



GSK Diversity Paper: The Challenges we Faced with Real World Data

Paper RE04

Carletha Blanding, Chris Leach and Harrison James
GlaxoSmithKline

Agenda

1. Introduction
2. Overview
3. Data sources
4. Data integration
5. Tables and figures
6. Examples of data issues
7. GSK Diversity paper status

Introduction

- In early 2021, Pfizer published a paper titled '*Demographic diversity of participants in Pfizer sponsored clinical trials in the United States*'¹
- The primary focus of their paper was to “provide a baseline upon which we can quantify the impact of our ongoing efforts to improve racial and ethnic diversity in clinical trials”
- Analyzed data for 213 US trials between 2011 and 2020
- Compared against US Census data
- Industry wide challenge for other big pharma companies to release their diversity data

Challenge accepted!

- We reproduced all displays and supplemental displays produced by Pfizer
- We reproduced all census level comparisons performed by Pfizer
- We went a step further and performed comparisons to real-world epidemiology data
- We summarized data for 495 US Trials conducted between 2002-2019

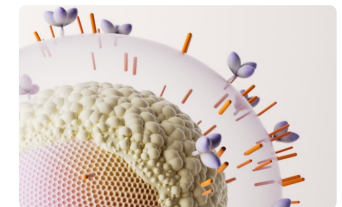
¹Rottas, M. et al. (2021) *Demographic diversity of participants in Pfizer sponsored clinical trials in the United States*. Elsevier Inc.

Overview

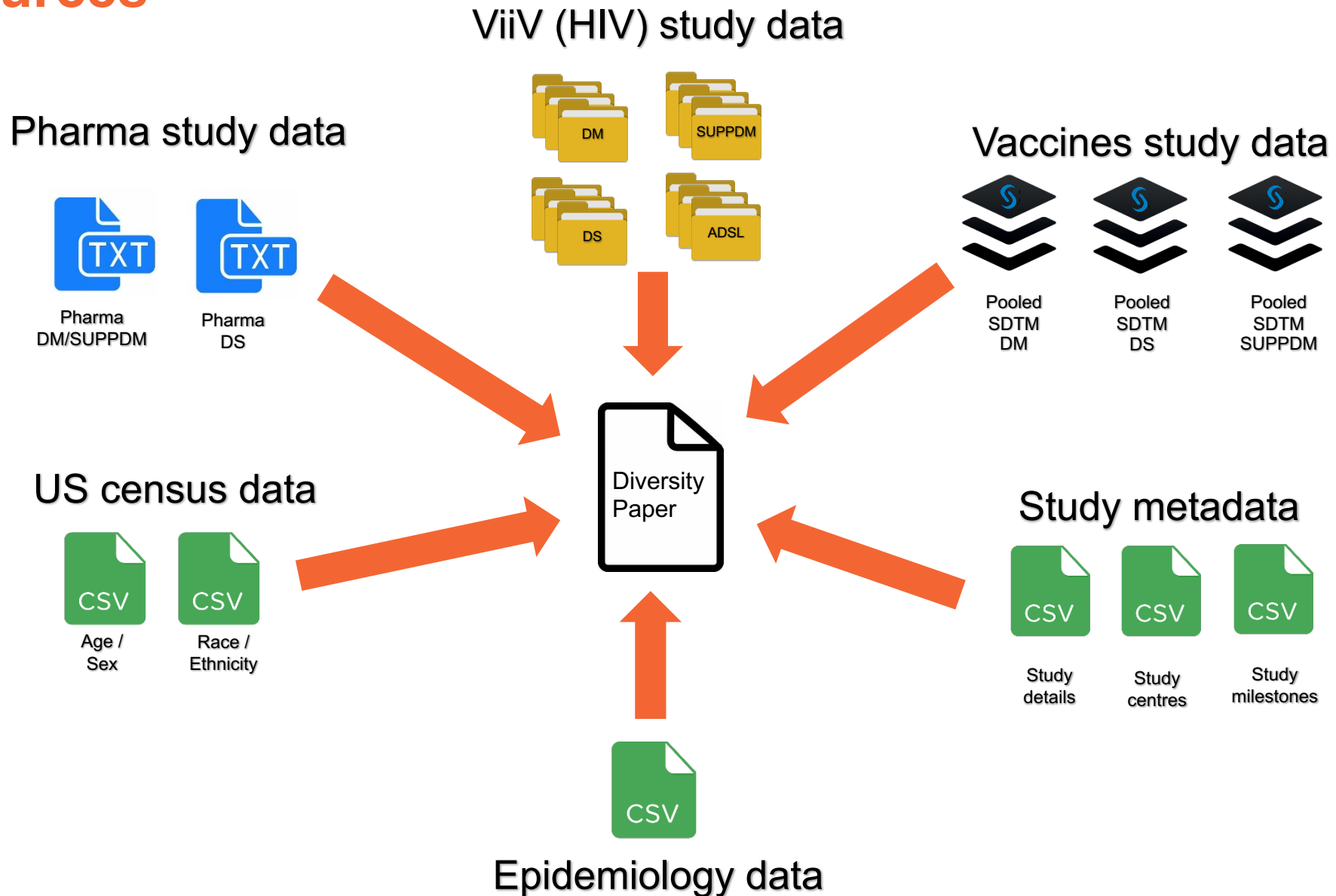
Create a paper summarizing subject demographics in US studies*. The paper would include:

- Three different entities of GSK:
 - Pharma
 - Vaccines
 - ViiV (HIV)
- Studies using CDISC standards
- US Census 2019 data
- Epidemiology results for specific therapeutic areas

* Specific criteria was used to select eligible US studies.



Data Sources



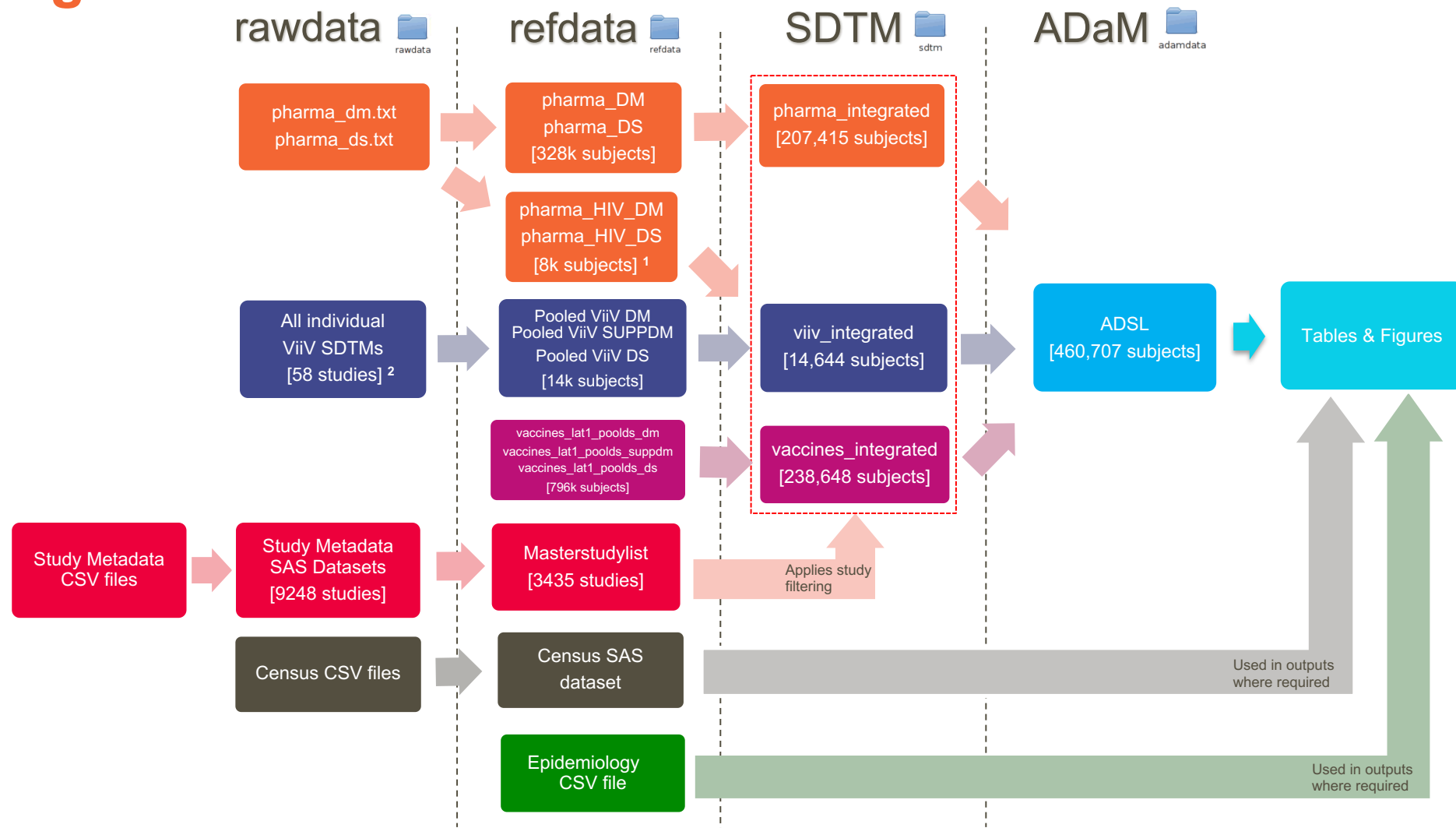
Data Sources

Criteria

Import, integrate and summarize study data for:

- GSK sponsored interventional trial data for Pharma, Vaccines, and ViiV (HIV)
 - Studies conducted from 2002-2019 (excluding extension studies)
 - Trials that were completed and archived
- } Study metadata
- Trials that collected **both** race and ethnicity data
- } DM
- All participants who passed screening
- } DS
- All participants who were assigned treatment
- } DM

Data Integration



1. The Pharma data included HIV studies prior to ViiV. For the purpose of integration, these HIV studies were moved to ViiV data.
2. For one ViiV study, an ADSL dataset was provided instead of SDTM datasets.

Summary of Studies

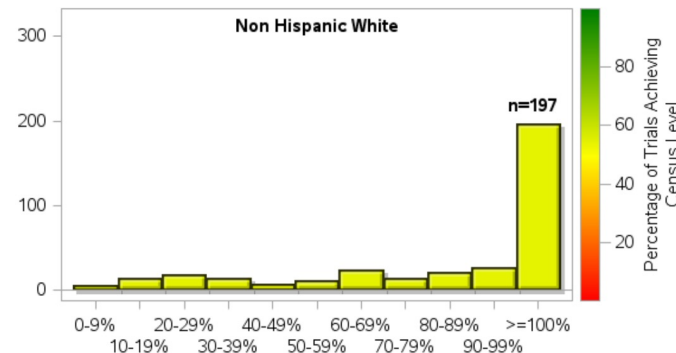
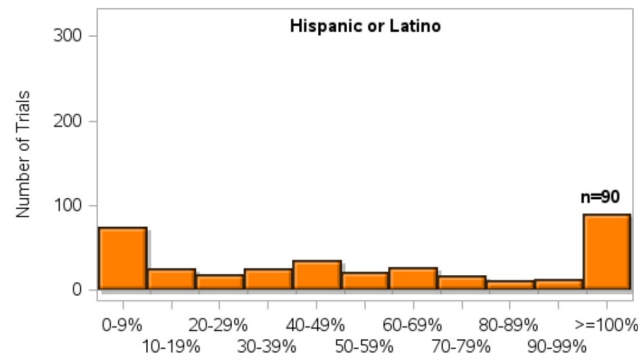
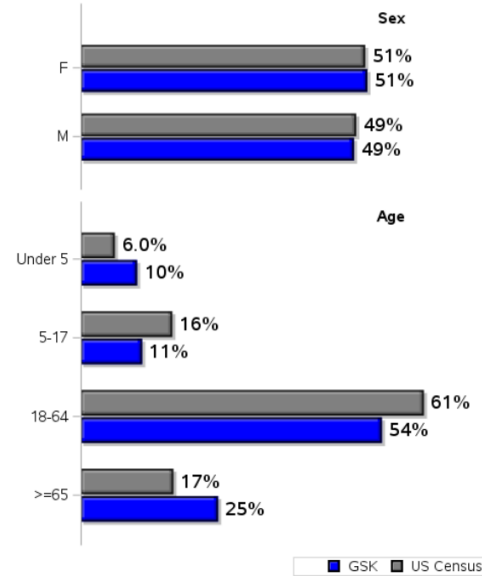
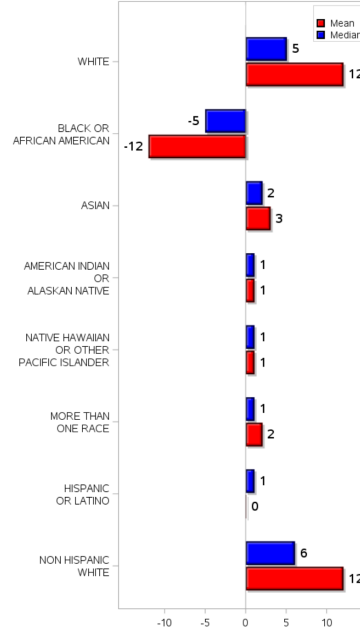
US Studies

	Number of Subjects [1]	% of Subjects	Number of Clinical Trials [1]	% of Clinical Trials
Pharma Therapy Area				
Autoimmune	0	0.00%	0	0.00%
Cardiovascular Diseases	9188	15.30%	52	14.57%
Dermatology	53	0.09%	4	1.12%
Gastrointestinal Diseases	245	0.41%	6	1.68%
Infectious Diseases (Non-HIV)	3161	5.26%	33	9.24%
Inflammation	534	0.89%	8	2.24%
Metabolic Diseases	4322	7.20%	43	12.04%
Musculoskeletal Diseases	567	0.94%	10	2.80%
Neurosciences	5687	9.47%	45	12.61%
Oncology	418	0.70%	14	3.92%
Ophthalmology	225	0.37%	2	0.56%
Other	28	0.05%	1	0.28%
Respiratory	34533	57.52%	121	33.89%
Urology	1077	1.79%	18	5.04%
Pharma Overall	60038	100.00%	357	100.00%
ViiV Therapy Area				
Infectious Diseases (HIV)	5720	100.00%	93	100.00%
Vaccines Therapy Area				
Vaccines Prophylactic	42448	99.87%	40	88.89%
Vaccines Therapeutic	55	0.13%	5	11.11%
Vaccines Overall	42503	100.00%	45	100.00%
Overall Therapy Area				
Overall	108261	100.00%	495	100.00%

[1] The number of subjects and number of clinical trials are based on the total number of subjects and clinical trials within each therapeutic area category.

Figures

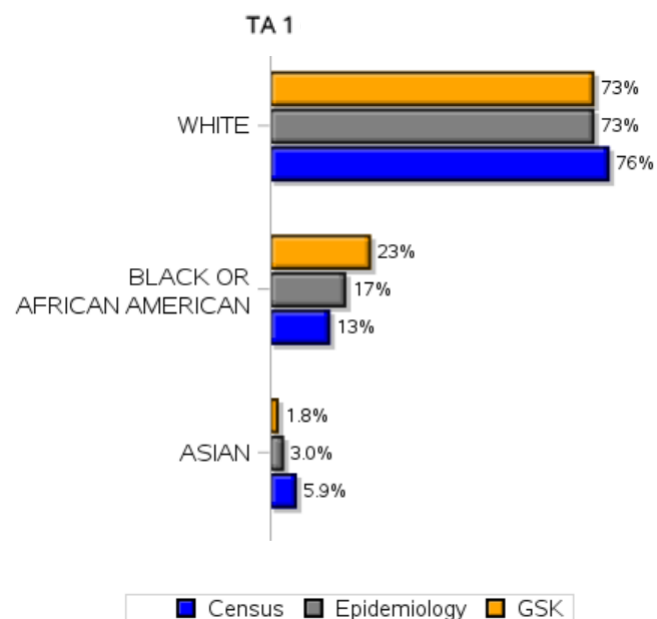
Overview



- Six figures were requested for the publication.
- These figures use various methods to compare the demographic makeup of the clinical trials to those found in the census and epidemiology data.
- The request involved replicating images of displays that were provided as examples of what was desired.
- SAS Graph Template Language was used to provide the options and control necessary to replicate the complex and non-standard nature of these displays.

Figures

Stacked Bar Chart of Ethnic Proportions of US GSK Trials, Epidemiology Data, and US Census

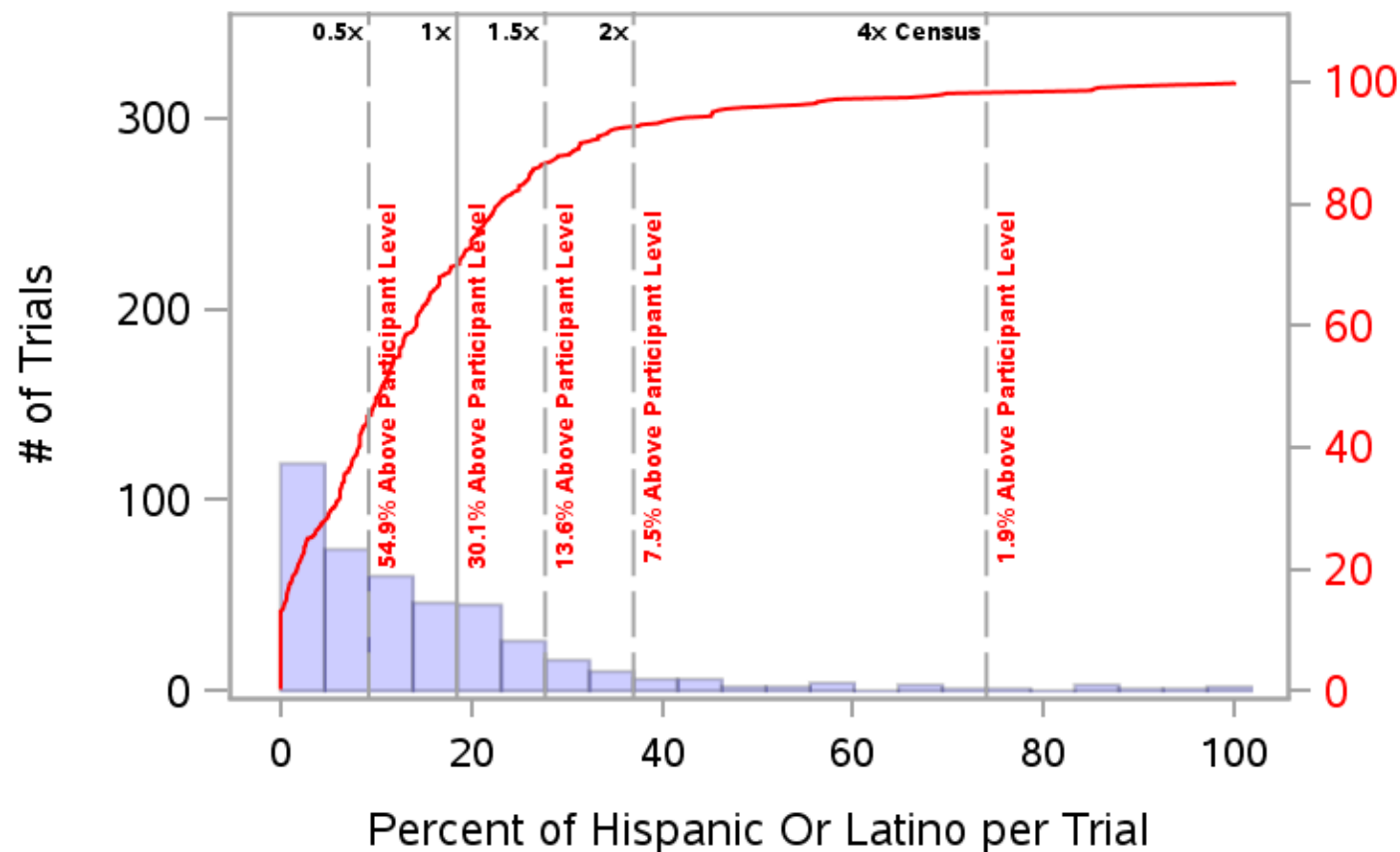


indication*	race ethnicity	percent
Therapeutic Area 1	White	72.85%
Therapeutic Area 1	Black or African American	17.02%
Therapeutic Area 1	Asian	2.97%
Therapeutic Area 1	American Indian or Alaskan Native	1.37%
Therapeutic Area 1	Native Hawaiian or Other Pacific Islander	-
Therapeutic Area 1	Mixed Race	5.80%
Therapeutic Area 1	Hispanic or Latino	14.37%
Therapeutic Area 2	White	85.66%
Therapeutic Area 2	Black or African American	7.10%
Therapeutic Area 2	Asian	2.14%
Therapeutic Area 2	American Indian or Alaskan Native	-
Therapeutic Area 2	Native Hawaiian or Other Pacific Islander	-
Therapeutic Area 2	Mixed Race	5.10%
Therapeutic Area 2	Hispanic or Latino	6.45%
Therapeutic Area 3	White	63.19%
Therapeutic Area 3	Black or African American	23.65%
Therapeutic Area 3	Asian/Pacific Islander*	5.60%
Therapeutic Area 3	American Indian or Alaskan Native	0.80%
Therapeutic Area 3	Native Hawaiian or Other Pacific Islander	-
Therapeutic Area 3	Other*	6.76%
Therapeutic Area 3	Hispanic or Latino	10.38%
Therapeutic Area 4	White	31.39%
Therapeutic Area 4	Black or African American	55.34%
Therapeutic Area 4	Asian	7.76%
Therapeutic Area 4	American Indian or Alaskan Native	0.63%
Therapeutic Area 4	Native Hawaiian or Other Pacific Islander	0.14%
Therapeutic Area 4	Mixed Race	4.73%
Therapeutic Area 4	Hispanic or Latino	35.72%

* Therapeutic areas redacted.

Figures

Histogram and Cumulative Distribution Curve of Trial Makeup

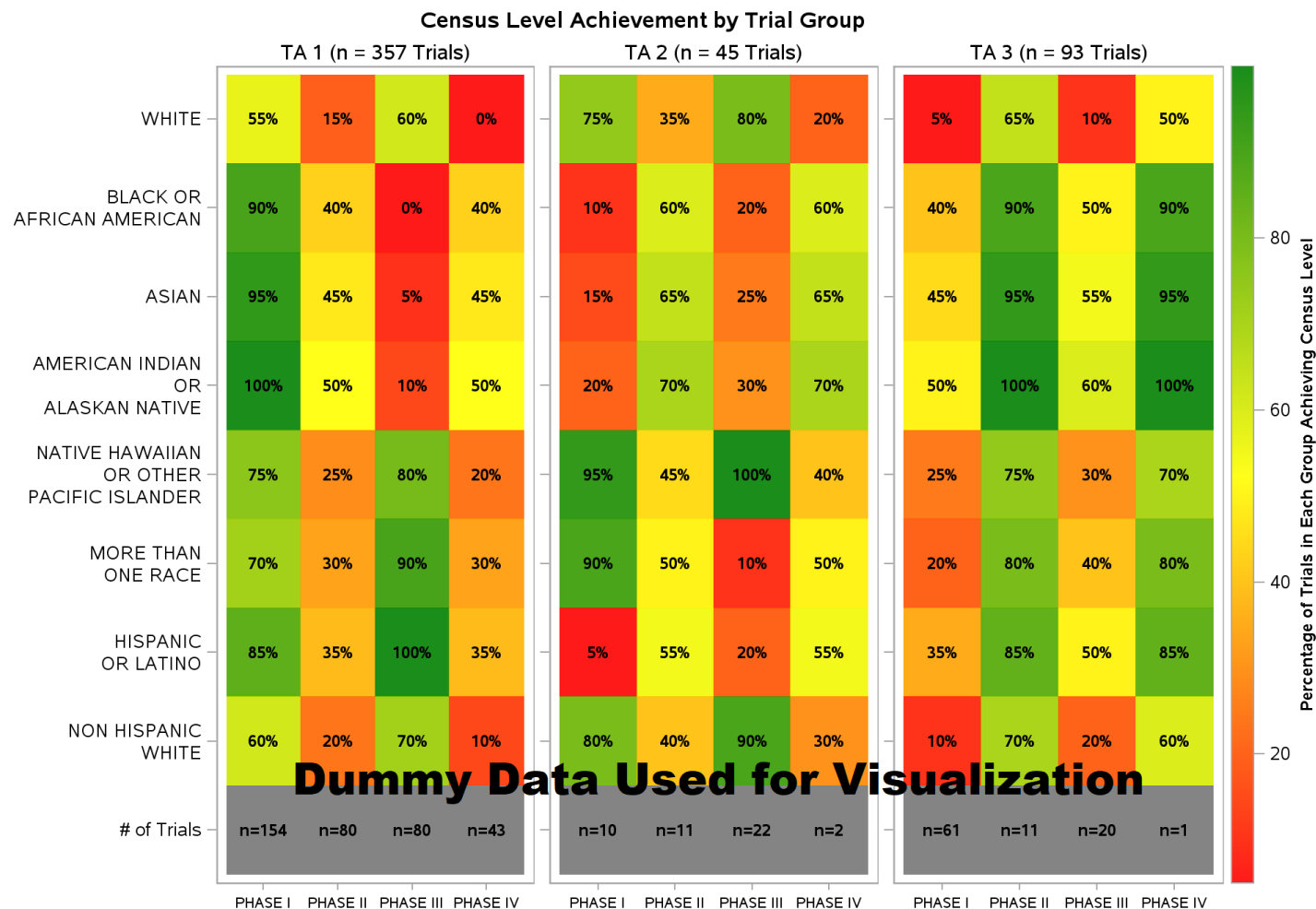


- Eight plots in a lattice layout
- Many aspects of the plot were programmatically derived from the data: Labels, bin size, reference lines, axis length, tick values
- Y1 axis was standardised across all plots to facilitate comparison between groups
- For every percent value found in the data the cumulative distribution was calculated as the proportion of studies having a lower percentage of study subjects in the race category
- Cumulative distribution values and bin counts had to be validated

Histogram of counts of trials by total percent with reference lines for multiples of census with an overlaid distribution curve showing Cumulative % of Trials at or Below Participant Level

Figures

Heatmap of Percent of Census by Therapeutic Area Category and Trial Phase



- A lattice layout plot was created with 1 row and 3 columns, one column per TA category.
- A common Y axis indicated the race/ethnicity category being summarised or the total number of trials included in each group.
- Each column has a scatterplot of a text version of the percentage value overlayed on a heatmap (heatmapparm) using the percentage value as the colorresponse and a colormodel=(red yellow green) option

Numbers and colours are produced with dummy data

Data Issues

Example 1 – Census files not machine-readable

	A	AI	AJ	AK	AL	AM	AN	AO	AP	AQ
1	table with row headers in column A and column headers in rows 3 through 5. (leading dots indicate sub-parts)									
2	Annual Estimates of the Resident Population for Selected Age Groups by Sex for the United States: April 1, 2010 to July 1, 2019									
3	Age									
4		2019								
5		Both Sexes	Male	Female						
6	Total	328,239,523	161,657,324	166,582,199						
7	.Under 5 years	19,576,683	10,009,207	9,567,476						
8	.5 to 9 years	20,195,895	10,322,762	9,873,133						
9	.10 to 14 years	20,798,268	10,618,261	10,180,007						
10	.15 to 19 years	21,054,570	10,745,607	10,308,963						
11	.20 to 24 years	21,632,940	11,064,752	10,568,188						
12	.25 to 29 years	23,509,016	12,004,570	11,504,446						

	A	M	N	O	P	Q	R	S
1	table with row headers in column A and column headers in rows 3 through 4. (leading dots indicate sub-parts)							
2	Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States: April 1, 2010 to July 1, 2019							
3	Sex, Race, and Hispanic Origin							
4		2019						
5	TOTAL POPULATION	328,239,523						
6	..One Race:							
7	...White	250,522,190						
8	...Black or African American	44,075,086						
9	...American Indian and Alaska Native	4,188,092						
10	...Asian	19,504,862						
11	...Native Hawaiian and Other Pacific Islander	806,937						
12	...Two or More Races	9,142,356						
13	..Race Alone or in Combination:1							
14	...White	258,646,488						
15	...Black or African American	48,221,139						
16	...American Indian and Alaska Native	6,945,552						
17	...Asian	22,861,985						
18	...Native Hawaiian and Other Pacific Islander	1,612,424						
19	..NOT HISPANIC	267,667,286						
20	...One Race:							
21	...White	197,309,822						
22	...Black or African American	41,147,488						
23	...American Indian and Alaska Native	2,434,908						

	Category	Value	Count	Total Population	Percentage
1	Ethnicity	HISPANIC	60572237	328239523	18.453669578
2	Ethnicity	NOT HISPANIC	267667286	328239523	81.546330422
3	Ethnicity	NOT HISPANIC - AMERICAN INDIAN AND ALASKA NATIVE	2434908	328239523	0.7418082922
4	Ethnicity	NOT HISPANIC - ASIAN	18905879	328239523	5.7597814021
5	Ethnicity	NOT HISPANIC - BLACK OR AFRICAN AMERICAN	41147488	328239523	12.535811539
6	Ethnicity	NOT HISPANIC - NATIVE HAWAIIAN AND OTHER PACIFIC ISLANDER	595908	328239523	0.1815466933
7	Ethnicity	NOT HISPANIC - TWO OR MORE RACES	7273281	328239523	2.2158455915
8	Ethnicity	NOT HISPANIC - WHITE	197309822	328239523	60.111536903
9	Race	AMERICAN INDIAN AND ALASKA NATIVE	4188092	328239523	1.2759255685
10	Race	ASIAN	19504862	328239523	5.9422649112
11	Race	BLACK OR AFRICAN AMERICAN	44075086	328239523	13.427720586
12	Race	NATIVE HAWAIIAN AND OTHER PACIFIC ISLANDER	806937	328239523	0.2458378542
13	Race	TWO OR MORE RACES	9142356	328239523	2.7852697068
14	Race	WHITE	250522190	328239523	76.322981374
15	Sex	M	161657324	328239523	49.249804692
16	Sex	F	166582199	328239523	50.750195308
17	Age	UNDER 18 YEARS	73039150	328239523	22.251784103
18	Age	UNDER 5 YEARS	19576683	328239523	5.9641455791
19	Age	5 TO 13 YEARS	36829704	328239523	11.22037458
20	Age	14 TO 17 YEARS	16632763	328239523	5.0672639443
21	Age	18 TO 64 YEARS	201142110	328239523	61.2790648
22	Age	18 TO 24 YEARS	30219206	328239523	9.2064495231
23	Age	25 TO 44 YEARS	87599465	328239523	26.687665215
24	Age	45 TO 64 YEARS	83323439	328239523	25.384950063
25	Age	65 YEARS AND OVER	54058263	328239523	16.469151096

www.census.gov

Data Issues

Example 2 – RACE = “OTHER” (Vaccines)

- The race category “OTHER”, had much higher counts in vaccine studies, than any other category.
- RACEOTH in SUPPDM contained the subject’s race information.

DM.RACE	SUPPDM.RACEOTH	Mapped Race
OTHER	BLACK AND WHITE	MIXED RACE
OTHER	AFRICAN AM/WHITE	MIXED RACE
OTHER	1/2BLACK,1/2WHITE	MIXED RACE
OTHER	CAUCASIAN/BLACK	MIXED RACE
OTHER	BLACK,CAUCASIAN	MIXED RACE
OTHER	MULTI RACIAL /MOM WHITE /DAD BLACK	MIXED RACE
OTHER	WHITE AND BLACK	MIXED RACE
OTHER	AFRICAN AMERICAN/CAUCASIAN	MIXED RACE
OTHER	MOTHER - AFRICAN AMERICAN,FATHER - WHITE	MIXED RACE
OTHER	BIRACIAL; WHITE AND AFRICAN AMERICAN	MIXED RACE
OTHER	BI-RACIAL/WHITE-CAUC & AFRICAN AMERICAN	MIXED RACE
OTHER	50%BLACK AND 50%CAUC	MIXED RACE
OTHER	AMERICAN	OTHER
OTHER	ASIAN	ASIAN
OTHER	UNKNOWN	NA
OTHER	EAST ASIAN/SOUTH EAST ASIAN	ASIAN
OTHER	ASIAN-EAST ASIAN/ASIAN-SOUTH EAST ASIAN	ASIAN
OTHER	DECLINE TO STATE	NA
OTHER	WHITE	WHITE

DM.RACE	DM.ETHNICITY	SUPPDM.RACEOTH	Mapped Race	Mapped Ethnicity
OTHER	null	HISPANIC	OTHER	HISPANIC OR LATINO

Data Issues

Example 3 – Unknown demography data (Pharma)

- There were very few cases of sex category “UNKNOWN”, except for one study which had a very large percentage of subjects with unknown sex.

Study X (N=42)	
Sex	
n	42
F	20 (48%)
M	8 (19%)
NA	14 (33%)

- One obstetric labour study which included the unborn foetus as a subject.

SUBJECT	USUBJID	BRTHDTC	AGE	AGEU	SEX	RACE	ETHNIC
Mother	111111-000001-M	1994	21	YEARS	F	WHITE	NOT HISPANIC OR LATINO
Foetus	111111-000001-F	null	null	null	U	null	null
New-born	111111-000001-N	2015-06-01	60	DAYS	M	WHITE	NOT HISPANIC OR LATINO

What comes next?

- Diversity paper going through final review stages
- Work underway to convert further non-CDISC studies to SDTM
- Additional analysis planned for the future

Thanks for listening!



You may contact the authors at:

Carletha Blanding

GlaxoSmithKline

Email: Carletha.x.blanding@gsk.com

Chris Leach

GlaxoSmithKline

Email: chris.x.leach@gsk.com

Harrison James

GlaxoSmithKline

Email: harrison.x.james@gsk.com