Example 5b: Generalized Linear Mixed Models for Count Clustered Outcomes using SAS PROC GLIMMIX and STATA MEPOISSON/MENBREG

These data were borrowed from Example 2 of the STATA 16 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 (no offset used). 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 for convenience given *n*=3) was addressed via fixed effects on the intercept and interactions with each UV predictor.

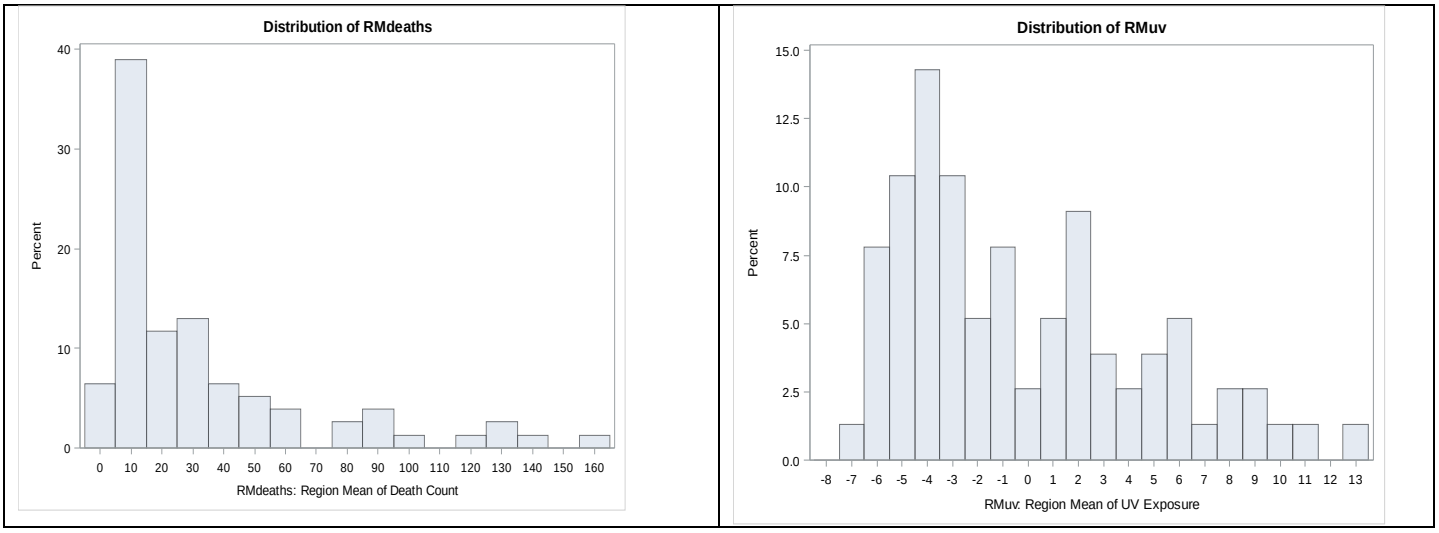
SAS Syntax for Data Import, Manipulation, and Description:

```
* Define global variable for file location to be replaced in code below;
* \\Client\ precedes actual path when using UIowa Virtual Desktop;
%LET filesave=C:\Dropbox\19_PSQF7375_Clustered\PSQF7375_Clustered_Example5b;
LIBNAME example "&filesave.";

* Import GSS_subsample STATA data file into SAS;
PROC IMPORT DATAFILE="&filesave.\skincancer_v11.dta"
            OUT=work.skincancer DBMS=DTA REPLACE; RUN;

* Label existing variables;
DATA work.skincancer; SET work.skincancer;
     LABEL region= "region: Region Nesting Variable"
           deaths= "deaths: Count of Deaths"
           uv=     "uv: Amount of UV Exposure";
     * Select cases that are complete for analysis variables;
     IF NMISS(region,deaths,uv,nation)>0 THEN DELETE;
     * Remove Luxemberg (N=3);
     IF nation=8 THEN DELETE; RUN;

* Get region means;
PROC SORT DATA=work.skincancer; BY region; RUN;
PROC MEANS NOPRINT N DATA=work.skincancer;
     BY region; VAR deaths uv;
     OUTPUT OUT=work.RegionMeans MEAN(deaths uv)= RMdeaths RMuv; RUN;

* Label new region mean variables;
DATA work.RegionMeans; SET work.RegionMeans;
     Nperregion = _FREQ_; * Saving N per region;
     DROP _TYPE_ _FREQ_;  * Dropping unneeded SAS-created variables;
     LABEL Nperregion= "Nperregion: # Students Contributing Data"
           RMdeaths=   "RMdeaths: Region Mean of Death Count"
           RMuv=       "RMuv: Region Mean of UV Exposure"; RUN;

* Merge region means back with individual data;
DATA work.skincancer; MERGE work.skincancer work.RegionMeans; BY region;
     * Center region mean uv (uncentered, but remember to center it);
     RMuv0 = RMuv - 0;
     LABEL RMuv0= "RMuv0: Region Mean of UV Exposure (0=0)";
     * Center to get within-region deaths and UV;
     WRdeaths = deaths - RMdeaths;
     WRuv = uv - RMuv;
     LABEL WRdeaths= "WRdeaths: Within-Region Deaths (0=RM)"
           WRuv= "WRuv: Within-region UV Exposure (0=RM)"; RUN;
```

```
TITLE "Region-Level Descriptives";
PROC MEANS NDEC=2 DATA=work.RegionMeans;
     VAR Nperregion RMdeaths RMuv; RUN;
```

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Nperregion	Nperregion: # Students Contributing Data	77	4.56	2.63	1.00	13.00
RMdeaths	RMdeaths: Region Mean of Death Count	77	33.20	35.98	1.20	160.00
RMuv	RMuv: Region Mean of UV Exposure	77	-0.15	4.80	-7.06	12.72

```
PROC UNIVARIATE NOPRINT DATA=work.RegionMeans;
     VAR RMdeaths RMuv;
     HISTOGRAM RMdeaths / MIDPOINTS=0 TO 160 BY 10;
     HISTOGRAM RMuv / MIDPOINTS = -8 TO 13 BY 1;
RUN; QUIT; TITLE;
```



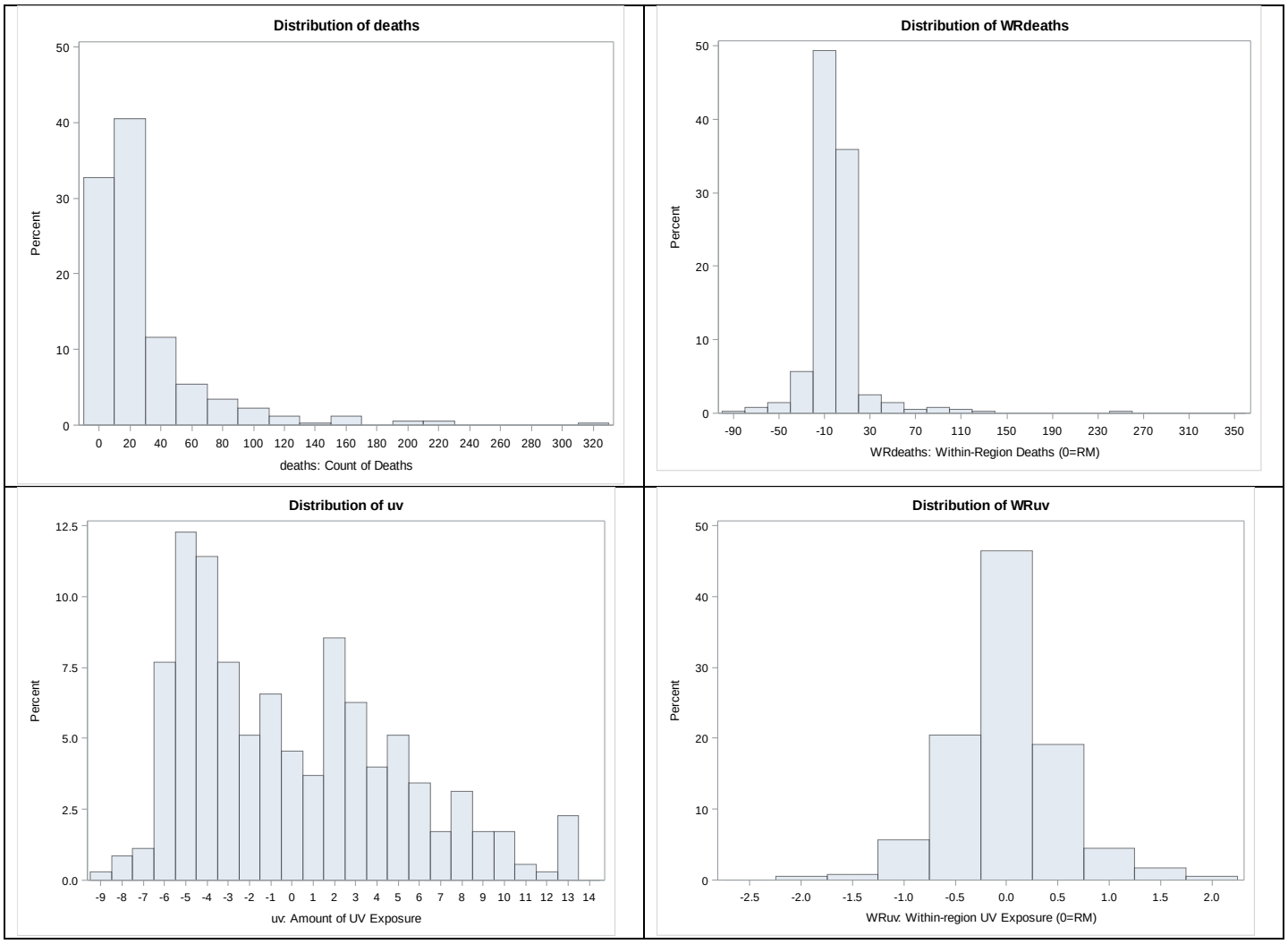

```
TITLE "County-Level Descriptives";
PROC FREQ DATA=work.skincancer;
     TABLE nation; RUN; TITLE;
```

nation				
nation	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Belgium	11	3.13	11	3.13
W. Germany	30	8.55	41	11.68
Denmark	14	3.99	55	15.67
France	94	26.78	149	42.45
UK	70	19.94	219	62.39
Italy	95	27.07	314	89.46
Ireland	26	7.41	340	96.87
Netherlands	11	3.13	351	100.00

```
PROC MEANS NDEC=2 DATA=work.skincancer;
     VAR deaths WRdeaths uv WRuv; RUN;
```

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
deaths	deaths: Count of Deaths	351	28.00	38.19	0.00	313.00
WRdeaths	WRdeaths: Within-Region Deaths (0=RM)	351	0.00	25.17	-84.00	251.15
uv	uv: Amount of UV Exposure	351	0.02	4.98	-8.90	13.36
WRuv	WRuv: Within-region UV Exposure (0=RM)	351	0.00	0.53	-2.07	1.94

```
PROC UNIVARIATE NOPRINT DATA=work.skincancer;
     VAR deaths WRdeaths uv WRuv;
     HISTOGRAM deaths / MIDPOINTS=0 TO 320 BY 20;
     HISTOGRAM WRdeaths / MIDPOINTS=-90 TO 360 BY 20;
     HISTOGRAM uv / MIDPOINTS = -9 TO 14 BY 1;
     HISTOGRAM WRuv / MIDPOINTS = -2.50 TO 2.00 BY 0.50;
RUN; QUIT; TITLE;
```




```
* Removing value labels from nation to get numeric order back;
DATA work.skincancer; SET work.skincancer;
     FORMAT nation; RUN;
```

STATA Syntax for Data Import, Manipulation, and Description:

```
// Define global variable for file location to be replaced in code below
// \\Client\ precedes actual path when using UIowa Virtual Desktop
global filesave "C:\Dropbox\19_PSQF7375_Clustered\PSQF7375_Clustered_Example5b"

// Import example stata data file
use "$filesave\skincancer.dta", clear

// Save results to separate file
log using $filesave\PSQF7375_Clustered_Example5b_STATA_Output.log, replace name(STATA_Example5b)

// Label existing variables
label variable region  "region: Region Nesting Variable"
label variable deaths  "deaths: Count of Deaths"
label variable uv      "uv: Amount of UV Exposure"
```

```
// Select cases complete for analysis variables
egen nmiss=rowmiss(region deaths uv nation)
drop if nmiss>0
drop if nation==8 // remove Luxemburg because N=3

// Get region means of variables and label them
egen RMdeaths = mean(deaths), by(region)
egen RMuv     = mean(uv),     by(region)
label variable RMdeaths "RMdeaths: Region Mean of Death Count"
label variable RMuv     "RMuv: Region Mean of UV Exposure"

// Get count per region and label it
egen Nperregion = count(deaths), by(region)
label variable Nperregion "Nperregion: Count per Region"

// Center region mean uv (uncentered, but remember to center it)
gen RMuv0 = RMuv
label variable RMuv0 "RMuv0: Region Mean of UV Exposure (0=0)"

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

display as result "STATA Region-Level Descriptives"
preserve  // Save for later use, then compute region-level dataset
collapse  Nperregion RMdeaths RMuv, by(region)
format    Nperregion RMdeaths RMuv  %4.2f
summarize Nperregion RMdeaths RMuv, format
histogram RMdeaths, percent discrete width(10) start(0)
histogram RMuv, percent discrete width(1) start(-8)
restore   // Go back to county-level dataset

display as result "STATA County-Level Descriptives"
format    deaths WRdeaths uv WRuv %4.2f
summarize deaths WRdeaths uv WRuv, format
tabulate  nation // Asking for it here to preserve value labels
histogram deaths, percent discrete width(20) start(0)
histogram WRdeaths, percent discrete width(20) start(-90)
histogram uv, percent discrete width(1) start(-9)
histogram WRuv, percent discrete width(0.5) start(-2.50)
```

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

Level 1: $UV_{cr} = \beta_{0r} + e_{cr}$
Level 2: Intercept: $\beta_{0r} = \gamma_{00} + U_{0r}$

```
TITLE "SAS Empty Means, Random Intercept Model for UV Exposure (predictor)";
PROC MIXED DATA=work.skincancer NOCLPRINT COVTEST NAMELEN=100 IC METHOD=REML;
     CLASS region;
     MODEL uv = / SOLUTION DDFM=Satterthwaite CHISQ;
     RANDOM INTERCEPT / VCORR=2 TYPE=UN SUBJECT=region; * VCORR gives ICC;
RUN; TITLE;

display as result "STATA Empty Means, Random Intercept Model for UV Exposure (predictor)"
mixed uv , ///
        || region: , variance reml covariance(unstructured) dfmethod(satterthwaite),
      estat ic, n(77), // get AIC and BIC equivalent to SAS
      estat icc        // compute Intraclass Correlation
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	region	22.9225	3.7353	6.14	<.0001
Residual		0.3623	0.03095	11.71	<.0001

ICC for the correlation of UV exposure for counties in the same region:

$$ICC = \frac{22.9225}{22.9225 + 0.3612} = .9844$$

Single-Level and Two-Level Models Predicting Number of Deaths (count outcome)

Single-Level Count Models:

Level 1: $\text{Log}(\text{Death}_{cr}) = \beta_{0r}$

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

Two-Level Count Models:

Level 1: $\text{Log}(\text{Death}_{cr}) = \beta_{0r}$

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

```
TITLE1 "SAS Model 1a: Empty Means, Single-Level Model for Deaths (outcome)";
TITLE2 "Log Link, Poisson Conditional Distribution";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region;
     MODEL deaths =  / SOLUTION LINK=LOG DIST=POISSON DDFM=Satterthwaite CHISQ;
     ESTIMATE "Intercept" intercept 1 / ILINK; * ILINK is inverse link (to un-log);
RUN; TITLE1; TITLE2;

display as result "STATA Model 1a: Empty Means, Single-Level Model for Deaths (outcome)"
display as result "Log Link, Poisson Conditional Distribution"
mepoisson deaths ,
          estat ic, n(77),     // get AIC and BIC equivalent to SAS
          nlcom exp(_b[_cons]) // fixed intercept in counts
```

```
          Fit Statistics
-2 Log Likelihood           13248.92
AIC  (smaller is better)    13250.92
AICC (smaller is better)    13250.93
BIC  (smaller is better)    13254.78
CAIC (smaller is better)    13255.78
HQIC (smaller is better)    13252.46
Pearson Chi-Square          18226.64
Pearson Chi-Square / DF        51.93
```

To go from logs to counts for predicted outcomes (i.e., to apply the inverse log link or exponentiation):

Count(y) = exp (3.322) = 28 (SE = 0.2824), which is equal to the mean of the deaths variable

```
                          Parameter Estimates
                       Standard
Effect       Estimate       Error       DF    t Value    Pr > |t|    Gradient
Intercept      3.3322     0.01009      350     330.34      <.0001    7.583E-6
```

```
                                  Estimates
                                                                    Standard
                       Standard                                        Error
Label        Estimate       Error       DF    t Value    Pr > |t|        Mean        Mean
Intercept      3.3322     0.01009      350     330.34      <.0001     28.0000      0.2824
```

```
TITLE1 "SAS Model 1b: Empty Means, Single-Level Model for Deaths (outcome)";
TITLE2 "Log Link, Negative Binomial Conditional Distribution";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region;
     MODEL deaths =  / SOLUTION LINK=LOG DIST=NEGBIN DDFM=Satterthwaite CHISQ;
     ESTIMATE "Intercept" intercept 1 / ILINK; * ILINK is inverse link (to un-log);
RUN; TITLE1; TITLE2;

display as result "STATA Model 1b: Empty Means, Single-Level Model for Deaths (outcome)"
display as result "Log Link, Negative Binomial Conditional Distribution"
menbreg deaths ,
        estat ic, n(77),     // get AIC and BIC equivalent to SAS
        nlcom exp(_b[_cons]) // fixed intercept in counts
```

```
          Fit Statistics
-2 Log Likelihood            3052.06
AIC  (smaller is better)     3056.06
AICC (smaller is better)     3056.09
BIC  (smaller is better)     3063.78
CAIC (smaller is better)     3065.78
HQIC (smaller is better)     3059.13
Pearson Chi-Square            577.68
Pearson Chi-Square / DF         1.65
```

This version of the negative binomial model predicts the variance to increase as a function of the mean (a multiplicative over-dispersion model, as opposed to an additive over-dispersion model).

SAS gives a "**scale**" parameter (in which 1 = Poisson) as the variance multiplier, whereas STATA gives the natural log of the scale parameter (labeled as "**/lnalpha**" in the output). You will see these decrease via level-1 predictors that explain variance, but they can't be used for pseudo-R^2.

```
                        Parameter Estimates
                     Standard
Effect       Estimate     Error       DF    t Value    Pr > |t|    Gradient
Intercept      3.3322   0.05666      350      58.81      <.0001    -1.13E-7
Scale          1.0911   0.07742        .          .           .    -2.61E-6
```

The mean count is the same as in the Poisson model (28), but its SE (based on the model-predicted variance) is greater in the negative binomial model (0.28 vs. 1.59).

```
                        Estimates
                     Standard                                    Standard Error
Label        Estimate     Error       DF    t Value    Pr > |t|        Mean        Mean
Intercept      3.3322   0.05666      350      58.81      <.0001     28.0000      1.5865
```

```
display as result "STATA Model 2a: Empty Means, Random Intercept Model for Deaths (outcome)"
display as result "Log Link, Poisson Conditional Distribution"
mepoisson deaths , || region: , covariance(unstructured) intpoints(15),
          estat ic, n(77),          // get AIC and BIC equivalent to SAS
          nlcom exp(_b[_cons])      // fixed intercept in counts
          estimates store Poisson2  // Save for LRT
```

```
TITLE1 "SAS Model 2a: Empty Means, Random Intercept Model for Deaths (outcome)";
TITLE2 "Log Link, Poisson Conditional Distribution";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region;
     MODEL deaths =  / SOLUTION LINK=LOG DIST=Poisson DDFM=BW CHISQ;
     RANDOM INTERCEPT / TYPE=UN SUBJECT=region;
     ESTIMATE "Intercept" intercept 1 / ILINK; * ILINK is inverse link (to un-log);
     COVTEST "Random Region Intercept?" 0;     * Test if G matrix (1,1)=0;
RUN; TITLE1; TITLE2;
```

```
          Fit Statistics
-2 Log Likelihood            5681.99
AIC  (smaller is better)     5685.99
AICC (smaller is better)     5686.03
BIC  (smaller is better)     5690.68
CAIC (smaller is better)     5692.68
HQIC (smaller is better)     5687.87
```

```
     Fit Statistics for Conditional
             Distribution
-2 log L(deaths | r. effects)     5269.49
Pearson Chi-Square                4623.95
Pearson Chi-Square / DF             13.17
```

The Pearson χ^2 / DF statistic is an index of fit to the conditional distribution (i.e., after including random effects). It should be close to 1 for good fit.

```
           Covariance Parameter Estimates
Cov                              Standard
Parm       Subject    Estimate       Error    Gradient
UN(1,1)    region       1.0116      0.1692    0.000134
```

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

```
                        Solutions for Fixed Effects
                     Standard
Effect       Estimate     Error       DF    t Value    Pr > |t|    Gradient
Intercept      3.0107    0.1159       76      25.98      <.0001    0.000057
```

```
                                Estimates
                                                                    Standard
                        Standard                                       Error
Label        Estimate       Error       DF    t Value    Pr > |t|        Mean        Mean
Intercept      3.0107      0.1159       76      25.98      <.0001     20.3017      2.3524

                        Tests of Covariance Parameters
                           Based on the Likelihood
Label                          DF    -2 Log Like      ChiSq    Pr > ChiSq    Note
Random Region Intercept?        1          13249    7566.93        <.0001    MI
MI: P-value based on a mixture of chi-squares.
```

```
display as result "STATA Model 2b: Empty Means, Random Intercept Model for Deaths (outcome)"
display as result "Log Link, Negative Binomial Conditional Distribution"
menbreg deaths , || region: , covariance(unstructured) intpoints(15),
        estat ic, n(77),          // get AIC and BIC equivalent to SAS
        nlcom exp(_b[_cons])      // fixed intercept in counts
        estimates store NegBin2   // save LL for LRT
        lrtest NegBin2 Poisson2   // LRT against fixed effect model
```

```
TITLE1 "SAS Model 2b: Empty Means, Random Intercept Model for Deaths (outcome)";
TITLE2 "Log Link, Negative Binomial Conditional Distribution";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region;
     MODEL deaths =  / SOLUTION LINK=LOG DIST=NEGBIN DDFM=BW CHISQ;
     RANDOM INTERCEPT / TYPE=UN SUBJECT=region;
     ESTIMATE "Intercept" intercept 1 / ILINK; * ILINK is inverse link (to un-log);
     COVTEST "Random Region Intercept?" 0 .;   * Test if G matrix (1,1)=0;
     COVTEST "Overdispersion?" . 1;            * Test if overdispersion=0 (1=Poisson);
     ODS OUTPUT CovParms=CovEmpty;             * Save random int var for pseudo-R2;
RUN; TITLE1; TITLE2;
```

```
          Fit Statistics
-2 Log Likelihood            2867.40
AIC  (smaller is better)     2873.40
AICC (smaller is better)     2873.47
BIC  (smaller is better)     2880.43
CAIC (smaller is better)     2883.43
HQIC (smaller is better)     2876.21

     Fit Statistics for Conditional
               Distribution
-2 log L(deaths | r. effects)     2645.96
Pearson Chi-Square                 374.61
Pearson Chi-Square / DF              1.07
```

The Pearson χ^2 / DF statistic is an index of fit to the conditional distribution (i.e., after including random effects). It should be close to 1 for good fit—hooray!

```
           Covariance Parameter Estimates
Cov                                Standard
Parm       Subject    Estimate       Error    Gradient
UN(1,1)    region       0.8466      0.1660    0.000055
Scale                   0.4292     0.04066    -0.00033

                        Solutions for Fixed Effects
                        Standard
Effect       Estimate       Error       DF    t Value    Pr > |t|    Gradient
Intercept      3.0493      0.1139       76      26.78      <.0001     0.00015
```

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

```
                                Estimates
                                                                    Standard
                        Standard                                       Error
Label        Estimate       Error       DF    t Value    Pr > |t|        Mean        Mean
Intercept      3.0493      0.1139       76      26.78      <.0001     21.1001      2.4025
```

```
                    Tests of Covariance Parameters
                       Based on the Likelihood

Label                          DF    -2 Log Like      ChiSq    Pr > ChiSq    Note
Random Region Intercept?        1        3052.06     184.66        <.0001    MI
Overdispersion?                 1        2939.89      72.49        <.0001    DF
DF: P-value based on a chi-square with DF degrees of freedom.
MI: P-value based on a mixture of chi-squares.
```

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(0.8466) = −1.25 to 4.85 in log counts, or 3.48 to 128.09 in counts.*

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

Model 3a: Add Fixed Slope of Between-Region UV Mean Predictor

$$\text{Level 1: } \operatorname{Log}\left(\text{Death}_{cr}\right) = \beta_{0r}$$
$$\text{Level 2: Intercept: } \beta_{0r} = \gamma_{00} + \gamma_{01}\left(\overline{\text{UV}}_{r}\right) + \text{U}_{0r}$$

```
TITLE1 "SAS Model 3a: Add Fixed Slope of Between-Region Mean UV Predictor";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region;
     MODEL deaths =  RMuv0 / SOLUTION LINK=LOG DIST=NEGBIN DDFM=BW CHISQ;
     RANDOM INTERCEPT / TYPE=UN SUBJECT=region;
     ESTIMATE "Intercept if RMuv0=-1"  intercept 1 RMuv0 -1 / ILINK;
     ESTIMATE "Intercept if RMuv0= 0"  intercept 1 RMuv0  0 / ILINK;
     ESTIMATE "Intercept if RMuv0= 1"  intercept 1 RMuv0  1 / ILINK;
     ODS OUTPUT CovParms=CovBRuv;              * Save random int var for pseudo-R2;
RUN; TITLE1;

* Calculate PseudoR2 relative to previous model 2b;
%PseudoR2G(NCov=2, CovFewer=CovEmpty, CovMore=CovBRuv);

display as result "STATA Model 3a: Add Fixed Slope of Between-Region Mean UV Predictor"
menbreg deaths c.RMuv0, || region: , covariance(unstructured) intpoints(15),
        estat ic, n(77),
        margins , at(c.RMuv0=(-1(1)1)) predict(xb) // predicted log counts
        margins , at(c.RMuv0=(-1(1)1))             // marginal predicted counts
```

Note that STATA converts log counts to counts that are marginal, not unit-specific like SAS does. So these outputs (from SAS ILINK and STATA margins without xb) will not match.

```
           Fit Statistics
-2 Log Likelihood            2854.32
AIC  (smaller is better)     2862.32
AICC (smaller is better)     2862.44
BIC  (smaller is better)     2871.70
CAIC (smaller is better)     2875.70
HQIC (smaller is better)     2866.07

     Fit Statistics for Conditional
              Distribution
-2 log L(deaths | r. effects)     2647.77
Pearson Chi-Square                 376.65
Pearson Chi-Square / DF              1.07

            Covariance Parameter Estimates
Cov                                Standard
Parm       Subject    Estimate       Error    Gradient
UN(1,1)    region       0.6895      0.1399    -0.00123
Scale                   0.4276     0.04048    0.000141
```

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Gradient
Intercept	3.0381	0.1044	75	29.10	<.0001	-0.00081
RMuv0	-0.08251	0.02187	75	-3.77	0.0003	-0.00259

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 3.081, which is a count = 20.87 (from estimates below).*

What does the main effect of region mean UV represent? *Without controlling for county UV, for every unit higher region mean UV, the log of the death count is significantly lower by 0.0825. This is the "total" between-region effect. This effect accounted for 18.56% of the level-2 region random intercept variance.*

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Mean	Standard Error Mean
Intercept if RMuv0=-1	3.1206	0.1061	75	29.41	<.0001	22.6600	2.4041
Intercept if RMuv0= 0	3.0381	0.1044	75	29.10	<.0001	20.8653	2.1785
Intercept if RMuv0= 1	2.9556	0.1073	75	27.56	<.0001	19.2127	2.0606

PsuedoR2 (% Reduction) for CovEmpty vs. CovBRuv

Name	CovParm	Subject	Estimate	StdErr	Gradient	Pseudo R2
CovEmpty	UN(1,1)	region	0.8466	0.1660	0.000055	.
CovEmpty	Scale		0.4292	0.04066	-0.00033	.
CovBRuv	UN(1,1)	region	0.6895	0.1399	-0.00123	0.18555
CovBRuv	Scale		0.4276	0.04048	0.000141	0.00367

Model 3b: Add Fixed Slope of Within-Region UV Mean Predictor

$$\text{Level 1: } \text{Log}\left(\text{Death}_{cr}\right) = \beta_{0r} + \beta_{1r}\left(\text{UV}_{cr} - \overline{\text{UV}}_{r}\right)$$

$$\text{Level 2: Intercept: } \beta_{0r} = \gamma_{00} + \gamma_{01}\left(\overline{\text{UV}}_{r}\right) + \text{U}_{0r}$$

$$\text{Within UV: } \beta_{1r} = \gamma_{10}$$

```
TITLE1 "SAS Model 3b: Add Fixed Slope of Within-Region UV Predictor";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region;
     MODEL deaths =  RMuv0 WRuv / SOLUTION LINK=LOG DIST=NEGBIN DDFM=BW CHISQ;
     RANDOM INTERCEPT / TYPE=UN SUBJECT=region;
     ESTIMATE "Contextual UV Effect"  WRuv -1  RMuv0 1;
     CONTRAST "Multivariate Wald test for Math Effects"  RMuv0 1, WRuv 1 / CHISQ;
     ESTIMATE "Intercept if WRuv=-1"  intercept 1 WRuv -1 / ILINK;
     ESTIMATE "Intercept if WRuv= 0"  intercept 1 WRuv  0 / ILINK;
     ESTIMATE "Intercept if WRuv= 1"  intercept 1 WRuv  1 / ILINK;
     ODS OUTPUT CovParms=CovWRuv;              * Save random int var for pseudo-R2;
RUN; TITLE1;

* Calculate PseudoR2 relative to previous model 3a;
%PseudoR2G(NCov=2, CovFewer=CovBRuv, CovMore=CovWRuv);

display as result "STATA Model 3b: Add Fixed Slope of Within-Region UV Predictor"
menbreg deaths c.RMuv0 c.WRuv, || region: , covariance(unstructured) intpoints(15),
        estat ic, n(77),
        lincom c.WRuv*-1 + c.RMuv0*1              // contextual UV effect
        margins , at(c.WRuv=(-1(1)1)) predict(xb) // predicted log counts
        margins , at(c.WRuv=(-1(1)1))             // marginal predicted counts
        estimates store Fixed                     // save LL for LRT
```

Fit Statistics	
-2 Log Likelihood	2851.53
AIC (smaller is better)	2861.53
AICC (smaller is better)	2861.70
BIC (smaller is better)	2873.25
CAIC (smaller is better)	2878.25
HQIC (smaller is better)	2866.21

Fit Statistics for Conditional Distribution	
-2 log L(deaths | r. effects)	2644.25
Pearson Chi-Square	378.02
Pearson Chi-Square / DF	1.08

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Gradient
UN(1,1)	region	0.6909	0.1399	0.000359
Scale		0.4236	0.04012	0.004868

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Gradient
Intercept	3.0368	0.1044	75	29.08	<.0001	-0.00085
RMuv0	-0.08265	0.02187	75	-3.78	0.0003	-0.00915
WRuv	0.1387	0.08297	273	1.67	0.0958	0.000341

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 and within-region UV exposure = 0 is 3.037, which is a count = 20.84 (from estimates below).*

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 death count is significantly lower by 0.0827. This is the "total" between-region effect. This effect is still significant after controlling for county UV (as indicated by a contextual between-region effect = −0.2213 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 death is nonsignificantly higher by 0.1387. We cannot (easily) compute a pseudo-R^2 for the residual variance, which remains a function of the mean (and thus is non-constant).*

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Mean	Standard Error Mean
Contextual UV Effect	-0.2213	0.08579	75	-2.58	0.0118	Non-est	.
Intercept if WRuv=-1	2.8981	0.1333	273	21.75	<.0001	18.1396	2.4174
Intercept if WRuv= 0	3.0368	0.1044	273	29.08	<.0001	20.8375	2.1759
Intercept if WRuv= 1	3.1754	0.1335	273	23.79	<.0001	23.9368	3.1948

Contrasts

Label	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Multivariate Wald test for Math Effects	2	273	17.08	8.54	0.0002	0.0003

PsuedoR2 (% Reduction) for CovBRuv vs. CovWRuv

Name	CovParm	Subject	Estimate	StdErr	Gradient	PseudoR2
CovBRuv	UN(1,1)	region	0.6895	0.1399	-0.00123	.
CovBRuv	Scale		0.4276	0.04048	0.000141	.
CovWRuv	UN(1,1)	region	0.6909	0.1399	0.000359	-.001999259
CovWRuv	Scale		0.4236	0.04012	0.004868	0.009483257

Model 3c: Add Random Slope of Within-Region UV Mean Predictor

$$\text{Level 1: } \text{Log}\left(\text{Death}_{cr}\right) = \beta_{0r} + \beta_{1r}\left(\text{UV}_{cr} - \overline{\text{UV}}_{r}\right)$$

$$\text{Level 2: Intercept: } \beta_{0r} = \gamma_{00} + \gamma_{01}\left(\overline{\text{UV}}_{r}\right) + \text{U}_{0r}$$

$$\text{Within UV: } \beta_{1r} = \gamma_{10} + \text{U}_{1r}$$

```
TITLE1 "SAS Model 3c: Add Random Slope of Within-Region UV Predictor";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region;
     MODEL deaths =  RMuv0 WRuv / SOLUTION LINK=LOG DIST=NEGBIN DDFM=BW CHISQ;
     RANDOM INTERCEPT WRuv / TYPE=UN SUBJECT=region;
     ESTIMATE "Contextual UV Effect"  WRuv -1  RMuv0 1;
     COVTEST "Random WRuv Slope?" . 0 0 .; * Leave (1,1) and OD, test if (2,1) and (2,2) =0;
     ODS OUTPUT CovParms=CovRandWRuv;      * Save random variances for pseudo-R2;
RUN; TITLE1;

display as result "STATA Model 3c: Add Random Slope of Within-Region UV Predictor"
menbreg deaths c.RMuv0 c.WRuv, || region: c.WRuv, covariance(unstructured) intpoints(15),
        estat ic, n(77),
        lincom c.WRuv*-1 + c.RMuv0*1 // contextual UV effect
        estimates store Random       // save LL for LRT
        lrtest Random Fixed          // LRT against fixed effect model
```

```
           Fit Statistics
-2 Log Likelihood            2843.94
AIC  (smaller is better)     2857.94
AICC (smaller is better)     2858.27
BIC  (smaller is better)     2874.35
CAIC (smaller is better)     2881.35
HQIC (smaller is better)     2864.50

     Fit Statistics for Conditional
               Distribution
-2 log L(deaths | r. effects)     2605.18
Pearson Chi-Square                 355.45
Pearson Chi-Square / DF              1.01

            Covariance Parameter Estimates
Cov                                Standard
Parm       Subject    Estimate       Error    Gradient
UN(1,1)    region       0.7021      0.1402    0.001919
UN(2,1)    region      -0.1150      0.1072     0.00042
UN(2,2)    region       0.1297     0.07497    -0.00187
Scale                   0.3916     0.03878    -0.00048

                        Solutions for Fixed Effects
                        Standard
Effect       Estimate       Error       DF    t Value    Pr > |t|    Gradient
Intercept      3.0240      0.1047       75      28.89      <.0001    -0.00121
RMuv0         -0.08425     0.02168       75      -3.89      0.0002    -0.00218
WRuv           0.1063      0.1036      273       1.03      0.3059    -0.00245

                                 Estimates
                                    Standard
Label                    Estimate       Error       DF    t Value    Pr > |t|
Contextual UV Effect      -0.1906      0.1072       75      -1.78      0.0795
```

```
                    Tests of Covariance Parameters
                       Based on the Likelihood
Label                  DF    -2 Log Like      ChiSq    Pr > ChiSq    Note
Random WRuv Slope?      2        2851.53       7.59        0.0142    MI
MI: P-value based on a mixture of chi-squares.
```

Does the level-2 random effect of level-1 within-region UV improve model fit?
Yes, −2ΔLL(mixture of df=1 and df=2) = 7.59, p = .0142

Calculate a 95% random effect confidence interval for the within-region UV slope:
*CI = fixed effect ± 1.96*SQRT(random slope variance)*
*CI = 0.1063 ± 1.96*SQRT(0.1297) = −0.60 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 4a: Add National as a Control Predictor—Main Effects Only

$$\text{Level 1: } \operatorname{Log}\left(\text{Death}_{cr}\right) = \beta_{0r} + \beta_{1r}\left(\text{UV}_{cr} - \overline{\text{UV}}_{r}\right)$$

$$\text{Level 2: Intercept: } \beta_{0r} = \gamma_{00} + \gamma_{01}\left(\overline{\text{UV}}_{r}\right) + \sum_{n=1}^{7}\gamma_{0n+1}\left(\text{Nation}_{r} = n\right) + U_{0r}$$

$$\text{Within UV: } \beta_{1r} = \gamma_{10} + U_{1r}$$

```
TITLE1 "SAS Model 4a: Nation as Control Predictor";
TITLE2 "Main Effect of Nation Only";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region nation;
     MODEL deaths =  RMuv0 WRuv nation / SOLUTION LINK=LOG DIST=NEGBIN DDFM=BW CHISQ;
     RANDOM INTERCEPT WRuv / TYPE=UN SUBJECT=region;
     ODS OUTPUT CovParms=CovNatMain;   * Save random variances for pseudo-R2;
RUN; TITLE1;

* Calculate PseudoR2 relative to previous model 3c;
%PseudoR2G(NCov=4, CovFewer=CovRandWRuv, CovMore=CovNatMain);

display as result "STATA Model 4a: Nation as Control Predictor"
display as result "Main Effect of Nation Only"
menbreg deaths c.RMuv0 c.WRuv i.nation, ///
           || region: c.WRuv, covariance(unstructured) intpoints(15),
        estat ic, n(77),
        contrast i.nation  // Omnibus test of df=7 nation on intercept
```

```
          Fit Statistics
-2 Log Likelihood           2744.64
AIC  (smaller is better)    2772.64
AICC (smaller is better)    2773.89
BIC  (smaller is better)    2805.45
CAIC (smaller is better)    2819.45
HQIC (smaller is better)    2785.76

     Fit Statistics for Conditional
               Distribution
-2 log L(deaths | r. effects)     2628.62
Pearson Chi-Square                 442.23
Pearson Chi-Square / DF              1.26
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Gradient
UN(1,1)	region	0.1252	0.03690	0.002358
UN(2,1)	region	-0.08597	0.04294	0.00047
UN(2,2)	region	0.1190	0.06591	0.001261
Scale		0.3872	0.03738	0.001886

Solutions for Fixed Effects

Effect	nation	Estimate	Standard Error	DF	t Value	Pr > |t|	Gradient
Intercept		3.5218	0.2894	68	12.17	<.0001	0.000494
RMuv0		-0.06071	0.02433	68	-2.50	0.0150	0.003772
WRuv		0.1311	0.1009	273	1.30	0.1950	-0.00076
nation	1	-0.01705	0.4000	68	-0.04	0.9661	-0.00386
nation	2	0.9069	0.3201	68	2.83	0.0061	0.001817
nation	3	-0.04006	0.3895	68	-0.10	0.9184	-0.00276
nation	4	-0.7702	0.3205	68	-2.40	0.0190	0.000468
nation	5	-0.5208	0.3037	68	-1.72	0.0909	-0.00151
nation	6	-0.5767	0.3766	68	-1.53	0.1303	-0.00011
nation	7	-2.9282	0.3754	68	-7.80	<.0001	0.006876
nation	9	0	.	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
RMuv0	1	68	6.23	6.23	0.0126	0.0150
WRuv	1	273	1.69	1.69	0.1939	0.1950
nation	7	68	179.19	25.60	<.0001	<.0001

PsuedoR2 (% Reduction) for CovRandWRuv vs. CovNatMain

Name	CovParm	Subject	Estimate	StdErr	Gradient	Pseudo R2
CovRandWRuv	UN(1,1)	region	0.7021	0.1402	0.001919	.
CovRandWRuv	UN(2,2)	region	0.1297	0.07497	-0.00187	.
CovRandWRuv	Scale		0.3916	0.03878	-0.00048	.
CovNatMain	UN(1,1)	region	0.1252	0.03690	0.002358	0.82163
CovNatMain	UN(2,2)	region	0.1190	0.06591	0.001261	0.08248
CovNatMain	Scale		0.3872	0.03738	0.001886	0.01125

Do the level-2 fixed main effects of nation improve model fit?
Yes, $F(7,68) = 25.60$, $p < .0001$ (or yes, $\chi^2(7) = 179.19$, $p < .0001$), pseudo-$R^2 = .822$

Model 4b: Add National as a Control Predictor—Interactions with BR and WR UV Slopes

Level 1: $\text{Log}\left(\text{Death}_{cr}\right) = \beta_{0r} + \beta_{1r}\left(\text{UV}_{cr} - \overline{\text{UV}}_{r}\right)$

Level 2: Intercept: $\beta_{0r} = \gamma_{00} + \gamma_{01}\left(\overline{\text{UV}}_{r}\right) + \sum_{n=1}^{7}\gamma_{0n+1}\left(\text{Nation}_{r} = n\right) + \sum_{n=1}^{7}\gamma_{0n+8}\left(\overline{\text{UV}}_{r}\right)\left(\text{Nation}_{r} = n\right) + U_{0r}$

Within UV: $\beta_{1r} = \gamma_{10} + \sum_{n=1}^{7}\gamma_{0n}\left(\text{Nation}_{r} = n\right) + U_{1r}$

```
TITLE1 "SAS Model 4b: Nation as Control Predictor";
TITLE2 "Add Nation Interactions with Between-Region UV and Within-Region UV";
PROC GLIMMIX DATA=work.skincancer NOCLPRINT NAMELEN=100 METHOD=QUAD(QPOINTS=15) GRADIENT;
     CLASS region nation;
     MODEL deaths =  RMuv0 WRuv nation nation*RMuv0 nation*WRuv
                    / SOLUTION LINK=LOG DIST=NEGBIN DDFM=BW CHISQ;
     RANDOM INTERCEPT WRuv / TYPE=UN SUBJECT=region;
```

```
     ODS OUTPUT CovParms=CovNatUV;   * Save random variances for pseudo-R2;
     ESTIMATE "Between-Region UV For Nation 1" RMuv0 1 RMuv0*nation 1 0 0 0 0 0 0 0;
     ESTIMATE "Between-Region UV For Nation 2" RMuv0 1 RMuv0*nation 0 1 0 0 0 0 0 0;
     ESTIMATE "Between-Region UV For Nation 3" RMuv0 1 RMuv0*nation 0 0 1 0 0 0 0 0;
     ESTIMATE "Between-Region UV For Nation 4" RMuv0 1 RMuv0*nation 0 0 0 1 0 0 0 0;
     ESTIMATE "Between-Region UV For Nation 5" RMuv0 1 RMuv0*nation 0 0 0 0 1 0 0 0;
     ESTIMATE "Between-Region UV For Nation 6" RMuv0 1 RMuv0*nation 0 0 0 0 0 1 0 0;
     ESTIMATE "Between-Region UV For Nation 7" RMuv0 1 RMuv0*nation 0 0 0 0 0 0 1 0;
     ESTIMATE "Between-Region UV For Nation 9" RMuv0 1 RMuv0*nation 0 0 0 0 0 0 0 1;
     ESTIMATE "Within-Region UV For Nation 1" WRuv 1  WRuv*nation 1 0 0 0 0 0 0 0;
     ESTIMATE "Within-Region UV For Nation 2" WRuv 1  WRuv*nation 0 1 0 0 0 0 0 0;
     ESTIMATE "Within-Region UV For Nation 3" WRuv 1  WRuv*nation 0 0 1 0 0 0 0 0;
     ESTIMATE "Within-Region UV For Nation 4" WRuv 1  WRuv*nation 0 0 0 1 0 0 0 0;
     ESTIMATE "Within-Region UV For Nation 5" WRuv 1  WRuv*nation 0 0 0 0 1 0 0 0;
     ESTIMATE "Within-Region UV For Nation 6" WRuv 1  WRuv*nation 0 0 0 0 0 1 0 0;
     ESTIMATE "Within-Region UV For Nation 7" WRuv 1  WRuv*nation 0 0 0 0 0 0 1 0;
     ESTIMATE "Within-Region UV For Nation 9" WRuv 1  WRuv*nation 0 0 0 0 0 0 0 1;
RUN; TITLE1;

* Calculate PseudoR2 relative to previous model 4a;
%PseudoR2G(NCov=4, CovFewer=CovNatMain, CovMore=CovNatUV);

* Calculate PseudoR2 for total nation effect;
%PseudoR2G(NCov=4, CovFewer=CovRandWRuv, CovMore=CovNatUV);

display as result "STATA Model 4b: Nation as Control Predictor"
display as result "Add Nation Interactions with Between-Region UV and Within-Region UV"
menbreg deaths c.RMuv0 c.WRuv i.nation i.nation i.nation#c.RMuv0 i.nation#c.WRuv, ///
          || region: c.WRuv, covariance(unstructured) intpoints(15),
        estat ic, n(77),
        contrast i.nation          // Omnibus test of df=7 nation on intercept
        contrast i.nation#c.RMuv0  // Omnibus test of df=7 nation on RMuv0 slope
        lincom  c.RMuv0*1 + i1.nation#c.RMuv0*1 // RMuv0 fixed slope per nation
        lincom  c.RMuv0*1 + i2.nation#c.RMuv0*1
        lincom  c.RMuv0*1 + i3.nation#c.RMuv0*1
        lincom  c.RMuv0*1 + i4.nation#c.RMuv0*1
        lincom  c.RMuv0*1 + i5.nation#c.RMuv0*1
        lincom  c.RMuv0*1 + i6.nation#c.RMuv0*1
        lincom  c.RMuv0*1 + i7.nation#c.RMuv0*1
        lincom  c.RMuv0*1 + i9.nation#c.RMuv0*1
        contrast i.nation#c.WRuv  // Omnibus test of df=7 nation on WRuv slope
        lincom  c.WRuv*1 + i1.nation#c.WRuv*1 // WRuv fixed slope per nation
        lincom  c.WRuv*1 + i2.nation#c.WRuv*1
        lincom  c.WRuv*1 + i3.nation#c.WRuv*1
        lincom  c.WRuv*1 + i4.nation#c.WRuv*1
        lincom  c.WRuv*1 + i5.nation#c.WRuv*1
        lincom  c.WRuv*1 + i6.nation#c.WRuv*1
        lincom  c.WRuv*1 + i7.nation#c.WRuv*1
        lincom  c.WRuv*1 + i9.nation#c.WRuv*1
```

```
          Fit Statistics
-2 Log Likelihood            2715.72
AIC  (smaller is better)     2771.72
AICC (smaller is better)     2776.76
BIC  (smaller is better)     2837.35
CAIC (smaller is better)     2865.35
HQIC (smaller is better)     2797.97

     Fit Statistics for Conditional
               Distribution
-2 log L(deaths | r. effects)     2626.19
Pearson Chi-Square                 459.69
Pearson Chi-Square / DF              1.31
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Gradient
UN(1,1)	region	0.06546	0.02735	0.040374
UN(2,1)	region	-0.04740	0.04384	0.042663
UN(2,2)	region	0.1209	0.06948	0.016882
Scale		0.3752	0.03643	0.004667

Solutions for Fixed Effects

Effect	nation	Estimate	Standard Error	DF	t Value	Pr > |t|	Gradient
Intercept		8.1930	2.3914	61	3.43	0.0011	0.016163
RMuv0		1.0699	0.5698	61	1.88	0.0652	-0.04177
WRuv		-3.5955	1.7130	266	-2.10	0.0368	-0.0066
nation	1	-14.2489	4.8006	61	-2.97	0.0043	0.001936
nation	2	-3.9113	2.4184	61	-1.62	0.1110	-0.02101
nation	3	-8.1571	5.7260	61	-1.42	0.1594	0.038949
nation	4	-5.4034	2.3932	61	-2.26	0.0275	-0.01437
nation	5	-3.8954	2.4332	61	-1.60	0.1146	-0.0185
nation	6	-5.0714	2.3983	61	-2.11	0.0386	-0.02087
nation	7	-3.5765	3.3157	61	-1.08	0.2850	0.046375
nation	9	0	.	.	.	.	.
RMuv0*nation	1	-4.4120	1.5422	61	-2.86	0.0058	0.00691
RMuv0*nation	2	-1.1666	0.5807	61	-2.01	0.0490	0.016641
RMuv0*nation	3	-1.7509	1.0897	61	-1.61	0.1133	-0.06055
RMuv0*nation	4	-1.1447	0.5712	61	-2.00	0.0495	-0.00181
RMuv0*nation	5	-0.8501	0.5773	61	-1.47	0.1460	-0.00855
RMuv0*nation	6	-1.1570	0.5705	61	-2.03	0.0469	-0.0332
RMuv0*nation	7	-0.3269	0.7290	61	-0.45	0.6554	0.025167
RMuv0*nation	9	0	.	.	.	.	.
WRuv*nation	1	2.8178	2.0104	266	1.40	0.1622	0.02152
WRuv*nation	2	4.0680	1.7476	266	2.33	0.0207	0.008597
WRuv*nation	3	2.4594	1.8616	266	1.32	0.1876	-0.03559
WRuv*nation	4	3.6824	1.7210	266	2.14	0.0333	0.010579
WRuv*nation	5	3.8786	1.7323	266	2.24	0.0260	0.005555
WRuv*nation	6	3.6630	1.7217	266	2.13	0.0343	0.010102
WRuv*nation	7	4.1527	1.8224	266	2.28	0.0235	0.025454
WRuv*nation	9	0	.	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
RMuv0	1	61	1.45	1.45	0.2291	0.2338
WRuv	1	266	3.15	3.15	0.0757	0.0769
nation	7	61	35.62	5.09	<.0001	0.0001
RMuv0*nation	7	61	22.77	3.25	0.0019	0.0053
WRuv*nation	7	266	10.57	1.51	0.1586	0.1639

Do the level-2 fixed simple main effects of nation improve model fit?
Yes, $F(7,61) = 5.09$, $p < .0001$ (or yes, $\chi^2(7) = 35.62$, $p < .0001$)

Do the level-2 fixed interaction effects of nation by region mean UV improve model fit?
Yes, $F(7,61) = 3.25$, $p = .0053$ (or yes, $\chi^2(7) = 22.77$, $p < .0001$)

Do the cross-level fixed interaction effects of nation by within-region UV improve model fit?
No, $F(7,266) = 1.51$, $p = .1639$ (or no, $\chi^2(7) = 10.57$, $p = .1586$)

However—given that these are omnibus interactions with df=7, it is still informative to see what the UV effects are for each nation, as follows:

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Between-Region UV For Nation 1	-3.3421	1.4340	61	-2.33	0.0231
Between-Region UV For Nation 2	-0.09667	0.1130	61	-0.86	0.3957
Between-Region UV For Nation 3	-0.6810	0.9295	61	-0.73	0.4666
Between-Region UV For Nation 4	-0.07480	0.03788	61	-1.97	0.0529
Between-Region UV For Nation 5	0.2198	0.09673	61	2.27	0.0266
Between-Region UV For Nation 6	-0.08711	0.02782	61	-3.13	0.0027
Between-Region UV For Nation 7	0.7430	0.4543	61	1.64	0.1071
Between-Region UV For Nation 9	1.0699	0.5698	61	1.88	0.0652
Within-Region UV For Nation 1	-0.7777	1.0493	266	-0.74	0.4592
Within-Region UV For Nation 2	0.4724	0.3451	266	1.37	0.1722
Within-Region UV For Nation 3	-1.1361	0.7245	266	-1.57	0.1180
Within-Region UV For Nation 4	0.08686	0.1595	266	0.54	0.5866
Within-Region UV For Nation 5	0.2831	0.2500	266	1.13	0.2584
Within-Region UV For Nation 6	0.06744	0.1925	266	0.35	0.7263
Within-Region UV For Nation 7	0.5572	0.6240	266	0.89	0.3727
Within-Region UV For Nation 9	-3.5955	1.7130	266	-2.10	0.0368

PsuedoR2 (% Reduction) for CovNatMain vs. CovNatUV

Name	CovParm	Subject	Estimate	StdErr	Gradient	PseudoR2
CovNatMain	UN(1,1)	region	0.1252	0.03690	0.002358	.
CovNatMain	UN(2,2)	region	0.1190	0.06591	0.001261	.
CovNatMain	Scale		0.3872	0.03738	0.001886	.
CovNatUV	UN(1,1)	region	0.06546	0.02735	0.040374	0.47723
CovNatUV	UN(2,2)	region	0.1209	0.06948	0.016882	-0.01596
CovNatUV	Scale		0.3752	0.03643	0.004667	0.03113

PsuedoR2 (% Reduction) for CovRandWRuv vs. CovNatUV

Name	CovParm	Subject	Estimate	StdErr	Gradient	Pseudo R2
CovRandWRuv	UN(1,1)	region	0.7021	0.1402	0.001919	.
CovRandWRuv	UN(2,2)	region	0.1297	0.07497	-0.00187	.
CovRandWRuv	Scale		0.3916	0.03878	-0.00048	.
CovNatUV	UN(1,1)	region	0.06546	0.02735	0.040374	0.90675
CovNatUV	UN(2,2)	region	0.1209	0.06948	0.016882	0.06783
CovNatUV	Scale		0.3752	0.03643	0.004667	0.04203

Sample Results Section using SAS Output [notes about what to add]

The extent to which UV exposure could predict death counts 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 marginal maximum likelihood (MML) using adaptive Gaussian quadrature with 15 points of integration per random effect dimension in SAS GLIMMIX. 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 *t*-test of the ratio of each estimate to its standard error using between–within 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.

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 binomial model, the two-level negative binomial model had significantly better fit, $-2\Delta LL(1) = 184.66$, $p < .001$, 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, $-2\Delta LL(1) = 72.49$, $p < .001$, indicating significant over-dispersion of the conditional variance (i.e., of the level-1 residuals). The ratio of the Pearson χ^2 to degrees of freedom was 1.07, indicating very

close fit of the outcome to the target distribution (in which the ratio = 1). A 95% random effects confidence interval, calculated as fixed intercept ± 1.96*SQRT(random intercept variance), revealed that 95% of the regions were predicted to have death counts between 3.48 and 128.09. The fixed intercept estimate for the log 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 group-mean-centering, in which the level-1 predictor was calculated 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 effect of region mean UV exposure was first added to the model. The fixed intercept indicated that the log death count for a county in a region with a random intercept = 0 and region mean UV exposure = 0 was 3.0381, or count = 20.865. The total between-region effect of UV exposure indicated that for every unit higher region mean UV, the log of the death count was significantly lower by 0.0825, which accounted for 18.56% of the level-2 region random intercept variance. Next, the effect of group-mean-centered within-region UV expsoure was added to the model. The fixed intercept indicated that the log death count for a county in a region with a random intercept = 0, region mean UV exposure = 0, and within-region UV exposure = 0 was 3.037, or a count = 20.838. The total 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 death was nonsignificantly higher by 0.1387. The level-2 between-region UV effect of −0.083 was significantly smaller 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, −2ΔLL(mixture of df =1 and df=2) = 7.59, p = .014, 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 −0.60 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 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, $F(7,61) = 5.09$, $p < .0001$, and in the between-region UV slopes, $F(7,61) = 3.25$, $p = .0053$, but not in the within-region UV slopes, $F(7,266) = 1.51$, $p = .1639$. Model-predicted simple slopes for each region were then requested via ESTIMATE statements. The between-region slopes for UV exposure were significantly positive in one nation, nonsignificant in five 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]