Example 6b: Generalized Multilevel Models for Count Two-Level Nested Outcomes
(complete data, syntax, and output available for STATA, R, and SAS electronically)

These data were borrowed from Example 2 of the STATA MENBREG manual: "Rabe-Hesketh and Skrondal (2012, exercise 13.7) describe data from the *Atlas of Cancer Mortality in the European Economic Community* (EEC) (Smans, Mair, and Boyle 1993). The data were analyzed in Langford, Bentham, and McDonald (1998) and record the number of deaths among males due to malignant melanoma during 1971–1980." The STATA 16 example features a three-level model, a single predictor for the effect of UV exposure, and an offset for predicted exposure, whereas the example below differs in several respects. This example begins with single-level and two-level models to distinguish the need for a random intercept variance across the 351 counties nested within for the 77 regions as well as the need for a multiplicative over-dispersion parameter in predicting the number of deaths; I did not use their "expected" offset in this example. It then unsmushes the effect of UV exposure via county-level and region-level fixed effects, followed by a test for random slope variance of the within-county exposure effect. Finally, the dependency of regions nested in nations (excluding Luxemburg given *n*=3) was addressed via fixed effects on the intercept and interactions with each UV predictor. All models used laplace estimation given the lack of an alternative via the R glmer.nb function. Effect sizes are given as risk ratios (= odds ratios) and for pseudo-R^2 for the random effects.

STATA Syntax for Importing and Preparing Data for Analysis:

```
// Define global variable for file location to be replaced in code below
// \\Client\ precedes path in Virtual Desktop outside H drive
   global filesave "C:\Dropbox\23_PSQF6272\PSQF6272_Example6b"

// Open trimmed example excel data file from sheet "skin cancer" and clear away existing data
   clear // clear memory in case of open data
   import excel "$filesave\Example6b_Data.xlsx", firstrow case(preserve) sheet("skin cancer")

// Add labels to original variables
   label variable nation    "nation: String-Format Nation"
   label variable regionID   "regionID: L2 Region ID Variable"
   label variable countyID   "countyID: L1 County ID Variable"
   label variable deaths     "deaths: Count Outcome of Actual Deaths"
   label variable expected   "expected: Offset for Expected Deaths"
   label variable uv         "uv: County Amount of UV Exposure"

// Select cases complete for analysis variables
   egen nmiss=rowmiss(region deaths uv nation)
   drop if nmiss>0

// Create cluster=regionID means
   egen regionN  = count(deaths), by(regionID)
   egen RMdeaths = mean(deaths),  by(regionID)
   egen RMuv     = mean(uv),      by(regionID)
   label variable regionN   "regionN: # Counties Contributing Data"
   label variable RMdeaths  "RMdeaths: Region Mean of Death Count"
   label variable RMuv      "RMuv: Region Mean of UV Exposure"

display "STATA Descriptive Statistics within Region-Level-2 Data"
preserve  // Save for later use, then compute region-level dataset
collapse  regionN RMdeaths RMuv, by(regionID)
summarize regionN RMdeaths RMuv
pwcorr    regionN RMdeaths RMuv, sig
restore   // Go back to county-level dataset
```

SAS Descriptive Statistics within Region-Level-2 Data

Variable	N	Mean	Std Dev	Minimum	Maximum
regionN	77	4.56	2.63	1.00	13.00
RMdeaths	77	33.20	35.98	1.20	160.00
RMuv	77	-0.15	4.80	-7.06	12.72

Pearson Correlation Coefficients, N = 77 Prob > |r| under H0: Rho=0			
	regionN	RMdeaths	RMuv
regionN regionN: # Counties Contributing Data	1.00000	-0.25400 0.0258	0.06058 0.6007
RMdeaths RMdeaths: Region Mean of Death Count	-0.25400 0.0258	1.00000	-0.42614 0.0001
RMuv RMuv: Region Mean of UV Exposure	0.06058 0.6007	**-0.42614** 0.0001	1.00000

```
// Cluster-MC to get within-region deaths and UV
    gen WRdeaths = deaths - RMdeaths
    gen WRuv = uv - RMuv
    label variable WRdeaths   "WRdeaths: Within-regionID Deaths (0=RM)"
    label variable WRuv       "WRuv: Within-regionID UV Exposure (0=RM)"

// Convert nation from string to numeric
    encode nation, gen(catnation)

display "STATA Descriptive Statistics within County-Level-1 Data"
summarize deaths WRdeaths uv WRuv, detail
pwcorr    deaths WRdeaths uv WRuv, sig
tabulate  catnation
histogram deaths, percent discrete width(20) start(0)
graph export "$filesave\STATA Deaths Histogram.png", replace
```
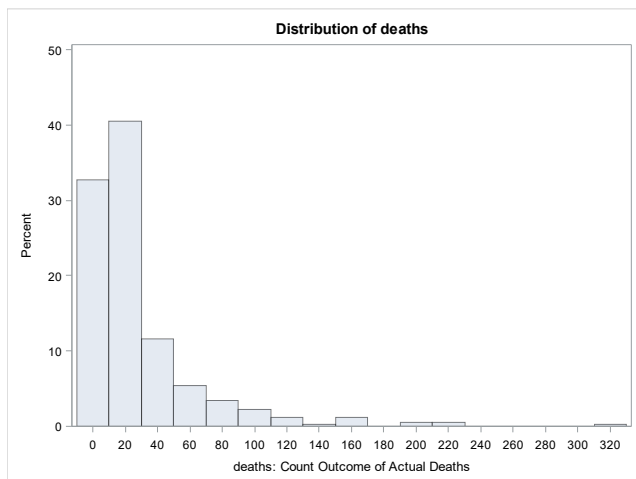
SAS Descriptive Statistics within County-Level-1 Data

Variable	N	Mean	Variance	Std Dev	Minimum	Maximum
deaths	351	28.000	1458.131	38.185	0.000	313.000
WRdeaths	351	0.000	633.691	25.173	-84.000	251.154
uv	351	0.020	24.754	4.975	-8.900	13.359
WRuv	351	-0.000	0.284	0.533	-2.070	1.936

Pearson Correlation Coefficients, N = 351; Prob > |r| under H0: Rho=0				
	deaths	WRdeaths	uv	WRuv
deaths deaths: Count Outcome of Actual Deaths	1.00000	0.65924 <.0001	-0.25289 <.0001	0.03545 0.5080
WRdeaths WRdeaths: Within-Region Deaths (0=RM)	0.65924 <.0001	1.00000	0.00576 0.9144	0.05378 0.3151
uv uv: County Amount of UV Exposure	**-0.25289** <.0001	0.00576 0.9144	1.00000	0.10705 0.0450
WRuv WRuv: Within-regionID UV Exposure (0=RM)	0.03545 0.5080	**0.05378** 0.3151	0.10705 0.0450	1.00000

Distribution of deaths

R Syntax for Importing and Preparing Data for Analysis (after loading packages *readxl*, *TeachingDemos*, *Hmisc*, *psych*, *lme4*, *lmerTest*, *performance*, *multcomp*, and *emmeans*):

```
# Define variables for working directory and data name -- CHANGE THESE
filesave = "C:\\Dropbox/23_PSQF6272/PSQF6272_Example6b/"
filename = "Example6b_Data.xlsx"
setwd(dir=filesave)

# Load Jonathan's custom R functions from folder within working directory
functions = paste0("R functions/",dir("R functions/"))
temp = lapply(X=functions, FUN=source)

#  Import trimmed example excel data file from sheet "skin cancer"
Example6b = read_excel(paste0(filesave,filename), sheet="skin cancer")
# Convert to data frame to use in analysis
Example6b = as.data.frame(Example6b)

# Filter to only cases complete on all variables to be used below (before cluster means)
Example6b = Example6b[complete.cases(Example6b[ , c("regionID","deaths","uv")]),]

# Create region=cluster means using Jonathan's function
Example6b = addUnitMeans(data=Example6b, unitVariable="regionID",
                         meanVariables=c("deaths","uv"), newNames=c("RMdeaths","RMuv"))

print("R Descriptive Statistics within Region-Level-2 Data")
regionMeans = unique(Example6b[,c("regionID","NperregionID","RMdeaths","RMuv")])
describe(x=regionMeans[ , c("NperregionID","RMdeaths","RMuv")])
rcorr(x=as.matrix(regionMeans[ , c("NperregionID","RMdeaths","RMuv")]), type="pearson")

# Cluster-MC to get within-region deaths and UV
Example6b$WRdeaths = Example6b$deaths - Example6b$RMdeaths
Example6b$WRuv = Example6b$uv - Example6b$RMuv

print("R Descriptive Statistics within County-Level-1 Data")
describe(x=Example6b[ , c("deaths","WRdeaths","uv","WRuv")])
rcorr(x=as.matrix(Example6b[ , c("deaths","WRdeaths","uv","WRuv")]), type="pearson")
var(Example6b$WRuv)
table(x=Example6b$nation, useNA="ifany")
# to save a plot: open a file, create the plot, then close the file
png(file = "R Deaths Histogram.png")  # open file
hist(x=Example6b$deaths, freq=FALSE,
     ylab="Density",xlab="deaths: Count Outcome of Actual Deaths") # axis labels
dev.off()  # close file
```

Empty Means, Random Intercept Model Predicting UV Exposure (continuous predictor)

Level 1: $UV_{cr} = \beta_{0r} + e_{cr}$

Level 2: $\beta_{0r} = \gamma_{00} + U_{0r}$

> For level-1 county c nested in level-2 region r

```
display "STATA Empty Means, Random Intercept Model for UV Exposure (predictor)"
mixed uv , || regionID: , reml dfmethod(satterthwaite) dftable(pvalue) nolog
estat icc   // Intraclass Correlation

print("R Empty Means, Random Intercept Model for UV Exposure (predictor)")
EmptyUV = lmer(data=Example6b, REML=TRUE, formula=uv~1+(1|regionID))
print("Show results using Satterthwaite DDF including -2LL as deviance")
llikAIC(EmptyUV, chkREML=FALSE); summary(EmptyUV, ddf="Satterthwaite")
print("Show intraclass correlation and its LRT")
icc(EmptyUV); ranova(EmptyUV)
```

```
Random effects:
 Groups    Name         Variance Std.Dev.
 regionID (Intercept) 22.92381 4.78788
 Residual              0.36231 0.60193
```

```
Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) -0.14218    0.54702 76.07480 -0.2599   0.7956

# Intraclass Correlation Coefficient
    Adjusted ICC: 0.984
  Unadjusted ICC: 0.984

ANOVA-like table for random-effects: Single term deletions
              npar    logLik    AIC     LRT Df Pr(>Chisq)
<none>           3  -530.628 1067.26
(1 | regionID)   2 -1061.133 2126.26 1061.01  1 < 2.22e-16
```

> ICC for the correlation of UV exposure
> for counties in the same region:
> $$ICC = \frac{22.9238}{22.9238 + 0.3612} = .9844$$

Model 1a. Empty Means, Single-Level Predicting Number of Deaths with Log Link and Poisson Conditional Distribution (in which Conditional Mean = Variance)

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r}$

Level 2: $\beta_{0r} = \gamma_{00}$

> For level-1 county c nested in level-2 region r
> g = gamma below

```
display "STATA Model 1a: Empty Means, Single-Level for Deaths"
display "Log Link, Poisson Conditional Distribution"
mepoisson deaths , nolog
------------------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   3.332205   .0100871   330.34   0.000     3.312434    3.351975  g00 in logs
------------------------------------------------------------------------------
```

```
display "-2LL = " e(ll)*-2  // Print -2LL for model
-2LL = 13248.921
```

> Expected count = 28.0 exactly matches the
> mean of deaths in the level-1 dataset

```
nlcom exp(_b[_cons])            // Fixed intercept in expected counts
------------------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _nl_1 |         28   .2824395    99.14   0.000     27.44643    28.55357  exp(g00)
------------------------------------------------------------------------------
```

```
print("R Model 1a: Empty Means, Single-Level for Deaths")
print("Log Link, Poisson Conditional Distribution")
Model1a = glm(data=Example6b, family=poisson(link="log"), formula=deaths~1)
-2*logLik(Model1a); summary(Model1a)

'log Lik.' 13248.921 (df=1) → -2LL for model

Coefficients:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 3.332205   0.010087  330.35 < 2.2e-16

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 11663.9  on 350  degrees of freedom
Residual deviance: 11663.9  on 350  degrees of freedom → not -2LL
```

```
print("Convert log count to expected count via inverse link")
Model1aCount=exp(coefficients(Model1a)); Model1aCount

(Intercept)
       28
```

Model 1b. Empty Means, Random Intercept Predicting Number of Deaths with Log Link and Poisson Conditional Distribution (in which mean = variance)

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r}$

Level 2: $\beta_{0r} = \gamma_{00} + U_{0r}$

> For level-1 county c nested in level-2 region r
> g = gamma below

```
display "STATA Model 1b: Empty Means, Random Intercept for Deaths"
display "Log Link, Poisson Conditional Distribution"
mepoisson deaths , || regionID: , intmethod(laplace) nolog
estimates store Model1b          // Save for LRT
```
```
-----------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+---------------------------------------------------------
       _cons |   3.010757   .1158271    25.99   0.000      2.78374    3.237774   g00 in logs
-------------+---------------------------------------------------------
regionID     |
   var(_cons)|   1.010885    .169026                        .7284099   1.402903   Var(U0r) in logs
-----------------------------------------------------------------------
LR test vs. Poisson model: chibar2(01) = 7566.71       Prob >= chibar2 = 0.0000
```
```
display "-2LL = " e(ll)*-2      // Print -2LL for model
-2LL = 5682.2127
```
```
nlcom exp(_b[_cons])            // Fixed intercept in expected counts
-----------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+---------------------------------------------------------
       _nl_1 |   20.30276    2.35161     8.63   0.000      15.69369   24.91184   exp(g00)
-----------------------------------------------------------------------
```

```
print("R Model 1b: Empty Means, Random Intercept for Deaths")
print("Log Link, Poisson Conditional Distribution, Laplace Estimation")
print("-2LL is not at all comparable, but results match SAS and STATA")
Model1b = glmer(data=Example6b, family=poisson(link="log"),
             deaths~1+(1|regionID))
print("Show -2LL with precision, results, and LRT for random intercept")
-2*logLik(Model1b); summary(Model1b); anova(Model1b,Model1a)
```

'log Lik.' **5682.2127** (df=2) → **–2LL for model**

```
Random effects:
 Groups    Name           Variance Std.Dev.
 regionID (Intercept) 1.0109    1.0054   Var(U0r)
```

> Note that the fixed intercept no longer is the exact sample mean (28)—this is because it is "unit-specific": it is the expected count for a region with $U_{0r} = 0$

```
Fixed effects:
           Estimate Std. Error z value  Pr(>|z|)
(Intercept)   3.0108     0.1158        26 < 2.2e-16   g00 in logs
```
```
        npar      AIC      BIC  logLik deviance   Chisq Df Pr(>Chisq)
Model1a    1 13250.92 13254.78 -6624.46 13248.92
Model1b    2  5686.21  5693.93 -2841.11  5682.21 7566.71  1 < 2.22e-16
```

```
print("Convert log count to expected count via inverse link")
Model1bCount=exp(fixef(Model1b)); Model1bCount
(Intercept)
  20.302871
```

```
print("ICC using icc function in performance package"); icc(Model1b)
# Intraclass Correlation Coefficient
    Adjusted ICC: 0.955
  Unadjusted ICC: 0.955
```

```
print("Compute ICC manually using residual variance as defined below")
intvar=(Model1b@theta)^2; resvar=log(1+(1/exp(fixef(Model1b))))
ICC=intvar/(intvar+resvar); ICC
 0.95459683
```

Model 1c. Empty Means, Random Intercept Predicting Number of Deaths with Log Link and <u>Negative Binomial</u> Conditional Distribution for Overdispersion (in which mean < variance)

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r}$

Level 2: $\beta_{0r} = \gamma_{00} + U_{0r}$

This negative binomial model adds a "scale" parameter (for overdispersion) in which the variance increases as a **quadratic function of the mean**.

```
display "STATA Model 1c: Empty Means, Random Intercept for Deaths"
display "Log Link, Negative Binomial Conditional Distribution"
menbreg deaths , || regionID: , intmethod(laplace) nolog
```

```
-------------------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
       _cons |   3.049335    .1158039    26.33   0.000     2.822363    3.276306   g00
-------------+-----------------------------------------------------------------
    /lnalpha |   -.846158    .0947238                     -1.031813   -.6605028   log(1/dispersion)
-------------+-----------------------------------------------------------------
regionID     |
   var(_cons)|   .8384467    .1635084                      .5721102    1.228772   Var(U0r)
-------------------------------------------------------------------------------
LR test vs. nbinomial model: chibar2(01) = 183.84     Prob >= chibar2 = 0.0000
```

```
display "-2LL = " e(ll)*-2      // Print -2LL for model
-2LL = 2868.2191
```

```
nlcom exp(_b[_cons])           // Fixed intercept in expected counts
-------------------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
       _nl_1 |    21.1013    2.443613     8.64   0.000     16.31191    25.89069   exp(g00)
-------------------------------------------------------------------------------
```

```
estimates store Model1c  // Save for LRT
lrtest Model1c Model1b   // LRT for overdispersion

LR chi2(1) = 2813.99
Prob > chi2 =  0.0000
```

The LRT = 183.84 above is for the random intercept variance, and the LRT = 2813.99 below is for the overdispersion scale parameter. Looks like we need both of them!

Calculate a 95% random effect confidence interval for the region random intercept:
*CI = fixed effect ± 1.96*SQRT(random intercept variance)*
*CI = 3.0493 ± 1.96*SQRT(0.8384) = −1.25 to 4.84 in log counts, or 3.51 to 126.97 in expected counts.*

```
print("R Model 1c: Empty Means, Random Intercept for Deaths")
print("Log Link, Negative Binomial Conditional Distribution")
Model1c = glmer.nb(data=Example6b, deaths~1+(1|regionID))
print("Show -2LL with precision, results, and LRT for overdispersion")
-2*logLik(Model1c); summary(Model1c); anova(Model1c,Model1b)

'log Lik.' 2867.5129 (df=3) → -2LL for model

    AIC      BIC   logLik deviance df.resid
  2873.5   2885.1   -1433.8   2867.5      348

Random effects:
 Groups    Name        Variance Std.Dev.
 regionID (Intercept) 0.83999  0.91651

Fixed effects:
           Estimate Std. Error z value  Pr(>|z|)
(Intercept)   2.9983     0.1131   26.51  < 2.2e-16

         npar     AIC     BIC   logLik deviance  Chisq Df Pr(>Chisq)
Model1b     2 5686.21 5693.93 -2841.11  5682.21
Model1c     3 2873.51 2885.10 -1433.76  2867.51 2814.7  1 < 2.22e-16
```

```
print("Convert log count to expected count via inverse link")
Model1cCount=exp(fixef(Model1c)); Model1cCount
   20.05107
```

```
print("Overdispersion from R")
getME(Model1c, "glmer.nb.theta")
[1] 2.3351857
```

<div style="border:1px solid black; padding:5px;">
In R the "**scale**" variance multiplier (in which 1 = Poisson) is given directly. In **SAS** it's given as **1/scale**, and in STATA it's given as **log(1/scale)**, labeled as "**/lnalpha**" in the output. You will see these decrease after adding level-1 fixed effects that explain variance, but they can't be used for purposes of pseudo-R^2.
</div>

```
print("Overdispersion that nearly matches STATA")
log(1/(getME(Model1c, "glmer.nb.theta")))
[1] -0.84809141
```

```
print("Overdispersion that nearly matches SAS GLIMMIX")
1/(getME(Model1c, "glmer.nb.theta"))
[1] 0.42823147
```

Continuing with two-level models with a log link and negative binomial conditional distribution…

Model 2a: Add Fixed Slope of Region Mean UV

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r}$

Level 2: $\beta_{0r} = \gamma_{00} + \gamma_{01}(\overline{UV_r}) + U_{0r}$

```
display "STATA Model 2a: Add Fixed Slope of Region Mean UV"
menbreg deaths c.RMuv, || regionID: ,  intmethod(laplace) nolog
```

```
------------------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        RMuv |  -.0825012   .0217735    -3.79   0.000    -.1251765   -.0398259   g01
       _cons |   3.038251   .1043409    29.12   0.000     2.833747    3.242756   g00
-------------+----------------------------------------------------------------
    /lnalpha |  -.8493021   .0947247                     -1.034959   -.6636451   log(1/dispersion)
-------------+----------------------------------------------------------------
regionID     |
   var(_cons)|   .6823389   .1377027                      .4594299     1.0134    Var(U0r)
------------------------------------------------------------------------------
LR test vs. nbinomial model: chibar2(01) = 138.10    Prob >= chibar2 = 0.0000
```

```
display "-2LL = " e(ll)*-2      // Print -2LL for model
-2LL = 2855.0334
```

```
display "STATA Model 2a: Risk Ratio Effect Sizes Instead "
menbreg deaths c.RMuv, || regionID: ,  intmethod(laplace) nolog eform
```

```
------------------------------------------------------------------------------
      deaths |     exp(b)   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        RMuv |   .9208103   .0200493    -3.79   0.000     .8823411    .9609567   exp(g01)
       _cons |   20.86872   2.177462    29.12   0.000     17.00907    25.60419   exp(g00)
-------------+----------------------------------------------------------------
```

Pseudo-R2 Relative to CovEmpty1c (from SAS)

Name	CovParm	Subject	Estimate	PseudoR2
CovEmpty1c	UN(1,1)	regionID	0.8384	.
CovBRuv2a	UN(1,1)	regionID	0.6823	0.18619

What does the fixed intercept NOW represent? *The log of the death count for a county in a region with a random intercept $U_{0r} = 0$ and region mean UV exposure = 0 is 3.038, which is a count = 20.867 (from exponentiation).*

What does the main effect of region mean UV represent? <u>*Without controlling for county UV*</u>*, for every unit higher region mean UV, the log of the expected death count is significantly lower by 0.0825 (RR = 0.9208). This between-region UV effect accounted for 18.62% of the level-2 region random intercept variance.*

```
print("R Model 2a:  Add Fixed Slope of Region Mean UV")
Model2a = glmer.nb(data=Example6b, deaths~1+RMuv+(1|regionID))
print("Show -2LL with precision, results, and overdispersion")
-2*logLik(Model2a); summary(Model2a)


'log Lik.' 2854.4829 (df=4) → -2LL for model


     AIC      BIC   logLik deviance df.resid
  2862.5   2877.9  -1427.2   2854.5      347

Random effects:
 Groups    Name         Variance Std.Dev.
 regionID (Intercept) 0.68504  0.82767  Var(U0r)

Fixed effects:
             Estimate Std. Error z value  Pr(>|z|)
(Intercept)  2.989560   0.103762 28.8117 < 2.2e-16  g00
RMuv        -0.081891   0.021738 -3.7672 0.0001651  g01

getME(Model2a, "glmer.nb.theta")
[1] 2.3439329 → dispersion

print("Get risk ratio effect size"); exp(fixef(Model2a))
(Intercept)          RMuv
  19.876929      0.921372  → exp(gamma)
```

Model 2b: Add Fixed Slope of Within-Region UV Mean Predictor

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r} + \beta_{1r}(uv_{cr} - \overline{UV_r})$

Level 2: $\beta_{0r} = \gamma_{00} + \gamma_{01}(\overline{UV_r}) + U_{0r} \qquad \beta_{1r} = \gamma_{10}$

```
display "STATA Model 2b: Add Fixed Slope of Within-Region UV"
menbreg deaths c.RMuv c.WRuv, || regionID: , intmethod(laplace) nolog
estimates store Model2b        // Save for LRT
-------------------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
        RMuv | -.0826372   .0218255    -3.79   0.000    -.1254143    -.03986   g01
        WRuv |   .138657   .0828537     1.67   0.094    -.0237333   .3010473   g10
       _cons |  3.036924   .1039221    29.22   0.000      2.83324   3.240607   g00
-------------+-----------------------------------------------------------------
    /lnalpha | -.8588796   .0947314                      -1.04455  -.6732096   log(1/dispersion)
-------------+-----------------------------------------------------------------
regionID     |
   var(_cons)|  .6837942   .1380328                       .4603618  1.015667   Var(U0r)
-------------------------------------------------------------------------------
LR test vs. nbinomial model: chibar2(01) = 139.33     Prob >= chibar2 = 0.0000

display "-2LL = " e(ll)*-2      // Print -2LL for model
-2LL = 2852.2365

lincom c.RMuv*1 +c.WRuv*-1      // UV Contextual slope
-------------------------------------------------------------------------------
      deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
         (1) | -.2212941    .085626    -2.58   0.010    -.3891179  -.0534703   g01-g10
-------------------------------------------------------------------------------
```

What does the fixed intercept NOW represent? *The log of the expected death count = 3.0370 (expected count = 20.841 from below) for a county in a region with a random intercept $U_{0r} = 0$ and region mean UV exposure = 0.*

What does the main effect of region mean UV NOW represent? *The interpretation is the same: without controlling for county UV, for every unit higher region mean UV, the log of the expected death count is significantly lower by 0.0827 (RR = 0.9207). This is the between-region effect. This effect is still significant after controlling for county UV, as indicated by a contextual effect = −0.2213 (RR =0.8015 from below).*

What does the main effect of within-region UV represent? *For every unit higher within-region county UV relative to the rest of the region, the log of the expected death count is nonsignificantly higher by 0.1387 (RR = 1.1487). We cannot compute a pseudo-R^2 for the residual variance, which is a function of the mean (and thus is non-constant).*

```
display "STATA Model 2b: Risk Ratio Effect Sizes Instead"
menbreg deaths c.RMuv c.WRuv, || regionID: , intmethod(laplace) nolog eform
------------------------------------------------------------------------
      deaths |    exp(b)   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------
        RMuv |  .9206851   .0200944    -3.79   0.000    .8821314    .9609239   exp(g01)
        WRuv |   1.14873   .0951765     1.67   0.094    .9765461    1.351273   exp(g10)
       _cons |  20.84103   2.165844    29.22   0.000    17.00046    25.54923   exp(g00)
-------------+----------------------------------------------------------
lincom c.RMuv*1 +c.WRuv*-1, irr    // UV Contextual slope
------------------------------------------------------------------------
      deaths |      IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------
         (1) |  .8014809   .0686276    -2.58   0.010    .6776544    .9479341   exp(g01-g10)
------------------------------------------------------------------------
```

```
print("R Model 2b:  Add Fixed Slope of Within-Region UV")
Model2b = glmer.nb(data=Example6b, deaths~1+RMuv+WRuv+(1|regionID))
print("Show -2LL with precision, results, and overdispersion")
-2*logLik(Model2b); summary(Model2b)

'log Lik.' 2851.6735 (df=5)   → –2LL for model

     AIC      BIC   logLik deviance df.resid
  2861.7   2881.0   -1425.8   2851.7      346

Random effects:
 Groups    Name        Variance Std.Dev.
 regionID (Intercept) 0.68628  0.82842   Var(U0r)

Fixed effects:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept)  2.988579   0.103772 28.7996 < 2.2e-16   g00
RMuv        -0.082029   0.021741 -3.7730 0.0001613   g01
WRuv         0.138264   0.082543  1.6751 0.0939212   g10

getME(Model2b, "glmer.nb.theta")
[1] 2.3662007   → dispersion

print("Get risk ratio effect sizes"); exp(getME(Model2b, "fixef"))
(Intercept)        RMuv        WRuv
19.85744946  0.92124508  1.14827896   → exp(gamma)

Model2bglht = summary(glht(model=Model2b, linfct=rbind(
  "UV Contextual Slope and Risk Ratio"= c(0,1,-1))),test=adjusted("none"))
Model2bglht; exp(Model2bglht$test$coefficients)
Linear Hypotheses:
                                     Estimate Std. Error z value Pr(>|z|)
UV Contextual Slope and Risk Ratio == 0 -0.220293   0.085353  -2.581 0.009852   g01-g10
(Adjusted p values reported -- none method)

> exp(Model2bglht$test$coefficients)
UV Contextual Slope and Risk Ratio
                   0.80228334   exp(g01-g10)
```

Model 2c: Add Random Slope of Within-Region UV Mean Predictor

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r} + \beta_{1r}(uv_{cr} - \overline{UV_r})$

Level 2: $\beta_{0r} = \gamma_{00} + \gamma_{01}(\overline{UV_r}) + U_{0r}$ $\beta_{1r} = \gamma_{10} + U_{1r}$

```
display "STATA Model 2c: Add Random Slope of Within-Region UV"
menbreg deaths c.RMuv c.WRuv, || regionID: c.WRuv, cov(un) intmethod(laplace) nolog
-------------------------------------------------------------------------------
       deaths | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------+----------------------------------------------------------------
         RMuv | -.0842463    .0215884    -3.90   0.000    -.1265587   -.0419339  g01
         WRuv |  .1074161    .102708      1.05   0.296    -.0938878    .3087201  g10
        _cons | 3.024315     .1040806    29.06   0.000     2.820321    3.228309  g00
--------------+----------------------------------------------------------------
      /lnalpha | -.9365029   .0991939                     -1.130919   -.7420865  log(1/disp)
--------------+----------------------------------------------------------------
regionID      |
    var(WRuv)|   .126337     .0735324                      .0403741    .3953291  Var(U1r)
   var(_cons)|   .6949277    .1384332                      .470301     1.026841  Var(U0r)
--------------+----------------------------------------------------------------
regionID      |
cov(WRuv,_cons)| -.1135761   .1055167    -1.08   0.282    -.3203851    .0932328  Cov(U0,U1)
-------------------------------------------------------------------------------
LR test vs. nbinomial model: chi2(3) = 146.74          Prob > chi2 = 0.0000

display "-2LL = " e(ll)*-2       // Print -2LL for model
-2LL = 2844.8212

estimates store Model2c         // Save for LRT
lrtest Model2c Model2b          // LRT for random slope

 LR chi2(2) =    7.42
Prob > chi2 = 0.0245

print("R Model 2c:  Add Random Slope of Within-Region UV")
Model2c = glmer.nb(data=Example6b, deaths~1+RMuv+WRuv+(1+WRuv|regionID))
print("Show -2LL with precision, results, and overdispersion")
-2*logLik(Model2c); summary(Model2c)

'log Lik.' 2846.0578 (df=7) → -2LL for model

    AIC      BIC   logLik deviance df.resid
  2860.1   2887.1  -1423.0   2846.1      344

Random effects:
 Groups    Name         Variance Std.Dev. Corr
 regionID (Intercept) 0.69571  0.83409           Var(U0r)
          WRuv        0.11237  0.33521  -0.428  Var(U1r) Corr(U0,U1)

getME(Model2c, "glmer.nb.theta")
[1] 2.544305  → dispersion

Fixed effects:
            Estimate Std. Error z value   Pr(>|z|)
(Intercept)  2.974474   0.104031 28.5922  < 2.2e-16  g00
RMuv        -0.084263   0.021492 -3.9206 0.00008833  g01
WRuv         0.119371   0.101244  1.1790     0.2384  g10

print("LRT for random slope"); anova(Model2c, Model2b)
        npar    AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
Model2b    5 2861.67 2880.98 -1425.84  2851.67
Model2c    7 2860.06 2887.08 -1423.03  2846.06 5.61576  2   0.060333
```

Does the level-2 random effect of level-1 within-region UV improve model fit?

Yes, $-2\Delta LL$(mixture of df=1 and df=2) = 7.42, mixture p-value = .0155 (see excel spreadsheet)

Calculate a 95% random effect confidence interval for the within-region UV slope:

*CI = fixed effect ± 1.96*SQRT(random slope variance)*
*CI = 0.1074 ± 1.96*SQRT(0.1263) = −0.59 to 0.81 in log counts (you cannot unlog or exponentiate slopes)*

So what does this mean? *The extent to which within-region UV differences predicts county death count varies significantly across regions, with some regions expecting positive slopes and others expecting negative slopes.*

Model 3a: Control for Main Effect of Nation on Intercept

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r} + \beta_{1r}(uv_{cr} - \overline{UV_r})$

Level 2: $\beta_{0r} = \gamma_{00} + \gamma_{01}(\overline{UV_r}) + \sum_{n=1}^{7} \gamma_{0,n+1}(Nation_r = n) + U_{0r}$

$\qquad \beta_{1r} = \gamma_{10} + U_{1r}$

Because we are adding 7 fixed effects, we need 7 new gammas in the fixed effects predicting the intercept. The summation sign conveys this efficiently (with specific parameters as labeled below adding commas). The first category is the default reference, and so each fixed effect is the difference for that country relative to the reference country.

```
display "STATA Model 3a: Control for Main Effect of Nation on Intercept"
menbreg deaths c.RMuv c.WRuv i.catnation, ///
    || regionID: c.WRuv, cov(un) intmethod(laplace) nolog baselevels difficult
```

deaths	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
RMuv	-.0607778	.0242398	-2.51	0.012	-.108287	-.0132686	**g0,1**
WRuv	.1310242	.1007302	1.30	0.193	-.0664035	.3284519	**g1,0**
catnation							
Belgium	0	(base)					
Denmark	-.0224265	.3854822	-0.06	0.954	-.7779577	.7331046	**g0,2**
France	-.7526957	.2966279	-2.54	0.011	-1.334076	-.1713156	**g0,3**
Ireland	-2.911847	.3698549	-7.87	0.000	-3.63675	-2.186945	**g0,4**
Italy	-.5587369	.3462026	-1.61	0.107	-1.237282	.1198076	**g0,5**
Netherlands	.0177251	.3766398	0.05	0.962	-.7204753	.7559255	**g0,6**
UK	-.5040486	.2960282	-1.70	0.089	-1.084253	.0761561	**g0,7**
W. Germany	.9234955	.3072133	3.01	0.003	.3213685	1.525622	**g0,8**
_cons	3.504924	.2677143	13.09	0.000	2.980214	4.029634	**g0,0**
/lnalpha	-.9468714	.0964652			-1.13594	-.757803	**log(1/disp)**
regionID							
var(WRuv)	.1174509	.065452			.0394009	.3501115	**Var(U1r)**
var(_cons)	.1239045	.0364014			.0696652	.2203728	**Var(U0r)**
regionID							
cov(WRuv,_cons)	-.0847918	.0424794	-2.00	0.046	-.1680498	-.0015338	**Cov(U0,U1)**

LR test vs. nbinomial model: chi2(3) = 42.40   Prob > chi2 = 0.0000

```
display "-2LL = " e(ll)*-2          // Print -2LL for model
-2LL = 2744.7977
```

```
contrast i.catnation    // Omnibus DF=7 test of nation on intercept
```

	df	chi2	P>chi2
deaths			
catnation	7	**179.98**	0.0000

This DF=7 chi-square tests the difference across all 8 nations simultaneously. The solution above only gives some of the possible model-implied differences—those relative to the reference nation (Belgium here). We could ask for any others of interest using margins or lincom.

```
margins  i.catnation, predict(xb)  // Predicted unit-specific log count by nation
```
```
-------------------------------------------------------------------------------
             |            Delta-method
             |    Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
     Belgium |   3.503701   .2678282    13.08   0.000     2.978767    4.028635   g0,0
     Denmark |   3.481274   .3072648    11.33   0.000     2.879046    4.083502   g0,0 + g0,2
      France |   2.751005   .1105273    24.89   0.000     2.534376    2.967635   g0,0 + g0,3
     Ireland |   .5918535   .2844883     2.08   0.037     .0342667     1.14944   g0,0 + g0,4
       Italy |   2.944964   .1762274    16.71   0.000     2.599565    3.290363   g0,0 + g0,5
 Netherlands |   3.521426   .2891778    12.18   0.000     2.954648    4.088204   g0,0 + g0,6
          UK |   2.999652   .1756335    17.08   0.000     2.655417    3.343888   g0,0 + g0,7
  W. Germany |   4.427196   .1813411    24.41   0.000     4.071774    4.782618   g0,0 + g0,8
-------------------------------------------------------------------------------
```
```
margins  i.catnation, predict(mu)  // Marginal expected count by nation
```
```
-------------------------------------------------------------------------------
             |            Delta-method
             |    Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
     Belgium |   37.46191    9.85986     3.80   0.000     18.13694    56.78688
     Denmark |   36.63112   10.80917     3.39   0.001     15.44553    57.81671
      France |   17.64812   2.196061     8.04   0.000     13.34392    21.95232
     Ireland |   2.036998   .5551304     3.67   0.000     .9489628    3.125034
       Italy |   21.42563   4.302979     4.98   0.000     12.99194    29.85931
 Netherlands |   38.13185   10.66749     3.57   0.000     17.22396    59.03973
          UK |   22.62999   3.568895     6.34   0.000     15.63509     29.6249
  W. Germany |   94.33198   16.25167     5.80   0.000     62.47929    126.1847
-------------------------------------------------------------------------------
```

```r
print("R Model 3a: Control for Main Effect of Nation on Intercept")
Model3a = glmer.nb(data=Example6b, deaths~1+RMuv+WRuv
                +as.factor(nation)+(1+WRuv|regionID))
print("Show -2LL with precision, results, and overdispersion")
-2*logLik(Model3a); summary(Model3a)
```

'log Lik.' **2746.4788** (df=14) → **-2LL for model**

```
     AIC       BIC    logLik deviance df.resid
  2774.5    2828.5   -1373.2   2746.5      337
```

```
Random effects:
 Groups    Name         Variance Std.Dev. Corr
 regionID (Intercept) 0.12286  0.35051            Var(U0r)
          WRuv         0.10724  0.32747  -0.773   Var(U1r)  Cor(U0,U1)
```

```
Fixed effects:
                              Estimate Std. Error z value  Pr(>|z|)
(Intercept)                   3.479382   0.297873 11.6808 < 2.2e-16   g0,0
RMuv                         -0.060062   0.024013 -2.5012   0.01238   g0,1
WRuv                          0.159419   0.100759  1.5822   0.11361   g1,0
as.factor(nation)Denmark     -0.021952   0.404586 -0.0543   0.95673   g0,2
as.factor(nation)France      -0.760268   0.325224 -2.3377   0.01940   g0,3
as.factor(nation)Ireland     -2.896492   0.388194 -7.4615 8.557e-14   g0,4
as.factor(nation)Italy       -0.559277   0.371741 -1.5045   0.13246   g0,5
as.factor(nation)Netherlands  0.013406   0.397263  0.0337   0.97308   g0,6
as.factor(nation)UK          -0.496019   0.320473 -1.5478   0.12168   g0,7
as.factor(nation)W. Germany   0.919973   0.332223  2.7691   0.00562   g0,8
```

```r
getME(Model3a, "glmer.nb.theta")
```
[1] **2.5678408** → **dispersion**

```r
print("Omnibus DF=7 test of nation on intercept"); anova(Model3a)
Analysis of Variance Table
                  npar  Sum Sq Mean Sq  F value
RMuv                 1  43.4928 43.4928 43.49276
WRuv                 1   2.1040  2.1040  2.10401
```
as.factor(nation)   7 183.1213 26.1602 26.16018 → **Btw, F = chi-square/#slopes**

```
print("Predicted unit-specific log count by nation")
emmeans(Model3a, specs="nation")
```

```
nation        emmean    SE  df asymp.LCL asymp.UCL
 Belgium       3.478 0.298 Inf    2.8941      4.06  g0,0
 Denmark       3.456 0.305 Inf    2.8576      4.05  g0,0 + g0,2
 France        2.718 0.110 Inf    2.5025      2.93  g0,0 + g0,3
 Ireland       0.582 0.282 Inf    0.0283      1.14  g0,0 + g0,4
 Italy         2.919 0.175 Inf    2.5767      3.26  g0,0 + g0,5
 Netherlands   3.492 0.288 Inf    2.9278      4.06  g0,0 + g0,6
 UK            2.982 0.174 Inf    2.6411      3.32  g0,0 + g0,7
 W. Germany    4.398 0.179 Inf    4.0466      4.75  g0,0 + g0,8

Results are given on the log (not the response) scale.
Confidence level used: 0.95
```

Pseudo-R2 Relative to CovRandWRuv2c (fromSAS)

Name	CovParm	Subject	Estimate	PseudoR2
CovRandWRuv2c	UN(1,1)	regionID	0.6949	.
CovRandWRuv2c	UN(2,2)	regionID	0.1263	.
CovNatMain3a	UN(1,1)	regionID	0.1239	0.82170
CovNatMain3a	UN(2,2)	regionID	0.1174	0.07041

The DF=7 main effect of nation explained **82.17%** of the remaining random intercept variance (which is what it should have targeted) and **7.04%** of the random within-region UV slope variance (which it should *not* have targeted).

Model 3b: Add National as a Control Predictor—Interactions with BR and WR UV Slopes

Level 1: $Log(E[deaths_{cr}]) = \beta_{0r} + \beta_{1r}(uv_{cr} - \overline{UV}_r)$

Level 2: $\beta_{0r} = \gamma_{00} + \gamma_{01}(\overline{UV}_r) + \sum_{n=1}^{7}\gamma_{0,n+1}(Nation_r = n) + \sum_{n=1}^{7}\gamma_{0,n+8}(\overline{UV}_r)(Nation_r = n) + U_{0r}$

$\beta_{1r} = \gamma_{10} \qquad\qquad + \sum_{n=1}^{7}\gamma_{1,n+1}(Nation_r = n) + U_{1r}$

We need 7 new interaction fixed effects with RMuv, as well as 7 cross-level interaction fixed effects predicting the within-region UV slope. The summation sign conveys this efficiently (with specific parameters as labeled below). The first category is the reference, and so each fixed effect is the difference for that country relative to the reference.

```
display "STATA Model 3b: Add Nation Interactions with UV Slopes"
menbreg deaths c.RMuv c.WRuv i.catnation i.catnation#c.RMuv i.catnation#c.WRuv, ///
       || regionID: c.WRuv, cov(un) intmethod(laplace) nolog baselevels diffult
-------------------------------------------------------------------------------
          deaths | Coefficient  Std. err.     z    P>|z|     [95% conf. interval]
-----------------+-------------------------------------------------------------
            RMuv | -3.340402    .1797739   -18.58   0.000    -3.692753   -2.988052  g0,1
            WRuv | -.7880363    1.141548    -0.69   0.490    -3.025429    1.449356  g1,0
       catnation |
         Belgium |         0    (base)
         Denmark |  5.727207    6.90222      0.83   0.407    -7.800894   19.25531   g0,2
          France |  8.84147     .5280026    16.75   0.000     7.806604    9.876336  g0,3
         Ireland | 10.53397     2.376984     4.43   0.000     5.875166   15.19277   g0,4
           Italy |  9.173619    .5648329    16.24   0.000     8.066567   10.28067   g0,5
     Netherlands | 14.22699     2.547394     5.58   0.000     9.234193   19.21979   g0,6
              UK | 10.35433     .7148347    14.48   0.000     8.95328    11.75538   g0,7
      W. Germany | 10.33474     .6313229    16.37   0.000     9.097373   11.57211   g0,8
                 |
 catnation#c.RMuv |
         Denmark |  2.595348    1.296        2.00   0.045     .0552339    5.135462  g0,9
          France |  3.265387    .1831522    17.83   0.000     2.906416    3.624359  g0,10
         Ireland |  4.057001    .4977696     8.15   0.000     3.081391    5.032612  g0,11
           Italy |  3.253224    .1796081    18.11   0.000     2.901199    3.60525   g0,12
     Netherlands |  4.406338    .6270512     7.03   0.000     3.17734     5.635335  g0,13
              UK |  3.561276    .2098728    16.97   0.000     3.149933    3.972619  g0,14
      W. Germany |  3.244261    .2127753    15.25   0.000     2.82723     3.661293  g0,15
                 |
```

```
catnation#c.WRuv |
         Denmark |   -.3449321   1.393106    -0.25   0.804    -3.075369    2.385505   g1,2
          France |    .8748838   1.154612     0.76   0.449    -1.388115    3.137883   g1,3
         Ireland |    1.343707   1.296319     1.04   0.300    -1.197033    3.884446   g1,4
           Italy |     .854335   1.158716     0.74   0.461    -1.416706    3.125376   g1,5
     Netherlands |   -2.732094    2.06809    -1.32   0.186    -6.785476    1.321288   g1,6
              UK |    1.071575   1.174819     0.91   0.362    -1.231029    3.374178   g1,7
      W. Germany |    1.259776    1.19667     1.05   0.292    -1.085654    3.605206   g1,8
                 |
           _cons |   -6.050965   .5165537   -11.71   0.000    -7.063391   -5.038538   g0,0
-----------------+----------------------------------------------------------------
        /lnalpha |   -.9788317   .0969886                     -1.168926   -.7887375   log(1/disp)
-----------------+----------------------------------------------------------------
regionID         |
        var(WRuv)|    .119931   .0690455                      .0388048     .370661   Var(U1r)
       var(_cons)|     .06482    .027076                      .0285859    .1469825   Var(U0r)
-----------------+----------------------------------------------------------------
regionID         |
 cov(WRuv,_cons)|   -.0465893   .0433843    -1.07   0.283    -.1316209    .0384423   Cov(U0,U1)
----------------------------------------------------------------------------------
LR test vs. nbinomial model: chi2(3) = 18.56              Prob > chi2 = 0.0003
```

display "-2LL = " e(ll)*-2   **// Print -2LL for model**
-2LL = 2715.7682

contrast i.catnation   **// Omnibus DF=7 test catnation on intercept**

```
-----------------------------------------------
                |      df       chi2     P>chi2
----------------+------------------------------
deaths          |
      catnation |       7      426.28     0.0000
-----------------------------------------------
```

> Do the level-2 fixed simple main effects of nation improve the model?
> *Yes, $\chi^2(7) = 426.28$, $p < .0001$*

contrast i.catnation#c.RMuv   **// Omnibus DF=7 test catnation on RMuv slope**

```
----------------------------------------------------
                    |      df       chi2     P>chi2
--------------------+-------------------------------
deaths              |
catnation#c.RMuv    |       7      529.58     0.0000
----------------------------------------------------
```

> Do the level-2 fixed interaction effects of nation by region mean UV improve the model?
> *Yes, $\chi^2(7) = 529.58$, $p < .0001$*

contrast i.catnation#c.WRuv   **// Omnibus DF=7 test catnation on WRuv slope**

```
----------------------------------------------------
                    |      df       chi2     P>chi2
--------------------+-------------------------------
deaths              |
catnation#c.WRuv    |       7       10.40     0.1669
----------------------------------------------------
```

> Do the cross-level fixed interaction effects of nation by within-region UV improve the model?
> *Yes $\chi^2(7) = 529.58$, $p < .0001$*

However—given that these are omnibus interactions with DF=7, it is still informative to see what the simple between-region and within-region UV effects are for each nation, as follows:

// UV Between fixed slope per catnation
margins, over(catnation) dydx(c.RMuv) predict(xb)

```
---------------------------------------------------------------------------------
                |            Delta-method
                |      dy/dx    std. err.      z     P>|z|     [95% conf. interval]
----------------+----------------------------------------------------------------
RMuv            |
      catnation |
        Belgium |   -3.340402   .1797739    -18.58   0.000    -3.692753   -2.988052   g0,1
        Denmark |   -.7450545   1.184545     -0.63   0.529     -3.06672    1.576611   g0,1 + g0,9
         France |   -.0750147    .037803     -1.98   0.047    -.1491073   -.0009221   g0,1 + g0,10
        Ireland |    .7165991   .4541163      1.58   0.115    -.1734524    1.606651   g0,1 + g0,11
          Italy |   -.0871779   .0278118     -3.13   0.002     -.141688   -.0326678   g0,1 + g0,12
    Netherlands |    1.065935   .5776112      1.85   0.065    -.0661618    2.198032   g0,1 + g0,13
             UK |    .2208734    .097686      2.26   0.024     .0294124    .4123344   g0,1 + g0,14
     W. Germany |   -.0961407   .1125284     -0.85   0.393    -.3166924     .124411   g0,1 + g0,15
---------------------------------------------------------------------------------
```

```
// UV Within fixed slope per catnation
margins, over(catnation) dydx(c.WRuv) predict(xb)
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
WRuv         |
  catnation  |
     Belgium |  -.7880363   1.141548    -0.69   0.490    -3.025429    1.449356   g1,0
     Denmark |  -1.132968   .7305234    -1.55   0.121    -2.564768    .2988312   g1,0 + g1,2
      France |   .0868475   .1593871     0.54   0.586    -.2255454    .3992404   g1,0 + g1,3
     Ireland |   .5556705   .6238094     0.89   0.373    -.6669735    1.778314   g1,0 + g1,4
       Italy |   .0662988   .1916148     0.35   0.729    -.3092594    .4418569   g1,0 + g1,5
 Netherlands |  -3.520131   1.713769    -2.05   0.040    -6.879056   -.1612047   g1,0 + g1,6
          UK |   .2835385   .2504386     1.13   0.258    -.2073121    .7743892   g1,0 + g1,7
  W. Germany |   .4717396   .3449692     1.37   0.171    -.2043876    1.147867   g1,0 + g1,8
------------------------------------------------------------------------------
```

```
print("R Model 3b: Add Nation Interactions with UV Slopes")
Model3b = glmer.nb(data=Example6b, deaths~1+RMuv+WRuv++as.factor(nation)
                   +as.factor(nation):RMuv+as.factor(nation):WRuv+(1+WRuv|regionID))
print("Show -2LL with precision, results, and overdispersion")
-2*logLik(Model3b); summary(Model3b)
```

'log Lik.' **2717.6391** (df=28) → **-2LL for model**

```
     AIC      BIC   logLik deviance df.resid
  2773.6   2881.7  -1358.8   2717.6      323
```

```
Random effects:
 Groups    Name            Variance Std.Dev. Corr
 regionID  (Intercept)     0.064438 0.25385          Var(U0r)
           WRuv            0.108801 0.32985  -0.622   Var(U1r)   Corr(U0,U1)
```

```
Fixed effects:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -6.20827    4.12815 -1.5039 0.132610   g0,0
RMuv                            -3.38779    1.42086 -2.3843 0.017111   g0,1
WRuv                            -0.76828    1.04226 -0.7371 0.461043   g1,0
as.factor(nation)Denmark         5.80158    6.59105  0.8802 0.378740   g0,2
as.factor(nation)France          8.97324    4.12772  2.1739 0.029713   g0,3
as.factor(nation)Ireland        10.67924    4.71914  2.2630 0.023638   g0,4
as.factor(nation)Italy           9.30751    4.13122  2.2530 0.024261   g0,5
as.factor(nation)Netherlands    14.35166    4.76224  3.0136 0.002581   g0,6
as.factor(nation)UK             10.44532    4.13686  2.5249 0.011572   g0,7
as.factor(nation)W. Germany     10.46199    4.14129  2.5263 0.011528   g0,8

RMuv:as.factor(nation)Denmark     2.63137   1.68958  1.5574 0.119372   g0,9
RMuv:as.factor(nation)France      3.31450   1.42258  2.3299 0.019811   g0,10
RMuv:as.factor(nation)Ireland     4.10415   1.49169  2.7513 0.005935   g0,11
RMuv:as.factor(nation)Italy       3.30143   1.42133  2.3228 0.020192   g0,12
RMuv:as.factor(nation)Netherlands 4.45098   1.52920  2.9107 0.003607   g0,13
RMuv:as.factor(nation)UK          3.59819   1.42044  2.5332 0.011304   g0,14
RMuv:as.factor(nation)W. Germany  3.28973   1.42449  2.3094 0.020921   g0,15

WRuv:as.factor(nation)Denmark    -0.35092   1.26477 -0.2775 0.781427   g1,2
WRuv:as.factor(nation)France      0.86666   1.05381  0.8224 0.410846   g1,3
WRuv:as.factor(nation)Ireland     1.33016   1.21012  1.0992 0.271680   g1,4
WRuv:as.factor(nation)Italy       0.86881   1.06333  0.8171 0.413890   g1,5
WRuv:as.factor(nation)Netherlands -2.71798  2.00298 -1.3570 0.174792   g1,6
WRuv:as.factor(nation)UK          1.07734   1.07019  1.0067 0.314088   g1,7
WRuv:as.factor(nation)W. Germany  1.25318   1.09551  1.1439 0.252656   g1,8
```

```
optimizer (Nelder_Mead) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.00489407 (tol = 0.002, component 1)
Model is nearly unidentifiable: large eigenvalue ratio
 - Rescale variables?
```

```
getME(Model3b, "glmer.nb.theta")
[1] 2.6689824 → dispersion

print("Omnibus DF=7 test of nation on intercept, each slope"); anova(Model3b)
print("Simple slopes per country for between and within UV effects")
Model3bglht = glht(model=Model3b, linfct=rbind(
  "Between-Region UV: Belgium"      = c(0,1,0, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0),
  "Between-Region UV: Denmark"      = c(0,1,0, 0,0,0,0,0,0,0, 1,0,0,0,0,0,0, 0,0,0,0,0,0,0),
  "Between-Region UV: France"       = c(0,1,0, 0,0,0,0,0,0,0, 0,1,0,0,0,0,0, 0,0,0,0,0,0,0),
  "Between-Region UV: Ireland"      = c(0,1,0, 0,0,0,0,0,0,0, 0,0,1,0,0,0,0, 0,0,0,0,0,0,0),
  "Between-Region UV: Italy  "      = c(0,1,0, 0,0,0,0,0,0,0, 0,0,0,1,0,0,0, 0,0,0,0,0,0,0),
  "Between-Region UV: Netherlands"= c(0,1,0, 0,0,0,0,0,0,0, 0,0,0,0,1,0,0, 0,0,0,0,0,0,0),
  "Between-Region UV: UK"           = c(0,1,0, 0,0,0,0,0,0,0, 0,0,0,0,0,1,0, 0,0,0,0,0,0,0),
  "Between-Region UV: W Germany " = c(0,1,0, 0,0,0,0,0,0,0, 0,0,0,0,0,0,1, 0,0,0,0,0,0,0),
  "Within-Region UV: Belgium"     = c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0),
  "Within-Region UV: Denmark"     = c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 1,0,0,0,0,0,0),
  "Within-Region UV: France"      = c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,1,0,0,0,0,0),
  "Within-Region UV: Ireland"     = c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,0,1,0,0,0,0),
  "Within-Region UV: Italy  "     = c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,0,0,1,0,0,0),
  "Within-Region UV: Netherlands"= c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,0,0,0,1,0,0),
  "Within-Region UV: UK"          = c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,0,0,0,0,1,0),
  "Within-Region UV: W Germany " = c(0,0,1, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0, 0,0,0,0,0,0,1)))
Summary(Model3bglht, test=adjusted("none"))
```

```
Linear Hypotheses:                     Estimate Std. Error z value Pr(>|z|)
Between-Region UV: Belgium == 0        -3.387791   1.420863 -2.3843 0.017111   g0,1
Between-Region UV: Denmark == 0        -0.756417   0.920976 -0.8213 0.411463   g0,1 + g0,9
Between-Region UV: France == 0         -0.073292   0.037567 -1.9510 0.051063   g0,1 + g0,10
Between-Region UV: Ireland == 0         0.716358   0.450651  1.5896 0.111924   g0,1 + g0,11
Between-Region UV: Italy   == 0        -0.086362   0.027542 -3.1357 0.001715   g0,1 + g0,12
Between-Region UV: Netherlands == 0     1.063191   0.567008  1.8751 0.060780   g0,1 + g0,13
Between-Region UV: UK == 0              0.210400   0.094416  2.2284 0.025852   g0,1 + g0,14
Between-Region UV: W Germany  == 0     -0.098065   0.112281 -0.8734 0.382454   g0,1 + g0,15

Within-Region UV: Belgium == 0         -0.768280   1.042257 -0.7371 0.461043   g1,0
Within-Region UV: Denmark == 0         -1.119204   0.717856 -1.5591 0.118974   g1,0 + g1,2
Within-Region UV: France == 0           0.098381   0.155043  0.6345 0.525727   g1,0 + g1,3
Within-Region UV: Ireland == 0          0.561883   0.613391  0.9160 0.359653   g1,0 + g1,4
Within-Region UV: Italy   == 0          0.100535   0.195599  0.5140 0.607264   g1,0 + g1,5
Within-Region UV: Netherlands == 0     -3.486256   1.708911 -2.0400 0.041346   g1,0 + g1,6
Within-Region UV: UK == 0               0.309061   0.242434  1.2748 0.202370   g1,0 + g1,7
Within-Region UV: W Germany  == 0       0.484901   0.336396  1.4415 0.149455   g1,0 + g1,8
(Adjusted p values reported -- none method)
```

Pseudo-R2 Relative to CovRandWRuv2c (from SAS)
Change in Pseudo-R2 for CovNatMain3a vs. CovNatUV3b

Name	CovParm	Subject	Estimate	PseudoR2	PseudoR2Change
CovRandWRuv2c	UN(1,1)	regionID	0.6949	.	.
CovRandWRuv2c	UN(2,2)	regionID	0.1263	.	.
CovNatMain3a	UN(1,1)	regionID	0.1239	0.82170	.
CovNatMain3a	UN(2,2)	regionID	0.1174	0.07041	.
CovNatUV3b	UN(1,1)	regionID	0.06469	0.90691	0.085211
CovNatUV3b	UN(2,2)	regionID	0.1195	0.05423	-0.016179

Sample Results Section for Count Multilevel Models using STATA Output

[indicates notes about what to customize or also include; note that SE and p-values are not needed if you provide tables for the model solutions]

The extent to which UV exposure could predict death counts *[uncorrected for expected differences in population size, which could have been controlled for as an offset]* was examined in a series of multilevel models in which the 351 counties were modeled as nested at level 1 within their 77 regions at level 2, and region differences were captured via region-level random effects. Based on the results from preliminary empty means models (as described below), the death count outcome was predicted in two-level models using a log link function and negative binomial conditional outcome distribution. All model parameters were estimated via full-information maximum likelihood (Laplace method) in STATA 17. Accordingly, all fixed effects should be interpreted as unit-specific (i.e., as the fixed effect specifically for regions in which the corresponding random effect = 0). The significance of fixed effects was evaluated with Wald tests (i.e., the z-test of the ratio of each estimate its standard error without denominator degrees of freedom), whereas the significance of random effects was evaluated via likelihood ratio tests (i.e., $-2\Delta LL$ with degrees of freedom equal to the number of new random effects variances and covariances). Effect size was evaluated via pseduo-R^2 values for the proportion reduction in each variance component for level-2 region variances as well as by incident risk ratios (exponentiated slope coefficients), which for negative slopes range from 0–1 and for positive slopes range from 1 to positive infinity.

We initially tested the need for a random intercept variance and for over-dispersion separately by comparing the fit of single-level and two-level models with either Poisson or negative binomial conditional distributions. Relative to a single-level negative Poisson model, the two-level Poisson model had significantly better fit, $-2\Delta LL(1) = 7566.71$, $p < .001$, indicating significant dependency (correlation) of the death count within regions. Likewise, relative to a two-level Poisson model, the two-level negative binomial model had significantly better fit, $-2\Delta LL(1) = 2,813.99$, $p < .001$, indicating significant over-dispersion of the conditional variance (i.e., of the level-1 residuals). A 95% random effects confidence interval, calculated as fixed intercept $\pm 1.96*SQRT$(random intercept variance), revealed that 95% of the regions were predicted to have expected death counts between 3.51 and 126.97. The fixed intercept estimate for the log expected death count in an average region (random intercept = 0) was 3.049, or count = 21.100.

We then examined the impact of UV exposure in predicting death counts. Given that previous analyses had revealed that approximately 98% of the variance in UV expsoure was between regions, the level-1 variance in county UV exposure was represented by cluster-mean-centering, in which the level-1 predictor was computed by substracting the region's mean UV exposure from each county's UV exposure. The level-2 region variance in UV exposure was then represented by the uncentered region mean UV exposure (given that zero was already the mean of the UV distribution). The level-2 effect of region mean UV exposure was first added to the model. The fixed intercept = 3.038 indicated that the log death count for a county in a region with a random intercept = 0 and region mean UV exposure = 0 (expected count = 20.869). The level-2 between-region effect of UV exposure indicated that for every unit higher region mean UV, the log of the expected death count was significantly lower by 0.083, which accounted for 18.62% of the level-2 region random intercept variance. Next, the effect of cluster-mean-centered within-region UV expsoure was added to the model. The fixed intercept = 3.037 indicated the log of the expected death count for a county in a region with a random intercept = 0, region mean UV exposure = 0, and within-region UV exposure = 0 (expected count = 20.838). The level-1 within-region effect of UV exposure indicated that every unit higher within-region county UV relative to the rest of the region, the log of the expected death count was nonsignificantly higher by 0.139. The level-2 between-region UV effect of -0.083 was significantly different than the within-region UV effect, as indicated by a significant level-2 contextual effect of -0.221. We then examined to what extent the within-region effect of UV exposure varied across regions. A level-2 random slope variance for the effect of level-1 within-region UV exposure resulted in a significant improvement in model fit, $-2\Delta LL$(mixture of df =1 and df=2) = 7.42, $p = .016$, indicating that the size of the within-region UV slope differed significantly across regions. A 95% random effects confidence interval for the within-region UV effect, calculated as fixed slope ± 1.96 *SQRT(random slope variance), revealed that 95% of the regions were predicted to have UV-related slopes on the log scale ranging from -0.59 to 0.81.

We then examined the potential for differences across the nations in which the regions were located. We first considered nation differences in the expected death count intercept, for which a significant omnibus effect was found. We then added interactions of nation with the between-region and within-region UV slopes. In total, significant differences across nations were found in the fixed intercepts, $\chi^2(7) = 426.28$, $p < .0001$, as well as in the level-2 between-region UV slopes, $\chi^2(7) = 529.58$, $p < .0001$, but not in the level-1 within-region UV slopes, $\chi^2(7) = 10.40$, $p < .0001$. These interactions explained 8.52% of the remaining level-2 region random intercept variance but none of the region random slope variance for within-region UV exposure. Model-predicted simple slopes for each region were then requested via MARGINS statements. The level-2 between-region slopes for UV exposure were significantly positive in one nation, nonsignificant in four nations, and significantly negative in two nations. The within-region slopes for UV exposure were significantly negative in one nation and nonsignificant in the other seven nations. [figures illustrating nation slopes might be useful] [table of results from final model would also be useful]