Supplementary Appendix A: Statistical analysis

1. Formula and examples of odd ratio calculation for ►Table 1

Odd ratio example 1:

	Video (A) $N=57{,}915$	Telephone (B) $N=46{,}224$	In-person (C) $N=453{,}848$	Odds ratio (A vs. B)	Odds ratio (A vs. C)	Odds ratio (B vs. C)
White	45,760	35,230	334,925	1.175	1.337	1.138
Non-White	12,155	10,994	118,923	0.851	0.748	0.879

We define that *A_White* is the number of White-race patient in Video(A) group, and A_non-White is the number of non-White-race patient in Video(A) group. Similarly, B_White is the number of White-race patient in Telephone(B) group, while B_non-White is the number of non-White-race patient in Telephone(B) group. C_White is the number of White-race patient in in-person(C) group, while C_non-White is the number of non-White-race patient in in-person(C) group.

The formula and results for odds ratio results A versus B for White are:

$$OR(A,B),WHITE = \frac{A_{White} \times B_{non-White}}{A_{non-White} \times B_{White}} = \frac{45{,}760 \times 10{,}994}{12{,}155 \times 35{,}230} = 1.175$$

Similarly, formula and results for odds ratio results A versus B for non-White is:

$$OR(A,B),NON-WHITE = \frac{A_{non-White} \times B_{White}}{A_{White} \times B_{non-White}} = \frac{12{,}155 \times 35{,}230}{45{,}760 \times 10994} = 0.851$$

For 95% confidence intervals of OR(A,B) White,

Upper 95% CI = e^[ln (OR(A,B)) + 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.14

Lower 95% CI = e^[ln (OR(A,B)) − 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.21

where $a = A_{White}$, $b = A_{non-White}$, $c = B_{non-White}$, $d = B_{White}$. 1.96 is the approximate *z*-value of the 95-percentile point of the standard normal distribution.

Statistical interpretation: The group of patients who use video-type telemedicine service have higher percentage of White race patient than those group of telephone service.

The odds of an individual who use video services being White race are 1.175 (95% CI: 1.14–1.21) times higher than the odds of an individual who using telephone call being a White race. Or, the odds of an individual who use telephone service being non-White race are 0.851 times lower than the odds of an individual who use video services being non-White race.

The formula and results for odds ratio results A versus C for White is:

$$OR(A,C),White = \frac{A_{White} \times C_{non-White}}{A_{non-White} \times C_{White}} = \frac{45{,}760 \times 118{,}923}{12{,}155 \times 334{,}925} = 1.337$$

For 95% confidence intervals,

Upper 95% CI = e^[ln (OR(A,C)) + 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.31

Lower 95% CI = e^[ln (OR(A,C)) − 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.37

where $a = A_{White}$, $b = A_{non-White}$, $c = C_{non-White}$, $d = C_{White}$. 1.96 is the approximate *z*-value of the 95-percentile point of the standard normal distribution.

The statistical interpretation: The group of patients who use video-type telemedicine service have higher percentage of White race patient than those group of in-person visits. The odds of an individual who use video services being White race are 1.337 (95% CI: 1.31–1.37) times higher than the odds of an individual who visit hospital in person being White race.

The formula and results for odds ratio results B versus C for White is:

$$OR(B,C),WHITE = \frac{B_{White} \times C_{non-White}}{B_{non-White} \times C_{White}} = \frac{35{,}230 \times 118{,}923}{10{,}994 \times 334{,}925} = 1.138$$

Similarly, formula and results for odds ratio results A versus B for non-White is:

$$OR(A,B),NONWHITE = \frac{B_{non-White} \times C_{White}}{B_{White} \times C_{non-White}} = \frac{12{,}155 \times 35{,}230}{45{,}760 \times 10{,}994} = 0.879$$

For 95% confidence intervals of OR(B,C) White,

Upper 95% CI = e^[ln (OR(B,C)) + 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.11

Lower 95% CI = e^[ln (OR(B,C)) − 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.16

where $a = A_{White}$, $b = A_{non-White}$, $c = B_{non-White}$, $d = B_{White}$. 1.96 is the approximate *z*-value of the 95-percentile point of the standard normal distribution.

Statistical interpretation: The group of patients who use phone-type telemedicine service have higher percentage of White race patient than those group of in-person visits. The odds of an individual who use phone services being White race are 1.138 (95% CI: 1.11–1.16) times higher than the odds of an individual who visit hospital in person being White race.

 © 2021. Thieme. All rights reserved.

Odd ratio analysis example 2:

	Video $N=57,915$	Telephone $N=46,224$	In-person = 453,848	Odds ratio (A vs. B)	Odds ratio (A vs. C)	Odds ratio (B vs. C)
Black	7,465	8,513	66,568	0.655	0.896	1.31
Non-Black	50,450	37,711	387,280	1.526	1.16	0.76

We define that *A_Black* is the number of Black-race patient in Video(A) group, and A_non-Black is the number of non-Black-race patient in Video(A) group. Similarly, B_Black is the number of Black-race patient in Telephone (B) group, while B_non-Black is the number of non-Black-race patient in Telephone(B) group. C_Black is the number of Black-race patient in in-person(C) group, while C_non-Black is the number of non-Black-race patient in in-person (C) group.

The formula and results for odds ratio results A versus B for Black is:

$$OR(A,B), Black = \frac{A_{Black} \times B_{non-Black}}{A_{non-Black} \times B_{Black}} = \frac{7{,}465 \times 37{,}711}{50{,}450 \times 8{,}513} = 0.655$$

Similarly, formula and results for odds ratio results A versus B for non-Black is:

$$OR(A,B), non-Black = \frac{A_{non-Black} \times B_{Black}}{A_{Black} \times B_{non-Black}} = \frac{50{,}450 \times 8{,}513}{7{,}465 \times 37{,}711} = 1.526$$

For 95% confidence intervals of OR(A,B) Black,

Upper 95% CI = e^[ln (OR(A,B)) + 1.96√ (1/a + 1/b + 1/c + 1/d)] = 0.63
Lower 95% CI = e^[ln (OR(A,B)) – 1.96√ (1/a + 1/b + 1/c + 1/d)] = 0.68

where $a = A_{Black}$, $b = A_{non-Black}$, $c = B_{non-Black}$, $d = B_{Black}$. 1.96 is the approximate *z*-value of the 95-percentile point of the standard normal distribution.

Statistical interpretation: The group of patients who use video-type telemedicine service have higher percentage of Black race patient than those group of telephone service.

The odds of an individual who use video services being Black race are 0.655 (95% CI: 0.63–0.68) times of the odds of an individual who using telephone call being a Black race.

Or, the odds of an individual who use telephone service being non-Black race are 1.526 times higher than the odds of an individual who use video services being non-Black race.

The formula and results for odds ratio results A versus C for Black is:

$$OR(A,C), Black = \frac{A_{Black} \times C_{non-Black}}{A_{non-Black} \times C_{Black}} = \frac{7{,}465 \times 387{,}280}{50{,}450 \times 66{,}568} = 0.86$$

For 95% confidence intervals,

Upper 95% CI = e^[ln (OR(A,C)) + 1.96√ (1/a + 1/b + 1/c + 1/d)] = 0.84
Lower 95% CI = e^[ln (OR(A,C)) – 1.96√ (1/a + 1/b + 1/c + 1/d)] = 0.88

where $a = A_{Black}$, $b = A_{Black}$, $c = C_{non-Black}$, $d = C_{Black}$. 1.96 is the approximate *z*-value of the 95-percentile point of the standard normal distribution.

The statistical interpretation: The group of patients who use video-type telemedicine service have higher percentage of Black race patient than those group of in-person visits. The odds of an individual who use video services being Black race are 0.86 (95% CI: 0.84–0.88) times of the odds of an individual who visit hospital in person being Black race.

The formula and results for odds ratio results B versus C for Black is:

$$OR(B,C), Black = \frac{B_{Black} \times C_{non-Black}}{B_{non-Black} \times C_{non-Black}} = \frac{8{,}513 \times 387{,}280}{37{,}711 \times 66{,}568} = 1.31$$

Similarly, formula and results for odds ratio results A versus B for non-Black is:

$$OR(A,B), non-Black = \frac{B_{non-Black} \times C_{Black}}{B_{Black} \times C_{non-Black}} = \frac{37{,}711 \times 66{,}568}{8{,}513 \times 387{,}280} = 0.76$$

For 95% confidence intervals of OR(B,C) Black,

Upper 95% CI = e^[ln (OR(B,C)) + 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.28
Lower 95% CI = e^[ln (OR(B,C)) – 1.96√ (1/a + 1/b + 1/c + 1/d)] = 1.35

where $a = A_{Black}$, $b = A_{non-Black}$, $c = B_{non-Black}$, $d = B_{Black}$. 1.96 is the approximate *z*-value of the 95-percentile point of the standard normal distribution.

Statistical interpretation: The group of patients who use phone-type telemedicine service have higher percentage of Black race patient than those group of in-person visits. The odds of an individual who use phone services being Black race are 1.31 (95% CI: 1.28–1.35) times higher than the odds of an individual who visit hospital in person being Black race.

Formula and examples of linear regression for ►Table 2

We used *R* as a statistical tool for linear regression analysis.

 © 2021. Thieme. All rights reserved.

Linear regression example 1: regression of video TAR across income

Zip code	Video TAR of all population in each zip code area	Median Income of all population in each zip code area
Zip 1	0.0287	72,206
Zip 2	0.0089	77,989
Zip 3	0.0526	97,202
Zip 4	0.0027	51,250
...		
Zip X	0.0003	47,031

R linear regression command and results:

lm(formula = zips$TAR_video – (zips$median_income))

In R language, lm is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance, and analysis of covariance.

Reference: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/lm

Results:
Coefficients:
Estimate std. Error *t*-value Pr (>|*t*|)
(Intercept) –23.4520 8.1921 –2.863 0.00495**
log(zips$income) 1.6472 0.5234 3.147 0.00021***
Residual standard error: 1.791 on 121 degrees of freedom
Multiple *R*-squared: 0.009888, Adjusted *R*-squared: 0.001706
F-statistic: 1.208 on 1 and 121 DF, *p*-value: 0.2738

Therefore, we fill the ►**Table 2** in the manuscript with following coefficient and *p*-values.

Coefficients	Median income
Video (Group A)	1.65 (p <0.001)

Statistical interpretation of the linear regression results:

A single regression with TAR of video group as an independent variable and median income as a dependent variable shows that TAR of video service is positively correlated with median income of a certain zip code area. A logged coefficient value of 1.65 on median income indicates that a 1% of median income increase will be expected to see increase an average of TAR increased by 1.65% from current level. The association is significant.

Linear regression example 2 regression of video TAR across college education

Zip code	Video TAR of all population in each zip code area	College educated rate of each zip code area
Zip 1	0.0287	0.168
Zip 2	0.0089	0.309
Zip 3	0.0526	0.574
Zip 4	0.0027	0.091
...		
Zip X	0.0003	0.116

R commands and results

lm(formula = zips$TAR_Video – log(zips $college_educated_rate))
Estimate std. error *t*-value Pr(>|*t*|)
(Intercept) –2.9799 1.3488 –2.209 0.029035*
log(zips$college_educated_rate) 1.5774 0.3975 3.968 0.000123***
Residual standard error: 1.997 on 121 degrees of freedom
Multiple *R*-squared: 0.1151, Adjusted *R*-squared: 0.1078
F-statistic: 15.75 on 1 and 121 DF, *p*-value: 0.0001233

Coefficients	College education
Video (Group A)	1.58 (p <0.001)

Statistical interpretation: A single regression with TAR of video group as an independent variable and college educated rate as a dependent variable shows that TAR of video service is positively correlated with college educated rate of a certain zip code area. A logged coefficient value of 1.58 on college educated rate indicates that a 1% of college education increase will be expected to see increase an average of TAR increased by 1.58% from current level. The association is significant.

2. Formula and examples of multiple linear regression for ►**Table 3**

We present an example of data table used for multiple linear regression:

Zip	Income	College_ educated	White	TAR_ Video	TAR_ Phone	TAR_ inperson
ZIP 1	72,206	0.168	0.9723	0.0292	0.0287	0.2144
ZIP 2	85,478	0.246	0.9815	0.0596	0.0436	0.4656
ZIP 3	55,500	0.088	1.0000	0.0000	0.0063	0.0570
ZIP 4	77,989	0.309	0.9367	0.0089	0.0071	0.0931
ZIP 5	97,202	0.574	0.9016	0.0526	0.0335	0.3056
ZIP 6	74,107	0.222	0.9912	0.0174	0.0139	0.0992
.....						
ZIPx	63,509	0.165	0.9743	0.0381	0.0397	0.3844

 © 2021. Thieme. All rights reserved.

R linear regression command and results:

Call:

lm(formula = df$TAR_video – log(df$income) + log(df$white) + log(df$college_educated_rate))

Coefficients:

Estimate std. error *t*-value Pr(>|*t*|)

(Intercept) −20.4343 10.5328 −1.940 0.05474.

log(df$income) 1.4161 0.7844 1.805 0.01356*

log(df$white) −0.9482 0.4155 −2.282 0.02425*

log(df$college_educated_rate) 1.4037 0.5287 2.655 0.00902**

—

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

Residual standard error: 1.967 on 119 degrees of freedom

Multiple *R*-squared: 0.1562, Adjusted *R*-squared: 0.135

F-statistic: 7.345 on 3 and 119 DF, *p*-value: 0.0001472

Therefore, the number we retrieved from the multiple regression results are partly filled into ►Table 3:

Social determinant (log value)	Telemedicine groups	Coefficients	Std error	*p*-Value (Significance: "**" 0.01 ; "*" 0.05)
Median income	Video	1.4161	0.7844	0.0135*
White ratio	Video	−0.9482	0.4155	0.0242*
College education rate	Video	1.4037	0.5287	0.0090**

Statistical interpretation of the multiple regression results:

We treat TAR video as independent variable and median income, White ratio, college education rate as dependent variable, excluding all other possible factors to the TAR of video. Our multiple regression model shows that these three factors have significant impact on TAR of video. Specifically, a logged coefficient value of 1.41 on median income indicates that a 1% of median income increase will be expected to see increase an average of TAR increased by 1.41% from current level. A logged coefficient value of −0.9482 on White ratio indicates that a 1% of White ratio increase will be expected to see decrease an average of TAR increased by 0.9482% from current level. Similarly, a logged coefficient value of 1.4037 on college education rate indicates that a 1% of college education increase will be expected to see increase an average of TAR increased by 1.4037% from current level.

3. Two-way ANOVA tests of median_income, median_age, college_educated_rate and White_rate variable

We first build baseline model and four adjusted models that takes one of the variables out:

model_basedline = lm(df$UR_A – log(df$income) + log(df$white) + log(df$college_educated_rate) + log(df$median_age))

model_no_median_income = lm(df$UR_A – log(df$white) + log(df$college_educated_rate) + log(df$median_age))

model_no_White_rate = lm(df$UR_A – log(df$income) + log(df$college_educated_rate) + log(df$median_age))

model_no_college_edu = lm(df$UR_A – log(df$income) + log(df$white) + log(df$median_age))

model_no_median_age = lm(df$UR_A – log(df$income) + log(df$white) + log(df$college_educated_rate))

summary(model_basedline)

Call:

lm(formula = df$UR_A – log(df$income) + log(df$white) + log(df$college_educated_rate) + log(df$median_age))

Residuals:

Min 1Q median 3Q max

−3.0243 −1.3871 −0.5834 0.6984 8.4618

Coefficients:

Estimate std. error *t*-value Pr(>|*t*|)

(Intercept) −20.2198 11.1059 −1.821 0.0712.

log(df$income) 1.3773 1.0008 1.376 0.1714

log(df$white) −0.9514 0.4205 – 2.262 0.0255*

log(df$college_educated_rate) 1.4135 0.5533 2.555 0.0119*

log(df$median_age) 0.1012 1.6191 0.062 0.9503

—

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

Residual standard error: 1.975 on 118 degrees of freedom

Multiple *R*-squared: 0.1563, Adjusted *R*-squared: 0.1276

F-statistic: 5.463 on 4 and 118 DF, *p*-value: 0.0004529

Note: the variable median_age was introduced so it changes a bit of the statistics and *p*-values on other variables.

From this multiple regression model results, only White_rate and college_educated rate are significant.

a. Two-way ANOVA: For median_income variable:

Hypothesis: β_zero, the estimate of median_income is 0, which means median_income variable, the median income of zip code area, have no effect to the TAR_video.

We perform an ANOVA *F*-test, compare two models with and without median income variable in the multiple regression model:

model_basedline = lm(df$UR_A – log(df$income) + log(df$white) + log(df$college_educated_rate) + log(df$median_age))

model_no_median_income = lm(df$UR_A – log(df$white) + log(df$college_educated_rate) + log(df$median_age))

The two-way ANOVA comparison and their statistics are performed below:

ANOVA(model_basedline, model_no_median_income)

A ANOVA: 2 × 6						
	Res.Df	RSS	Df	Sum of sq	*F*	Pr(>*F*)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	118	460.2164	NA	NA	NA	NA
2	119	467.6031	−1	−7.386664	1.893949	0.1713633

The result shows a nonsignificant result (*p*-value of *F*-test is 0.1713633). We do not reject the hypothesis: median income of zip code area has no effect to the TAR_video. It means

 © 2021. Thieme. All rights reserved.

when the White_rate, college_educated_rate, median_age variable are given in the model, we cannot associate median_income predictor with TAR_video.

b. Two-way ANOVA: For White variable:

Hypothesis: β_zero, the estimate of White_rate is 0, which means White_rate variable makes no effect to the TAR_video.

We perform an ANOVA *F*-test, compare two models with and without median income variable in the multiple regression model:

```
model_basedline = lm(df$UR_A – log(df$income) + log
(df$white) + log(df$college_educated_rate) + log(df
$median_age))
model_no_w_rate = lm(df$UR_A – log(df$income) + log
(df$college_educated_rate) + log(df$median_age))
```

The two-way ANOVA comparison and their statistics are performed below:

ANOVA(model_basedline, model_no_White_rate)

A ANOVA: 2 × 6						
	Res.Df	RSS	Df	Sum of sq	*F*	Pr(>*F*)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	118	460.2164	NA	NA	NA	NA
2	119	480.1793	−1	−19.96283	5.11849	0.02550193

The result shows a significant result (*p*-value of *F* test is 0.0255). We reject the hypothesis.

The test indicates that White_rate has negative effect to the TAR_video. It means when the median_income, college_educated_rate, median_age variables are set in the model, the White race rate of an area is negatively associated with TAR_video.

c. Two-way ANOVA: For college_edu variable:

Similarly, we perform another two ANOVA tests for college_education and median_age variable:

ANOVA(model_basedline, model_no_college_edu)

A ANOVA: 2 × 6						
	Res.Df	RSS	Df	Sum of sq	*F*	Pr(>*F*)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	118	460.2164	NA	NA	NA	NA
2	119	485.6702	−1	−25.45371	6.526359	0.01190119

The result shows a significant outcome (*p*-value of *F* test is 0.0119). We reject the hypothesis that the college education rate variable makes no effect on the TAR_video variable.

The test indicates that college_edu variable has positive effect to the TAR_video. It means when the median_income, White_rate, and median_age variables are set in the model, the college education rate of an area is positively associated with TAR_video.

d. Two-way ANOVA: For median_age variable:

ANOVA(model_basedline, model_no_median_age)

A ANOVA: 2 × 6						
	Res.Df	RSS	Df	Sum of sq	*F*	Pr(>*F*)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	118	460.2164	NA	NA	NA	NA
2	119	460.2317	−1	−0.01522277	0.003903135	0.9502902

The result shows a nonsignificant result (*p*-value of *F* test is 0.95). We do not reject the hypothesis: Median age of zip code area has no effect to the TAR_video. It means when the White_rate, college_educated_rate, median_income variables are given in the model, we cannot associate median_age predictor with TAR_video.

Note: $^{*}p \leq 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

 © 2021. Thieme. All rights reserved.