Supplementary Materials

The Association of Stay-at-Home Orders and COVID-19 Incidence and Mortality in Rural and Urban United States

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

Model Regression

eMethods

Analysis of the data was approached utilizing the following mixed effects count data models: Poisson (lm1glmmrelev); zero-inflated Poisson (lm2relev, lm2catziprelev, lm2catziprelev_cs_cdate, lm2catziprelev_toep_cdate, lm2catziprelev_toep_date2, and lm2catziprelev_us_date2); zero-inflated Poisson with random intercept and slope (lm2catziprelev_randslope_cdate and lm2catziprelev_randslope_date2); negative binomial (lm3glmmrelev); negative binomial with random intercept and slope (lm3glmmRandslope); zero-inflated negative binomial (lm4catziprelev). "Catzip" refers to only using the categories of dates (during and after stay-at-home orders) and their interactions with county type for the zero inflated model, instead of all of the variables used in the conditional model.

All models used the same variables for the fixed effects, as all are necessary to account for the time varying nature of stay-at-home orders. In addition, all models were offset by the population of the county divided by 100,000 to standardize the results per 100,000 people. All models were ran using the glmmTMB package in R.[1]Summary results of each model are detailed below, where URBinary represents the rurality status (a dummy variable that is 0 for rural counties and 1 for urban counties), c_daterelevduring SaH represents the stay-at-home order status (a dummy variable that is 0 for not during stay-at-home orders and 1 for during stay-at-home orders), c_daterelevafter SaH represents another indication of the stay-at-home order status (a dummy variable that is 0 for after stay-at-home orders and 1 for after stay-at-home orders), Date2 represents the number of days since January 22, 2020, dsahcarried represents the number of days under stay-at-home orders at a given time and the total number of days under stay-at-home orders while c_daterelevafter SaH is 1, asahcarried represents the number of days since the end of stay-at-home orders, URBinary:c_daterelevduring SaH represents the interaction term between the rurality status and stay-at-home order status (a dummy variable that is 0 for rural counties and for urban counties not under stay-at-home orders, and 1 for urban counties under stay-at-home orders), URBinary:c_daterelevafter SaH represents another interaction term between the rurality status and stay-at-home order status (a dummy variable that is 0 for rural counties and for urban counties not after stay-at-home orders, and 1 for urban counties after stay-at-home orders), URBinary:Date2 represents the interaction term between the number of days since January 22, 2020 and the rurality status (0 for rural counties and 1 through 142 for urban counties), URBinary:dsahcarried represents the interaction term between the number of days under stay-at-home orders and the rurality status (0 for rural counties and 0 for urban counties before stay-at-home orders), URBinary:asahcarried represents the interaction term between the number of days after stay-at-home orders and the rurality status (0 for rural counties and 0 for urban counties before the end of stay-at-home orders).

The models were compared on the basis of model diagnostics, Akaike information criterion (AIC), and parsimony (preferring non zero-inflated models where appropriate and prioritizing model diagnostics). All models were consistent in terms of estimate signs and significance.

Model diagnostics were performed examining the model's simulated quantile scaled residuals using the DHARMA package in R.[2] The models were assessed for over-dispersion, zero-inflation, and expected distribution of the residuals. The mixed effects negative binomial model with random intercept by county was found to be statistically significantly not zero-inflated and having normally distributed residuals, but over-dispersed and having outliers. To examine if this over-dispersion was due to the presence of outliers, the model was rerun after outliner counties (369 of 3142) were removed, but this restricted model was still over-dispersed. The models were also assessed for temporal autocorrelation using the Durbin-Watson test in the DHARMa package.[2] The zero inflated Poisson model (lm2catziprelev) was found to only be temporally auto correlated and thus was chosen to be the best model. It was examined further using variance-covariance structures in an attempt to remove the temporal autocorrelation (lm2catziprelev_cs_cdate, lm2catziprelev_toep_cdate, lm2catziprelev_toep_date2, and lm2catziprelev_us_date2). Compound symmetry (cs_cdate) and Toeplitz (toep_cdate) structures where the only structures out of AR(1), compound symmetry, Toeplitz, and unstructured to converge using categorical date. Similarly, Toeplitz (toep_date2) and unstructured (us_date2) were the only structures able to converge using days since January 22nd. All attempts to remove temporal autocorrelation were inadequate and detrimental to the overall fit of the model. Temporal autocorrelation was thus deemed unavoidable. Moreover, it did not have a significant effect on the results because of the long follow-up time, the significance of the results, and the large number of counties.[3]

The final model chosen was the zero inflated Poisson model using the categories of dates and their interactions with county type for the zero inflation model (lm2catziprelev). The equations of the final model are:

$$\Pr(Y_{ij} = y_{ij}) = \begin{cases} \pi_{ij} + (1-\pi_{ij})\exp(-\mu_{ij}), & if\ y_{ij} = 0 \\ (1-\pi_{ij})\frac{\mu_{ij}^{y_{ij}}\exp(-\mu_{ij})}{y_{ij}!}, & if\ y_{ij} > 0 \end{cases} \qquad \text{(Equation 1)}$$

$$logit(\pi_{ij}) = a_0 + a_1 Rurality_i + a_2 Under\_SAH_{ij} + a_3 After\_SAH_{ij} + a_4 Rurality_i * Under_{SAH_{ij}} + a_5 Rurality_i * After_{SAH_{ij}} \qquad \text{(Equation 2)}$$

$$\text{Log}(\mu_{ij}) = \log\left(\frac{Population_i}{100{,}000}\right) + \beta_0 + \beta_1 Rurality_i + \beta_2 Under\_SAH_{ij} + \beta_3 After\_SAH_{ij} + \beta_4 Days_{ij} + \beta_5 Days\_Under\_SAH_{ij} + \beta_6 Days\_After\_SAH_{ij} + \beta_7 Rurality_i * Under\_SAH_{ij} + \beta_8 Rurality_i * After\_SAH_{ij} + \beta_9 Rurality_i * Days_{ij} + \beta_{10} Rurality_i * Days\_Under\_SAH_{ij} + \beta_{11} Rurality_i * Days\_After\_SAH_{ij} + b_{1i} \qquad \text{(Equation 3)}$$

where Equation 1 is the probability distribution, Equation 2 is the zero inflation model, and Equation 3 is the Poisson model. Y_{ij} represents the 14-day lagged incidence of COVID-19 in the i^{th} county on the j^{th} day (technically the $(j+14)^{th}$ day) represents the probability of being 0 for the i^{th} county on the j^{th} day, μ_{ij} represents the 14-day lagged incidence of COVID-19 in the i^{th} county on the j^{th} day, b_i represents the random effect of the i^{th} county, $Population_i$ represents the population of the i^{th} county, Rurality_i represents the rurality status of the i^{th} county (a dummy variable that is 0 for rural counties and 1 for urban counties), Under_SAH_{ij} represents the stay-at-home order status of the i^{th} county on the j^{th} day (a dummy variable that is 0 for not during stay-at-home orders and 1 for during stay-at-home orders), After_SAH_{ij} represents another indication of the stay-at-home order status of the i^{th} county on the j^{th} day (a dummy variable that is 0 for after stay-at-home orders and 1 for after stay-at-home orders), Days_{ij} represents the number of days since January 22, 2020 for the i^{th} county on the j^{th} day, Days_Under_SAH$_{ij}$ represents the number of days under stay-at-home orders for the i^{th} county on the j^{th} day, Days_After_SAH$_{ij}$ represents the number of days since the end of stay-at-home orders for the i^{th} county on the j^{th} day, Rurality*Under_SAH$_{ij}$ represents the interaction term between the rurality status of the i^{th} county and stay-at-home order status for the i^{th} county on the j^{th} day (a dummy variable that is 0 for rural counties and for urban counties not under stay-at-home orders, and 1 for urban counties under stay-at-home orders), Rurality*After_SAH$_{ij}$ represents another interaction term between the rurality status of the i^{th} county and stay-at-home order status for the i^{th} county on the j^{th} day (a dummy variable that is 0 for rural counties and for urban counties not after stay-at-home orders, and 1 for urban counties after stay-at-home orders), Rurality*Days$_{ij}$ represents the interaction term between the number of days since January 22, 2020 and the rurality status for the i^{th} county on the j^{th} day (0 for rural counties and 1 through 142 for urban counties), Rurality*Days_Under_SAH$_{ij}$ represents the interaction term between the number of days under stay-at-home orders and the rurality status for the i^{th} county on the j^{th} day (0 for rural counties and 0 for urban counties before stay-at-home orders), Rurality*Days_After_SAH$_{ij}$ represents the interaction term between the number of days after stay-at-home orders and the rurality status for the i^{th} county on the j^{th} day (0 for rural counties and 0 for urban counties before the end of stay-at-home orders).

Therefore, a_0 represents the baseline log odds of being a “zero” day for a typical county at j = 0 (in that the zero inflated model assumes two zero generating processes, the first generating zeros, the top half of equation 1, and the second a Poisson process that generates counts including zeros, the bottom half of equation 1. In this case a “zero” day is one that never had the chance of being a count), a_1 represents the change in the log odds of being a zero for urban counties, a_2 represents the change in the log odds during stay-at-home orders, a_3 represents the change in the log odds after stay-at-home orders, a_4 represents the additional change in the log odds during stay-at-home orders for urban counties, a_5 represents the additional change in the log odds after stay-at-home orders for urban counties, β_0 represents the baseline outcome (i.e. 14-day lagged new daily cases of COVID-19) for a typical county at j = 0, b_i represents the random effects (the random intercept) which is the change in baseline outcome from the typical county for the i^{th} county (that is $\beta_0 + b_1$ is the baseline outcome for the 1^{st} county), $\beta_1$represents the change in the outcome for urban counties, β_2 represents the change in the outcome during stay-at-home orders, β_3 represents the change in the outcome after stay-at-home orders, β_4 represents the change in the outcome for each day since j =0 (January 22, 2020), β_5 represents the change in the outcome for each day a county was under stay-at-home orders, β_6 represents the change in the outcome for each day a county was out of stay-at-home orders, β_7 represents the additional change in the outcome for urban counties during stay-at-home orders (that is for urban counties the “actual β_2” is $\beta_2 + \beta_7$), β_8 represents the additional change in the outcome for urban counties after stay-at-home orders, β_9 represents the additional change in the outcome for each day since j =0 (January 22, 2020), β_{10} represents the additional change in the outcome for urban counties for each day it was under stay-at-home orders, β_{11} represents the additional change in the outcome for urban counties for each day it was out of stay-at-home orders.

Figure 2 Generation

Figure 2 was generated by inputting the estimates of fixed effects and the urban and rural averages of stay-at-home orders start and end dates. The outcome was divided by the offset to

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

standardize the results per 100,000 population. The respective offsets for urban and rural counties were calculated using urban and rural counties respective population averages. Similarly, the extrapolations were generated by using the conditional model only with intercept and variables: Rurality, Days, and Rurality*Days. The extrapolations represent continuation of the before stay-at-home order trends.

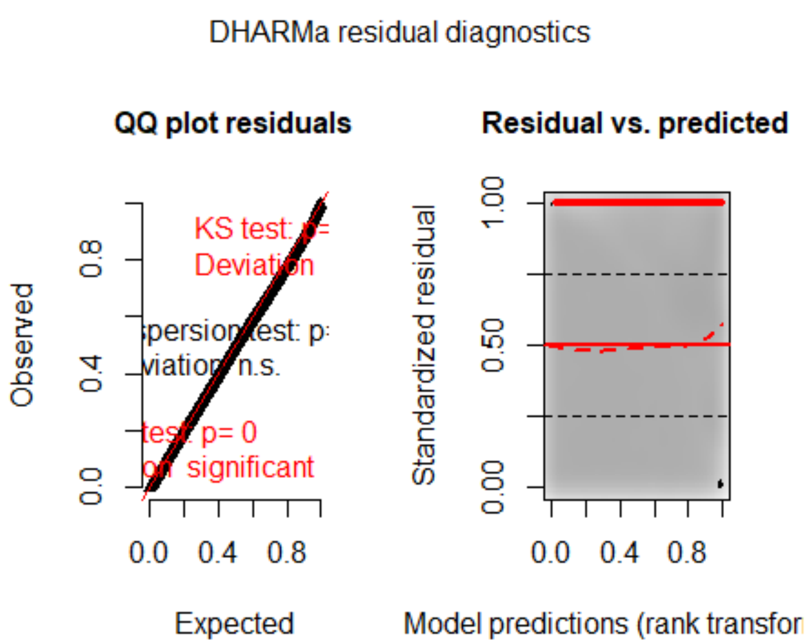
County Level Cases Analysis

```
##  Family: poisson  ( log )
## Formula:
## newcase_nst_14 ~ offset(popoff) + URBinary * c_daterelev + URBinary *
##     Date2 + URBinary * dsahcarried + URBinary * asahcarried + (1 | c_FIPS)
## Zero inflation:                  ~URBinary * c_daterelev
## Data: df_14
##
##      AIC      BIC   logLik deviance df.resid
##  2220521  2220730 -1110242  2220483   446145
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  c_FIPS (Intercept) 1.389    1.179
## Number of obs: 446164, groups:  c_FIPS, 3142
##
## Conditional model:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -0.6452101  0.0342097  -18.86  < 2e-16 ***
## URBinary                       -0.9052841  0.0496647  -18.23  < 2e-16 ***
## c_daterelevafter SaH            0.3348040  0.0152451   21.96  < 2e-16 ***
## c_daterelevduring SaH           0.4726759  0.0111321   42.46  < 2e-16 ***
## Date2                           0.0208674  0.0002576   81.02  < 2e-16 ***
## dsahcarried                    -0.0184094  0.0003317  -55.50  < 2e-16 ***
## asahcarried                    -0.0054122  0.0004206  -12.87  < 2e-16 ***
## URBinary:c_daterelevafter SaH  -0.5307952  0.0159890  -33.20  < 2e-16 ***
## URBinary:c_daterelevduring SaH -0.1659420  0.0115200  -14.40  < 2e-16 ***
## URBinary:Date2                  0.0215742  0.0003019   71.46  < 2e-16 ***
## URBinary:dsahcarried           -0.0309239  0.0003699  -83.60  < 2e-16 ***
## URBinary:asahcarried            0.0023423  0.0004701    4.98 6.28e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Zero-inflation model:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     0.83343    0.01385   60.19   <2e-16 ***
## URBinary                       -0.59437    0.01740  -34.15   <2e-16 ***
## c_daterelevafter SaH           -1.37483    0.02111  -65.13   <2e-16 ***
## c_daterelevduring SaH          -1.08682    0.01874  -57.99   <2e-16 ***
## URBinary:c_daterelevafter SaH  -0.55237    0.03219  -17.16   <2e-16 ***
## URBinary:c_daterelevduring SaH -0.80902    0.02630  -30.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

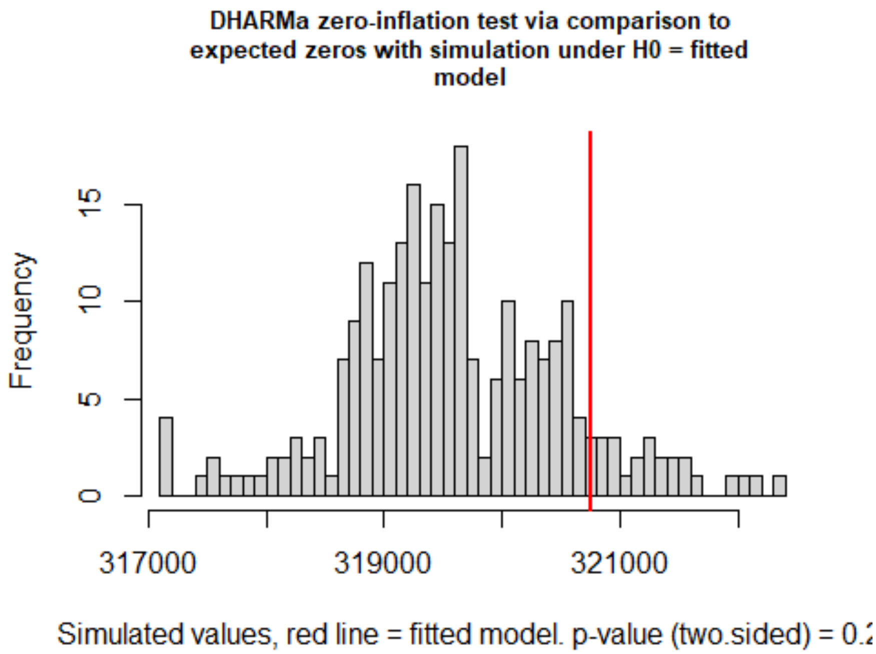
Cases Analysis Diagnostics

```
SimOut_lm2catziprelev <- simulateResiduals(fittedModel = lm2catziprelev, plot
= T)

## DHARMa:plot used testOutliers with type = binomial for computational
reasons (nObs > 500). Note that this method may not have inflated Type I
error rates for integer-valued distributions. To get a more exact result, it
is recommended to re-run testOutliers with type = 'bootstrap'. See
?testOutliers for details

plot(SimOut_lm2catziprelev)

## DHARMa:plot used testOutliers with type = binomial for computational
reasons (nObs > 500). Note that this method may not have inflated Type I
error rates for integer-valued distributions. To get a more exact result, it
is recommended to re-run testOutliers with type = 'bootstrap'. See
?testOutliers for details
```



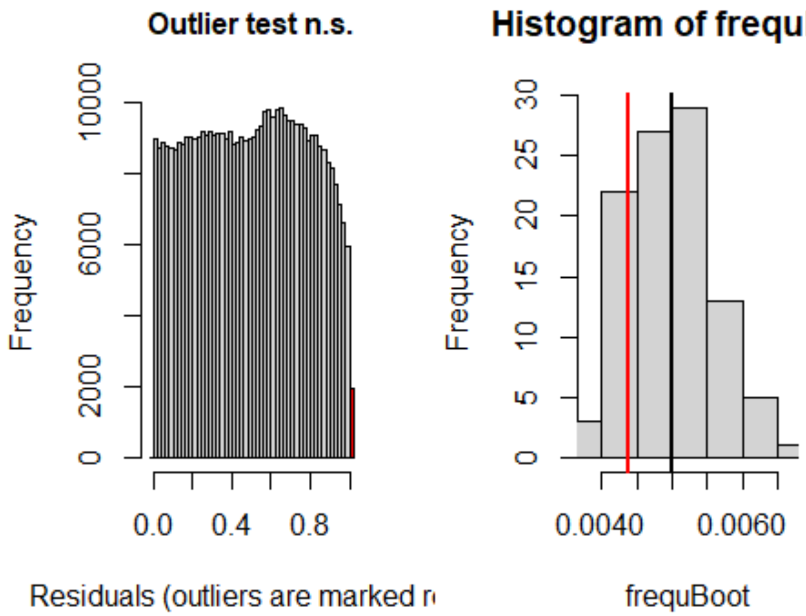

```
testZeroInflation(SimOut_lm2catziprelev)
```


Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791



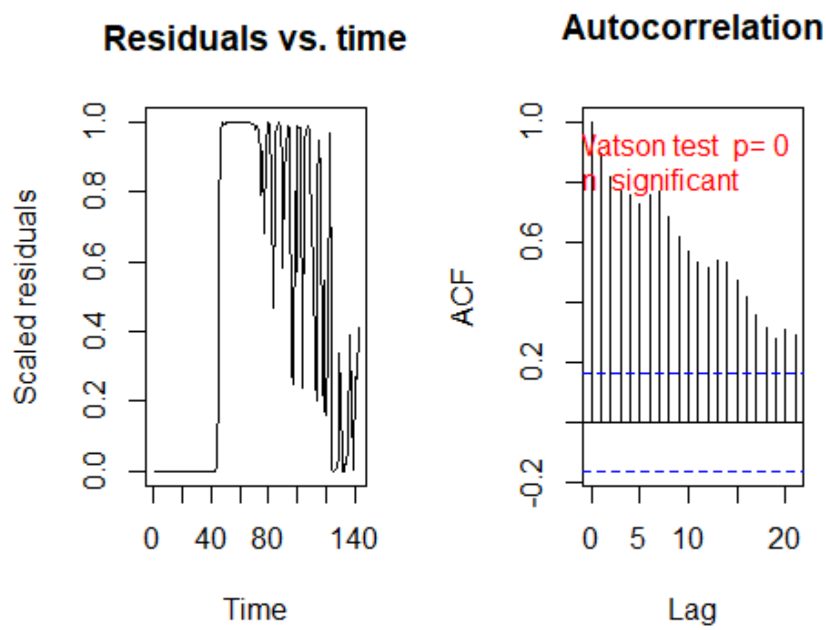

```
##
##  DHARMa zero-inflation test via comparison to expected zeros with
##  simulation under H0 = fitted model
##
## data:  simulationOutput
## ratioObsSim = 1.0037, p-value = 0.2
## alternative hypothesis: two.sided
```

```
testOutliers(SimOut_lm2catziprelev, type= 'bootstrap')
```




```
##
##  DHARMa bootstrapped outlier test
##
## data:  SimOut_lm2catziprelev
## outliers at both margin(s) = 1953, observations = 446164, p-value =
## 0.28
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.003989508 0.006183937
## sample estimates:
## outlier frequency (expected: 0.00499551734339839 )
##                                         0.004377314
```

```
simoutrecalc <- recalculateResiduals(SimOut_lm2catziprelev, group =
df_14$Date2)
testTemporalAutocorrelation(simoutrecalc, time = unique(df_14$Date2))
```




```
##
##  Durbin-Watson test
##
## data:  simulationOutput$scaledResiduals ~ 1
## DW = 0.19769, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is not 0
```

```
#
```

Death Analysis

```
> summary(lm3glmmRandslope)
 Family: nbinom2  ( log )
Formula:          newcase_nst_14 ~ offset(popoff) + URBinary * c_daterelev + URBi
nary *
    Date2 + URBinary * dsahcarried + URBinary * asahcarried +      (1 + c_daterel
ev | c_FIPS)
Data: df_14

      AIC       BIC    logLik  deviance  df.resid
 873199.8  873409.0 -436580.9  873161.8    446145

Random effects:

Conditional model:
 Groups Name                   Variance Std.Dev. Corr
 c_FIPS (Intercept)            8.582    2.9295
        c_daterelevbefore SaH 6.205    2.4911   0.41
        c_daterelevduring SaH 0.911    0.9545   0.62 0.88
Number of obs: 446164, groups:  c_FIPS, 3142

Overdispersion parameter for nbinom2 family (): 15.7

Conditional model:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -4.4667557  0.0893028  -50.02  < 2e-16 ***
URBinary                      -0.7347610  0.1240524   -5.92 3.16e-09 ***
c_daterelevbefore SaH         -3.5596847  0.1540933  -23.10  < 2e-16 ***
c_daterelevduring SaH         -0.9052158  0.0492548  -18.38  < 2e-16 ***
Date2                          0.0446022  0.0002627  169.77  < 2e-16 ***
dsahcarried                   -0.0234901  0.0003332  -70.50  < 2e-16 ***
asahcarried                   -0.0308345  0.0004076  -75.65  < 2e-16 ***
URBinary:c_daterelevbefore SaH  2.2134155  0.1452720   15.24  < 2e-16 ***
URBinary:c_daterelevduring SaH  0.6358044  0.0504837   12.59  < 2e-16 ***
URBinary:Date2                 0.0424600  0.0003560  119.27  < 2e-16 ***
URBinary:dsahcarried          -0.0441241  0.0004229 -104.34  < 2e-16 ***
URBinary:asahcarried          -0.0440375  0.0005319  -82.79  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

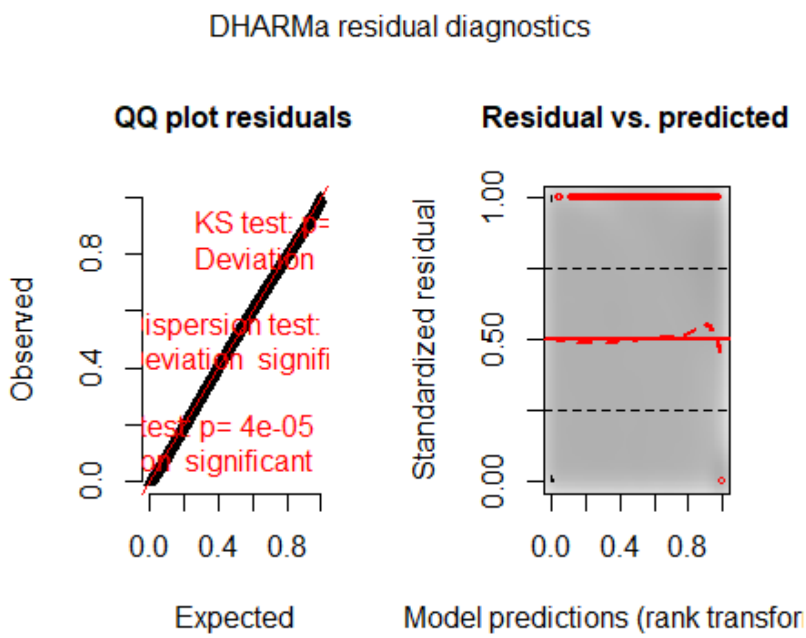
Death Analysis Diagnostics

```
SimOut_lm3glmmRandSlope <- simulateResiduals(fittedModel = lm3glmmRandslope,
plot = T)
```

```
## DHARMa:plot used testOutliers with type = binomial for computational
reasons (nObs > 500). Note that this method may not have inflated Type I
error rates for integer-valued distributions. To get a more exact result, it
is recommended to re-run testOutliers with type = 'bootstrap'. See
?testOutliers for details
```

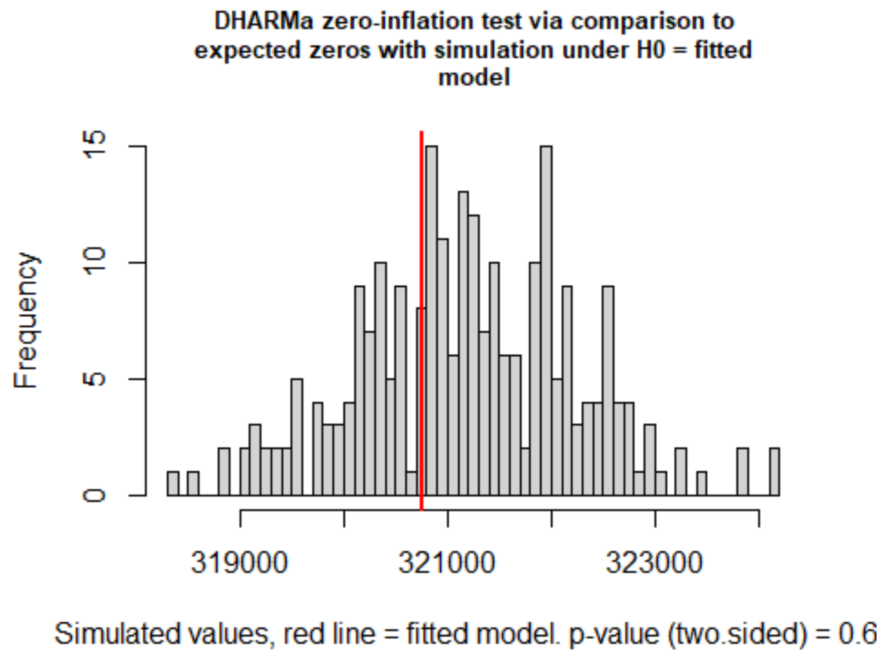
```
plot(SimOut_lm3glmmRandSlope)
```

```
## DHARMa:plot used testOutliers with type = binomial for computational
reasons (nObs > 500). Note that this method may not have inflated Type I
error rates for integer-valued distributions. To get a more exact result, it
is recommended to re-run testOutliers with type = 'bootstrap'. See
?testOutliers for details
```



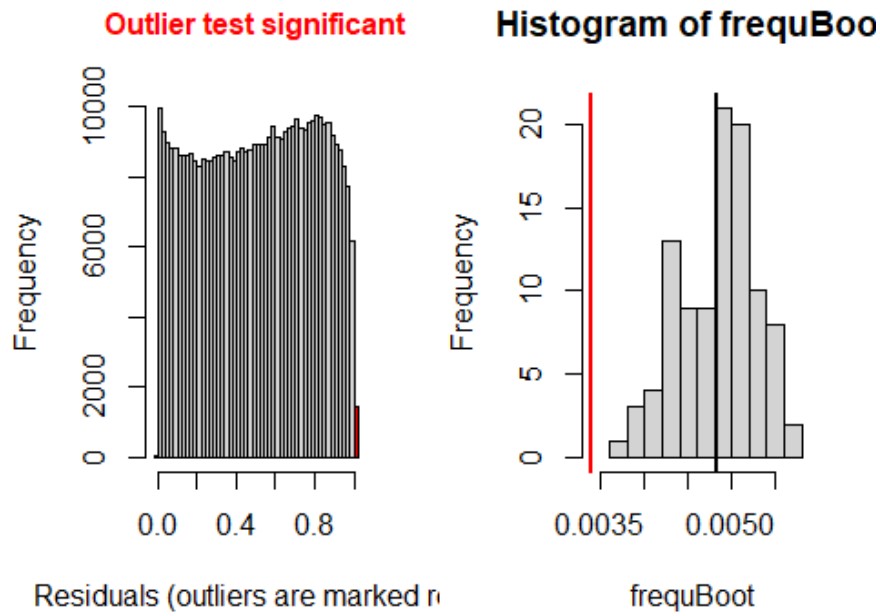

```
testZeroInflation(SimOut_lm3glmmRandSlope)
```


Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791



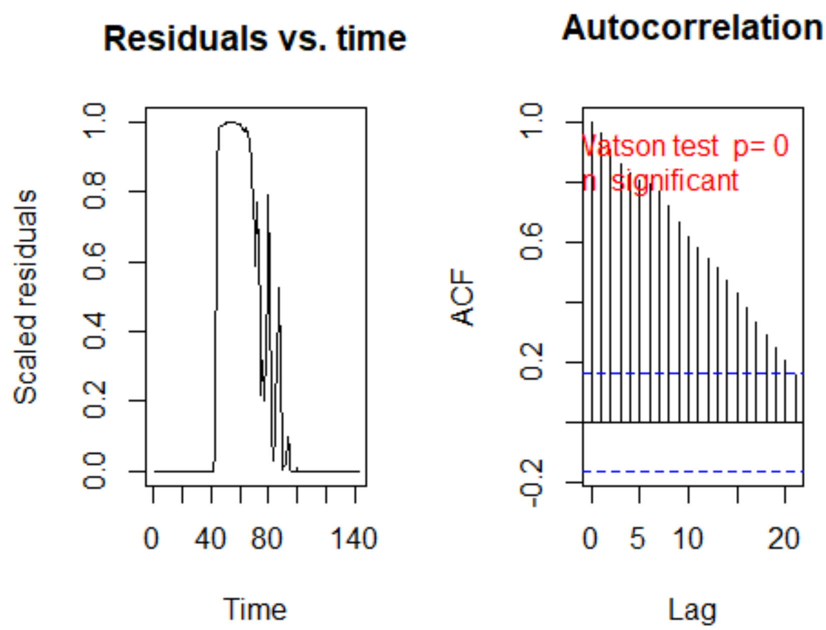

```
##
##  DHARMa zero-inflation test via comparison to expected zeros with
##  simulation under H0 = fitted model
##
## data:  simulationOutput
## ratioObsSim = 0.88331, p-value = 0.64
## alternative hypothesis: two.sided
```

```
testOutliers(SimOut_lm3glmmRandSlope, type= 'bootstrap')
```




```
##
##  DHARMa bootstrapped outlier test
##
## data:  SimOut_lm3glmmRandSlope
## outliers at both margin(s) = 9, observations = 446164, p-value <
## 2.2e-16
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.003867636 0.008386714
## sample estimates:
## outlier frequency (expected: 0.00617835594086479 )
##                                         2.017195e-05
```

```
simoutrecalc <- recalculateResiduals(SimOut_lm3glmmRandSlope, group =
df_14$Date2)
testTemporalAutocorrelation(simoutrecalc, time = unique(df_14$Date2))
```




```
##
##  Durbin-Watson test
##
## data:  simulationOutput$scaledResiduals ~ 1
## DW = 0.072204, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is not 0
```

```
##
```

Sensitivity Analysis

We conducted a sensitivity analysis surrounding the lag time between daily new cases and time reported. The dependent variable, daily new cases, in this case must be lagged for proper analysis because of several reasons. First, it is well known that the potential incubation period for SARS-CoV-2 is upwards of 14 days, which would imply that an individual being tested positive for the virus could have been exposed to the virus some two weeks earlier, potentially placing them out of range of a particular stay-at-home order, and thus complicating analysis. Second, while stay-at-home orders are declared and in place, it takes time for the orders to be adhered to and enforced for a measurable effect. We initially used the longer 14-day lag due to its being the incubation period. However, other studies have utilized five-to-ten-day lags. Therefore, it becomes necessary to conduct sensitivity analysis, the result of which we report below.

Results of the sensitivity analysis did not change any of the study inferences of conclusions. Moreover, the five-day and ten-day lag analyses exhibited significant overdispersion and zero-inflation, whereas the 14-day lag does not exhibit these characteristics

Five-Day Lag

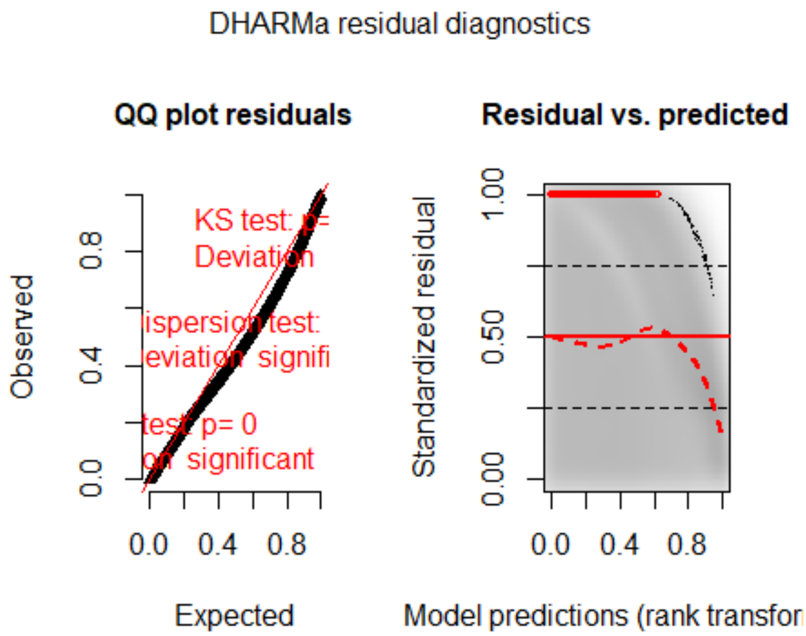
```
#reads in data
setwd("C:\\Users\\Jake\\Desktop\\MAYO\\COVID RURALITY")
df_14 <- read.csv("df_14.csv",header=T)

#installs packages then loads them into the session
library(glmmTMB)

## Warning: package 'glmmTMB' was built under R version 3.6.3

library(DHARMa)

## Warning: package 'DHARMa' was built under R version 3.6.3

## This is DHARMa 0.3.3.0. For overview type '?DHARMa'. For recent changes, t
ype news(package = 'DHARMa') Note: Syntax of plotResiduals has changed in 0.3
.0, see ?plotResiduals for details

# Releveling
df_14$c_daterelev <- relevel(df_14$c_date, ref = "before SaH")


#Five Day Lag
n <- 142
D <- 5
for (i in 1:n){
  df_14$newcase_nst_5[df_14$Date2 == i] <- ifelse( i > (n-D), df_14$newcase_n
st_14[df_14$Date2 == (i-(14-D))], df_14$newcase_nst[df_14$Date2 == (i+D)])
}

#RENAMING THE VARIABLE TO ALLOW the implementation of the lag
```


Jiang DH, et al. BMJ Open 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

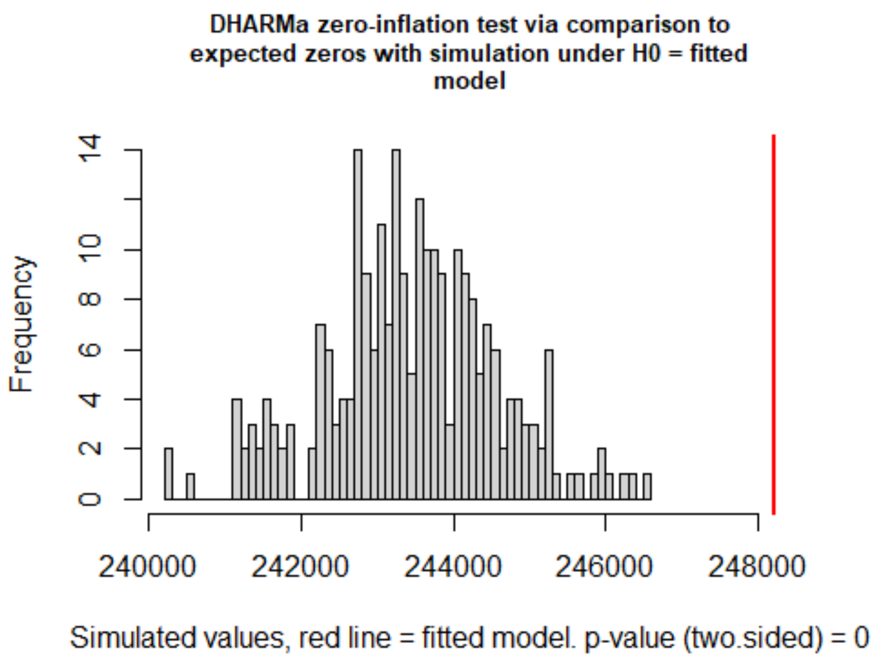
```
df_14$newcase_nst_14 <- df_14$newcase_nst_5


load("C:/Users/Jake/Desktop/MAYO/COVID RURALITY/5Day.RData")
###########################################
############# SUMMARY RESULTS #############
###########################################



# Zero inflated poisson mixed effects (zero inflated using the rurality and d
ates)
summary(lm2catziprelev)

##  Family: poisson  ( log )
## Formula:
## newcase_nst_14 ~ offset(popoff) + URBinary * c_daterelev + URBinary *
##     Date2 + URBinary * dsahcarried + URBinary * asahcarried +      (1 | c_
FIPS)
## Zero inflation:                  ~URBinary * c_daterelev
## Data: df_14
##
##       AIC       BIC    logLik  deviance  df.resid
## 1385307.0 1385516.1 -692634.5 1385269.0    446145
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  c_FIPS (Intercept) 1.546    1.243
## Number of obs: 446164, groups:  c_FIPS, 3142
##
## Conditional model:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.3991715  0.0305346  -45.82  < 2e-16 ***
## URBinary                     -1.7737967  0.0540949  -32.79  < 2e-16 ***
## c_daterelevafter SaH          1.1083193  0.0122324   90.61  < 2e-16 ***
## c_daterelevduring SaH         0.9587011  0.0109879   87.25  < 2e-16 ***
## Date2                         0.0404765  0.0001284  315.14  < 2e-16 ***
## dsahcarried                  -0.0087794  0.0002006  -43.76  < 2e-16 ***
## asahcarried                  -0.0167820  0.0001979  -84.82  < 2e-16 ***
## URBinary:c_daterelevafter SaH  -0.3972812  0.0231576  -17.16  < 2e-16 ***
## URBinary:c_daterelevduring SaH -0.3212096  0.0215502  -14.91  < 2e-16 ***
## URBinary:Date2                0.0024167  0.0003101    7.79 6.53e-15 ***
## URBinary:dsahcarried         -0.0014551  0.0003835   -3.79 0.000148 ***
## URBinary:asahcarried         -0.0017310  0.0003994   -4.33 1.46e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Zero-inflation model:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -1.54573    0.02600  -59.45   <2e-16 ***
## URBinary                          1.05664    0.04693   22.52   <2e-16 ***
## c_daterelevafter SaH             -1.62265    0.03622  -44.80   <2e-16 ***
## c_daterelevduring SaH             0.40845    0.02847   14.35   <2e-16 ***
## URBinary:c_daterelevafter SaH  -1.23150    0.06411  -19.21   <2e-16 ***
## URBinary:c_daterelevduring SaH -1.15857    0.05027  -23.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


SimOut_lm2catziprelev <- simulateResiduals(fittedModel = lm2catziprelev, plot
= T)

## DHARMa:plot used testOutliers with type = binomial for computational reaso
ns (nObs > 500). Note that this method may not have inflated Type I error rat
es for integer-valued distributions. To get a more exact result, it is recomm
ended to re-run testOutliers with type = 'bootstrap'. See ?testOutliers for d
etails

plot(SimOut_lm2catziprelev)

## DHARMa:plot used testOutliers with type = binomial for computational reaso
ns (nObs > 500). Note that this method may not have inflated Type I error rat
es for integer-valued distributions. To get a more exact result, it is recomm
ended to re-run testOutliers with type = 'bootstrap'. See ?testOutliers for d
etails
```

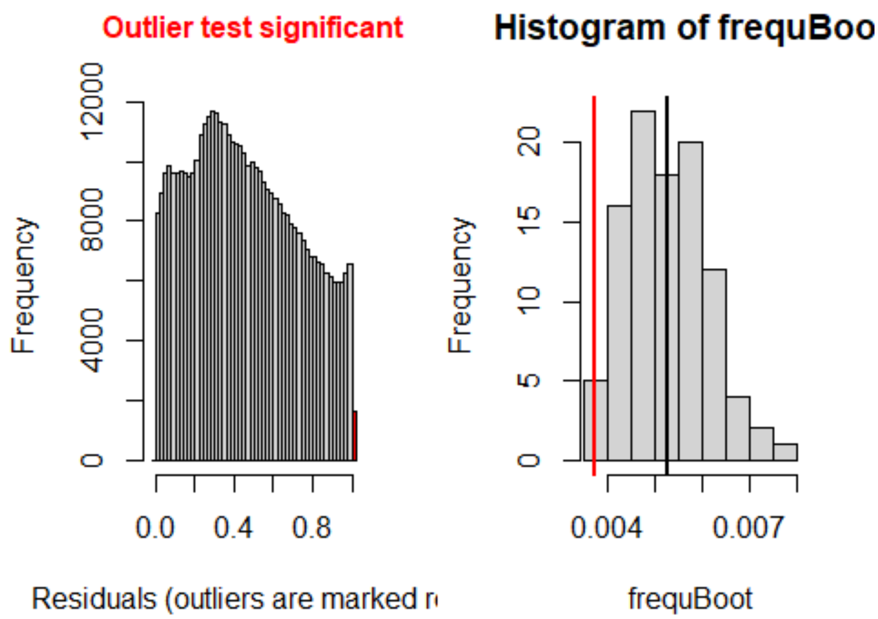

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791



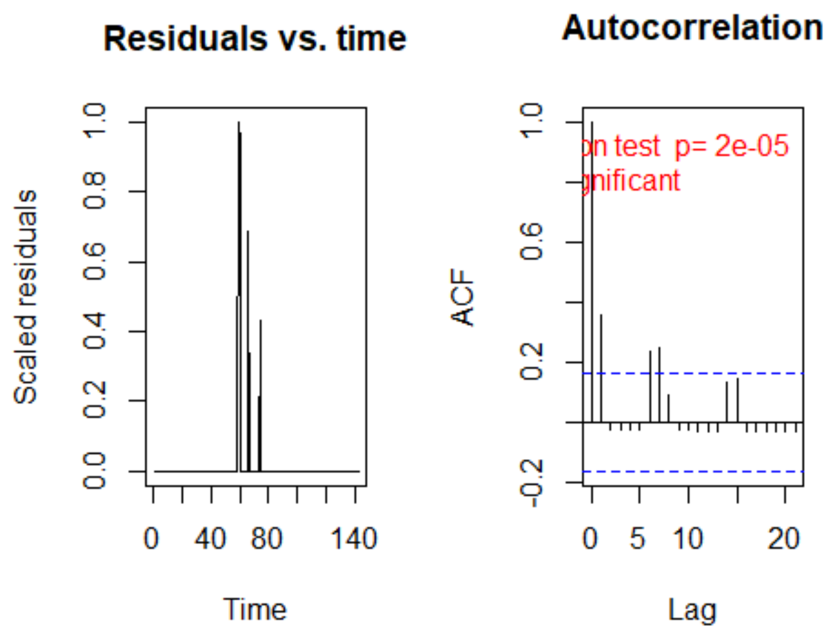

```
testZeroInflation(SimOut_lm2catziprelev)
```



```
##
##  DHARMa zero-inflation test via comparison to expected zeros with
##  simulation under H0 = fitted model
##
## data:  simulationOutput
## ratioObsSim = 1.0196, p-value < 2.2e-16
## alternative hypothesis: two.sided

testOutliers(SimOut_lm2catziprelev, type= 'bootstrap')
```




```
##
##  DHARMa bootstrapped outlier test
##
## data:  SimOut_lm2catziprelev
## outliers at both margin(s) = 1660, observations = 446164, p-value =
## 0.04
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.003838611 0.007088649
## sample estimates:
## outlier frequency (expected: 0.0052525080463686 )
##                                         0.003720605

simoutrecalc <- recalculateResiduals(SimOut_lm2catziprelev, group = df_14$Dat
e2)
testTemporalAutocorrelation(simoutrecalc, time = unique(df_14$Date2))
```




```
##
##  Durbin-Watson test
##
## data:  simulationOutput$scaledResiduals ~ 1
## DW = 1.2925, p-value = 2.181e-05
## alternative hypothesis: true autocorrelation is not 0
```

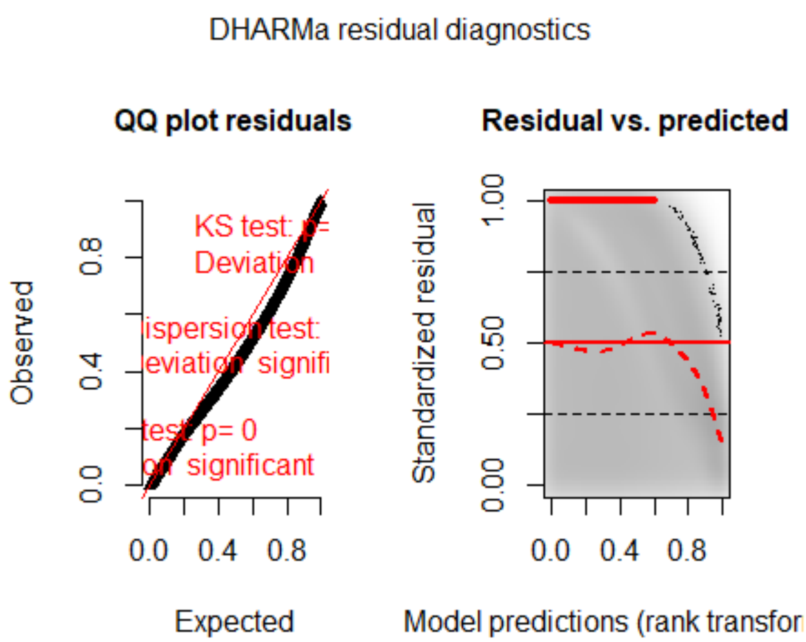
Ten-Day Lag

```
#reads in data
setwd("C:\\Users\\Jake\\Desktop\\MAYO\\COVID RURALITY")
df_14 <- read.csv("df_14.csv",header=T)

#installs packages then loads them into the session
library(glmmTMB)

## Warning: package 'glmmTMB' was built under R version 3.6.3

library(DHARMa)

## Warning: package 'DHARMa' was built under R version 3.6.3

## This is DHARMa 0.3.3.0. For overview type '?DHARMa'. For recent changes, t
ype news(package = 'DHARMa') Note: Syntax of plotResiduals has changed in 0.3
.0, see ?plotResiduals for details

# Releveling
df_14$c_daterelev <- relevel(df_14$c_date, ref = "before SaH")


#ten Day Lag
n <- 142
D <- 10
for (i in 1:n){
  df_14$newcase_nst_10[df_14$Date2 == i] <- ifelse( i > (n-D), df_14$newcase_
nst_14[df_14$Date2 == (i-(14-D))], df_14$newcase_nst[df_14$Date2 == (i+D)])
}

#RENAMING THE VARIABLE TO ALLOW the implementation of the lag

df_14$newcase_nst_14 <- df_14$newcase_nst_10


load("C:/Users/Jake/Desktop/MAYO/COVID RURALITY/10day.RData")
###########################################
############ SUMMARY RESULTS ############
###########################################



# Zero inflated poisson mixed effects (zero inflated using the rurality and d
ates)
summary(lm2catziprelev)

##  Family: poisson  ( log )
## Formula:
## newcase_nst_14 ~ offset(popoff) + URBinary * c_daterelev + URBinary *
##     Date2 + URBinary * dsahcarried + URBinary * asahcarried +      (1 | c_
FIPS)
```

```
## Zero inflation:                  ~URBinary * c_daterelev
## Data: df_14
##
##       AIC       BIC    logLik  deviance  df.resid
## 1580358.4 1580567.6 -790160.2 1580320.4    446145
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  c_FIPS (Intercept) 1.526    1.236
## Number of obs: 446164, groups:  c_FIPS, 3142
##
## Conditional model:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -0.8388483  0.0300594   -27.9  < 2e-16 ***
## URBinary                        -1.5289178  0.0517319   -29.6  < 2e-16 ***
## c_daterelevafter SaH             0.8518576  0.0111095    76.7  < 2e-16 ***
## c_daterelevduring SaH            0.7121948  0.0099498    71.6  < 2e-16 ***
## Date2                            0.0377610  0.0001168   323.2  < 2e-16 ***
## dsahcarried                     -0.0059850  0.0001798   -33.3  < 2e-16 ***
## asahcarried                     -0.0165647  0.0001840   -90.0  < 2e-16 ***
## URBinary:c_daterelevafter SaH  -0.5690112  0.0200518   -28.4  < 2e-16 ***
## URBinary:c_daterelevduring SaH -0.4933506  0.0184905   -26.7  < 2e-16 ***
## URBinary:Date2                   0.0016374  0.0002762     5.9 3.07e-09 ***
## URBinary:dsahcarried            -0.0013485  0.0003407    -4.0 7.56e-05 ***
## URBinary:asahcarried            -0.0007427  0.0003652    -2.0    0.042 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Zero-inflation model:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -1.25889    0.02310  -54.50  < 2e-16 ***
## URBinary                         1.34931    0.03367   40.08  < 2e-16 ***
## c_daterelevafter SaH            -2.00159    0.03468  -57.71  < 2e-16 ***
## c_daterelevduring SaH           -0.18711    0.02623   -7.13 9.84e-13 ***
## URBinary:c_daterelevafter SaH  -1.47527    0.05563  -26.52  < 2e-16 ***
## URBinary:c_daterelevduring SaH -1.43643    0.03877  -37.05  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# ###########################################################################
##
# ############### DISPERSION, RESIDUALS, AND ZERO-INFLATION #################
##
# ###########################################################################
##
#
#
```

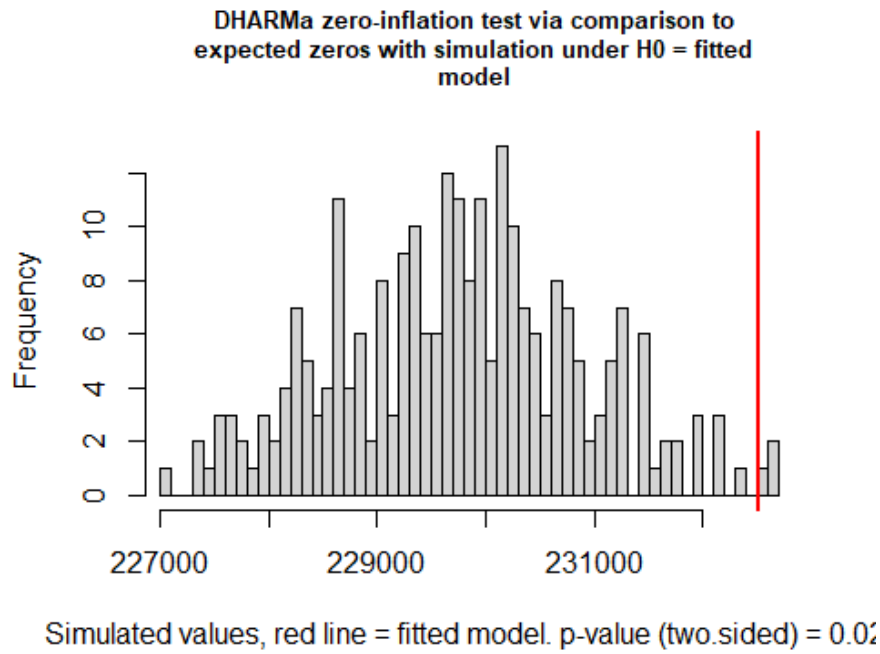

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

```
SimOut_lm2catziprelev <- simulateResiduals(fittedModel = lm2catziprelev, plot
= T)

## DHARMa:plot used testOutliers with type = binomial for computational reaso
ns (nObs > 500). Note that this method may not have inflated Type I error rat
es for integer-valued distributions. To get a more exact result, it is recomm
ended to re-run testOutliers with type = 'bootstrap'. See ?testOutliers for d
etails

plot(SimOut_lm2catziprelev)

## DHARMa:plot used testOutliers with type = binomial for computational reaso
ns (nObs > 500). Note that this method may not have inflated Type I error rat
es for integer-valued distributions. To get a more exact result, it is recomm
ended to re-run testOutliers with type = 'bootstrap'. See ?testOutliers for d
etails
```



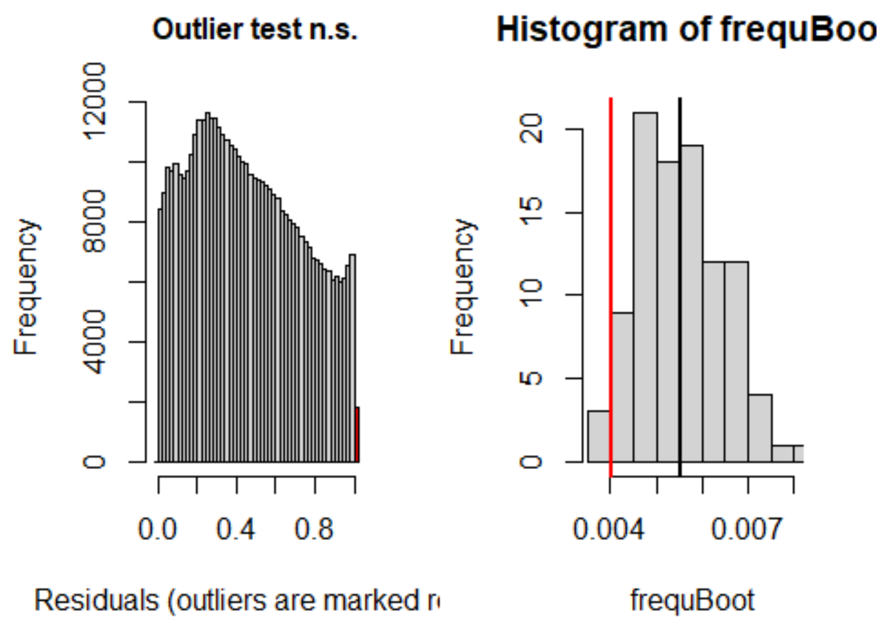

```
testZeroInflation(SimOut_lm2catziprelev)
```


Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791



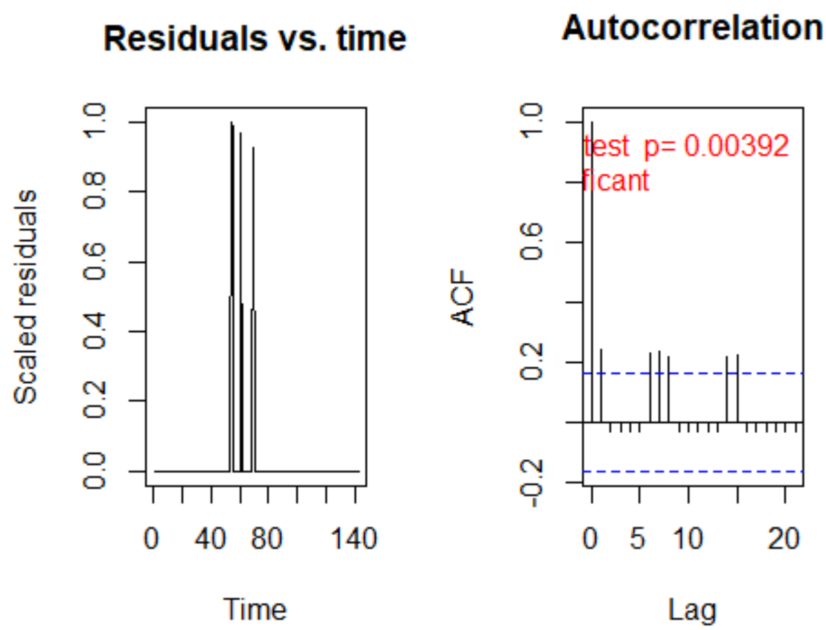

```
##
##  DHARMa zero-inflation test via comparison to expected zeros with
##  simulation under H0 = fitted model
##
## data:  simulationOutput
## ratioObsSim = 1.012, p-value = 0.024
## alternative hypothesis: two.sided

testOutliers(SimOut_lm2catziprelev, type= 'bootstrap')
```


Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791




```
##
##  DHARMa bootstrapped outlier test
##
## data:  SimOut_lm2catziprelev
## outliers at both margin(s) = 1802, observations = 446164, p-value =
## 0.06
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.004060951 0.007104675
## sample estimates:
## outlier frequency (expected: 0.00551676065303341 )
##                                          0.004038874
```

```
simoutrecalc <- recalculateResiduals(SimOut_lm2catziprelev, group = df_14$Dat
e2)
testTemporalAutocorrelation(simoutrecalc, time = unique(df_14$Date2))
```




```
##
##  Durbin-Watson test
##
## data:  simulationOutput$scaledResiduals ~ 1
## DW = 1.5192, p-value = 0.003916
## alternative hypothesis: true autocorrelation is not 0
```

Mobility Data Analysis

Community Mobility Reports from Google Inc[4] were used to examine county-level mobility trends .The data shows movement trends by individuals within U.S. counties across several categories of places as well as the percent change of movement relative to a baseline period. According to Google, "The data shows how visitors to (or time spent in) categorized places change compared to our baseline days. A baseline day represents a *normal* value for that day of the week. The baseline day is the median value from the 5-week period Jan 3 – Feb 6, 2020." The categories of places include grocery & pharmacy, parks, transit stations, retail & recreation, residential, and workplaces. However, due to the fact that not every county reports parks and transit stations, those were not included in our analysis.

Google did not report a change in baseline for every county for every day. However, since the measured outcome is the change from baseline for each individual county relative to itself, we were able to average the percent changes across county types (i.e. rural and urban counties). For each day, the numbers of counties included in the analysis each day by county type are shown in the table below. There are a total of 1,976 rural and 1,166 urban counties in the United States.

Date	Rural Counties	Urban Counties
2/15/20	1450	1111
2/16/20	1355	1098
2/17/20	1594	1150
2/18/20	1577	1146
2/19/20	1583	1146
2/20/20	1580	1147
2/21/20	1567	1141
2/22/20	1449	1113
2/23/20	1352	1096
2/24/20	1564	1142
2/25/20	1579	1145
2/26/20	1574	1146
2/27/20	1573	1146
2/28/20	1562	1141
2/29/20	1442	1109
3/1/20	1332	1090
3/2/20	1559	1142
3/3/20	1572	1146
3/4/20	1576	1146
3/5/20	1572	1146
3/6/20	1566	1142
3/7/20	1437	1109
3/8/20	1337	1092
3/9/20	1560	1143
3/10/20	1570	1146
3/11/20	1577	1147
3/12/20	1575	1146
3/13/20	1568	1142
3/14/20	1442	1109
3/15/20	1338	1090
3/16/20	1575	1146
3/17/20	1603	1151
3/18/20	1612	1151
3/19/20	1611	1152
3/20/20	1607	1150
3/21/20	1460	1110
3/22/20	1375	1102
3/23/20	1612	1151
3/24/20	1623	1152
3/25/20	1625	1152
3/26/20	1629	1152
3/27/20	1609	1150

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

3/28/20	1470	1116
3/29/20	1387	1102
3/30/20	1618	1152
3/31/20	1632	1152
4/1/20	1630	1152
4/2/20	1637	1152
4/3/20	1617	1151
4/4/20	1471	1115
4/5/20	1395	1102
4/6/20	1607	1152
4/7/20	1628	1152
4/8/20	1627	1152
4/9/20	1628	1153
4/10/20	1606	1152
4/11/20	1177	1075
4/12/20	1124	1061
4/13/20	1609	1152
4/14/20	1621	1152
4/15/20	1624	1152
4/16/20	1620	1152
4/17/20	1587	1151
4/18/20	1165	1072
4/19/20	1087	1054
4/20/20	1604	1152
4/21/20	1616	1152
4/22/20	1622	1152
4/23/20	1619	1152
4/24/20	1586	1151
4/25/20	1156	1072
4/26/20	1078	1051
4/27/20	1597	1151
4/28/20	1616	1152
4/29/20	1621	1152
4/30/20	1613	1152
5/1/20	1581	1151
5/2/20	1137	1069
5/3/20	1067	1046
5/4/20	1593	1152
5/5/20	1613	1152
5/6/20	1610	1152
5/7/20	1607	1152
5/8/20	1579	1151
5/9/20	1139	1067
5/10/20	1055	1044
5/11/20	1589	1151
5/12/20	1611	1152
5/13/20	1606	1152
5/14/20	1603	1152
5/15/20	1574	1150
5/16/20	1135	1067
5/17/20	1058	1042
5/18/20	1581	1152
5/19/20	1605	1152
5/20/20	1608	1152
5/21/20	1602	1152
5/22/20	1574	1151
5/23/20	1126	1062
5/24/20	1052	1042
5/25/20	1610	1146
5/26/20	1607	1151
5/27/20	1605	1152
5/28/20	1597	1152
5/29/20	1567	1149
5/30/20	1111	1054
5/31/20	1015	1028
6/1/20	1577	1152
6/2/20	1604	1152
6/3/20	1602	1152
6/4/20	1596	1152
6/5/20	1567	1148
6/6/20	1112	1055
6/7/20	1020	1036
6/8/20	1580	1151
6/9/20	1605	1152
6/10/20	1599	1152
6/11/20	1597	1152
6/12/20	1571	1148
6/13/20	1102	1055

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

6/14/20	1012	1031

```
#information on the number of counties on which data was collected
counties <- NA
for (i in 1:130) {
  counties[i] <- sum(mob$X_FREQ_[mob$date2==i])
}
#Urban and rural combined
mean(counties)
```

```
## [1] 2615.846
```

```
median(counties)
```

```
## [1] 2729
```

```
#UR separated
mean(mob$X_FREQ_[mob$URBinary=="Urban"])
```

```
## [1] 1127.9
```

```
median(mob$X_FREQ_[mob$URBinary=="Rural"])
```

```
## [1] 1579.5
```

```
mean(mob$X_FREQ_[mob$URBinary=="Urban"])
```

```
## [1] 1127.9
```

```
median(mob$X_FREQ_[mob$URBinary=="Rural"])
```

```
## [1] 1579.5
```

Repeated measures ANOVA analysis was performed on the Google mobility data using the rstatix package.[5] The dependent variable was the mean % change from baseline mobility on a given day (mean of the counties with data on a given day). The "subjects" were the individual days and the "within-subject factor" was the county type (urban or rural). This approach was chosen because each outcome is the change from baseline (each county acts as its own control and null hypothesis that all change equally) and thus minimizes the bias of treating outcomes of rural and urban counties on the same day as independent.

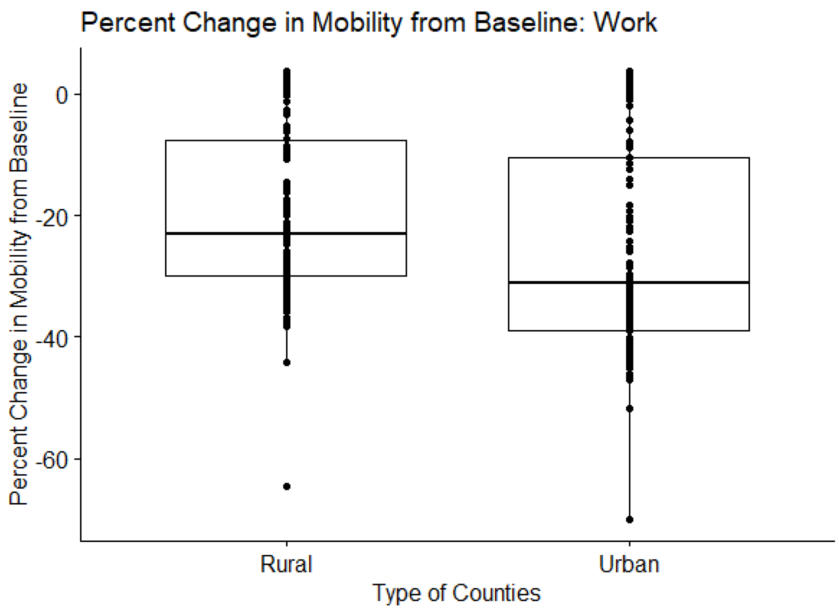
Each category of mobility data were tested for outliers and normality. The anova_test function of the rstatix package tests for sphericity and automatically applies the Greenhouse-Geisser sphericity correction.

Outliers were classified as observations outside of 1.5 times the interquartile range (IQR) of their respective distribution (mobility type and rurality). Grocery/pharmacy and workplace were the

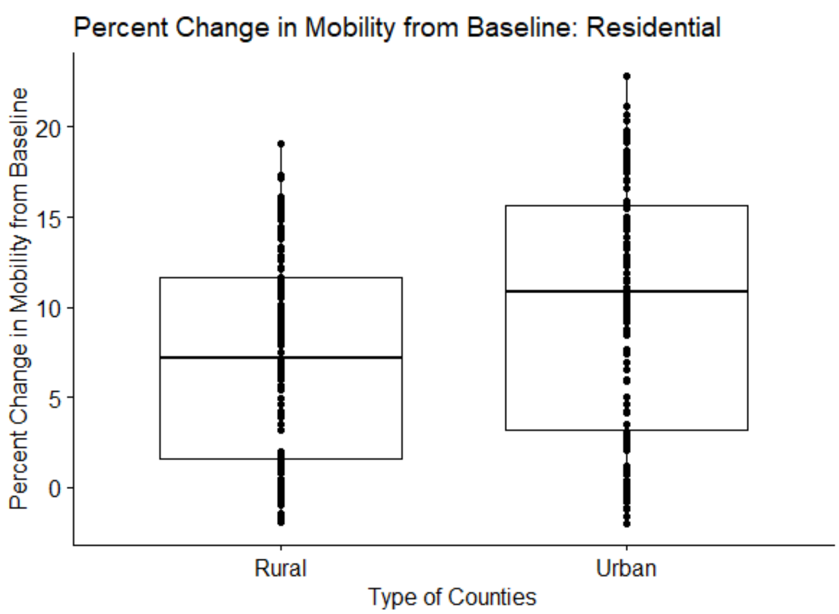
Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

only categories with outliers, with 8 outliers (4 days) and 2 outliers (1 day) removed for these categories, respectively. Below are the boxplots of every mobility category by type of county.

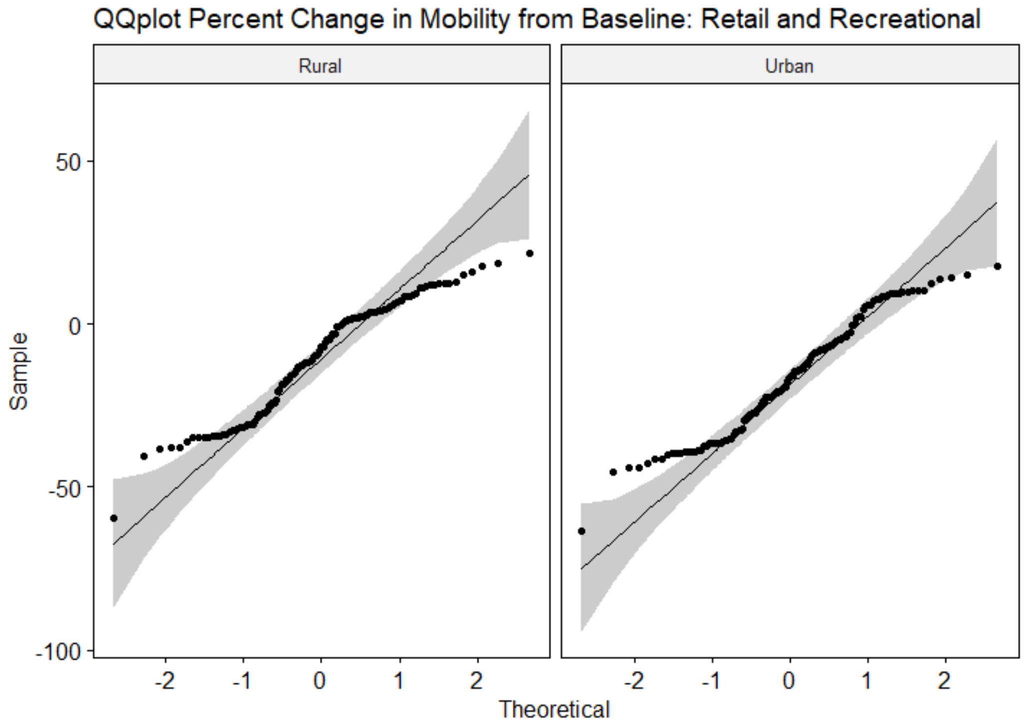
Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

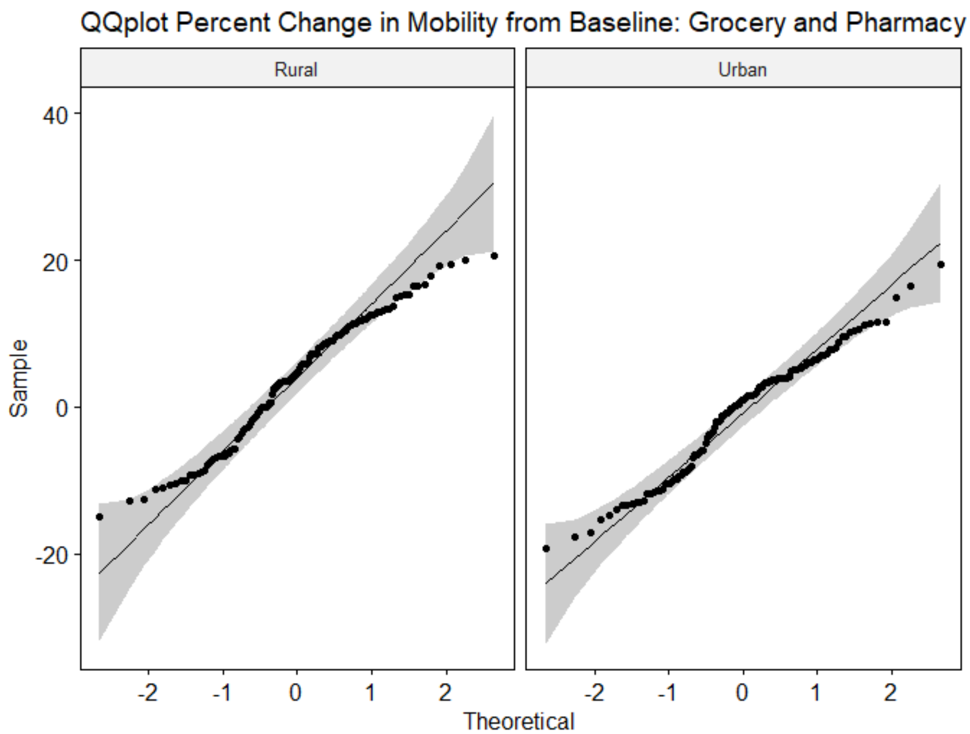


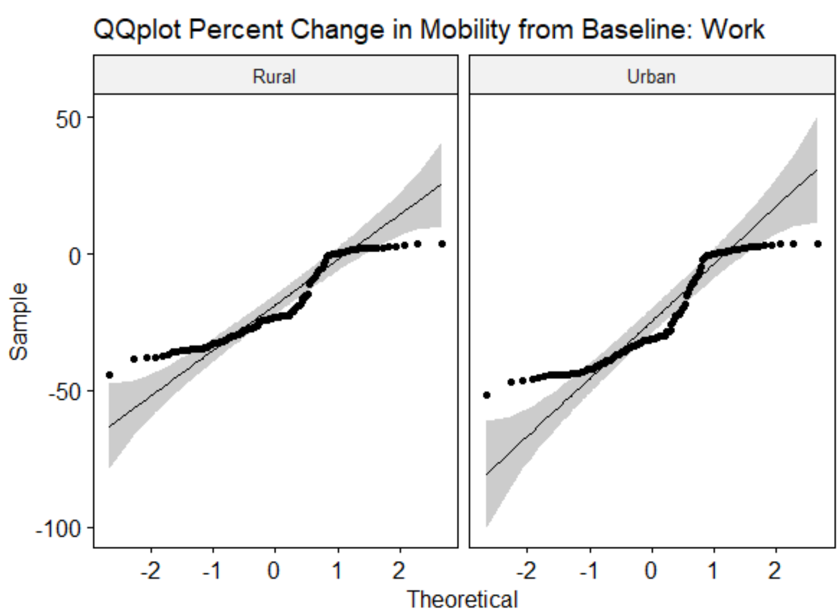

```
out3 <- boxplot(mob$work ~ mob$URBinary )$out
```

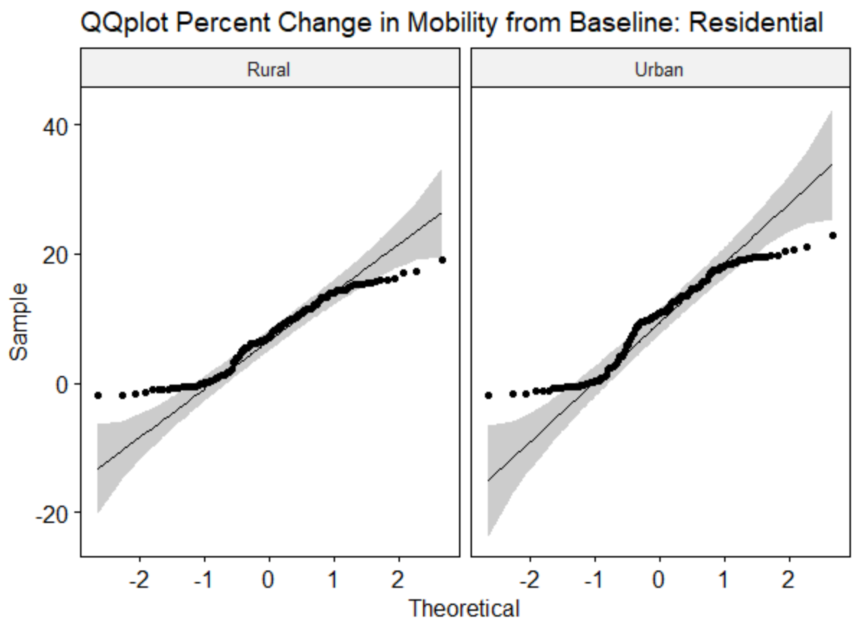


The assumption of normality in the case of this analysis is a given based on the large number of observations, 130 days for each of the mobility types. To ensure that this was not incorrectly assumed normality was assessed by county type and mobility type using QQ-plots. Based on these QQ-plots, residential and work seem to not be perfectly normally distributed, but there are a large number of observations (n>50) thus alleviating this concern. Below are the QQ-plots.




QQplot Percent Change in Mobility from Baseline: Grocery and Pharmacy
Rural
Urban
Sample
Theoretical
40
20
0
-20
-2
-1
0
1
2


QQplot Percent Change in Mobility from Baseline: Work
Rural
Urban
Sample
Theoretical
50
0
-50
-100
-2
-1
0
1
2




```
ret.aov <- anova_test(data = mob_ret, dv = retail_rec, wid =date2 , within =
URBinary)
```

```
## ANOVA Table (type III tests) Retail and Recreation
##
##     Effect DFn DFd       F        p p<.05   ges
## 1 URBinary   1 129 415.405 3.71e-42     * 0.038
```

```
groc.aov <- anova_test(data = mob_groc, dv = groc_pha, wid =date2 , within =
URBinary)
```

```
## ANOVA Table (type III tests) Grocery and Pharmacy
##
##     Effect DFn DFd       F        p p<.05   ges
## 1 URBinary   1 125 317.158 4.28e-36     * 0.072
```

```
work.aov <- anova_test(data = mob_work, dv = work, wid =date2 , within =
URBinary)
```

```
## ANOVA Table (type III tests) Work
##
##     Effect DFn DFd       F       p p<.05   ges
## 1 URBinary   1 128 340.928 6.7e-38     * 0.035
```

```
res.aov <- anova_test(data = mob_res, dv = residential, wid =date2 , within =
URBinary)
get_anova_table(res.aov)
```


Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

```
## ANOVA Table (type III tests) Residential
##
##     Effect DFn DFd       F        p p<.05   ges
## 1 URBinary   1 129 381.282 2.44e-40     * 0.042
```

All of the repeated measures ANOVA tests resulted in a significant p-value indicating that all of the percentage change in mobility from baseline categories are statistically significantly different between Rural and Urban counties.

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

Stay-at-Home Orders Start and End Dates

Individual state governments started stay-at-home at different times and ended at different times, ascertained by review of each state's executive order by the study team. Four states (Arkansas, Iowa, North Dakota, and South Dakota) did not issue stay at home orders. Three others (Oklahoma, Utah, and Wyoming) allowed the county and local governments to make such determinations. The following table displays the start and end dates of statewide stay-at-home orders, while the subsequent table displays that of locales.

State	Start	End
Alabama	4/4/20	4/30/20
Alaska	3/28/20	4/24/20
Arizona	3/31/20	5/15/20
Arkansas	Did Not Issue SAH	
California	3/19/20	Ongoing
Colorado	3/26/20	4/26/20
Connecticut	3/23/20	5/20/20
Delaware	3/24/20	5/31/20
District of Columbia	4/1/20	5/29/20
Florida	4/3/20	5/4/20
Georgia	4/3/20	4/30/20
Hawaii	3/25/20	5/31/20
Idaho	3/25/20	4/30/20
Illinois	3/21/20	5/29/20
Indiana	3/24/20	5/4/20
Iowa	Did Not Issue SAH	
Kansas	3/30/20	5/3/20
Kentucky	3/26/20	Ongoing
Louisiana	3/23/20	5/15/20
Maine	4/2/20	5/31/20
Maryland	3/30/20	5/15/20
Massachusetts	3/24/20	5/18/20
Michigan	3/24/20	6/1/20
Minnesota	3/27/20	5/13/20
Mississippi	4/3/20	4/27/20
Missouri	4/6/20	5/3/20
Montana	3/28/20	4/26/20
Nebraska	Did Not Issue SAH	
Nevada	4/1/20	4/29/20
New Hampshire	3/27/20	Ongoing
New Jersey	3/21/20	6/9/20
New Mexico	3/24/20	5/31/20
New York	3/22/20	5/28/20
North Carolina	3/30/20	5/22/20
North Dakota	Did Not Issue SAH	
Ohio	3/23/20	5/29/20
Oklahoma	Local Decision	
Oregon	3/23/20	Ongoing
Pennsylvania	4/1/20	6/4/20
Rhode Island	3/28/20	5/8/20
South Carolina	4/7/20	5/4/20
South Dakota	Did Not Issue SAH	
Tennessee	3/31/20	4/30/20
Texas	4/2/20	4/30/20
Utah	Local Decision	
Vermont	3/25/20	5/10/20
Virginia	3/30/20	6/10/20
Washington	3/23/20	5/31/20
West Virginia	3/24/20	5/3/20
Wisconsin	3/25/20	5/13/20
Wyoming	Local Decision	

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791

County	State	FIPS	Start	End
Carter County	OK	40019	4/6/20	4/24/20
Rogers County	OK	40131	4/6/20	4/24/20
Cleveland County	OK	40027	3/25/20	4/24/20
Seqouyah County	OK	40135	4/4/20	4/24/20
Payne County	OK	40119	3/30/20	4/24/20
Tulsa County	OK	40143	3/28/20	4/24/20
Oklahoma County	OK	40109	3/28/20	4/24/20
Davis County	UT	49011	4/1/20	5/1/20
Salt Lake County	UT	49035	3/30/20	5/1/20
Summit County	UT	49043	3/27/20	5/1/20
Teton County	WY	56039	3/28/20	5/1/20

References

1. Brooks ME, Kristensen K, Van Benthem KJ, et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal 2017;9:378-400.
2. Hartig F, Lohse L. Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. CRAN 2020.
3. Mitchell DJ, Dujon AM, Beckmann C, Biro PA. Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity. Behavioral Ecology 2019;31:222-31.
4. Google COVID-19 Community Mobility Reports. Online: Google LLC; 2020.
5. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. CRAN 2020.

Jiang DH, *et al. BMJ Open* 2022; 12:e055791. doi: 10.1136/bmjopen-2021-055791