Local Control Regression: Comparison of Local Measures

Local Control strategy is ideal for observational studies where:

- **outcome Y-variables are continuous measures,**
- **several pre-treatment X-confounder covariates (likely to be predictive of Y) are available, and**
- **the number of experimental units (patients, subjects, etc.) is as large as possible.**

Especially when the "Treatment" indicator variable has many ordered levels and/or is a continuous measure, it just makes common sense to do local (within X-space cluster) analyses by regressing the Y-outcome upon the Treatment variable. One obvious advantage is then that all clusters of size greater than 2 tend to be "informative." Meanwhile, a disadvantage is that having the within-cluster range of observed Treatment levels being large is highly desirable; a cluster becomes "uninformative" when this range is zero!

The fitted regression coefficient (slope) is then of the form:

Beta-hat = (Correlation of Y with T) * (Std. Dev. of Y) / (Std. Dev. of T).

Neither Beta-hat nor the Correlation of interest are properly defined when (Std. Dev. of T) = 0.

While the Correlation between Y and T is independent of the scales of measurement for Y and T, Beta-hat is not. And the p-Value (observed significance level) of Beta-hat is, by definition, the same as that of the Correlation of Y with T. Similarly, the R-square Goodness-of-Fit measure is just (Correlation of Y with T)^2.

The Y-intercept at T = 0 is of interest in radiation studies because the "Linear No-Threshold" assumption requires a Zero intercept. However, T = 0 is frequently outside of the observed range of (positive) T values. On the other hand, the fitted regression line always passes through Y = Y-bar at T = T-bar ...a pair of values always guaranteed to be within the observed within-cluster ranges of both Y and T. Luckily, at least for our Radon case-study, it appears that Y = Lung Cancer Mortality is best regressed upon T = Ln[Rn]. T = 0 then corresponds to a Radon level of 1 pCi/m^3, which is within the observed T-range for 48 of the 50 X-covariate clusters that we think are most relevant.

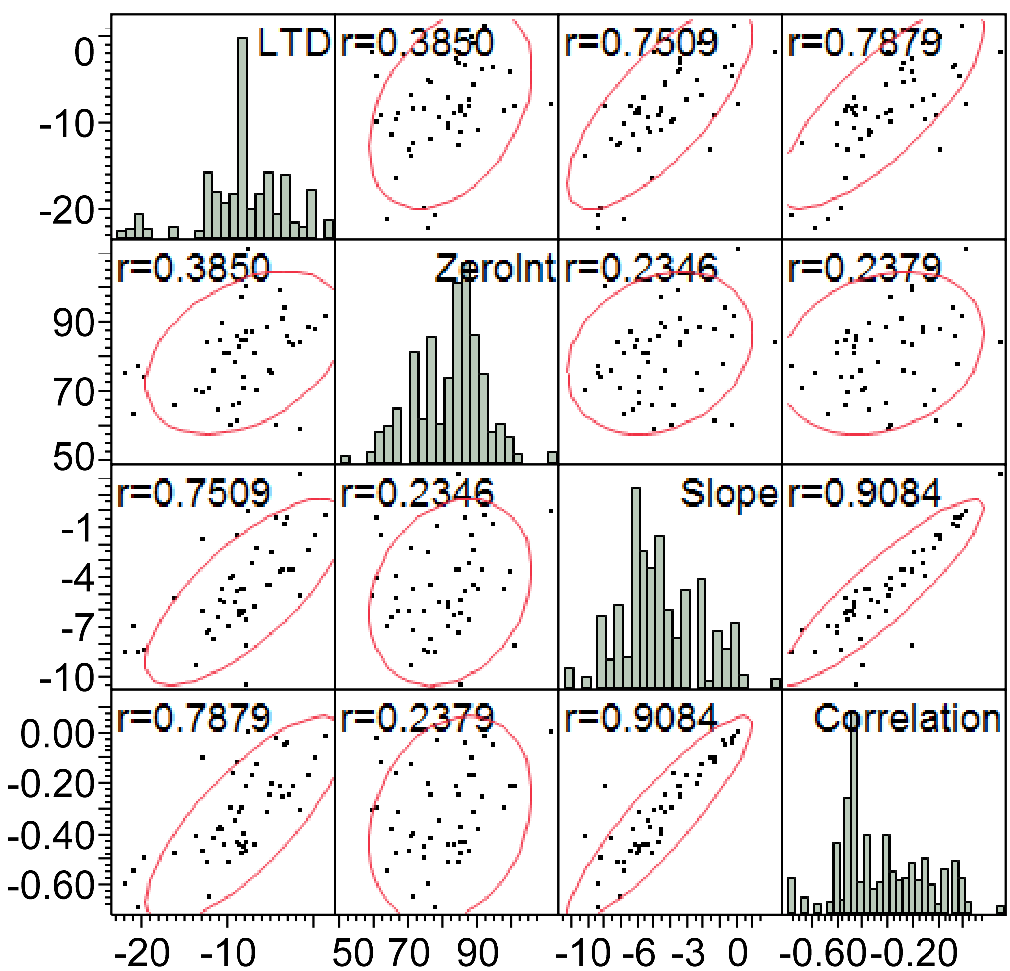
Correlations between Local Measures:

	LTD	ZeroInt	Slope	Correlation
LTD	1.0000	0.3850	0.7509	0.7879
ZeroInt	0.3850	1.0000	0.2346	0.2379
Slope = Beta-hat	0.7509	0.2346	1.0000	0.9084
Correlation	0.7879	0.2379	0.9084	1.0000

Correlations are estimated by REML method; 2,881 values except for 2,870 LTD values ==> these inter-correlations are thus weighted by Cluster Size (measured in number of US Counties).

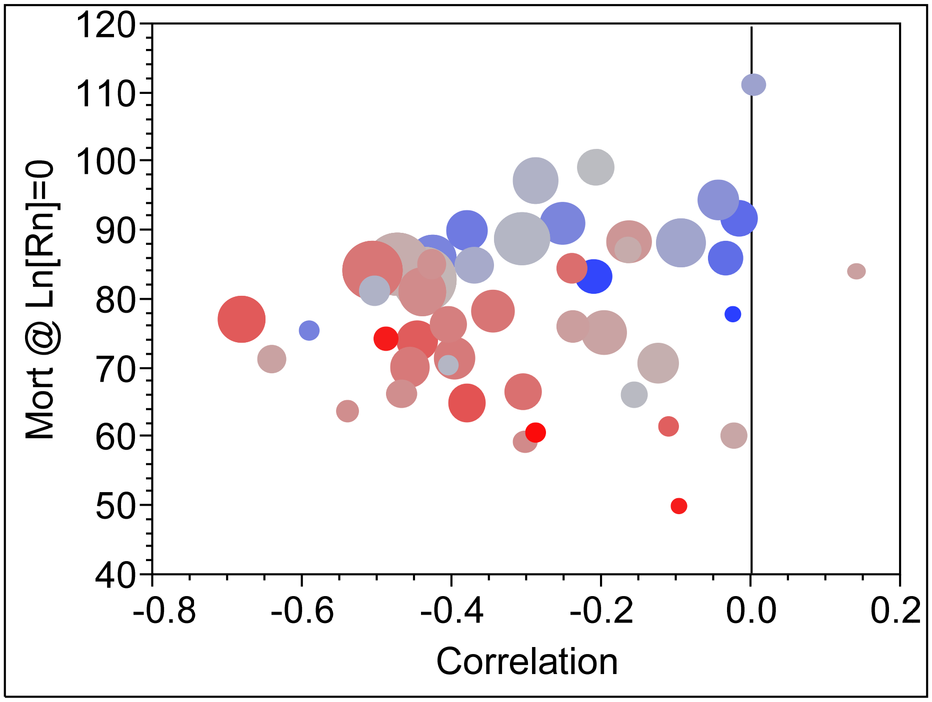
Confidence Intervals

Variable	by Variable	Correlation	Lower 95%	Upper 95%
ZeroInt	LTD	0.3850	0.3533	0.4157
Slope	LTD	0.7509	0.7345	0.7664
Slope	ZeroInt	0.2346	0.1997	0.2688
Correlation	LTD	0.7879	0.7736	0.8014
Correlation	ZeroInt	0.2379	0.2031	0.2721
Correlation	Slope	0.9084	0.9018	0.9146



Bubble Plot of ZeroInt vs Correlation for 50 LC clusters

- **Area of Circle proportional to Cluster Size in # of US Counties**
- **Color of Circle indicates Ln[Rn] on a Blue-Gray-Red scale**
  - **The more Red a Circle, the Higher the average Radon level**
  - **The more Blue a Circle, the Lower the average Radon level**



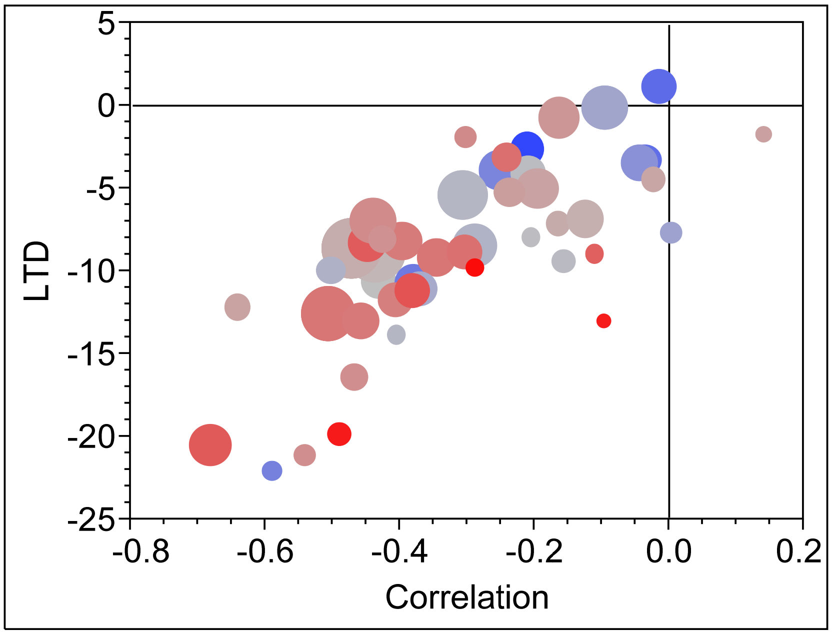
Note that the Cluster (#16) with the most positive local Correlation (+0.14) has a Mortality Intercept of 84.1 and contains only 13 counties.

Meanwhile, the other Cluster (#15) with a positive Correlation (+0.069) has the highest Mortality Intercept (111.3) and contains 23 counties.

Cluster (#10) is Uninformative about its LTD, has a negative local Correlation (-0.025), a Mortality Intercept of 78, and contains only 11 counties with low Average Ln[Rn] (-0.39).

Bubble Plot of LTD vs Correlation for 49 <u>Informative</u> LC clusters

- **Area of Circle proportional to Cluster Size in # of US Counties**
- **Color of Circle indicates Ln[Rn] on a Blue-Gray-Red scale**
  - **The more Red a Circle, the Higher the average Radon level**
  - **The more Blue a Circle, the Lower the average Radon level**



Note that the Cluster (#16) with the most positive Correlation (+0.14) has a negative LTD (-1.76) and contains only 13 counties.

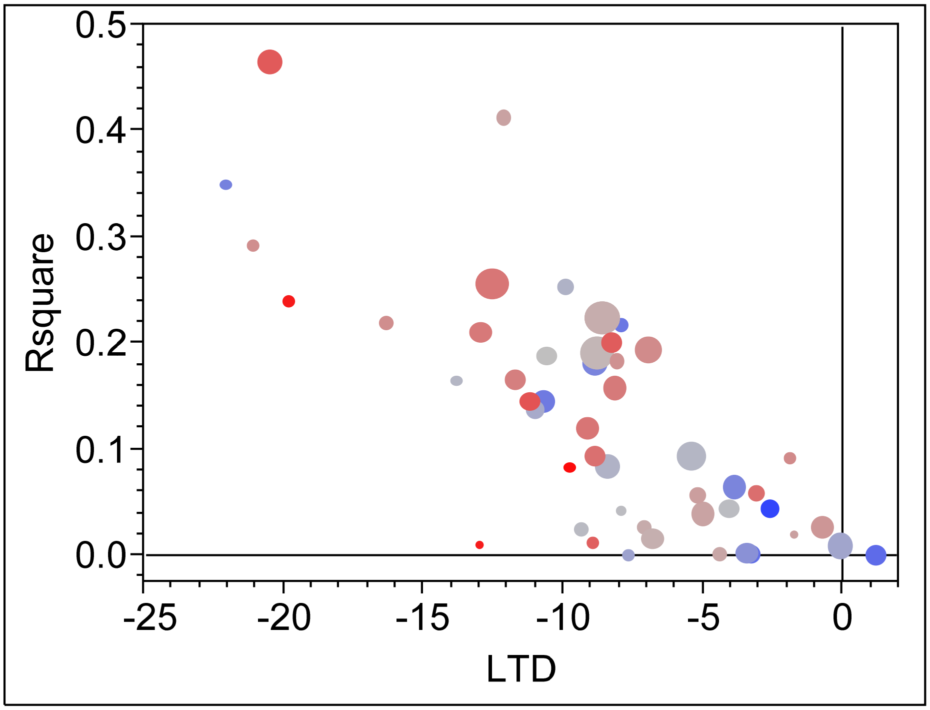
Meanwhile, the other Cluster (#15) with a positive Correlation (+0.069) has a strongly negative LTD (-7.69) and contains 23 counties.

Similarly, the Cluster (#2) with a positive LTD (+1.18) has a negative Correlation (-0.015) and 53 of its 57 counties have low average Radon.

The "uninformative" Cluster (#10) has a negative Correlation (-0.25) and contains 11 counties ...but <u>is not displayed here</u> because its LTD is a missing value.

Bubble Plot of Rsquare vs LTD for 49 Informative LC clusters

- **Area of Circle proportional to Cluster Size in # of US Counties**
- **Color of Circle indicates Ln[Rn] on a Blue-Gray-Red scale**



This plot provides an excellent summary of the strong agreement between results from two alternative Local Control (LC) analyses of the effects of Indoor Radon data on Lung Cancer Mortality:

- **When Radon "treatment" level is viewed as dichotomous (High minus Low), almost all of the resulting Local Treatment Differences (LTDs) in Mortality (the horizontal co-ordinates) are Negative.**

- **When Radon "treatment" level is viewed as continuous (log scale), almost all of the resulting within-cluster Correlations of Radon with Mortality are also Negative. The vertical co-ordinates displayed here are the (positive) Squares of these correlations = Local values for the Coefficient of Determination (Goodness-of-Fit to a strictly Log-Linear Relationship.)**

In summary, for 2,881 US Counties, increases in Indoor Radon level are consistently associated with decreases in Lung Cancer Mortality. This is totally inconsistent with current State regulations of Radon level as well as EPA policy that prevalent (low) levels of Indoor Radon cause Lung Cancer Mortality.