

Background:

Utility of Race and Ethnicity Data in Drug Review: Good clinical review requires that drug outcomes be assessed by intrinsic (age, gender and race) and extrinsic factors to identify potential modifiers of drug dose or effect¹. Further, Section 907 of Food and Drug Safety Innovation Act (FDASIA) requires that FDA report on clinical trial participation and safety and effectiveness outcomes by demographic subgroups². The Office of Management and Budget’s^{3,4} (OMB) minimum categories (5 racial / 2 ethnicity) may suffice for reporting needs, whereas more detailed race and ethnicity data may be needed to explain observed subgroup differences (e.g. Human Leukocyte Antigen (HLA) may describe severe skin reactions to allopurinol in Han Chinese^a, efficacy of antihypertensive medications may vary based on CYP metabolizing enzyme prevalence by race).

Standards for electronic submission of race and ethnicity data: Clinical trial data is subject to seemingly inconsistent collecting and reporting requirements from various organizations (Health Level 7⁵ (HL7), National Institutes of Health (NIH)⁶, Department of Health and Human Services (DHHS)⁷, Agency for Healthcare Research and Quality⁸ (AHRQ)). FDA’s guidance⁹ “Collection of Race and Ethnicity Data in Clinical Trials” recommends adherence to the OMB categories but also states that individuals “*be permitted to designate a multiracial identity*”. The non-extensible Study Data Tabulation Model (SDTM) controlled terminology¹⁰ for race (C74457) and ethnicity (C66790) (Table 4) adhere to OMB’s 5/2 categories for race and ethnicity, respectively and nonstandard values in the corresponding DM domains would be flagged as errors in conformance validation. However, the SDTM Implementation Guide¹¹ (IG) v 3.2 provides for alternatives for submission of multiple responses and a mechanism to handle nonstandard variable values (such as “*other, specify*” and could provide the granularity of data that would enable insightful review.

Aim: We illustrate limitations and successful use of the race and ethnicity standard assessed for technical conformance to SDTM. These examples could inform how sponsors might employ the SDTM IG to provide granular race and ethnicity data in NDA submissions.

FINDINGS and DISCUSSION:

We reviewed race and ethnicity data from 94 studies (35 New Drug Applications (NDA)) assessed through the JumpStart service. We found various race and ethnicity terms beyond the 5 OMB defined reporting. Fifty seven (62%) of studies had Race terms that did not conform to the CDISC Race value set; the terms “*Other*” and “*multiracial*” were the most prevalent violations (Table 1). More recent studies used a SUPPDM to handle the nonstandard values (23 studies from 12 applications). The use of a SUPPDM was not necessarily useful in many instances(Table 3) In many instances country substituted for either a RACE or ETHNICITY term (e.g. Arabic, Australian, Russian). To accommodate more granular race and ethnicity data “*multiracial*”, “*unknown*” or “*other*” terms were provided as RACE and ETHNICITY values and a SUPPDM variable included race or ethnicity values that are not valid CDISC controlled terms. Inconsistencies in the terminology and grouping of race and ethnicity in studies assessed in JumpStart represent the data collected in legacy trial case report forms that allowed either free text or that allowed multiple race or ethnicity responses. Some of these differences are more marked in studies that enroll patients for diseases that are not prevalent in the US, such as a rare parasitic disease. Even for trials for a prevalent disease such as diabetes, racial and ethnic self identification was observed to vary between immigrant groups in the United States (e.g. RACE “*White*”, SUPPDM “*Mexican*”, Ethnicity “*Hispanic*”) compared to the subjects enrolled from sites in Mexico (e.g. RACE “*Mexican*”, SUPPDM “*Mestizo*”, ethnicity “*non Hispanic*”. Unique terms for race and ethnicity were observed in trials conducted in the US versus those conducted in other regions, indicating differences in patterns of self identification based on country of residence. A summary of the racial and ethnic reporting requirements revealed a progressive interest in obtaining more granular race and ethnicity data, represented in a SUPPDM with an effort to map to the minimum 5 and 2 racial and ethnic classifications in the original OMB directive.

Table 1 Non-Standard RACE Terms (DataFit Validation Failures)	
Term	Number (n/94)
Multiple	14 (15%)
Multiracial	1 (1%)
Not applicable	1 (1%)
Not permitted	3 (3.2%)
Other	52 (56%)
Other Asian	1 (1%)
Unknown	7 (7.6%)

Table 3 Frequently used terms in SUPPDM that could be challenging to map to OMB RACE/Ethnicity Terms	
Term	Number (n/23*)
Mestizo	10 (23%)
Crossbred	1
Data not recorded	1
Gipsy	1
Mixed (variations)	8 (34.8)
Mixed Latino	1
Not recorded in France	3
Other	16 (69.5%)
Not reported	8 (34.8%)
Unknown	4 (17%)

* total could exceed 100%

REFERENCES: 1. 21 CFR 314.50(d)(5)(v)

1. Food and Drug Safety and Innovation Act <http://www.gpo.gov/fdsys/pkg/PLAW-112publ144/pdf/PLAW-112publ144.pdf> Accessed 2/10/2015

3. OMB. 1977, excerpt from “Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity http://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/re_app-a-update.pdf Accessed 2/8/2015

4. OMB. 2000, “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity”in Implementation of the 1997Standards for Federal Data on Race and Ethnicity, Accessed 2/8/2015 http://www.whitehouse.gov/omb/fedreg_1997standards.pdf Accessed 2/7/2015

5 NIH, 2014. Use of New Inclusion Management System Accessed 2/7/2015 <http://grants.nih.gov/grants/funding/phs398/CumulativeInclusionEnrollmentReport.pdf>.

6 NIH. Policy on reporting race and ethnicity data: subjects in clinical research. [Fhttp://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-005.html](http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-005.html) Accessed 2/7/2015

7 HHS US Dept of HHS implementation guidance on data collection standards for race, ethnicity, sex, primary language and disability status. <http://aspe.hhs.gov/datacncl/standards/ACA/4302/index.shtml> Accessed 2/8/2015

8. AHRQ Race Ethnicity and Language Date Standardization for Health Care Quality and Improvement. http://www.Ahrq.gov/research/findings/final-reports/io_racereport/iomracereport.pdf Accessed 2/8/2015

9.Providing Regulatory Submissions in Electronic Format: Standardized Study Data <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292334.pdf>

10 SDTM Controlled Terminology v 9/28/2014. (<http://evs.nci.nih.gov/ftp1/CDISC/SDTM/SDTM%20Terminology.pdf>). Accessed 2/10/2015

11Study Data Tabulation Model Implementation Guide v 3.2 http://www.cdisc.org/system/files/all/standard_category/application/pdf/sdtmig_20v3.2_20nonportfolio.pdf Accessed 2/10/2015

a. *Journal of Human Genetics* (2013) **58**, 317–326; doi:10.1038/jhg.2013.37; published online 2 May 2013 Pharmacogenomics of severe cutaneous adverse reactions and drug-induced liver injury Nahoko Kaniwa¹ and Yoshiro Saito

Table 2 Non-Standard ETHNICITY Terms from one NDA		
ETHNICITY	FREQUENCY	PERCENT
EAST ASIAN	24	7.1
HISPANIC OR LATINO	20	5.9
MIXED ETHNICITY	2	0.6
NOT REPORTED	82	24.2
OTHER	170	50.2
RUSSIAN	12	3.5
SOUTH ASIAN	1	0.3
SOUTHEAST ASIAN	1	0.3
UNKNOWN	27	8.0

Reporting Race and Ethnicity by Country

- Self identification of race and ethnicity can vary by country
- In trials conducted in the US, Hispanics commonly identify their race as “White”
- In trials conducted outside the US, some Hispanics report their race as “Mestizo”
- In a trial that enrolled patients from Mexico, many identified their race as “Other – Mexican” and their ethnicity as non-Hispanic
- Country specific regulations could preclude data collection

Table 4 RACE/ETHNICITY REPORTING and DATA STANDARDS vs SDTM Terminology							
OMB ³	OMB ⁴	Guidance ⁱⁱⁱ	HL7 ⁵	NIH ⁶	HHS ⁷	SDTM ^v	Hypothetical SUPPDM codes
Race Values						RACE	
AI or AN	AI or AN	AI or AN	AI or AN	AI or AN	AI or AN	AI or AN	AI or AN
							#
Asian or PI	Asian	Asian	Asian	Asian	Other Asian	Asian	Other Asian
					Asian Indian		Asian Indian
					Chinese		Chinese
					Filipino		Filipino
					Japanese		Japanese
					Korean		Korean
					Vietnamese		Vietnamese
							#
	NH/PI	NH/PI	NH/PI	NH/PI	Other PI	NH/PI	Other PI
					Native Hawaiian		Native Hawaiian
					Guamanian/ Chamorro		Guamanian/ Chamorro
					Samoan		Samoan
							#
Black	Black or AA	Black or (AA) [®]	Black or AA	Black or AA	Black or AA	Black or AA	Black or AA
							#
White	White	White	White	White	White	White	White
							#
		Multiracial		> 1 race			#
			Other				#
				Unknown or not reported			#
Ethnicity						ETHNIC	
Non-Hispanic	Non-Hispanic or Latino	Non-Hispanic	Non-Hispanic	Hispanic N	Not of Hispanic, Latin/a or Spanish origin	Not Hispanic or Latino	Not of Hispanic, Latin/a or Spanish Origin
Hispanic	Hispanic or Latino	Hispanic	Hispanic	Hispanic Y	Yes, Mexican, Mexican American, Chicano/a	Hispanic or Latino	Yes, Mexican, Mexican American, Chicano/a
					Yes, Puerto Rican		Yes, Puerto Rican
					Yes, Cuban		Yes, Cuban
					Yes, another Hispanic, Latino/a or Spanish origin		Yes, another Hispanic, Latino/a or Spanish origin
				Unknown or Not reported		Not reported	#
						Unknown	

[®] in US AI Al American Indian AN Alaskan Native PI Pacific Islander NH Native Hawaiian AA African American

consider terminology extension for "other "to include multiracial, details in SUPPDM (to include HHS and AHRQ categories)

The OMB reporting categories are compared in the table above with the reporting terms specified by HL7, the NIH and DHHS. Other Agency specific terms are more extensive and not represented above, such as those specified by AHRQ and the Centers for Disease Control and Prevention. These terms do provide more granular information and are designed to map to the 5/2 OMB reporting codes through the choice of a primary race or ethnicity and could provide useful information if adopted uniformly in Case Report Forms (CRF). Additional insight on individual subject characteristics could be obtained by combining Race and SUPPDM data in a by subject analysis to assess subgroup outcomes. Similarly, data from COUNTRY and SITE ID could provide more granular data regarding subgroups.

CONCLUSION:

Granular demographic data (race, ethnicity, region, country of origin) has important analytic value and their collection in CRFs & study tabulations is encouraged. Inclusion of detailed responses for Race and Ethnicity could be included in a SUPPDM and accommodate for the need to represent subjects who self identify as having multiple races; the use of other detailed standard terms for race and ethnicity in case report forms could enhance the ability to map the SUPPDM data to the OMB reporting terms as well as provide detailed data needed for analyses.

Acknowledgement: Dr. Armando Oliva for insightful review .