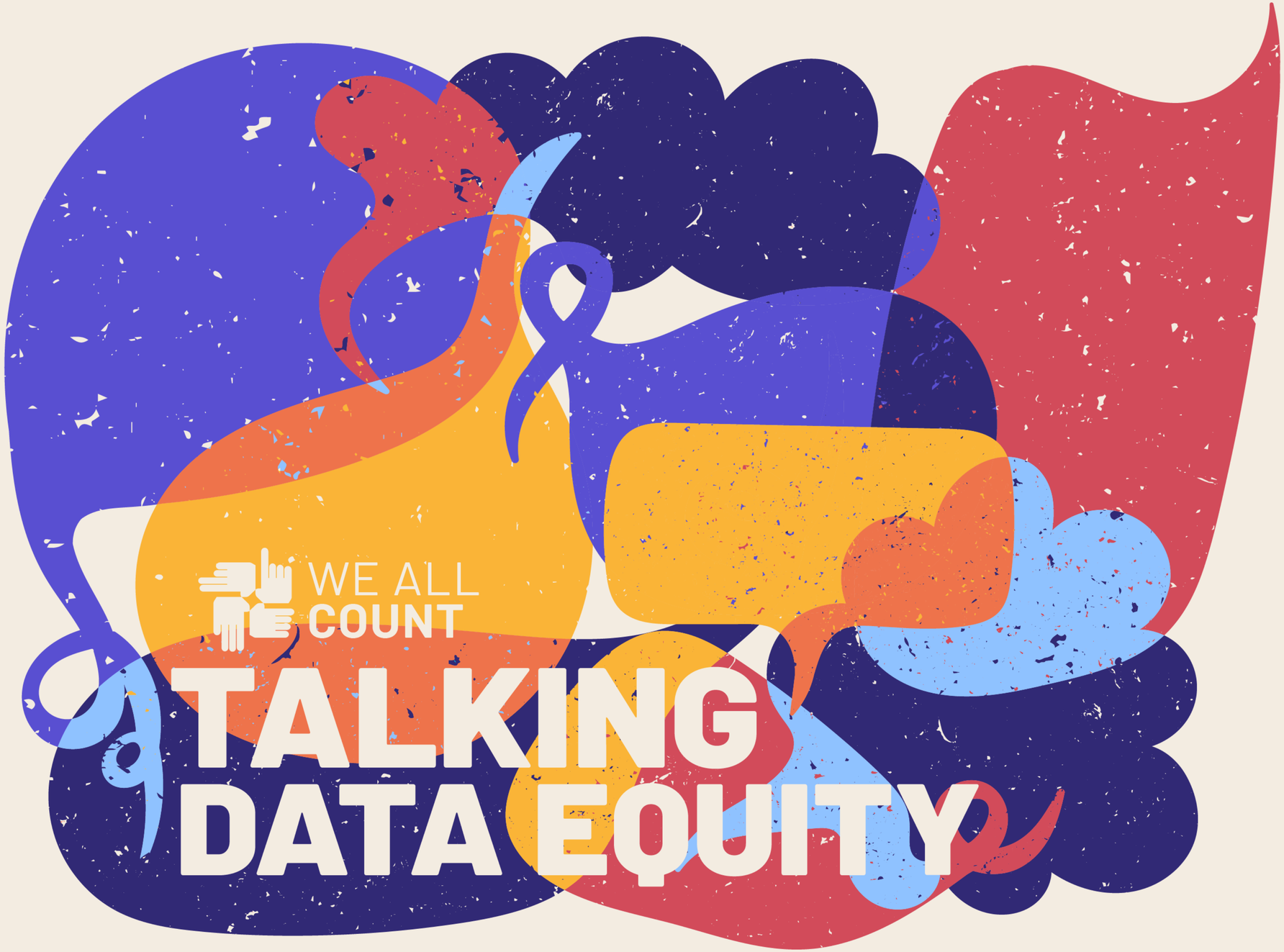

WE ALL
COUNT
TALKING
DATA EQUITY

Measuring Gaps and Data Equity

WE ALL COUNT

project for equity in data science

What even IS a gap?

When does a difference become a...

...gap

...equity gap

...disproportionality

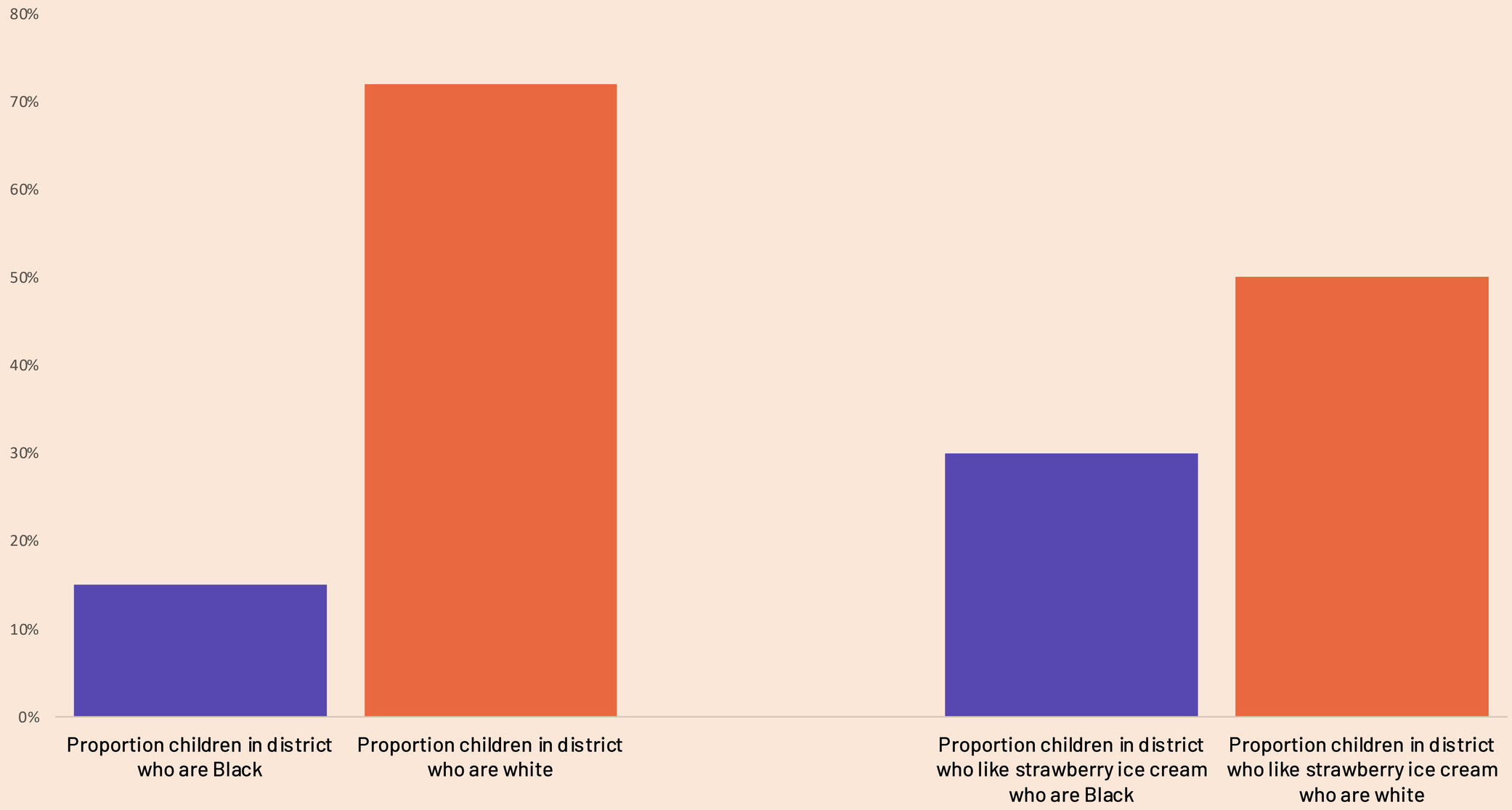
...disparity

When does a difference become a...

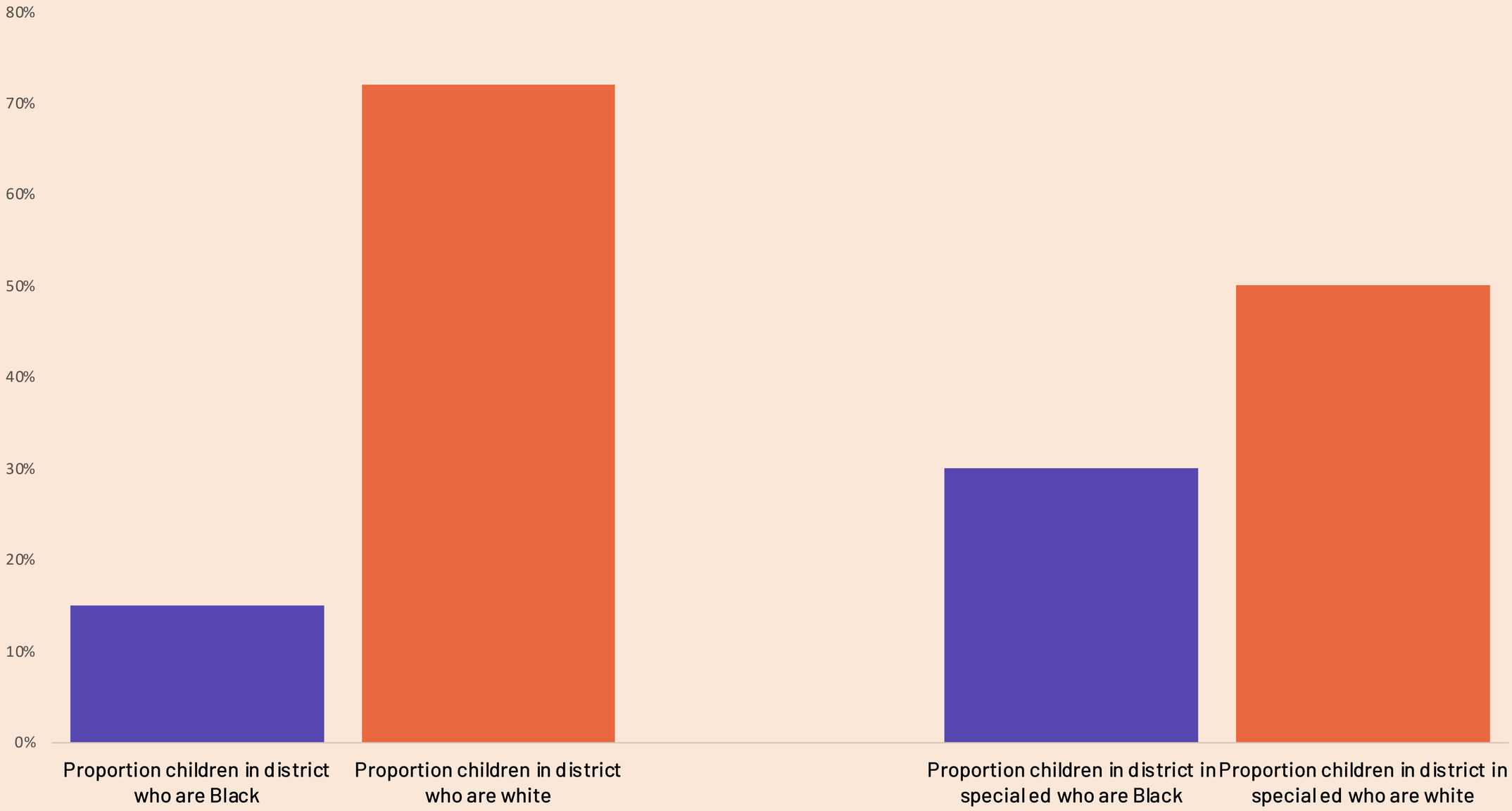
...disproportionality

From an equity POV, we only want to use these when we can safely assume that everyone wants the same thing.

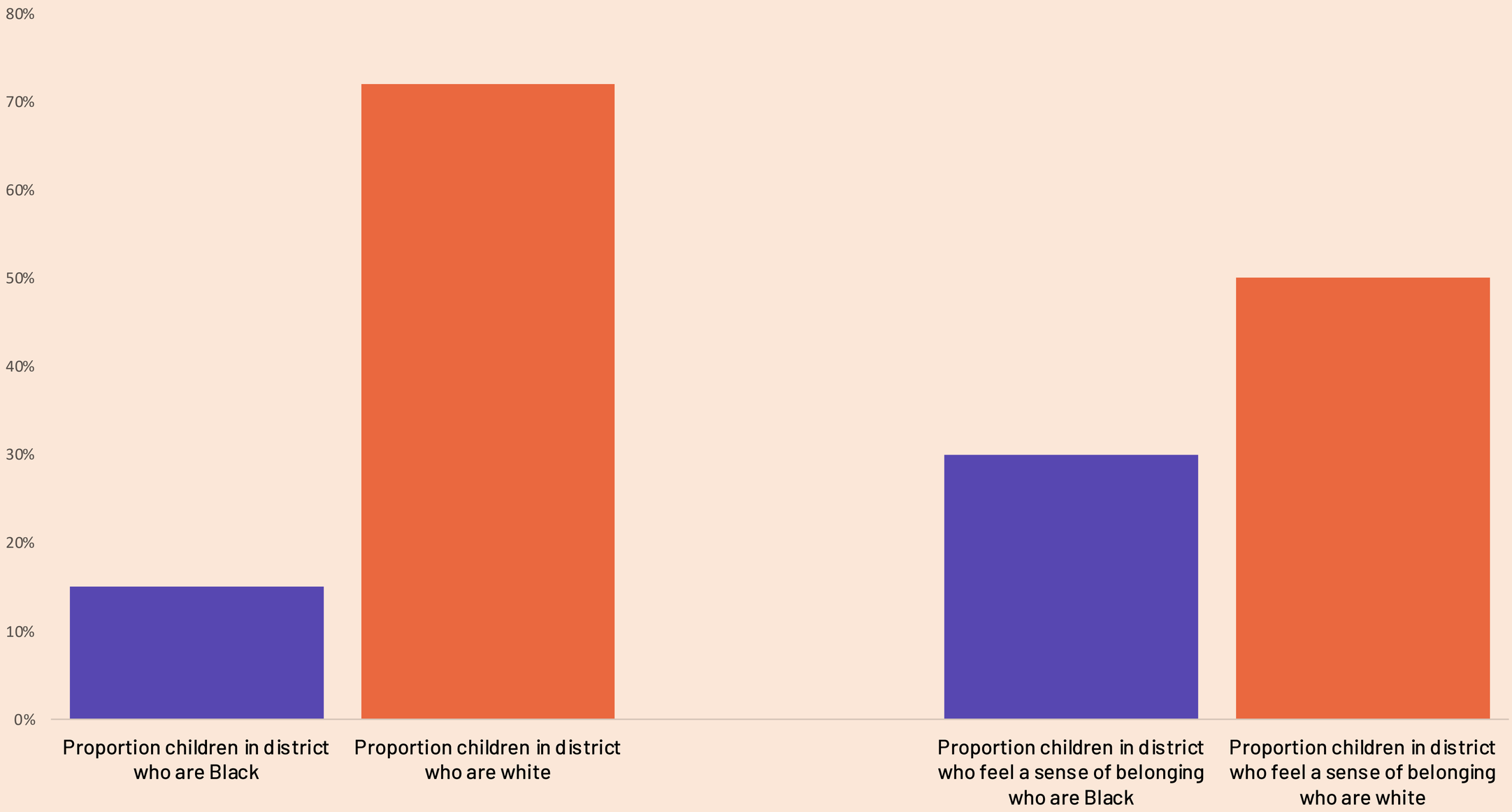
...disproportionality



...disproportionality



...disproportionality



When does a difference become a...

...disparity

From an equity POV, this is a difference that exists due to a lack of access to resources, opportunity, and similar.

Why does any of this matter for equity?

Why does any of this matter for equity?

The End of “Equity Gaps” in Higher Education?

By — Estela Mara Bensimon, Yolanda Watson Spiva

Aug 24, 2022

Why does any of this matter for equity?

US blacks 3 times more likely than whites to get COVID-19

Black people and other people of colour make up 83% of reported COVID-19 cases in Toronto

Coronavirus: Ethnic minorities 'over-exposed' to Covid-19

“It’s not that Black people genetically have a higher risk of getting COVID-19. It’s about the social determinants of health. Blacks often have higher risk factors that make them more susceptible to this disease.”

— Michael Lucien

Why does any of this matter for equity?

Highlighting COVID-19 Racial Disparities Can Reduce Support for Safety Precautions Among White U.S. Residents

AUTHORS
Allison L. Skinner-Dorkenoo, Apoorva Sarmal, Kasheena Rogbeer, Chloe Andre, Bhumi Patel, Leah Cha

Some key early considerations in measurement of differences include:

(a) the point from which differences will be measured (the reference point)
(b) whether differences will be measured on an absolute or relative scale
(c) whether the size of the social group will be accounted for
(d) whether differences will be evaluated with a pair-wise or summary approach
(e) whether extra emphasis should be given to social groups that are of greater concern through differential weighting
(f) how you’re going to draw attention to the root cause

Reference point

There are several choices for a reference point—the rate or other estimate from which differences are measured. Some of the options for a reference point include the largest social group, the social group with the most favorable rate for a health outcome or determinant, the group with the greatest social advantage, the total population, and a target chosen through a planning process.

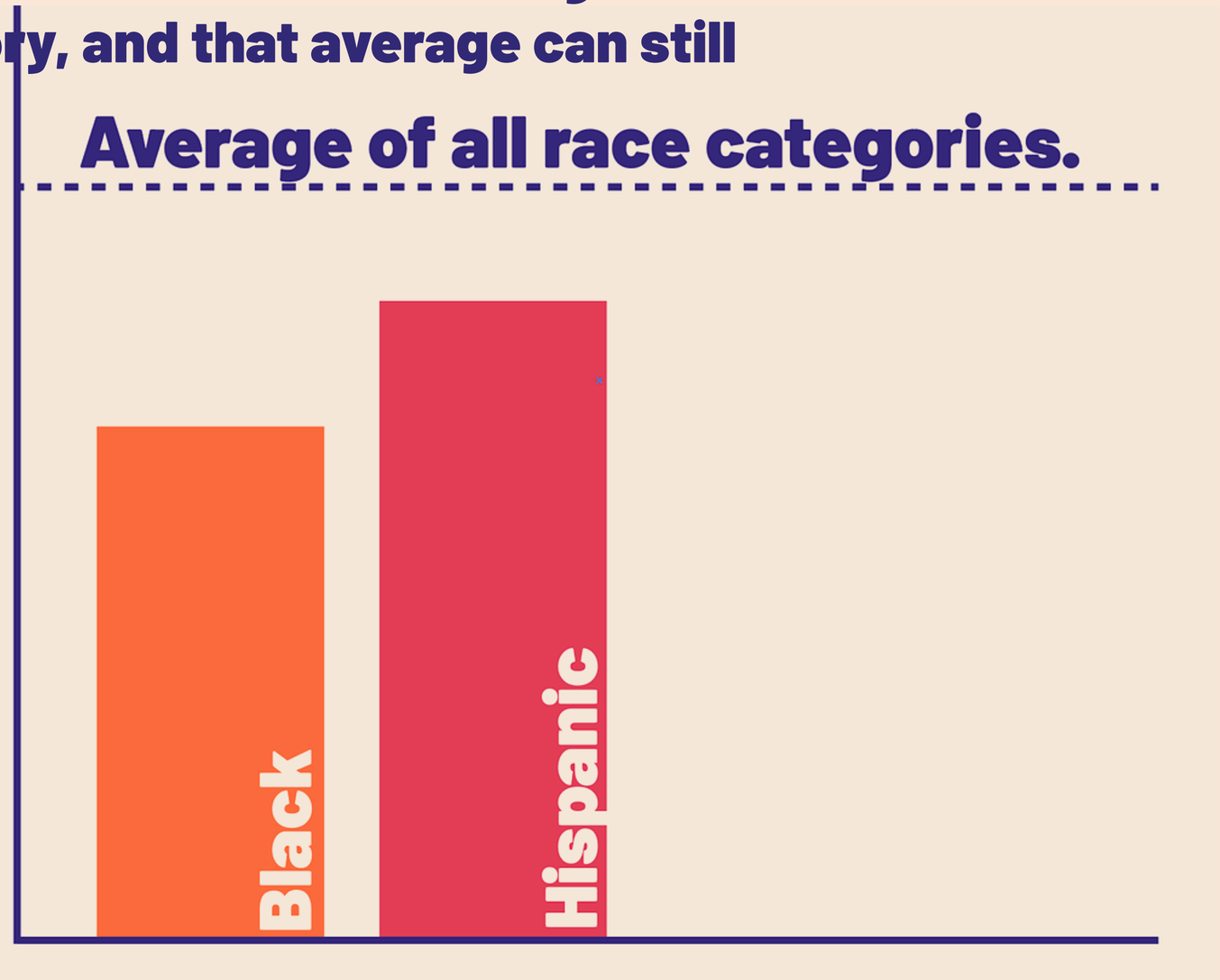
Let’s say I have this data:

Average satisfaction by Race/Ethnicity

White: 86%
Black: 45%
Hispanic: 56%
Asian: 80%

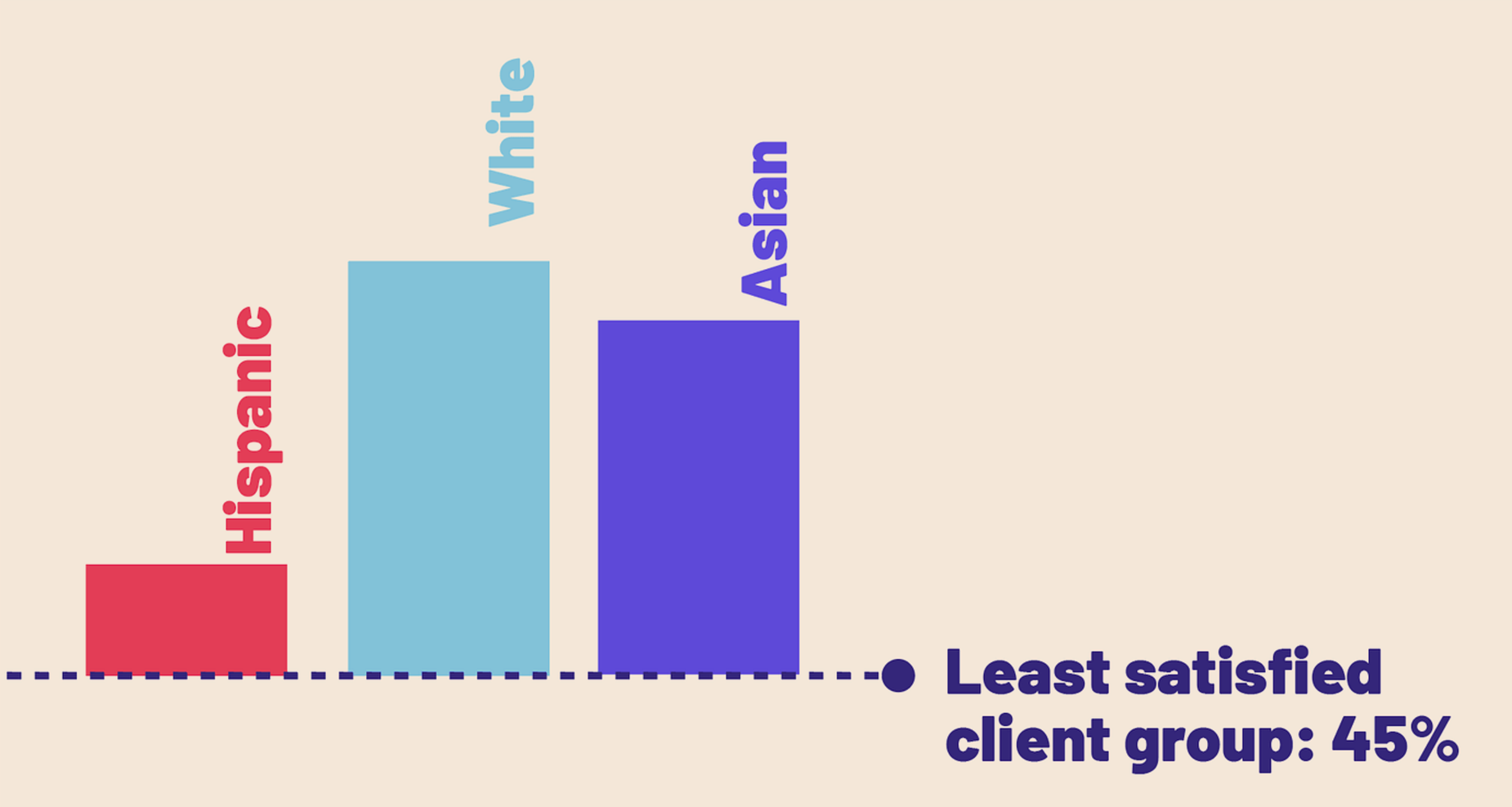
If I only care about certain categories:

I can talk about those groups in relation to an average instead of a reference category, and that average can still



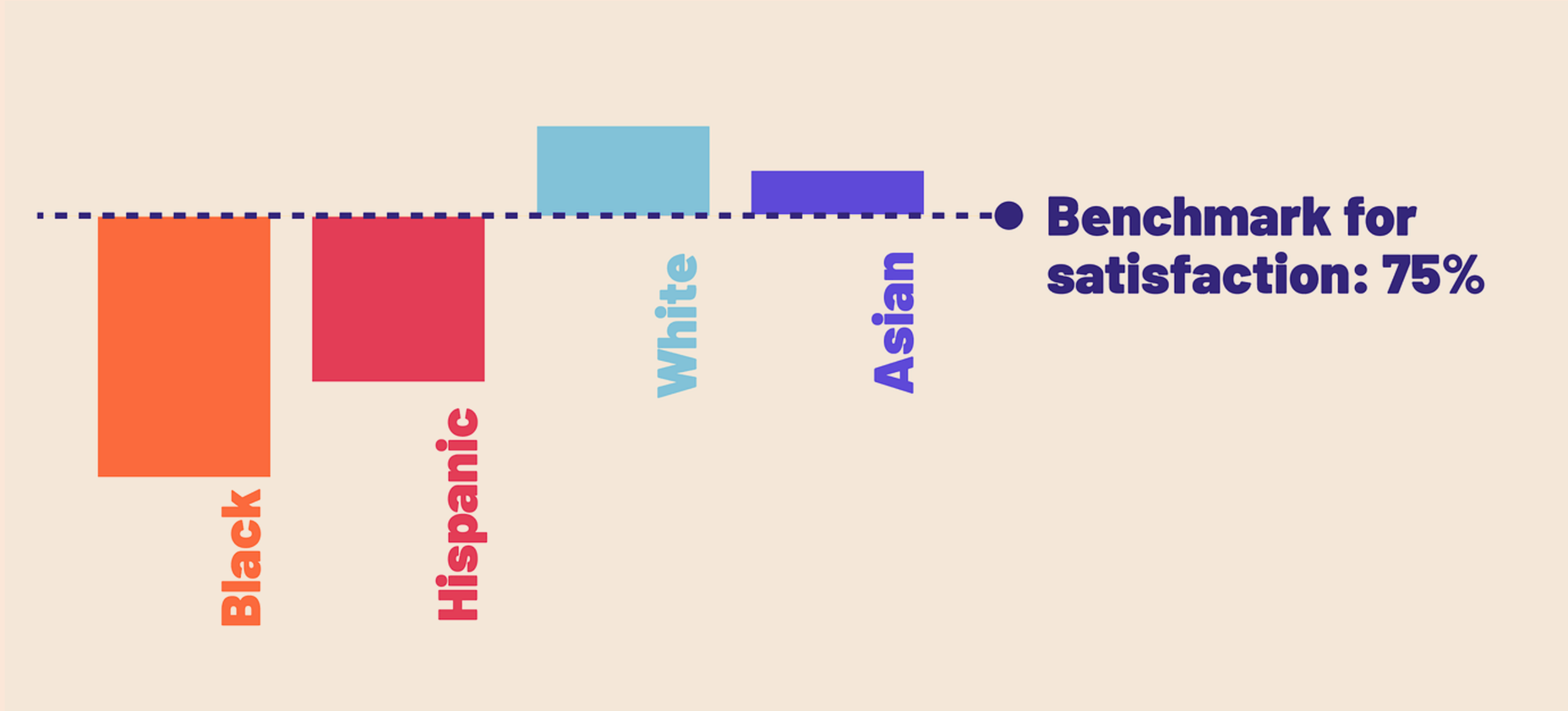
If want to talk about differences above a category:

If I want to put focus on the difference between certain categories and the lowest performing category, I can report it that way, regardless of what the lower category is (and in fact, it might change over time across reports/projects). This way you aren't arbitrarily choosing to always set up a prevailing dichotomy (in this example Black vs White), you are comparing categories against the worst and you don't have to name it.



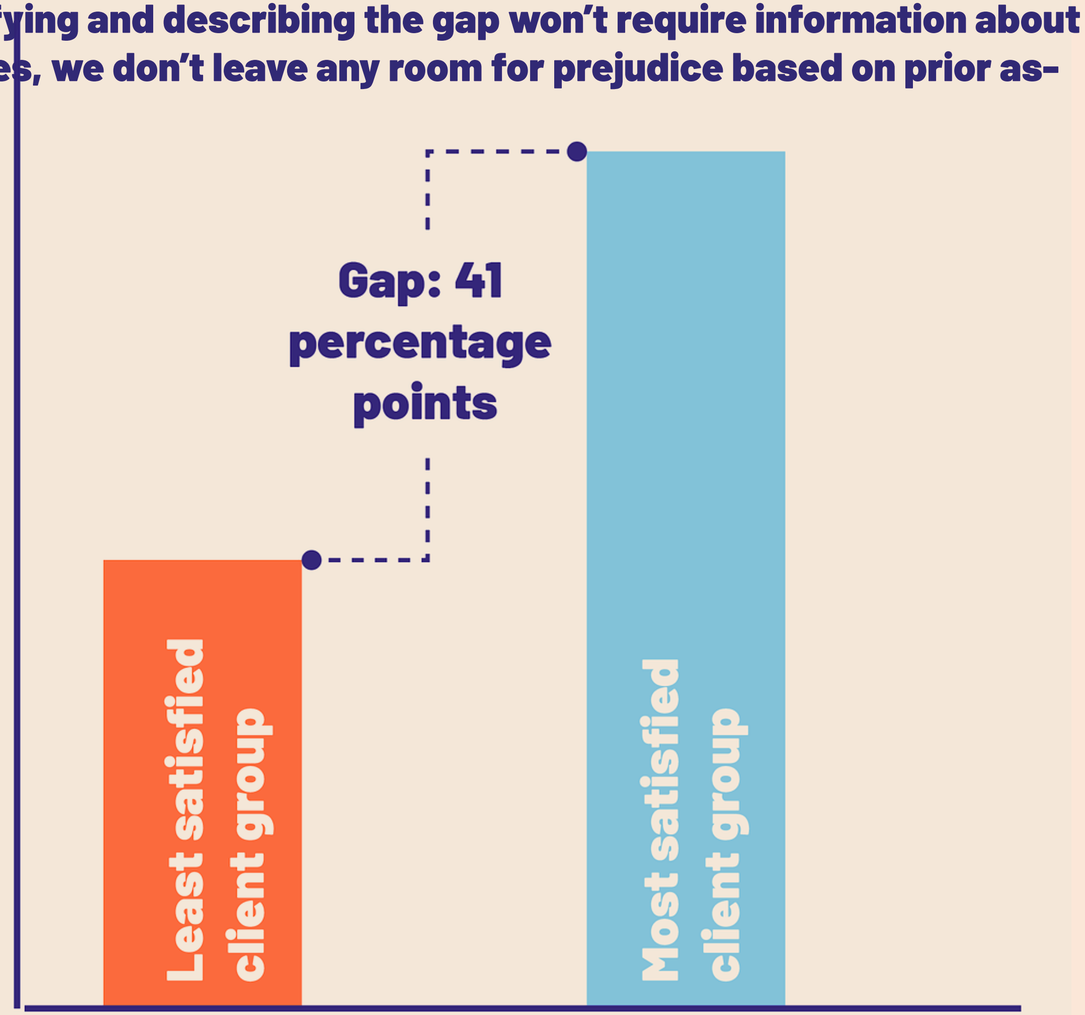
If I want to compare against a meaningful benchmark:

Instead of looking at gap between categories, we can compare under/over performance to a meaningful benchmark to our organization. I really like this one for its subtle yet powerful suggestion that we are measuring ourselves not the characteristics of clients/types of people.



If I care about gaps, but don’t need to emphasize specific pair by pair groups:

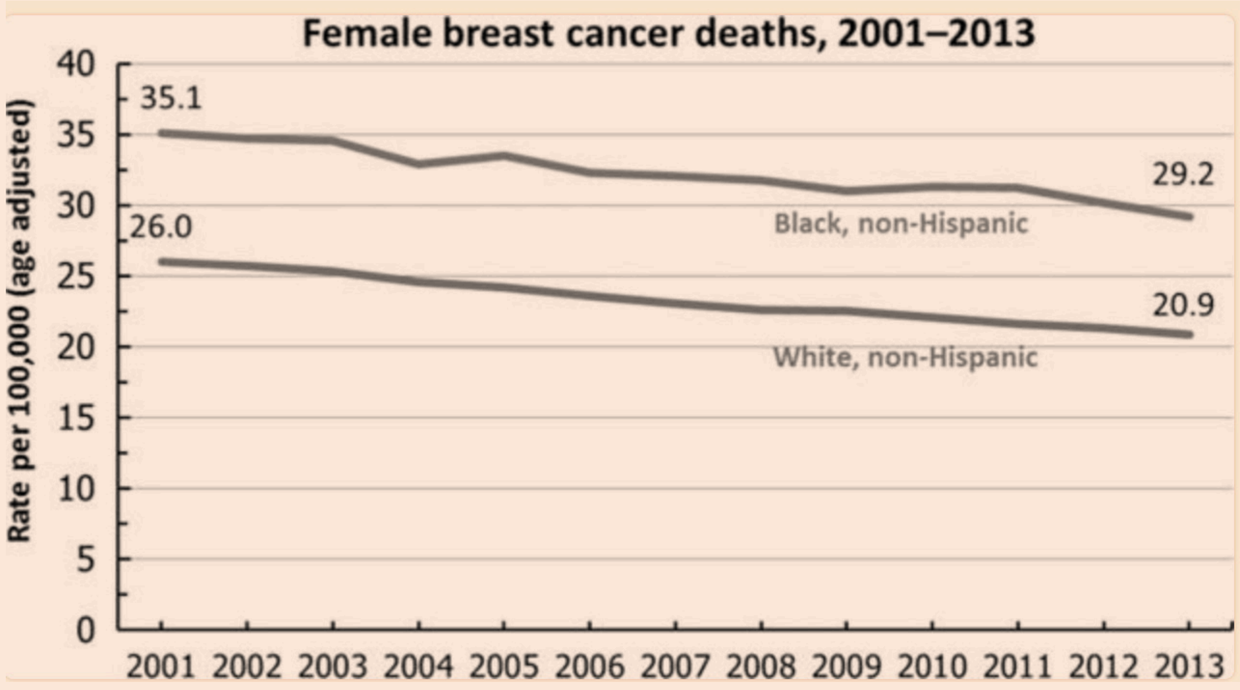
We might just care about the existence and magnitude of a gap regardless of who it’s between, especially at the beginning of exploring these issues, or if the consequences of identifying and describing the gap won’t require information about which groups they are between. By burying the group names, we don’t leave any room for prejudice based on prior assumptions about people.



Absolute versus relative scale

Considering one scale versus the other can lead to different conclusions. For example, in the Figure, rates are decreasing for both non-Hispanic white and non-Hispanic black women. The simple difference (an absolute measure of disparity) between the rates also has decreased, by 0.8 per 100,000 (35.1 – 26.0 = 9.1 in 2001 vs 29.2 – 20.9 = 8.3 in 2013)

However, when one considers the rate ratio, the relative disparity between the 2 groups has increased (rate ratio 35.1 ÷ 26.0 = 1.35 in 2001 is less than the ratio 29.2 ÷ 20.9 = 1.40 in 2013)



Accounting for group size

Measures of absolute or relative disparity are not sufficient for assessing the societal impact or burden of disparities, because the assessment of impact requires a “head count”—the number of persons affected in each group—a concept closely related to the size of groups.

Is there a real difference in rates of infants with disabilities receiving early intervention by race and ethnicity?

Gap between White Infants and Black Infants 35%

Gap between White Infants and American Indian Infants is not statistically significant.

Is there a real difference in rates of infants with disabilities receiving early intervention by race and ethnicity?

Gap between White Infants and Black Infants 35%

P-value = .02

Gap between White Infants and American Indian Infants 45%

P-value = .19

Pair-wise versus summary approach

Population breakdown by race and ethnicity, education, or income often results in 3 or more groups for comparison relative to the reference point. The resulting pair-wise differences between groups can be combined using summary measures.

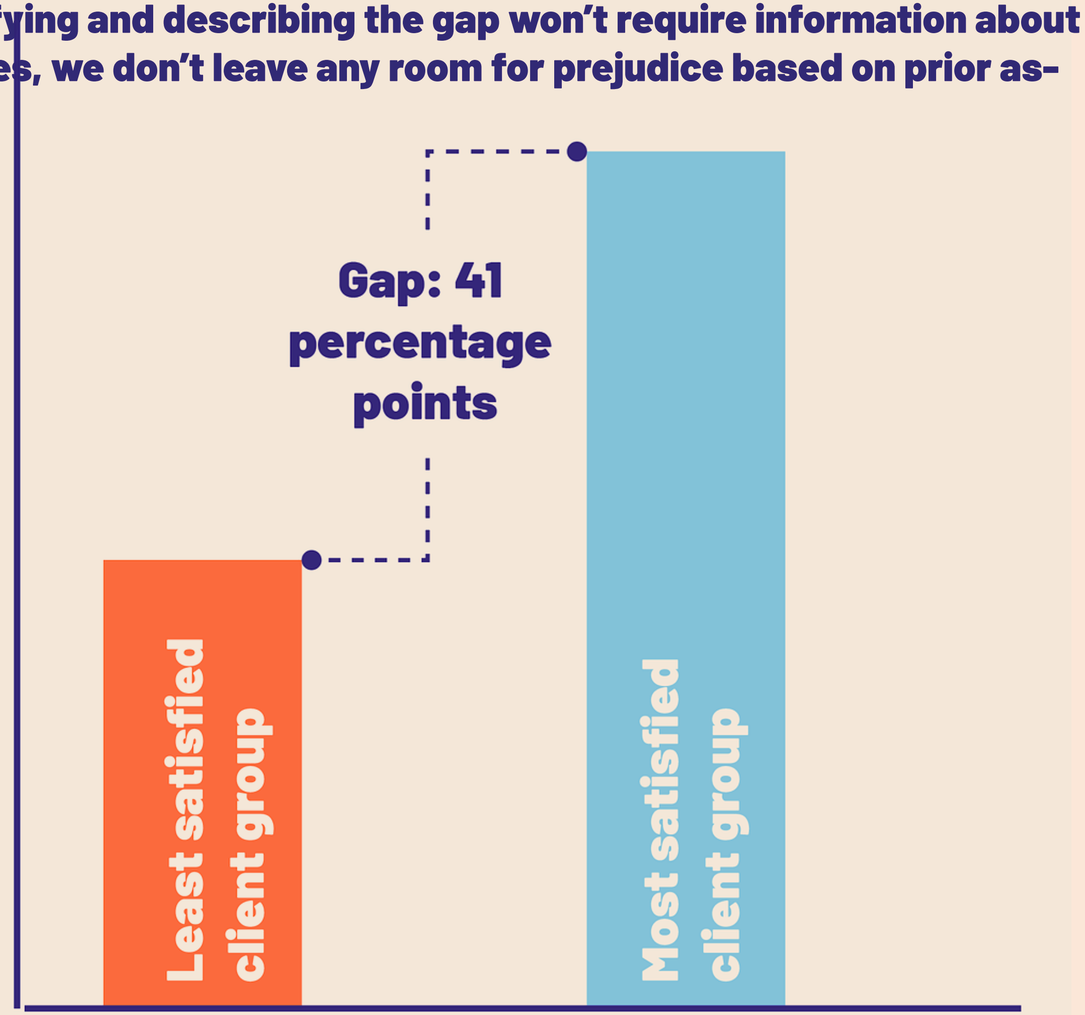
Summary measures are useful in that they quantify the relationship between the distributions of health outcomes and population shares.

Summary measures are especially useful when comparisons among numerous populations as well as among different time periods are desired, because the alternative option of multiple pair-wise comparisons over time can be difficult to interpret.

As an example, if non-Hispanic whites, non-Hispanic blacks, Hispanics, Asians (non-Hispanic), Pacific Islanders (non-Hispanic), and American Indians/Alaska Natives (non-Hispanic) were compared with the population average on a particular outcome over time, there would be 6 different comparisons to monitor, whereas a summary measure monitored over time could indicate whether, overall, racial and ethnic differences for that outcome were increasing or decreasing over time.

If I care about gaps, but don’t need to emphasize specific pair by pair groups:

We might just care about the existence and magnitude of a gap regardless of who it’s between, especially at the beginning of exploring these issues, or if the consequences of identifying and describing the gap won’t require information about which groups they are between. By burying the group names, we don’t leave any room for prejudice based on prior assumptions about people.

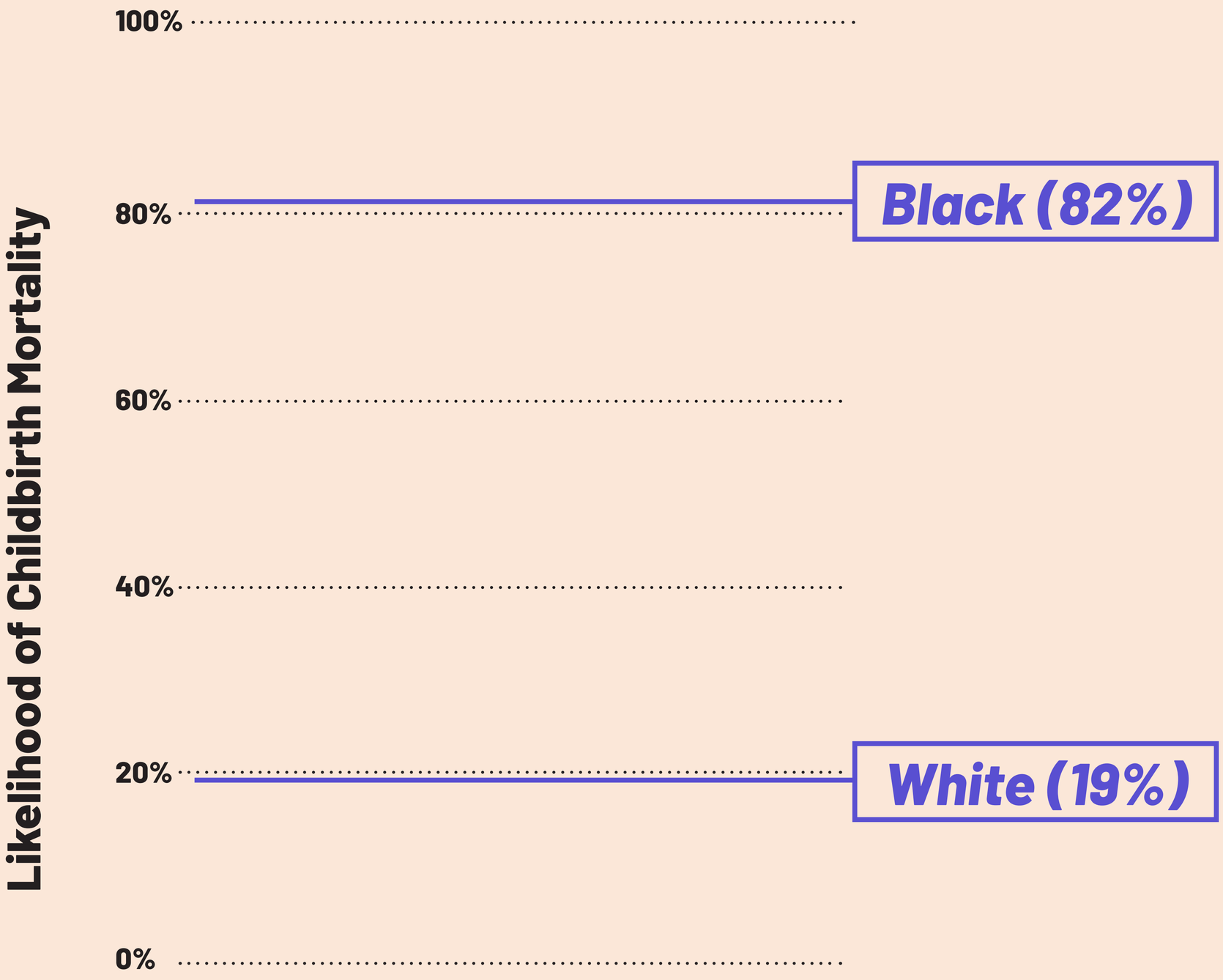


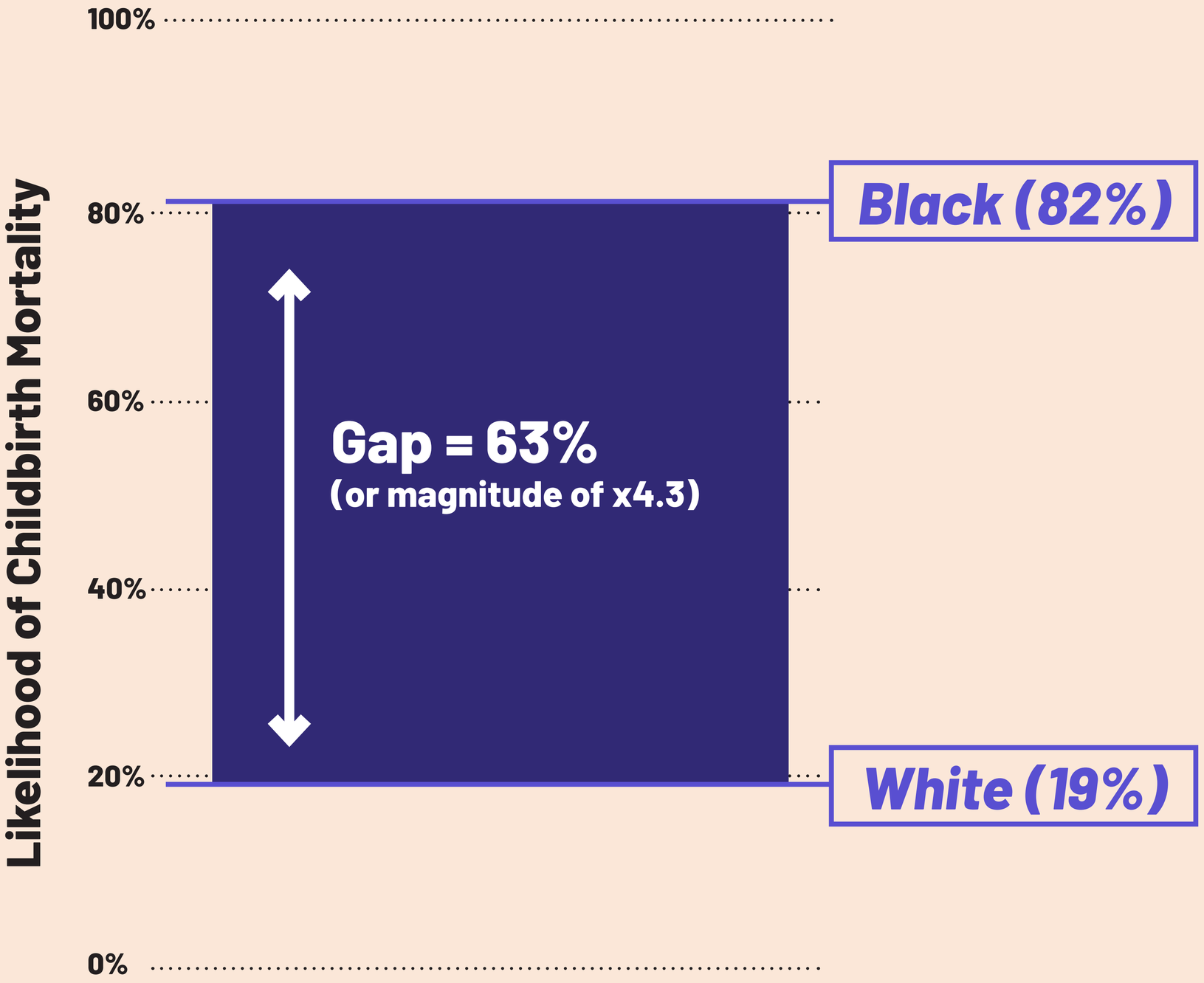
Differential weighting

Because of the “unfavorable social, economic, or political conditions that some groups of people systematically experience based on their relative position in social hierarchies, some social groups may be deemed to be of greatest concern and merit extra emphasis (eg, in the design of interventions, applying a tailored approach to complement a universal approach. Differential weighting can provide the analyst with the measurement tool to account.

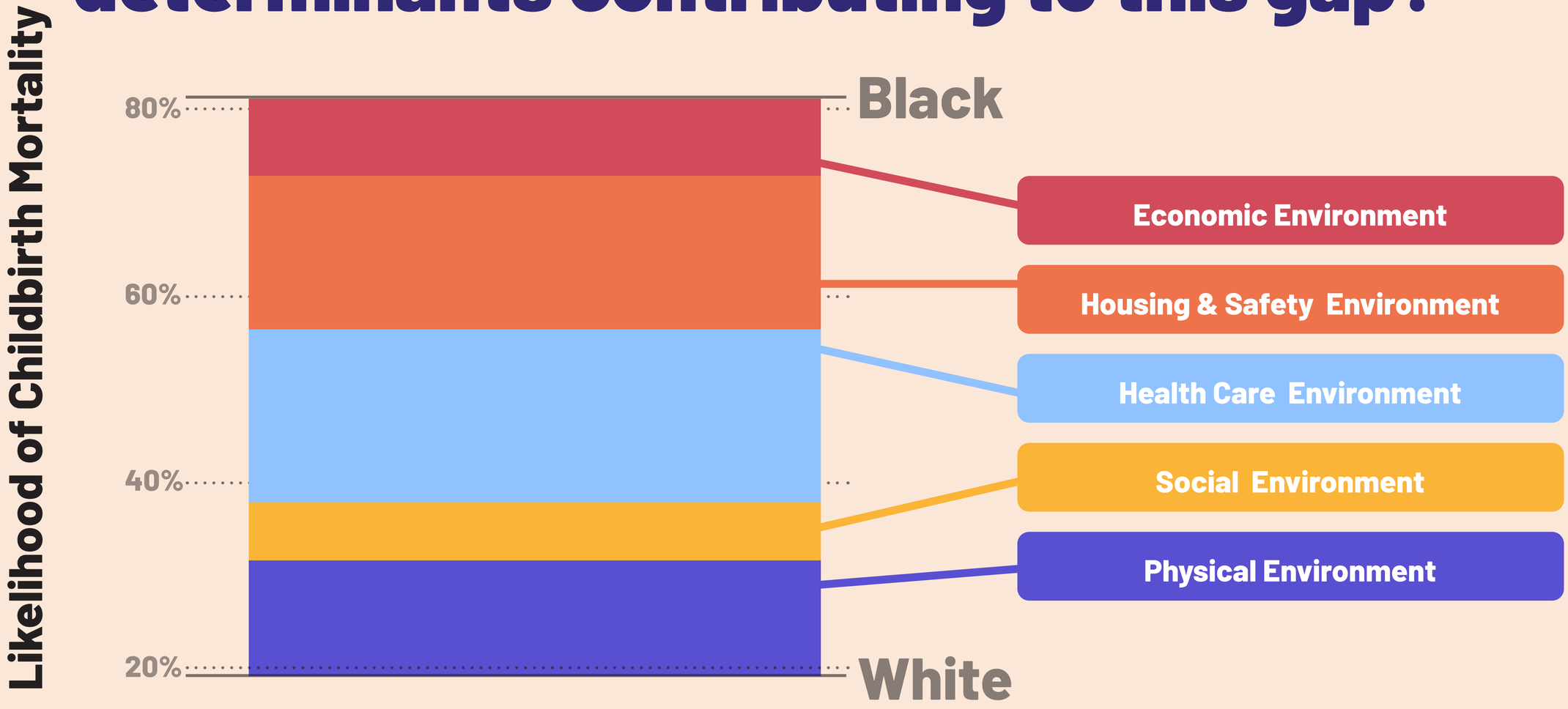
How you're going to draw attention to the root cause

Growing body of research that simple descriptive depiction of "gaps" promotes stereotypes


Likelihood of Childbirth Mortality
100%
80%
60%
40%
20%
0%
Black (82%)
White (19%)


Likelihood of Childbirth Mortality
100%
80%
60%
40%
20%
0%
Gap = 63%
(or magnitude of x4.3)
Black (82%)
White (19%)

How much are each of our structural determinants contributing to this gap?



Gaps in Special Education Placement Journey

Referred for Evaluation

Evaluated for Special Ed

Recommended for Special Ed

Enrolled in Special Ed

Some key early considerations in measurement of differences include:

(a) the point from which differences will be measured (the reference point)
(b) whether differences will be measured on an absolute or relative scale
(c) whether the size of the social group will be accounted for
(d) whether differences will be evaluated with a pair-wise or summary approach
(e) and whether extra emphasis should be given to social groups that are of greater concern through differential weighting

(f) how you’re going to draw attention to the root cause

Absolute measures:

Range Difference

RD is the largest value, among all possible pairwise comparisons, of the absolute differences between two age-adjusted rates for two social groups. RD compares the social group with the worst health status with the social group with the best health status, which represents the largest gap possible. An RD value away from zero indicates disparity.

Absolute Concentration Index

ACI measures whether any particular social group received concentrated health or illness on the absolute scale.

Slope Index of Inequality

SII captures the difference in the average health status between a person in the highest social group and a person in the lowest social group. SII is derived from a simple linear regression model

Between Group Variance

Measures that are based on variance are often used to summarize deviations from a population mean. BGV calculates the weighted sum of squares of the differences between the group-average rates and the population average.

Relative measures:

Range Ratio
RR is calculated by dividing the min and max values so that the rate of the social group with the worst health status and that of the social group with the best health status are compared. The further that RR is from unity, the more disparities exist.

Index of Disparity
IDisp is calculated by first summing the absolute differences between social group rates and a reference rate, then dividing it by the reference rate.

Theil Index
The T index measures general disproportionality. It is a summary of the relative distance that the population is away from the egalitarian state when everyone has the same health status.

Mean Log Deviation
MLD is a population-weighted measure of relative disparity, and it is sensitive to differences from the average rate

Relative Concentration Index
RCI measures the extent to which health or illness is concentrated among particular social groups. RCI was originally intended for social groups that have an inherent ranking, such as income or education groups.

Absolute Numbers

In our school, we suspended 437 white students, 16 American Indian students, 26 Asian students, and 333 Black students.

Pro: Can help understand whether discipline is overused that a rate may not show. Provides information about the number or count of a group or category of interest; it can help you gain a perspective that is very grounded.

Con: Does not show relative differences equally across groups. Is not comparable over time if the population changes.

Rate

In our school, we suspended 1.6% of white students, 3.7% of American Indian students, 0.7% of Asian students, and 6.7% of Black students.

Pro: Provides you with an idea of proportions across groups so can be compared.

Con: Does not provide information about disproportionality between groups.

Relative rate ratio

In our school, we suspended American Indian students 2.3 more times than white students; Asian students 0.4 more times than white students, and we suspended Black students 4.2 times more frequently than white students.

Pro: Identifies disproportional rates by examining the relative difference between groups.

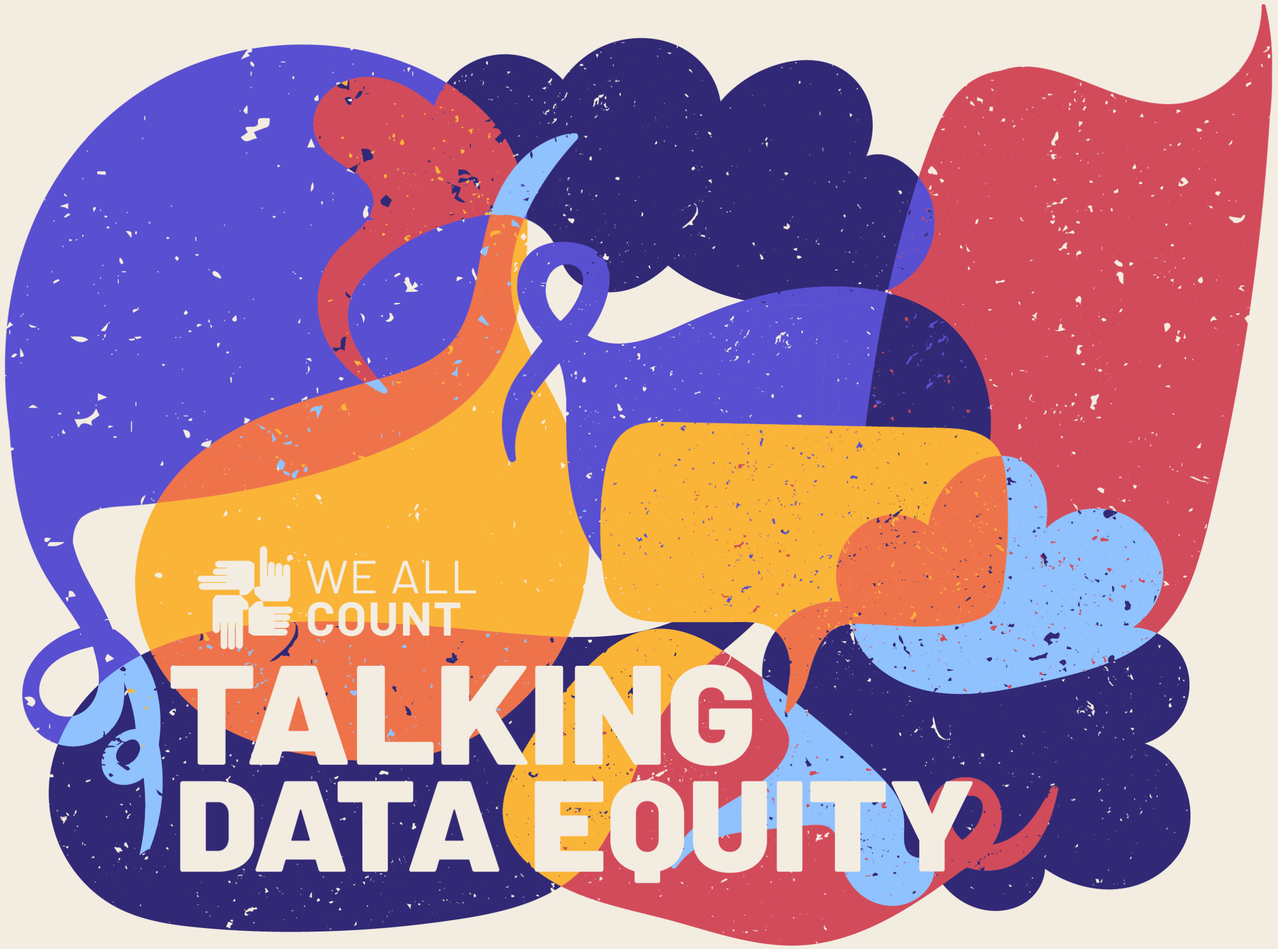
Con: Highly sensitive to small group size and centers the group that is chosen as the baseline. Does not tell you the number or rate of events (in this example suspensions) so you are unable to determine if these are at an acceptable level.

Composition Index

White students represent 56% of our school and 39% of the students in our school who were suspended.
American Indian students represent 1% of our school and 1.4% of the students in our school who were suspended.
Asian students represent 8% of our school and 2% of the students in our school who were suspended.
Black students represent 10% of our school and 30% of the students in our school who were suspended.

Pro: Provides information about the proportion of a group that is disciplined

Con: Does not tell you if the proportion of the group is disproportionate compared to another group. Does not tell you the number or rate of events (in this example suspensions) so you are unable to determine if these are at an acceptable level.



No TDE next week
Back March 24
with Howard Shih from AAPI Data

Disparity Measure	Absolute or Relative	Reference Group	All Social Groups	Reflect SES Gradient	Social Group Weighting	Inequality Aversion Parameter	Graphi Analc
Absolute difference	Absolute	Best	No	Yes	No	No	Yes
Relative difference	Relative	Best	No	Yes	No	No	Yes
Slope index of inequality	Absolute	Average	Yes	Yes	Yes	No	Yes
Relative index of inequality	Relative	Average	Yes	Yes	Yes	No	Yes
Index of disparity	Relative	Best	Yes	No	No	No	No
Relative concentration index	Relative	Average	Yes	Yes	Yes	Yes	Yes
Absolute concentration index	Absolute	Average	Yes	Yes	Yes	Yes	Yes
Between group variance	Absolute	Average	Yes	No	Yes	Yes	No
Theil index	Relative	Average	Yes	No	Yes	Yes	No
Mean log deviation	Relative	Average	Yes	No	Yes	Yes	No

Odds ratio (black vs. white VLBW neonatal mortality)

Unadjusted 1.19

Adiusted for unavoidable factors (sex of infant, congenital anomaly, multiple birth,ear of birth) 1.17

Adjusted for unavoidable factors, and choice (marital status, prenatal care, tobacco use, alcohol use) 0.95

Adjusted for unavoidable factors, choice, and system factors (insurance, delivery type, hospital of birth) 1.00

Adjusted for unavoidable factors, choice, system factors, and pathways (birthweight, maternal education, maternal complications,Apgar score) 0.62

The Racial Disparity Index (also known as a risk ratio or relative risk index)

Because the size of the risk ratio is affected by the school-level racial/ethnic demographics of the comparison group. The risk for the comparison group is jointly influenced by the racial/ethnic composition of the comparison group and the risk for each of those racial/ethnic. The district's rates of disproportionality will be inflated due to a failure to disaggregate schools with large populations of children of color from schools with small populations.

1. Attributable Disparity (or Risk Difference)
The amount of an event that is specifically due to social identity factor of interest.
(Subtract the absolute risk between group A and All Else)

2. Weighted Disparity Ratio
A weighted risk ratio adjusts for school variability in the racial/ethnic composition of the comparison group. Because the weighted formula accounts for variance in the size of ethnic groups in each school, it is more accurate and more stable over time.

3. Alternate Disparity Ratio
This one addresses the problems associated with small numbers.
The alternate risk ratio uses school-level data to calculate the risk for the racial/ethnic group and district-level data to calculate the risk for the comparison group.

4. Raw Differential Representation
The cost is that risk-based metrics do not provide information about the actual number of children and youth impacted by disproportionality– information that is potentially critical for understanding disproportionality in terms of the extent of harm that it is causing or resources that may be necessary to address it.

The RDR, is a metric for computing disproportionality, is the estimated number of students in a target group who did experience discipline (or any other metric) but who would not have if students in that group were subject to discipline at the same rate as students in the reference group. Calculated by the number of Group A students multiplied by the (Group A – Reference Disparity Difference).

5. Racial Disparity Index
This is defined as the average of the absolute difference between rates of disparity for specific groups within the population. One of the big problems in all other measures in that they require a comparison between two groups. These pairwise comparisons are frequently inappropriate. To describe the level of actual disparity in the system, a more inclusive summary measure of the magnitude of the difference is needed. (Similar but not the same as coefficient of variation.)

6. Odds Ratio
The odds ratio is the measure of choice in a case-control study.
So, what does an OR mean? Here it is in plain language.

- **An OR of 1.2 means there is a 20% increase in the odds of an outcome with a given exposure.**
- **An OR of 2 means there is a 100% increase in the odds of an outcome with a given exposure. Or this could be stated that there is a doubling of the odds of the outcome. Note, this is not the same as saying a doubling of the risk.**
- **An OR of 0.2 means there is an 80% decrease in the odds of an outcome with a given exposure.**

The Racial Disparity Index (also known as a risk ratio or relative risk index)

Because the size of the risk ratio is affected by the school-level racial/ethnic demographics of the comparison group. The risk for the comparison group is jointly influenced by the racial/ethnic composition of the comparison group and the risk for each of those racial/ethnic. The district's rates of disproportionality will be inflated due to a failure to disaggregate schools with large populations of children of color from schools with small populations.

1. Attributable Disparity (or Risk Difference)
The amount of an event that is specifically due to social identity factor of interest.
(Subtract the absolute risk between group A and All Else)

2. Weighted Disparity Ratio
A weighted risk ratio adjusts for school variability in the racial/ethnic composition of the comparison group. Because the weighted formula accounts for variance in the size of ethnic groups in each school, it is more accurate and more stable over time.

3. Alternate Disparity Ratio
This one addresses the problems associated with small numbers.
The alternate risk ratio uses school-level data to calculate the risk for the racial/ethnic group and district-level data to calculate the risk for the comparison group.

4. Raw Differential Representation
The cost is that risk-based metrics do not provide information about the actual number of children and youth impacted by disproportionality– information that is potentially critical for understanding disproportionality in terms of the extent of harm that it is causing or resources that may be necessary to address it.

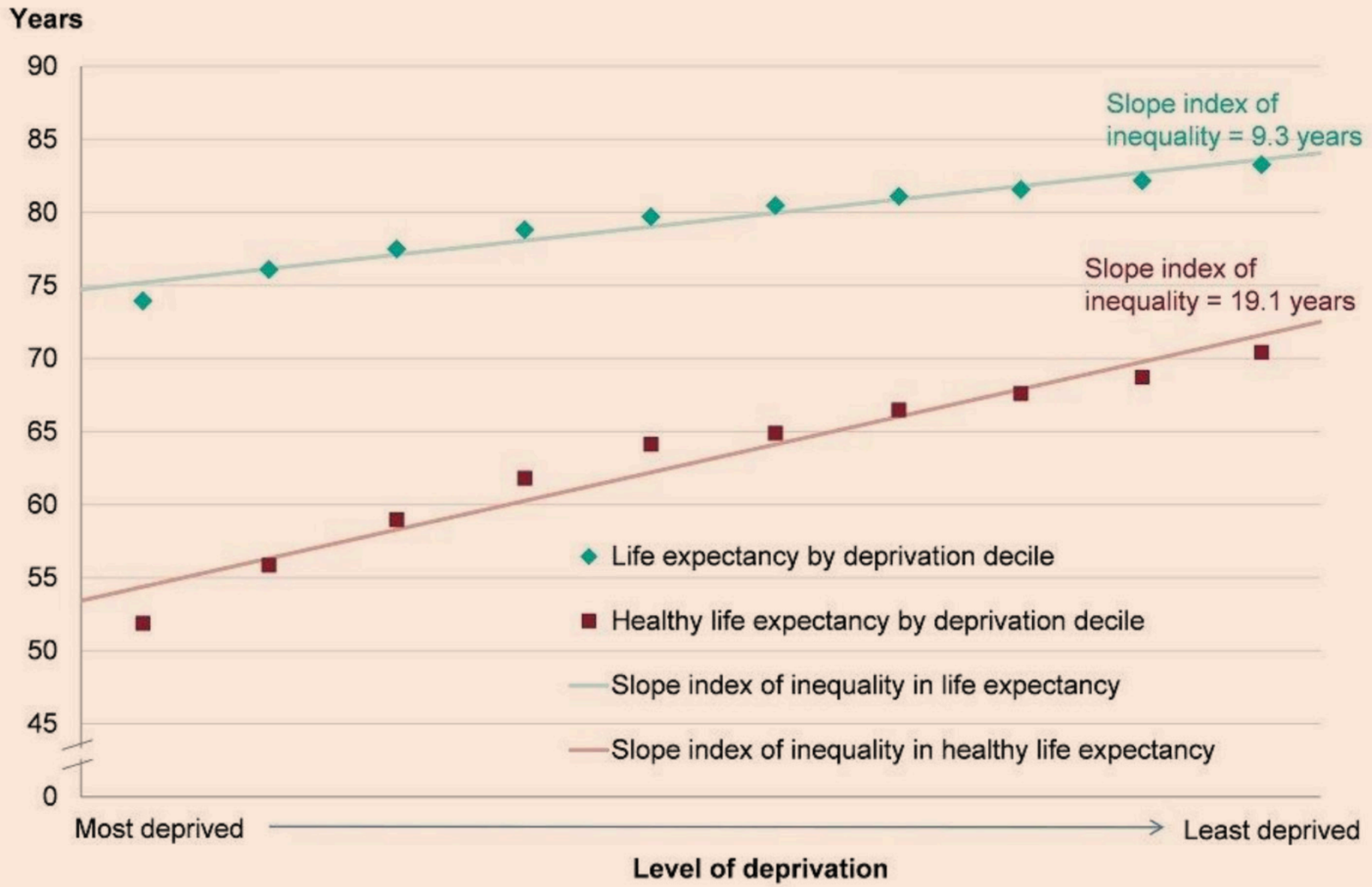
The RDR, is a metric for computing disproportionality, is the estimated number of students in a target group who did experience discipline (or any other metric) but who would not have if students in that group were subject to discipline at the same rate as students in the reference group. Calculated by the number of Group A students multiplied by the (Group A – Reference Disparity Difference).

5. Racial Disparity Index
This is defined as the average of the absolute difference between rates of disparity for specific groups within the population. One of the big problems in all other measures in that they require a comparison between two groups. These pairwise comparisons are frequently inappropriate. To describe the level of actual disparity in the system, a more inclusive summary measure of the magnitude of the difference is needed. (Similar but not the same as coefficient of variation.)

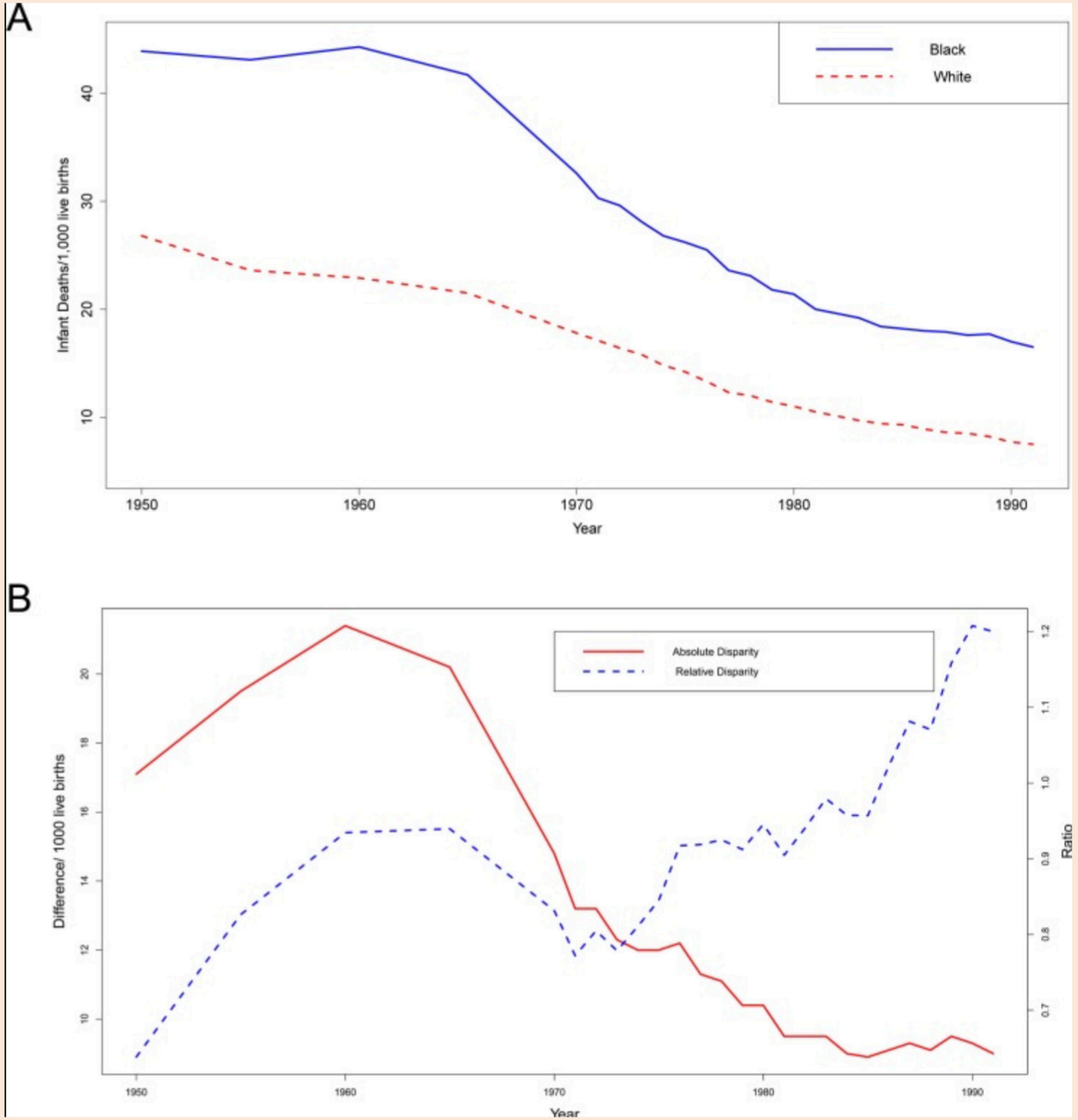
6. Odds Ratio
The odds ratio is the measure of choice in a case-control study.
So, what does an OR mean? Here it is in plain language.

- **An OR of 1.2 means there is a 20% increase in the odds of an outcome with a given exposure.**
- **An OR of 2 means there is a 100% increase in the odds of an outcome with a given exposure. Or this could be stated that there is a doubling of the odds of the outcome. Note, this is not the same as saying a doubling of the risk.**
- **An OR of 0.2 means there is an 80% decrease in the odds of an outcome with a given exposure.**

Figure 1: slope index of inequality in life expectancy and healthy life expectancy at birth, males, England, 2014 to 2016



Source: Public Health Outcomes Framework.


A
Black
White
Infant Deaths/1,000 live births
10
20
30
40
1950
1960
1970
1980
1990
Year
B
Absolute Disparity
Relative Disparity
Difference/ 1000 live births
10
12
14
16
18
20
Ratio
0.7
0.8
0.9
1.0
1.1
1.2
1950
1960
1970
1980
1990
Year