Washington Ave, Suite 209B, Lexington, KY 40536-0003, USA.
Tel.:+1.606.304.1878. E-mail: courtney.walker@uky.edu

Key words: Geocode; Kentucky; USA; birth; logistic regression; rural-urban continuum codes.

Acknowledgements: the authors would like to thank Dr. John Barton for his feedback on this manuscript and the Kentucky Office of Vital Statistics and the Central Appalachian Regional Education and Research Center for their support of this project.

Funding sources: this manuscript was supported in part by the Education Research Center (ERC) through Grant 6U54OH007547 and by Berea College through the Olive Ruth Russel Fellowship. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIOSH/CDC or Berea College.

Ethical and other information: the Medical Institutional Review Board at the University of Kentucky and the Kentucky Cabinet for Health and Family Services (CHFS) Institutional Review Board approved this protocol. This study is part of a larger research project to examine environmental risk factors associated with gestational hypertension.

Received for publication: 18 May 2021.
Revision received: 14 October 2021.
Accepted for publication: 14 October 2021.

Geospatial Health 2022; 17:1020
doi:10.4081/gh.2022.1020

Introduction

To assess retrospective environmental exposures during pregnancy, health researchers are increasingly employing US birth records as a source of residential information. However, studies evaluating birth record data quality predominantly focus on health information - the quality of address data are less frequently discussed in applied health research, although many of these studies employ geocoded address information to assess geospatial patterns of disease or evaluate residential environmental exposures (Cromley, 2019; Goldberg, 2008; Krieger Nancy *et al.*, 2002; Zandbergen, 2007). Understanding the quality of residential address information, specifically the precision of geocoded addresses, is vitally important, as imprecise geocoding can result in records being assigned to an incorrect geographic location-leading to biased results and erroneous conclusions (Krieger Nancy *et al.*, 2002; Zandbergen, 2007). Of further concern is the evidence suggesting maternal demographic characteristics, notably race and ethnicity, are associated with the geocoding precision (Ha *et al.*, 2016).

Geocoding assigns geographic coordinates to an address, using components such as address number, street name, city, state, and/or ZIP code (Buttling *et al.,* 2021; Jacquez 2012; Krieger *et al.,* 2001). Input addresses are compared to a reference dataset that contains street segments with address ranges and verified coordinates. Coordinates from the reference dataset are matched to the corresponding record of the input dataset (Goldberg, 2008). These matches may be as precise as the 'rooftop' of the actual residential structure or within a small range, *e.g.*, a location along a street segment based on the address range of that segment. Addresses with missing components, such as the street address or apartment number, or containing errors that prevent a sufficiently probable address match, are assigned less precise coordinates that correspond to the mid-point (centroid) of the spatial resolution to which they were matched. Issues that can impede geocoding include spelling errors, use of special characters, neglected or incorrect suffixes (drive, lane, *etc.*), incorrect ZIP codes, inaccurate or absent apartment numbers or complex names. Rural areas can pose additional challenges as rural or hired contractor routes have limited coordinate data. However, this may be less of an issue in future geocoding projects as the precision of emergency 911 telephone (E911) databases continue to expand and the increasingly common practice to assign specific addresses (Mask, 2014). Overall, geocoding 85% or more addresses to either a rooftop or a street segment is considered a benchmark for high quality (Andresen *et al.,* 2020).

Many studies that employ geocoding to assign exposure status do not often report the proportion of records that geocoded imprecisely or the methods used to address imprecision and limit potential bias (Cromley, 2019; Ha *et al.,* 2016; Schinasi *et al.,* 2018). There have also been limited analyses of maternal characteristics that could be associated with poor geocoding precision-such as race, ethnicity, and education-which are factors that are also associated with disease status and severity. This possibility of cartographic confounding association between disease status and exposure can bias spatial analyses as imprecise records may be assigned incorrect geographic locations resulting in systematic exposure misclassification if one group has a higher probability of imprecisely geocoding.

To address existing gaps in the literature, this study sought to describe the demographic characteristics of imprecisely geocoded records, identify covariates associated with imprecise geocoding, and explore regional variations in geocoding precision by examining maternal residential data from Kentucky birth records from 2008 through 2017. We believe there may be a difference among metropolitan status regions, demographic characteristics, and between the Appalachian and non-Appalachian regions.

Materials and methods

Data source

Kentucky Vital Statistics provided individual records of all live and stillbirths of Kentucky residents from January 1, 2008, through December 31, 2017. These records contained maternal addresses; maternal and paternal demographic information (age, race, ethnicity and education), self-reported marital status, maternal health information (height and pre-pregnancy weight); previous pregnancy history (parity), and characteristics of current pregnancy (birthplace, prenatal care, and plurality). Although live and stillbirth forms differ slightly, all variables used in this study were captured on both certificates. Non-singleton birth records were excluded to reduce the number of duplicate addresses (n=18,628), leaving 538,117 records for analysis.

Address-matched geocoding

We geocoded the maternal addresses using the ESRI Address Coder 10.7.1 (ESRI, Redlands, CA, USA). Address match score, or the level of agreement needed to match an address, was set at the program default of 93%. Geocoding outcomes were classified as precise (address point/street address range) or less precise (street name, ZIP code, city, or unknown). Addresses matched to a street name, city, or ZIP code were reviewed for spelling errors, inappropriate characters (for example, “0’Mally” rather than “O’Mally”), or neglected or incorrect designations (*e.g.*, road *vs* boulevard) (N=30,879). After corrections and re-geocoding, 14.5% (N=4608) of addresses that had initially been imprecisely geocoded were geocoded to a point or street segment.

Covariates

Maternal age in years was grouped into quartiles (<23, 23-25, 26-30, >30, and unknown), as was paternal age (<24, 25-28, 29-33, >33, and unknown). Due to small cell counts, race was classified into four categories (Black, White, Other, and unknown). Maternal body mass index (BMI) was calculated using height and pre-pregnancy weight recorded on the birth record. Using standard classifications, BMI was then grouped into four categories: underweight (<18.5 kg/m^2), normal (18.5-24.9 kg/m^2), overweight (25.0-29.9 kg/m^2), and obese (>30.0 kg/m^2). Appalachian designation, defined by the Appalachian Regional Commission (2020) (ARC), was assigned based on the geocoded county of residence. Rural-metro continuum codes (RUCC) from the United States Department of Agriculture were classified into three categories: rural (7-9), non-metro (4-6), and metro (1-3). (Alexander, 2011; United States Department of Agriculture, 2021) Other covariates considered in the model included ethnicity (Hispanic, non-Hispanic, and unknown), education (less than high school, high school, some college to above, and unknown); marital status (yes/no, or not stated); child birthplace (hospital, home birth, or other); and parity (0; 1; and 2+ previous births).

To classify prenatal care, we used the revised graduated prenatal care utilization index (R-GINDEX). Based on ACOG recommendations for the frequency of prenatal visits, this index assigns a score based on the percent of prenatal visits attended adjusting for the date of the first prenatal visit (obtained from the birth record) and gestational age at delivery (Kotelchuck, 1994). To adjust for local (*i.e.*, micro block-group level) socioeconomic disadvantage, the area deprivation index (ADI) was used (Health, 2021; Kind and Buckingham, 2018). The ADI values, which are scaled from 1 to 10, were collapsed into deprivation quartiles: lowest (1-3), low (4-5), mid-range (ADI 6-7), and highest (8-10). Out of the 3285 micro-block groups in Kentucky, 81 did not have an assigned ADI score due to low population counts or high group population quarters (such as a large apartment complex).

Analysis

Demographic characteristics were summarized with counts and row percentages by RUCC (rural; non-metro, and metro). A Chi-Square test was used to assess the distribution of frequencies of demographic covariates in geocoding precision. We used a multivariate binary logistic regression as the final model and a backward elimination method to identify covariates for the model. Our final model included maternal age, race, BMI, marital status, father’s age, race, education, ethnicity; infant birthplace parity, adequacy of gestational care, year of birth, ADI, and the Appalachian region. All non-spatial analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). To visualize the temporal trends, we created a line plot of the percentage of addresses imprecisely geocoded by rural-urban classification.

Adjusted odds ratio (aOR) with confidence intervals (CI) were calculated as measures of association.

Results

In this cross-sectional study, we found that 59% of birth records (n=317,279) geocoded to an address point and 31%

(n=166,752) geocoded to a street segment, yielding a precise geocoding rate of 90% (Figure 1).

Figure 2 displays the proportion of records that geocoded imprecisely by RUCC over the study period. Rural areas experienced the sharpest decline in imprecision, whereas metro areas had consistent geocoding imprecision over the study period.

Table 1 presents the number of records among each subgroup and the percent of records that geocoded imprecisely among all RUCCs and the entire sample. Overall, the population of Kentucky births occurred in women predominantly between the ages of 26-30 years old (n=154,789), who were non-Hispanic (n=509,425), White (n=450,425), had some college (n=293,878), and were married (n=311,817).

In all RUCCs, the prevalence of imprecise geocoding was highest among non-Hispanic, unmarried women with less than a high school degree (Table 1). As educational attainment increased, the proportion of addresses that geocoded imprecisely decreased across all RUCCs. Among records with known maternal and paternal age, mothers ≥23 years old and fathers who were ≤24 years old had the highest proportion of imprecise records across all RUCCs as well. However, among RUCCs, the proportion of imprecisely geocoded records varied substantially among maternal and paternal racial groups. In metro areas, White mothers had a higher proportion of records that were not geocoded (5%). However, in non-metro and rural areas, a higher proportion of Black mothers did not geocode precisely (non-metro: 16%, rural: 39%), which was similar to proportions observed within father's race. In metro areas, White fathers had the highest proportion of imprecise records (5%), while in non-metro and rural areas, we observed similar proportions as seen with Black mothers, *i.e*., –14% and 34%, respectively. Records with unknown characteristics, such as maternal age, paternal age, race, and ethnicity, did not have consistent geocoding precision patterns across RUCCs, except unknown paternal ethnicity, which had the highest proportion of imprecisely geocoded records across all RUCCs. In non-metro areas, mothers with unknown maternal and paternal age and maternal BMI had the highest prevalence of geocoding imprecision. In contrast, records with fathers of unknown age and education in rural areas had the highest proportion for their respective subgroups. Areas with low economic deprivation had higher geocoding imprecision in metro (6%) and non-metro (24%) areas. However, in rural areas, those without an assigned ADI had the highest proportion of imprecision (33%).

Figure 3 illustrates county-level geocoding imprecision across the state. For additional context, Figure 4 displays the RUCC for each county and identifies larger cities (>50,000 inhabitants) and the state capital, Frankfort. The Appalachian region had multiple

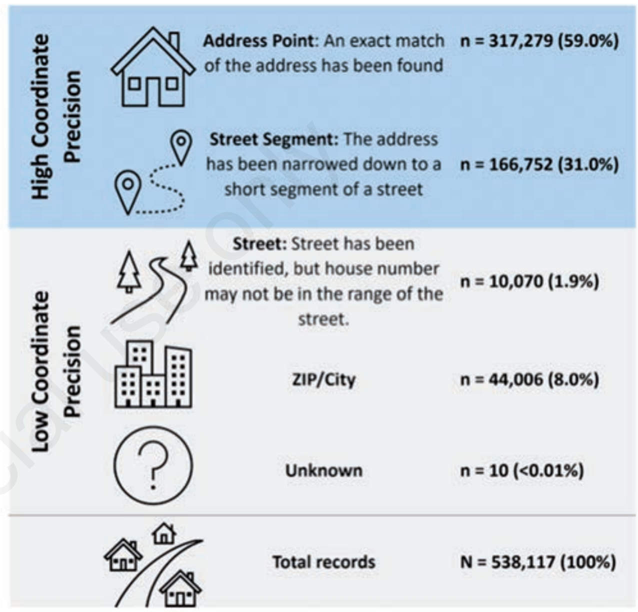

Figure 1. Description and summary of geocoding precision for Kentucky birth records, 2008-2017.

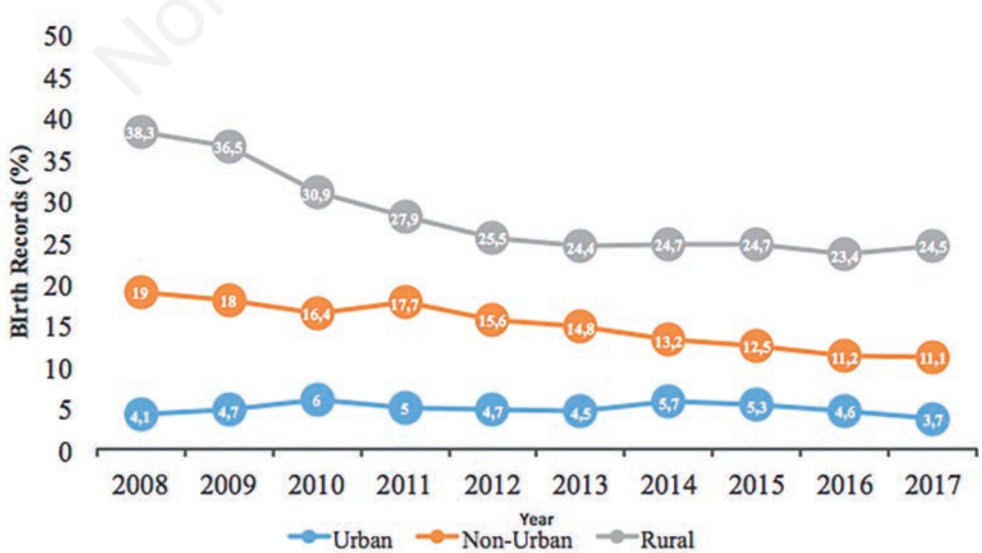

Figure 2. Prevalence of imprecise geocoding of Kentucky birth records by RUCC, 2008-2017.

counties characterized by high rates of imprecise geocoding. Counties with a high population density and their contiguous neighbours, on the other hand, had much higher percentages of addresses that geocoded precisely.

The multivariable logistic regression, presented in Table 2, displays notable differences in the odds of imprecise geocoding among RUCCs. In metro areas, Black mothers had a 48% lower odds of imprecise geocoding, adjusting for other covariates (aOr=0.48, 95% CI: 0.48, 0.56); however, in rural areas, Black women had 96% higher odds of imprecise geocoding (aOr=1.96, 95% CI: 1.68, 2.28). Maternal ethnicity also varied by RUCC, with non-Hispanic mothers in metro areas having 74% higher odds of imprecise geocoding (aOr=1.74, 95% CI: 1.58, 1.94), 20% higher in non-metro areas (aOr=1.20, 95% CI: 1.06, 1.38), but 14% lower in rural areas, although this last difference was not statistically significant (aOr=0.86, 95% CI: 0.70, 1.04). Patterns within father's race and ethnicity were slightly different, with Black fathers in metro areas having lower odds of imprecision (aOr=0.72, 95% CI: 0.66, 0.78); however, Black fathers in non-metro areas had higher, although non-significant, odds of imprecise geocoding, compared to White fathers (aOr=1.04, 95% CI: 0.94, 1.12). In rural areas, like maternal race, the odds of imprecision were higher among Black fathers than White (aOr=1.32, 95% CI: 1.14, 1.54).

As maternal age increased, the odds of imprecise geocoding decreased in metro areas, with those between the ages of 31-41 years experiencing the lowest odds (aOr=0.84, 95% CI: 0.78, 0.90). However, in non-metro and rural areas, as maternal age increased, the odds of imprecision increased, with 10% higher odds of imprecision among those aged 31-41 years in non-metro areas (aOr=1.10, 95% CI: 1.04, 1.16) and 12% in rural areas, although non-significant (aOr=1.12, 95% CI: 0.90, 1.42). Paternal age also significantly impacted geocoding precision. In non-metro areas, fathers age ≥33 years had a significant 6% increase in the odds of imprecise geocoding (aOr=1.06, 95% CI: 1.00, 1.12). In

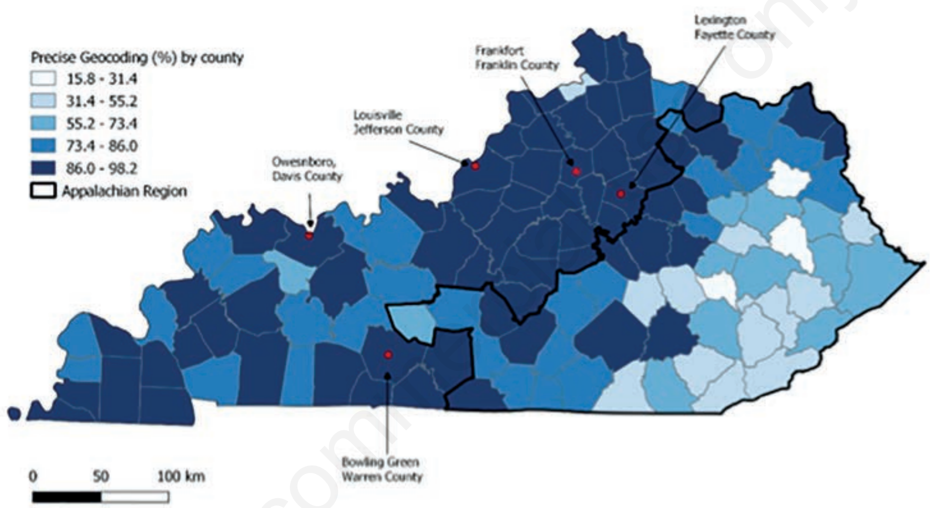

Figure 3. Percent of Kentucky birth record addresses that geocoded to an address point or street segment by county, 2008-2017.

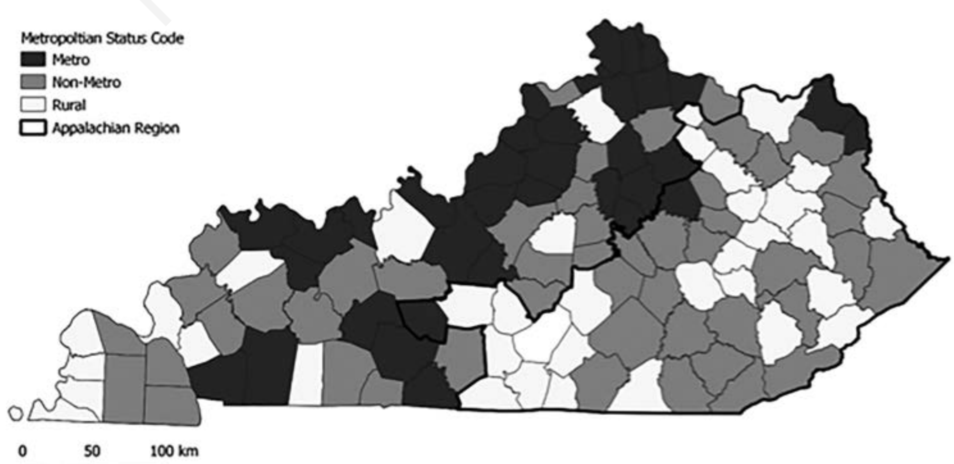

Figure 4. Kentucky rural-urban classification areas (RUCC), 2010.

non-metro (aOr=1.24, 95% CI: 1.12, 1.34) and rural (aOr=1.26, 95% CI: 1.10, 1.42) areas, unknown age was associated with a significantly increased odd of imprecise geocoding among fathers, compared to fathers ≤24 years.

Across all RUCCS, higher maternal education was associated with lower odds of imprecision compared to those with no high school degree. Fathers' education was protective against imprecision in metro and non-metro areas.

Compared to mothers who were married, unmarried mothers had 30% lower odds of imprecise geocoding (aOr=0.70, 95% CI: 0.68, 0.74), although, in rural areas, those who were married had 10% higher odds (aOr=1.10, 95% CI: 1.06, 1.16). Inadequate care was associated with higher odds of geocoding imprecision in all RUCCs, compared to adequate care started in the first trimester. Although non-significant in rural areas, in metro areas, intensive care also had substantially higher odds of imprecision (aOr 1.26, 95% CI: 1.18, 1.34). The impact of socioeconomic disadvantage, as measured by ADI, is also noteworthy concerning geocoding precision. Those in areas without ADI scores had the highest odds of imprecise geocoding in metro (aOr=12.22, 95% CI: 11.22, 13.32) and non-metro areas (aOr=8.5, 95% CI: 7.68, 9.40).

Over the study period, the odds of imprecise geocoding remained unchanged in metro areas. However, in non-metro areas, the odds declined by 8% (OR=0.92, 95% CI: 0.92, 0.94) and by 8% in rural areas (OR: 0.92, 95% CI: 0.92, 0.94).

Discussion

Racial and ethnic disparities have been found in other studies assessing geocoding precision in birth records. In a study using birth data from Florida, Ha and colleagues (2016) reported that Black women had higher odds of not geocoding than White women. Gilboa and colleagues (2006) reported a higher proportion of Latinas in non-geocoded records in an assessment of the precision of Texas birth records and birth defects registry data (Gilboa *et al.,* 2006). In this study, although almost all records were geocoded, we found both parents' race and ethnicity influenced the odds of precise geocoding after adjusting for other relevant factors, such as RUCC classification. Although records of Black women and men were more likely to geocode precisely in metro areas, this may be due to greater resources for developing GIS-related infrastructure and data layers in more populous municipalities that also tend to have a more diverse population, as well as a legacy of environmental injustice and 'redlining' (Ash and Fetter, 2004).

Rural residency has been cited as an important factor in the inability to geocode in Florida and Virginia (Ha *et al.,* 2016; Jones *et al.,* 2014; Oliver *et al.,* 2005). However, certain conventions in rural areas have shifted, such as introduction of E911 system that allow for more accurate positioning of addresses (The New York Times, 1996). This study used the national address coder provided through the software manufacturer, but future work exploring other sources, including historical records, for geospatial coordinates is needed.

This study also found that educational and economic factors were associated with geocoding imprecision. Both maternal and paternal education was associated with decreased odds of imprecision, compared to the respective parent not completing high school. The protective effect of education, which often serves as a proxy for income, may reflect financial means to choose more attentive hospitals or reflect data collection form literacy; however, further work is needed. We also found that addresses geocoded in areas without an ADI score (whether due to low population count or high proportion of multiple family housing) had a much higher odds of imprecise geocoding than areas with low economic deprivation. Across all RUCCs, however, an increase in economic deprivation was associated with increased odds of imprecise geocoding. As our geocoder contained only one year of street information, it is possible that older addresses, particularly for those who live in lower income areas, no longer exist due to redevelopment, and for that reason, records would be unmatchable when compared to our existing database.

Cartographic confounding, or the association demographic factors (such as race and rurality), with both the outcome of interest and geocoding precision, is an under-discussed but pressing concern in studies of health geography (Oliver *et al.,* 2005). In an assessment of research employing geographic information systems (GIS) methodologies Cromley (2019) noted that 25% of articles assessing health literature did not describe their techniques. In a meta-analysis of electronic health records using GIS methods, Shinasi and colleagues (Schinasi *et al.,* 2018) noted that less than one-third of journal articles described their geocoding process. This lack of transparency in disclosing geocoding methodologies leaves other health researchers unable to evaluate a study's merits and limitations fully.

Limitations of this study include noted variations among hospitals in vital statistics collection and reporting. We were also unable to assess the records for accuracy or identify addresses changes during the pregnancy or preconception period, as only current residential information is captured on the birth record. As we used the currently available geocoding reference file through ESRI, older records may contain addresses that no longer exist, due to redevelopment or other reasons, and would not have been able to be precisely identified. However, a noted strength of this study was the inclusion of all singleton birth records-still and live birth-in Kentucky over ten years. Further, we were able to assess maternal and paternal demographic factors, economic factors, and regional differences among all Kentucky births.

Future work may consider exploring hospital-based interventions-such as verifying the accuracy of a larger pool of randomly selected records or offering additional support to women who report having less than a high school education or do not report paternal information on the birth record. Both of these factors were associated with an increased odds of imprecision in rural areas. However, using state specific geocoders and employ historical records to improve geocoding of older records, may address many of the issues we found with older records. Moving forward, however, it is important to explore other solutions to improving geocoding for all races and ethnicities, particularly in rural regions.

Conclusions

To more adequately characterize potential limitations for studies that employ geocoding to assess disease prevalence and exposure status, further review is needed. This study used Kentucky birth records from 2008-2017 to characterize the prevalence of geocoding imprecision among RUCCs, identify characteristics associated with imprecise geocoding, and assess the geospatial distribution of geocoding imprecision in the state. We found that the proportion of addresses that geocoded imprecisely in rural and non-metro areas declined over the study period. We also found maternal and paternal race, age, ethnicity, and education, along with marital status, prenatal care, and ADI were significantly asso-

Table 1. Total population count and percent of Kentucky birth records imprecisely geocoded among rural-urban classifications, 2008-2017.

	Metro N (%)°	Non-metro N (%)°	Rural N (%)°	Total N (%)°
Mother's age (years)	*	*	*	*
<23	75,540 (6.2)	54,255 (16.7)	15,590 (30.0)	145,385 (12.7)
23-25	54,212 (5.4)	32,763 (15.0)	9194 (28.0)	96,169 (10.8)
26-30	97,265 (4.4)	45,658 (13.8)	11,866 (27.4)	154,789 (8.9)
31-41	92,288 (3.9)	34,664 (14.1)	8712 (27.3)	135,664 (8.0)
Unknown	2985 (5.1)	1381 (17.0)	473 (29.8)	4839 (10.9)
Mother's race	*	*	*	*
White	250,694 (5.3)	156,255 (15.2)	43,476 (28.1)	450,425 (10.9)
Other	30,619 (3.3)	6785 (11.4)	1283 (31.2)	38,687 (5.7)
Black	40,977 (3.3)	5681 (15.8)	1076 (39.4)	47,734 (5.6)
Mother's ethnicity	*	*	*	*
Hispanic	21,743 (3.0)	4668 (11.1)	1010 (32.5)	27421 (5.4)
Not Hispanic	300,547 (5.0)	164,053 (15.1)	44,825 (28.3)	509,425 (10.3)
Mother's education	*	*	*	*
Less than HS	47,345 (6.6)	30,443 (20.0)	10,530 (31.5)	88,318 (14.2)
High School	77,265 (6.1)	54,422 (15.5)	15,463 (27.8)	147,150 (11.9)
Some College	193,311 (3.9)	81,418 (12.9)	19,149 (27.2)	293,878 (7.9)
Unknown	4369 (5.0)	2438 (14.2)	693 (27.8)	7500 (10.1)
Mother's BMI (kg/m^2)	*	*	*	*
Underweight (<18)	13,293 (6.0)	8387 (18.1)	2459 (32.4)	24,139 (12.9)
Normal (18-24.9)	137,596 (4.5)	65,378 (14.3)	17,466 (27.6)	220,440 (9.2)
Overweight (25-29.9)	79,540 (4.8)	41,048 (14.6)	11,170 (28.4)	131,758 (9.9)
Obese (>30)	79,093 (5.4)	48,884 (15.5)	13,337 (28.5)	141,314 (11.1)
Unknown	12,768 (4.7)	5024 (18.1)	1403 (30.9)	19,195 (10.1)
Marital status	*	*	*	*
Yes	187,167 (4.6)	97,443 (14.1)	27,207 (26.4)	311,817 (9.5)
No or not stated	135,123 (5.1)	71,278 (16.3)	18,628 (31.4)	225,029 (10.8)
Father's age	*	*	*	*
<24	48,649 (6.4)	37,215 (15.7)	10,646 (28.4)	96,510 (12.4)
25-28	57,280 (5.0)	33,017 (13.8)	8998 (26.1)	99,295 (9.8)
29-33	78,730 (3.9)	35,148 (12.8)	9033 (25.4)	122,911 (8.0)
>33	73,810 (4.0)	31,462 (14.5)	8420 (27.4)	113,692 (8.7)
Unknown	63,821 (5.7)	31,879 (18.6)	8738 (34.9)	104,438 (12.1)
Father's race	*	*	*	*
White	206,370 (5.1)	127,882 (14.8)	36,132 (27.2)	370,384 (10.6)
Unknown	1738 (4.4)	1024 (14.4)	266 (20.3)	3028 (9.2)
Other	80,693 (4.8)	33,323 (16.1)	8292 (33.2)	122,308 (9.8)
Black	33,489 (3.3)	6492 (13.6)	1145 (34.1)	41,126 (5.8)
Father's ethnicity	*	*	*	*
Hispanic	19,136 (3.1)	4839 (11.0)	1106 (31.2)	25,081 (5.9)
Non-Hispanic	239,191 (4.7)	132,001 (14.5)	36,138 (27.1)	407,330 (9.9)
Unknown	63,963 (5.8)	31,881 (17.9)	8591 (33.7)	104,435 (11.8)
Father's education	*	*	*	*
Less than HS	32,285 (6.7)	24,573 (18.5)	9330 (28.3)	66,188 (14.1)
High School	83,242 (5.9)	60,518 (14.9)	17,306 (26.9)	161,066 (11.5)
Some College	140,488 (3.4)	50,757 (11.9)	10,387 (26.9)	201,632 (6.7)
Unknown	66,275 (5.6)	32,873 (17.6)	8812 (33.4)	107,960 (11.5)
Birthplace	*	*	*	*
Hospital	318,892 (4.8)	166,600 (15.1)	44,421 (29.1)	529,913 (10.1)
Home	2947 (6.4)	1817 (9.4)	1241 (6.3)	6005 (7.3)
Other	451 (11.5)	304 (15.5)	173 (6.4)	928 (11.9)
R-GINDEX	*	*	*	*
Adequate (80%-109%)	142,625 (4.4)	80,499 (13.6)	20,110 (27.9)	243,234 (9.4)
Inadequate (<50%)	21,281 (7.2)	13,382 (18.8)	4549 (27.9)	39,212 (13.6)
Intermediate (50-79%)	114,774 (4.8)	48,528 (16.4)	14,988 (28.7)	178,290 (10.0)
Intensive (>110%)	23,579 (5.6)	18,811 (13.2)	4042 (26.9)	46,432 (10.5)
Missing/No care	20,031 (4.8)	7501 (19.2)	2146 (35.6)	29,678 (10.6)
Parity	*	*	*	*
0	130,713 (4.6)	68,068 (15.1)	17,662 (28.9)	216,443 (9.9)
1	102,585 (4.7)	54,893 (14.5)	14,497 (28.4)	171,975 (9.9)
2 +	88,992 (5.3)	45,760 (15.6)	13,676 (27.7)	148,428 (10.5)
ADI	*	*	*	*
Lowest (1-3)	143,561 (3.2)	25,357 (4.6)	1189 (20.9)	170,107 (3.6)
Mid-low (4-5)	65,378 (5.0)	36,980 (6.2)	4735 (20.1)	107,093 (6.1)
Mid-high (6-7)	59,047 (6.4)	41,800 (14.1)	11,865 (20.0)	112,712 (10.7)
Highest (8-10)	51,115 (5.9)	61,757 (24.4)	27,098 (33.3)	139,970 (19.4)
Missing ADI	3189 (27.8)	2827 (33.7)	948 (45.6)	6964 (32.6)
Stillbirths			*	
No or not stated	320,552 (4.8)	167,697 (15.0)	45,569 (28.5)	533,818 (10.1)
Yes	1738 (4.4)	1024 (14.4)	266 (20.3)	3028 (9.2)
Appalachian	*	*	*	*
Not Appalachia	307,406 (4.7)	79,709 (6.6)	12,658 (15.2)	399,773 (5.4)
Appalachia	14,884 (8.2)	89,012 (22.6)	33,177 (33.5)	137,073 (23.6)
Total	322,290 (4.8)	168,721 (15.0)	45,835 (28.4)	536,846 (10.1)

BMI, body mass index; R-GINDEX, Revised Graduated Prenatal Care Index; ADI, Area Deprivation Index. °Percent imprecisely geocoded.

Table 2. Adjusted bivariate logistic regression of imprecise geocoding of Kentucky birth records by rural-urban classification 2007-2018.

	Metro aOR (95% CI)	Non-metro aOR (95% CI)	Rural aOR (95% CI)
Mother's age (years)			
<23	Reference	Reference	Reference
23-25	0.94 (0.88, 0.98)	1.02 (0.98, 1.06)	1.06 (0.98, 1.12)
26-30	0.88 (0.82, 0.92)	1.04 (1.00, 1.10)	1.06 (1.00, 1.14)
31-41	0.84 (0.78, 0.90)	1.10 (1.04, 1.16)	1.12 (1.02, 1.22)
Unknown	0.98 (0.82, 1.18)	1.16 (0.98, 1.36)	1.12 (0.90, 1.42)
Mother's race			
White	Reference	Reference	Reference
Other	0.80 (0.74, 0.86)	1.16 (1.04, 1.30)	1.16 (0.98, 1.38)
Black	0.52 (0.48, 0.56)	1.42 (1.30, 1.54)	1.96 (1.68, 2.28)
Mother's ethnicity			
Hispanic	Reference	Reference	Reference
Not Hispanic	1.74 (1.58, 1.94)	1.20 (1.06, 1.38)	0.86 (0.70, 1.04)
Mother's education			
Less than HS	Reference	Reference	Reference
High School	0.90 (0.86, 0.94)	0.84 (0.80, 0.86)	0.82 (0.78, 0.88)
Some College +	0.76 (0.72, 0.80)	0.84 (0.80, 0.88)	0.84 (0.80, 0.90)
Unknown	0.84 (0.72, 1.00)	0.68 (0.60, 0.78)	0.68 (0.54, 0.82)
Mother's BMI (kg/m^2)			
Normal (18-24.9)	Reference	Reference	Reference
Underweight (<18)	1.14 (1.06, 1.24)	1.14 (1.08, 1.22)	1.12 (1.02, 1.24)
Overweight (25-29.9)	1.08 (1.04, 1.12)	1.06 (1.02, 1.08)	1.04 (1.00, 1.10)
Obese (>30)	1.14 (1.10, 1.18)	1.10 (1.06, 1.14)	1.06 (1.00, 1.12)
Unknown	1.02 (0.94, 1.12)	1.28 (1.18, 1.38)	1.22 (1.08, 1.38)
Marital status			
Yes	Reference	Reference	Reference
No or not stated	0.70 (0.68, 0.74)	1.02 (1.00, 1.06)	1.10 (1.06, 1.16)
Father's age			
<24	Reference	Reference	Reference
25-28	0.90 (0.86, 0.96)	0.96 (0.92, 1.00)	0.92 (0.86, 0.98)
29-33	0.80 (0.76, 0.86)	0.96 (0.90, 1.00)	0.88 (0.82, 0.94)
>33	0.88 (0.82, 0.94)	1.06 (1.00, 1.12)	0.96 (0.90, 1.06)
Unknown	0.84 (0.76, 0.94)	1.24 (1.12, 1.34)	1.26 (1.10, 1.42)
Father's race			
White	Reference	Reference	Reference
Other	0.70 (0.64, 0.74)	0.78 (0.72, 0.84)	0.94 (0.84, 1.06)
Black	0.72 (0.66, 0.78)	1.04 (0.94, 1.12)	1.32 (1.14, 1.54)
Unknown	0.50 (0.38, 0.64)	0.68 (0.54, 0.82)	0.52 (0.38, 0.72)
Father's ethnicity			
Hispanic	Reference	Reference	Reference
Non-Hispanic	1.28 (1.16, 1.44)	0.98 (0.86, 1.12)	0.82 (0.68, 0.98)
Unknown	2.38 (2.02, 2.78)	1.28 (1.10, 1.50)	0.88 (0.70, 1.08)
Father's education			
Less than HS	Reference	Reference	Reference
High School	0.84 (0.78, 0.88)	0.90 (0.86, 0.94)	0.96 (0.90, 1.04)
Some College +	0.54 (0.50, 0.56)	0.84 (0.80, 0.88)	1.06 (0.98, 1.14)
Unknown	0.72 (0.64, 0.82)	0.78 (0.70, 0.86)	0.92 (0.78, 1.08)
Birthplace			
Hospital	Reference	Reference	Reference
Home	0.88 (0.76, 1.04)	0.66 (0.56, 0.78)	0.18 (0.14, 0.24)
Other	2.28 (1.70, 3.06)	0.78 (0.56, 1.08)	0.16 (0.08, 0.30)
Parity			
0	Reference	Reference	Reference
1	1.06 (1.02, 1.12)	0.96 (0.94, 1.00)	1.02 (0.96, 1.06)
2 +	1.14 (1.10, 1.20)	1.00 (0.96, 1.04)	1.02 (0.96, 1.08)
R-GINDEX			
Adequate (80%-109%)	Reference	Reference	Reference
Inadequate (<50%)	1.40 (1.32, 1.48)	1.12 (1.06, 1.18)	1.04 (0.96, 1.12)
Intermediate (50-79%)	1.04 (1.00, 1.08)	1.10 (1.06, 1.14)	1.00 (0.96, 1.06)
Intensive (>110%)	1.26 (1.18, 1.34)	0.98 (0.94, 1.04)	1.00 (0.92, 1.08)
Missing/No care	0.98 (0.92, 1.06)	1.18 (1.10, 1.26)	1.36 (1.24, 1.52)
ADI			
Lowest (1-3)	Reference	Reference	Reference
Mid-low (4-5)	1.68 (1.60, 1.76)	4.46 (4.20, 4.76)	0.68 (0.58, 0.80)
Mid-high (6-7)	1.38 (1.32, 1.46)	1.30 (1.20, 1.40)	0.58 (0.48, 0.68)
Highest (8-10)	1.78 (1.70, 1.88)	2.66 (2.50, 2.84)	0.42 (0.36, 0.50)
Missing ADI	12.22 (11.22, 13.32)	8.50 (7.68, 9.40)	1.20 (0.98, 1.46)
Appalachian region			
Non-Appalachian	Reference	Reference	Reference
Appalachia	1.38 (1.30, 1.48)	3.16 (3.04, 3.26)	2.72 (2.56, 2.90)
Year	1.00 (1.00, 1.02)	0.92 (0.92, 0.94)	0.92 (0.92, 0.94)

HS, High School; BMI, body mass index; ADI, Area Deprivation Index; R-GINDEX, Revised Graduated Prenatal Care Index.

ciated with imprecision, but the magnitude and direction varied among RUCCs. Further, we identified that rural regions, particularly in the Appalachian region, had the highest proportion of imprecisely geocoded addresses.

References

Andresen MA, Malleson N, Steenbeek W, Townsley M, Vandeviver C, 2020. Minimum geocoding match rates: an international study of the impact of data and areal unit sizes. Int Geo Inf Sci 34:1306-22.

Ash M, Fetter TR, 2004. Who lives on the wrong side of the environmental tracks? Evidence from the Epa's risk-screening environmental indicators model*. Soc Sci Q 85:441-62.

Buttling LG, Mcknight MX, Kolivras KN, Ranganathan S, Gohlke JM, 2021. Maternal proximity to Central Appalachia Surface Mining and Birth Outcomes. Environ Epidemiol 5:e128.

Cromley EK, 2019. Using Gis to address epidemiologic research questions. Curr Epidemiol Rep 6:162-73.

Gilboa SM, Mendola P, Olshan AF, Harness C, Loomis D, Langlois PH, Savitz DA, Herring AH, 2006. Comparison of residential geocoding methods in population-based study of air quality and birth defects. Environ Res 101:256-62.

Goldberg D, Swift JN, Wilson JP, 2008. Geocoding best practices: reference data, input data, and feature matching. University of Southern California GIS Research Laboratory, Techincal Report No. 9, Los Angeles, CA, USA.

Ha S, Hu H, Mao L, Roussos-Ross D, Roth J, Xu X, 2016. Potential selection bias associated with using geocoded birth records for epidemiologic research. Ann Epidemiol 26:204-11.

Health UOWSOMP, 2021. 2015 Area Deprivation Index V2.0. Inc.

Jacquez GM, 2012. A research agenda: Does geocoding positional error matter in health Gis studies? Spat Spatiotemporal Epidemiol 3:7-16.

Jones RR, Dellavalle CT, Flory AR, Nordan A, Hoppin JA, Hofmann JN, Chen H, Giglierano J, Lynch CF, Beane Freeman LE, Rushton G, Ward MH, 2014. Accuracy of residential geocoding in the agricultural health study. Int J Health Geogr 13:37.

Kind AJH, Buckingham WR, 2018. Making neighborhood-disadvantage metrics accessible - the neighborhood atlas. N Engl J Med 378:2456-58.

Kotelchuck M, 1994. The adequacy of prenatal care utilization index: its us distribution and association with low birthweight. Am J Public Health 84:1486-89.

Krieger N, Waterman P, Chen JT, Soobader M-J, Subramanian SV, Carson R, 2002. Zip Code Caveat: Bias Due to Spatiotemporal Mismatches between Zip Codes and Us Census-Defined Geographic Areas - the Public Health Disparities Geocoding Project. Am J Public Health 92:1100-02.

Krieger N, Waterman P, Lemieux K, Zierler S, Hogan JW, 2001. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. Am J Public Health 91:1114-16.

Mask D, 2014. Where the streets have no name. Atlantic.

Oliver MN, Matthews KA, Siadaty M, Hauck FR, Pickle LW, 2005. Geographic bias related to geocoding in epidemiologic studies. Int J Health Geogr 4:29.

Schinasi LH, Auchincloss AH, Forrest CB, Diez Roux AV, 2018. Using electronic health record data for environmental and place based population health research: a systematic review. Ann Epidemiol 28:493-502.

The New York Times 1996. Kentucky faces tradition of offensive road names. The New York Times Company, p 13.

Zandbergen PA, 2007. Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. BMC Public Health 7:37.