

Are You Turned On? A Hierarchical Modeling Approach for Estimating Lighting Hours of Use

*Lisa Wilson-Wright, NMR Group, Inc., Somerville, MA
David Barclay, NMR Group, Inc., Jacksonville Beach, FL
Andrew Correia, SessionM, Boston, MA*

ABSTRACT

When fitting statistical models to data that span multiple geographic areas for an evaluation, the researcher often prefers to provide both an overall regional savings or consumption estimate and a separate estimate for each individual area represented in the data. This works well when each area is represented by a large sample. However, project budgets and time often dictate smaller sample sizes which may be insufficient in some areas. In this scenario, the evaluator is faced with a handful of options about what data to use to report an estimate in the under-represented areas.

This paper presents a more attractive alternative using hierarchical regression models to obtain both overall and area-specific estimates for the data structure described above. Under the distributional assumptions of this approach, the model is able to use the information from each area to form the basis for the estimate in that area, while also borrowing information from each of the other areas in the study to help inform the estimate. In this way, the authors obtain more robust estimates than would have been the case had they fit separate models to each area, yet also allowed each area to maintain its own unique characteristics in the model. The study, one of the largest metering studies of its kind, relied on a sample of over 4,600 loggers from 845 homes across four states. Other regional evaluations may wish to consider the approach, increasing opportunities to leverage resources while preserving information on individual areas.

Background

The Northeast Residential Lighting HOU study was designed as one of the largest and most comprehensive residential lighting HOU studies ever conducted. Owing to the complexity and comprehensiveness of the study, this paper is one of three companion papers presented at IEPEC Long Beach 2015. This paper focuses on modeling techniques and methods. Each of the other two papers also has a specific focus; as such, this paper makes references to material that is covered in more depth in the companion papers:

- *A Lighting Study to Stand the Test of Time: Exploring the Results of a Residential Lighting Study Designed to Produce Lasting Data*, Barclay, et al., focuses on overall approaches and results.
- *What Light through Yonder Window Breaks? Methods to Study the Effects of Urban Canyons on Lighting Usage*, Walker, et al., focuses on the results of solar shading analysis performed for high-rise apartments in Manhattan.

Overview

The objective of this study was to provide updated load shapes, coincidence factors (CFs), and HOU estimates for the Connecticut Energy Efficiency Board (in cooperation with Eversource and United Illuminating), the Massachusetts Electric Program Administrators (Cape Light Compact,

National Grid Massachusetts, Eversource, and Unitil), National Grid Rhode Island, and the New York State Energy Research and Development Authority (hereafter “the Sponsors”) to assist in the calculations of demand and energy savings for lighting programs.

While all of the Sponsors were involved in project design and oversight, funding levels varied. As such, the number of households included in the final sample also varied greatly. Massachusetts and New York each contributed over 300 homes for inclusion in the study (Table 1), Connecticut and Rhode Island provided fewer than 100 each. Differences were based on available funding and the ability of Massachusetts and New York to leverage the resources of multiple previously planned saturation studies (NMR Massachusetts 2013, NMR New York 2013). Note that the study examines Upstate and Downstate New York separately and also provides estimates specific to Manhattan.

Table 1. Sample by State

Area	Homes	Loggers Modeled
Connecticut	90	549
Massachusetts	398	2,175
New York	319	1,686
Rhode Island	41	232
Total	848	4,642

Key Takeaways

This paper concentrates on the methods specific to the modeling of HOU estimates, focusing on the potential benefits of using hierarchical modeling under certain conditions. Also known as “multilevel models,” hierarchical models provide a method for analyzing clustered, multilevel data (Fitzmaurice, Laird, & Ware 2011). In the case of this study, the levels include bulbs in rooms, rooms in homes, homes in states, and states in regions, with the last level being the most critical for this paper. The advantage of hierarchical modeling is that it produces cluster-specific results but also takes information from the aggregated data into account (Stevens 2007). Thus, researchers gain both an understanding of the unique nature of each cluster without losing the additional information gained by looking at the study group as a whole.

Additional details on sample recruitment and data collection can be found in the companion paper also being presented at IEPEC 2015 (Barclay et al.). Key takeaways from this paper include the following:

- *Hierarchical modeling can serve as a tool to improve estimates for each individual area in a larger regional study.* Under certain conditions, including but not limited to individual areas exhibiting similar estimates, researchers may be able to use information from the entire sample to inform and improve estimates from the others. Given the time and cost of conducting metering studies, when appropriate, hierarchical models allow numerous program administrators within a region to pool resources that provide for a strong estimate at both the regional and area-specific levels.
- *Hierarchical modeling is most beneficial to those specific areas with smaller sample sizes.* Smaller sample sizes lead to higher standard errors and wider confidence intervals. By drawing on regional information to inform an area-specific estimate, hierarchical modeling reduces standard errors and narrows confidence intervals.

Data Preparation

The authors took several steps to prepare the data prior to fitting statistical models to estimate HOU. The steps include data cleaning, understanding reasons for sample attrition, and outlier detection. This section also addresses the time period addressed in this study.

Data Cleaning, Sample Attrition, and Outlier Detection

The team originally installed 5,730 loggers, but, as is common in studies of this size, some attrition took place due to loggers being damaged, stolen, or being otherwise unrecoverable. In total, the authors obtained data for 5,494 loggers—2,627 specifically placed for the HOU study (hereafter, “the base study”) and 2,867 from the following three studies combined: the Massachusetts Low-Income Study, the National Grid NY *EnergyWise* Study, and the NYSERDA High-Rise Study. For each logger, the HOU for each day of the study period was calculated.

Analysts performed quality assurance and quality control on the daily logger data. While some loggers did record very high or low usage over the study period, the percentage of these loggers was small. In addition, it is reasonable to expect that different households can exhibit very different usage patterns for any number of reasons, and it is not unlikely that the loggers exhibiting higher than ordinary usage represent some small portion of the actual population. Therefore, authors adopted a very conservative approach, and the only loggers removed were those for which it was not reasonable to assume the recorded data were correct—namely, those that exhibited obvious flickering or that were on continuously for over three consecutive weeks *and* whose unexpectedly high observed usage did not agree with self-reported usage for the bulb in question. Preliminary data cleaning ultimately resulted in the removal of 364 loggers, leaving 5,130 loggers across all areas.

Of the 5,130 loggers included after cleaning, an additional 488 loggers were dropped because they were missing one or more of the other variables (e.g., demographic characteristics) that contributed to the regression analysis, or because the logger ID could not be correctly matched to the on-site data. This left us with a total of 4,642 loggers for analysis.

Logging Period

Figure 1 shows when the loggers included in the final analysis (4,642 loggers) were deployed. February through July of 2013 (six months) marked the period with the greatest number of loggers deployed, and a substantial number of loggers (greater than 1,500) were in the field in each month from December 2012 through July 2013 (eight months). On average, loggers were installed for 143 days, with 84% of loggers in the field for at least 121 days and 31% in the field for at least 151 days. Loggers were installed on average for the following number of days in each area: CT – 147 days, MA – 145 days, RI – 216 days, Upstate NY – 123 days, and Downstate NY – 132 days.

This approach to logging a partial year is consistent with the guidelines recommended by the Uniform Methods Protocol for upstream lighting programs (Dimetrosky). According to the protocols,

Due to the seasonality of lighting usage, logging should (1) be conducted in total for at least six months and (2) capture summer, winter, and at least one shoulder season—fall or spring. At a minimum, loggers should be left in each home for at least three months (that is, two waves of three-month metering will attain six months of data). All data should be annualized using techniques such as sinusoidal modeling to reflect a full year of usage.

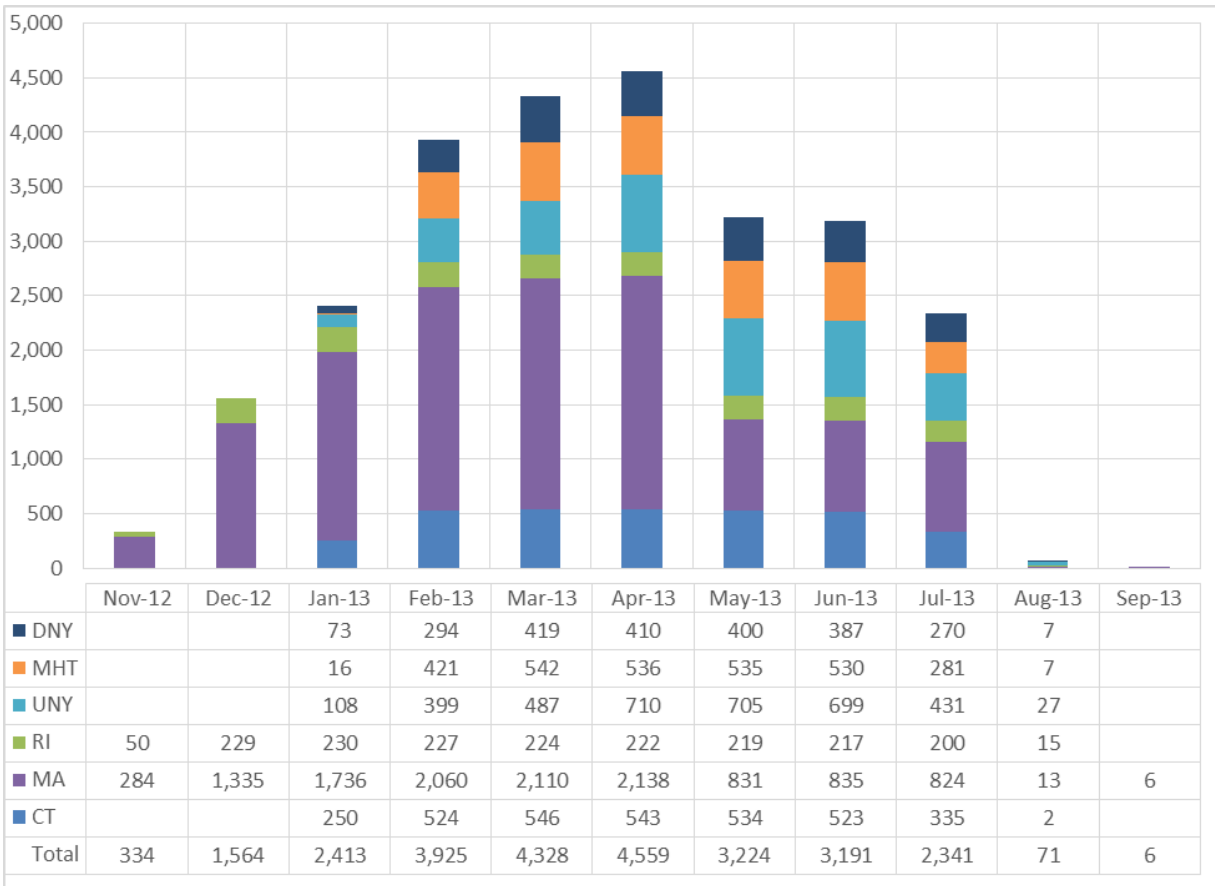


Figure 1. A Substantial Number of Loggers were in the field December 2012 through July 2013

HOU Modeling

Developing HOU estimates consisted of three modeling steps: 1) Creating annual datasets, 2) Adjusting HOU estimates, and 3) Applying a hierarchical model.

Creating Annual Data Sets

Since each logger was installed for only a portion of the year, analysts had to annualize the data. This was accomplished by fitting a sinusoid model individually to each logger. Authors drew upon the methods outlined in past studies (KEMA & Cadmus 2010) and NMR and DNV GL (2014) detail the approach used in the current study. In summary, the authors fitted separate weekend and weekday models for each logger. For any loggers not conforming well to the sinusoid model, the analysts took additional steps to prepare annualized estimates based on average daily usage over the period logged (described below). The sinusoid model for each logger took the following form:

$$h_d = \alpha + \beta \sin(\theta_d) + \epsilon_d$$

Where

h_d = hours of use on day d ,

θ_d = angle for day d , where θ_d is 0 and the spring and fall equinox, $\pi/2$ for d = December 21, and $-\pi/2$ for d = June 21,

α and β are regression coefficients,

ϵ_d is the residual from the regression.

In each model, α represents the average weekday (or weekend day) use for a given logger. Because the authors fitted a weekday model and a weekend model for each logger, the overall average usage for the year for each logger represented the weighted average of the α from the weekday model and the α from the weekend model. After assessing the goodness of fit and assigning average yearly values to poor fitting models (NMR & DNG GL 2014), the authors then calculated the overall average annual daily hours of use for each logger by averaging the weekend and weekday specific averages in proportion to the number of weekend/weekday days over the course of the year. Specifically:

$$avg. hou_i = \frac{(n_{wd}\alpha_{wd,i} + n_{we}\alpha_{we,i})}{n_{wd} + n_{we}}$$

Where i indexes each logger, n_{wd} is the number of weekdays over the year, n_{we} is the number of weekend days over the year, $\alpha_{wd,i}$ is the average weekday usage for logger i , and $\alpha_{we,i}$ is the average weekend usage for logger i .

After annualizing the data for each logger, the authors merged logger data with household demographic data. Household demographic data included information on education level, income, single- or multifamily status, own/rent status, and whether there was anyone under 18 years of age in the household.

As described in the full report (NMR & DNV GL 2014), the model performed well for most loggers, and the average amplitude of the sine curve across all good-fitting models (the average estimate of the slope term, β) was very similar to those of other comparable studies, suggesting that the overall effect of season is relatively similar in the two regions. Average estimates for poor-fitting models were more extreme in absolute value and exhibited much higher uncertainty than their good-fitting counterparts, again consistent with previous studies.

Adjusting HOU Estimates

Next, the authors used the annualized estimates as the dependent variable in a weighted regression analysis to estimate the adjusted average HOU for each room in each area of the study. Table 2 on the next page describes the variables that contributed to the regression analysis as predictors.

The authors retained variables in the model if they were statistically significant at 90% confidence, allowing for a more parsimonious model. The authors made two exceptions to this rule: they retained income level and housing type despite the lack of statistical significance, as one of the goals of this study was to quantify the association between usage and income/housing type.¹

At this point in the process, the model used only loggers for each individual area to develop area-specific estimates, while the regional model included all loggers in the study. Based on outputs from this model, the results indicated that Connecticut, Massachusetts, Rhode Island, and Upstate New York exhibited comparable usage patterns, while those for Downstate New York (including Manhattan) differed from the other areas. Table 3 on the next page presents the HOU estimates from these area-specific regressions.

¹ Additional variables considered for inclusion in the model that did not prove to be statistically significant included saturation, fixture type, bulb shape, socket type, and control type.

Table 2. Variables Used as Predictors in HOU Regression Models

Variable	Description	Levels
Room Type	Room/location the bulb was located.	Bedroom
		Bathroom
		Kitchen
		Living Space
		Dining Room
		Exterior
		Other
Efficient Bulb	Whether the bulb was efficient or non-efficient.	Yes
		No
Income	Household income.	Low Income
		Non-Low Income
Education	Education level of the respondent.	Less than High School
		High School or GED
		Some College
		Bachelor's Degree
		Advanced or Graduate Degree
Rent/Own	Whether household is owned or rented	Rent
		Own
Under 18	Anyone under 18 years of age in the household	Yes
		No
Home Type	Single or multi-family residence	Multi Family
		Single Family

Table 3. Overall Estimated HOU from Preliminary Models

Area	Estimated Overall HOU	90% Confidence Interval
Connecticut	2.9	(2.5, 3.2)
Massachusetts	2.6	(2.4, 2.8)
Rhode Island	2.9	(2.2, 3.5)
Upstate New York	2.4	(2.1, 2.8)
Downstate New York	4.1	(3.5, 4.7)
Manhattan	3.9	(3.4, 4.4)

Applying Hierarchical Model

Due to the similar use patterns in Connecticut, Massachusetts, Rhode Island, and Upstate New York, the authors sought a way to leverage data from them to refine area-specific estimates. The team treats Downstate New York and Manhattan differently, as described below. The structure of the data—loggers nested in homes, nested in areas—is well suited for a multi-level hierarchical model. This modeling approach offers the advantage of using information from all four areas to help inform area-specific estimates. In a hierarchical model, the observations specific to an area form the basis of the estimates for that area, while observations from outside that area also inform and help refine the area-specific estimates (Cnaan, Laird, & Slasor 1997, Fitzmaurice, Laird, & Ware 2011). Figure 2 provides a

visual representation of how the estimate for Rhode Island is informed by loggers in Connecticut, Massachusetts, and Upstate New York. The hierarchical model particularly benefits areas that had fewer loggers installed, thus providing more refined (tighter precision and adjusted means) HOU estimates compared to individual models fit to each area separately.

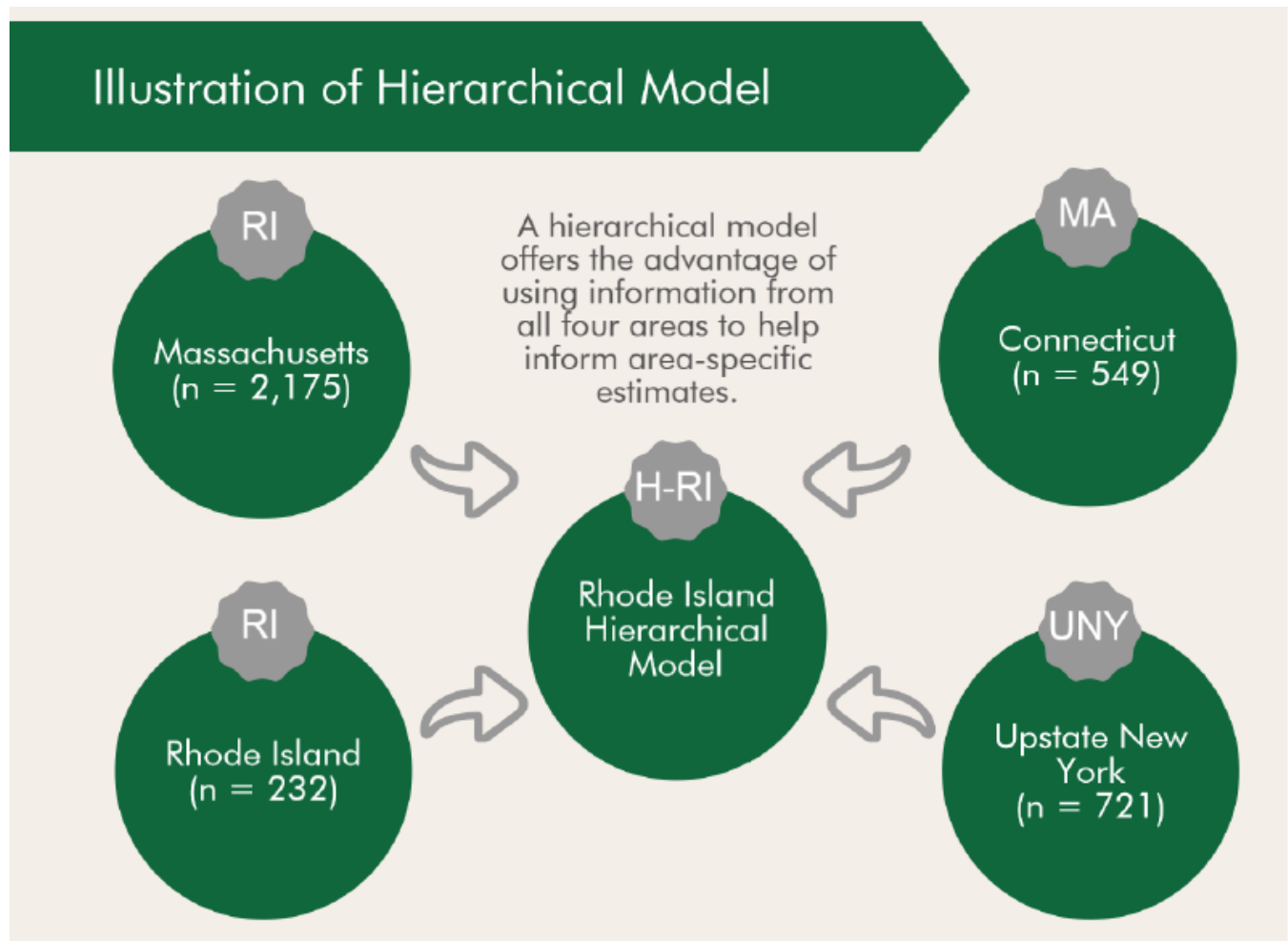


Figure 2. Illustration of Hierarchical Model

To account for potential correlation among loggers in the same household or area (e.g., HOU may be correlated based on when people leave or return to their home, or go on vacation), the model included a random intercept term at the site ID level, dependent on the area in which the site ID is nested. This dependence is established at another level in the modeling framework. Additionally, to estimate area-specific HOU estimates for all rooms, the model included random area-specific regression coefficients for the room type variable, allowing for information from other areas to help inform the area-specific HOU estimate of each room. Premise and room weights were applied directly in the likelihood of the model (Graubard & Korn 1996, Rabe-Heskett & Skrondal 2006). The exact form of the hierarchical model is presented below.

$$\begin{aligned}
E(hou_{ijk}) = & (\beta_0 + b_{0,j}) + (\beta_1 + b_{1,k}) \times I(\text{Room}_{ijk} = \text{Bathroom}) + (\beta_2 + b_{2,k}) \\
& \times I(\text{Room}_{ijk} = \text{Bedroom}) + (\beta_3 + b_{3,k}) \times I(\text{Room}_{ijk} = \text{Dining}) + (\beta_4 + b_{4,k}) \\
& \times I(\text{Room}_{ijk} = \text{Kitchen}) + (\beta_5 + b_{5,k}) \times I(\text{Room}_{ijk} = \text{Living}) + (\beta_6 + b_{6,k}) \\
& \times I(\text{Room}_{ijk} = \text{Other}) + I(\text{Bulb.type}_{ijk} = \text{Efficient}) + \beta_8 I(\text{Income}_{ijk} = \text{LI}) \\
& + \beta_9 I(\text{Education}_{ijk} = \text{HS}) + \beta_{10} I(\text{Education}_{ijk} = \text{Some college}) \\
& + \beta_{11} I(\text{Education}_{ijk} = \text{Bach.}) + \beta_{12} I(\text{Education}_{ijk} = \text{Adv/Grad Deg.}) \\
& + \beta_{13} I(\text{Own/Rent}_{ijk} = \text{Rent}) + \beta_{14} I(\text{Under18}_{ijk} = \text{yes}) \\
& + \beta_{15} I(\text{Home.type}_{ijk} = \text{MF})
\end{aligned}$$

Where i indexes the loggers, j indexes the homes, k indexes the areas, and:

$$\begin{aligned}
b_{0,j} & \sim N(b_k, \sigma_{b_k}^2), \quad \forall \text{site}_j \in \text{region}_k \quad \forall k, \\
b_k & \sim N(0, \sigma_{reg}^2), \quad \text{for } k = 1, \dots, n_{\text{regions}}, \\
b_{l,k} & \sim N(0, \sigma_l^2), \quad \text{for } l = 1, \dots, 6 \text{ and } \forall k,
\end{aligned}$$

Table 3 above shows that Downstate New York (including Manhattan) and Manhattan by itself had different usage patterns—specifically, higher HOU—than the other four areas in the study.² Thus, separate robust linear regression models were fit for Downstate New York, for the subset of Downstate New York in Manhattan, and for all of the NYSEDA area (all of Upstate and Downstate combined). Downstate regression models incorporated the same variables listed above. After fitting the regression models, the authors used the fitted values of the appropriate regression to calculate adjusted HOU estimates by area and room.

For all areas, the authors also fitted models for eight sub-categories based on home type (single vs. multifamily) and income level (low-income vs. all others) and the mixture of the two. Again, they fitted a hierarchical model with data from Connecticut, Massachusetts, Rhode Island, and Upstate New York but treating Downstate New York and Manhattan separately within each sub-category. The authors urge the reader to review these models in the full study.

Regression Model Coefficients

Table 4 shows the overall regression coefficients from the hierarchical model fitted to all loggers in Connecticut, Massachusetts, Rhode Island, and Upstate New York. These coefficients were relatively consistent across models, so the table only presents the overall hierarchical model coefficients. Not only does the hierarchical model allow information from across regions to help inform each region-specific estimate, it also performs better than its non-hierarchical counterpart. The pseudo- R^2 for the overall hierarchical regression model, as calculated according to Xu (2003), is 0.26, compared to an R^2 value of 0.14 for the stand-alone regression model fit at the overall level, suggesting a nearly two-fold improvement in the amount of explained variance from fitting a standard linear regression model to this data. Blank cells in this table represent the baseline level of each variable in the model, and all coefficients should be interpreted as relative to the corresponding baseline level for each variable. For example, controlling for other factors, efficient bulbs were used for about 0.6 hours more than less efficient models,³ and renters used bulbs for about 0.5 hours more than owners did.

² We do not present Downstate New York minus Manhattan due to the NYSEDA program structure. They treat Downstate—comprising all of New York City, most of Westchester County, and a few towns in other counties—as one unit in their program planning and implementation.

³ This coefficient—and a discussion of “take back”—is addressed in the companion paper: *A Lighting Study to Stand the Test of Time: Exploring the Results of a Residential Lighting Study Designed to Produce Lasting Data*, Barclay, et al. also presented at the 2015 IEPEC. This paper—as well as the full study report—provide more detail on these coefficients and on **2015 International Energy Program Evaluation Conference, Long Beach**

Table 4. Overall Regression Coefficients from Hierarchical Model

Variable	Level	Coefficient	90% Confidence Interval*
Efficient Bulb	Yes	0.631	(0.455, 0.806)
	No		
Income	Low Income	0.007	(-0.261, 0.273)
	Non-Low Income		
Education	Grad/Adv. Degree	-0.635	(-1.288, -0.082)
	Bachelor's Degree	-0.587	(-1.253, -0.019)
	Some College	-0.778	(-1.420, -0.248)
	HS or GED	-0.728	(-1.362, -0.176)
	Less than HS		
Own/Rent	Rent	0.532	(0.249, 0.821)
	Own		
Under 18	Yes	0.598	(0.362, 0.824)
	No		
Home Type	Multi Family	-0.157	(-0.470, 0.154)
	Single Family		

* Intervals that do not contain zero correspond to statistical significance at 90% confidence.

HOU Estimates and Confidence Intervals Derived from Hierarchical Models

Figure 3 compares the hours of use estimates and confidence intervals derived from the preliminary models and the hierarchical models. These estimates include all bulbs in the home; as noted above use varied between efficient and inefficient bulbs, with efficient bulbs seeing more use.⁴ While the point estimates change very little between models within the state, the confidence intervals derived from the hierarchical models are considerably tighter than those developed from the preliminary models. Thus, taking advantage of additional information from other similar areas allowed the authors to reduce standard errors, thereby yielding estimates with a smaller error band around them. Note that the estimates for Downstate New York and Manhattan did not change from those reported in Table 3 as they were not included in the hierarchical modeling effort.

This approach clearly has the most impact on reducing the standard error for the two areas with the fewest loggers, namely Connecticut and Rhode Island. However, it is worth noting that the error band in Rhode Island remains larger even after fitting the hierarchical model than that of the other states; this reflects the small number of loggers in that state compared to the others. This fact serves to emphasize that, while the approach certainly assists in narrowing confidence intervals for areas with fewer loggers, it still requires an adequately large sample size—and Rhode Island just made the cut for this analysis.

the issue of take back.

⁴ *A Lighting Study to Stand the Test of Time: Exploring the Results of a Residential Lighting Study Designed to Produce Lasting Data*, Barclay, et al., also presented at the 2015 IEPEC, addresses these issues in more details.



Figure 3. Comparison of Preliminary and Hierarchical HOU Estimates and Confidence Intervals

Conclusions

The research described in this paper provided area-specific and regional estimates of HOU for numerous residential lighting program administrators in the Northeast, with recommended HOU estimates ranging from 2.6 for Rhode Island and Upstate New York to 4.1 for Downstate New York. However, the focus of this paper was not on the estimates themselves but on the statistical modeling approach the evaluators used to derive them.⁵ Importantly, by fitting the data to a hierarchical model, the authors succeeded in producing estimates that used logger data from the individual area and the region overall. This not only means that the resulting HOU estimates took more information into account, but it also reduced the standard error for each area, ultimately narrowing the confidence intervals. The approach provides the most benefit to areas with smaller sample sizes. The successful application of this approach suggests that future evaluations that rely on statistical modeling may want to explore the use of hierarchical models, particularly when the study requires that they produce estimates for individual areas as well as an overall estimate. The approach also provides program administrators with smaller budgets to pool resources into a larger study rather than expending limited evaluation resources on individual, expensive studies that may still lack adequate sample sizes to meet desired levels of confidence and precision.

⁵ See *A Lighting Study to Stand the Test of Time: Exploring the Results of a Residential Lighting Study Designed to Produce Lasting Data*, Barclay, et al., focuses on overall approaches and results also presented at the 2015 IEPEC for details on the study and the results.

References

- Barclay, D., Nelson, M., Prah, R., Walker, S., von Trapp, K., Correia, A., Hoefgen, L., Dimetrosky, S., and Barclay, D. 2015. "A Lighting Study to Stand the Test of Time: Exploring the Results of a Residential Lighting Study Designed to Produce Lasting Data." *In Proceedings of the IEPEC*, Long Beach, CA: International Energy Program Evaluation Conference.
- Walker, S., Barclay, D., Correia, A., Engel-Fowles, V., and Prah, R. 2015. "What Light through Yonder Window Breaks? Methods to Study the Effects of Urban Canyons on Lighting Usage." *In Proceedings of the IEPEC*, Long Beach, CA: International Energy Program Evaluation Conference.
- Barclay, D., Walker, S., and von Trapp, K. 2015. "We Know What You Did Last Summer: Revelations of a Multi-Year Lighting Panel Study." *In Proceedings of the IEPEC*, Long Beach, CA: International Energy Program Evaluation Conference.
- Cadmus Group, Inc. 2014. "Massachusetts Low Income Metering Study." Delivered to the Massachusetts Program Administrators and Energy Efficiency Advisory Council Consultants, May.
- Cnaan, A. Laird, N.M., and Slasor, P. 1997. "Tutorial in Biostatistics: Using the Generalized Linear Mixed Model to Analyze Unbalanced Repeated Measure and Longitudinal Data." *Statistics in Medicine* 16: 2349-2380.
- Dimetrosky, S. 2013. "Chapter 6: Residential Lighting Evaluation Protocol." *In The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Golden, CO: National Renewable Energy Lab.
- Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. 2011. *Applied Longitudinal Analysis, 2nd Ed.* New York: Wiley.
- Graubard, B.I. & Korn E.L. "Modeling the Sampling Design in the Analysis of Health Surveys." *Statistical Methods in Medical Research* 5 (1996): 263-281
- KEMA, Inc. and the Cadmus Group, Inc. *Final Evaluation Report: Upstream Lighting Program Volume 1*. Prepared for California Public Utilities Commission, Energy Division. February 8, 2010.
- NMR Group, Inc. 2013. "Massachusetts Onsite Lighting Saturation Report." Delivered to the Massachusetts Program Administrators.
- NMR Group, Inc. and DNV GL. 2014. "Northeast Residential Lighting Hours-of-Use Study." Delivered to the Program Administrators and Energy Efficiency Advisory Council, May.
- NMR Group, Inc., Research Into Action, and Apex Analytics. 2013. "Market Effects, Market Assessment, Process and Impact Evaluation of the NYSERDA Statewide Residential Point-of-Sale Lighting Program: 2010-2012." Delivered to NYSERDA. Albany, NY: New York State Energy Research Development Authority..

- NMR, RLW, and GDS, 2010. "Residential Lighting Markdown Impact Evaluation." Delivered to the Massachusetts Program Administrators and Energy Efficiency Advisory Council, January.
- Rabe-Hesketh, S. & Skrondal A. "Multilevel Modeling of Complex Survey Data." Journal of the Royal Statistical Society, Series A 169 (2006): 805-827.
- Stevens, J. 2007. "Hierarchical Linear Models." Presentation available at <http://pages.uoregon.edu/stevensj/workshops/hlmslides.pdf>. Accessed May 26, 2015.
- United States Naval Observatory. "Duration of Daylight/Darkness Table for One Year." http://aa.usno.navy.mil/data/docs/Dur_OneYear.php.
- Xu R. "Measuring Explained Variance in Linear Mixed Effects Models." Statistics in Medicine 22 (2003):3527-3541.