Spatial and temporal patterns in preterm birth in the United States

John Byrnes¹, Richard Mahoney², Cele Quaintance^{3,4}, Jeffrey B. Gould^{3,4}, Suzan Carmichael^{3,4}, Gary M. Shaw^{3,4}, Amy Showen⁵, Ciaran Phibbs^{3,4}, David K. Stevenson^{3,4} and Paul H. Wise^{3,4,6}

BACKGROUND: Despite years of research, the etiologies of preterm birth remain unclear. In order to help generate new research hypotheses, this study explored spatial and temporal patterns of preterm birth in a large, total-population dataset.

METHODS: Data on 145 million US births in 3,000 counties from the Natality Files of the National Center for Health Statistics for 1971–2011 were examined. State trends in early (<34 wk) and late (34–36 wk) preterm birth rates were compared. K-means cluster analyses were conducted to identify gestational age distribution patterns for all US counties over time.

RESULTS: A weak association was observed between state trends in <34 wk birth rates and the initial absolute <34 wk birth rate. Significant associations were observed between trends in <34 wk and 34–36 wk birth rates and between white and African American <34 wk births. Periodicity was observed in county-level trends in <34 wk birth rates. Cluster analyses identified periods of significant heterogeneity and homogeneity in gestational age distributional trends for US counties.

CONCLUSION: The observed geographic and temporal patterns suggest periodicity and complex, shared influences among preterm birth rates in the United States. These patterns could provide insight into promising hypotheses for further research.

Preterm birth, defined by the World Health Organization as the delivery of an infant before 37 completed weeks of gestational duration (1), is a major cause of infant mortality and long-term child morbidity in the United States (2–4). The search for the causes of preterm birth and its prevention has engaged multiple disciplines including the basic, clinical, and social sciences. While these research initiatives have documented associations between preterm birth and a wide range of biologic, demographic, social, and clinical influences (5–7), definitive insight into the primary etiologic pathways of preterm birth remains elusive (8). There has been growing interest in the use of new analytic frameworks and techniques (9) to explore broad spatial and temporal patterns of preterm birth in the hope that these approaches might provide new clues to important etiologies or preventive opportunities (10–11). Often colloquially labeled "big data" analyses, these strategies take advantage of expanding computing power and machine learning techniques to illuminate relationships or patterns in very large genetic, clinical, or epidemiologic datasets (12–13). Among the most useful of these approaches are cluster analyses which have been employed to identify complex patterns in health service utilization (14) and a variety of health outcomes, including chronic respiratory disease (15), schizophrenia (16), and necrotizing enterocolitis (17).

This study assesses recent spatial and temporal trends in preterm birth in the United States. Of special concern is the exploration of patterns in these trends, particularly among different gestational age groupings, geographic regions, and social groups over time.

RESULTS

Temporal Trends

Trends of preterm birth in the United States are presented in **Figure 1**. Total births of <37 wk of gestation rose steadily between 1981 and 2006, increasing 35.6% over this time period. However, after 2006, the percentage of births born at <37 wk began to decline, falling 8.4% between 2006 and 2011. This trend was largely accounted for by changes in births of between 34 and 37 wk gestation. The percentage of births of <32 wk gestation was generally more stable over the study period but did rise 12.7% from 1981 through 2006 and then declined by 5.4% between 2007 and 2011.

Rates of preterm birth and trends for the 1995 to 2010 period and for the 2011 year for each of the 50 states and the District of Columbia are presented in **Table 1**. Considerable variation in trends over these periods was noted. States in the highest quintile of <34 wk birth rates in 1995, showed little change between 1995 and 2005 but averaged a 4.6% decline in this category

Received 13 August 2014; accepted 26 November 2014; advance online publication 15 April 2015. doi:10.1038/pr.2015.55

¹Artificial Intelligence Center, SRI International, San Diego, California; ²Robotics Program, SRI International, Menlo Park, California; ³Division of Neonatology, Department of Pediatrics, Stanford University School of Medicine, Stanford, California; ⁴March of Dimes Center for Prematurity Research at Stanford, Stanford University School of Medicine, Stanford, California; ⁵David Geffen School of Medicine, University of California, Los Angeles, California; ⁶Center for Policy, Outcomes and Prevention, Center for Health Policy/ Center for Primary Care and Outcomes Research, Stanford University, Stanford, California. Correspondence: Paul H. Wise (pwise@stanford.edu)



Figure 1. Trends in preterm birth rates for the United States by gestational age groups <32 wk, 32–37 wk, and total births <37 wk gestation: 1981–2011.

between 2005 and 2010. States in the lowest quintile of <34 wk birth rates in 1995 experienced modest increases between 1995 and 2005 and a 3.0% decline in the 2005–2010 period.

The examination of gestational age trends between 2000 and 2010 revealed a general inverse relationship between the absolute rate of <34 births in 2000 and the percent reduction in this rate over the decade. However, this relationship was significant only for births to white women (P = 0.037). The relative weakness of the association between initial absolute <34 wk birth rates and subsequent trends was also true when births <31 wk gestation were assessed independently.

The association between trends for the <34 wk and 34–36 wk gestational age groups are presented in **Figure 2**. Significant associations were found for total (P < 0.05) (see **Figure 2a**), white (P < 0.05) (see **Figure 2b**), and African American births (P < 0.01) (see **Figure 2c**). Note that only states reporting at least 1,000 African American births were included in the figure and that the scale of % change was greater for African American births than that for whites. In addition, a significant relationship was also observed between trends in <34 wk gestation birth rates for white and African American births ($P \le 0.001$) (see **Figure 2d**).

Gestational age-specific birth rates were then calculated for all US counties between the years 1971 and 2008. 905 counties reported more than 10 births of <34 wk gestation annually for this period. However, the calculation of total national counts of year-to-year changes in county preterm birth rates was examined only for the 1985–2008 period, the years for which all states reported county-based data. **Figure 3** presents the smoothed, moving average annual change in the percent of births occurring <34 wk for the 905 counties. The number of counties reporting >5% increase in the birth rate of <34 wk infants was generally similar to the number of counties reporting >5% decrease. However, counties reporting relative stability in the <34 wk birth rate generally outnumbered those reporting >5% changes. The figure also suggests that some



periodicity may exist in the number of counties reporting increased rates of <34 wk births. Significant peaks occurring at \sim 7-y intervals were noted.

Cluster Analysis

The k-means cluster analysis generated 10 patterns that best described the year-to-year changes in the distribution of births among nine gestational age groups. In this manner, each county was assigned to one of these characteristic clusters for each year during the study period. Figure 4 presents both the gestational age definitions generated by the k-means cluster analysis (Figure 4a) and the number of counties that were characterized by each of these 10 clusters for each year under study (Figure 4b). The colors were assigned randomly to each of the 10 clusters but are consistent in both Figure 4a,b. Figure 4a suggests considerable variation in the gestational age patterns generated by the cluster analysis, primarily in the gestational age groups > 36 wk. Figure 4b documents that while the cluster analysis was structured to identify 10 unique clusters, only 3 clusters accounted for more than half of all the county-years. The prevalence of these three clusters, cluster 1 (color-coded as pink), cluster 4 (color-coded as orange), and cluster 7 (color-coded at tan) are shown in Figure 4b. Note that the y-axis for the prevalence of each of the clusters in the figure varies, given the wide variation in the number of counties falling into each cluster in any given year. Cluster 4 (orange) and cluster 7 (tan) which predominated in the earlier years under study were characterized by relative increases in late preterm births. Cluster 1 (pink), which predominated in the latter years of study, was characterized by relative declines in mid and late preterm births. More broadly, the early years of the study were characterized by considerable heterogeneity in cluster assignments with cluster 4 (orange) accounting for the largest number of counties. However, by the early 1980's, cluster 7 (tan) began to dominate the county assignments, creating a far more homogeneous cluster pattern. The 2000's were characterized by greater heterogeneity in county assignments. During this latter period, cluster 1 (pink) was the most common cluster assignment.

The geographic distribution of the cluster assignments is depicted in Figure 5. Each of the national maps presents the cluster assignments (designated by the assigned colors) for 5 representative years. The map for 1972 reflects areas of missing data as county data were not available for a considerable number of states during the early years of the study period. These were omitted from the cluster analyses until data became available from all these states in the mid 1970's, except from Arizona which reported data beginning in the mid-1980's. Missing county data are depicted in the geo-temporal mapping as blank areas. However, for the majority of states reporting county data, there was substantial heterogeneity in cluster patterns throughout the United States. The maps for 1982, 1992, and 2002, on the other hand, depict the predominance of cluster 7 (tan) and cluster 1 (pink) throughout the country. This suggests that over this time period, the trends in gestational age distribution noted in national data reflected

Articles Byrnes et al.

Table 1.	Trends in preterm births and birth rates: US states and the District of Columbia

	2011			%change 1995–2000		%change 2000–2005		%change 2005–2010	
State	Births	%<34 wk	%34–36 wk	<34 wk	34–36 wk	<34 wk	34–36 wk	<34 wk	34–36 wk
Alabama	59,354	4.6%	10.3%	8.4%	12.2%	9.8%	12.2%	-9.6%	-5.5%
Alaska	11,456	2.8%	7.5%	12.1%	31.2%	-6.8%	11.6%	-7.8%	-9.3%
Arizona	85,543	2.9%	9.2%	7.0%	2.3%	10.7%	13.7%	-3.5%	-8.5%
Arkansas	38,715	3.6%	9.5%	14.7%	9.4%	-2.3%	1.9%	-7.9%	-3.6%
California	502,120	2.7%	7.1%	-1.8%	-1.6%	7.6%	9.0%	-3.9%	-3.9%
Colorado	65,055	3.0%	7.3%	1.2%	0.7%	11.1%	14.9%	-9.0%	-13.4%
Connecticut	37,281	3.1%	6.9%	8.8%	9.0%	7.8%	0.1%	-1.7%	-1.0%
Delaware	11,257	3.7%	7.4%	4.1%	7.1%	7.6%	12.1%	-0.5%	-12.1%
District of Columbia	9,295	4.5%	9.1%	-19.8%	-7.2%	-23.6%	-8.6%	21.3%	-13.0%
Florida	213,414	3.9%	9.1%	2.7%	6.5%	11.8%	11.0%	0.5%	-5.2%
Georgia	132,409	3.9%	9.3%	-3.8%	5.0%	4.0%	18.0%	7.3%	-1.5%
Hawaii	18,956	3.6%	8.7%	5.8%	3.9%	18.8%	7.6%	-2.8%	1.2%
Idaho	22,305	2.6%	7.6%	2.2%	11.2%	11.5%	12.7%	-4.8%	-11.0%
Illinois	161,312	3.7%	8.4%	1.0%	3.7%	7.0%	8.9%	-5.3%	-8.2%
Indiana	83,701	3.3%	8.2%	6.6%	12.4%	14.3%	13.7%	-13.5%	-12.7%
lowa	38,214	3.1%	8.0%	8.3%	10.6%	11.4%	12.2%	-2.4%	-2.5%
Kansas	39,642	3.0%	8.2%	13.1%	10.6%	5.7%	12.3%	-11.1%	-13.8%
Kentucky	55,370	3.7%	9.6%	4.1%	10.9%	14.8%	22.2%	-5.8%	-11.0%
Louisiana	61,888	4.9%	10.6%	2.8%	9.7%	10.7%	9.5%	-10.8%	-7.5%
Maine	12,704	2.8%	6.8%	0.9%	5.1%	6.2%	14.6%	0.3%	-11.7%
Maryland	73,093	3.8%	8.6%	-3.1%	2.2%	1.0%	6.0%	-1.5%	-5.1%
Massachusetts	73,166	3.1%	7.2%	22.4%	20.3%	12.8%	13.7%	-8.2%	-4.4%
Michigan	114,008	3.7%	8.3%	3.8%	7.8%	5.0%	7.9%	2.9%	-3.6%
Minnesota	68,409	2.8%	7.1%	6.9%	6.7%	10.5%	14.0%	-6.1%	-4.3%
Mississippi	39,860	5.0%	11.9%	5.6%	5.8%	9.6%	17.9%	-6.2%	-6.1%
Missouri	76,117	3.3%	8.3%	5.7%	3.7%	12.6%	12.9%	-4.3%	-10.4%
Montana	12,069	2.9%	7.9%	11.7%	12.5%	-6.7%	12.3%	23.8%	-0.9%
Nebraska	25,720	2.9%	7.7%	8.7%	5.2%	-50.4%	107.4%	-9.2%	-5.1%
Nevada	35,296	3.5%	9.7%	6.9%	10.2%	4.4%	-4.7%	11.1%	14.6%
New Hampshire	12,851	2.8%	6.7%	24.7%	4.1%	5.6%	22.6%	-8.5%	-11.4%
New Jersey	105,883	3.7%	7.9%	13.4%	9.2%	7.8%	8.5%	-3.2%	-8.9%
New Mexico	27,289	3.0%	8.8%	18.8%	5.7%	1.7%	9.4%	-13.9%	-7.3%
New York	241,312	3.2%	7.7%	0.1%	5.3%	8.7%	10.6%	-2.3%	-6.5%
North Carolina	120,389	4.1%	8.4%	0.9%	3.8%	2.6%	6.9%	-3.5%	-8.7%
North Dakota	9,527	2.9%	7.0%	2.7%	6.3%	13.1%	6.3%	-4.7%	-5.3%
Ohio	137,918	3.9%	8.2%	1.4%	6.4%	10.9%	10.7%	1.3%	-9.1%
Oklahoma	52,272	3.7%	9.4%	9.8%	10.8%	7.2%	13.3%	8.1%	5.3%
Oregon	45,155	2.4%	6.7%	5.2%	3.3%	5.1%	12.5%	2.5%	-3.8%
Pennsylvania	143,178	3.3%	7.7%	-0.3%	4.3%	9.6%	6.8%	-3.9%	-3.6%
Rhode Island	10,960	3.1%	7.2%	22.7%	10.8%	6.0%	13.9%	-5.4%	-12.4%
South Carolina	57,393	4.5%	9.5%	6.3%	10.2%	14.6%	15.0%	-6.3%	-10.2%
South Dakota	11,846	2.9%	8.2%	9.6%	15.9%	-7.4%	11.3%	30.2%	-9.7%
Tennessee	79,588	3.6%	9.1%	-1.3%	5.5%	10.9%	7.5%	-16.5%	-10.2%
Texas	377,445	3.6%	9.2%	9.7%	10.8%	7.2%	8.5%	-3.4%	-3.4%
Utah	51,223	2.6%	8.2%	14.2%	13.7%	3.9%	11.1%	-2.1%	-5.9%
Vermont	6,078	2.7%	6.1%	-5.4%	0.8%	20.7%	7.0%	2.4%	-10.9%
Virginia	102,652	3.4%	7.7%	-2.9%	-1.8%	-0.4%	7.7%	2.2%	-8.2%
Washington	86,976	2.8%	7.0%	5.3%	13.7%	14.8%	11.7%	-0.6%	-5.0%
West Virginia	20,717	3.8%	8.9%	6.1%	16.1%	3.8%	17.2%	-12.3%	-16.9%
Wisconsin	67,810	2.9%	7.5%	9.3%	8.6%	8.5%	8.9%	-2.5%	-6.1%
Wyoming	7,399	2.5%	7.7%	-0.8%	13.6%	7.2%	13.4%	-5.4%	-19.0%

Patterns in preterm birth

Articles



Figure 2. The relationships between the percent change in gestational age <34 wk birth rates and the percent change in gestational age 34–36 wk birth rates between 2000 and 2010 in all US states and the District of Columbia and for White and African American births. Each data point represents one state or the District of Columbia (**a**). Births in all states and the District of Colombia; (**b**) White births; (**c**) African American births; (**d**) Percent change in gestational age <34 wk birth rates for African American births. Note that states reporting less than 500 African American births for any year during the study period are not plotted on the figure.

fairly generalized trends with relatively little geographic variation among different counties or regions. However, the map for 2008 reveals greater variation with cluster 2 (red), cluster 5 (purple), and cluster 9 (dark blue) growing in prevalence, particularly in parts of the mid-west.

DISCUSSION

This study utilized four decades of national birth data and machine learning algorithms to examine spatial and temporal patterns in preterm births for the total population of United States. States with high absolute rates at the beginning of the study period tended to have more precipitous declines over subsequent years; however, this was not uniformly observed as some states with high initial rates experienced little improvement. While early and late preterm birth rates may reflect different phenotypes or etiologies, our observations indicate that trends in these rates tend to move together within states. Moreover, while the disparity in preterm birth rates between African American and white births in any given state remains profound, the absolute rates also tend to move together over time. County-level birth rates of preterm infants also revealed some temporal synchrony with the suggestion of a possible multi-year periodicity. When machine learning techniques were used to assess temporal and spatial patterns in preterm births, a dynamic picture emerged characterized by periods of both heterogeneity and homogeneity among US counties.

Articles Byrnes et al.



Figure 3. Number of counties reporting an annual >5 percent and >10 percent increase in the birth rate of <34 wk gestation births: United States, moving 2-y average (total counties = 905).

There was considerable variation in state preterm birth rate trends over the study period. Overall, jurisdictions with the highest absolute preterm birth rates in 2000, such as the District of Columbia, experienced on average the largest declines over the subsequent decade. This observation could reflect the general tendency of initially aberrant rates to regress to the mean. However, the relationship between initial absolute rates and subsequent trends was not strong to begin with and there were states that did not conform to this general observation. The reasons for this variation are unclear; however, trends in a variety of demographic or obstetrical management strategies are not likely to have been responsible for the observed variations (18,19).

While it has been difficult in large datasets to accurately distinguish between the wide variety of clinical phenotypes associated with preterm birth (20), the stratification of preterm births into early and late preterm births has served as a useful proxy (21). While there was considerable differentiation in the state-specific trends in early and late preterm birth rates,



Figure 4. K-means cluster assignments by year: All counties, United States. (**a**) The gestational age definitions derived from the cluster analysis. (**b**) The prevalence of each cluster among US counties (note differences in prevalence scales for each cluster). Each color represents a different cluster and is consistently assigned in both **a** and **b**. The colors in descending order in the Figure are: pink, red, green, orange, purple, light purple, tan, light green, turquoise, and light blue.

Patterns in preterm birth





Figure 5. Maps of county cluster assignments for selected years. (a) 1972; (b) 1982; (c) 1992; (d) 2002; (e) 2008. Colors represent one of the 10 cluster assignments as defined in Figure 4 (note gray areas represent counties that did not report sufficient data for the specified year).

the analyses presented here suggest that these two rates do not move independently within states. Rather, there was a significant relationship in the movement of these rates over time raising the question of whether these gestational age categories may share some common influences.

Periodicity in preterm birth rates has been noted for some time. However, this has largely taken the form of seasonal patterns, including patterns that may be related to annual infectious outbreaks, such as influenza (10). There has been little evidence suggesting significant periodicity beyond 1 y. The finding in this study of some periodicity in multi-year time scales warrants confirmation and is undergoing more detailed analysis. Nevertheless, these findings do raise questions regarding potential infectious etiologies that tend to operate on multi-year cycles (22). In addition, social phenomena, including economic cycles (23,24), demographic shifts, including trends in maternal age (25), or other environmental influences (26,27) could produce complex, multi-year cycles and are worthy of more directed research into the specific role these potential influences have on trends in preterm birth.

The cluster analysis revealed both periods of relative congruity and incongruity in preterm birth trends across the United States. The use of unsupervised, agnostic clustering techniques are increasingly being used to identify heretofore unrecognized

Articles

Byrnes et al.

patterns in large, multidimensional datasets concerned with complex health disorders (28,29). In this analysis clusters were determined by seeking the 10 patterns that best described yearto-year changes in gestational age distributions for all counties in the United States from 1971 through 2008. While the predominance of shifts in late preterm and term births appears to have influenced the most prevalent clusters, the generalized distribution of these trends at the county level was of special interest. The transformation over time from a highly heterogeneous pattern to a long, homogeneous pattern and then most recently, back to a more heterogeneous pattern raises important questions about the broad influences shaping gestational age patterns in specific geographic areas. There is a large literature documenting a wide variety of individual social, behavioral, and environmental contributors to preterm birth, many of which vary substantially by geographic location. Periods of substantial spatial heterogeneity in preterm birth patterns could reflect a time frame in which such local, variable influences predominate. However, during periods of substantial homogeneity in preterm birth patterns, these local variations in etiologic influences may become relatively aligned or combined with broad currents of influence that affect gestational age distributions across large geographic areas, transcending county, state and regional borders. While this analysis cannot identify the distinct contribution of any specific influence, it is useful to document that such broad, spatial, and temporal influences exist and may be important in shaping patterns of preterm birth. Such influences could include rapidly disseminated changes in clinical practice (30,31), regional or national trends in economic well-being (23,24), or wide-scale environmental exposures, some of which remain somewhat speculative (27).

The findings of this analysis should be interpreted with caution. The determination of gestational age in US Natality Files has traditionally been based on reported last menstrual period, a variable long recognized as being susceptible to error, particularly for low gestational ages (32). In addition, the relatively long period under study in this analysis would likely involve a changing capacity to assess gestational age correctly. However, for this concern to introduce biased estimation, the error in gestational classification would need to vary with the geographic and spatial clusters. Nevertheless, the US Natality Files have been used successfully to identify a variety of important risk factors for preterm birth as well as in documenting longterm trends in adverse birth outcomes (33). Moreover, the focus of this analysis was on broad patterns of preterm birth using relatively large geographic aggregations of births or clusters of births. The presented analyses also utilize both state and county-level data. These spatial jurisdictions are hierarchal in nature (states are composed of counties) and, while the county and state-level analyses are presented separately, some caution should be used in assessing potential influences that may vary or be shared by individuals or across different spatial aggregations (34).

The cluster analysis should also be interpreted with care. The construction of 10 clusters in the k-means analysis was an arbitrary decision and it is possible that levels of heterogeneity could change if more or fewer clusters were constructed. In addition, unlike the temporal trend analyses, the cluster analyses were confined to singleton births in order to seek patterns generally unassociated with factors related to multiple gestations. The clustering models utilizing three temporal periods could generate anomalous results at the earliest and latest margins of the period under study. However, a sensitivity analysis was conducted to assess this concern and confirmed the robustness of the presented cluster assignments. It should also be noted that there have been a variety of new techniques in the spatial and temporal analysis of complex datasets (35), some of which have been applied to the study of birth outcomes (36). The current analysis used k-means cluster strategies because of its computational efficiency for high-dimensional analyses in very large datasets; however, these new statistical refinements could provide enhanced analytic approaches to spatial and temporal health-related research.

While the exploration of broad spatial and temporal patterns of preterm birth using very large databases and machine learning techniques cannot substitute for detailed epidemiologic, clinical, and basic investigations, the analyses presented here utilize new analytic techniques to explore broad patterns of preterm birth in the United States. The findings suggest potential periodicity and common influences among different gestational age groups and disparate geographic jurisdictions over time. These broad temporal and spatial patterns could help frame new analytic perspectives and promising investigative hypotheses into preterm birth, one of the most profound and persistent adverse health outcomes in the United States today.

METHODS

Study Population

Our analyses utilized birth certificate information for all births in the United States between 1971 and 2011 and were derived from the National Center for Health Statistics Natality Files (37). These files provide information on a variety of demographic, maternal and infant characteristics, including maternal and infant race\ethnicity, estimated gestational age, and county and state of maternal residence. All counties across 51 regions (50 states and the District of Columbia) were included in the analysis. The analyzed data set consisted of more than 3,000 counties over 40 y, yielding about 120,000 county-year data points representing ~145 million live births.

Temporal Trends

Gestational age-specific birth rates (i.e., birth prevalences) were calculated for each year in the study period. These rates were generated by dividing the number of births in a specified gestational age category in a given year by the total number of births for that population and year and presented as a percentage of all births. For the secular trend analyses, gestational age-specific birth rates were calculated for all live-births (including all singleton and multiple births; stillbirths were not included) in each state and the District of Columbia. State and county designation was allocated based on the residence of the mother at the time of delivery and not on the state or county location of birth. Similarly, the designation of a birth as white or African American was based on maternal report. Analyses were performed by using SAS 9.1 (SAS Institute, Cary, NC) statistical software. Simple trends were calculated for 5-y periods between 1995 and 2010. In addition, analyses were conducted to assess whether states with relatively high absolute rates of preterm births in 2000 were more or less likely to have experienced a decline in these rates over the subsequent 10-y period. Analyses were also conducted to examine the relationship between trends in <34 wk and 34–36 wk categories and between different racial-ethnic groups.

County-Level Trends

Distinct from the state-based analyses, trends were calculated for all US counties for the 1985 through 2008 period. County-level data were not available subsequent to 2008. Total counts of all counties experiencing a >5% and/or >10% year-to-year increase of <34 wk births were calculated for the 1985–2008 period, the years for which all states reported county-based data. Trends were calculated as the percent change in preterm births over the examined period as $(R_{t2}-R_{t1})/(100)$, where R_{t2} represents the latter year or years and R_{t1} the initial study year or years. Simple and smoothed, moving average plots of counties experiencing annual changes of >5% and >10% increases in <34 wk births were calculated.

Cluster Analysis

We used cluster analysis to characterize temporal-spatial patterns of gestational age at delivery for all singleton births in the United States between 1971 and 2008. Clustering is used to detect latent structures or regularities within a dataset. We used the k-means procedure, a clustering strategy that requires the choice of the desired number of cluster centers with the k-means procedure iteratively estimating the cluster centers by minimizing the total variance within each cluster (38). The first step of the analysis was the calculation of gestational age-specific birth rates for singleton births for each county in the United States for the period 1971 through 2008. The percentage of births that fell into each of the following gestational age categories for each year in each county was computed: Under 20, 20-27, 28-31, 32-35, 36, 37-39, 40, 41, 42 wk and over. Next temporal trends for each category were characterized using a model which breaks the date range 1971–2008 into three periods 1971-A, A-B, B-2008 (where the model selects A and B independently for each county and each age category) and fits a straight line to the data in each time range, forcing the lines to meet at the boundary years A and B. The slope of the trend line in each year was taken to characterize the trend for that age group in that county for that year. Each year in each county was thus characterized by nine gestational age trend slopes. Each county-year pair was assigned to this vector of nine values. The number of clusters is set *a priori* and here the vectors were associated into 10 clusters via the k-means clustering methodology (26). The above analysis was carried out in python (Python Software Foundation, Beaverton, Oregon) using the SciPy library (SciPy.org). The presented analyses utilized publicly available vital statistics datasets and were approved by the Human Subjects Institutional Review Board of Stanford University.

STATEMENT OF FINANCIAL SUPPORT

This study was funded in part by the March of Dimes (Grant # 50185), White Plains, New York

REFERENCES

- World Health Organization. The prevention of perinatal mortality and morbidity. Geneva, Switzerland: WHO Technical Report Series, 1970. Report 457.
- Chang HH, Larson J, Blencowe H, et al.; Born Too Soon preterm prevention analysis group. Preventing preterm births: analysis of trends and potential reductions with interventions in 39 countries with very high human development index. Lancet 2013;381:223–34.
- Anderson PJ, Doyle LW. Cognitive and educational deficits in children born extremely preterm. Semin Perinatol 2008;32:51–8.
- Kugelman A, Colin AA. Late preterm infants: near term but still in a critical developmental time period. Pediatrics 2013;132:741–51.
- Behrman RE, Butler AS. Sociodemographic and community factors contributing to preterm birth. In: *Preterm Birth: Causes, Consequences, and Prevention*. Washington, DC: National Academies Press, 2007:124–47.
- Kramer MR, Hogue CR. Place matters: variation in the black/white very preterm birth rate across U.S. metropolitan areas, 2002-2004. Public Health Rep 2008;123:576–85.



- Klebanoff MA, Keim SA. Epidemiology: the changing face of preterm birth. Clin Perinatol 2011;38:339–50.
- Muglia LJ, Katz M. The enigma of spontaneous preterm birth. N Engl J Med 2010;362:529–35.
- Lackritz EM, Wilson CB, Guttmacher AE, et al.; Preterm Birth Research Priority Setting Group. A solution pathway for preterm birth: accelerating a priority research agenda. Lancet Glob Health 2013;1:e328–30.
- 10. Currie J, Schwandt H. Within-mother analysis of seasonal patterns in health at birth. Proc Natl Acad Sci USA 2013;110:12265–70.
- Padula AM, Mortimer KM, Tager IB. Traffic-related air pollution and risk of preterm birth in the San Joaquin Valley of California. Ann Epidem 2014;24:888–95e4.
- Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform 2014;9:8–13.
- Larson EB. Building trust in the power of "big data" research to serve the public good. JAMA 2013;309:2443–4.
- Vencloviene J, Babarskiene R, Dobozinskas P, Siurkaite V. Effects of weather conditions on emergency ambulance calls for acute coronary syndromes. Int J Biometerol 2014. Online First. http://link.springer.com/article/10.100 7%2Fs00484-014-0921-6. 26 October 2014.
- Lee JH, Cho MH, McDonald ML, et al.; COPDGene Investigators. Phenotypic and genetic heterogeneity among subjects with mild airflow obstruction in COPDGene. Respir Med 2014;108:1469–80.
- Lin LA, Rosenheck R, Sugar C, Zbrozek A. Comparing antipsychotic treatments for schizophrenia: A health state approach. Psychiatr Q. 2015;86:107–21.
- Brower-Sinning R, Zhong D, Good M, et al. Mucosa-associated bacterial diversity in necrotizing enterocolitis. PLoS One 2014;9:e105046.
- Shapiro-Mendoza CK, Lackritz EM. Epidemiology of late and moderate preterm birth. Semin Fetal Neonatal Med 2012;17:120–5.
- Tita AT, Landon MB, Spong CY, et al.; Eunice Kennedy Shriver NICHD Maternal-Fetal Medicine Units Network. Timing of elective repeat cesarean delivery at term and neonatal outcomes. N Engl J Med 2009;360: 111–20.
- Laughon SK, Albert PS, Leishear K, Mendola P. The NICHD Consecutive Pregnancies Study: recurrent preterm delivery by subtype. Am J Obstet Gynecol 2014;210:131.e1–8.
- Savitz DA, Blackmore CA, Thorp JM. Epidemiologic characteristics of preterm delivery: etiologic heterogeneity. Am J Obstet Gynecol 1991;164: 467–71.
- 22. White LJ, Waris M, Cane PA, Nokes DJ, Medley GF. The transmission dynamics of groups A and B human respiratory syncytial virus (hRSV) in England & Wales and Finland: seasonality and cross-protection. Epidemiol Infect 2005;133:279–89.
- Olson ME, Diekema D, Elliott BA, Renier CM. Impact of income and income inequality on infant health outcomes in the United States. Pediatrics 2010;126:1165–73.
- 24. Kirkegaard JK. Structural and cyclical trends in net employment over us business cycles, 1949–2009: Implications for the next recovery and beyond. Washington, DC: Peterson Institute for International Economics, 2009.
- Centers for Disease Control and Prevention. Vital signs: teen pregnancy – United States, 1991–2009. MMWR 2011;60:414–20.
- Stoupel E, Sirota L, Osovsky M, et al. Monthly number of preterm births and environmental physical activity. J Basic Clin Physiol Pharmacol 2007;18:149–57.
- Davis GE Jr, Lowell WE. Solar cycles and their relationship to human disease and adaptability. Med Hypotheses 2006;67:447–61.
- 28. Fitzpatrick AM, Teague WG, Meyers DA, et al.; National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. J Allergy Clin Immunol 2011;127:382–389.e1–13.
- Bhaskar H, Hoyle DC, Singh S. Machine learning in bioinformatics: a brief survey and recommendations for practitioners. Comput Biol Med 2006;36:1104–25.

Articles

Byrnes et al.

- Bassil KL, Yasseen AS 3rd, Walker M, et al. The association between obstetrical interventions and late preterm birth. Am J Obstet Gynecol 2014;210:538.e1–9.
- MacDorman MF, Declercq E, Zhang J. Obstetrical intervention and the singleton preterm birth rate in the United States from 1991-2006. Am J Public Health 2010;100:2241–7.
- Behrman RE, Butler AS. Measurement of fetal and infant maturity. In: Preterm Birth: Causes, Consequences, and Prevention. Washington, DC: National Academies Press, 2007:55–83.
- Martin JA, Osterman MJK, Sutton PD. Are preterm births on the decline in the United States? Recent data from the national vital statistics system. NCHS data brief, no 39. Hyattsville, MD: National Center for Health Statistics, 2010.
- Blakely TA, Woodward AJ. Ecological effects in multi-level studies. J Epidemiol Community Health 2000;54:367–74.
- Lawson AB. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. 2nd edn. Boca Raton, FL: CRC Press, 2013:253–76.
- Neelon B, Anthopolos R, Miranda ML. A spatial bivariate probit model for correlated binary data with application to adverse birth outcomes. Stat Methods Med Res 2014;23:119–33.
- Centers for Disease Control and Prevention. Birth Data: National Vital Statistics System. http://www.cdc.gov/nchs/births.htm. 14 July 2014.
- Ava Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2009:460–1.