Hierarchical Modeling of the association between water temperature and water lead level

Gérard NGUETA, BS, MSc, MPH
PhD Candidate
CHUQ Research Center,
Public Health Research Unit;
Faculty of Medicine, Laval University (Canada)

Gerard.Ngueta@crchuq.ulaval.ca

➔Drinking water : source of Lead exposure:

- Relative contribution of water is more important due to 'worst' primary prevention
- 63-76% of Lead poisoning (BLLs ≥ 10 µg/dL) can be attributed to water lead concentration (Watt et al, 1996)
- Most valuable predictor for BLLs given the new CDC guideline (CDC, 2012)
- High-risk Homes: The problem of Lead pipes

Gerard.Ngueta@crchuq.ulaval.ca

➔Association between water temperature and water lead concentration

- Lead solubility increases with Water temperature (Schock, 1990)
- Water from Hot-faucets contains more lead than cold-water (Bryant, 2004; Mesch et al, 1996)
- 1°C in water temperature is associated with 5% increase in lead concentrations in flushed water (Cartier et al, 2011)
- Proportion of homes with high WLLs is strongly associated with Water temperature (Massey & Steele, 2012)

Gerard.Ngueta@crchuq.ulaval.ca

→ Limitations of previous studies

- All previous studies used standard linear regression (OLS) or discrete data analysis approach
- None of published studies adjusted for Neighborhood-level variables
- Factors affecting the transfer of lead from plumbing systems may vary spatially:
  - Different treatment plant (especially true in large metropolitans)
  - Population density AND/OR presence of activity centers may vary from community to community

Gerard.Ngueta@crchuq.ulaval.ca

➔My hypothesis and main objective of this study

Hypothesis

Neighborhood cluster analysis would be more appropriate than standard linear regression to describe the association between water lead levels and factors affecting lead leaching from pipes.

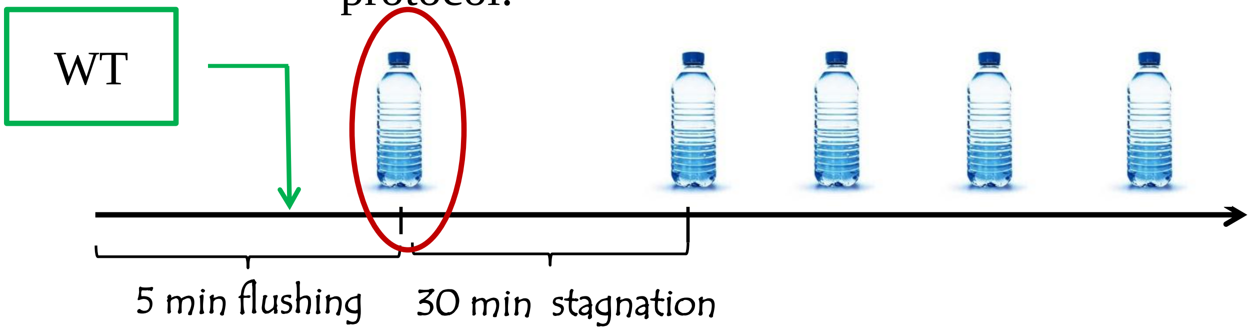
Main objective

To determine the association between water temperature (after 3 minutes of flushing) and water lead concentration in the first-draw (flushed) sample by using **neighborhood cluster analysis**

Gerard.Ngueta@crchuq.ulaval.ca

→ Data from Montreal Lead Study

- Undertaken in 2009 and firstly designed to assess lead exposure in 1-5 years old children living in Montreal city (INSPQ, 2011)
- N= **313** included homes from four old neighborhoods
- Water sampling from September, 10 (2009) to March, 27 (2010) according to U.S. EPA protocol:



Gerard.Ngueta@crchuq.ulaval.ca

→ Statistical analysis

- Shapiro Wilk test → To assess gaussian distribution of outcome
- Standard linear regression and Hierachical linear model were consecutively used for investigate the association between water temperature and water lead levels

Standard Linear regression

$$\mathbf{WLL} = \beta_{\mathbf{0}} + \beta_{\mathbf{1}}*\mathbf{WT} + \sum\beta_{\mathbf{k}}*X_{\mathbf{k}} + \varepsilon$$

Gerard.Ngueta@crchuq.ulaval.ca

➔ Statistical analysis

- Shapiro Wilk test → To assess gaussian distribution of outcome
- Standard linear regression and Hierachical linear model were consecutively used for investigate the association between water temperature and water lead levels

Hierarchical Linear modeling

Level 2: neighborhood (j)

Level 1: Household (i)

$$Y_{ij} = [X_{ij}]*\beta + [Z_{ij}]*b_j + e_{ij}$$

Gerard.Ngueta@crchuq.ulaval.ca

➔Statistical analysis

Variance-covariance matrix

```
Proc Mixed data=  Covtest Noclprint=10 Method=REML;
Class ID neighborhood X1 X2;
Model Y= Maximum /s chisq;
Random intercept/Subject=neighborhood Type= ;
Run;
```

UN, CS, CSH, TOEP, Ar(1), Arh(1)

Gerard.Ngueta@crchuq.ulaval.ca

Table 2 : Unadjusted regression coefficients based on hierarchical linear modeling of the relation between water temperature and Log (WLLF5)

	Fixed regression coefficients			Random regression coefficients			
	Estimate	SE(Estimate)	p-value	Intercept variance	Random slope variance	Residual variance	-2Log(Likelihood)
'Naive' regression model[a]							
Intercept	-0.7666	0.1266	<0.0001			1.5880	960.4
Water temperature	0.0579	0.0089	<0.0001				

-

-

-

Gerard.Ngueta@crchuq.ulaval.ca

Table 2 : Unadjusted regression coefficients based on hierarchical linear modeling of the relation between water temperature and Log (WLLF5)

	Fixed regression coefficients			Random regression coefficients			
	Estimate	SE(Estimate)	p-value	Intercept variance	Random slope variance	Residual variance	-2Log(Likelihood)
'Naive' regression model[a]							
Intercept	-0.7666	0.1266	<0.0001			1.5880	960.4
Water temperature	0.0579	0.0089	<0.0001				

Table 3 : Estimated regression coefficients based on 'standard' linear modeling of the association between water temperature and Log(WLLF5)

	Fixed regression coefficients				
	Estimate	SE(Estimate)	p-value	Adjusted R-Square	Root Mean square error
Model 1[a]					
Intercept	-0.7722	0.1238	<0.0001	0.1224	1.2616
Water temperature	0.0576	0.0088	<0.0001		

As expected !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

Gerard.Ngueta@crchuq.ulaval.ca

Table 2 : Unadjusted regression coefficients based on hierarchical linear modeling of the relation between water temperature and Log (WLLF5)

	Fixed regression coefficients			**Random regression coefficients**			
	Estimate	SE(Estimate)	p-value	Intercept variance	Random slope variance	Residual variance	-2Log(Likelihood)
'Naive' regression model[a]							
Intercept	-0.7666	0.1266	<0.0001			1.5880	960.4
Water temperature	0.0579	0.0089	<0.0001				
Model 1[b]							
Intercept	-0.8043	0.1907	0.0244	0.0837		1.5138	952.8
Water temperature	0.0567	0.0087	<0.0001				

$$960.4 - 952.8 = 7.6 \sim X^2_{df=1} \ (p<0.001)$$

Gerard.Ngueta@crchuq.ulaval.ca

Table 2 : Unadjusted regression coefficients based on hierarchical linear modeling of the relation between water temperature and Log (WLLF5)

	Fixed regression coefficients			Random regression coefficients			
	Estimate	SE(Estimate)	p-value	Intercept variance	Random slope variance	Residual variance	-2Log(Likelihood)
'Naive' regression model[a]							
Intercept	-0.7666	0.1266	<0.0001			1.5880	960.4
Water temperature	0.0579	0.0089	<0.0001				
Model 1[b]							
Intercept	-0.8043	0.1907	0.0244	0.0837		1.5138	952.8
Water temperature	0.0567	0.0087	<0.0001				

Table 3 : Estimated regression coefficients based on 'standard' linear modeling of the association between water temperature and Log(WLLF5)

	Fixed regression coefficients				
	Estimate	SE(Estimate)	p-value	Adjusted R-Square	Root Mean square error
Model 1[a]					
Intercept	-0.7722	0.1238	<0.0001	0.1224	1.2616
Water temperature	0.0576	0.0088	<0.0001		
Model 2[b]					
Intercept	0.1992	0.1025	0.0530	0.0845	1.0445
Water temperature	0.0390	0.0073	<0.0001		

Gerard.Ngueta@crchuq.ulaval.ca

Table 2 : Unadjusted regression coefficients based on hierarchical linear modeling of the relation between water temperature and Log (WLLF5)

	Fixed regression coefficients			Random regression coefficients			
	Estimate	SE(Estimate)	p-value	Intercept variance	Random slope variance	Residual variance	-2Log(Likelihood)
'Naive' regression model[a]							
Intercept	-0.7666	0.1266	<0.0001			1.5880	960.4
Water temperature	0.0579	0.0089	<0.0001				
Model 1[b]							
Intercept	-0.8043	0.1907	0.0244	0.0837		1.5138	952.8
Water temperature	0.0567	0.0087	<0.0001				
Model 2[c]							
Intercept	-0.7653	0.2089	0.0351	0.1132	0.0014	1.4296	942.7
Water temperature	0.0518	0.0211	0.0909				
Model 3[d]							
Intercept	-0.6130	0.0713	0.0033	1.0657	5.3902	0.3377	535.6
Water temperature	0.0154	0.0044	0.0005				

Δ (-2Likelihood) >>> 3.84 ~ $X^2_{df=1}$ (p<0.000)

Gerard.Ngueta@crchuq.ulaval.ca

	Fixed coefficients			Random coefficients			-2L
	β	SE (β)	p	Intercept variance	Random slope variance	Residual variance	
Unadjusted model							
Intercept	-0.6130	0.0713	0.0033	1.0657	5.3902	0.3377	535.6
Water temperature	0.0154	0.0044	0.0005				
Adjusted model †							
Intercept	-1.6228	0.1294	0.0011	0.0502	0.2158	0.3382	523.3
Water temperature	0.0158	0.0044	0.0004				

†*Adjusted for LSLs, Dwelling-type, flow-rate, length of inner pipe*

Gerard.Ngueta@crchuq.ulaval.ca

MAIN LESSONS

1) If all other factors are kept stable, then for each additional increase of 1°C , WLL in the first-draw(flushed) sample is expected to increase 1.02 µg/dL ($p=0.0004$)

2) About 75.9% of variability in WLLs occur at neighborhood level

3) The great amount of neighborhood variability is attributable to the difference in the LSL distribution patterns (75.9% vs 12.9%)

4) HLM is more appropriate to capture the great amount of variability in the relation between water temperature and WLLs

Gerard.Ngueta@crchuq.ulaval.ca

THANKS FOR YOUR ATTENTION !

Questions, Suggestions, Contributions

Gerard.Ngueta@crchuq.ulaval.ca

ANNEXES

Gerard.Ngueta@crchuq.ulaval.ca

	Fixed coefficients			Random coefficients			-2L
	β	SE (β)	p	Intercept variance	Random slope variance	Residual variance	
Unadjusted model							
Intercept	-0.6130	0.0713	0.0033	1.0657	5.3902	0.3377	535.6
Water temperature	0.0154	0.0044	0.0005				
Adjusted model							
Intercept	-1.6228	0.1294	0.0011	0.0502	0.2158	0.3382	523.3
Water temperature	0.0158	0.0044	0.0004				

†Adjusted for LSLs, Dwelling-type, flow-rate, length of inner pipe

Gerard.Ngueta@crchuq.ulaval.ca

	Fixed coefficients			Random coefficients			-2L
	β	SE (β)	p	Intercept variance	Random slope variance	Residual variance	
Unadjusted model							
Intercept	-0.6130	0.0713	0.0033	1.0657	5.3902	0.3377	535.6
Water temperature	0.0154	0.0044	0.0005				
Adjusted model							
Intercept	-1.6228	0.1294	0.0011	0.0502	0.2158	0.3382	523.3
Water temperature	0.0158	0.0044	0.0004				

2) About 75.9% of variability in WLLs occur at neighborhood level

Gerard.Ngueta@crchuq.ulaval.ca

	Fixed coefficients			Random coefficients			-2L
	β	SE (β)	p	Intercept variance	Random slope variance	Residual variance	
Unadjusted model							
Intercept	-0.6130	0.0713	0.0033	1.0657	5.3902	0.3377	535.6
Water temperature	0.0154	0.0044	0.0005				
Adjusted model							
Intercept	-1.6228	0.1294	0.0011	0.0502	0.2158	0.3382	523.3
Water temperature	0.0158	0.0044	0.0004				

3) The great amount of neighborhood variability is attributable to the difference in the LSL distribution patterns (75.9% vs 12.9%)

Gerard.Ngueta@crchuq.ulaval.ca

	Fixed coefficients			Random coefficients			-2L
HLM	β	SE (β)	p	Intercept variance	Random slope variance	Residual variance	
Intercept	-0.6130	0.0713	0.0033	1.0657	5.3902	0.3377	535.6
Water temperature	0.0154	0.0044	0.0005				
OLS				**Adjusted R^2**		**RMSE**	
Intercept	-1.6228	0.1294	0.0011	75.9 %		0.6167	
Water temperature	0.0158	0.0044	0.0004				

4) HLM is more appropriate to capture the great amount of variability in the relation between water temperature and WLLs

Gerard.Ngueta@crchuq.ulaval.ca

➔ Descriptive analysis

Table 1 : Descriptive analysis

	Neighborhoods			
	Mercier-Hochelage-Maisonneuve	**Saint-Laurent**	**Verdun**	**Villeray**
N	84	47	63	100
Continuous variables‡				
WLL**F5** (µg/L) †	2.41 (0.37 – 7.00)	0.46 (0.29 – 0.78)	1.99 (0.16 – 6.30)	1.79 (0.26 – 5.14)
Length of inner pipes (m)	14.00 (10.90 – 17.61)	11.50 (9.19 – 14.90)	13.80 (12.30 – 16.50)	15.10 (12.20 – 18.21)
Flow rate (liter/s)	0.09 (0.09 – 0.10)	0.10 (0.09 – 0.10)	0.09 (0.09 – 0.10)	0.09 (0.09 – 0.11)
Water temperature (°C)	9.30 (3.05 – 21.20)	6.90 (3.60 – 14.71)	9.70 (3.20 – 21.01)	11.0 (3.70 – 21.50)
Dichotomous variables [N(%)]				
LSL+ Homes ††	58 (69.05)	15 (31.91)	34 (53.97)	62 (62.00)
Built before 1950	18 (21.43)	12 (25.53)	54 (85.71)	43 (43.00)
Single-family homes	65 (77.38)	12 (25.53)	40 (63.49)	83 (83.00)

Gerard.Ngueta@crchuq.ulaval.ca

➔Statistical analysis

$$Y_{ij} = [X_{ij}]*\beta + [Z_{ij}]*b_j + e_{ij}$$

Denotes the water lead levels in the Home *i* located in the neighborhood *j*

Gerard.Ngueta@crchuq.ulaval.ca

➔Statistical analysis

$$Y_{ij} = [X_{ij}]*\beta + [Z_{ij}]*b_j + e_{ij}$$

Fixed effect : Describe effects of all covariates included on the populational mean WLLs

Gerard.Ngueta@crchuq.ulaval.ca

➔ Statistical analysis

$$\mathbf{Y_{ij} = [X_{ij}]*\beta + [Z_{ij}]*b_j + e_{ij}}$$

How WLLs randomly vary across homes of different neighborhoods

Gerard.Ngueta@crchuq.ulaval.ca

➔Statistical analysis

$$\mathbf{Y_{ij} = [X_{ij}]*\beta + [Z_{ij}]*b_j + e_{ij}}$$

How WLLs randomly vary over time

Gerard.Ngueta@crchuq.ulaval.ca

It was not feasible to introduce at the same time LSL and Water temperature in the random statement…Model did not converge !

Gerard.Ngueta@crchuq.ulaval.ca