

# *Variability in Collection and Use of Race/Ethnicity and Language Data in 93 Pediatric Hospitals*

**John D. Cowden, Gabriela Flores, Tiffany Chow, Patricia Rodriguez, Tracy Chamblee, Megan Mackey, Anne Lyren & Michael F. Gutzeit**

**Journal of Racial and Ethnic Health Disparities**

ISSN 2197-3792

J. Racial and Ethnic Health Disparities  
DOI 10.1007/s40615-020-00716-8



**Your article is protected by copyright and all rights are held exclusively by W. Montague Cobb-NMA Health Institute. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**



# Variability in Collection and Use of Race/Ethnicity and Language Data in 93 Pediatric Hospitals

John D. Cowden<sup>1</sup> · Gabriela Flores<sup>1</sup> · Tiffany Chow<sup>1</sup> · Patricia Rodriguez<sup>2</sup> · Tracy Chamblee<sup>2</sup> · Megan Mackey<sup>3</sup> · Anne Lyren<sup>4</sup> · Michael F. Gutzeit<sup>5</sup>

Received: 23 August 2019 / Revised: 15 January 2020 / Accepted: 2 February 2020  
© W. Montague Cobb-NMA Health Institute 2020

## Abstract

**Objective** To describe how pediatric hospitals across the USA and Canada collect race/ethnicity and language preference (REaL) data and how they stratify quality and safety metrics using such data.

**Methods** Pediatric hospitals from the Solutions for Patient Safety network (125 US, 6 Canadian) were surveyed between January and March 2018 on collection and use of patient/family race/ethnicity data and patient/family language preference data. The study team created the survey using a formal process including pretesting. Responses were analyzed using descriptive statistics.

**Results** Ninety-three of 131 (71%) hospitals completed the survey (87/125 [70%] US, 6/6 [100%] Canadian). Patient race/ethnicity was collected by 95%, parent/guardian race/ethnicity was collected by 31%, and 5/6 Canadian hospitals collected neither. Minimum government race/ethnicity categories were used without modification/addition by 68% of US hospitals. Eleven hospitals (13%) offered a multiracial/multiethnic option. Most hospitals reported collecting language preferences of parent/guardian (81%) and/or patient (87%). A majority provided formal training on data collection for race/ethnicity (70%) and language preferences (70%); fewer had a written policy (41%, 51%). Few hospitals stratified hospital quality and safety measures by race/ethnicity (20% readmissions, 20% patient/family experience, 16% other) or language preference (21% readmissions, 21% patient/family experience, 8% other).

**Conclusions** The variability of REaL data collection practices among pediatric hospitals highlights the importance of examining the validity and reliability of such data, especially when combined from multiple hospitals. Nevertheless, while improvements in data accuracy and standardization are sought, efforts to identify and eliminate disparities should be developed concurrently using existing data.

**Keywords** Race/ethnicity · Language · Safety · Quality improvement · Pediatrics

## Abbreviations

REaL Race/ethnicity and language  
SPS Solutions for Patient Safety

OMB Office of Management and Budget  
IOM Institute of Medicine  
HHS Health and Human Services

✉ John D. Cowden  
jdcowden@cmh.edu

## Introduction

Disparities related to race, ethnicity, and language have been described throughout adult and pediatric medicine, affecting access to health care, health care processes, and health outcomes [1–4]. To be successful, efforts to identify and address such disparities must be based on accurate collection and effective use of race/ethnicity and language preference (REaL) data in the health care setting [5–7]. Challenges to such data collection and use include diverse definitions and categories for race and ethnicity, variable approaches to understanding and recording language needs and preferences, and a lack of

<sup>1</sup> Children's Mercy Kansas City, 2401 Gillham Road, Kansas City, MO 64108, USA

<sup>2</sup> Children's Medical Center Dallas, 1935 Medical District Drive, Mailstop K4.01, Dallas, TX 75235, USA

<sup>3</sup> Central Connecticut State University, 1615 Stanley St, New Britain, CT 06053, USA

<sup>4</sup> UH Rainbow Babies and Children's Hospital, 11100 Euclid Avenue, Cleveland, OH 44106, USA

<sup>5</sup> Children's Hospital of Wisconsin, 8915 W. Connell Court, Milwaukee, WI 53226, USA

standardized procedures [5, 8]. In pediatrics, there are additional challenges, including the predominance of adult-oriented electronic health records (that do not allow for data to be recorded for multiple family members, such as parents) and limited precedent for stratifying quality and safety metrics by REaL data in pediatric hospitals [9].

A problem fundamental to REaL data is the lack of uniformity in how race, ethnicity, and language variables are defined. Although in the USA there is an official set of race and ethnicity categories defined for federal government use (the Office of Management and Budget categories) [10], they are variably used outside of government and have been criticized as inadequate to accurately describe individuals or groups [8, 11]. Concepts related to language needs and preferences vary, as well, potentially including spoken and/or written preferences, primary language spoken vs. preferred language for health care interactions, English proficiency level, interpreter need, and all these types of language data for multiple family members (which can be different for each member, but not reflected in the electronic health record) [5, 9]. As a result of these inconsistencies, hospitals working with REaL data might not only have different sets of response options for race/ethnicity and language variables, but might have different concepts represented within them.

Three sets of recommendations have been published that seek to standardize REaL data collection in health care settings, one by Health Research and Education Trust ("The Disparities Toolkit," 2007) [12], the second by the Institute of Medicine (*Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*, 2009) [5], and the third by the US Department of Health and Human Services ("Data Collection Standards for Race, Ethnicity, Primary Language, Sex, and Disability Status," 2011) [13]. The effects of these efforts are unclear, as the majority of studies on REaL data collection practices in health care settings were published before or just after these publications appeared [14–21]. In them, almost all hospitals collected REaL data, but there was wide variability in collection practices. In addition, all of the studies reported practices in adult health care settings; to our knowledge, no such studies have been published for pediatric hospitals. A more comprehensive understanding of current REaL data collection and use at pediatric institutions would provide a foundation for on-going standardization of these processes, an essential step in improving data quality and comparability between hospitals and across systems so that inequities in health and health care can be identified and eliminated.

The purpose of this study is to describe current practices in REaL data collection and use in pediatric hospitals throughout the USA and Canada so that researchers and those working to improve the quality and safety of pediatric care might better understand how such practices vary. Ultimately, such

understanding could lead to more effective identification and elimination of disparities in children's health and health care.

## Methods

### Study Population and Sampling Methodology

The study population included all children's hospitals with inpatient services in the USA and Canada. At the time of the study, there were at least 220 such hospitals in the USA according to the Children's Hospital Association [22], and 15 in Canada (Children's Healthcare Canada, Emily Gruenwoldt, written communication, January 2019). To sample from this population, we utilized The Children's Hospitals' Solutions for Patient Safety (SPS), a network of children's hospitals focused on improving in-hospital safety outcomes through shared data and the adoption of standard safety practices. At the time of the study, SPS included 131 children's hospitals (125 in the USA, 6 in Canada), all of which were surveyed between January and March 2018 on their REaL data practices. The survey was designated as "required" for SPS network members.

### Survey Description

The survey was created by the study team using the following process: (1) creation of candidate questions by 2 team members (JC & GF), (2) refinement of the question set by study team, (3) survey testing with 10 volunteer respondents in positions related to REaL data collection and use at 4 children's hospitals, and (4) finalization of questions and survey format based on feedback from respondents and all study team members.

The final survey asked for hospital name, followed by questions related to the hospital's collection and use of race/ethnicity and language preference data. Because of the evolving debate about definitions of race and ethnicity, including whether and how they differ [23–28], the survey included the phrase "race and/or ethnicity" throughout. Here we have chosen to use "race/ethnicity" for efficient expression of these complicated and controversial concepts. For language, the survey asked about "language preference," which relates to the information commonly sought by hospitals for purposes of language support during clinical encounters.

Survey questions about data collection addressed whose race/ethnicity and language preference data were collected (child and/or parent/guardian), what categories were used (race/ethnicity), and whether formal training and policies existed for collection of REaL data. Data use questions asked whether specific quality metrics were stratified by race/ethnicity or language and allowed respondents to list such metrics. Due to branching logic, respondents might have

answered as few as 4 questions and as many as 15. The full survey is included as [Appendix](#). The final survey was created in REDCap (Research Electronic Data Capture) hosted by Children's Mercy Kansas City [29].

### Survey Methodology

The SPS contact person at each member hospital was sent an email including introductory text explaining the purpose of the survey, a link to the online survey, and a printable worksheet to use in preparing survey answers. The worksheet was included because answers to some questions needed to be sought from various hospital leaders, including those in quality and safety, inclusion/diversity/health equity, and admissions/patient intake. Respondents were notified in the introductory email that this was a required survey for SPS network members. They were given 4 weeks to complete the survey, with reminders sent at 7, 14, 21, and 25 days after initial contact. At the 4-week point, any hospitals that had not responded were sent a personalized email from the SPS coordinator to check for problems in completing the survey.

### Statistical Analysis

Descriptive statistics (i.e., counts, percentages) were used to summarize survey responses and the chi-square test for associations between variables.

The study was determined to be non-human subject research by the Institutional Review Board at Children's Mercy Kansas City.

### Results

Ninety-three of 131 (71%) hospitals completed the survey, including 87/125 (70%) from the USA and 6/6 (100%) from Canada. Characteristics of responding and non-responding hospitals are summarized in [Table 1](#). Response rates differed by hospital type and size, with standalone and larger (> 300 beds) hospitals more likely to respond ( $P = .008$  and  $.04$ , respectively). For hospitals in the USA, response rates did not differ by region of the country.

#### Data Collection—Race/Ethnicity

Routine data collection on patient race/ethnicity was reported by 88/93 (95%) hospitals. In contrast, parent/guardian race/ethnicity was collected by 29/93 (31%). There was no relationship between hospital type or size and the collection of parent/guardian race/ethnicity ( $P > .05$  for both). The 5 hospitals that did not collect race/ethnicity for patients were in Canada and none of them collected parental/guardian

race/ethnicity either. In contrast, one Canadian hospital reported collecting both patient and parent/guardian race/ethnicity.

Just over two-thirds (60/88, 68%) of hospitals collecting race/ethnicity data reported using the Office of Management and Budget (OMB) categories set forth by the US federal government [10] without modification. The other 28 hospitals described in free text how their categories differed from the OMB category set. A summary of these differences is shown in [Table 2](#). Most (20/28, 71%) added options such as "other," "unknown," "declined," "unavailable," or similar. Only 11 hospitals (13% of all survey respondents) reported including either a multiracial and/or multiethnic option (with or without free text to describe the races/ethnicities) or allowed the family member to select multiple races. Two hospitals had open-ended questions with no specific menu of options offered, and one did not ask a race/ethnicity question, instead offering drop-down menus for race (55 options) and ethnicity (45 options) for the family to select on their own. The single Canadian hospital that reported collecting race/ethnicity data used a set of options different from the US OMB categories and the Statistics Canada census categories [30] ([Table 3](#)).

#### Data Collection—Language

Routine collection of parent/guardian language preference was reported by 75/93 (81%) hospitals, with an even higher number (81/93, 87%) reporting collection of patient language preference separate from parent/guardian language information. Nine hospitals (10%) collected language preference for parent/guardian, but not patient, and 15 (16%) for patient, but not parent/guardian. Three hospitals (3%, 2 in Canada) reported collecting no language preference information (neither patient nor parent/guardian).

#### Written Policies and Formal Training

More hospitals reported having written policies related to language preference data collection (46/90, 51%) than related to race/ethnicity data collection (36/88, 41%). Similar numbers of hospitals reported having formal training for staff related to language preference (63/90, 70%) and race/ethnicity (62/88, 70%) data collection.

#### Data Use

A minority of hospitals reported stratifying hospital outcomes by race/ethnicity (readmissions 18/88, 20%; patient/family experience 18/88, 20%; other quality or safety measure 14/88, 16%) or by language (readmissions 19/90, 21%; patient/family experience 19/90, 21%; other quality or safety measure 7/90, 8%). Those reporting "other" were asked to name the measures, which are summarized in [Table 4](#).

**Table 1** Characteristics of hospitals responding and not responding to the REaL<sup>a</sup> Data Survey

	Responding <i>N</i> = 93 (%)	Non-responding <i>N</i> = 38 (%)	Total <i>N</i> = 131	<i>P</i> value
HHS region, <sup>b</sup> no. (%)				.30 <sup>c</sup>
1 (CT, ME, MA, NH, RI, VT)	5 (5)	4 (11)	9 (7)	
2 (NJ, NY)	9 (10)	6 (16)	15 (11)	
3 (DE, DC, MD, PA, VA, WV)	9 (10)	1 (3)	10 (8)	
4 (AL, FL, GA, KY, MS, NC, SC, TN)	19 (20)	6 (16)	25 (19)	
5 (IL, IN, MI, MN, OH, WI)	17 (18)	6 (16)	23 (18)	
6 (AR, LA, NM, OK, TX)	13 (14)	5 (13)	18 (14)	
7 (IA, KS, MO, NE)	5 (5)	1 (3)	6 (5)	
8 (CO, MT, ND, SD, UT, WY)	1 (1)	2 (5)	3 (2)	
9 (AZ, CA, HI, NV)	7 (8)	4 (11)	11 (8)	
10 (AK, ID, OR, WA)	3 (3)	2 (5)	5 (4)	
Canada	6 (6)	0 (0)	6 (5)	
Number of beds, no. (%)				.04
≤ 300	71 (76)	35 (92)	106 (81)	
≤ 100	14	8	22	
101–200	41	18	59	
201–300	16	9	25	
> 300	22 (24)	3 (8)	25 (19)	
301–400	12	1	13	
401–500	4	2	6	
> 500	6	0	6	
Hospital type, no. (%)				.008
Standalone children's hospital	40 (43)	7 (18)	47 (36)	
Other (e.g., within a larger hospital)	53 (57)	31 (82)	84 (67)	

<sup>a</sup> REaL race/ethnicity and language<sup>b</sup> HHS Health and Human Services; regions used to show regional representation among respondents and non-respondents<sup>c</sup> Regions combined for statistical analysis into East (1, 2, 3); South (4, 6); Midwest (5, 7); West (8, 9, 10); Canada

## Discussion

To our knowledge, this is the first report of REaL data collection practices in a large set of pediatric hospitals in the USA and Canada. We found that while almost all hospitals collected REaL data, specific practices related to data collection and use varied widely among responding hospitals, as well as between countries.

The difference in REaL data collection rates between US and Canadian hospitals appears to reflect differences in government and compliance pressures in the two countries. In the USA, the publication of *Unequal Treatment* in 2003 by the Institute of Medicine [1] highlighted racial and ethnic disparities throughout the US health care system, leading to a swell of interest in describing and reducing such disparities. Subsequent legislation, including the 2009 American Recovery and Reinvestment Act and the 2010 Patient Protection and Affordable Care Act, along with accreditation requirements from the Joint Commission and the National Committee for Quality Assurance Patient-Centered Medical

Home Recognition Program, have provided incentives and penalties to increase the collection of high-quality REaL data in health care settings. In contrast, there is a common perception in Canada that the human rights legislative framework does not allow collection of racial, ethnic, or language information, leading hospitals to avoid this practice [31]. The Ontario Human Rights Commission has stated that this is incorrect and that such collection is allowed, as long as it is for purposes consistent with Canada's *Human Rights Code* [31]. Comments from Canadian respondents to our survey showed this contrast in viewpoints. One respondent wrote, "in Canada, we cannot ask questions on the patient's race, ethnicity, language as it is unconstitutional as per our federal laws," while another commented, "...[we] led in the development of the Paediatric Health Equity Tool which was designed to collect patient/parent/guardian data including race, ethnicity, language as well as additional sociodemographic data..." This latter effort reflects a relatively new approach in Canada, where the historical lack of race/ethnicity focus in health efforts recently has been criticized [32, 33]. Although the

**Table 2** List of ways hospital race/ethnicity options differed from US Office of Management and Budget (OMB) minimum standards<sup>a</sup> among 28 children's hospitals

Difference from OMB minimum standards	No. of hospitals
Included "Decline" or "Prefer Not to Answer" (or similar)	17
Included "Other"	16
Included "Unknown" (or similar)	12
Multiple races could be selected or included "Multiracial" (or similar)	11
Specific additional races/ethnicities listed	5
Included "Unavailable" or "Respondent Not Available" (or similar)	4
Oral open question—no specific options given	2
No question—only a drop-down menu (55 race, 45 ethnicity options)	1
Completely different option set (Canadian setting)	1

<sup>a</sup> OMB minimum standards: ethnicity = Hispanic or Latino/not Hispanic or Latino; race = American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White

remainder of this discussion relates primarily to the US experience and perspective, the concepts and conclusions may be applicable to emerging efforts in Canada and elsewhere.

Similar to our study, reports of REaL data collection and use in adult hospitals in the USA have found high rates of race/ethnicity data collection (ranging from 78 to 97%), but marked inconsistency between hospitals in collection procedures [14–19]. For language data collection, 3 national surveys from 2003 to 2005 showed lower levels of patient language data collection than in our survey (39–60% vs 81–87%) [14, 16], but we found no similar studies from recent years. Despite the growing attention paid to multiple aspects of language and literacy that might be measured in the clinical

setting (e.g., primary vs. preferred language, spoken/understood vs. written/read, English proficiency in addition to primary language, interpreter need, health literacy in one or more languages), the health care literature includes little discussion of how such elements are being collected across broad groups of hospitals and other health care settings. A 2012 review of primary/preferred language data collection in health care included 10 studies that drew language data from hospital medical records, showing broad heterogeneity in language categories and options [21].

Whether variability in data collection is good or bad is a matter of debate. As efforts to find and eliminate disparities grow, two opposing trends in REaL data collection have

**Table 3** US and Canadian governmental guidance on race/ethnicity categories

US Office of Management and Budget Minimum Standard Categories <sup>9</sup>	Statistics Canada Census Questions and Categories <sup>16</sup>
Race:	What were the ethnic or cultural origins of this person's ancestors? (Specify as many origins as applicable)
American Indian or Alaska Native	Is this person an Aboriginal person, that is, First Nations (North American Indian), Métis, or Inuk (Inuit)?
Asian	1: No, not an Aboriginal person. Continue with the next question.
Black or African American	2: Yes, First Nations (North American Indian).
Native Hawaiian or Other Pacific Islander	3: Yes, Métis.
White	4: Yes, Inuk (Inuit).
Ethnicity:	(If "Yes" then respondent skips next question)
Hispanic or Latino	Is this person:
Not Hispanic or Latino	1: White
	2: South Asian (e.g., East Indian, Pakistani, Sri Lankan, etc.)
	3: Chinese
	4: Black
	5: Filipino
	6: Latin American
	7: Arab
	8: Southeast Asian (e.g., Vietnamese, Cambodian, Laotian, Thai, etc.)
	9: West Asian (e.g., Iranian, Afghan, etc.)
	10: Korean
	11: Japanese
	12: Other—specify:

**Table 4** Summary list of other measures stratified by race/ethnicity and/or language among hospitals responding to the REaL<sup>a</sup> Data Survey

Other measure	No. of hospitals
Disease-related (e.g., asthma, diabetes, cancer, infections, pain)	7
Processes of care (e.g., length of stay, discharge instructions, readmissions)	7
Preventive care or health maintenance (e.g., well-child care, vaccinations, cancer screening)	3
Access to care or care utilization	2
Employee or family experience	2
Safety events	1

<sup>a</sup> REaL race/ethnicity and language

emerged—(1) standardization and (2) customization and disaggregation. In order to compare data across populations and between studies, a common set of questions and categories is needed. The US federal government has led the way with an evolving set of standards (most recently updated in 2011 to comply with the Affordable Care Act) [13] that are required for national health surveys and other federal efforts. Yet, because the racial/ethnic and linguistic diversity of each locality is unique, customized categories are commonly needed to best describe individuals and communities specific to an area. Furthermore, the traditional minimum federal categories (5 races, 2 ethnicities; Table 3) have been increasingly criticized as too general, inappropriately combining diverse populations with differing health and health care experiences under the terms “Asian,” “Hispanic,” “Black,” etc. while leaving other groups unrepresented (e.g., various communities of Middle Eastern or North African background) [11, 34–36]. A movement towards disaggregating data into more precise identities has begun, with emerging proposals for more accurate terms selected in partnership with diverse community members and based on recent research by the US Census [11, 36].

The challenge of balancing standardization with customization and disaggregation has been addressed in the two most recent sets of REaL data recommendations from the US government—the 2009 IOM report, *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement* [5], and the 2011 Department of Health and Human Services (HHS) “Data Collection Standards for Race, Ethnicity, Primary Language, Sex, and Disability Status,” created in response to requirements in the Affordable Care Act [13]. The IOM recommendations are expansive, suggesting the use of dozens or even hundreds of granular race/ethnicity options depending on the populations being described, as well as a set of multiple language questions (English proficiency, spoken language preference, written language preference). The HHS requirements do not go so far, but expand substantially on the OMB categories, providing multiple options under the “Hispanic,” “Asian,” and “Native Hawaiian or other Pacific Islander” categories (total of 19 race/ethnicity options, including “other”) and including one required and two optional spoken language questions.

Both reports emphasize the importance of including specific race/ethnicity and language options relevant to a given area or community in addition to the minimum set of standard categories, as well as the option to select multiple races/ethnicities. Such specific options could then be rolled up into the basic set of HHS or OMB categories for comparison with other data sets.

While the HHS standards dictate how REaL data will be collected at the federal level, individual hospitals face the challenge of deciding which REaL data recommendations to follow, if any, and how to customize categories for their local communities. In our study sample, we found that a majority (68%) of US hospitals that collected race/ethnicity data used the OMB set without modification. Although this provides consistency of categories, the inability to choose from disaggregated categories, to choose “other” and name the other race/ethnicity, or to choose more than one race/ethnicity means that these hospitals are likely collecting data with compromised validity. In contrast, the 28 hospitals using a modification of the OMB set or an entirely different approach likely have a more detailed understanding of the communities they serve, but need to roll up their granular data into the HHS or OMB categories for comparison with other hospitals or with national data. Procedures for such aggregation are described in the 2009 IOM REaL data recommendations [5].

Because the focus of interaction in pediatrics is the family, children’s hospitals face a special challenge not addressed in national recommendations or in the larger literature on REaL data collection and use: how to collect, store, and use REaL data from multiple family members. In 2018, the Pediatric Health Equity Collaborative described this challenge in a white paper and proposed best practices for pediatric health care settings, including the collection of data related to race/ethnicity and language preference (including spoken, written, and English proficiency) for both parents/guardians and patients [9]. This practice would require significant changes to the electronic health record, which to date has been structured primarily for the adult health care setting. It would allow, though, for a more complete understanding of a family’s clinical needs and for the measurement of disparities based on

characteristics of family members currently not captured. In our study sample, language preference was collected for both patient and parent/guardian by more than 80% of hospitals, but race/ethnicity was not commonly collected for parents/guardians (31%).

The ultimate purpose of collecting REaL data is to use them to improve care provided to diverse populations. Our findings highlight a disconnect between collection and use, with almost all respondents collecting REaL data, but only one-fifth reporting their use in evaluating quality or safety outcomes. This is consistent with multiple studies of adult hospitals [14, 16, 37, 38], where authors have observed that compliance pressures from federal and other mandates have increased collection of data, but often not their use. Leaders from hospitals not using REaL data have identified various barriers to data use, including lack of perceived need, resources, and know-how. In contrast, hospitals already using such data described successful efforts to identify and address disparities, concluding that the effort to overcome barriers has been worth it [16, 37]. An important barrier to the use of REaL data for analyses across multiple hospitals, such as in the SPS network, is a concern for accuracy, reliability, and validity, as outlined above. Yet waiting for perfect data to appear risks a delay in finding and tackling disparities that affect individuals and communities served by all hospitals. In a set of recommendations for REaL data collection and use in hospitals, Regenstein emphasized that “[e]fforts to strengthen the accuracy and consistency of data collection should continue, but should not take center stage....The most significant and sustained efforts should focus on encouraging hospitals to use the information they currently collect” [16].

This data set has important limitations. Although we had a high response rate of 71%, just over half of the pediatric hospitals with inpatient services in the USA and Canada are included in the SPS network. Practices among pediatric hospitals that did not have a chance to participate in this survey remain undescribed, and our lack of detailed information about children’s hospitals that are not part of SPS makes it unclear how representative our sample is of all children’s hospitals in the USA and Canada. Although the study was not designed to statistically test for differences between subgroups of hospitals, the small number of Canadian children’s hospitals in this sample (due to the small overall number of children’s in Canada) results in a statistical mismatch when comparing US and Canadian rates—confidence intervals around estimates would be significantly larger for Canadian hospital rates. The roles of respondents at each hospital likely varied, with some knowing more than others about REaL data processes. Although we recommended that they contact the appropriate staff and leaders to find the requested

information, challenges in getting all details collected may have led to incomplete or inaccurate responses in some cases. We did not ask in our survey how REaL data were obtained (e.g., staff asking families to self-report, direct entry by families via paper or computer, or staff determination without asking) nor whether hospitals collected only spoken language or included written language preferences, as well. Collection method is a major concern in REaL data use due to the possibility that staff might record race/ethnicity without asking the respondent, which deviates from the recommended practice of having respondents to self-identify. As a follow-up to the current study, our group has the opportunity to assess such practices in a subset of SPS hospitals and develop recommended interventions to lower the risk of staff-identification of respondent race/ethnicity. Finally, we did not include questions about how complete or accurate REaL data were at each hospital. Such aspects of REaL data collection are important to further understanding inter-hospital differences that affect data validity, comparability, and use and will be part of future efforts within the SPS network.

## Conclusions

The variability of REaL data collection practices among pediatric hospitals highlights the importance of examining the validity and reliability of such data, especially when combined from multiple hospitals. Nevertheless, while improvements in data accuracy and standardization are sought, efforts to identify and eliminate disparities should continue to be developed using existing data.

**Contributors’ Statement Page** Dr. Cowden conceptualized and designed the study (including design and/or revision of the data collection instrument), coordinated and supervised data collection, carried out the initial analysis, drafted the initial manuscript and finalized the manuscript. Drs. Rodriguez, Chamblee, Mackey, Lyren, and Gutzeit and Ms. Flores and Ms. Chow conceptualized and designed the study (including design and/or revision of the data collection instrument) and reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

**Funding Information** This study received funding from the Center for Medicare & Medicaid Services (CMS) Partnership for Patients Program.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

## Appendix

### SPS REaL (Race, Ethnicity, & Language) Data Survey Questions

#### Race and ethnicity data collection

1. Does your hospital routinely gather information on the race and/or ethnicity of patients for inclusion in the medical record (for example, during patient registration or intake)? Y/N
2. Does your hospital routinely gather information on race and/or ethnicity of parents/guardians for inclusion in the medical record (for example, during patient registration or intake)? Y/N
3. Does your hospital only use the US government's standard "OMB" (Office of Management and Budget) options for race and ethnicity as shown below without modification or addition (for example, no "multiracial," "other," "declined," etc.): Y/N

Question 1—Hispanic (yes/no)

Question 2—Race:

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White

Please provide the questions and options for race and/or ethnicity used at your hospital (exact wording, if possible):

4. Does your hospital have a written policy on how to ask families about race and/or ethnicity? Y/N
5. Does your hospital have formal training for staff on how to ask families about race and/or ethnicity? Y/N

#### Race and ethnicity data use

6. Does your hospital stratify readmissions measures by race and/or ethnicity? Y/N
7. Does your hospital stratify patient and family experience measures by race and/or ethnicity? Y/N
8. Does your hospital stratify any other quality or safety measures by race and/or ethnicity? Y/N

Please list other measures stratified by race and/or ethnicity: \_\_\_\_\_

#### Language data collection

9. Does your hospital routinely gather information on language preferences of patients (apart from those of parents/

guardians) for inclusion in the medical record (for example, during patient registration or intake)? Y/N

10. Does your hospital routinely gather information on the language preferences of parents/guardians for inclusion in the medical record (for example, during patient registration or intake)? Y/N
11. Does your hospital have a written policy on how to ask families about their language preferences? Y/N
12. Does your hospital have formal training for staff on how to ask families about their language preferences? Y/N

#### Language data use

13. Does your hospital stratify readmissions measures by language preferences?
14. Does your hospital stratify patient and family experience measures by language preferences?
15. Does your hospital stratify any other quality or safety measures by language preferences?

Please list other measures stratified by language preferences: \_\_\_\_\_

## References

1. Smedley BD, Stith AY, Nelson AR. Unequal treatment: confronting racial and ethnic disparities in health care. Institute of Medicine (US) committee on understanding and eliminating racial and ethnic disparities in health care. Washington, DC: National Academies Press; 2003.
2. Flores G, Committee on Pediatric Research. Technical report—racial and ethnic disparities in the health and health care of children. *Pediatrics*. 2010;125(4):e979–e1020.
3. Goenka PK. Lost in translation: impact of language barriers on children's healthcare. *Curr Opin Pediatr*. 2016;28(5):659–66.
4. 2018 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; September 2019. AHRQ Pub. No. 19-0070-EF.
5. Institute of Medicine. Race, ethnicity, and language data: standardization for health care quality improvement. Washington, DC: National Academies Press; 2009.
6. National Research Council Panel on DHHS Collection of Race and Ethnic Data. *Eliminating Health Disparities: Measurement and Data Needs*. Ver Ploeg M, Perrin E, editors. Washington, DC: National Academies Press; 2004.
7. Cheng TL, Goodman E, Committee on Pediatric Research. Race, ethnicity, and socioeconomic status in research on child health. *Pediatrics*. 2015;135(1):e225–37.
8. Mays VM, Ponce NA, Washington DL, Cochran SD. Classification of race and ethnicity: implications for public health. *Annu Rev Public Health*. 2003;24:83–110.
9. Tan-McGrory A, Bennett-AbuAyyash C, Gee S, Dabney K, Cowden JD, Williams L, et al. A patient and family data domain collection framework for identifying disparities in pediatrics: results from the pediatric health equity collaborative. *BMC Pediatr BMC Pediatr*. 2018;18(1):18.
10. Office of Management and Budget, Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, Office of

- Management and Budget. [https://obamawhitehouse.archives.gov/omb/fedreg\\_1997standards](https://obamawhitehouse.archives.gov/omb/fedreg_1997standards). Accessed 18 Apr 2019.
11. PolicyLink. 2018. Counting a Diverse Nation: Disaggregating data on race and ethnicity to advance a culture of health. <http://www.policylink.org/resources-tools/counting-a-diverse-nation>. Accessed 20 Nov 2018.
  12. Hasnain-Wynia R, Pierce D, Haque A, Hedges Greising C, Prince V, Reiter J. Health Research and Educational Trust Disparities Toolkit. 2007. <http://hretdisparities.org>.
  13. Department of Health and Human Services. Data Collection Standards for Race, Ethnicity, Primary Language, Sex, and Disability Status. <https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=2&lvlid=23>. Accessed 18 Apr 2019.
  14. Hasnain-Wynia R, Pierce D, Pittman MA. Who, when, and how: the current state of race, ethnicity, and primary language data collection in hospitals. Commonwealth Fund May 1, 2004.
  15. Polednak AP. Collecting information on race, Hispanic ethnicity, and birthplace of cancer patients: policies and practices in Connecticut hospitals. *Ethn Dis*. 2005;15(1):90–6.
  16. Regenstein M, Sickler D. Race, ethnicity, and language of patients: hospital practices regarding collection of information to address disparities in health care. Princeton: Robert Wood Johnson Foundation; 2006.
  17. Gomez SL, Stariano W, Le GM, Weeks P, McClure L, West DW. Variability among hospitals and staff in collection of race, ethnicity, birthplace, and socioeconomic information in the greater San Francisco Bay Area. *J Registry Manag*. 2009;36(4):105–10.
  18. Hasnain-Wynia R, Weber DM, Yonek JC, Pumarino J, Mittler JN. Community-level interventions to collect race/ethnicity and language data to reduce disparities. *Am J Manag Care*. 2012;18(6 Suppl):S141–7.
  19. Gomez SL, Lichtensztajn DY, Parikh P, Hasnain-Wynia R, Ponce N, Zingmond D. Hospital practices in the collection of patient race, ethnicity, and language data: a statewide survey, California, 2011. *J Health Care Poor Underserved*. 2014;25(3):1384–96.
  20. Bonham VL, Umeh NI, Cunningham BA, Abdallah KE, Sellers SL, Cooper LA. Primary care physicians' collection, comfort, and use of race and ethnicity in clinical practice in the United States. *Health Equity*. 2017;1(1):118–26.
  21. Duong LM, Singh SD, Buchanan N, Phillips JL, Cerlach K. Evaluation of primary/preferred language data collection. *J Registry Manag*. 2012;39(3):121–32.
  22. Children's Hospital Association. About the Association Web site. <https://www.childrenshospitals.org/About-Us/About-the-Association>. Accessed 15 Mar 2019.
  23. Frank R. What to make of it? The (re)emergence of a biological conceptualization of race in health disparities research. *Soc Sci Med*. 2007;64(10):1977–83.
  24. Collins FS. What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era. *Nat Genet*. 2004;36(11 Suppl):S13–5.
  25. Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med*. 2003;348(12):1166–70.
  26. Corbie-Smith G, Henderson G, Blumenthal C, Dorrance J, Estroff S. Conceptualizing race in research. *J Natl Med Assoc*. 2008;100(10):1235–43.
  27. Baer RD, Arteaga E, Dyer K, Eden A, Gross R, Helmy H, et al. Concepts of race and ethnicity among health researchers: patterns and implications. *Ethn Health*. 2013;18(2):211–25.
  28. Travassos C, Williams DR. The concept and measurement of race and their relationship to public health: a review focused on Brazil and the United States. *Cad Saude Publica*. 2004;20(3):660–78.
  29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81.
  30. Statistics Canada. Census 2A-L – 2016. [http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&lang=en&Item\\_Id=295122#qb295150](http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&lang=en&Item_Id=295122#qb295150).
  31. Ontario Human Rights Commission. Count me in! Collecting human rights-based data. Government of Ontario, 2010. ISBN: 987-1-4435-2357-8.
  32. Grant T, Balkissoon D. How Canada's racial data gaps can be hazardous to your health. *The Globe and Mail* February 6, 2019. <http://www.theglobeandmail.com>. Accessed 18 Apr 2019.
  33. Boshra S. Why Canada voids asking about race, and why that's a problem. *The Globe and Mail* October 8, 2014. <http://www.theglobeandmail.com>. Accessed 18 Apr 2019.
  34. Srinivasan S, Guillermo T. Toward improved health: disaggregating Asian American and Native Hawaiian/Pacific Islander data. *Am J Public Health*. 2000;90(11):1731–4.
  35. Hawkins SS, Cohen BB. Affordable Care Act standards for race and ethnicity mask disparities in maternal smoking during pregnancy. *Prev Med*. 2014;65:92–5.
  36. Mathews K, Phelan J, Jones NA, Konya S, Marks R, Pratt BM et al. 2015 National content test race and ethnicity analysis report. Washington, DC: US Census Bureau; 2017.
  37. Pathman DE, Teal R. Use of data by hospitals in North Carolina to identify disparities in the care and outcomes of minority patients. *NC Med*. 2011;72(3):177–82.
  38. Wilson-Stronks A, Galvez E. Hospitals, language and culture: a snapshot of the nation. Joint Commission: Oakbrook Terrace, IL; 2007.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.