



The Harvard Center for Population and Development Studies

Working Paper Series

US racial/ethnic data for COVID-19 cases: Still missing-in-action

Nancy Krieger, PhD¹, Christian C. Testa, BS¹, Jarvis T. Chen, ScD¹

September 16, 2020

HCPDS Working Paper Volume 20, Number 1

The views expressed in this paper are those of the author(s) and do not necessarily reflect those of the Harvard Center for Population and Development Studies.

Affiliations

1) Author affiliations: All authors are affiliated with the Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health.

2) Corresponding author: **Nancy Krieger**, PhD, Professor of Social Epidemiology, American Cancer Society Clinical Research Professor, Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Kresge 717, Boston, MA, 02115; office: 617-432-1571; email: nkrieger@hsph.harvard.edu

3) Author contributions: NK conceptualized the study, contributed to locating and interpreting the data, and drafted the manuscript; JTC and CT both obtained, analyzed, and contributed to interpreting the data; all authors contributed to revising the text and approving the final version.

No IRB approval was required, since the study analyzed de-identified aggregated publicly available data.

Title: US racial/ethnic data for COVID-19 cases: still missing-in-action

Authors: Nancy Krieger, Christian Testa, Jarvis T. Chen,

Word Count: 385 words

Tables: 1

References: 5

Final Version: September 16, 2020

TEXT

On June 4, 2020, the Trump Administration, responding to criticisms regarding the incompleteness of racial/ethnic and other demographic data for COVID-19 test results, including as reported by the US Centers for Disease Control and Prevention (CDC), released new reporting requirements [1]. The guidance stated: “this information should be made available in all reporting (including through methods using existing technical infrastructure such as an HIE) to state and local public health departments and subsequently the CDC as soon as possible, but no later than August 1, 2020” [2].

To our knowledge, no reports have evaluated whether reporting of racial/ethnic data for US COVID-19 cases has improved since August 1, 2020. On September 15, 2020, i.e., 6 weeks after the data for cases was required by law, we checked the publicly available websites. Our key findings are that:

1) The current CDC demographic tracker for COVID-19 cases first reported data on August 28, 2020, at which time 2,263,017 (51%) of its 4,458,258 cases were missing data on race/ethnicity [3]. As of September 16, 2020, 2,445,731 (50%) of its 4,880,315 reported cases were missing these data [3]. Mathematically, this means that 43% of the 422,057 cases added between these two dates were missing racial/ethnic data.

2) The COVID-19 Racial Data Tracker, which uses publicly available data from state and local health departments, as well as the CDC, reported on September 13, 2020 that among the notably larger tally of 6,448,573 cases, 37.5% were missing data on race (Table 1); the range spanned from 0% (MN, SD, WV) to 100% (NY), with a median of 21% [4]. For ethnicity, the range of percent missing extended from 0% (MN, SD) to 100% (LA, ND), with a median of 24% [4]. However, several states do not appear to report any ethnicity data at all (WV, HI, NY) [4]. As also shown in Table 1, states that are the base for leaders of the US government – President, Vice President, Senate Majority Leader, and Speaker of the US House of Representatives – have upwards of a quarter to over a third of COVID 19 cases missing data on race and ethnicity.

These findings suggest that compliance with regulations to report data on race/ethnicity for US COVID-19 cases is inadequate and continues to hamper understanding of and efforts to mitigate racial/ethnic inequities in COVID-19 [5].

Table 1. COVID-19 cases: number and percent missing data on race, ethnicity, or race/ethnicity, by political geographic unit in the United States, as of September 13, 2020.			
Political geographic unit	N of cases	Percent missing data	
		Race	Ethnicity
US	6,448,573	38%	44%
Florida (state of President Trump)	663,994	26%	32%
Indiana (state of Vice President Pence, in charge of the federal COVID-19 task force)	105,804	22%	54%
Kentucky (state of Mitch McConnell, Senate Majority Leader)	56,945	35%	37%
California (state of Nancy Pelosi, Speaker of the US House of Representatives)	754,923	32%	32%
Source of data: reference 4			

REFERENCES

1. Weiland N, Mandavalli A. Trump Administration sets demographic requirements for coronavirus reports. New York Times, June 4, 2020. <https://www.nytimes.com/2020/06/04/us/politics/coronavirus-infection-demographics.html?searchResultPosition=2> ; accessed September 14, 2020.

2. US Dept of Health and Human Services. COVID-19 Pandemic Response, Laboratory Data Reporting: CARES Act Section 18115. June 4, 2020; p. 4.
<https://www.hhs.gov/sites/default/files/covid-19-laboratory-data-reporting-guidance.pdf> ; accessed September 14, 2020.
3. CDC COVID Data Tracker. Demographic trends of COVID-19 cases and deaths in the US reported to CDC. Updated September 15, 2020, 12:16 pm.
https://covid.cdc.gov/covid-data-tracker/?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcases-updates%2Fcases-in-us.html#demographics ; accessed September 15, 2020.
4. The Covid Tracking Project at the Atlantic. The COVID Racial Data Tracker.
<https://covidtracking.com/race/dashboard> ; accessed September 15, 2020.
5. Krieger N, Gonsalves G, Bassett MT, Hanage W, Krumholz HM. The fierce urgency of now: closing glaring gaps in US surveillance data on COVID-19. *Health Affairs Blog*, April 14, 2020.
<https://www.healthaffairs.org/doi/10.1377/hblog20200414.238084/full/> ; accessed September 14, 2020.

SUPPLEMENTAL MATERIALS FOR:

US racial/ethnic data for COVID-19 cases: Still missing-in-action
Nancy Krieger, PhD, Christian C. Testa, BS, Jarvis T. Chen, ScD

*Trends in missing data for race and ethnicity by US States and Census Divisions, using
data from the COVID-19 Racial Data Tracker (through September 13, 2020)*

Supplemental Material

In order to gain further insight into how different US state and territorial health departments are reporting race and ethnicity data in relation to COVID-19 cases and deaths, we accessed data from the COVID Racial Data Tracker.¹ The COVID Racial Data Tracker is a collaboration between the Atlantic's COVID Tracking Project and the Boston University Center for Antiracist Research. The project compiles race and ethnicity data based on what is reported by local and/or state and territorial public health departments on their public facing websites.²

We extracted data from the COVID Racial Data Tracker website through September 13, 2020.³ We calculated the proportion of cases with unknown race by dividing Cases_Unknown by Cases_Total. Similarly, we calculated the proportion of cases with unknown ethnicity by dividing Cases_Ethnicity_Unknown by Cases_Total. However, we noticed that in some cases, states and territories any not report any ethnic data at all and Cases_Ethnicity_Unknown was missing. In these situations, we set Cases_Ethnicity_Unknown to equal Cases_Total. We calculated similar quantities for the proportion of deaths with unknown race and unknown ethnicity.

As noted by the COVID Tracking Project, the data are extracted by volunteers and datasets are updated by hand once a day.⁴ As a result, and because state reporting practices may vary and data may be backfilled/backdated, the data can be messy. Individual state data may also be subject to data suppression in case of small numbers, in which case sums of racial/ethnic categories may not sum to total cases or deaths. We have not attempted to further clean the

COVID Racial Data Tracker data even where discrepancies exist, except as noted above for the proportion with ethnicity unknown when all ethnicity data are unreported. Occasionally this results in improbable number (for example, on August 1, North Dakota reports the proportion of cases with unknown race in excess of 100%).

To visualize temporal trends in the reporting of race and ethnicity by state, we have plotted the data by state grouped by US Census Division, as listed in Table S.1. We additionally plot data for US Territories and for the United States as a whole (based on aggregating the full dataset). Data on unknown race and ethnicity among cases are presented in Figures S.1 and S.2; data on unknown race and ethnicity among deaths are presented in Figures S.3 and S.4. Time series for states are colored to help distinguish them from one another; the color of the line has no inherent meaning, however. We also demarcate August 1 (the date by which states were mandated to report case data by race and ethnicity) with a dotted line.

Since population sizes vary substantially across states, we also present proportions of cases and deaths with unknown race and unknown ethnicity based on aggregating over states within US Census Divisions. These data are presented in Figures S.5(a)-(d).

Results

States continue to be variable in the completeness of their data by race and ethnicity. While isolated improvements in the completeness of race data for cases are seen for some states (e.g. Louisiana, North Dakota, Nebraska), other cases reporting of race data remains absent (e.g. NY) or has increased dramatically (e.g. Texas, Florida, Alabama). With the exception of Louisiana,

no state appears to have improved dramatically around the mandated deadline of August, and the overall completeness of race data for COVID-19 cases in the US as of September 13, 2020 was 62.3%. Data on ethnicity reporting are considerably messier but show similar patterns, with the percent with recorded ethnicity at 56.1% as of September 13, 2020.

Completeness of reporting of race in COVID-19 death data is considerably better across the states (Figure S.3) with 5.1% of reported COVID-19 deaths with unknown race across the United States and 7.8% with unknown ethnicity. Notably, the proportion of deaths with unknown race is higher than that reported in the CDC's dataset of Deaths involving coronavirus disease 2019 (COVID-19) by race and Hispanic origin group and age, by state.⁵ This may reflect the fact that the CDC dataset is based on confirmed death certificates reported to NCHS, which requires race/ethnicity to be recorded. The data in the COVID Racial Data Tracker is instead based on data extracted from public facing websites and may reflect deaths where filing of the death certificate is incomplete or state dashboards have not yet routinized the reporting of race/ethnicity.

Table S.1: US States and US Census Divisions

US Census Division	States
Division 1:	New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont)
Division 1:	Mid-Atlantic (New Jersey, New York, and Pennsylvania)
Division 1:	East North Central (Illinois, Indiana, Michigan, Ohio, and Wisconsin)
Division 1:	West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)
Division 1:	South Atlantic (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, and West Virginia)
Division 1:	East South Central (Alabama, Kentucky, Mississippi, and Tennessee)
Division 1:	West South Central (Arkansas, Louisiana, Oklahoma, and Texas)
Division 1:	Mountain (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming)
Division 1:	Pacific (Alaska, California, Hawaii, Oregon, and Washington)
Territories:	American Samoa, Guam, Northern Mariana Islands, Puerto Rico, US Virgin Islands

References

1. The COVID Tracking Project. COVID Racial Data Tracker, available at <https://covidtracking.com/race>. Accessed September 21, 2020.
2. The COVID Tracking Project. Data Sources, available at <https://covidtracking.com/about-data/sources>. Accessed September 21, 2020.
3. The COVID Tracking Project, About the Racial Data Tracker, available at: <https://covidtracking.com/race/about#download-the-data>. Accessed September 15, 2020.
3. The COVID Tracking Project. Data FAQ, available at <https://covidtracking.com/about-data/faq>. Accessed September 21, 2020.

Figure S.1: Trend in proportion of cases with unknown race by US State and Census Division

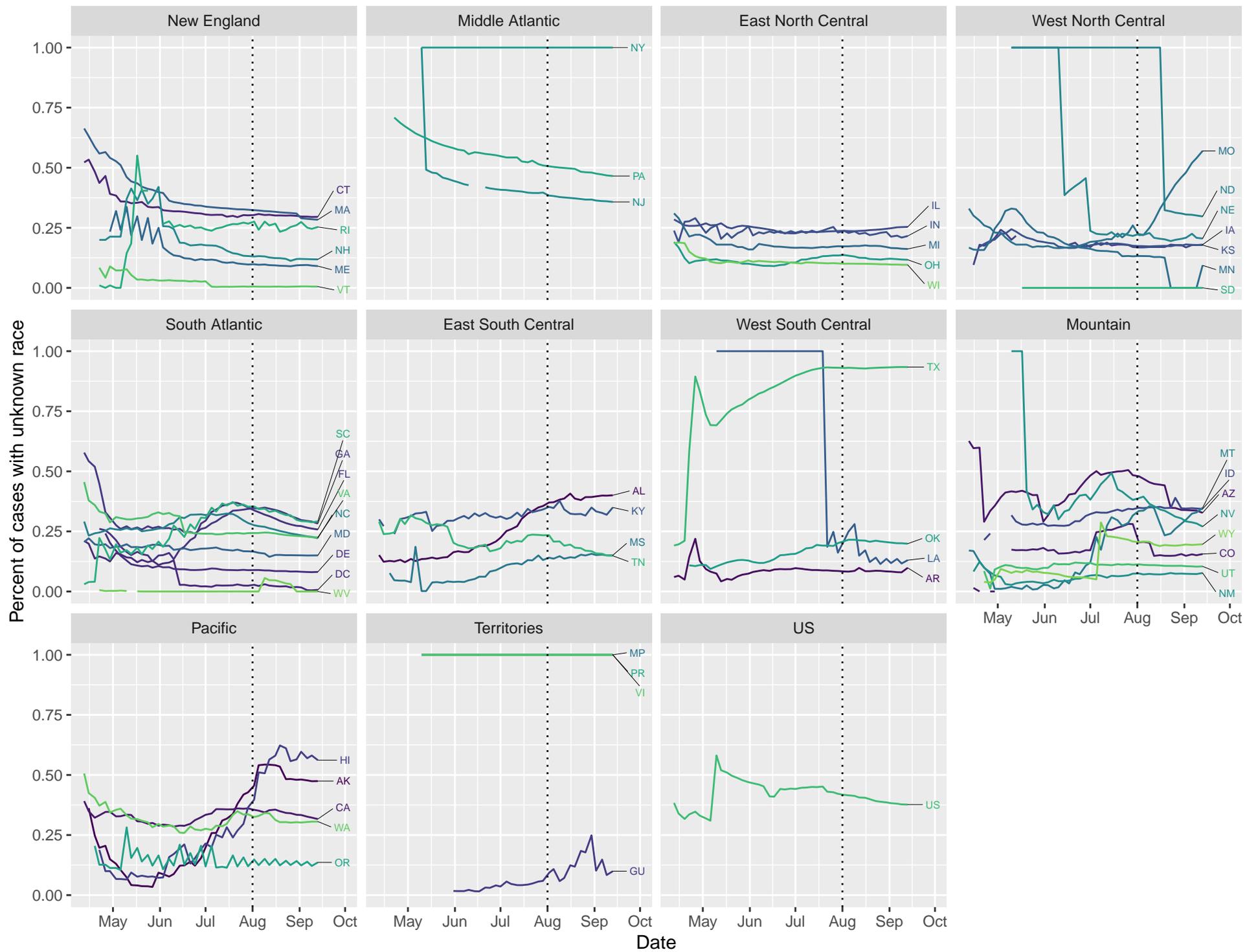


Figure S.2: Trend in proportion of cases with unknown ethnicity by US State and Census Division

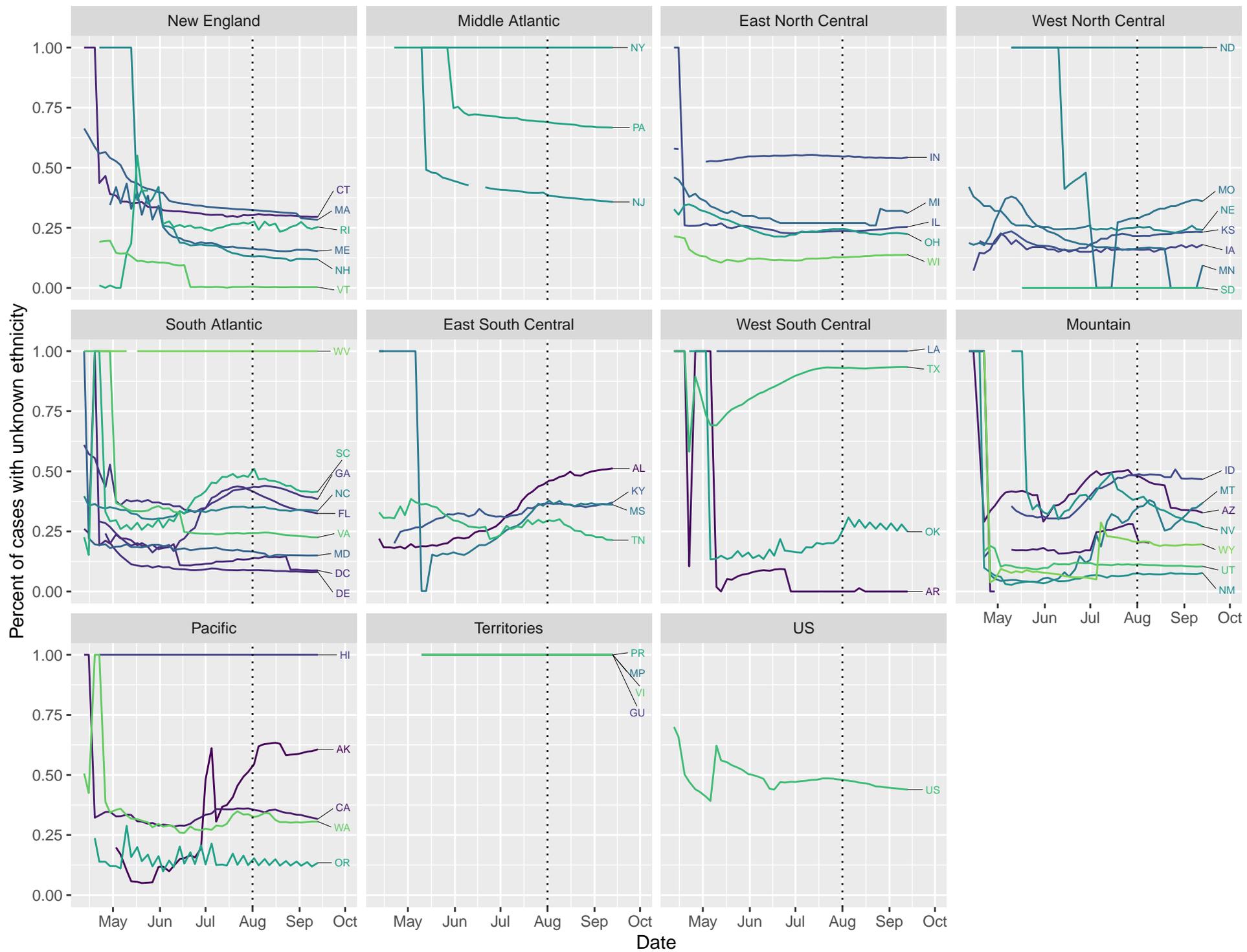


Figure S.3: Trend in proportion of deaths with unknown race by US State and Census Division

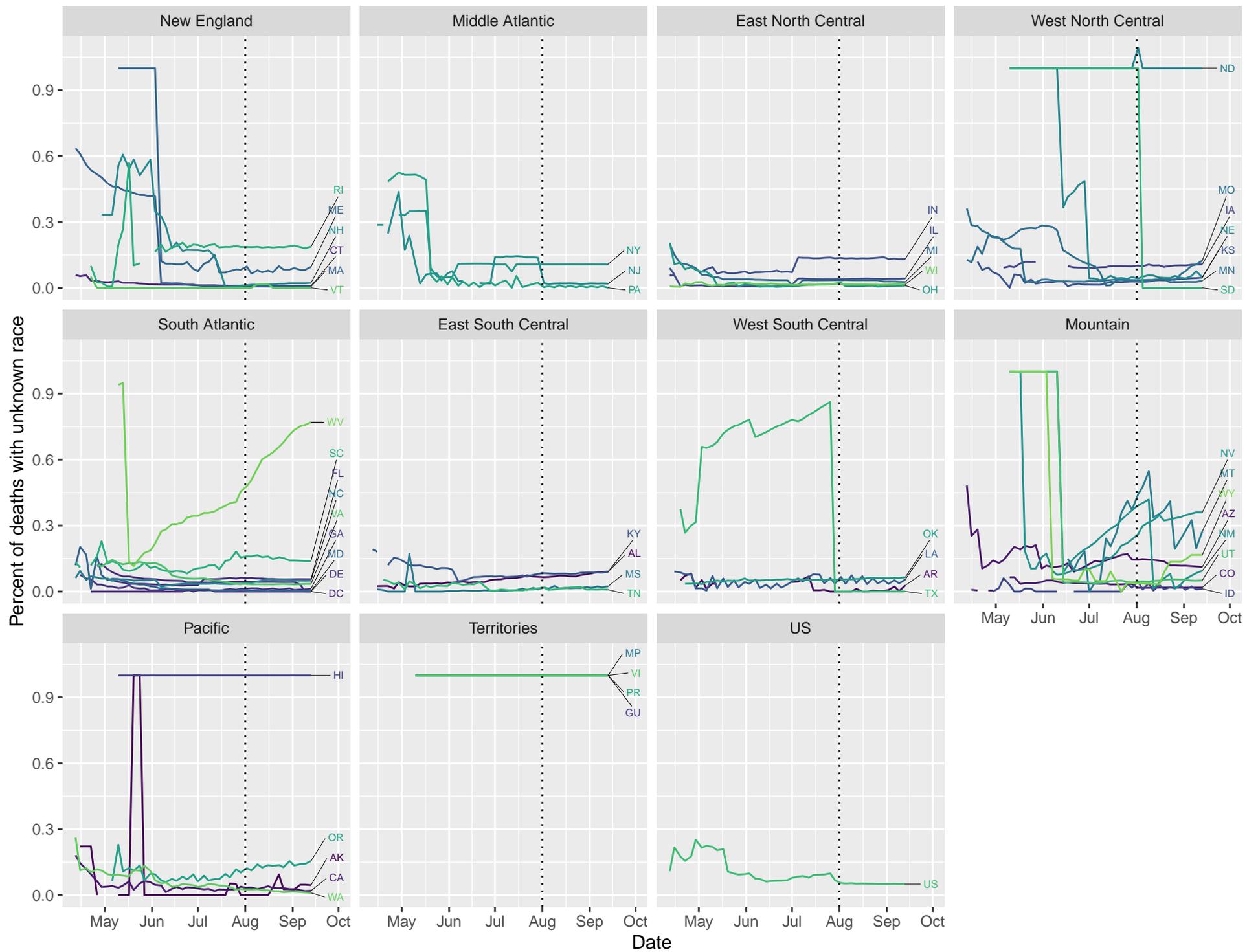


Figure S.4: Trend in proportion of deaths with unknown ethnicity by US State and Census Division

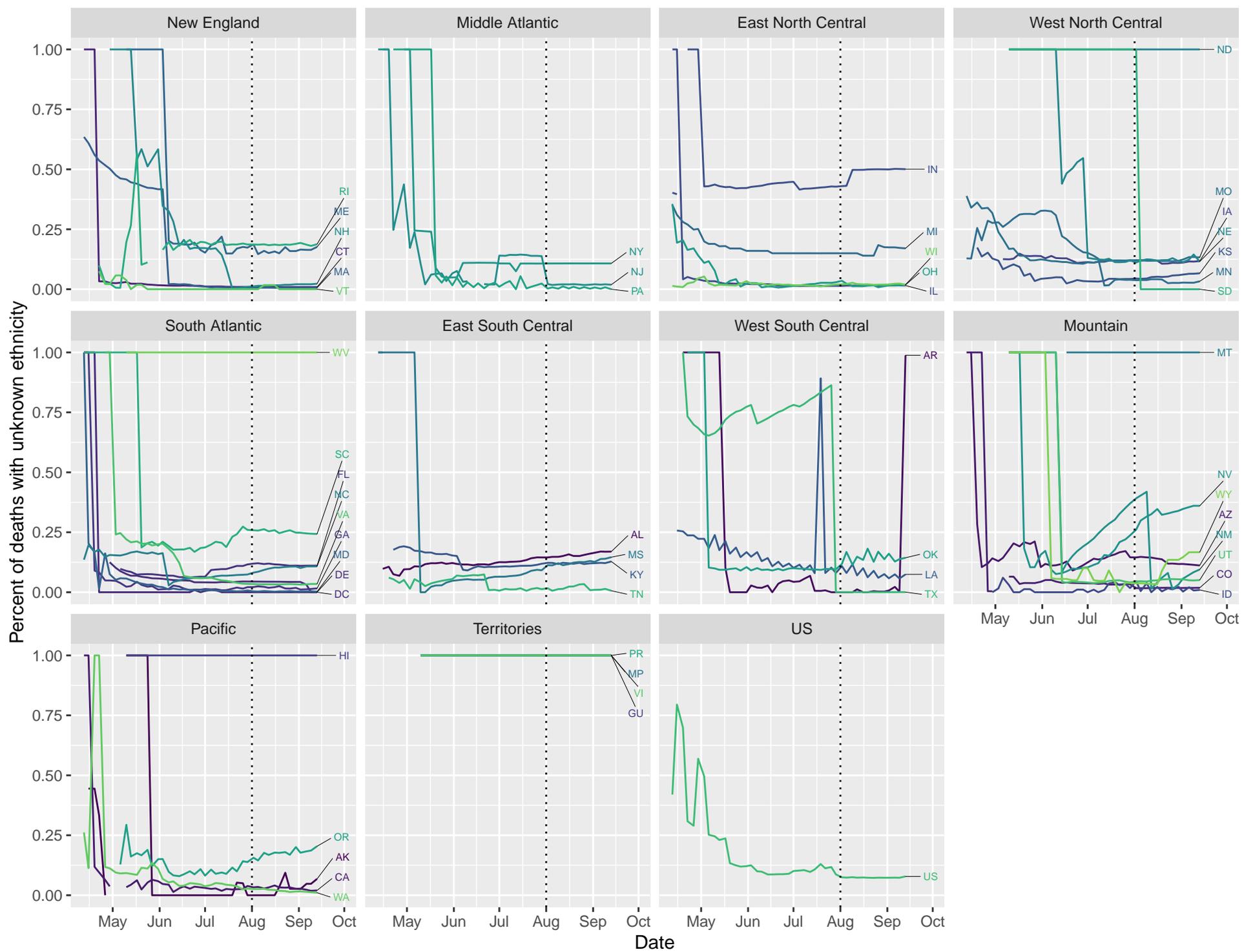


Figure S.5(a): Trend in proportion of cases with unknown race by US Census Division

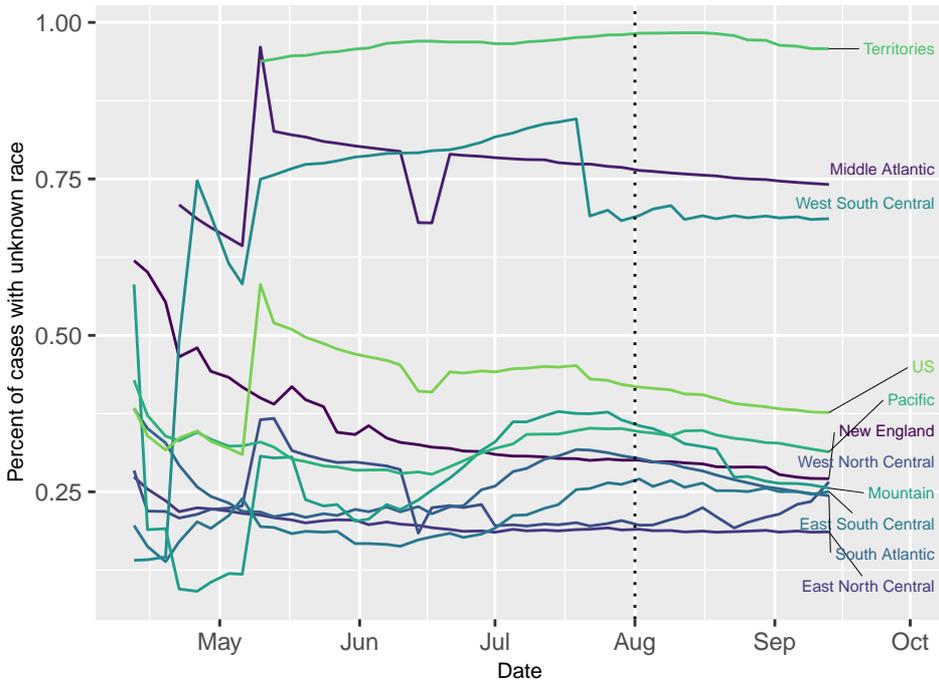


Figure S.5(b): Trend in proportion of cases with unknown ethnicity by US Census Division

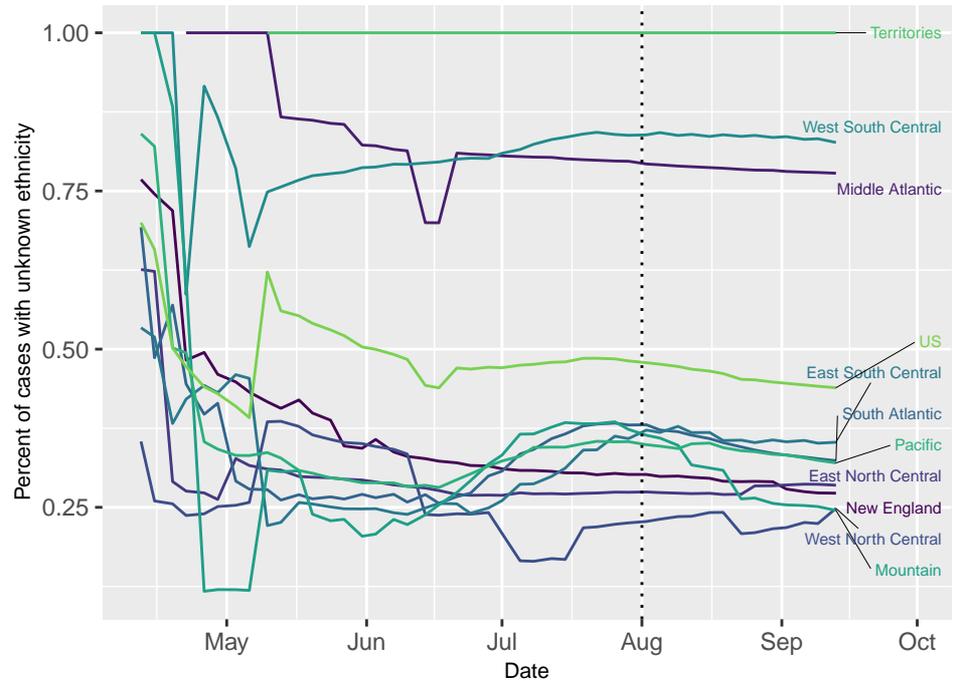


Figure S.5(c): Trend in proportion of deaths with unknown race by US Census Division

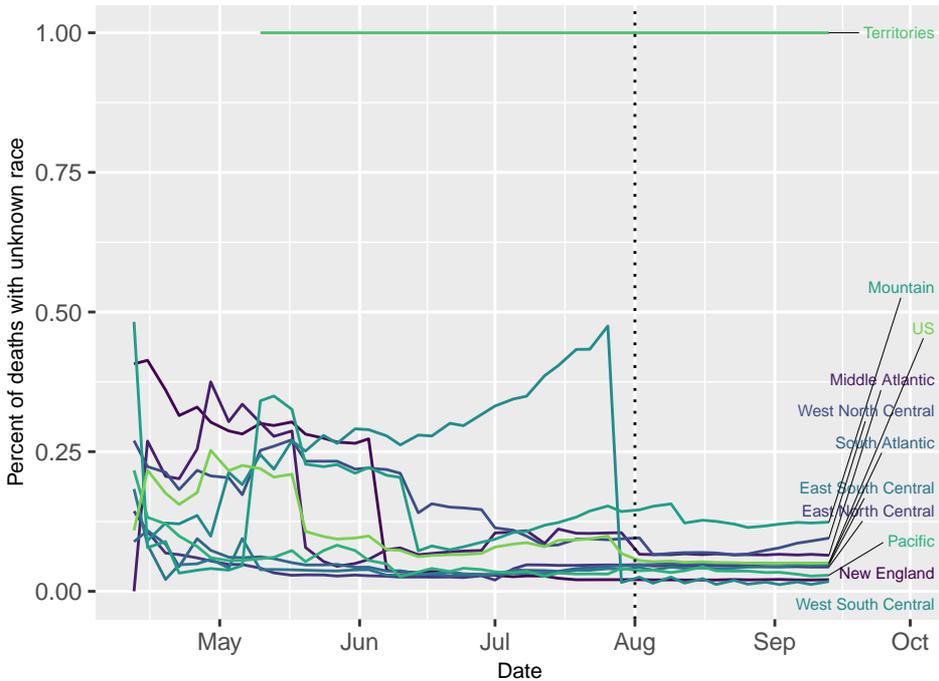


Figure S.5(d): Trend in proportion of deaths with unknown ethnicity by US Census Division

