U.S. EPA ENERGY STAR benchmarking scores for medical office buildings lack scientific validity

John H. Scofield[1]
Department of Physics & Astronomy, Oberlin College, Oberlin, OH 44074, USA

Abstract

The *U.S. Environmental Protection Agency* (EPA) issues ENERGY STAR building scores for benchmarking 11 conventional building types following a similar methodology. Here we describe this methodology, then examine its implementation for one specific building type, medical offices. We find the EPA's description of the dataset that underpins this model to be erroneous. Correcting published errors we replicate the stated properties of the EPA's model dataset and subsequent model regression. The standard error in the predicted building source energy derived from this regression is shown to be so large as to produce ENERGY STAR scores that are uncertain by ±35 points – rendering such scores meaningless. Two tests of the EPA's model dataset and regression indicate that four of the five regression variables are not statistically reliable predictors of medical office building energy use. The root problem is shown to be that the standard relative error in the CBECS subset on which the model is based is too large to provide adequate characterization of the U.S. medical office building stock at the level required for model purposes. Preliminary analysis of other building models suggests half have similar problems and all produce scores that are uncertain by 10 points or more.

[1] Email: john.scofield@oberlin.edu or Telephone: +1.440.775.8333

1. Introduction

In 1998 the *U.S. Environmental Protection Agency* (EPA) introduced its ENERGY STAR building rating system or ENERGY STAR score [1] based, in part, on benchmarking methodology developed at US Department of Energy (DOE) laboratories [2, 3]. This so-called energy benchmarking system allows a commercial building manager to track the energy performance of his/her building over time and also provides a measure of the building's energy performance relative to similar buildings nationally. The score, based on measured building energy data, is a number from 1-100 that is supposed to represent the percentage of similar buildings nationally that use more source energy than the building under consideration – adjusting for building size, climate, and key operational factors.

ENERGY STAR building benchmarking began as a voluntary process that engaged relatively few buildings. Data were self-reported and scores were confidential – publicly disclosed only for those buildings that sought and received ENERGY STAR certification – requiring a minimum score of 75. The EPA does not grant external access to data gathered by *Portfolio Manager*, its web-based tool for calculating building ENERGY STAR scores [4]. Instead the EPA releases selected statistics for the purpose of promoting the ENERGY STAR benchmarking program. In this tightly-controlled, compartmentalized environment it is not possible for anyone outside the EPA's ENERGY STAR commercial buildings group to validate the science behind these ENERGY STAR scores or even to observe trends in the scores other than those revealed by the EPA. Moreover, the EPA retroactively revises building ENERGY STAR scores whenever it revises its building models. This was the case, for instance, in 2007 when it updated its office/bank/finance/courthouse model. No record of older ENERGY STAR scores is retained – except for certified buildings whose scores, of course, are 75 or higher. This makes it difficult if not impossible to see what errors, if any, were corrected by the EPA in revising its building models.

In recent years green building certification programs administered by both the *US Green Building Council* (USGBC) [5] and *Green Building Initiative* (GBI) [6] have adopted the EPA's ENERGY STAR score as a measure of energy efficiency, dramatically expanding the use of the ENERGY STAR building benchmark. Green buildings seeking certification under these increasingly popular programs must submit energy performance data to *Portfolio Manager*; energy efficiency points towards certification are tied to the ENERGY STAR score received. More importantly, as a result of Energy Benchmarking laws recently passed in upwards of 10 major US cities, numerous commercial building owners are now required to use the EPA's *Portfolio Manager* for scoring the energy efficiency of their buildings [7]. In many cases the resulting ENERGY STAR scores are made public.

If ENERGY STAR scores have the interpretation the EPA claims for them then one expects the mean ENERGY STAR scores for the eligible[2] US building stock of any particular type of building to be 50. Moreover, these scores are expected to be uniformly distributed: 10% of eligible buildings should have scores from 1-10, another 10% scores from 11-20, etc. This is a trivial consequence of the manner in which a rank-order distribution is defined. The EPA has offered no evidence that its building models, when applied to the entire eligible buildings stock, produce such a distribution. To the contrary, evidence is emerging that national scores do not demonstrate this property. The expanded use of ENERGY STAR benchmarking has resulted in publication of mean ENERGY STAR scores for several large sets of buildings – and in all cases, these means are significantly higher than 50. A study of California office buildings found a mean ENERGY STAR of 64 [8]. This same mean value was observed by the *New Buildings Institute* (NBI) in their 2008 analysis of LEED-certified buildings [9]. The EPA itself reports

[2] The EPA has not developed a scoring system for all building types, and for the types it does score, in most cases it will only score buildings with 5,000 ft^2 (465 m^2) or more. Roughly half the commercial buildings in the US building stock are smaller and therefore ineligible for the ENERGY STAR scoring process.

that the mean score issued for all office buildings whose data have been entered into Portfolio Manager is 62 – well above the putative mean of 50 [10]. The LEED building mean is particularly puzzling since the LEED buildings in that study have been shown to have no significant reduction in source energy consumption relative to non-LEED buildings [11].

For this author the tipping point came with the release of 2011 energy benchmarking data from New York City for more than 4,000 commercial buildings. Analysis of nearly 1,000 office buildings in that data set found their mean ENERGY STAR score to be 68 while their source energy consumption was no different from that given by CBECS national averages [12]. Furthermore, within the set of NYC office buildings a subset of 21 LEED-certified buildings was identified. This LEED-certified subset had a mean ENERGY STAR score of 78 – yet their mean source energy use intensity was the same as that of other NYC office buildings and the CBECS average for large office buildings. This raises a number of questions regarding the validity of these particular ENERGY STAR scores and, more generally, the EPA's scoring methodology [12].

The EPA has developed mathematical models for scoring 15 different building types; 11 of these are for conventional building types included by the *Energy Information Administration* (EIA) in its *Commercial Building Energy Consumption Survey* (CBECS). These 11 models (mostly) follow a common methodology. The models for the other four types (data centers, swimming pools, parking lots, and waste-water treatment plants) are more diverse. The prescription for each of these building models is provided in technical documents published by the EPA [14]. There is, however, almost no published documentation developing or justifying the science behind this methodology.

We have looked closely at the methodology behind 10 of the EPA's 11 conventional building models. For each of these 10 models we have obtained the underlying data on which the EPA's model is built and subsequently attempted to replicate the EPA's regression and analysis. The results for all 10 models investigated are too extensive and varied to be discussed in a single paper. Here in this first paper we consider the general methodology behind these buildings models then focus on one particular building type –medical office buildings. This model is one of the EPA's oldest building models. It was released in 2004 and is based on 1999 CBECS data. It has not been subsequently revised.

The ENERGY STAR scores issued by the EPA using this medical office model will be shown to have little scientific validity. The uncertainties (or standard errors) in these ENERGY STAR scores (±35 pts) are so large as to render these scores of little quantitative value. Only one of the five variables identified by the EPA as significant predictors of source energy consumption for these buildings passes consistency tests – suggesting that much of the observed correlation with the other four variables is accidental. It will be shown that building data on which this model is constructed are insufficient to provide meaningful characterization of the gross source energy use intensity of US medical office buildings, let alone provide information that allows for the prediction of energy consumption from other variables.

2. General Methodology for building ENERGY STAR models

Before looking at the details of the medical office ENERGY STAR model we first look at the general approach used for all 11 of the ENERGY STAR conventional building models [14].

The basic idea behind the EPA's ENERGY STAR building score is this. First, for a particular building type (e.g., medical offices) the EPA must acquire a statistically meaningful dataset that accurately represents the properties of the national building stock for this particular type of building. Data collected include building size, annual energy purchases (electricity, natural gas, etc.), and a host of other parameters that describe building systems, function and operation. Ideally, representative data should be gathered from hundreds if not thousands of buildings spanning every U.S. climate zone, age category, size category, and demographic region. Each building in the dataset has associated with it a “weighting factor” that indicates how many similar buildings it represents in the building stock. The properties of larger building stock are obtained by combining these weighting factors with the properties of the sampled buildings. Here I will refer to this dataset as the “*model dataset*.”

The energy consumed at a building, called site energy, is readily obtained from its annual utility bills. Annual amounts of electricity, natural gas, and any other forms of energy that flow into the building are converted to a common energy unit – Joules (or Btu in the U.S.) – and summed. Energy associated with district heat (steam or hot water) or chilled water are included here, as well, though metering these energy flows can be challenging. Since site energy scales with building size it is common practice to divide a building’s annual site energy by its gross floor area – measured in square meters (or square feet in the U.S.) – to obtain its site *energy use intensity* or *energy utilization index* (EUI) – expressed in J/m^2 (or Btu/ft^2 in the U.S.). EUI figures are subject to considerable uncertainty, not so much from errors in measured energy as much as errors/inconsistency in gross floor area figures [15].

Site energy, however, does not account for off-site energy losses associated with producing the energy and transporting it to the building – indirect consumption of energy that would not occur if the building did not exist. These off-site losses vary with fuel type and are particularly important for electric energy for which, on average, roughly three units of primary energy are consumed for every unit of electric energy delivered. Source energy is defined to account for both on- and off-site energy consumption associated with building operation. A building’s source energy is calculated from components of its site energy by multiplying the site energy for each fuel consumed by a standardized fuel-dependent site-to-source energy conversion factor. The adjusted energy figures are then summed to yield the building’s source energy [16].

The EPA calculates the annual source energy consumed by each building in its *model dataset*. For models developed or revised since 2007 the EPA divides the annual source energy by the gross floor area to obtain the source EUI for each building. This source EUI becomes the focus of the model regression. For older models the EPA, instead, used the natural log (Ln) of the annual source energy (E) as the dependent variable for its regression. The EPA looks for variables in the model dataset that are highly-correlated with the building’s source EUI (or LnE for older models). The *model dataset* may contain hundreds of variables for each sampled building – many that are not useful for this purpose. Specifically, EPA staff (or contractor acting on their behalf) experiment with numerous multi-point regressions where the annual source EUI (or LnE for older models) for each sampled building is the dependent variable and various building parameters in the dataset are independent variables. The choices of independent variables are driven by statistical significance rather an underlying physical building model. Independent variables are retained in the regression if they show significant correlation with the dependent variable and rejected otherwise. After trying dozens of variables and combinations of variables the EPA settles in on a final building model in which anywhere from 4-11 building variables are retained in the regression. The R^2 for these regressions vary from 0.22 to 0.93, meaning that, depending on the building model, the independent variables “explain” from 22% to 93% of the observed variation in the dependent variable (EUI for newer models or LnE for older models) found in the *model dataset*.[3] Any deviation in a building’s *measured source energy*[4] from that “predicted” by the regression model is assumed to reflect the energy efficiency of the building.

---

[3] For all buildings the total energy scales roughly with the size of the building. This obvious scaling is removed in models revised after 2007 by using the EUI as the dependent variable. For older models the LnE is the independent variable and Ln(Area) is one of the independent variables. This factor alone adds considerably to the R^2 for the older regression models.

[4] The term “measured source energy” is used here to distinguish from the “predicted source energy.” Of course, source energy is not measured but rather calculated from measured site energy.

Note there can be significant difference between the distribution of properties in the model dataset (sampled buildings) from the distribution of the properties of the national building stock the sample represents. Proper utilization of the sampling weights included in the model dataset is crucial for accurately inferring the properties of the national building stock. For its newer models (2007 onward) the EPA uses these sampling weights to perform a "weighted linear regression." For older models the regressions are unweighted.

Once the regression coefficients have been determined these coefficients may be combined with data from any building of this type to "predict" its source energy consumption given the appropriate independent variables. The building's *measured* source energy consumption is then compared to its *predicted* consumption. The EPA defines the building's *Energy Efficiency Ratio* (EER) to be the ratio of its *measured* source EUI to this *predicted* source EUI (or, for older models, the ratio of the *measured LnE* to the *predicted LnE*). The lower the EER value the more efficient the building is judged to be.

In order to translate the EER into an ENERGY STAR score the EPA looks again to the *model dataset*. EER's are calculated for each building in the *model dataset*. The *model dataset* records are then sorted in order of increasing EER. In order to represent the national building stock each sampled building must be combined with a "weighting factor" that indicates the number of similar buildings in the building stock that it represents. The sorted list and weighting factors associated with each building are then used to form a cumulative EER distribution for the target national building stock. The ENERGY STAR score associated with a specific EER is found by determining from this list the percentage of buildings in this national stock that have higher EER's. In its 11 building models the EPA has been inconsistent in its use of building weights for this step. For its newer models the EPA employs the building weights to determine the cumulative EER distribution which defines ENERGY STAR scores. For older models – like the medical office model – it does not.

For 9 of the 11 conventional building models the EPA has obtained its data from the EIA's *Commercial Building Energy Consumption Survey* (CBECS).[5] The EPA performs a query on the CBECS database to extract sampled records for a particular building type. The building type (or activity) is identified by several CBECS variables including PBA (principal building activity) or PBAPLUS. This initial building set is further refined by applying a series of "filters" to remove records for sampled buildings that are problematic for one reason or another. For almost all of its models the EPA claims not to include buildings whose gross floor area is less than 5,000 ft^2 (465 m^2), though its justification for this decision is questionable.[6] The filtered dataset then becomes the model dataset for this particular building ENERGY STAR model.[7]

In summary, the key steps in the methodology are: 1) The EPA must acquire a statistically valid database that represents a type of building in the commercial stock, 2) through regression analysis the EPA determines a set of key independent variables that are statistically significant for predicting building source energy consumption, 3) the EPA uses this to generate the cumulative distribution of EER's for the

[5] For the remaining two models – *Senior Care* facilities and *Hospitals*, the EPA developed its own data sets through collaboration with industry associations.

[6] For the medical office model considered here the justification is inconsistent with the facts.

[7] In its most recent models/revisions the EPA has recognized that the model dataset for regression purposes may actually be different from that used for representing the entire building stock. For instance, a building that was sampled in CBECS may have such high or low energy consumption that it is eliminated from the regression as an "outlier." But once the regression is established, it may be used to produce an EER for such an outlier and it does contribute to the cumulative EER distribution of the building stock. This distinction was not recognized in models before 2010.

building stock, determining where a particular EER ranks with respect to other similar buildings, and 4) to score a particular building you must submit the relevant building data (energy use and operating parameters) with regression parameters to calculate a predicted energy, EER, and subsequent ENERGY STAR score. For a building's ENERGY STAR score to be valid with the stated meaning – all four of the above steps must be carried out with accuracy and validity.

3. Medical Office Model

The EPA's building model for medical offices was released in January 2004 and has not been subsequently revised [17]. 93 building records were extracted from the 1999 CBECS with the variable PBAPLUS7 = '08' (meaning "Doctor/Dentist office"). The EPA's technical document claims to filter out 11 records by removing any for which (a) operating hours were less than 30 per week, (b) number of workers was less than 1, (c) source EUI was less than 38 kBtu/ft^2 (430 MJ/m^2) or greater than 575 kBtu/ft^2 (6500 MJ/m^2) and d) floor area was less than 5,000 ft^2 (465 m^2).

We have extracted these 93 records from CBECS 1999 but, upon application of the above filters, were unable to replicate the EPA's results. Filters (a) and (b) eliminated one record and filter (c) eliminated five records. Of the 87 records remaining 16 had floor area less than 5,000 ft^2 (465 m^2). Hence applying the EPA's filter (d) produced a dataset with only 71 records. Furthermore, continued processing of these 71 records yielded results inconsistent with data means, minimum, and maxima published in the Table 2 of the EPA's technical methodology document [17]. It is clear that the EPA's document does not accurately describe the data on which the EPA's medical office model is constructed.

To resolve this issue this we requested from EPA staff a list of the CBECS 1999 building ID's which constitute the medical office dataset. That request drew no response [18]. We subsequently filed a *Freedom of Information Act* (FOIA) request with the EPA to obtain this list. The EPA closed this case claiming no such document exists [19]. Apparently the EPA does not know what data form the basis for its medical office ENERGY STAR model.

Sufficient clues were provided in Table 2 of the EPA's Technical Methodology document to determine the subset of the 1999 CBECS medical office buildings that form the 82-building model dataset. Table 2 lists the maxima, minima, and means for various variables in the dataset. These values indicated that buildings with as little as 1,000 ft^2 (93 m^2) were included in the dataset, buildings with fewer than 35 or more than 124 hrs/wk were not included, and buildings with fewer than 2 workers were not included. Three of the buildings in the list of 87 obtained after applying filters (a)-(c) had weekly hours of 168, one had 30 hrs/wk, and one had just one worker. 82 buildings remained after removing these five buildings – and these collectively had mean, maximum, and minimum properties nearly identical to those listed in Table 2 of the EPA's *Technical Methodololgy* document. 14 of the 82 buildings (17%) had floor area ranging from 1,001 ft^2 (93 m^2) to 4,000 ft^2 (372 m^2). It is clear that the EPA's statement that "Analysis could not model behavior for buildings smaller than 5,000 ft^2" is not correct [17]. The 1999 CBECS building identification numbers (PUBID7) for the 82 medical office dataset are listed in Table 1 below.

1, 7, 75, 116, 135, 164, 368, 539, 836, 878, 979, 988, 1066, 1082, 1128, 1129, 1191, 1200, 1223, 1255, 1288, 1361, 1658, 1679, 1681, 1694, 1706, 1800, 1925, 2023, 2114, 2227, 2248, 2269, 2276, 2295, 2333, 2448, 2513, 2640, 2695, 2784, 2888, 2898, 2948, 3092, 3096, 3229, 3281, 3289, 3340, 3342, 3394, 3494, 3656, 3888, 3919, 4004, 4008, 4019, 4035, 4086, 4108, 4113, 4204, 4268, 4409, 4442, 4694, 4719, 4778, 4853, 4889, 4902, 4972, 4999, 5179, 5300, 5308, 5314, 5341, 5352

Table 1. List of the 1999 CBECS building identification numbers (PUBID7) for the 82 sampled buildings that apparently make up the EPA's ENERGY STAR medical office building model dataset.

With the medical office model dataset determined the next step is to reproduce the EPA's model regression. While building source energy use intensity (EUI) is used as the dependent variable in regressions for most of the building models, regressions for older models (that predate 2007) including this one use the natural log of the building source energy as the independent variable. The EPA performed an un-weighted linear regression of the natural log of the measured source energy (dependent variable in kBtu/ft^2 units) on five independent variables: the natural log of the gross floor area (LnArea, Area expressed in ft^2), the natural log of the number of weekly hours of operation (LnWkHrs), natural log of the number of full-time workers (LnNWker), the number of *Heating Degree Days* multiplied by the percentage of the building that is heated (HDDpH), and the number of *Cooling Degree Days* multiplied by the percentage of the building that is cooled (CDDpC) [17][8]. The linear regression is represented by the following equation:

$$y = \sum_{j=0}^{5} a_j x_j \text{ , where y = LnE and } \vec{x} = \begin{bmatrix} 1 \\ \text{LnArea} \\ \text{LnWkHrs} \\ \text{LnNWker} \\ \text{HDDpH} \\ \text{CDDpC} \end{bmatrix}. \qquad (1)$$

The regression coefficients (a_j) are those which minimize the sum of the squares of the differences between the y's calculated above and the natural logs of the measured source energies (in kBtu).

The EPA indicates that these final regression variables were chosen after looking at the statistical significance of many potential variables and combinations of variables, keeping those that were statistically significant [17].

j	x	a	Δa	t-value	Pr(>|t|)	95%
0	1	2.79	1.20	2.33	0.0223	
1	LnArea	0.92	0.10	9.15	7.0E-14	x
2	LnWkHrs	0.46	0.30	1.54	0.1289	
3	LnNWker	0.22	0.09	2.33	0.0225	x
4	HDDpH/10^4	0.56	0.37	1.50	0.1382	
5	CDDpC/10^4	2.00	0.74	2.71	0.0082	x

Table 2. Results of EPA's medical office model regression. The independent variable is the natural log of the source energy (in kBtu). Values listed include the regression coefficients (a_j), their uncertainties (or standard errors), t-value, and the probability that such an association would arise by chance. The model produced an R^2 value of 93%. The last column indicates variables with significance at the 95% confidence level.

We have duplicated the EPA's regression using the "R" software environment for statistical computing and graphics [20]. The results of this linear regression using data from the 82 CBECS 1999 records are shown in Table 2. These results are in close agreement with those published by the EPA [17]. Listed are the regression coefficients a_j, the standard errors in these estimates, the "t-value" associated with each variable, and the likelihood or probability that such an association would arise by chance. The model R^2 is 0.9336, indicating that 93% of the variation in the LnE data are "explained" by the five regression

[8] Both HDD and CDD are expressed in degrees Fahrenheit and are based on the difference between the average daily temperature and 65 ºF (18.3 ºC).

variables – the vast majority of this associated with the single variable LnSqft. Three of the regression variables – LnArea, LnNWker, and CDDpC (highlighted in blue) – show correlation with statistical significance at the 95% confidence level or higher. (The coefficients a_4 and a_5 are multiplied by 10^4 for easier comparison.)

With the regression coefficients determined the next step is to use the regression formula to calculate predicted LnE for all the buildings in the model dataset and to combine these with measured source energy to calculate each building's energy performance ratio

$$EER = \frac{Ln\left(E_{\mathrm{Meas}}\right)}{LnE_{\mathrm{Predict}}} \ . \qquad (2)$$

Finally, the dataset is sorted in order of increasing EER. The sorted list is used to calculate the cumulative EER distribution, corresponding to the percentage of buildings with EER value less than or equal to a particular EER value. For the sampled buildings there are 82 records. Each record represents 1/82 = 1.22% of the total number of sampled buildings. The cumulative EER distribution for the sample buildings is graphed in Figure 1 as the solid blue squares.

The cumulative EER distribution for the 1999 US medical office building stock is calculated using CBECS weights (w_j, j = 1,...,82) associated with the n = 82 buildings.[9] The number of medical offices in the building stock is $N_{Tot} = \sum w_j = 86,714$, and the percentage of these buildings represented by the j^{th} building is just w_j/N_{Tot}. These numbers are combined to determine the cumulative EER distribution for all US medical offices, graphed as the solid red triangles in Figure 1. Note that large vertical discontinuities occur in the red triangles for buildings with relatively large weighting factors. For most of the EPA's building models these two distributions differ significantly – not so for this case. The solid green line represents the EPA's "smooth" fit to these distributions.[10] In their technical document the EPA claims the fit is accomplished with a 2-parameter cumulative gamma distribution [17]. We are not able to confirm this result and believe it to be false. The EPA responded to our FOIA request to learn the values of the parameters for this gamma distribution with "no documents found [21]." The EPA apparently cannot justify this claim.

To determine an ENERGY STAR score for a building, then, one would simply calculate its EER then use the above graph to look up the cumulative percentage that goes with that EER. For example, if the EER = 1.0, the EPA's green curve gives a cumulative percentage of 36%, indicating that this building's EER is lower than those for 64% of other buildings – yielding an ENERGY STAR score of 64. Of course one wonders why the EPA's green curve does not match the solid red triangles – the cumulative EER distribution of the medical office building stock. If one employs this curve (instead of the green curve) one finds that 42% of buildings have lower EER than 1.0, or, equivalently, 58% of buildings have higher EER – corresponding to an ENERGY STAR score of 58. The red curve is the correct curve to use.

But before making too much of this discrepancy we should consider the uncertainties or standard errors in the predicted LnE's for these buildings – an issue the EPA does not address for any of its ENERGY STAR building models. The standard errors or uncertainties in the regression coefficients (Table 1) lead

[9] This language is not quite correct. All of the medical office buildings in the 1999 stock are represented by the 93 CBECS records obtained prior to filtering. These 82 records those medical office buildings that are eligible for ENERGY STAR scoring.

[10] The fit should be to the solid red triangles which represent the national medical office building stock. Clearly the fit does not match either distribution. No acknowledgement or explanation of this discrepancy is offered in the EPA documentation.

to uncertainties in the predicted LnE values. Using the "R" software employed to produce entries in Table 1 we have calculated the standard errors in each "predicted" LnE. Since this number appears in the denominator of the EER, these errors translate into uncertainties in the EER's, shown as horizontal error bars in Figure 1.[11]

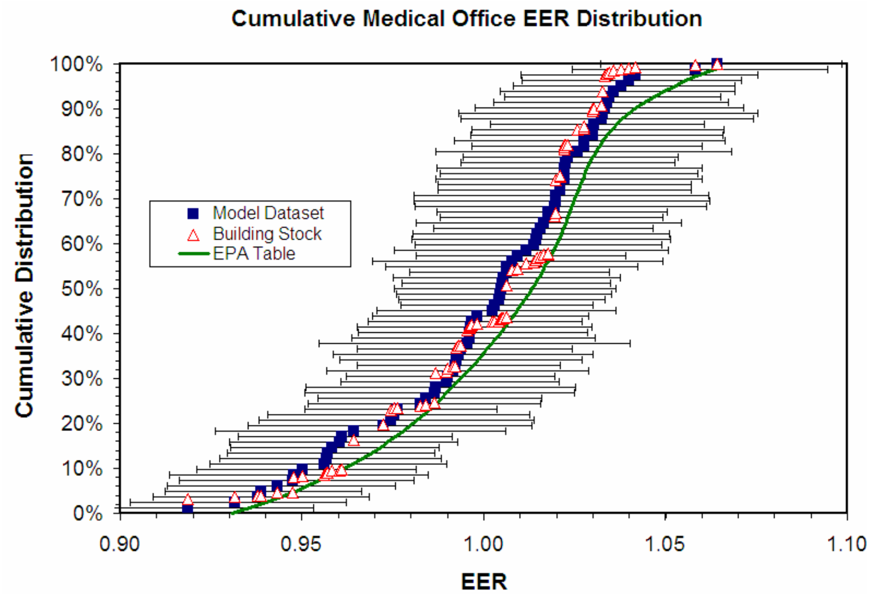

Figure 1. Graph of the medical office cumulative EER distribution for sampled buildings (solid blue squares) and the building stock they represent (open red triangles). The solid green curve represents the EPA's fit to the data used for assigning medical office ENERGY STAR scores. The horizontal bars represent the standard errors (uncertainties) in EER's (see text).

Since the issue of regression uncertainty is also important for the other 10 building models it deserves some discussion here. Consider one building in the model dataset. The building with PUBID7 = 3656 has 20,000 ft^2 (1860 m^2), measured source EUI = 131,6 kBtu/ft^2 (1490 MJ/m^2), WkHrs = 45, NWker = 20, HDD = 4558, CDD=1123, and has 100% of its space both heated and cooled. Taking logs we have LnE = 14.78, LnArea = 9.90, LnWkHrs = 3.81 and LnNWker = 3.00. Combining these independent variables with the regression coefficients in Table 1 we obtain a predicted LnE = 2.79 + 9.11 + 1.75 + 0.66 + 0.26 + 0.22 = 14.79. Dividing 14.78 by this value yields an EER of 1.002 which, as discussed in connection with Figure 1, is assigned an ENERGY STAR score of 64. But this value for the predicted LnE is calculated using the best estimates of the regression coefficients. Consider the impact of just the uncertainty in a_2 on this prediction. Replacing a_2 with $a_2+\Delta a_2$ in this calculation yields a predicted LnE, of 15.93 corresponding to an EER of 0.928. Looking at Figure 1 we now begin to understand the large error bars. Of course this considers only the effect of uncertainty in one coefficient, All of the coefficients have varying amounts of error. Since the errors in the various regression coefficients are correlated they cannot simply be added. Proper consideration of the covariance matrix yields the error bars shown in Figure 1.

---

[11] The standard errors for the regression predictions are calculated using the R-command "predict.lm(fit,medical.frame,interval="prediction", level=0.683). They represent one standard deviation in the regression prediction for particular values of the independent variables. This SE is larger than the SE obtained using the "confidence" interval, instead of "prediction" interval. We have investigated this issue using simulations and found that the "prediction" choice is indeed the correct one for this application.

It is clear that, given the uncertainties, it doesn't matter which of the three curves in Figure 1 are used to define the ENERGY STAR scores. Consider an EER = 1.0. The error bars in Figure 2 show that with 68% probability the EER value actually ranges from 0.957 to 1.030, which, using the EPA's green curve, implies a cumulative probability ranging from 10% to 80% and, corresponding to ENERGY STAR score that ranges from 20 to 90. The scores are so uncertain as to be meaningless. When the EPA issues an ENERGY STAR score of 64 for a medical office building it fails to disclose that the uncertainty in this scores is ±35 points – that is, the score is statistically no different from scores ranging from 20 to 90.

Readers may be confused by the fact that the data plotted in Figure 1 "hit" almost in the middle of each error bar – suggesting that the uncertainties in EER are not nearly so large as the error bars indicate. This is actually a consequence of the sorting process that results in the cumulative EER distribution. One could (literally) use a random number generator to produce the "predicted LnE" – and once sorted, the cumulative EER curve would be smooth and monotonically increasing. But this would not remove the uncertainty introduced by dividing the measured LnE by an arbitrary and uncertain predicted LnE.

The t-values and p-values in Table 1 demonstrate that three of the five independent variables are statistically significant in "predicting" the LnE for the 82 buildings in this *model dataset*. The other two variables also have predictive value, but not at the 95% confidence threshold. But this does not mean that these variables are generally useful in predicting source energy use for the estimated 90,000 medical offices represented by the *model dataset*. The key unanswered question is to what extent these 82 buildings adequately represent that larger building stock? Would a similar regression applied to a totally different building sample produce the same results?

The EPA's medical office model is built upon data from the 1999 CBECS. The EIA conducted another CBECS in 2003 – which provides an opportunity to use independent data to validate the EPA's regression model. 73 records in the 2003 CBECS correspond to medical office buildings. If the correlations observed between source energy and the five variables identified by the EPA are real then a regression on these 2003 data should yield similar results – consistent values for the regression coefficients with the variables showing similar statistical significance.

We have extracted 73 records from the 2003 CBECS for medical office buildings.[12] Three buildings were eliminated after being identified as "outliers." The EPA's regression was applied to the remaining 70 records and the results are listed in Table 3. The R^2 for the regression is 0.92, similar to that for the regression in Table 2.

j	x	a	Δa	t-value	Pr(>|t|)	95%
0	1	6.95	1.18	5.90	1.5E-07	
1	LnArea	0.78	0.11	7.34	4.6E-10	x
2	LnWkHrs	-0.37	0.25	-1.51	0.1350	
3	LnNWker	0.43	0.10	4.15	0.0001	x
4	$HDDpH/10^4$	0.51	0.37	1.38	0.1720	
5	$CDDpC/10^4$	0.87	0.68	1.29	0.2010	

Table 3. Results of regression applied to medical office records extracted from CBECS 2003. The independent variable is the natural log of the source energy (in kBtu). Values listed include the regression coefficients (a_j), their uncertainties (or standard errors), t-value, and probability that such an association would arise by chance. The model produced an R^2 value of 92%.

[12] These buildings correspond to (PBA = 2 AND PBAPLUS=5) OR (PBA = 8 AND PBAPLUS = 18).

The coefficients produced by the two regressions (with uncertainties) are compared in Figure 2. Coefficients for three of the regression variables agree within their uncertainties; the other two, a_2 (LnWkHrs) and a_3 (LnNWker) do not. For LnWkHrs the data sets yield coefficients with opposite sign – which cannot possibly make sense!

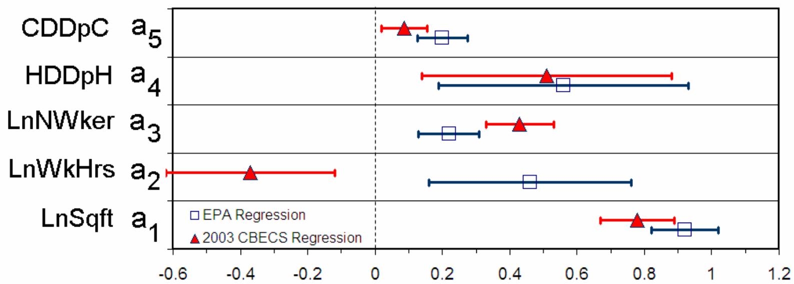

Figure 2. Direct comparison of the regression coefficients from the 1999 CBECS data (solid red triangles) and the 2003 CBECS data (open blue squares). Coefficients a_1, a_2, and a_5 agree within error while coefficients a_2 and a_3 do not.

Looking to the three variables for which the regression coefficients are in agreement we see that only building size (LnSqft) is significant at the 95% confidence level in both datasets. Cooling degree days, identified by the EPA as the most significant variable after LnSqft, is the least significant variable for the 2003 data set. Heating degree days (HDDpH) is not significant for either regression. In short, of the five independent variables identified by the EPA only building size (LnSqft) emerges as a consistent, significant predictor of source energy in both the 1999 and 2003 CBECS medical office data sets.

The differences in these two regressions is further illustrated by using each to calculate the "predicted source EUI" for the combined set of 82 + 70 medical office buildings from the 1999 and 2003 CBECS.[13] These predictions are compared in Figure 3 where one is plotted versus the other for each of the 152 buildings in the combined 1999 and 2003 CBECS medical office data subsets.

The significant scatter in the graph shows that two, presumably equally valid regressions predict very different source EUI for the same buildings. On average, predicted EUI from the regression to the 2003 data are 17 kBtu/ft^2 (190 MJ/m^2) higher than those predicted from the 1999 regression. The rms-difference in the predicted source EUI for the two models is 64 kBtu/ft^2 (730 MJ/m^2) – roughly 45% of the mean source EUI for either of the two datasets.

So which of these two regressions yields the correct "predicted" source EUI? Or, more precisely, which of these two model datasets accurately captures the properties of the larger medical office building stock? The answer, of course, is neither. If we included error bars on these predictions we simply confirm what is shown in Figure 1 – that both predictions are so uncertain as to be meaningless.

Another method for testing the validity of a regression and its underlying data is to randomly subdivide the model dataset into two halves and perform the same regression on each subset to see if similar results are obtained for each. If the results found for the entire data set are replicated in each subset the results are confirmed. If, on the other hand, one or more of the independent variables do not demonstrate

[13] For earlier graphs and regressions the site-to-source energy conversion factors were different for the 1999 and 2003 data. For example, the electric factor used for 1999 was 3.01 while for 2003 was 3.34. For this particular analysis all the source energies were calculated using the same 2003 conversion factors and the EPA's regression on 1999 medical offices was re-calculated using these source energies.

significance in both subsets, it suggests the observed correlation for that particular variable in one set is accidental.

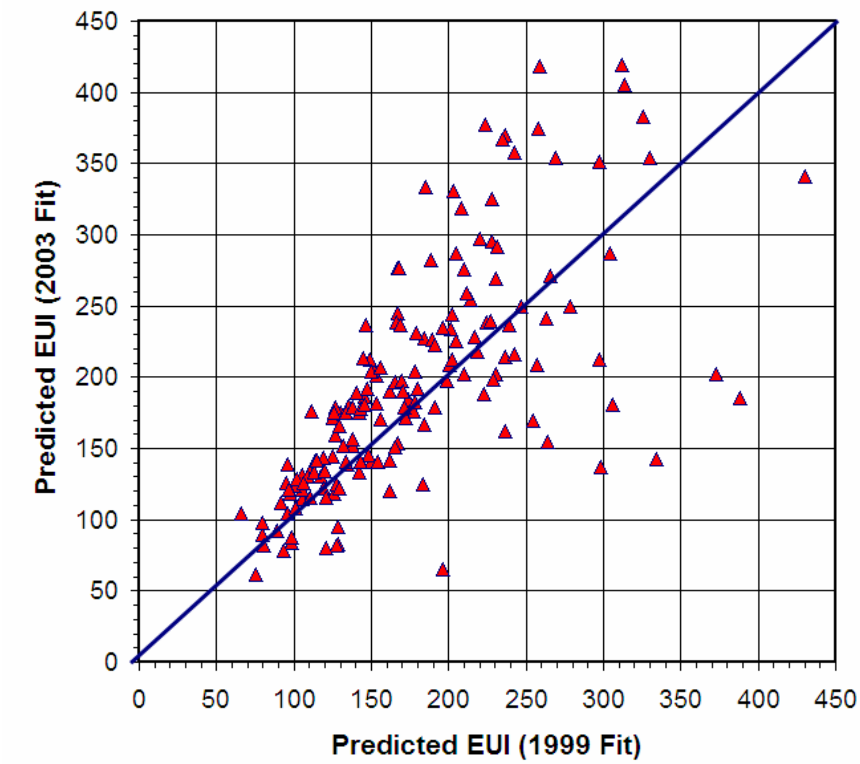

Figure 3. Graph of the predicted source EUI for combined 1999 and 2003 CBECS medical office building records using regression 2 (2003 data, Table 2) versus that predicted by regression 1 (1999 data, Table 3). The graph shows the "randomness" of the predicted source EUI that is associated with the two regressions.

We have carried out this process using the EPA's *model dataset*. The results are summarized in Table 4. Subsets A and B had 43 and 39 buildings, respectively. The significance of the variable LnArea is again confirmed – all three regressions (full data set, subset A and subset B) yield consistent regression variables a_1. For all three regressions t-values and p-values confirm the significance of this correlation. The significance of the variable LnWkHrs does not reach the 95% confidence level for any of the three sets, and clearly is not a key indicator for either of the two subsets – as indicated by p-values of 31% and 49% respectively. The variable LnNWker almost reaches the 95% confidence level for subset B (p-value 6%) but shows very low significance for subset A (p-value 41%). Heating degree days shows very low significance in both subsets A and B (p-values of 20 and 24%) and Cooling degree days exceeds the 95% confidence level in subset A while showing marginal significance in subset B (p-value 23%). In short, four of the five key variables identified by the EPA as being significant predictors of source energy consumption in medical office buildings do not hold up under this scrutiny.

The EIA employs a "jack-knife" method for estimating the standard relative error (SRE) in statistics calculated from CBECS building subsets [22]. The SRE is the percentage uncertainty in a particular statistic. In its own reports the EIA does not publish statistics for any subset for which the SRE is greater than 10% – judging such statistics to be insufficiently robust. For this reason the EIA does not include summary data for medical office buildings in any of its reports – their SRE is too large. We have implemented the EIA's method for estimating SRE for statistics for these two building subsets. The means and standard errors for three statistics are displayed in Table 5. These statistics are total number of medical office buildings, total gsf of these medical office buildings, and gross source EUI for these buildings. We see that for the two CBECS surveys the SRE's for total number of medical office buildings are 17-18% and for total gross floor area are 12-17%, and for source EUI are 7-19%. In short, the sampling of these types of buildings in either of the CBECS surveys is insufficient to determine these gross numbers to better than 7-18% accuracy – let alone to provide more detailed information regarding the distribution of their properties required for a statistically valid regression.[14]

j	0 - intercept			1 - LnArea			2 - LnWkHrs			3 - LnNWker			4 - HDDpH/10^4			5 - CDDpC/10^4		
	full	A	B	full	A	B	full	A	B	full	A	B	full	A	B	full	A	B
a	2.79	1.83	3.69	0.92	1.00	0.89	0.46	0.52	0.27	0.22	0.14	0.24	0.56	0.75	0.61	2.02	3.54	1.17
Δa	1.20	1.97	1.53	0.10	0.18	0.13	0.30	0.50	0.39	0.09	0.16	0.12	0.37	0.58	0.51	0.74	1.30	0.96
t-value	2.33	0.93	2.41	9.15	5.67	6.89	1.54	1.03	0.70	2.33	0.84	1.95	1.50	1.30	1.20	2.71	2.73	1.22
Pr(>|t|)	2%	36%	2%	0%	0%	0%	13%	31%	49%	2%	41%	6%	14%	20%	24%	1%	1%	23%

Table 4. Results of regressions on full EPA data set and randomly chosen subdivisions A and B (see text). Regressions yield consistent results for LnSqft but not for other four variables. p-values (last row) are rounded to nearest percent for ease of display.

CBECS	# Bldgs (1000's)	Gross Floor Area 10^6 ft^2	Gross Floor Area 10^6 m^2	Gross Floor Area SRE	Source EUI kBtu/ft^2	Source EUI MJ/m^2	Source EUI SRE	# Records
1999	115 ± 20	1050	100	12%	160	1800	19%	93
2003	85 ± 15	700	70	17%	140	1600	7%	74

Table 5. Medical Office building stock as represented by 1999 and 2003 CBECS data. Standard relative errors (SRE) are calculated using jacknife technique employed by EIA (see text).

These data simply are not sufficiently robust to provide any meaningful characterization of the properties if the U.S. medical office building stock.

5. Discussion

All of the above results – (1) the large uncertainty in the predicted source energy and resulting ENERGY STAR scores, (2) the significant variation in regression results from an equivalent building set drawn from the 2003 CBECS, and (3) the inconsistent results of regressions on two halves of the EPA's model dataset – all demonstrate that there is little substance behind the predictive power of the EPA's regression analysis for medical office buildings. The root problem is that the number of buildings in the model dataset is too few to offer any statistically significant characterization of the U.S. medical building stock. This is true of the 93 medical office records extracted from the 1999 CBECS (or 82 after filtering) used by the EPA and it is also true of the 73 medical office records extracted here from the 2003 CBECS used for the regression shown in Table 3.

The EPA has used this medical office model for nine years and continues to use it today. From 2006 through 2012 the model was used to issue scores for 4,622 medical office buildings [23]. At least 184 different medical office buildings have received ENERGY STAR certification using this model – many for multiple years [24]. New York City just released its 2012 public energy benchmarking data that includes ENERGY STAR scores for at least 30 medical office buildings [25]. These building owners are, by law, required to submit their data to *Portfolio Manager* for scoring and this score is made public.

[14] Note that the SRE's are even larger for the filtered medical office subsets used for the regressions.

In its 2012 DataTrend publication the EPA claims energy savings by 35,000 buildings that used *Portfolio Manager* four consecutive years, 2008-2011[26]. The EPA points to the fact that average ENERGY STAR scores for these buildings increased by 6 points over this period as evidence of energy savings for these buildings. The 35,000 buildings include 803 medical office buildings whose average ENERGY STAR score went up by less than four points during this period [27]. As is demonstrated in Figure 1 – these ENERGY STAR scores have standard errors of 30-40 points. An increase of four points in the average score is statistically meaningless.

The number and gross floor area associated with medical office buildings are small when compared with the totals for all buildings scored by ENERGY STAR building models. But all of the problems identified here for the EPA's medical office model are present in its residence hall/dormitory model – also introduced in 2004. Analysis of the dormitory model dataset exactly parallels that presented here – with identical conclusions [28]. The EPA's dormitory building model is built on 79 records from CBECS 1999 – assumed to represent an estimated 35,000 residence halls in 50 states. ENERGY STAR scores issued for residence hall/dormitory buildings are no more meaningful than those for medical offices. Moreover, the size of the model datasets for three additional ENERGY STAR building models – Hotel, Retail, and Supermarket – suggest these suffer from similar deficiencies. And preliminary analysis of the impact of uncertainties in regression coefficients for 10 of the 11 EPA ENERGY STAR models for conventional building types suggests they all yield large (no less than ±10 pts) uncertainties in the ENERGY STAR scores they produce [28].

Errors in the EPA's *Medical Office Technical Methodology* document combined with the EPA's inability to supply relevant model data in response to *Freedom of Information Act* requests demonstrate that the EPA has failed to document its own methodology. The most important error is that the EPA claims this model does not apply to buildings under 5,000 ft^2 (465 m^2).in size. Yet it is clear 17% of the buildings in the EPA's *model dataset* – which with their combined weights represent 47% of the buildings in the medical office building stock – are buildings with such floor area. For many of its building models the EPA claims to filter out such small buildings saying, as they did for medical office buildings, "analysis could not model behavior for buildings smaller than 5,000 ft^2." This text appears to be no more than "boiler plate" pasted into most of the EPA's technical methodology documents without justification. Such errors go uncorrected largely because the EPA operates in secret – its science is not subjected to peer-review and its data are not made accessible to others for independent analysis.

Many of the EPA's older ENERGY STAR building models have been revised to correct errors; it is hard to understand how errors in the Medical Office model – particularly documentation errors – went undiscovered after 2007 while the EPA reviewed and revised other building models. But fixing the medical office model would require that the EPA obtain new, statistically valid data for the medical office building stock. As shown above the 2003 CBECS did not provide such data. When the EIA cancelled their 2007 CBECS this would have left the EPA in an awkward position in which it should have withdrawn the medical office model. That did not happen. Instead, the EPA has continued to issue scientifically meaningless ENERGY STAR scores for medical offices.

6. Summary and Conclusions

Here we have outlined the EPA's methodology for calculating ENERGY STAR scores for its 11 conventional building models. We have further looked closely at the EPA's implementation of this methodology for one particular building type, medical offices. In doing so we find that the EPA's document that describes this model contains key errors and provides an inaccurate description of its own model data set, a description the EPA was unable to correct in response to a FOIA request. Specifically we find that the EPA data filters for hours of operation, number of employees, and building gross floor area are incorrectly specified. One consequence of this error is that the EPA has, for nine years, claimed that the model does not apply to buildings under 5,000 ft^2 (465 m^2) in size when, in fact, buildings this small provide 17% of the data on which the model regression was built.

Our analysis of the EPA's regression suggests that strong correlations observed for four of the five key variables identified by the EPA as "predictors" of energy consumption are "coincidental" to this specific data set – not found in either a second dataset or even in randomly selected "halves" of this dataset. In any case we show that the standard error in the "predicted source energy" from the EPA's regression model is so large that ENERGY STAR scores it produces are uncertain by ±35 points, rendering them meaningless – nothing more than placebos.

The fundamental problem with this building model is that the underlying data, from just 82 samples, are inadequate for representing the gross properties of the estimated 100,000 medical office buildings in the commercial building stock, let alone providing sufficient detail to understand how medical office energy use varies with other building variables. The model is based on CBECS data judged by the EPA to be of insufficient accuracy to be used in their own reports. The model should never have been introduced in 2004 and surely should have been withdrawn in 2007 when other models were revised. The EPA's continued use of this model raises serious questions regarding the quality of the science behind the EPA's ENERGY STAR building program. Preliminary investigation suggests that the fundamental problem identified here is not limited to the medical office model but is also present in as many as half of the other 10 conventional building ENERGY STAR models.

Acknowledgements

The author expresses appreciation to Alexandra Sullivan (EPA) for assistance in understanding ENERGY STAR modeling and Michael Zatz (EPA) for motivation to learn how to file FOIA requests. Thanks also to Jeff Witmer (Oberlin College) for assistance with statistical issues and Larry Spielvogel for helpful discussions.

References

1. Thomas W. Hicks, "The ENERGY STAR Building Label: Building Performance through Benchmarking and Recognition," *ACEEE Summer Study on Energy Efficiency in Buildings*, (4) pp.205-210 (1998).
2. Terry Sharp, "Energy Benchmarking in Commercial Office Buildings," *ACEEE Summer Study on Energy Efficiency in Buildings*, (4) pp.321-329 (1996).
3. Terry Sharp, "Energy Benchmarking Energy Use in Schools," *ACEEE Summer Study on Energy Efficiency in Buildings*, (3) pp.305-316 (1998).
4. http://www.energystar.gov/buildings/facility-owners-and-managers/existing-buildings/use-portfolio-manager
5. http://www.usgbc.org/
6. http://www.thegbi.org/
7. http://www.imt.org/policy/building-energy-performance-policy (accessed 10-3-2013).
8. Johnson, J. 2002, "Is What They Want What They Get? Examining Field Evidence for Links between Design Intent and As-Built Energy Performance of Commercial Buildings." In

Proceedings of the 2002 *ACEEE Summer Study on Energy Efficiency in Buildings*, 4:161-170. Washington, D.C.

9. Turner C. and M. Frankel. 2008. "Energy Performance of LEED for New Construction Buildings – Final Report." *New Buildings Institute*, White Salmon, WA.
10. EPA – DataTrends: Energy in Office Buildings, downloaded 2013-09-21, http://www.energystar.gov/buildings/tools-and-resources/datatrends-energy-use-office-buildings
11. John H Scofield, "A Re-examination of the NBI LEED Building Energy Consumption Study," *International Energy Program Evaluation Conference*, Portland, OR, August 12-15, 2009.
12. John H. Scofield, "Efficacy of LEED-certification in reducing energy consumption and greenhouse gas emission for large New York City office buildings," *Energy & Buildings*, vol. 67, pp. 517-524 (December 2013).
13. See http://www.energystar.gov/buildings/tools-and-resources/technical-documentation
14. http://www.energystar.gov/ia/business/evaluate_performance/General_Overview_tech_methodology.pdf
15. Woods JE, Sweetser R, Novosel D. Scientific Outreach Program Pilot. Alexandria, VA : National Center for Energy Management and Building Technologies, 2009. Final Report 090717 to US DOE; Section 4.1: Predominant Building Energy Performance Metrics.
16. See http://www.energystar.gov/ia/business/evaluate_performance/site_source.pdf?84c1-a195
17. http://www.energystar.gov/buildings/tools-and-resources/energy-star-score-medical-offices
18. Alexandra Sullivan, private communication, July 21, 2013.
19. FOIA request EPA-HQ-2013-009270, filed August 21, 2013, final disposition October 4, 2013.
20. http://www.r-project.org/
21. EPA-HQ-2013-009668, filed September 5, 2013, final disposition October 23, 2013.
22. ftp://ftp.eia.doe.gov/pub/consumption/commercial/micro.data/ce92tech.pdf
23. "ENERGY STAR Portfolio Manager DataTrends – Energy Use Benchmarking," http://www.energystar.gov/buildings/sites/default/uploads/tools/DataTrends_Energy_20121002.pdf?f83d-fb07
24. http://www.energystar.gov/index.cfm?fuseaction=labeled_buildings.locator
25. http://www.nyc.gov/html/gbee/downloads/excel/2013_nyc_ll84_disclosure.xlsx
26. "ENERGY STAR Portfolio Manager DataTrends – Benchmarking and Energy Savings," http://www.energystar.gov/buildings/sites/default/uploads/tools/DataTrends_Savings_20121002.pdf?f473-0dbc
27. FOIA request EPA-HQ-2013-004969, filed March 28, 2013, closed May 2, 2013.
28. John H. Scofield, unpublished.