Model the relation between oil/gas production and annual Lead-210 measurement

Longxiang Li

2018-10-25

Contents

Introduction

In this report, we'll try to explore the correlation between Lead-210 with oil/gas drilling.

Background

- Long-term exposure to low-level radon is dangerous. Because radon is carcinogenic.

- Direct long-term measurement of radon is rare.
- Lead-210 has been used as marker for radon exposure. This is applied in multiple studies.
- If there's an increase of Lead-210 in the particles, local residents are exposed to higher risk of lung cancer and other disease.

Data

- Study period: from 2014 to 2016. Because we don't have the annual Lead-210 measuremtn before 2014 and after 2016. I've reached out to RadNet for data.
- Study region: Lower 48 states of the United States. There're 139 RedNet monitors. But only a small fraction of these monitors have gas/oil activities nearby.
- Data sources:

Lead-210 data: Lead-210 are determined by the analysis of annually composited samples (air filters) collected from the airborne particulate samplers. Concentrations are determined by alpha-particle spectrometry following chemical separation. The total volume of air represented by all the samples received from one sampling location during a year typically ranges from 120,000 m^3 to 500,000 m^3.

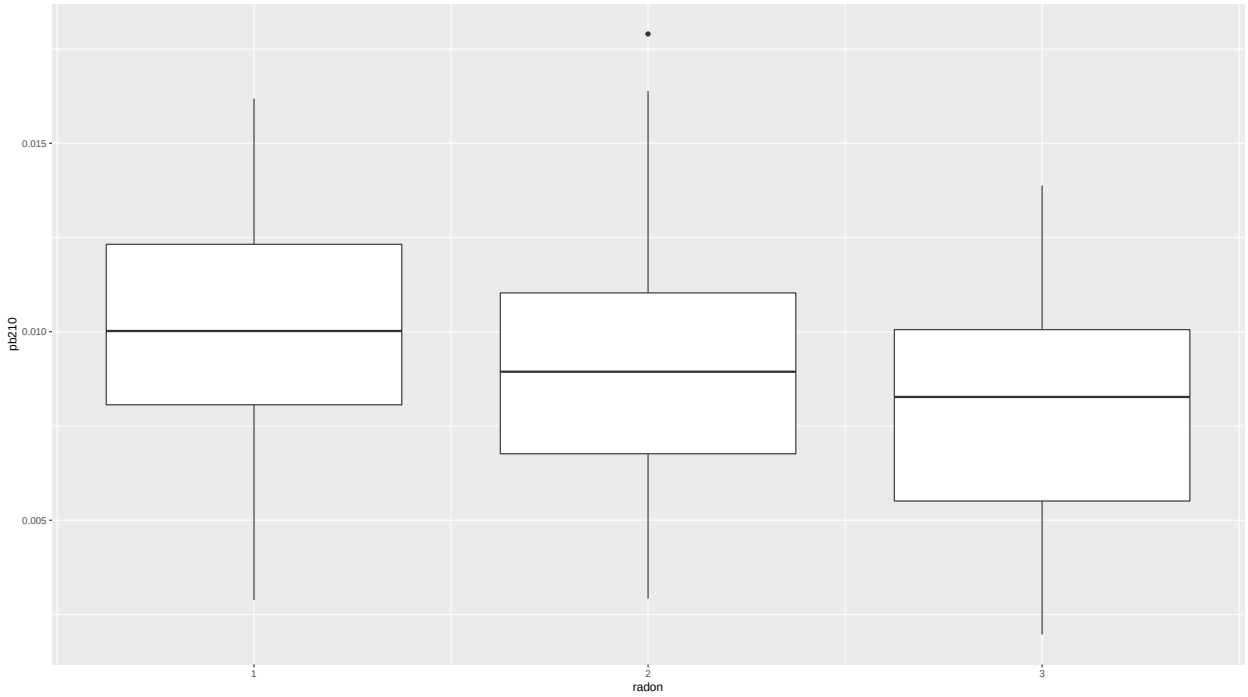
Drilling data: We colleced drilling information from drillinginfo.com. From the database, we extracted the monthly gas/oil production during the study period, both horizontal and vertical drillings are included. Since the production data is reported by month, we need to aggregate them by year. Then we have annual gas/oil production, number of active gas/oil wells, categorical and uncategorical within a radius of 25km from the assumed location of RadNet monitors.

Background radon data: We downloaded the EPA radon map data and join it to the assumed location of RadNet monitor. All counties of the U,S are catergorized into three classes ranging from 1 with the highest radon level and 3 with the lowest. The background radon level is calculated based on soil type, weather and other information.

In this report, we first set the radius as 25km. If there's any oil/gas production within this circle in the study period, this RadNet monitor is categorized as within gas/oil field. Otherwise, this RadNet monitor is categorized as clean ones.

Descriptive statistic of the data

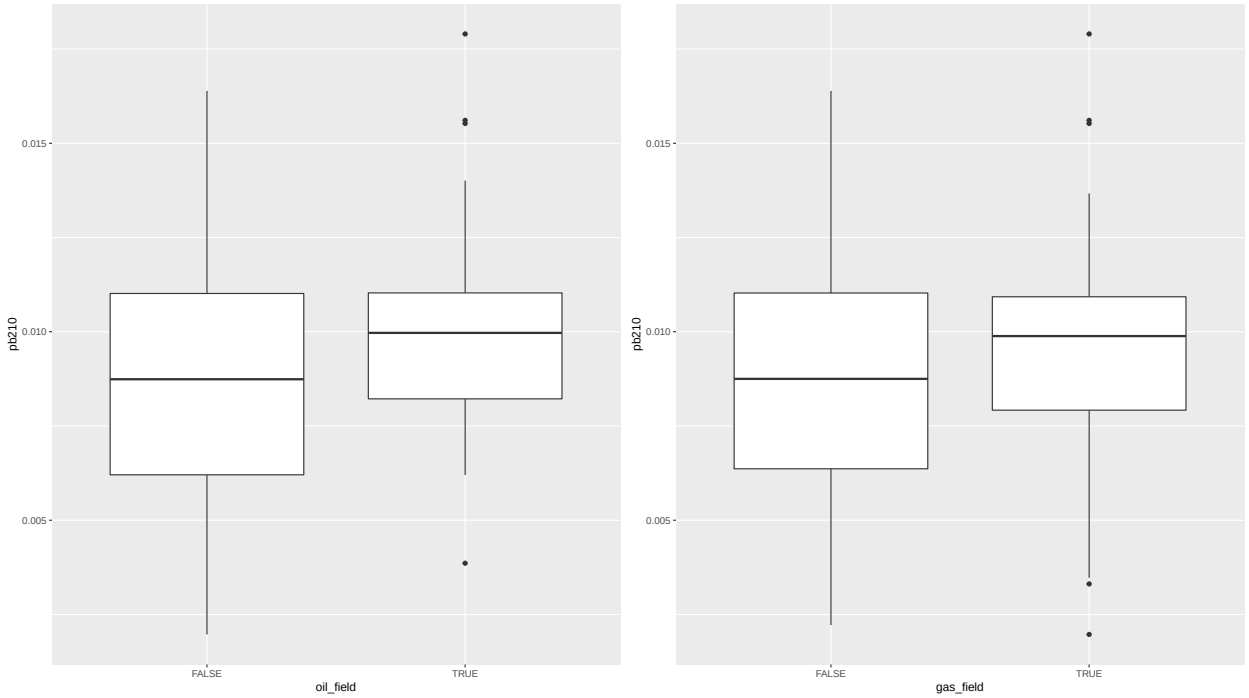
- Annual Pb210 is negatively related with EPA radon zone. Here we treat the ordered categorical radon zone as continuous.



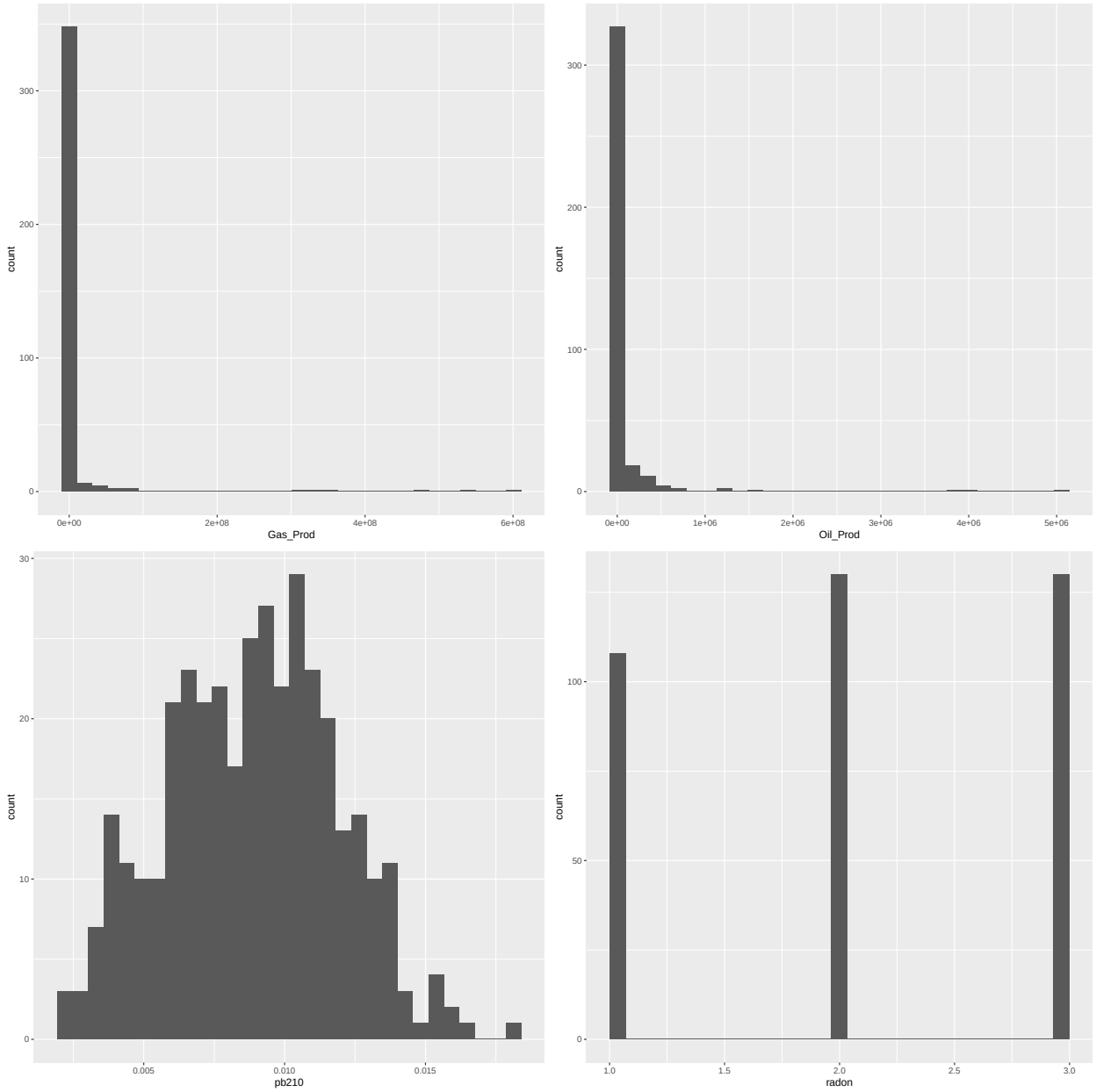
- Even though oil fields has lower radon level, but they have higher Pb210 level. This relation also applies to the natural gas drilling.

```
## # A tibble: 2 x 2
##   oil_field mean_radon
##   <lgl>          <dbl>
## 1 FALSE           1.97
## 2 TRUE            2.44
```

```
## # A tibble: 2 x 2
##   gas_field mean_radon
##   <lgl>          <dbl>
## 1 FALSE           1.93
## 2 TRUE            2.53
```



- The oil/gas data is skewed. But the pb210 measurement is almost normal. Radon zone is almost evenly distributed.



Models of 25km cases.

Mixed effects models are used in this report to model the correlation between our variable of interest and the Lead-210. Our variables of interest are always set as fixed effect while random intercepts are assigned to each RadNet monitor and basin. In addition, radon zone is also set as fixed effect. To check the significance of our fixed effect, a bootstrp confidence interval is calculated. In addition, a likelihood-ratio test is also applied here.

Gross Oil Production

Oil production is the sum of monthly oil production from all wells within 25km away from the monitor. Based on the summary of models and test, we can see that, gross oil production is not significantly correlated with the annual Lead-210. Adding oil production doesn't influence the slope of radon remarkably.

```
model_basic_25<-lmer(pb210~radon+Umeans+Thmeans+(1|city_state)+(1|YEAR:basin_name),data=rad_all,REML=T)
model_oil_prod_25<-lmer(pb210~radon+Umeans+Thmeans+Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=rad
confint(model_oil_prod_25,"Oil_Prod",method="boot")
```

```
##                   2.5 %       97.5 %
## Oil_Prod -1.844732e-11 1.940734e-09
```

```
anova(model_basic_25,model_oil_prod_25)
```

```
## Data: rad_all
## Models:
## model_basic_25: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_oil_prod_25: pb210 ~ radon + Umeans + Thmeans + Oil_Prod + (1 | city_state) +
## model_oil_prod_25:     (1 | YEAR:basin_name)
##                   Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_25     7 -3451.8 -3424.5 1732.9  -3465.8
## model_oil_prod_25  8 -3453.6 -3422.3 1734.8  -3469.6 3.7488      1
##                   Pr(>Chisq)
## model_basic_25
## model_oil_prod_25    0.05285 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fixed.effects(model_oil_prod_25)
```

```
##   (Intercept)         radon        Umeans       Thmeans      Oil_Prod
##  7.009526e-03 -4.843512e-04 -1.355834e-04  5.827870e-04  9.655162e-10
```

Gross Gas Production

Gas production is the sum of monthly gas production from all wells within 25km away from the monitor. Based on the summary of models and test, we can see that, gross gas production is not significantly correlated with the annual Lead-210.

```
model_gas_prod_25<-lmer(pb210~radon+Umeans+Thmeans+Gas_Prod+(1|city_state)+(1|YEAR:basin_name),data=rad
confint(model_gas_prod_25,"Gas_Prod",method="boot")
```

```
##                   2.5 %       97.5 %
## Gas_Prod -9.561232e-12 6.695407e-12
```

```
anova(model_basic_25,model_gas_prod_25)
```

```
## Data: rad_all
## Models:
## model_basic_25: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_gas_prod_25: pb210 ~ radon + Umeans + Thmeans + Gas_Prod + (1 | city_state) +
## model_gas_prod_25:     (1 | YEAR:basin_name)
##                   Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_25     7 -3451.8 -3424.5 1732.9  -3465.8
## model_gas_prod_25  8 -3449.9 -3418.6 1733.0  -3465.9 0.0793      1
##                   Pr(>Chisq)
```

```
## model_basic_25
## model_gas_prod_25     0.7782
```

Horizontal Oil Production

Horizontal oil production is the sum of monthly oil production from all horizontal wells within 25km away from the monitor. Based on the summary of models and test, we can see that, horizontal oil production is not significantly correlated with the annual Lead-210.

```
model_h_oil_prod_25<-lmer(pb210~radon+Umeans+Thmeans+H_Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=
confint(model_h_oil_prod_25,"H_Oil_Prod",method="boot")
```

```
##                    2.5 %      97.5 %
## H_Oil_Prod -7.329892e-10 1.94944e-09
```

```
anova(model_basic_25,model_h_oil_prod_25)
```

```
## Data: rad_all
## Models:
## model_basic_25: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_h_oil_prod_25: pb210 ~ radon + Umeans + Thmeans + H_Oil_Prod + (1 | city_state) +
## model_h_oil_prod_25:     (1 | YEAR:basin_name)
##                     Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_25       7 -3451.8 -3424.5 1732.9  -3465.8
## model_h_oil_prod_25  8 -3450.9 -3419.6 1733.4  -3466.9 1.0309      1
##                     Pr(>Chisq)
## model_basic_25
## model_h_oil_prod_25     0.3099
```

```
fixed.effects(model_h_oil_prod_25)
```

```
##   (Intercept)         radon        Umeans       Thmeans    H_Oil_Prod
##  7.096072e-03 -4.605541e-04 -1.709259e-04  5.874603e-04  6.646327e-10
```

Vertical Oil Production

Vertical oil production is the sum of monthly oil production from all vertical wells within 25km away from the monitor. Based on the summary of models and test, we can see that, vertical oil production is significantly correlated with the annual Lead-210. Adding vertical oil production doesn't influence the slope of radon and intercept remarkably.

```
model_v_oil_prod_25<-lmer(pb210~radon+Umeans+Thmeans+V_Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=
confint(model_v_oil_prod_25,method="boot")
```

```
##                     2.5 %        97.5 %
## .sig01       1.852464e-03  2.497553e-03
## .sig02       4.232665e-04  1.380802e-03
## .sigma       1.341394e-03  1.622529e-03
## (Intercept)  4.926283e-03  9.093858e-03
## radon       -1.127722e-03 -2.586210e-05
## Umeans      -1.312252e-03  1.183292e-03
## Thmeans      1.862754e-04  9.205807e-04
## V_Oil_Prod   1.609282e-09  6.650871e-09
```

```
anova(model_basic_25,model_v_oil_prod_25)
```

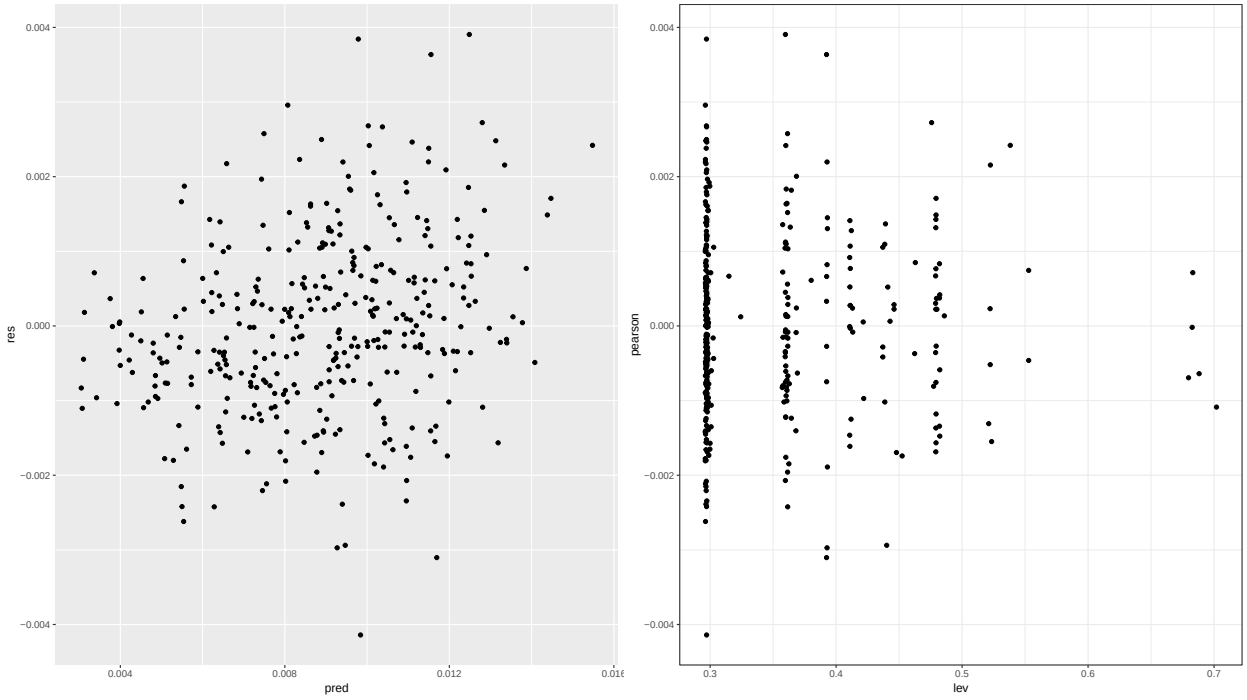
```
## Data: rad_all
## Models:
## model_basic_25: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_v_oil_prod_25: pb210 ~ radon + Umeans + Thmeans + V_Oil_Prod + (1 | city_state) +
## model_v_oil_prod_25:     (1 | YEAR:basin_name)
##                      Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_25        7 -3451.8 -3424.5 1732.9  -3465.8
## model_v_oil_prod_25   8 -3459.3 -3428.1 1737.7  -3475.3 9.4872      1
##                      Pr(>Chisq)
## model_basic_25
## model_v_oil_prod_25   0.002069 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fixed.effects(model_v_oil_prod_25)
```

```
##   (Intercept)         radon        Umeans       Thmeans    V_Oil_Prod
##  7.057864e-03 -5.749510e-04 -2.165374e-05  5.402743e-04  4.067843e-09
```

A tentative diagnostic based largely on leverage is applied here to check whether this correlation is stable. Otherwise, it can be influenced by few influential measurements. Based on the qqnorm and leverage plot, we can see that the residual of this model is largely normally distributed except for some limit values. After removing the measurements with very big leverage value, the slope of vertical oil production doesn't change remarkbly, only 5% of the standard deviation. The updated confidence interval after removing these measurements still doesn't cover 0, meansing this correlation is stable and significant.

```
g1<-ggplot(data=cbind.data.frame(res=resid(model_v_oil_prod_25),pred=fitted(model_v_oil_prod_25)),aes(y
g2<-ggplot(data.frame(lev=hatvalues(model_v_oil_prod_25),pearson=residuals(model_v_oil_prod_25,type="pea
           aes(x=lev,y=pearson)) +
  geom_point() +
  theme_bw()
cowplot::plot_grid(g1,g2)
```



```
levId <- which(hatvalues(model_v_oil_prod_25) >= .5)
rad_all[levId,c("pb210","radon","Umeans","Thmeans","V_Oil_Prod","YEAR","city_state")]
```

```
##       pb210 radon   Umeans  Thmeans V_Oil_Prod YEAR      city_state
## 24  0.01790     2 2.652282 7.674246     460545 2015  BAKERSFIELD,CA
## 25  0.01010     2 2.652282 7.674246     235449 2016  BAKERSFIELD,CA
## 35  0.01040     2 1.905638 5.382776          0 2014   BIRMINGHAM,AL
## 36  0.00872     2 1.905638 5.382776          0 2016   BIRMINGHAM,AL
## 117 0.00723     2 1.165853 4.227986          0 2016   ELLENSBURG,WA
## 121 0.01136     1 1.955704 6.656968          0 2016 FORT MADISON,IA
## 131 0.01550     2 2.109550 7.260149          0 2015       FRESNO,CA
## 132 0.00912     2 2.109550 7.260149          0 2016       FRESNO,CA
## 200 0.00597     2 1.647297 5.674589          0 2014    LYNCHBURG,VA
## 215 0.00606     3 1.140704 3.627871          0 2015       MOBILE,AL
## 216 0.00708     3 1.140704 3.627871          0 2016       MOBILE,AL
## 334 0.00480     1 1.336446 5.017595          0 2014     SYRACUSE,NY
## 368 0.00717     2 2.782820 7.727965          0 2016         YUMA,AZ
```

```
model_v_oil_diag_25 <- lmer(pb210 ~ radon+Umeans+Thmeans + V_Oil_Prod +(1|city_state)+(1|YEAR:basin_name
LevCD <- data.frame(effect=fixef(model_v_oil_prod_25),
                    change=(fixef(model_v_oil_diag_25) - fixef(model_v_oil_prod_25)),
                    se=sqrt(diag(vcov(model_v_oil_prod_25)))
)
rownames(LevCD) <- names(fixef(model_v_oil_diag_25))
LevCD$multiples <- abs(LevCD$change / LevCD$se)
LevCD
```

```
##                    effect        change           se  multiples
## (Intercept)  7.057864e-03  1.694633e-04 1.144064e-03 0.14812398
## radon       -5.749510e-04 -1.915668e-05 2.836242e-04 0.06754249
## Umeans      -2.165374e-05 -1.013448e-04 6.508817e-04 0.15570382
## Thmeans      5.402743e-04 -5.268175e-06 1.827533e-04 0.02882670
## V_Oil_Prod   4.067843e-09 -5.127609e-10 1.319930e-09 0.38847591
```

```
confint(model_v_oil_diag_25)
```

```
##                     2.5 %        97.5 %
## .sig01       1.839189e-03  2.486245e-03
## .sig02       2.986511e-04  1.091887e-03
## .sigma       1.351632e-03  1.634280e-03
## (Intercept)  4.930009e-03  9.513928e-03
## radon       -1.158884e-03 -3.085173e-05
## Umeans      -1.433005e-03  1.194152e-03
## Thmeans      1.737111e-04  8.941937e-04
## V_Oil_Prod   9.480451e-10  6.170906e-09
```

Horizontal Gas Production

Horizontal gas production is the sum of monthly gas production from all horizontal wells within 25km away from the monitor. Based on the summary of models and test, we can see that, without log-transformation, horizontal gas production is weakly correlated with the annual Lead-210. After log-transformation, the horizontal gas production is not significantlyrelated with log(Pb210).

```
model_h_gas_prod_25<-lmer(pb210~radon+Umeans+Thmeans+H_Gas_Prod+(1|city_state)+(1|YEAR:basin_name),data=
confint(model_h_gas_prod_25,method="boot")
```

```
##                     2.5 %       97.5 %
## .sig01       1.850596e-03 2.516623e-03
## .sig02       4.963434e-04 1.437139e-03
## .sigma       1.344437e-03 1.611026e-03
## (Intercept)  5.105486e-03 9.696592e-03
## radon       -1.088973e-03 8.229929e-05
## Umeans      -1.450205e-03 1.079222e-03
## Thmeans      1.794201e-04 9.312229e-04
## H_Gas_Prod  -6.250897e-12 9.610615e-12
```

```
anova(model_basic_25,model_h_gas_prod_25)
```

```
## Data: rad_all
## Models:
## model_basic_25: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_h_gas_prod_25: pb210 ~ radon + Umeans + Thmeans + H_Gas_Prod + (1 | city_state) +
## model_h_gas_prod_25:     (1 | YEAR:basin_name)
##                     Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_25       7 -3451.8 -3424.5 1732.9  -3465.8
## model_h_gas_prod_25  8 -3450.0 -3418.7 1733.0  -3466.0 0.1306      1
##                     Pr(>Chisq)
## model_basic_25
## model_h_gas_prod_25     0.7178
```

Vertical Gas Production

Vertical gas production is the sum of monthly gas production from all vertical wells within 25km away from the monitor. Based on the summary of models and test, we can see that, without log-transformation, vertical gas production is weakly correlated with the annual Lead-210. After log-transformation, the vertical gas production is not significantly related with log(Pb210).

```
model_v_gas_prod_25<-lmer(pb210~radon+Umeans+Thmeans+V_Gas_Prod+(1|city_state)+(1|YEAR:basin_name),data=
confint(model_v_gas_prod_25,method="boot")
```

```
##                     2.5 %       97.5 %
## .sig01       1.840665e-03 2.563349e-03
## .sig02       4.921960e-04 1.462449e-03
## .sigma       1.334802e-03 1.622075e-03
## (Intercept)  4.882347e-03 9.668691e-03
## radon       -1.042078e-03 1.069403e-04
## Umeans      -1.517779e-03 1.144900e-03
## Thmeans      2.447819e-04 9.879175e-04
## V_Gas_Prod  -2.800948e-11 6.381774e-12
```

```
anova(model_basic_25,model_v_gas_prod_25)
```

```
## Data: rad_all
## Models:
## model_basic_25: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_v_gas_prod_25: pb210 ~ radon + Umeans + Thmeans + V_Gas_Prod + (1 | city_state) +
## model_v_gas_prod_25:     (1 | YEAR:basin_name)
##                     Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_25       7 -3451.8 -3424.5 1732.9  -3465.8
## model_v_gas_prod_25  8 -3451.5 -3420.3 1733.8  -3467.5 1.6909      1
##                     Pr(>Chisq)
```

```
## model_basic_25
## model_v_gas_prod_25     0.1935
```

Partial Conclusion

Oil production especially oil production from vertical wells is significantly related with local Lead-210. Natural gas production is not significantly correlated with local local Lead-210. So, vertical oil drilling may significantly increase the local residents' exposure to radon.

Tentative interperation

Vertical wells were mostly completed before financial crisis. They're much older than the currently dorminant directional drilling. At the end of lifttime, the fraction of produced water is always higher than the new drill. Produced water (is not pumped back to the formation) may serve as the medium for radon leakage.

Models of 50km cases

After running the models based on aggregated gas/oil production within 25km buffer. We enlarge the radius to 50km and run similar models. It's expected that the slope is still significant but the value is a little smaller.

Gross Oil Production

After enlarging the buffer to 50km. The main effect of gross oil production is not significant with or withour log-trasformation.

```
model_basic_50<-lmer(pb210~radon+Umeans+Thmeans+(1|city_state)+(1|YEAR:basin_name),data=rad_all,REML=T)
fixed.effects(model_basic_50)
```

```
##   (Intercept)         radon        Umeans       Thmeans
##  0.0072062940 -0.0004652606 -0.0001704985  0.0005779429
```

```
model_oil_prod_50<-lmer(pb210~radon+Umeans+Thmeans+Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=rad
confint(model_oil_prod_50,method="boot")
```

```
##                    2.5 %       97.5 %
## .sig01       1.886054e-03 2.532011e-03
## .sig02       5.458580e-04 1.409958e-03
## .sigma       1.334260e-03 1.624848e-03
## (Intercept)  4.445030e-03 9.380708e-03
## radon       -1.036161e-03 1.181936e-04
## Umeans      -1.350929e-03 1.375000e-03
## Thmeans      1.954639e-04 9.143048e-04
## Oil_Prod    -3.273828e-11 1.405480e-10
```

```
anova(model_basic_50,model_oil_prod_50)
```

```
## Data: rad_all
## Models:
## model_basic_50: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_oil_prod_50: pb210 ~ radon + Umeans + Thmeans + Oil_Prod + (1 | city_state) +
## model_oil_prod_50:     (1 | YEAR:basin_name)
```

```
##                   Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_50     7 -3451.8 -3424.5 1732.9  -3465.8
## model_oil_prod_50  8 -3451.6 -3420.3 1733.8  -3467.6 1.7596      1
##                   Pr(>Chisq)
## model_basic_50
## model_oil_prod_50     0.1847
```

```
fixed.effects(model_oil_prod_50)
```

```
##   (Intercept)         radon        Umeans       Thmeans      Oil_Prod
##  7.060637e-03 -4.551873e-04 -1.002070e-04  5.614041e-04  5.567957e-11
```

Vertical Oil Production

After enlarging the radius to 50km. The main effect of vertical oil production is still significant. But the slope is smaller, approximately 1/4 of the slope in 25km case. Considering the mean of sum vertical oil production increased 7 times. The overall explanation power of vertical oil production increased.

```
model_v_oil_prod_50<-lmer(pb210~radon+Umeans+Thmeans+V_Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=
confint(model_v_oil_prod_50,method="boot")
```

```
##                    2.5 %       97.5 %
## .sig01       1.851851e-03 2.516940e-03
## .sig02       5.267386e-04 1.414268e-03
## .sigma       1.329974e-03 1.616084e-03
## (Intercept)  4.633982e-03 9.194038e-03
## radon       -1.054236e-03 3.883665e-05
## Umeans      -1.419386e-03 1.144024e-03
## Thmeans      2.287581e-04 9.183520e-04
## V_Oil_Prod   1.925226e-10 9.397345e-10
```

```
anova(model_basic_50,model_v_oil_prod_50)
```

```
## Data: rad_all
## Models:
## model_basic_50: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_v_oil_prod_50: pb210 ~ radon + Umeans + Thmeans + V_Oil_Prod + (1 | city_state) +
## model_v_oil_prod_50:     (1 | YEAR:basin_name)
##                     Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_50       7 -3451.8 -3424.5 1732.9  -3465.8
## model_v_oil_prod_50  8 -3458.7 -3427.4 1737.3  -3474.7 8.8415      1
##                     Pr(>Chisq)
## model_basic_50
## model_v_oil_prod_50   0.002945 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fixed.effects(model_v_oil_prod_50)
```

```
##   (Intercept)         radon        Umeans       Thmeans    V_Oil_Prod
##  6.871133e-03 -5.056702e-04 -8.028301e-05  5.776467e-04  5.710382e-10
```

Models of 75km cases

After running the models based on aggregated gas/oil production within 25km and 50km buffer. We enlarge the radius to 75km and run similar models. It's expected that the slope is still significant but the value is a little smaller.

Gross Oil Production

After enlarging the buffer to 75km. The main effect of gross oil production is not significant with or withour log-trasformation.

```
model_basic_75<-lmer(pb210~radon+Umeans+Thmeans+(1|city_state)+(1|YEAR:basin_name),data=rad_all,REML=T)
fixed.effects(model_basic_75)
```

```
##   (Intercept)         radon        Umeans       Thmeans
##  0.0072062940 -0.0004652606 -0.0001704985  0.0005779429
```

```
model_oil_prod_75<-lmer(pb210~radon+Umeans+Thmeans+Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=rad
confint(model_oil_prod_75,method="boot")
```

```
##                     2.5 %       97.5 %
## .sig01       1.847358e-03 2.497062e-03
## .sig02       4.880134e-04 1.483282e-03
## .sigma       1.347518e-03 1.631383e-03
## (Intercept)  4.684391e-03 9.548100e-03
## radon       -1.084984e-03 5.283494e-05
## Umeans      -1.452079e-03 1.086396e-03
## Thmeans      1.775247e-04 9.308269e-04
## Oil_Prod    -1.707747e-11 4.864934e-11
```

```
anova(model_basic_75,model_oil_prod_75)
```

```
## Data: rad_all
## Models:
## model_basic_75: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_oil_prod_75: pb210 ~ radon + Umeans + Thmeans + Oil_Prod + (1 | city_state) +
## model_oil_prod_75:     (1 | YEAR:basin_name)
##                   Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_75     7 -3451.8 -3424.5 1732.9  -3465.8
## model_oil_prod_75  8 -3451.0 -3419.7 1733.5  -3467.0 1.1862      1
##                   Pr(>Chisq)
## model_basic_75
## model_oil_prod_75     0.2761
```

Vertical Oil Production

After enlarging the radius to 75km. The main effect of vertical oil production is still significant. But the slope is smaller, approximately 1/4 of the slope in 25km case. Considering the mean of sum vertical oil production increased 7 times. The overall explanation power of vertical oil production increased.

```
model_v_oil_prod_75<-lmer(pb210~radon+Umeans+Thmeans+V_Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=
confint(model_v_oil_prod_75,method="boot")
```

```
##                     2.5 %       97.5 %
## .sig01       1.822748e-03 2.473933e-03
```

```
## .sig02       4.283640e-04 1.396151e-03
## .sigma       1.311964e-03 1.602063e-03
## (Intercept)  4.744830e-03 8.988300e-03
## radon       -1.104554e-03 5.566355e-06
## Umeans      -1.374830e-03 1.179639e-03
## Thmeans      2.265201e-04 9.554878e-04
## V_Oil_Prod   9.836768e-11 4.376097e-10
```

```
anova(model_basic_75,model_v_oil_prod_75)
```

```
## Data: rad_all
## Models:
## model_basic_75: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_v_oil_prod_75: pb210 ~ radon + Umeans + Thmeans + V_Oil_Prod + (1 | city_state) +
## model_v_oil_prod_75:     (1 | YEAR:basin_name)
##                      Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_75        7 -3451.8 -3424.5 1732.9  -3465.8
## model_v_oil_prod_75   8 -3459.5 -3428.2 1737.8  -3475.5 9.6745      1
##                      Pr(>Chisq)
## model_basic_75
## model_v_oil_prod_75    0.001868 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fixed.effects(model_v_oil_prod_75)
```

```
##   (Intercept)         radon        Umeans       Thmeans    V_Oil_Prod
##  6.878643e-03 -5.204755e-04 -4.443031e-05  5.664022e-04  2.796190e-10
```

Models of 100km cases

After running the models based on aggregated gas/oil production within 25km, 50km and 75km buffer. We enlarge the radius to 75km and run similar models. It's expected that the slope is still significant but the value is a little smaller.

Gross Oil Production

After enlarging the buffer to 75km. The main effect of gross oil production is not significant with or withour log-trasformation.

```
model_basic_100<-lmer(pb210~radon+Umeans+Thmeans+(1|city_state)+(1|YEAR:basin_name),data=rad_all,REML=T
fixed.effects(model_basic_100)
```

```
##   (Intercept)         radon        Umeans       Thmeans
##  0.0072062940 -0.0004652606 -0.0001704985  0.0005779429
```

```
model_oil_prod_100<-lmer(pb210~radon+Umeans+Thmeans+Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data=ra
confint(model_oil_prod_100,method="boot")
```

```
##                     2.5 %       97.5 %
## .sig01       1.816838e-03 2.512213e-03
## .sig02       5.171166e-04 1.432642e-03
## .sigma       1.327312e-03 1.623939e-03
## (Intercept)  4.946476e-03 9.434042e-03
```

```
## radon       -1.053274e-03 9.902681e-05
## Umeans      -1.399209e-03 1.202927e-03
## Thmeans      1.683096e-04 9.248350e-04
## Oil_Prod    -8.722922e-12 2.318344e-11
```

```
anova(model_basic_100,model_oil_prod_100)
```

```
## Data: rad_all
## Models:
## model_basic_100: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_oil_prod_100: pb210 ~ radon + Umeans + Thmeans + Oil_Prod + (1 | city_state) +
## model_oil_prod_100:     (1 | YEAR:basin_name)
##                    Df     AIC     BIC logLik deviance  Chisq Chi Df
## model_basic_100     7 -3451.8 -3424.5 1732.9  -3465.8
## model_oil_prod_100  8 -3450.8 -3419.5 1733.4  -3466.8 0.9659      1
##                    Pr(>Chisq)
## model_basic_100
## model_oil_prod_100     0.3257
```

```
fixed.effects(model_oil_prod_100)
```

```
##   (Intercept)         radon        Umeans       Thmeans      Oil_Prod
##  7.138946e-03 -4.880977e-04 -1.181367e-04  5.669643e-04  7.324637e-12
```

Vertical Oil Production

After enlarging the radius to 75km. The main effect of vertical oil production is still significant. But the slope is smaller, approximately 1/4 of the slope in 25km case. Considering the mean of sum vertical oil production increased 7 times. The overall explanation power of vertical oil production increased.

```
model_v_oil_prod_100<-lmer(pb210~radon+Umeans+Thmeans+V_Oil_Prod+(1|city_state)+(1|YEAR:basin_name),data
confint(model_v_oil_prod_100,method="boot")
```

```
##                     2.5 %        97.5 %
## .sig01       1.813256e-03  2.467343e-03
## .sig02       4.754841e-04  1.387084e-03
## .sigma       1.332373e-03  1.612516e-03
## (Intercept)  4.389512e-03  9.152760e-03
## radon       -1.099647e-03 -9.140363e-06
## Umeans      -1.406607e-03  1.336059e-03
## Thmeans      2.143113e-04  8.854184e-04
## V_Oil_Prod   6.504775e-11  3.119329e-10
```

```
anova(model_basic_100,model_v_oil_prod_100)
```

```
## Data: rad_all
## Models:
## model_basic_100: pb210 ~ radon + Umeans + Thmeans + (1 | city_state) + (1 | YEAR:basin_name)
## model_v_oil_prod_100: pb210 ~ radon + Umeans + Thmeans + V_Oil_Prod + (1 | city_state) +
## model_v_oil_prod_100:     (1 | YEAR:basin_name)
##                       Df     AIC     BIC logLik deviance Chisq Chi Df
## model_basic_100        7 -3451.8 -3424.5 1732.9  -3465.8
## model_v_oil_prod_100   8 -3460.0 -3428.7 1738.0  -3476.0 10.16      1
##                       Pr(>Chisq)
## model_basic_100
## model_v_oil_prod_100   0.001435 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fixed.effects(model_v_oil_prod_100)
```

```
##   (Intercept)         radon        Umeans       Thmeans    V_Oil_Prod
##  6.902822e-03 -5.567218e-04  1.117388e-05  5.520218e-04  1.909261e-10
```

Questions

Q1: Are the models and diagnostic process valid?

Q2: Current models use all the data, but only 1/6 (25km case) of the RadNet monitors are located within oil/gas field. Do I need to model based on the "contaminated" area only?

Q3: To control for the spatial confounder, a random intercept is assigned to each RadNet monitor . In additon, the EPA radon zone is also included. Is this sufficient, redundent or insufficient? Or can I use this as sensitivity analysis?

Q4: Concerning the significance part, can we state that we need to use the drilling information to update the radon zone data?

Q5: The temporal confounding is only controlled for by adding the year in the model. In the context of short study period, do you think this's sufficient?