Understanding and Addressing Missing Race and Ethnicity Data in the Electronic Health Record

Summary of Final Project Results | July 2023

SONNET
Social Needs Network for Evaluation and Translation

Why we did our project

Data about race and ethnicity are more likely to be missing from the electronic health record (EHR) compared to other demographic data like sex and insurance status. Improving the accuracy and completeness of race and ethnicity data in the EHR helps us better understand racial and ethnic health disparities — and increases our ability to address them.

What we did

We interviewed 23 experts to explore experiences and approaches for addressing missing race and ethnicity data.

- 21 experts were from Kaiser Permanente (KP).
- 2 experts were from external organizations.

10 Social health experts

4 Clinical health experts

7 Statisticians & analysts

2 Health equity experts

Our interview guide covered the following content areas:

- Experience with and use of race and ethnicity data
- Analytic strategies
- Health disparities research
- Race and ethnicity data in clinical encounters
- HR Membership Services

We conducted interviews from February to May 2022. Each interview was audio recorded, transcribed, and compared to the original recording for accuracy. We used NVivo Software to analyze and group text from the interviews based on pre-existing descriptive labels.

What we learned

Recommended strategies to address and prevent missing data

- ❑ Consistent data collection
- ❑ Clinical staff trained in collecting and retrieving data
- ❑ Systems ready to catch missing race and ethnicity data in the EHR

> "I still think that we're in a place in our society where although the tide seems to be changing and people seem to be paying more attention to it, I am still constantly having to pitch this work and the importance of this work, to really get it embedded, deep, deep, deep into our processes." – Social health expert

> "But I do know from a practical standpoint that the point of data collection, whether a clinician or from membership during application, may actually cloud, bias, slant, tint the data that we have." – Statistician

How we can use this work to advance social health practice at KP and beyond

Our results point to solutions for addressing missing with race and ethnicity data at multiple levels. (See our more detailed results and recommendations on pages 2-3 of this summary.)

Staff level

Train providers and staff to collect and correct race and ethnicity data

Administrative level

Fix systems that erase and replace data in Health Connect monthly, like employer updates through HR system

Organizational level

Improve members' comfort by helping them understand **why** KP wants to correct and include race and ethnicity data

Project lead: Teaniese "Tina" Davis, PhD, MPH
KP Georgia Center for Research and Evaluation

KAISER PERMANENTE®

Key findings and recommendations

Data collection

- Data collection is different across KP regions and ranges above and below KP standard of 80%.
- Staff collection/correction of member race and ethnicity data differs by and within KP regions.
- KP regions collected data by phone and survey in the past.
- Most data are collected during a care visit.
- Many collection systems do not connect to each other; monthly employer-based data erases any corrections manually entered by KP employees and replaces with prior information.

Solutions and recommendations

- ❑ Review for missing data
- ❑ Train staff in skills that improve and increase their comfort with asking about race and ethnicity
- ❑ Create similar data collection tools across departments and KP regions
- ❑ Prevent employer-based data from erasing data manually corrected by KP employees in the EMR

Statistical methods

- Statistical methods like Bayesian Improved Surname Geocoding (BISG) are used to fill in missing data by using existing data (last name and residential address) to predict a person's race and ethnicity.
- Use of KP Geographically Enriched Member Sociodemographic (GEMS) data (a CESR table).
- Race is labeled as "unknown" for people missing race or ethnicity or both data.

Solutions and recommendations

- ❑ Do not include missing data in analyses
- ❑ Predict missing data based on other characteristics
- ❑ If race and ethnicity are not a key part of an analysis, include people with missing data as "unknown"
- ❑ Continue to use statistical methods like BISG, but also clearly note why it is problematic to predict race and ethnicity using surnames and addresses

Member concerns

Based on staff perspectives of member concerns

- Members are unsure about how their race and ethnicity data will be used (e.g., denying care, immigration concerns) based on historical harms and privacy concerns.
- Members may feel sharing race and ethnicity will lead to different treatment, discrimination, or deportation.
- Some members give KP their race and ethnicity data without hesitation.

Solutions and recommendations

- ❑ Train staff to answer member questions about race and ethnicity and have race-based discussions
- ❑ Use videos and role play to train staff in the best ways to collect data (very different across departments and KP regions)

Race & ethnicity data grouping

- Race and ethnicity pairings are dynamic and changing.
- Grouping race and ethnicity categories can hide disparities and do harm (e.g., combining Asian and Pacific Islander members).
- Analyses should be used to look at differences between race and ethnicity categories across patient characteristics.
- People select race and ethnicity categories available to them (may not be their real race or ethnicity).
- KP data has 277 categories for race and ethnicity, with regional differences in data grouping and collection.

Solutions and recommendations

- ❑ Use more categories that separate smaller race and ethnicity groups (e.g., Pacific Islander, Alaskan Native)
- ❑ Use multiple data sources (EMR and surveys) to capture race and ethnicity
- ❑ When sharing data about a small number of people (small cell size), acknowledge the general makeup of race and ethnicity groups in the sample
- ❑ More publications to cite in the areas of race and ethnicity reporting and analyses

Staff trainings

- Staff do not feel prepared and comfortable asking questions about members' race and ethnicity.
- Staff have a low level of comfort talking about race and ethnicity in general.
- Leadership understands importance of better data collection and staff training. But solutions requiring a software or IT change present a large financial barrier.

Solutions and recommendations

- ❑ Train staff to ask questions about race and ethnicity and have race-based discussions
- ❑ Support staff experiencing microaggressions, anger, and direct discrimination from members; de-escalation training
- ❑ Use tools to support conversations between providers, staff, and members including providers sharing about themselves

Acknowledgments

Funding: This project was funded by the Kaiser Permanente (KP) Social Needs Network for Evaluation and Translation (SONNET). SONNET is a learning network that is committed to supporting KP leaders, clinicians, and staff as they design and evaluate effective interventions to address the social needs of our members. The Network includes researchers and evaluators from KP's 8 regional markets, KP's Bernard J Tyson School of Medicine, KP's Office of Community Health, and from the KP and Robert Wood Johnson Foundation-supported Social Interventions Research and Evaluation Network (SIREN). SONNET is supported by KP's Office of Community Health.

Project team: Teaniese "Tina" Davis, PhD, MPH; Roxan Andre, MPH; Breana Berry, MPH; Rachel Neenan, MPH