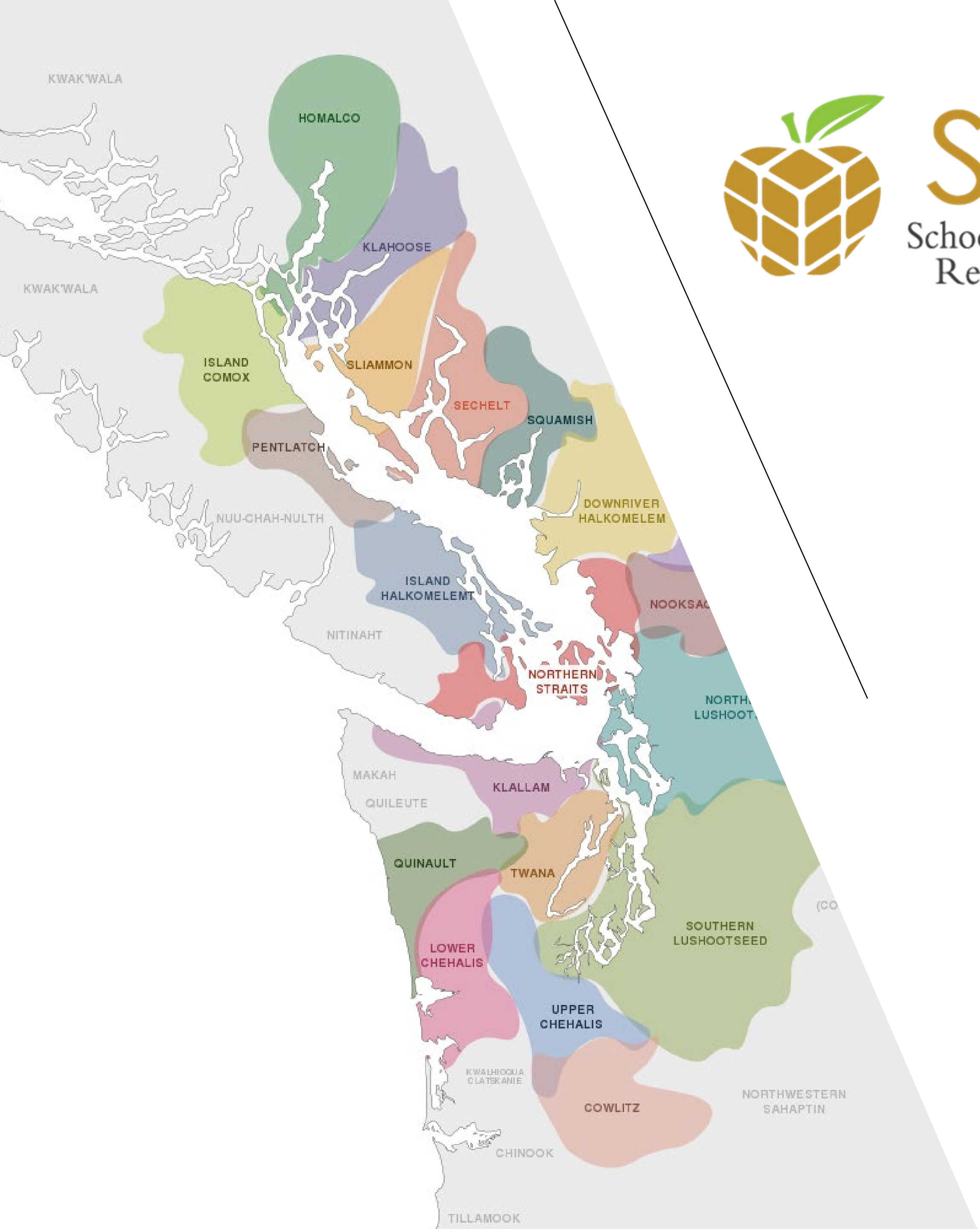
SMART
School Mental Health Assessment
Research & Training Center

HARBORVIEW
INJURY PREVENTION
& RESEARCH CENTER
UNIVERSITY of WASHINGTON

KEITH HULLENAAR, PHD
QUANTITATIVE METHODS "ENJOYER"
UNIVERSITY OF WASHINGTON
SMART CENTER & HIPRC
PSYCHIATRY AND BEHAVIORAL SCIENCE

SMARTSTATS:
FOUNDATIONS OF APPLIED STATISTICAL MODELING



SMART
School Mental Health Assessment
Research & Training Center

HARBORVIEW
INJURY PREVENTION
& RESEARCH CENTER
UNIVERSITY of WASHINGTON

LAND ACKNOWLEDGEMENT

We live and work on the **unceded** ancestral lands of the **Coast Salish people**, the land which touches the shared waters of all tribes and bands within the **Duwamish**, **Puyallup**, **Suquamish**, **Tulalip** and **Muckleshoot** nations, and pay our respects to elders past and present.

We work to create inclusive and respectful partnerships that **honor Indigenous cultures, histories, identities, and sociopolitical realities**, that dismantle ongoing legacies of settler colonialism, and that recognize the hundreds of Indigenous Nations who continue to resist, live, and uphold their sacred relations across their lands.

WELCOME TO SMARTSTATS

Our **mission** is to make quantitative methodologies freely accessible to **all who want to learn**.

Keith Hullenaar, PhD
SMARTstats Founder
Postdoc

Bethlehem Kebede, BS
SMART Center
Research Analyst

Casey Ehde, BA
SMART Center
Research Coordinator

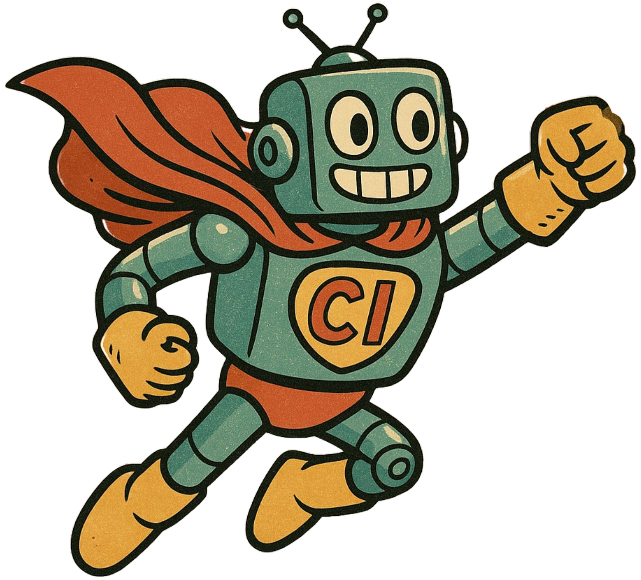
Mahima Joshi, MPH
SMART Center
Research Scientist I

WELCOME TO SMARTSTATS

Our **mission** is to make quantitative methodologies freely accessible to **all who want to learn**.

Keith Hullenaar, PhD
SMARTstats Founder
Postdoc

Bethlehem Kebede, BS
SMART Center
Research Analyst



Casey Ehde, BA
SMART Center
Research Coordinator

Mahima Joshi, MPH
SMART Center
Research Scientist I

Causal Inference Man!

AGENDA

Unpacking the regression model

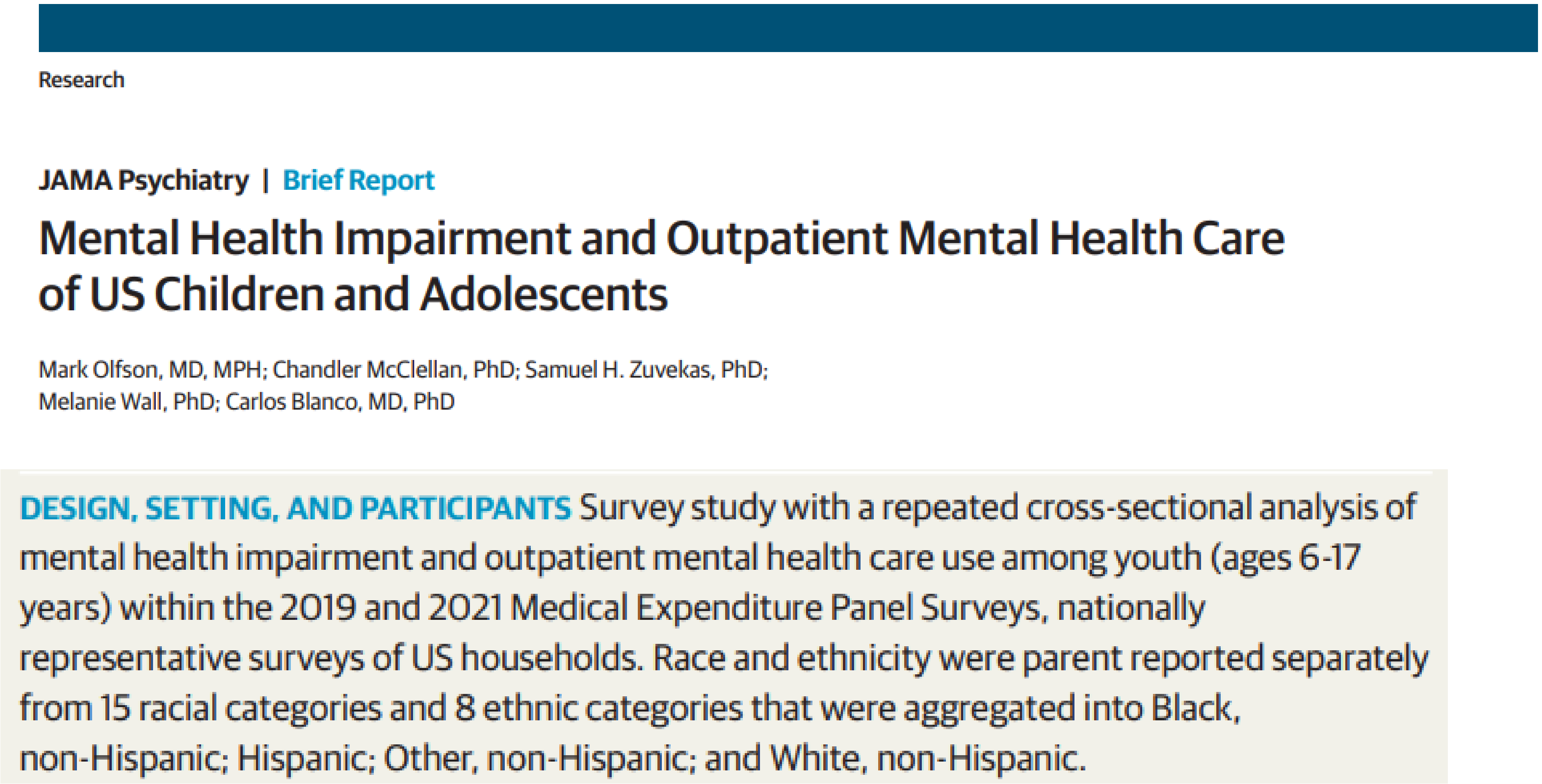
Key assumptions overview

Discussion

QUESTION TIME!

Join mentimeter.com

Type in code: **5363 3331**



Research

JAMA Psychiatry | Brief Report

Mental Health Impairment and Outpatient Mental Health Care of US Children and Adolescents

Mark Olfson, MD, MPH; Chandler McClellan, PhD; Samuel H. Zuvekas, PhD;
Melanie Wall, PhD; Carlos Blanco, MD, PhD

DESIGN, SETTING, AND PARTICIPANTS Survey study with a repeated cross-sectional analysis of mental health impairment and outpatient mental health care use among youth (ages 6-17 years) within the 2019 and 2021 Medical Expenditure Panel Surveys, nationally representative surveys of US households. Race and ethnicity were parent reported separately from 15 racial categories and 8 ethnic categories that were aggregated into Black, non-Hispanic; Hispanic; Other, non-Hispanic; and White, non-Hispanic.

HOW DID **OUTPATIENT YOUTH MENTAL HEALTHCARE USE** CHANGE DURING COVID?

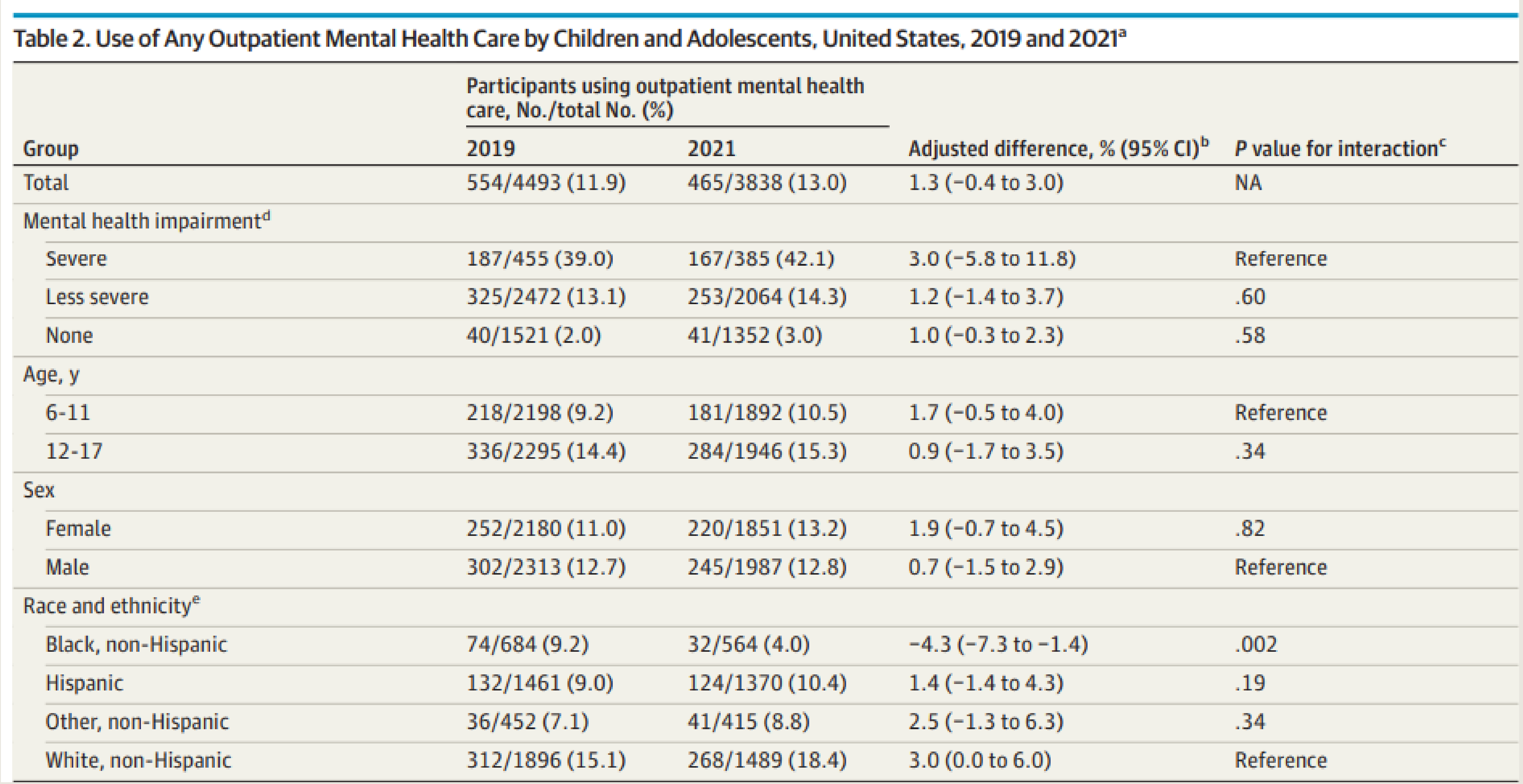


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Group	Participants using outpatient mental health care, No./total No. (%)			
	2019	2021	Adjusted difference, % (95% CI)[b]	P value for interaction[c]
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

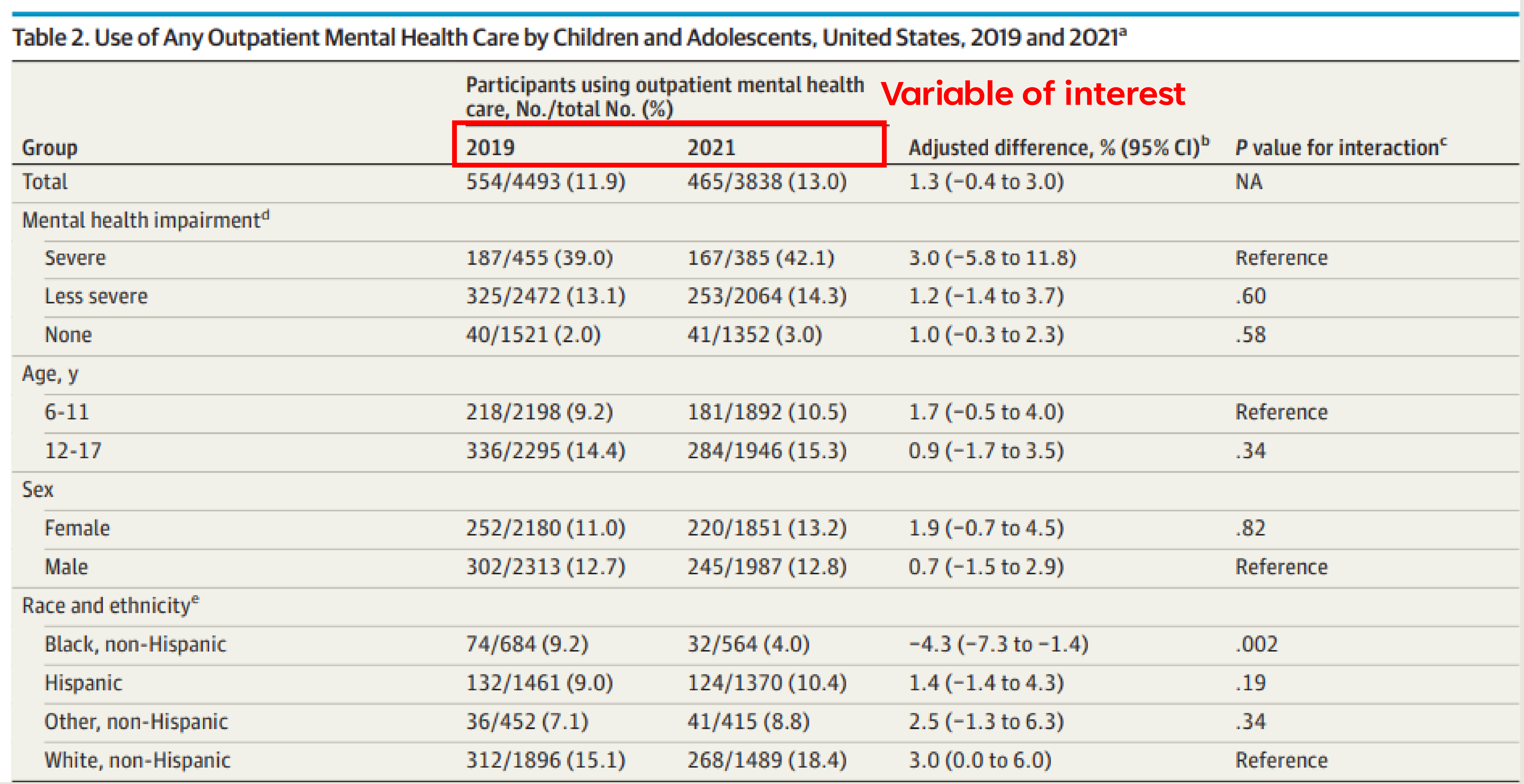


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Variable of interest

Group	Participants using outpatient mental health care, No./total No. (%)		Adjusted difference, % (95% CI)[b]	P value for interaction[c]
	2019	2021		
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

STRATIFICATION (GROUPS)

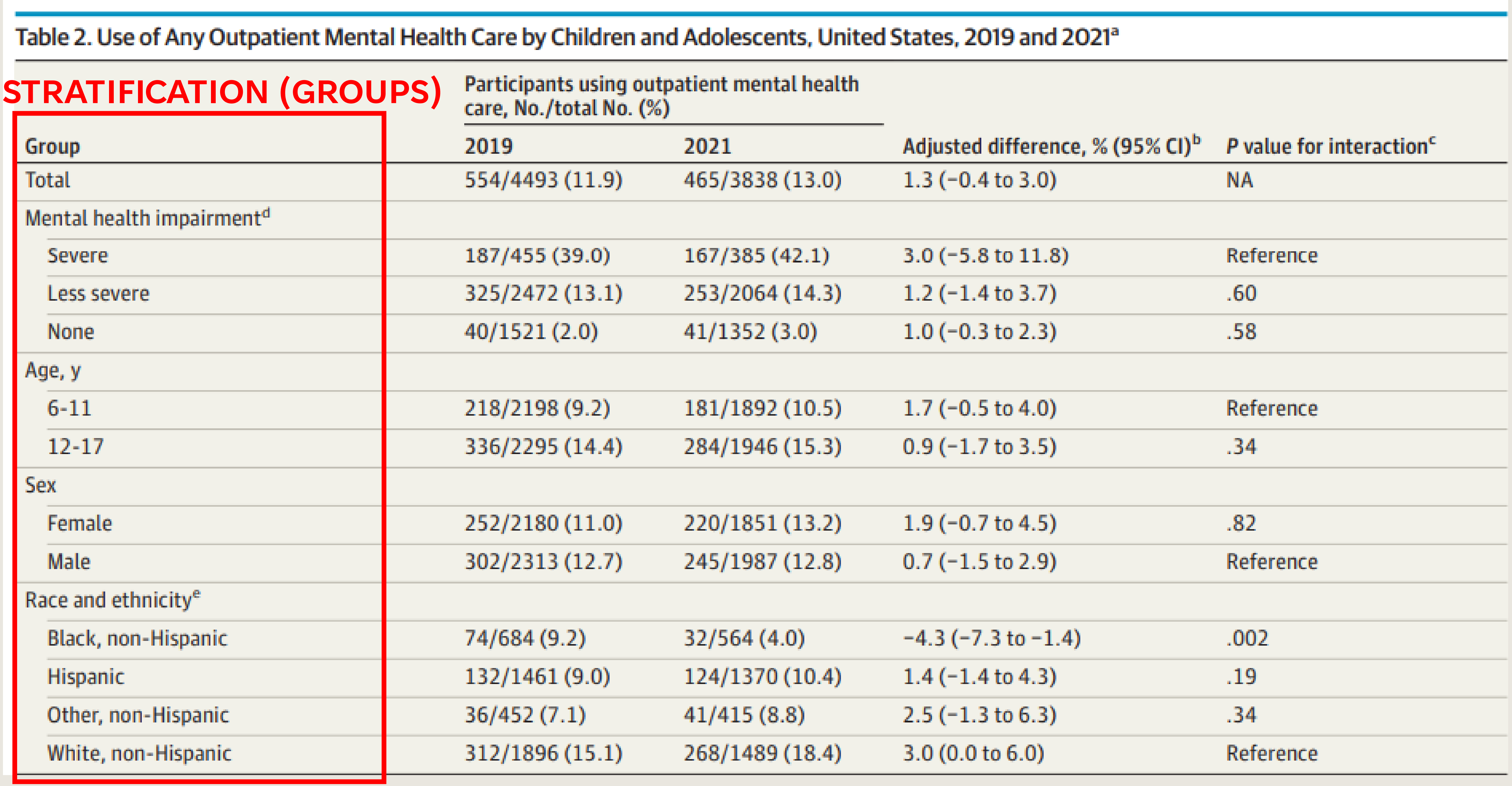


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Group	Participants using outpatient mental health care, No./total No. (%)			
	2019	2021	Adjusted difference, % (95% CI)[b]	P value for interaction[c]
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

(VARIABLES AS WELL!)

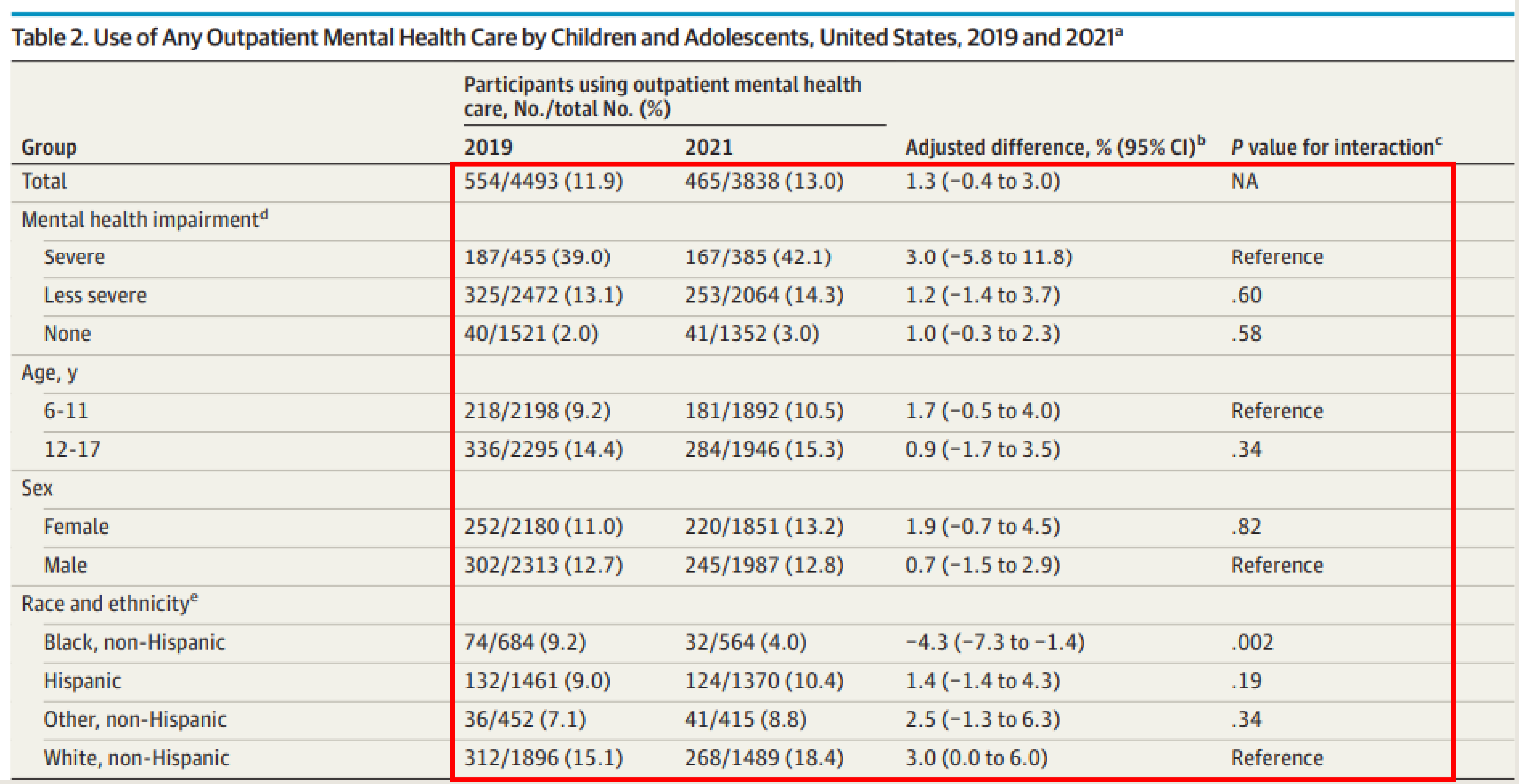


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Group	Participants using outpatient mental health care, No./total No. (%)			
	2019	2021	Adjusted difference, % (95% CI)[b]	P value for interaction[c]
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

INFERENTIAL STATISTICS

INFERENTIAL VERSUS DESCRIPTIVE STATISTICS

Descriptive statistics: Summarize features of a dataset, such as percentages, means, or medians. **They do not inform about the population or account for other variables.**

Inferential statistics: Summarize features of a population using a **probability-based sample**. These are the **things** estimated by statistical models.

In inference: all models are wrong, but some are useful.

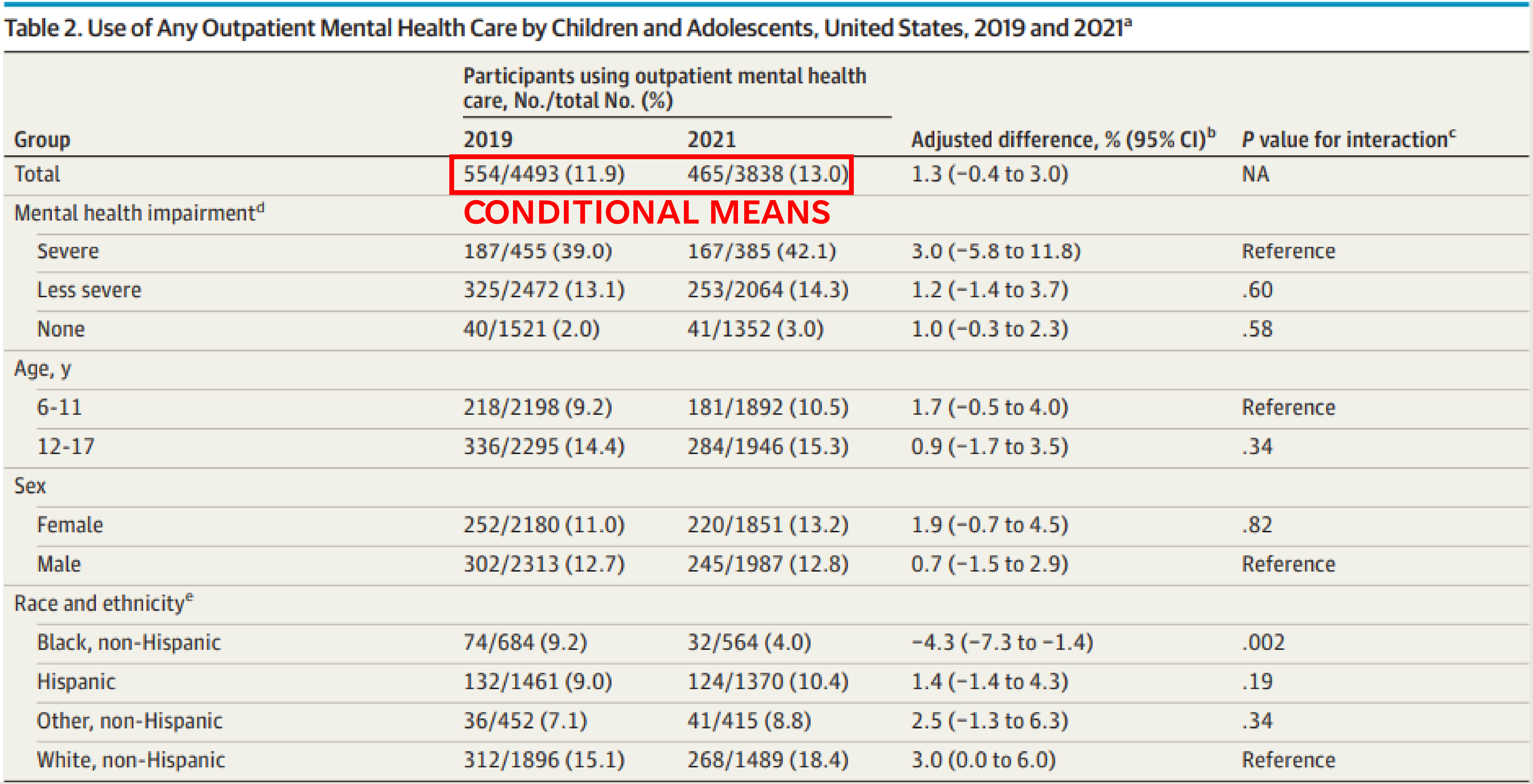


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Group	Participants using outpatient mental health care, No./total No. (%) 2019	2021	Adjusted difference, % (95% CI)[b]	P value for interaction[c]
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

CONDITIONAL MEANS

CONDITIONAL MEANS AND EXPECTATIONS

Conditional mean is the **expected (average) value** of the dependent variable (Y) given specific values of the independent variables (X).

"On average, what outcome do we expect to observe for a particular group defined by certain characteristics or conditions?"

$$\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}) = \beta^0 + \beta^1 x + \varepsilon$$

Conditional means are at the heart of popular models

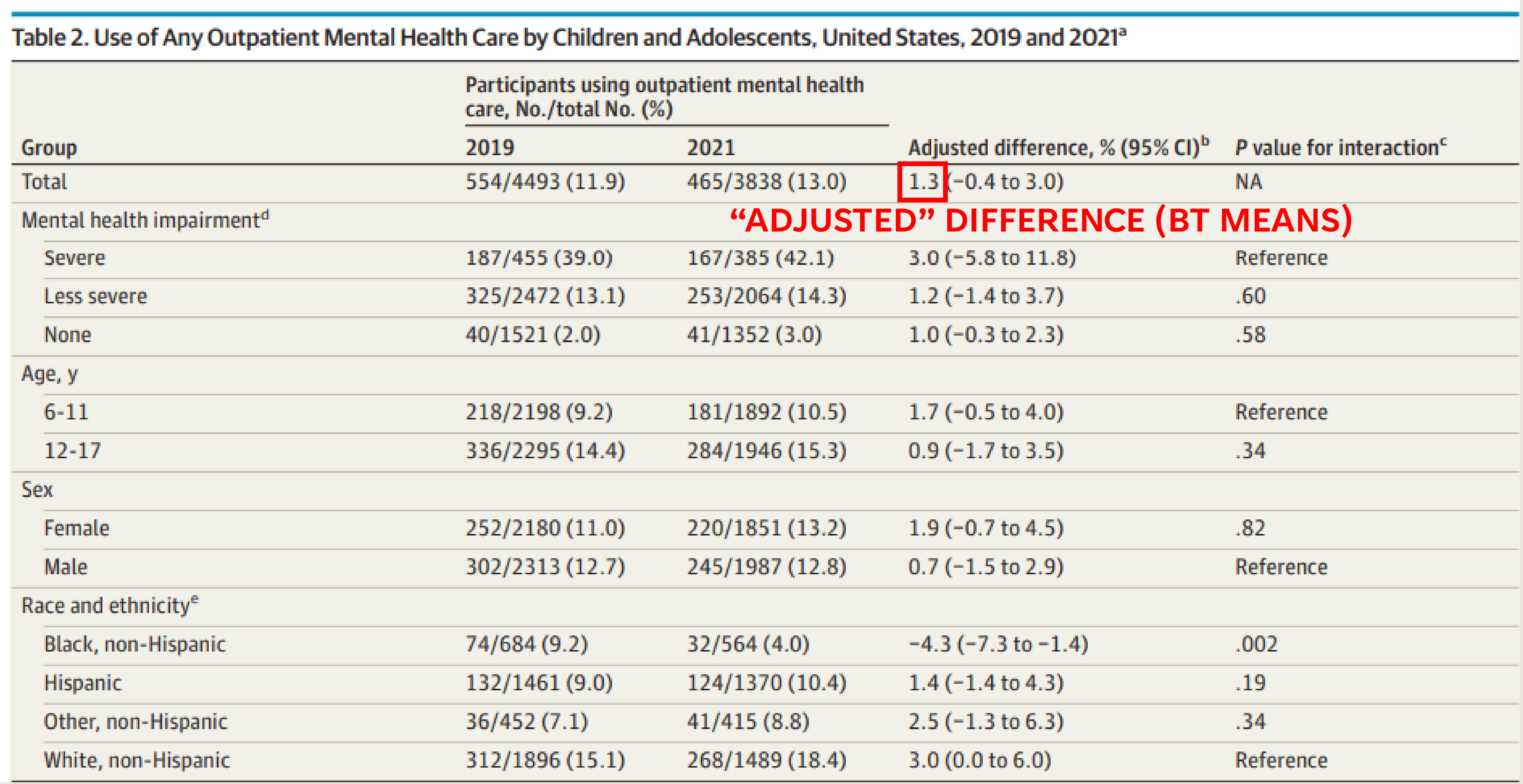


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Group	Participants using outpatient mental health care, No./total No. (%)		Adjusted difference, % (95% CI)[b]	P value for interaction[c]
	2019	2021		
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

“ADJUSTED” DIFFERENCE (BT MEANS)

COEFFICIENTS, ADJUSTED DIFFERENCES

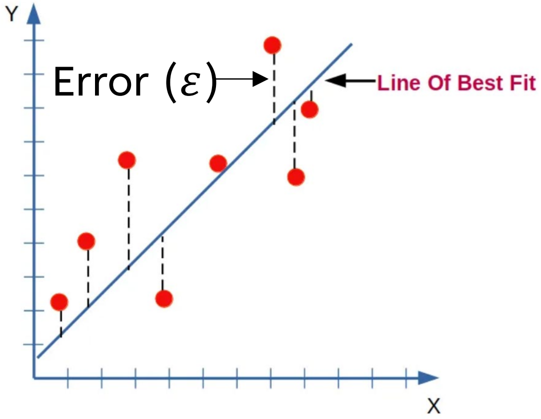
Adjusted differences (or $\boldsymbol{\beta^1}$) refer to the **expected change** in the dependent variable **(Y)** associated with a **one-unit change** in an independent variable **(X)**, after accounting for (or adjusting for) ALL other variables **(Z)** included in the model.

$$E(Y|X = x) = \beta^0 + \boldsymbol{\beta^1} x + \beta^2 \boldsymbol{z} + \varepsilon$$

How we estimate this difference depends on our outcome & model choice: linear regression, logistic regression, Poisson regression…etc…

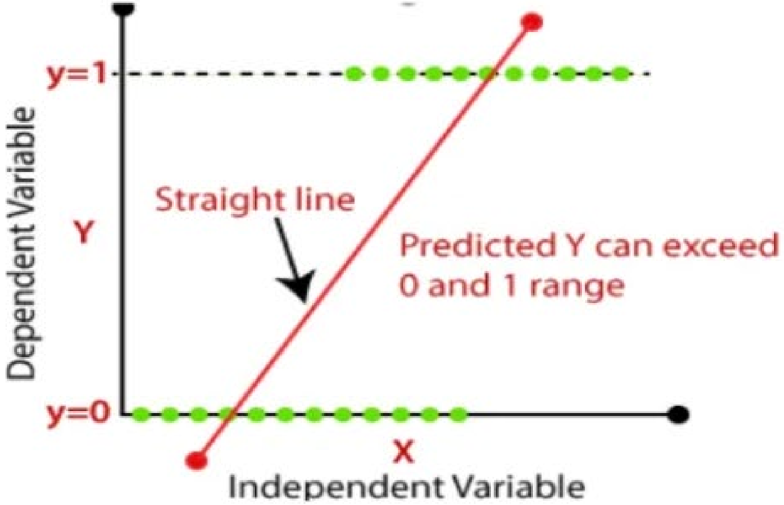
SMALL DETOUR

Linear regression models estimate a **best-fitted line,** which is the **estimated difference** in our study.

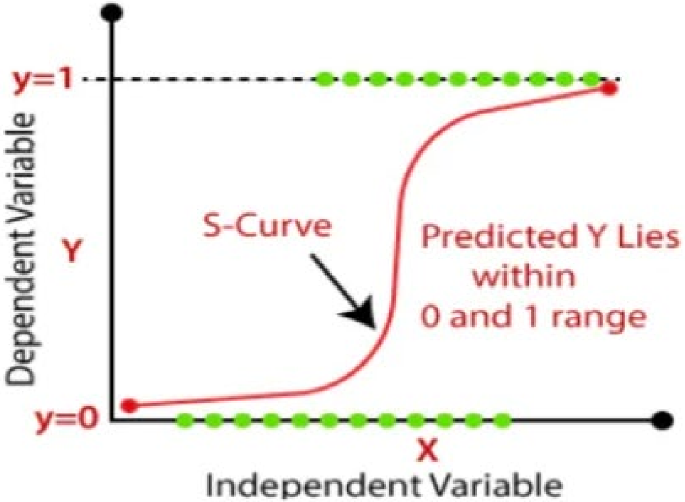


Estimated by **minimizing sum of squared errors.**

LINEAR REGRESSION



LOGISTIC REGRESSION



BACK ON TRACK…

HOW DO COVARIATES EVEN WORK?

Imagine we're studying the effect of **mental health treatment (X; 1=Yes, 0=No)** on a **mental health score (Y).** We suspect that **high SES (Z; 1=Yes, 0=No)** might also influence Y and could differ across groups defined by X (More High SES people get treatment).

To accurately estimate the effect of X on Y, we want to **account for differences in Z across the groups defined by X.**

$$E(Y|X = x) = \beta^0 + \beta^1 x + \boldsymbol{\beta^2 z} + \varepsilon$$

HOW DO COVARIATES EVEN WORK?

x (Treatment)	z (High SES)	% High SES within x	Expected value of Y if X = x & Z = z *E(Y| X,Z)*
0 (No)	0 (No)	50%	40
0 (No)	1 (Yes)	50	55
1 (Yes)	0 (No)	20%	45
1 (Yes)	1 (Yes)	80	60

1. In the total sample: 65% is **High SES**; X is evenly split

2. Calculate the expected Y values of X=0 & X=1, after standardizing the distribution of Z to its mean in the total sample:

- **Adjusted mean for X=0**: (40×0.35) + (55×0.65) = 14+35.75 = 49.75
- **Adjusted mean for X=1**: (45×0.35) + (60×0.65) = 15.75+39 = 54.75

3. Calculate the adjusted difference (X=1 vs. X=0): 54.75 − 49.75 = 5

After controlling for differences in **Z**, the adjusted difference in the expected outcome (**Y**) between treated **(X=1)** and not treated (**X=0**) is **5 points**.

HOW DO COVARIATES EVEN WORK?

x (Treatment)	z (High SES)	% High SES within x	Expected value of Y if X = x & Z = z *E(Y| X,Z)*
0 (No)	0 (No)	50%	40
0 (No)	1 (Yes)	50	55
1 (Yes)	0 (No)	20%	45
1 (Yes)	1 (Yes)	80	60

1. In the total sample: 65% is **High SES**; X is evenly split

2. Calculate the expected Y values of X=0 & X=1, after standardizing the distribution of Z to its mean in the total sample:

- **Adjusted mean for X=0**: (40×0.35) + (55×0.65) = 14+35.75 = 49.75
- **Adjusted mean for X=1**: (45×0.35) + (60×0.65) = 15.75+39 = 54.75

3. Calculate the adjusted difference (X=1 vs. X=0): 54.75 − 49.75 = 5

After controlling for differences in **Z**, the adjusted difference in the expected outcome (**Y**) between treated **(X=1)** and not treated (**X=0**) is **5 points**.

HOW DO COVARIATES EVEN WORK?

x (Treatment)	z (High SES)	% High SES within x	Expected value of Y if X = x & Z = z *E(Y| X,Z)*
0 (No)	0 (No)	50%	40
0 (No)	1 (Yes)	50	55
1 (Yes)	0 (No)	20%	45
1 (Yes)	1 (Yes)	80	60

1. In the total sample: 65% is **High SES**; X is evenly split

2. Calculate the expected Y values of X=0 & X=1, after standardizing the distribution of Z to its mean in the total sample:

- **Adjusted mean for X=0**: (40×0.35) + (55×0.65) = 14+35.75 = 49.75
- **Adjusted mean for X=1**: (45×0.35) + (60×0.65) = 15.75+39 = 54.75

3. Calculate the adjusted difference (X=1 vs. X=0): 54.75 − 49.75 = 5

After controlling for differences in **Z**, the adjusted difference in the expected outcome (**Y**) between treated **(X=1)** and not treated (**X=0**) is **5 points**.

HOW DO COVARIATES EVEN WORK?

x (Treatment)	z (High SES)	% High SES within x	Expected value of Y if X = x & Z = z *E(Y| X,Z)*
0 (No)	0 (No)	50%	40
0 (No)	1 (Yes)	50	55
1 (Yes)	0 (No)	20%	45
1 (Yes)	1 (Yes)	80	60

If we didn't adjust...

- **Unadjusted mean for X=0**: (40×0.5) + (55×0.5) = 20+27.5 = 47.5
- **Unadjusted mean for X=1**: (45×0.2) + (60×0.8) = 9+48 = 57

Calculate the uadjusted difference (X=1 vs. X=0): 57−47.5 = 9.5

Nearly **2x the effect** size (5 versus 9.5)....

WHAT COVARIATES SHOULD WE USE?

Somewhat debated but important principles to follow..

1. **Covariates should occur prior to exposure variable** (ie, pre-treatment)
2. **Covariates should be informed by prior work and theory**
3. **Exclude true instrumental variables** (ie, TRULY exogenous shocks to exposure that only indirectly affect Y; *advanced thinking required*)

Debate: Should we include all pretreatment covariates or only those that we think affect X and Y? (MORE ON THIS IN THE FUTURE)

WHAT COVARIATES SHOULD WE USE?

Somewhat debated but important principles to follow..

1. **Covariates should occur prior to exposure variable** (ie, pre-treatment)
2. **Covariates should be informed by prior work and theory**
3. **Exclude true instrumental variables** (ie, TRULY exogenous shocks to exposure that only indirectly affect Y; *advanced thinking required*)

Debate: Should we include all pretreatment covariates or only those that we think affect X and Y? (MORE ON THIS IN THE FUTURE)

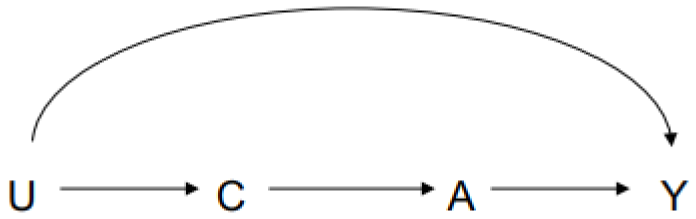


Fig. 3 Controlling for measured covariate C, even in the presence of unmeasured variable U, eliminates confounding of the relationship between exposure A and outcome Y, even though C itself is not a common cause of A and Y

PULSE CHECK

Join mentimeter.com

Type in code: **5363 3331**

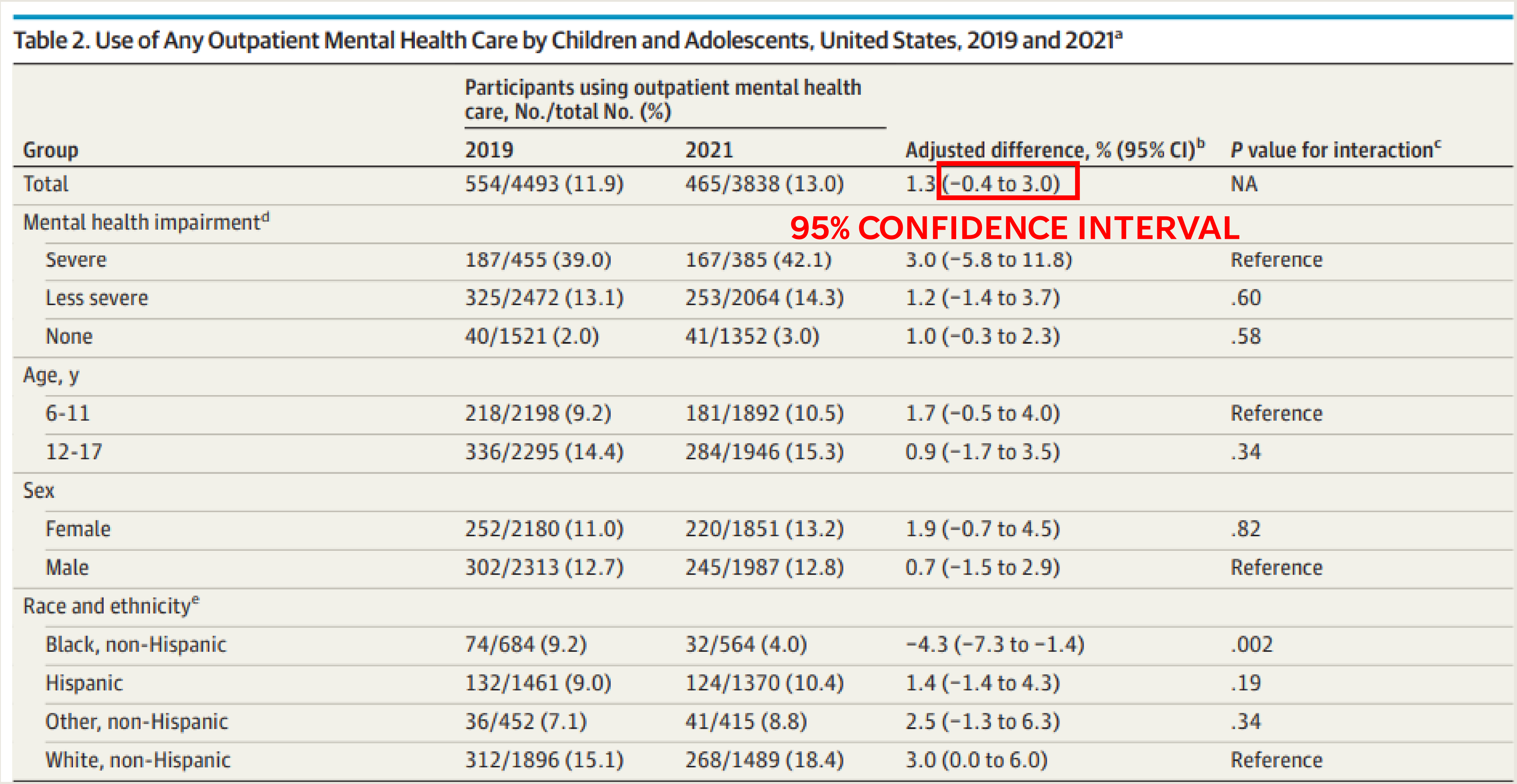


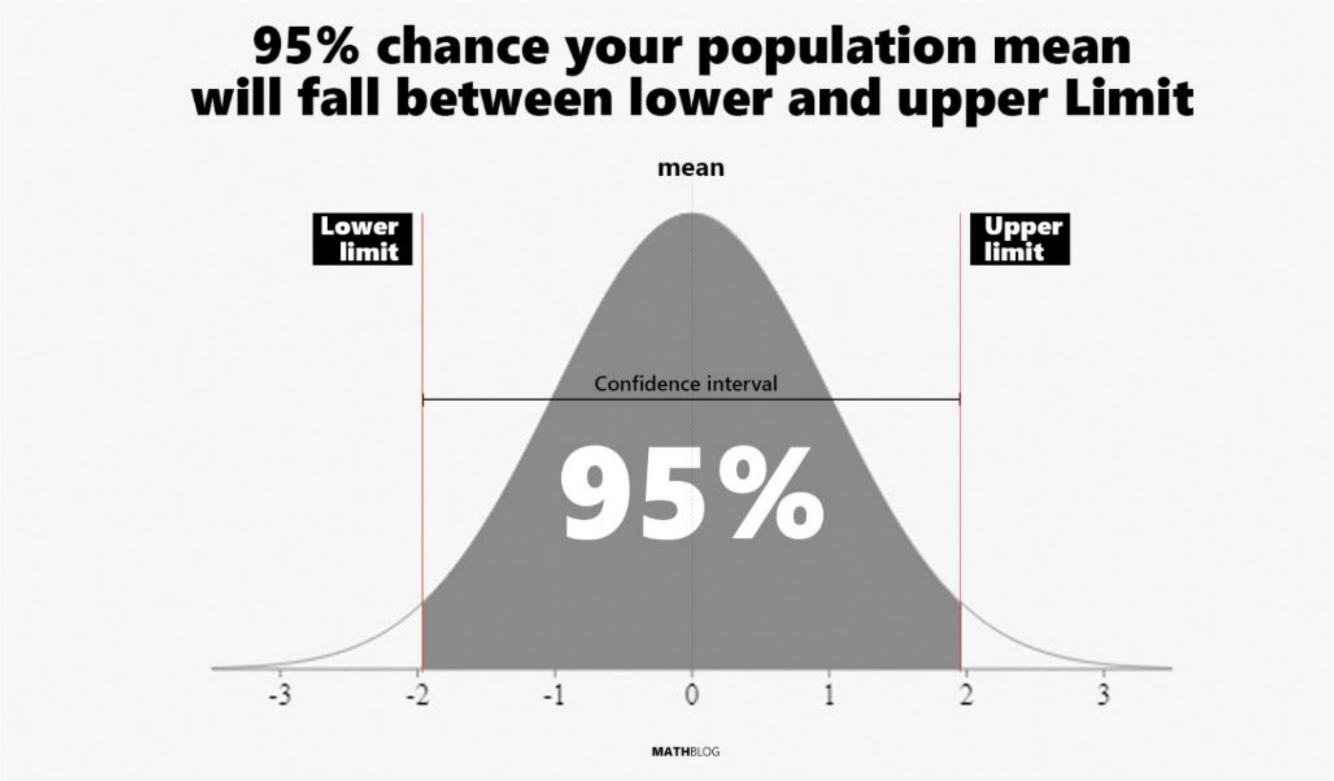
Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Group	Participants using outpatient mental health care, No./total No. (%)			
	2019	2021	Adjusted difference, % (95% CI)[b]	P value for interaction[c]
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

95% CONFIDENCE INTERVAL

WHAT ARE 95% CONFIDENCE INTERVALS?

A **95% confidence interval** is a numerical range which, upon repeated sampling, will contain the population value 95% of the time


95% chance your population mean
will fall between lower and upper Limit
mean
Lower limit
Upper limit
Confidence interval
95%
-3
-2
-1
0
1
2
3
MATHBLOG

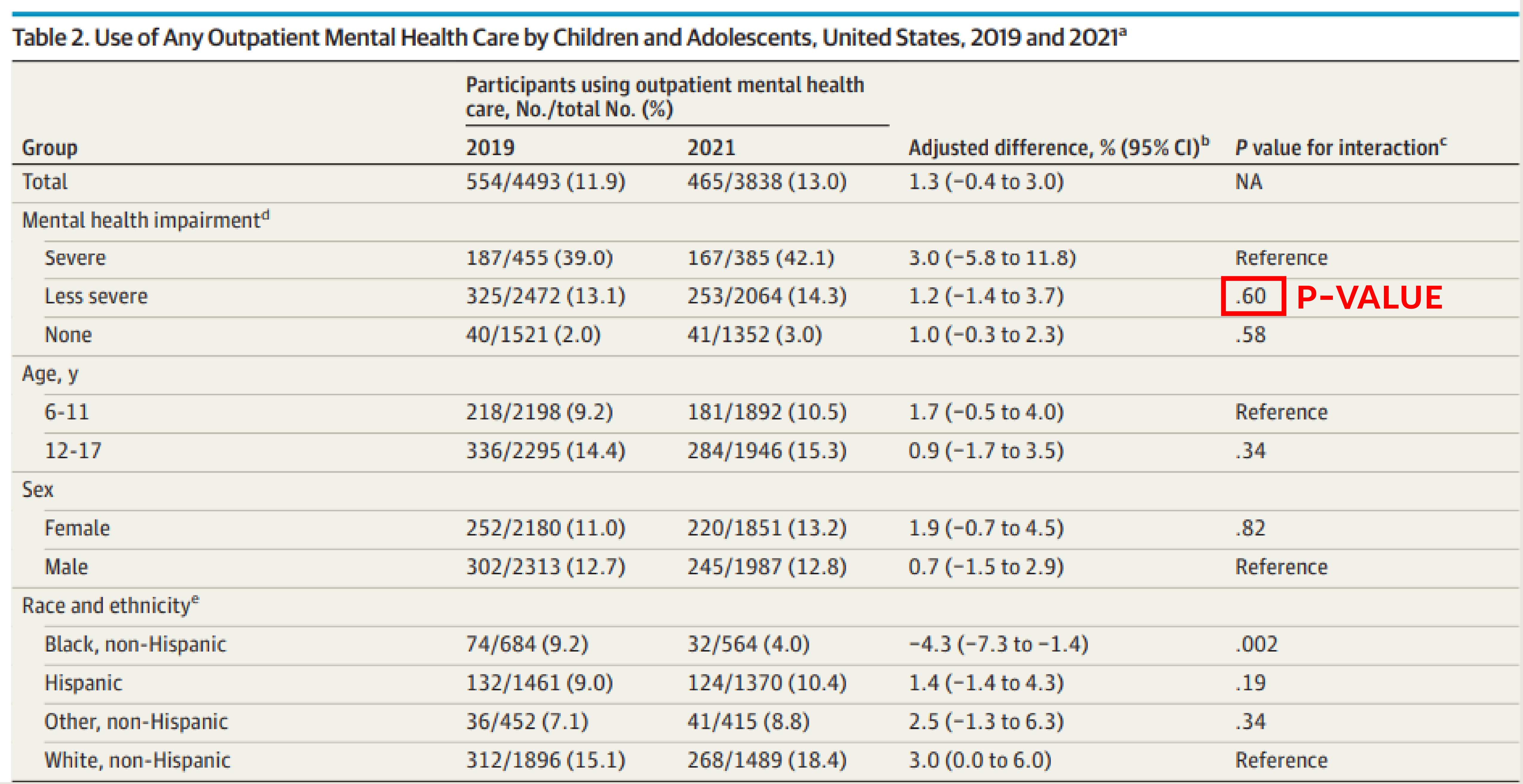


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

Group	Participants using outpatient mental health care, No./total No. (%)			
	2019	2021	Adjusted difference, % (95% CI)[b]	P value for interaction[c]
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

P-VALUE

THE DREADED P < .05

This **probability value (p value)** indicates the probability of observing differences as large (or larger in your sample), given that the **null hypothesis** (ie, no differences) was true.

Scholars vary in their value of p-values (particulary $p < 0.05$)

Pros: Unbiased, reasonably high & low bar, easy, practical

Cons: Arbitrary, masks clinical significance, p-hacking

Emerging scholars, $p < 0.05$ and 95% CIs are a good start.

You will eventually come across results that are
$p > 0.05$ that deserve serious consideration…I promise ☺

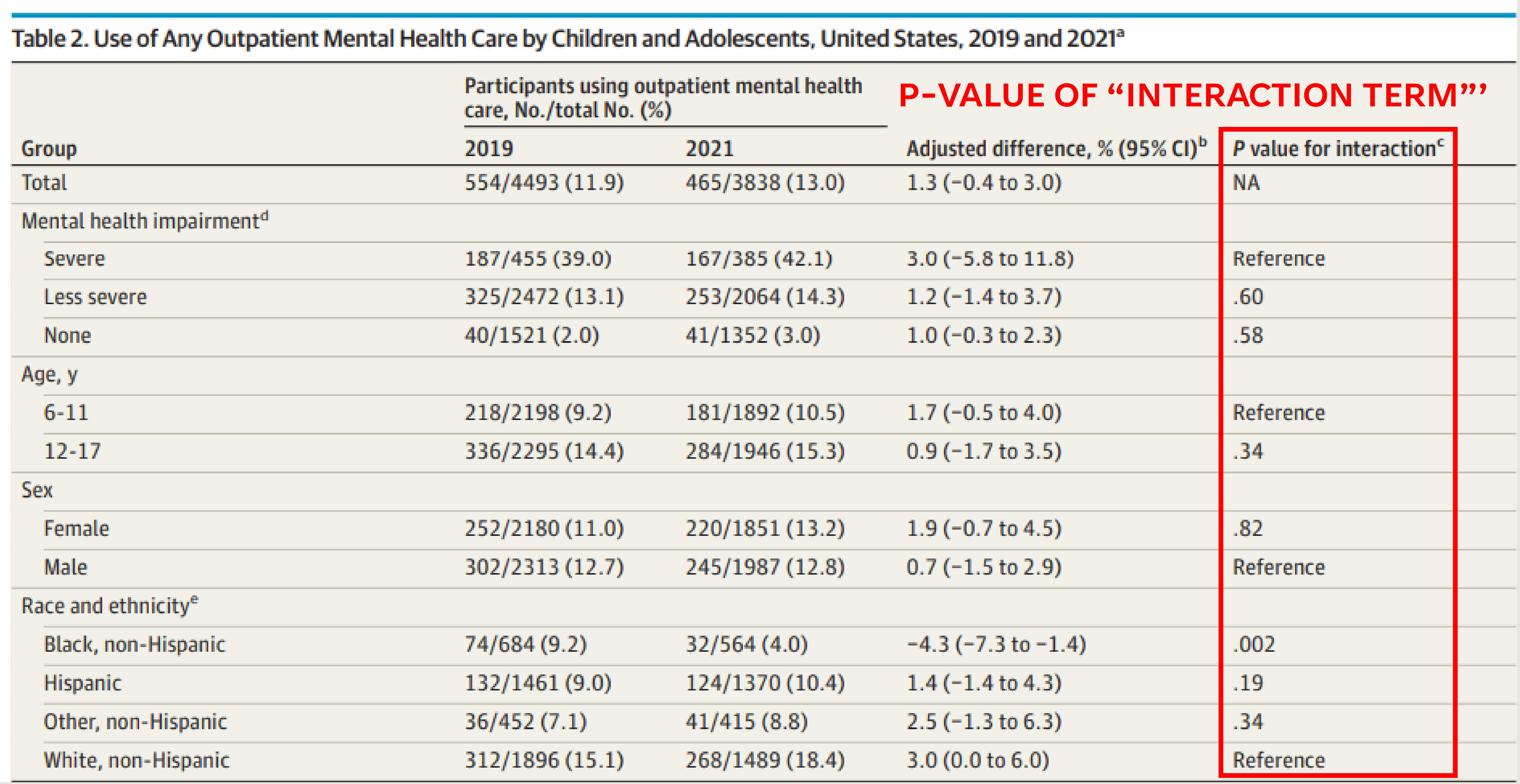


Table 2. Use of Any Outpatient Mental Health Care by Children and Adolescents, United States, 2019 and 2021[a]

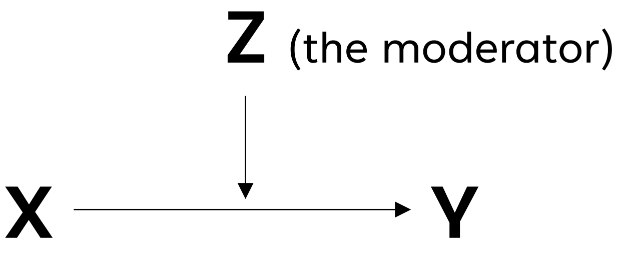
P–VALUE OF “INTERACTION TERM”’

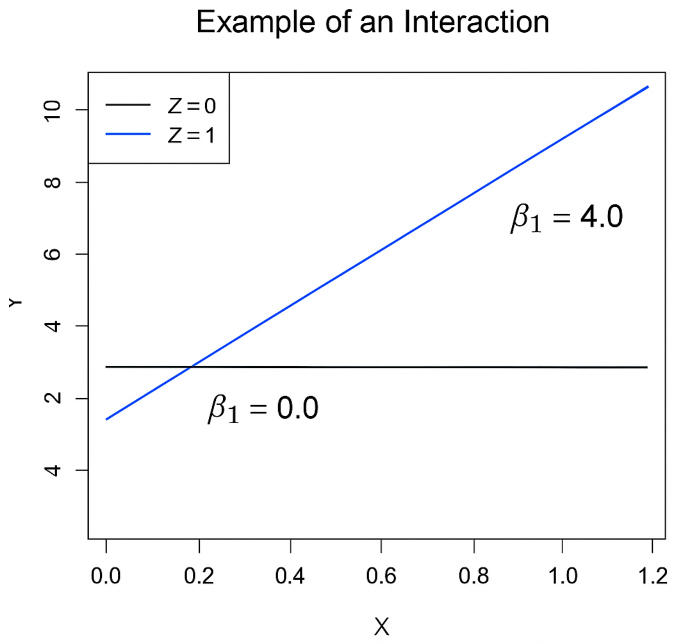
	Participants using outpatient mental health care, No./total No. (%)			
Group	2019	2021	Adjusted difference, % (95% CI)[b]	P value for interaction[c]
Total	554/4493 (11.9)	465/3838 (13.0)	1.3 (−0.4 to 3.0)	NA
Mental health impairment[d]				
Severe	187/455 (39.0)	167/385 (42.1)	3.0 (−5.8 to 11.8)	Reference
Less severe	325/2472 (13.1)	253/2064 (14.3)	1.2 (−1.4 to 3.7)	.60
None	40/1521 (2.0)	41/1352 (3.0)	1.0 (−0.3 to 2.3)	.58
Age, y				
6-11	218/2198 (9.2)	181/1892 (10.5)	1.7 (−0.5 to 4.0)	Reference
12-17	336/2295 (14.4)	284/1946 (15.3)	0.9 (−1.7 to 3.5)	.34
Sex				
Female	252/2180 (11.0)	220/1851 (13.2)	1.9 (−0.7 to 4.5)	.82
Male	302/2313 (12.7)	245/1987 (12.8)	0.7 (−1.5 to 2.9)	Reference
Race and ethnicity[e]				
Black, non-Hispanic	74/684 (9.2)	32/564 (4.0)	−4.3 (−7.3 to −1.4)	.002
Hispanic	132/1461 (9.0)	124/1370 (10.4)	1.4 (−1.4 to 4.3)	.19
Other, non-Hispanic	36/452 (7.1)	41/415 (8.8)	2.5 (−1.3 to 6.3)	.34
White, non-Hispanic	312/1896 (15.1)	268/1489 (18.4)	3.0 (0.0 to 6.0)	Reference

STATISTICAL INTERACTIONS

A **statistical interaction ($\boldsymbol{\beta}_3$)** occurs when the effect of one variable on an outcome depends on the level of another variable.

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \boldsymbol{\beta}_3 XZ + \varepsilon$$

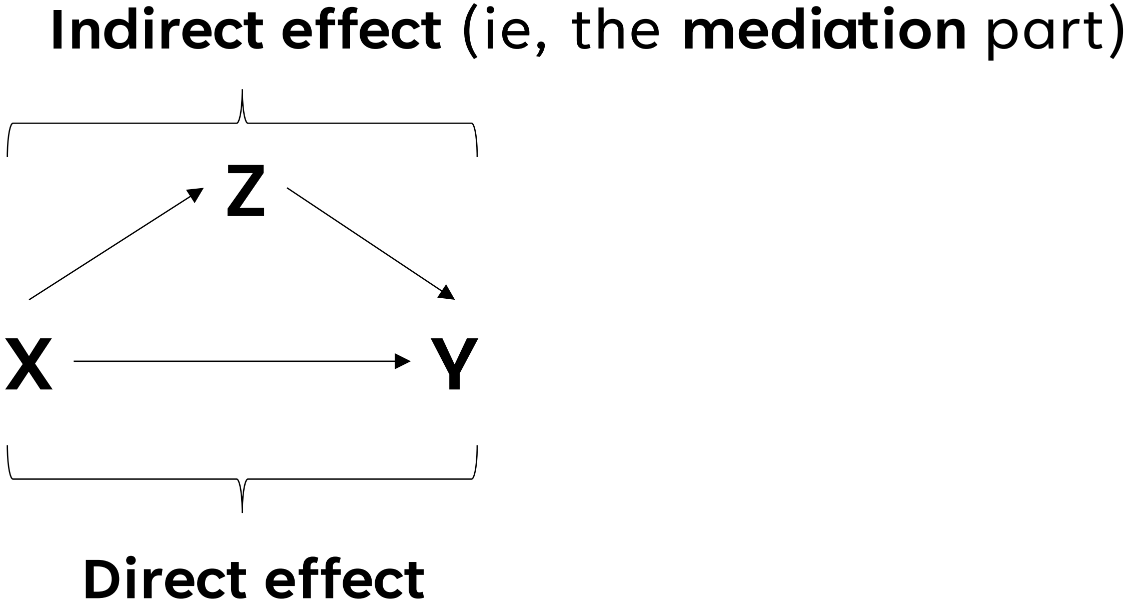




IMPORTANT DETOUR:
STATISTICAL MEDIATION

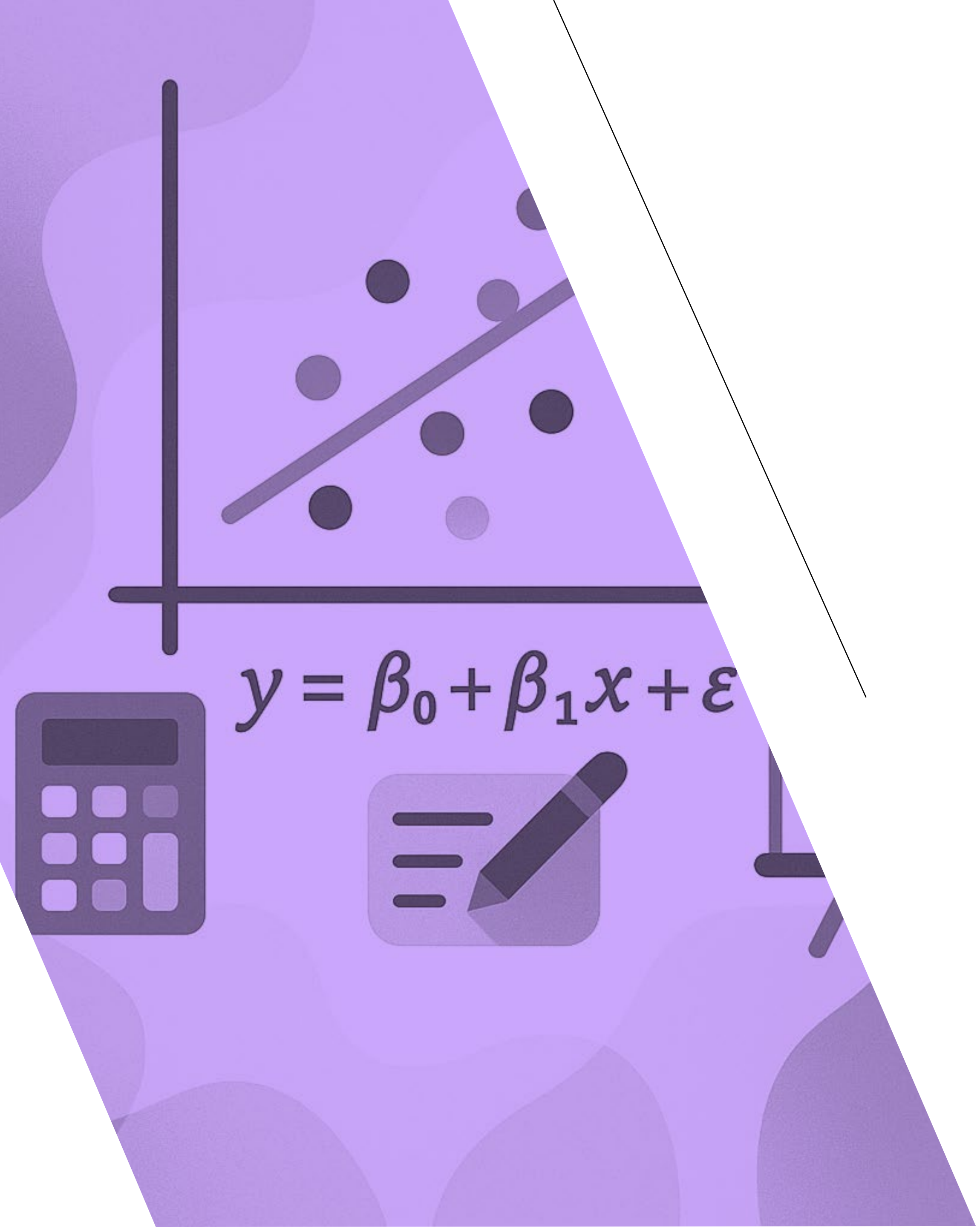
A **statistical mediation** occurs when a variable (ie, **mediator**) is on the causal pathway between an independent and dependent variable

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$$



Indirect effect (ie, the **mediation** part)

Direct effect



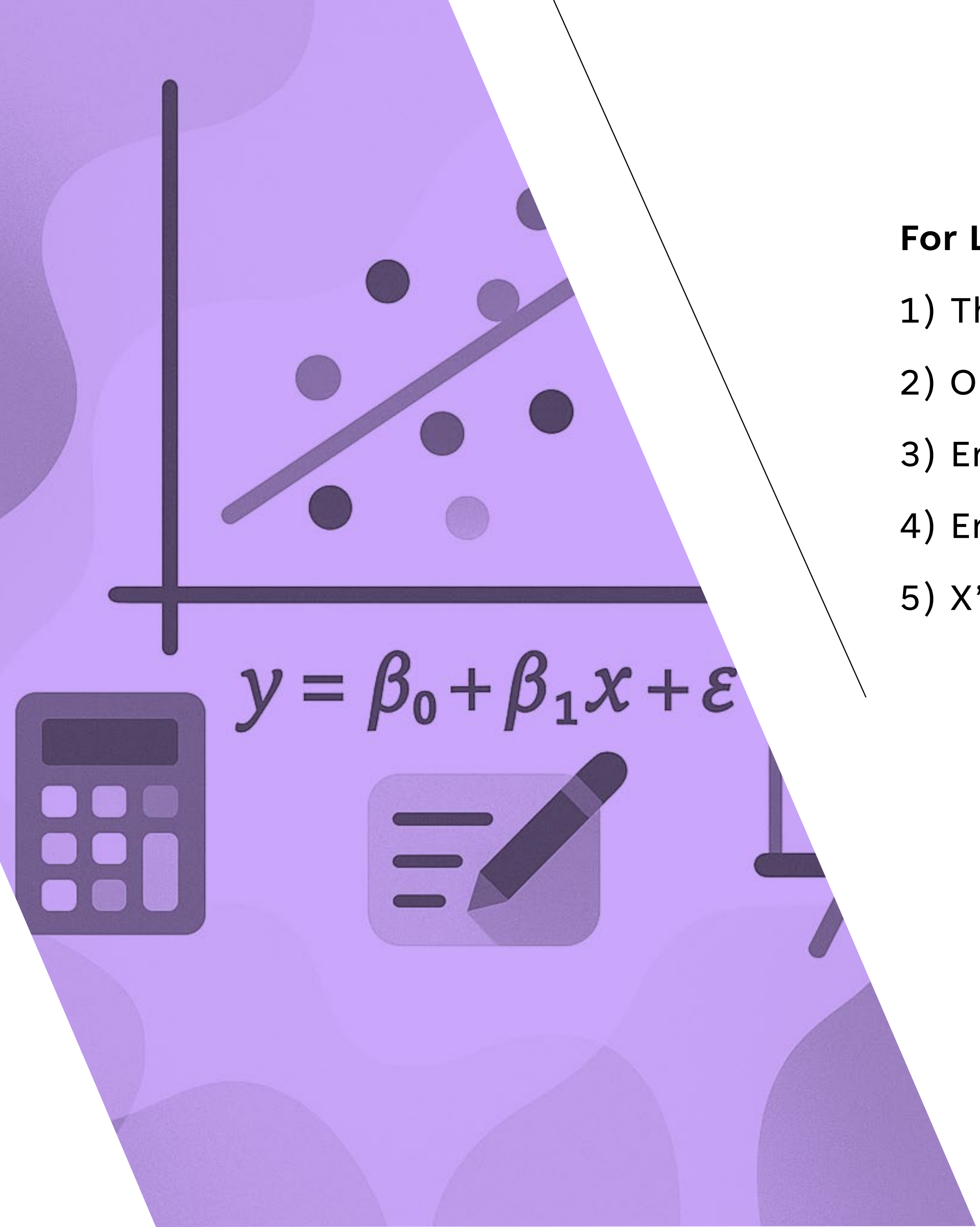
REGRESSION MODEL ASSUMPTIONS

Statistical modeling assumptions are things we assume to be true about the data and model to ensure **consistency**, **unbiasedness**, and **efficiency**

A model is **consistent** if estimates approach the population value as the sample size increases.

A model is **unbiased** if the difference between the estimate and population value is zero

A model is more **efficient** if the standard error (and 95% CI) is lower/tighter



REGRESSION MODEL ASSUMPTIONS

For Linear Regression (Ordinary Least Squares)…

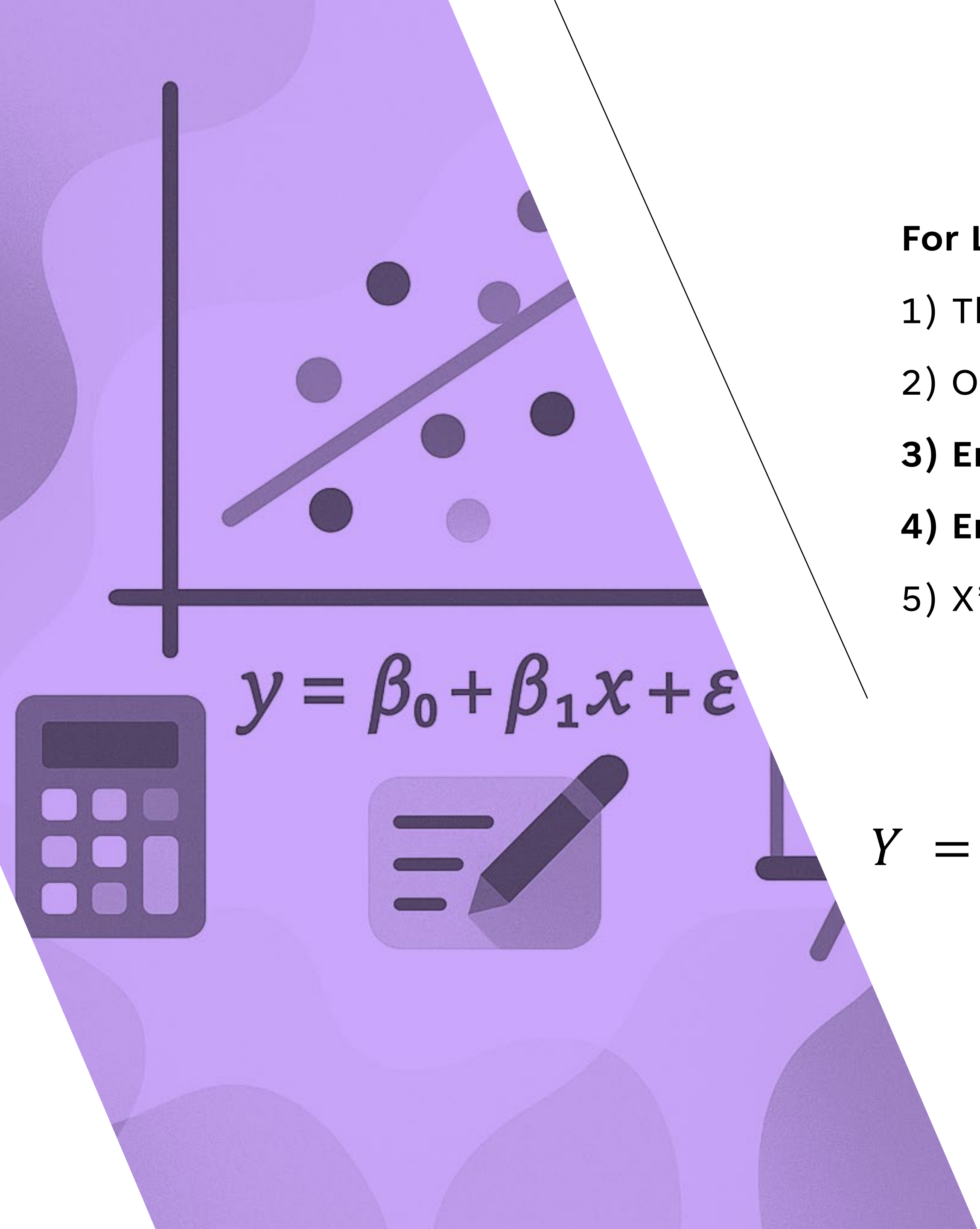
1) The relationships between X’s and Y are linear. (Linearity)

2) Observations are independent—errors are uncorrelated. (IID)

3) Error variance is constant across levels of X (Heteroskedastic)

4) Errors are normal distributed (Normality of errors)

5) X’s are not perfectly correlated (No multicollinearity)



REGRESSION MODEL ASSUMPTIONS

For Linear Regression (Ordinary Least Squares)…

1) The relationships between X’s and Y are linear. (Linearity)

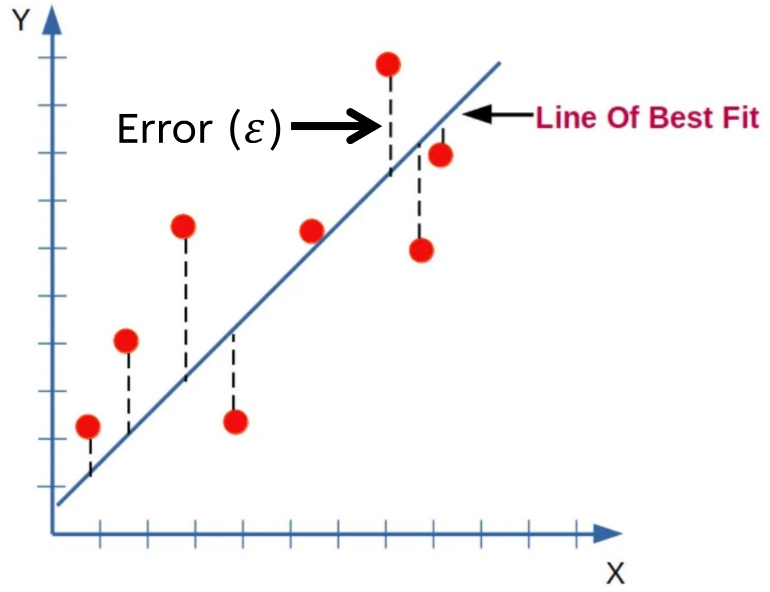
2) Observations are independent—**errors are uncorrelated**. (IID)

3) Error variance is constant across levels of X (Heteroskedastic)

4) Errors are normal distributed (Normality of errors)

5) X’s are not perfectly correlated (No multicollinearity)

$$Y = \beta_0 + \beta_1 X + \boldsymbol{\varepsilon}$$



KNOWLEDGE CHECK & REVIEW

Join mentimeter.com

Type in code: **5363 3331**

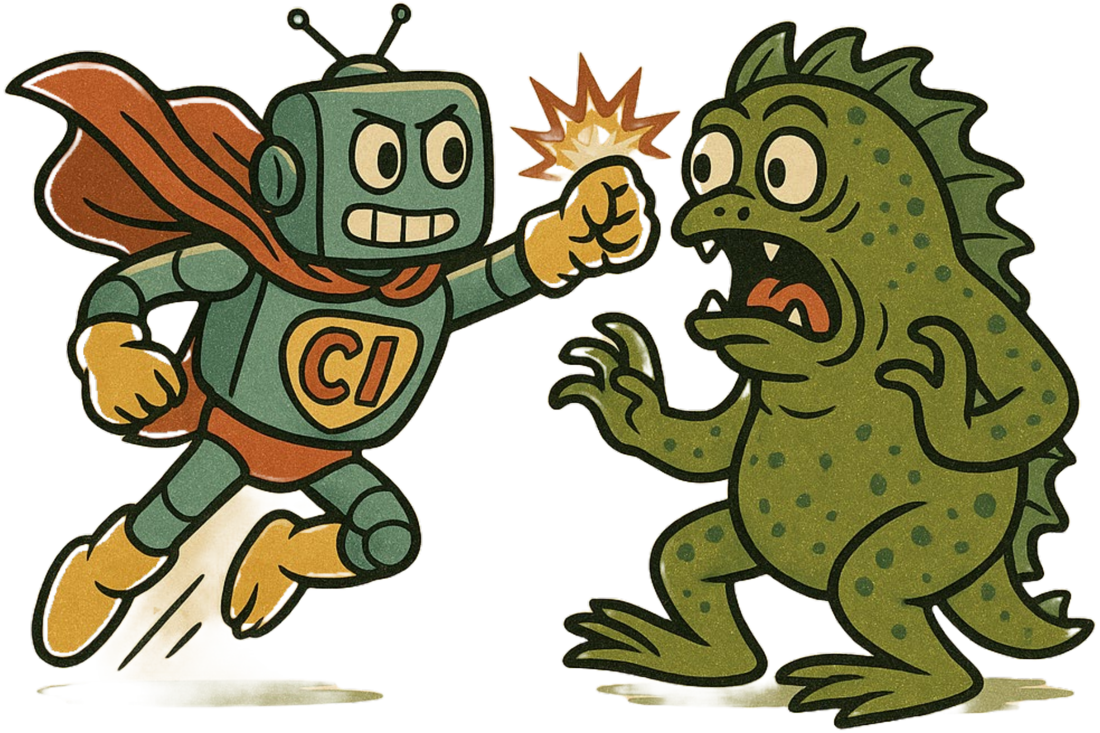
TIME TO RELAX…
YOU LEARNED A THING!

A LOOK INTO THE FUTURE…
HOW DO WE GO FROM REGRESSION TO CAUSAL ANALYSIS?

Causal assumptions are very simple in theory, but VERY complex in practice

1. X must be correlated with Y.
2. The X and Y correlation must not be due to chance alone (see 95% CI and p-value).
3. There must be no other variables that cause X and Y (confounding).
4. Y must not cause X (reverse causation).

We will delve into this in future workshops and you will see…



Causal Inference man battles Confounding Creature

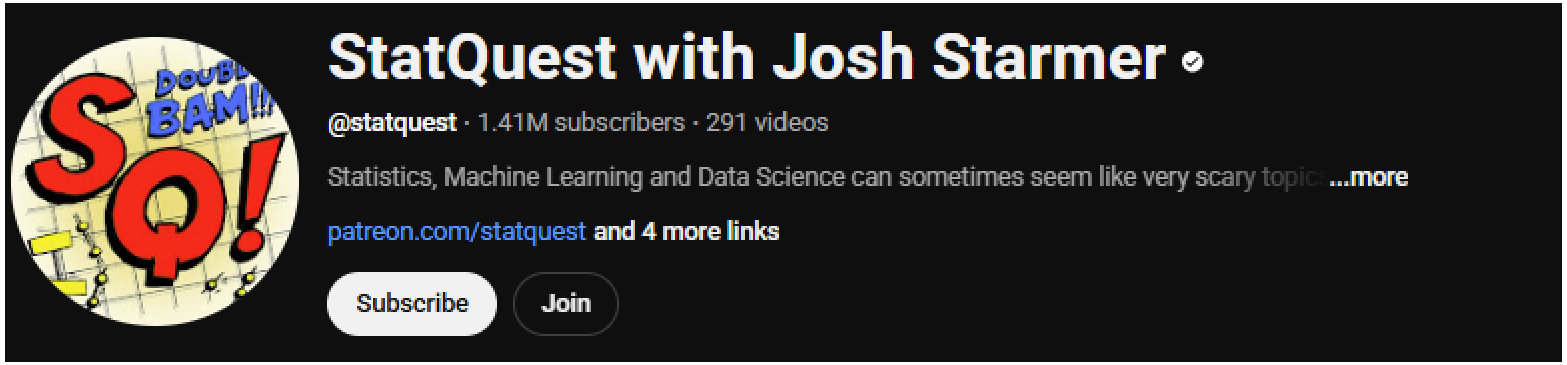
FUTURE SMARTSTATS WORKSHOPS

All will (more or less) rely on fundamentals covered today

1. Statistical software primers (R Studio, **June 25, Wednesday, 11a-12:15p)**
2. Causal inference (DiD, Synthetic control, Econometrics)
3. Multi-level modelling
4. Missing data analysis
5. Structural equation modelling
6. Longitudinal data analysis (fixed effects, random effects, and beyond)
7. Machine learning

RESOURCES FOR FURTHER LEARNING

Beginner friendly



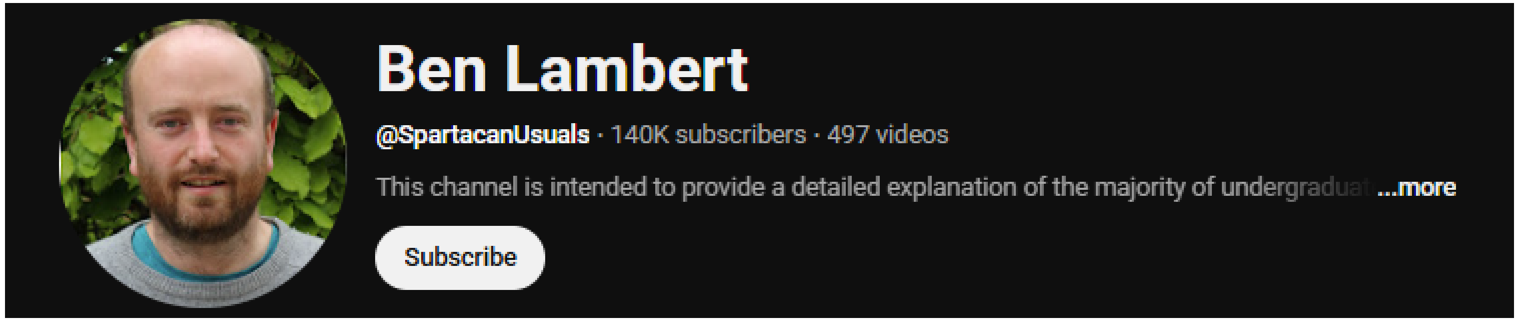
StatQuest with Josh Starmer

@statquest · 1.41M subscribers · 291 videos

Statistics, Machine Learning and Data Science can sometimes seem like very scary topic ...more

patreon.com/statquest and 4 more links

Intermediate, more technical (econometrics)



Ben Lambert

@SpartacanUsuals · 140K subscribers · 497 videos

This channel is intended to provide a detailed explanation of the majority of undergradua ...more

DISCUSSION AND QUESTIONS