Better Health through Data: Opportunities and Limitations of Health Data

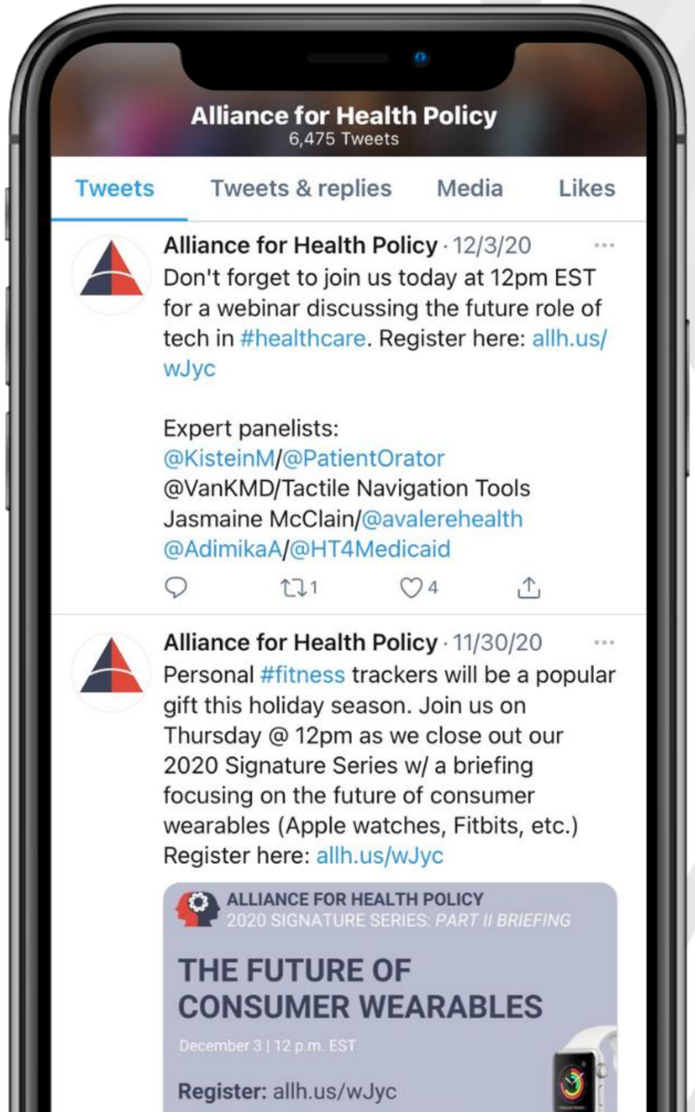
OCTOBER 6, 2022

ALLIANCE FOR HEALTH POLICY | 30 YEARS

JOIN THE CONVERSATION

@AllHealthPolicy

Alliance for Health Policy



Alliance for Health Policy
6,475 Tweets

Alliance for Health Policy · 12/3/20
Don't forget to join us today at 12pm EST for a webinar discussing the future role of tech in #healthcare. Register here: allh.us/wJyc

Expert panelists:
@KisteinM/@PatientOrator
@VanKMD/Tactile Navigation Tools
Jasmaine McClain/@avalerehealth
@AdimikaA/@HT4Medicaid

Alliance for Health Policy · 11/30/20
Personal #fitness trackers will be a popular gift this holiday season. Join us on Thursday @ 12pm as we close out our 2020 Signature Series w/ a briefing focusing on the future of consumer wearables (Apple watches, Fitbits, etc.)
Register here: allh.us/wJyc

ALLIANCE FOR HEALTH POLICY
2020 SIGNATURE SERIES: PART II BRIEFING

THE FUTURE OF CONSUMER WEARABLES

December 3 | 12 p.m. EST

Register: allh.us/wJyc

#AllHealthLive

PARTICIPATING

To mute yourself, click the microphone icon. The icon will appear orange when muted.

To ask a question, click the ? icon and enter your question in the chat box below.

PARTNERS

ALLIANCE
FOR HEALTH POLICY

Arnold
Ventures

@ArielleMir02

Arielle Mir,

Vice President of Health Care (Complex Care)

Arnold Ventures

SPEAKERS

Samantha Artiga, MHSA
Director, Racial Equity and Health Policy Program
KFF
@KFF | @SArtiga2

Elizabeth Lukanen, MPH
Deputy Director
State Health Access Data Assistance Center (SHADAC)
@SHADAC

Niall Brennan, MPP
Chief Analytics and Privacy Officer
Clarify Health
@ClarifyHealth | @N_Brennan

Irene Dankwa-Mullan, M.D., MPH
Chief Health Equity Officer, Deputy Chief Health Officer, Clinical Affairs
Merative
@Merative | @DankwaIrene

@SArtiga2

Samantha Artiga, MHSA

Vice President and Director,
Racial Equity and Health Policy Program
KFF

Data and Health Equity

Samantha Artiga
Vice President and Director, Racial Equity and Health Policy
Kaiser Family Foundation

KFF
Filling the need for trusted information on national health issues.

Figure 9

KFF Racial Equity and Health Policy Program

- **KFF** (Kaiser Family Foundation) is a nonprofit organization focusing on national health issues, as well as the U.S. role in global health policy
  - Policy analysis
  - Polling/survey research
  - Journalism (Kaiser Health News)
  - Public health information campaigns (Greater Than COVID)
- The **KFF Racial Equity and Health Policy Program** focuses on the intersection of racism and discrimination, social and economic inequities, and health
  - Timely and reliable data and analysis of health and health care disparities
  - Education to increase awareness and understanding of disparities
  - Analysis of implications of policy changes on disparities and efforts to advance equity

Figure 10

Comprehensive High-Quality Data is Central to Efforts to Advance Equity

- Having a complete picture of the status of disparities
- Understanding how intersectional factors drive disparities
- Guiding resources and efforts to advance equity
- Measuring progress and establishing accountability toward achieving equity
- Identifying best practices or strategies to advance equity

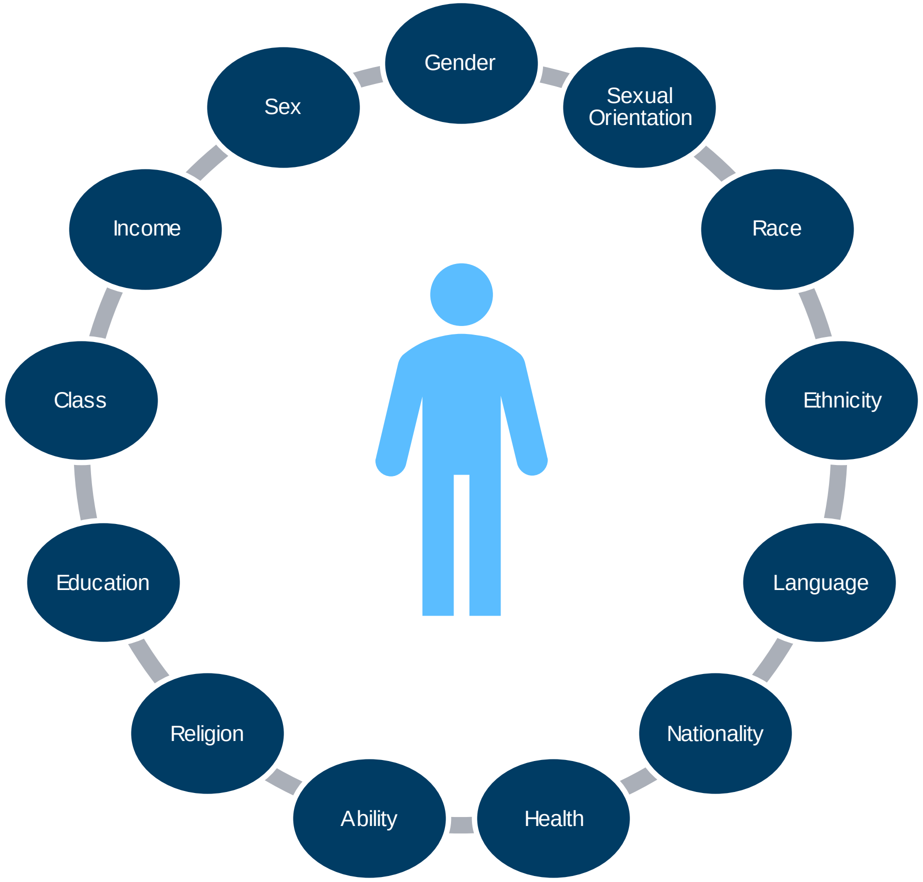


Figure 11

Key Sources of Data for Health and Health Care Disparities

- Survey Data
  - Federal surveys (e.g., American Community Survey, Current Population Survey, National Health Interview Survey, Behavioral Risk Factor Surveillance System)
  - State, local, or private organization surveys
- Administrative Data
  - Federal data (e.g., Centers for Disease Control and Prevention, Centers for Medicare and Medicaid Services (T-MSIS Analytic Files for Medicaid and CHIP))
  - State and local level data
- Claims and clinical encounter data

Figure 12

Federal Minimum Race/Ethnicity Classification Standards

- Minimum race categories
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Native Hawaiian or Other Pacific Islander
  - White
- Minimum ethnicity categories
  - Hispanic or Latino
  - Not Hispanic or Latino
- Additional granularity is encouraged where supported by sample size

Figure 13

Gaps and Limitations in Racial/Ethnic Health Data

- Lack of publicly reported data disaggregated by race/ethnicity in some datasets
- Insufficient data for smaller population groups and subgroups of broad racial/ethnic categories
- Inconsistencies in racial/ethnic classifications
- Growing shares of people categorized in multiracial or “other” categories
- Incompleteness of data (e.g., missing/unknown race/ethnicity)
- Lags in data reporting
- Inability to conduct intersectional analysis across multiple factors

Figure 14

Data Gaps and Limitations: COVID-19 Vaccination Data

- Federal Data
  - CDC has not reported state-level data on COVID-19 vaccinations by race/ethnicity
  - Race/ethnicity missing for roughly a quarter of vaccinations
  - Relatively high shares of vaccinations categorized as Multiple/Other race
  - Data not reported in a way that allows for analysis by multiple demographic factors (e.g., race/ethnicity and age or race/ethnicity and gender)
- State Data
  - Some states have not reported vaccination data by race/ethnicity
  - Racial and ethnic classifications vary across states
  - Many states have not reported data for AIAN or NHOPI people
  - Few states have reported data in a way to allow for analysis by race/ethnicity and age (e.g., vaccinations by race/ethnicity among children)

Figure 15

Data Gaps and Limitations: Medicaid Administrative Data

- T-MSIS Analytic files provide beneficiary-level data for Medicaid and the Children's Health Insurance Program, including demographic information and utilization data
- All Medicaid agencies collect self-reported data on race, ethnicity, and language during the enrollment process, but beneficiaries are not required to report it
- States are required to follow the federal minimum standard, but states may opt to collect more granular data
- There are challenges with the completeness and quality of the data
  - Missing race/ethnicity data. A dozen states had data that differed from benchmark survey data by >10% (SHADAC analysis)
  - The race/ethnicity data quality for 36 states is considered medium concern or above, with 22 states having data classified as high concern or unusable (DQ Atlas)
  - There may be additional data quality issues with the "multiracial" variable (CMS)
  - Quality has not been assessed for "language" data

Figure 16

Implications of Data Gaps and Limitations: Examples

- Inability to examine racial disparities in COVID-19 vaccination rates for children
- Limited understanding of health and health care experiences for smaller groups and subgroups within broader racial/ethnic categories
- Lack of data to identify disparities in coverage losses when the public health emergency ends
- Limited insight into how Medicaid managed care quality varies by race/ethnicity

Figure 17

Addressing Data Gaps and Limitations

- Prioritizing data collection and reporting
- Increasing availability and accessibility of public data
- Expanding data for smaller groups and subpopulations
- Modernizing data collection and reporting systems and structures
- Engaging community in design and implementation of data collection and reporting standards
- Increasing standardization of data reporting
- Conducting outreach and education to providers and individuals about importance of data collection and reporting

@SHADAC

Elizabeth Lukanen, MPH

Deputy Director

State Health Access Data Assistance Center (SHADAC)

30 YEARS

shadac

STATE HEALTH ACCESS DATA ASSISTANCE CENTER

Better Health through Data: Improved Collection of Demographic Data in Medicaid

Elizabeth Lukanen, MPH

Alliance for Health Policy Virtual Event

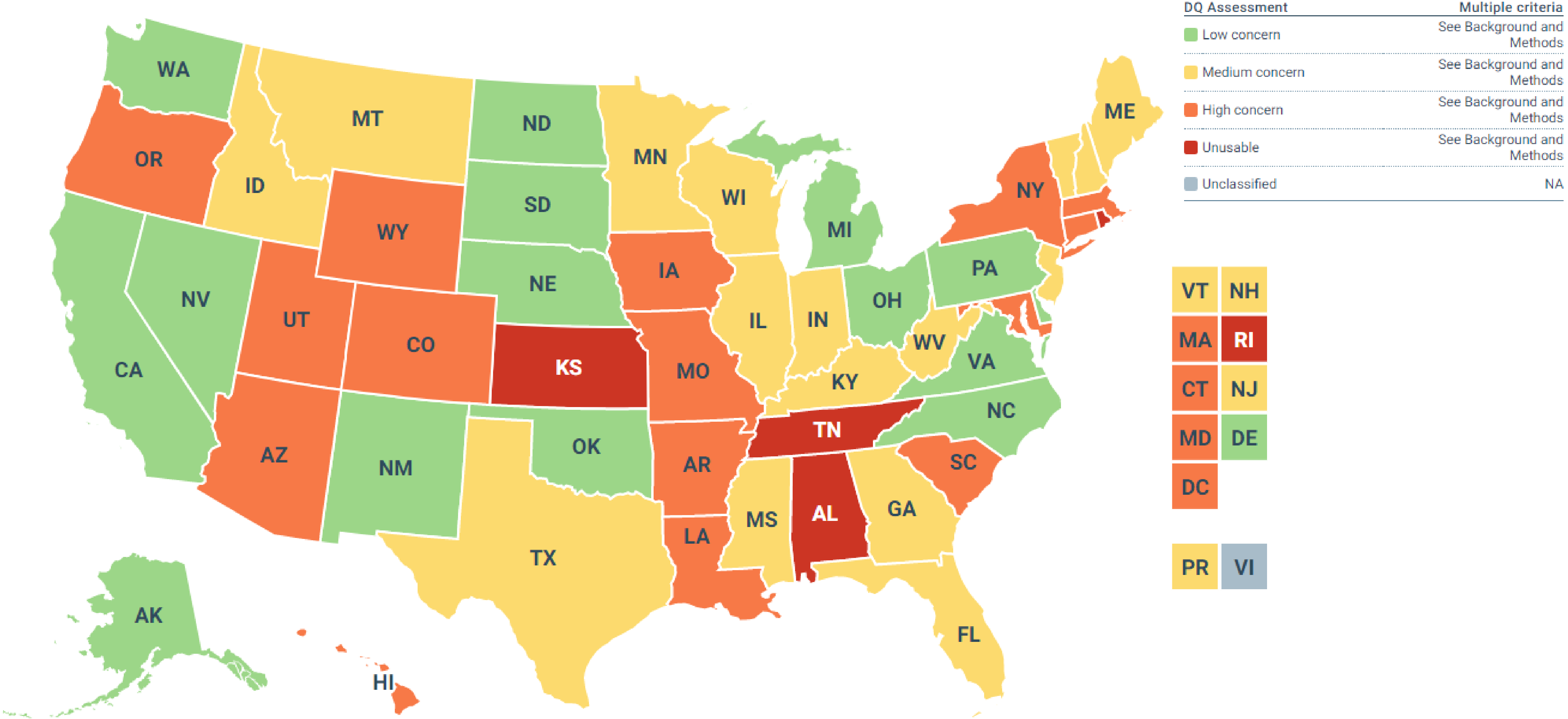
Better Health through Data: Opportunities and Limitations of Health Data

October 6, 2022

Reasons for Missing Race and Ethnicity Data in Medicaid

- **Guidance is not consistent with most recent evidence**
  Guidance recommends a two-question series with only a Hispanic race option. Evidence indicates that a combined race and ethnicity question—as well as offering "Middle Eastern or North African" (MENA) as a response option—decreases nonresponse and improves accuracy.
- **Voluntary reporting**
  States cannot require race and ethnicity information as a condition of eligibility.
- **Question design**
  The question wording, along with the broad racial and ethnic response option categories, may not resonate with the local population.
- **Mistrust about how data will be used**
  There are concerns about how data collection might relate to immigration status, be used to discriminate, etc.

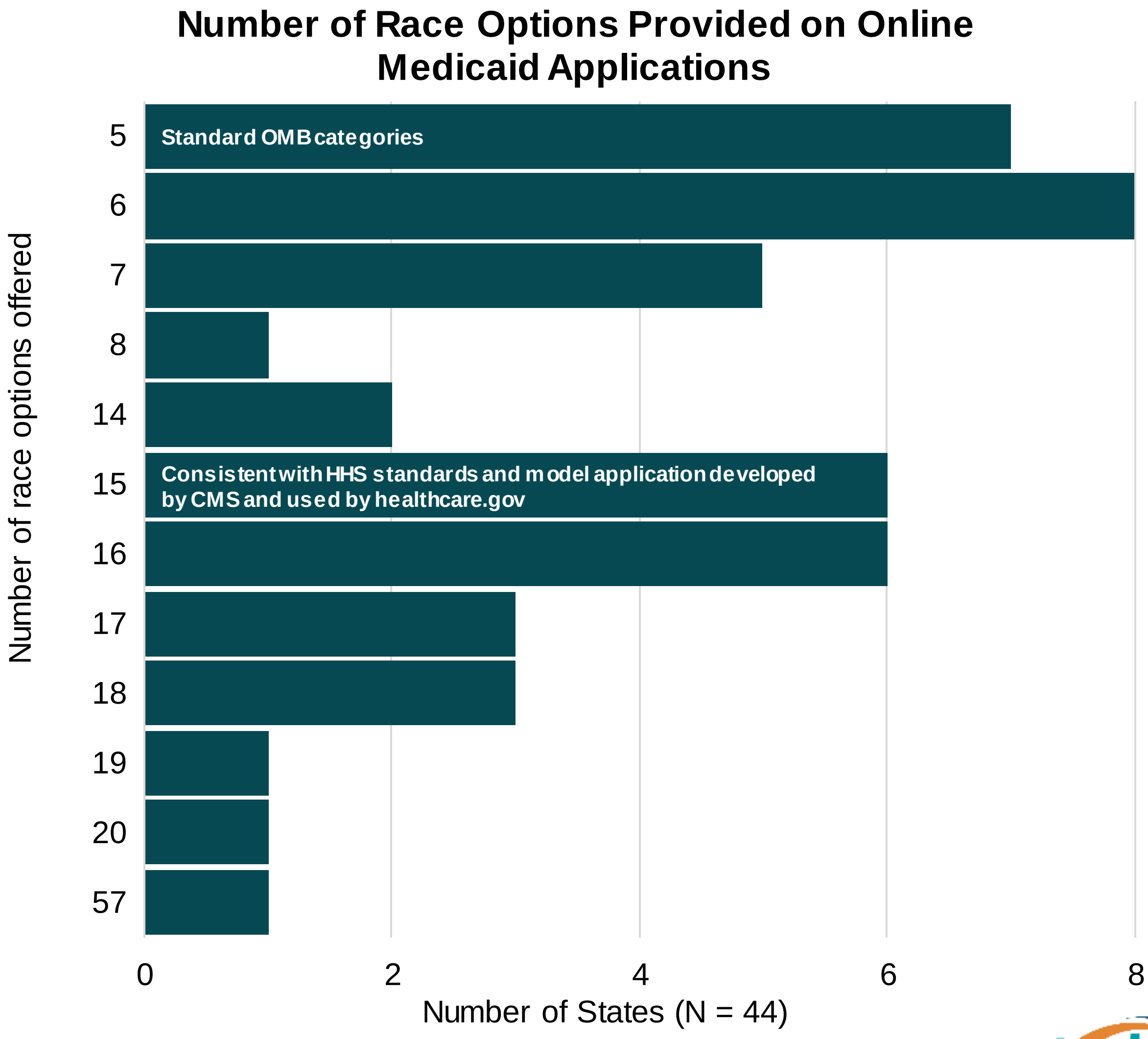
Race and Ethnicity Data Quality in 2020 T-MSIS Analytic File (TAF)



Source: Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2021c. DQ Atlas background and methods: Race and ethnicity. https://www.medicaid.gov/dq-atlas/landing/topics/single/map?topic=g3m16&tafVersionId=25

Findings from 50-State Medicaid Application Review

- Majority of states’ applications have two separate questions to ascertain race and ethnicity
- Across all 50 states, we identified 62 variations of racial response options
- Only Maryland, Oregon, and Vermont offered alterative race options such as Middle Eastern, Northern African, Eastern and Western European, etc.



Source: Forthcoming revision of. https://www.shvs.org/resource/collection-of-race-ethnicity-language-rel-data-in-medicaid-applications-a-50-state-review-ofthe-current-landscape/

Data Improvement: Enhance Existing Data

- **Leverage alterative data sources:** Identify individuals in other systems (other public programs, health records, provider data, etc.) and pull race and ethnicity data from those systems
- **Imputation:** Fill in gaps using surname and/or geography-based imputation
- **Data validation:** Compare distribution against data from the American Community Survey

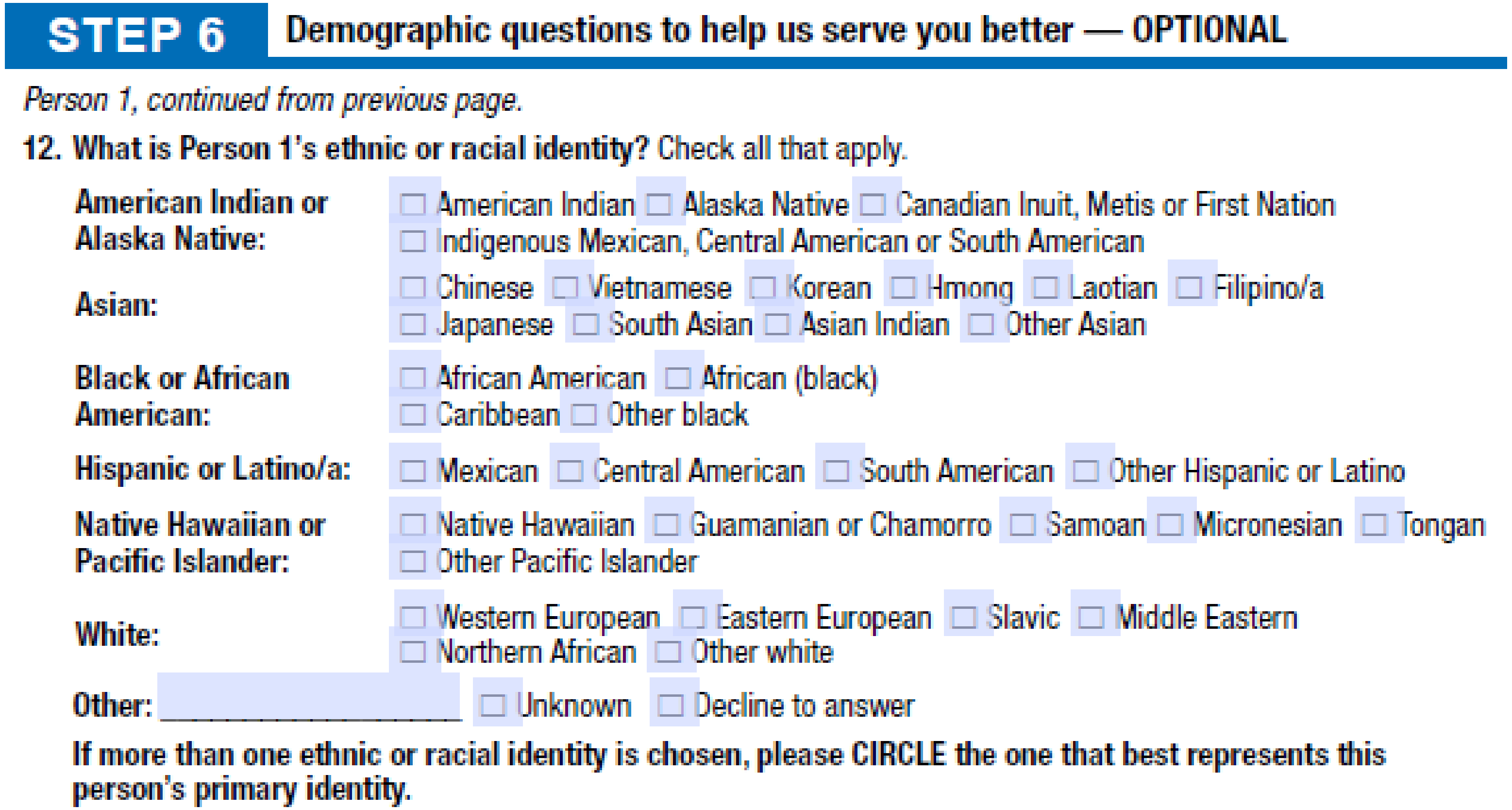
States tend to focus on improving existing data because it's easier than other tactics and doesn't contradict or conflict with federal guidance.

Data Improvement: Activities Aimed at Improving Response Rate and Quality

- **Mandatory response:** For online forms, require enrollees to provide an answer before proceeding to the next screen, including options such as “decline to answer” or “unknown”
- **Improve question format:** Use a single, combined question with response options that are informed by knowledge of population and community input
- **Engage enrollees:** Proactively seek advice from stakeholders about why data are not being provided and how to improve data collection
- **Engage navigators/enrollment assisters:** Develop communications and training strategies focused on importance of data collection
- **Post-eligibility activities:** Prompt for missing race and ethnicity at every login or at renewal

States are starting to consider these activities in response to other improvement efforts failing and communities demanding better data.

Oregon Informed Question Structure



STEP 6 **Demographic questions to help us serve you better — OPTIONAL**

Person 1, continued from previous page.

12. What is Person 1's ethnic or racial identity? Check all that apply.

American Indian or Alaska Native:	☐ American Indian ☐ Alaska Native ☐ Canadian Inuit, Metis or First Nation ☐ Indigenous Mexican, Central American or South American
Asian:	☐ Chinese ☐ Vietnamese ☐ Korean ☐ Hmong ☐ Laotian ☐ Filipino/a ☐ Japanese ☐ South Asian ☐ Asian Indian ☐ Other Asian
Black or African American:	☐ African American ☐ African (black) ☐ Caribbean ☐ Other black
Hispanic or Latino/a:	☐ Mexican ☐ Central American ☐ South American ☐ Other Hispanic or Latino
Native Hawaiian or Pacific Islander:	☐ Native Hawaiian ☐ Guamanian or Chamorro ☐ Samoan ☐ Micronesian ☐ Tongan ☐ Other Pacific Islander
White:	☐ Western European ☐ Eastern European ☐ Slavic ☐ Middle Eastern ☐ Northern African ☐ Other white

Other: ____________________ ☐ Unknown ☐ Decline to answer

If more than one ethnic or racial identity is chosen, please CIRCLE the one that best represents this person's primary identity.

Oregon's question structure is the result of ongoing assessment and improvement efforts that started in 2014 and include community-based engagement and education. They monitor results and have seen a decline in missing data.

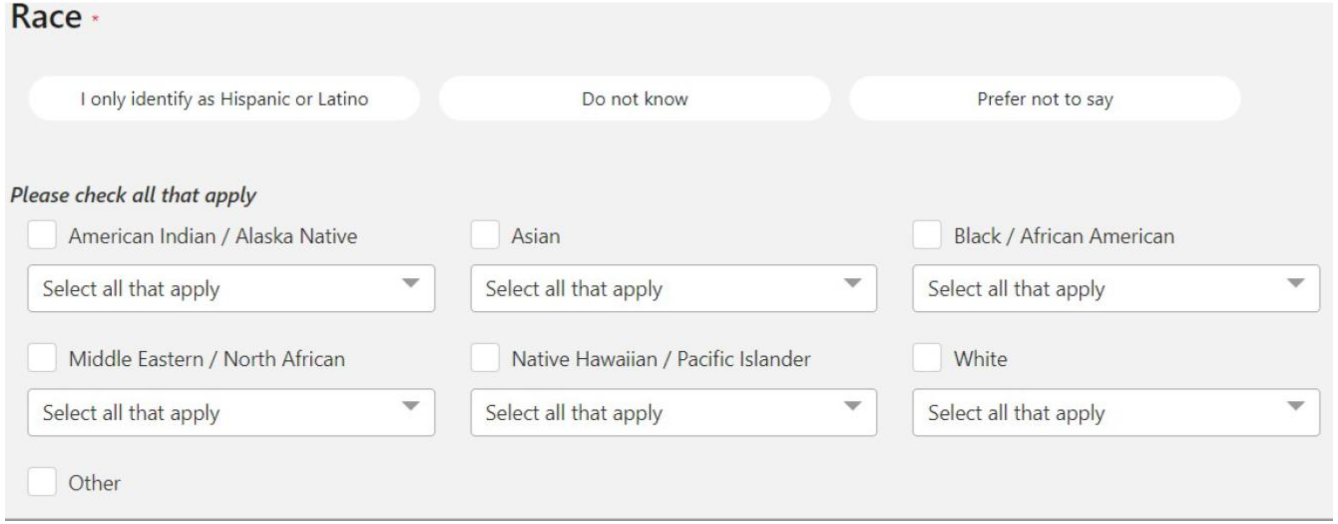
New York State of Health Pilot to Improve Race and Ethnicity Response Rates

- **Education**: Enhanced explanation on the importance of the question for applicants
- **Training:** Provided new training for assistors and navigators
- **Requiring a response*:*** Applicants did not have to share their race or ethnicity, but they could not leave the question blank; instead, they could respond with "don't know" or "choose not to answer"
- **Results:** Race response rate grew by 20 percentage points, and ethnicity grew by 8 percentage points; response rates for a comparison group saw minimal change
- **Expanding changes:** New York is now expanding changes to the race and ethnicity questions system-wide

New York has worked with application assisters who in turn engage and educate respondents to improve data quality.

Maryland Enhanced Race and Ethnicity Response Options

- **Leverage data:** Analyzed U.S. Census Bureau data to identify state-specific racial and ethnic groups and non-English primary language clusters
- **Engage stakeholders:** Connected with state staff and other experts who have a strong understanding of the relevant communities and the language/terminology they use



Race *

Please check all that apply

☐ American Indian / Alaska Native

☐ Asian

☐ Black / African American

☐ Middle Eastern / North African

☐ Native Hawaiian / Pacific Islander

☐ White

☐ Other

Through data analysis and consultation with community experts, the state increased the number of racial response options it provides on its online application from 15 to 57.

Good Data Starts with Trust

- At the end of the day, in order to ask individuals for personal information about their history, lives, and how they identify as a person, trust and good communication is needed.
    - Community input should shape questions and response options.
- Respondents should know how their information will be used and have that output shared back with them.

Examples from instructional language on Medicaid applications:

👎 **What is your race? (optional)**

👍 **California:** Tell us about your race. This information is confidential and will only be used to make sure that everyone has the same access to health care. It will not be used to decide what health insurance you qualify for.

👍 **New Hampshire:** Providing information about ethnicity and race is voluntary and will not affect your benefits. The reason we ask for the information is to assure that program benefits are distributed without regard to race, color, or national origin.

👍 **Wisconsin:** You don't have to answer these questions. We are asking these questions to improve our programs and make sure they are fair for all members, no matter what ethnicity or race they are. Your answers will not be used to make decisions about your programs or benefits.

Resources

SHADAC's understating of this issue is based on technical assistance in addition to an ongoing convening of states focused on improving race, ethnicity, and language data supported by State Health and Value Strategies (SHVS).

- SHVS has created an accessible one-stop source of health equity information aimed at states:

Collection of Race, Ethnicity, Language (REL) Data in Medicaid Applications:
A 50-state Review of the Current Landscape

Issue Brief:
Collection of Race, Ethnicity, Language (REL) Data in Medicaid Applications: A 50-state Review of the Current Landscape (*update forthcoming*)

Collection of Sexual Orientation and Gender Identity (SOGI) Data:
Considerations for Medicaid and Spotlight on Oregon

Introduction

Issue Brief:
Collection of Sexual Orientation and Gender Identity (SOGI) Data: Considerations for Medicaid and Spotlight on Oregon (*update forthcoming*)

Expert Perspective:
Exploring Strategies to Fill Gaps in Medicaid Race, Ethnicity, and Language Data

Expert Perspective:
New York State of Health Pilot Yields Increased Race and Ethnicity Question Response Rates

Thank you!

Elizabeth Lukanen

elukanen@umn.edu

Check out our website at www.shadac.org and follow us on Twitter: @shadac

STATE HEALTH ACCESS DATA ASSISTANCE CENTER

@DankwaIrene

Irene Dankwa-Mullan, M.D., MPH

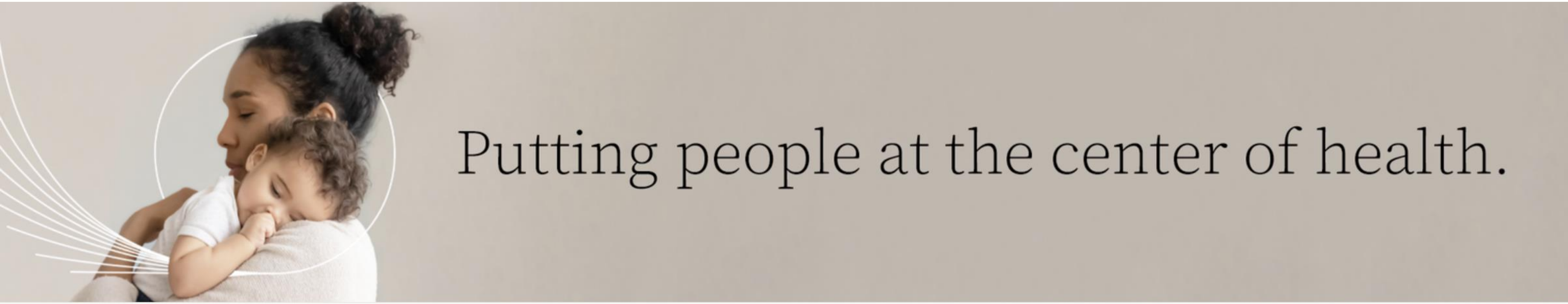
Chief Health Equity Officer,
Deputy Chief Health Officer, Clinical Affairs
Merative

Better Health through Data: Opportunities and Limitations of Health Data- Industry perspective

October 6, 2022

Irene Dankwa-Mullan MD MPH
Chief Health Equity Officer, Deputy Chief Health Officer
Clinical and Scientific Affairs
Merative (*formerly IBM Watson Health)*

Merative is a data, analytics, and technology partner for the health industry, including providers, health plans, governments, employers, and life sciences companies.



Putting people at the center of health.

BACKGROUND: Our commitment to addressing health equity and inclusive technologies

Promoting **inclusive language** in technology development and design

Hospital Measures: equity and social needs index: Incorporating community health measures and identifying social needs indicators

Health Insights: Building **health equity dashboards** and metrics

Providing timely access, holistic coordinated care and **social services** to individuals with high social needs

Thought Leadership
TechQuity: Integrating health equity, racial justice and ethical AI into technology development

Design justice to promote inclusive design of products, technologies and solutions

Addressing bias in artificial intelligence and machine learning algorithms

Leading with clinical expertise and science evidence to promote fairness, equity and technologies for social good

Social determinants of health mapping and precision geography with demographics

Examples of Artificial intelligence - Machine Learning applications in health

Pre-clinical and translational research e.g., drug discovery and genomic medicine, biomarkers for precision medicine

Clinical decision support, disease pathways, e.g., diagnostics, making predictions (predictive models) and identifying risk (patient stratification) or allocating resources

Patient-facing or consumer-facing applications e.g., delivery of therapies or the provision of information, management decision-making

Population-level applications e.g., understanding non-communicable chronic diseases, identifying epidemics

Interpretation of Medical Images Improved accuracy for radiology tasks such as screening, diagnosis, risk prediction and early intervention

Optimizing healthcare delivery processes Process optimization e.g., procurement, case management and logistics, and staff scheduling

AI-ML Data Bias

AI-ML bias is a general concept that refers to the fact that an AI system has been designed, intentionally or not, in a way that may make the system's decisions or use unfair.

Algorithm Model Assessments:

- - Label Bias
- - Modeling Bias
- - Population Bias
- - Measurement Bias
- - Missing validation bias
- - Human Use bias

Bias can be present both in the modeling of the **algorithm** and in the **data** used to train and test it. AI bias can emerge in an AI system as a result of **cultural, social, or institutional expectations**; because of **technical limitations** of its **design**;

- by being **used in unanticipated contexts** or
- by **making decisions about communities that are not considered in the initial design**.

Five Broad Aspects of Bias Across the Data Generation and Technology Development Continuum

Data and Health Equity

Evidence

Research bias: Lack of equitable standards around how our science is funded, conducted, reviewed, published and disseminated; lack of inclusion in clinical trials and researcher diversity, evidence-base & real-world data to inform health decisions

Experience/ Expertise

Provider bias: Provider expertise and experience; cognitive biases and in-group biases; Lack of health data insights and evidence; unconscious biases, pre-existing stereotypes or discriminatory practices from providers or health professionals that can feed into HER and claims data

Exclusion

Embedded data bias: incomplete health data, e.g., missing data or incomplete data in EHR's Favoring those groups who have robust health data profiles; Data bias in sample selection, modeling structure and selection of metrics for predictions Lack of cohort diversity; training data not representative

Environment

Data invisibility: Lack of data on the complete etiological context and exposures of a patient's health; including important factors – such as the social determinants of health, technology resources, broadband connectivity or other environmental factors that can trigger discriminatory outcomes

Empathy

Data empathy: Lack of knowledge, understanding and/or experience about the people, places, factors that make up the data – unable to recognize the bias and optimize analysis; lack of knowledge of data source and real-world evidence or social implications

Ref: Dankwa-Mullan I, Weeraratne D., Artificial Intelligence and Machine Learning Technologies in Cancer Care: Addressing Disparities, Bias and Data Diversity, Cancer Discov. 2022 12(6) Dankwa-Mullan, I. Perez-Stable E et. al eds The Science of Health Disparities Research; 2021 Wiley Dankwa-Mullan, I - Eliminating bias in health science data. Dankwa-Mullan I, Examining health disparities in precision medicine

SDoH Data and Informatics Challenges - Opportunities for Public Health

Data Collection – disparate data

- Variation in SDoH screening requirement,
- Surveys, questionnaires and SDoH tools, social media
- SDoH captured in EHR but devoid in claims data

Data Curation - standardization

- Context, ambiguity of measures
- Coding & ontology mapping challenges, terminologies
- Standardization of aggregate data

Data Choice & Compatibility – variability

- Determining best predictors for health & outcomes
- Analytic sets, EHR patient-level vs neighborhood /census level data (diverse zips)

Dankwa-Mullan I, McKillop MM, Solomon M, Preininger AM, Roebuck MC, Arriaga YE, George J, Jackson GP, South BR: **Identifying and Leveraging Public Data Sources with Structured Social Determinants of Health Information for Observational Health Research.** AMIA 2020

Ethical AI Framework to Promote Better Data for Health Equity and Racial Justice

Ethical AI
Goal is to understand how to optimize AI's beneficial impact, while reducing risks and adverse outcomes for all stakeholders,
in a way that prioritizes human agency and well-being, as well as environmental flourishing

Ethical AI Dimensions

- Accountability
- Impact of Algorithms
- Data Responsibility
- Design equity
- Discrimination and Bias
- Empathy
- Explainability
- Fairness
- Human Oversight

- Human Autonomy
- Inclusion
- Social Cohesion
- Inclusive Technology
- Moral Agency
- Privacy Protection
- Robustness, Safety
- Transparency and Trust
- Value Alignment

Dankwa-Mullan, I. Scheufele E., et al. "A Proposed Framework on Integrating Health Equity and Racial Justice into the Artificial Intelligence Development Lifecycle." *Journal of Health Care for the Poor and Underserved*, vol. 32 no. 2, 2021, p. 300-317..

Thank you

Irene Dankwa-Mullan MD MPH

idankwa@merative.com

Twitter: @DankwaIrene

LinkedIn: irene-dankwa-mullan-md-mph

merative™

© Merative 2022. All Rights Reserved.

The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on Merative's current product plans and strategy, which are subject to change by Merative without notice. Merative shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from Merative or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of Merative products.

Product release dates, availability and/or capabilities referenced in this presentation may change at any time at Merative's sole discretion, and are not intended to be a commitment to future product or feature availability in any way. Not all Merative products are available in all jurisdictions in which Merative operates. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific performance results, and an individual user may achieve results different than any stated here.

Any customer examples described are presented as illustrations of how those customers have used Merative products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Merative, the Merative logo, and merative.com are trademarks of Merative, registered in many jurisdictions worldwide. Other product and service names might be trademarks of Merative or other companies.

merative™

@N_Brennan

Niall Brennan, MPP

Chief Analytics and Privacy Officer

Clarify Health

30 YEARS

TAKE OUR SURVEY

Please fill out the evaluation survey you will receive immediately after this presentation, or via email this afternoon!

www.allhealthpolicy.org

UPCOMING EVENT

October 12, 2022 | 1:00 p.m. ET – 2:15 p.m. ET

The State of HIV/AIDS Progress & Applying Lessons Learned to Current Public Health Emergencies

The COVID-19 pandemic and more recently, the monkeypox global public health emergency have further complicated the lives of individuals living with HIV/AIDS. This event will explore lessons for public health from the HIV/AIDS experience; and how they map to COVID-19 and monkeypox response efforts.

allh.us/events

THANK YOU FOR ATTENDING!

ALLIANCE
FOR HEALTH POLICY | 30 YEARS