

External Peer-Review of Thermostat Collection Study
Nicholas P. Jewell, Ph.D., University of California, Berkeley
October 15, 2012

This is my formal response to your letter of September 24, 2012 charging me with providing a confidential and written external review of the general approach and methodology associated with a study conducted for thermostat manufacturers pursuant to the Thermostat Collection Study. The purpose of the study itself was estimation of the number of mercury-added thermostats becoming waste each year in California (in the future). The study is described in two documents provided to me directly^{1,2}. Specifically you requested that I address three specific issues: (i) the sample methodology, (ii) the survey response rate and bias reduction analysis, and (iii) the method for estimating waste flows.

Executive Summary

In my view, the thermostat collection study performed by Skumatz Economic Research Associates does not meet the standards of rigorous scientific methodology with regards to statistical estimation. It fails to account for serious sources of bias both due to data collection procedures and statistical methodology employed. It fails to report important information regarding the collected data. It fails to use up-to-date statistical approaches to lifetime distribution estimation. Finally, it fails to report a range of estimates that accounts fully for uncertainty in estimation. Many of the issues could be addressed by a more professional approach to the relevant statistical issues although the quality of the survey likely cannot be improved. While the study's reported range of the flow of mercury-containing thermostats is already broad, accounting for all of these issues in an appropriate scientific manner would undoubtedly make the range of uncertainty substantially greater, plausibly up to at least +/-250,000 thermostats. It is unclear to me that this large range of uncertainty is of value to the California Department of Toxic Substances but an answer to that question is outside the range of my expertise. Alternative methods for estimation should be considered, in addition to a thorough re-analysis of the data provided by the existing survey instrument. The analysis of available data could be substantially improved with resulting estimates and conclusions made significantly more reliable even in the absence of further data collection.

Specific conclusions requested from peer-reviewers now follow:

1. Sample methodology: **the selection of random samples of possible respondents was reasonable;**

¹ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Buildings, Technical Approach Summary, Skumatz Economic Research Associates, Inc., March 1, 2009.

² Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009.

2. Survey response rate and bias reduction analysis: **the non-response rate was apparently extremely high undermining the value of all the data presented, particularly in the absence of an analysis of non response bias. The validation samples were small but used correctly although the variability in estimated underestimation and misclassification rates was not used in subsequent estimation. The issue that some of the missing data is not likely to be missing at random was not addressed satisfactorily.**
3. Method for estimating waste flows: **the methodology for estimation of the lifetime distributions of thermostats containing mercury suffered from several deficiencies that are likely to lead to significant bias in both estimation of the flows and the uncertainty associated with such. Existing statistical methodology to address these issues was not employed.**

Introduction

The basic strategy used by Skumatz Economic Research Associates, as outlined in the two versions of the study report, is basically as follows:

- (A) Split the problem into two parts: California (i) businesses, and (ii) households
(it is unclear how the overlap of businesses that use household addresses was accommodated);
- (B) Use a statewide survey to both estimate the existing number of mercury-containing thermostats in current use (and thus subject to removal in the future), and to obtain data on natural lifetimes of said thermostats;
- (C) Use the data on the lifetimes to estimate an age-dependent yearly rate (or *hazard*) of removal;
- (D) Apply the estimated hazard function to the estimated number of existing mercury-containing thermostats to yield an estimate of the anticipated number of removals each year out into the future (the *anticipated flow by year*).

The overall strategy is basically sound and reasonable assuming sufficiently accurate data can be collected and associated sampling biases accommodated. The remaining scientific concern therefore is whether each of these steps can be carried out effectively, and to assess the biases and uncertainty associated with each step of the estimation strategy.

Sampling Design and Implementation

The sampling design used sampling frames provided by InfoUSA, a commercial organization. In part (i), the business survey, the investigators used a stratified

sampling design based on the number of employees associated with a business. This is a sensible approach to sampling since stratification may provide increased precision (when the size of the business is closely related to the number of thermostats in the business' properties), and can never be less precise on the average assuming that stratum sample sizes are selected appropriately. One standard practice would be to choose strata sample size proportions to be the same (e.g. 5% of each stratum) so that the resulting sample would be representative (or "self-weighting" in survey jargon). An alternative approach (and more desirable with regard to precision) would account for sampling frame information regarding the variability in the thermostat counts within the strata (with the implication to choose more sampled units from strata with high variability in thermostat counts). In this direction, the study chose a greater sampling fraction for the "larger" businesses based on the supposition that small businesses were likely to be more homogeneous³. This approach is entirely reasonable and, even if the sample allocation were not absolutely optimal, only precision in estimates would be affected as no bias can be introduced by inappropriate sample allocations.

For the household survey, part (ii), simple random sampling was employed, presumably because there was no useful frame information available for any form of stratification. This choice was based on the assumption that thermostat use in households does not vary substantially by geographic location, the age of a home, nor the type of home (single versus multi-family building, for example). (The type of home variable raises the issue of whether the sampling frame was based on addresses or household names, etc.: this issue was not described in the detailed descriptions of the survey.) To the extent that these assumptions are incorrect, only an opportunity for additional precision might be lost, as no bias would be introduced by simple random sampling.

The sample sizes ultimately selected were intended to generate from 270-380 responses for both parts of the survey intended to achieve a margin of error of around $\pm 5-10\%$, assuming a count of 724,000 businesses and more than 13 million households in California⁴. Of course, a greater number of units needs to be sampled in each case to allow for non-response, to which we return below. The final number of survey responses obtained was 595 residential and 267 commercial⁵.

³ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p. 10.

⁴ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, Table 3.1, p. 10.

⁵ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, Table 4.1, p. 15. There is a slight discrepancy in that 594 residential responses are noted on p. 10.

The survey was implemented through a postcard outreach with information to be provided via a website (with a phone survey possible to any respondent reluctant to use the website)⁶. There were two attempts to contact respondents. Apparently, 30 large businesses were called by phone to improve the response rate in this stratum⁷. It is unclear how these particular businesses were selected.

Measurement Error/Validation

The key variables obtained from survey respondents were the number and type of thermostats currently employed, and information regarding their current age and about the lifetimes of past thermostats⁸.

A small number of respondents were selected for on-site validation of reported thermostat counts (5 businesses, 4 apartment units, and 21 single family households)⁹. It is unclear how these respondents were selected for validation. Both validation sets indicated some undercounting of total number of thermostats (an estimated 18% “undercount” overall for the 5 commercial properties selected for validation, and an estimated 3% undercount for the 25 residences¹⁰). Regarding the type of thermostats reported, there were no errors in the commercial validation sample and four household errors resulting in an estimated total classification error of 5%¹¹.

A small separate validation was attempted by sending disposable cameras to 44 businesses and residences randomly selected from respondents to the initial survey.

⁶ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p. 11.

⁷ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p. 11.

⁸ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p. 12 and Appendix C.

⁹ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p.14.

¹⁰ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, Table 1.6.

¹¹ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p. 14.

Only 12 responses of these had been received at the time of the writing of the reports. From this small sample, the estimated overall undercount rate was 8%¹².

Combining the validation information from both approaches, the investigators report an estimated overall undercounting of thermostats of about 13.5%, and type misclassification of 3.4% (the latter all applying to residential responses)¹³.

Response Rates

As might be anticipated, there is non-response as indicated by the variation in respondents in Table 4.3 of the longer study report, for example. While the investigators describe the impact of non-response on the representativeness of the sample, I did not find quoted the overall response rate for either sample. That is, how many cards were sent out to achieve the 267 commercial and 595 residential respondents? The absence of these response rate figures is unacceptable as the size of the response rate drives the level of bias anticipated by non-response (in addition to the differences between respondents and non-respondents). Attachment 2¹⁴ (amongst documents sent to me for review) suggested that postcards were sent to lists of 10,000 residences and 12,000 businesses. This suggests that the non-response rates were astronomically high, opening the door to extraordinarily high levels of non-response bias. While this may represent “the best that we can do or expect” in such surveys, in my view this level of non-response begs for some form of assessment of non-response bias, perhaps including on-site visits to a sample of non-respondents.

The non-response has an immediate effect of disturbing the representativeness of both samples at least as viewed through the strata used to assess response rates. Thus, some strata may be less represented in the sample than they should be given the population distribution. This is straightforward to correct through re-weighting as is done here both for stratified samples (businesses) and through post-stratification (residences). On the other hand this only corrects for bias caused by non-respondents differing from respondents on the basis of the stratification factor. Without knowing the overall response rate it is simply impossible to speculate as to the extent of additional bias caused by using data from respondents to represent the entire population (or strata to be specific). This is not addressed by either validation method as they simply look at respondents. Nor is it addressed by the missing data analysis that again only concerns properties of respondents. (As it happens, 27% of the data on the number of existing thermostats is missing for the commercial

¹² Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, Table 1.6.

¹³ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p.15.

¹⁴ Attachment 2: Mercury-Containing Thermostats: Description of Scientific Conclusions to be addressed Reviewers

sample, with 8% being the corresponding number for households¹⁵.) The missing data, of course, not only reduces precision because of the smaller sample size but also introduces the potential for a different bias, since respondents with missing data may differ from those who provide such information. The investigators deal with this in an identical manner to non-response by adjusting the sampling weights within the described strata. Of course, this will not deal with missing data bias associated with variation according to other variables. The failure to report response rate and a qualitative and quantitative assessment of non-response and missing data biases represent a serious deficiency of the study. In summary, the potential for bias caused by the sampling scheme and implementation is potentially very large.

Estimation of the Number of Mercury-Containing Thermostats

It appears as if estimates of the number of thermostats existing in California structures, based on this survey, only used the discussed pre- and post-stratification factors and not other attributes of the buildings covered by respondents. Other stratification, or regression, adjustments are, of course, logically possible assuming appropriate available data, as used, for example in an analogous survey carried out in King County, Washington¹⁶. This kind of approach is not discussed in the reports. The impact of omitting useful such adjustments relates to precision rather than bias, however.

Analysis of Lifetimes of Mercury-Containing Thermostats

The data on thermostat “lifetimes” is obtained from the website questionnaire that is intended to elicit lifetimes of current thermostats in place, and potentially the lifetimes of previous thermostats that preceded current versions. In the parlance of lifetime data analysis, the ages of current thermostats provide *censored* observations (since their lifetimes are not yet complete), whereas data on previous thermostat lifetimes are uncensored observations (since the lifetimes of these thermostats are assumed known). The investigators use these data and the standard Kaplan-Meier method to estimate the distribution of the natural lifetimes of thermostats¹⁷. It is not clear that the lifetime distribution in Figure 4.1 of the longer report applies to all thermostats, or is restricted to mercury-containing thermostats only. This should be clarified as otherwise bias may be introduced absent evidence that the lifetimes of mercury-containing thermostats are identical (in general) to those not containing mercury.

¹⁵ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, Table 4.1.

¹⁶ Attachment 6: Watson, Taylor, *Mathematical Model Estimating Thermostats in Commercial Buildings, Final Report*. King County, Washington Local Hazardous Waste Program. June 2006.

¹⁷ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, Figure 4.1.

Subsequently, the study fits a so-called gamma distribution to these data but this seems entirely unnecessary as the Kaplan-Meier estimator natural can produce a year-by-year risk (hazard) that can estimate the probability that a thermostat of a certain age will fail in the next year (without any bias that might be induced by a parametric model like the gamma distribution). It is possible that the tail for the estimated distribution covers lifetimes that are only observed in exiting thermostats (the censored observations) thereby requiring a parametric model to extrapolate to longer lifetimes. As emphasized below, descriptive statistics should be reported requiring the number and properties of both censored and uncensored lifetime measurements. Finally these estimated hazards are applied to the estimated counts of existing thermostats by age to estimate the numbers expected to fail in any given year¹⁸. This is satisfactory since no new mercury-containing thermostats are being introduced due to a ban commencing in 2006¹⁹. One concern however hearkens back to the issue of non-response and missing data in that lifetime information may very likely differ between respondents and non-respondents.

Unfortunately, there are three major methodological problems with this approach to estimation of the lifetime estimation of thermostats. There is a very significant literature on the estimation of lifetime distributions based on what is known as *recurrent event data*, exactly of the kind considered in this data analysis. The first problem is that the censored observations are well known to suffer from *length bias*²⁰ (necessarily, thermostats with shorter lifetimes are less likely to be still current and thus appear as censored observations), and so cannot be directly used with either a nonparametric (Kaplan-Meier) or parametric (here, gamma) approach to estimation. Ignoring this fact overestimates average lifetimes, and therefore will underestimate the flow of thermostats from businesses and residences. This phenomenon can also be viewed as a form of *dependent censoring*. (A very quick simulation that I performed (not reported here) showed that estimates of cumulative probabilities of the kind reported in Appendix B of the longer study report could be biased by more than 10% in the middle of the distribution.) There are published methods to address this form of sampling bias. In addition, in another quick simulation that I carried out, using (i) all of the uncensored observations, or (ii) only the last known full prior lifetime, produced a much less unbiased estimator than one that uses the censored data on current thermostats as well. However, this remains to be fully investigated and I did not allow for measurement error or

¹⁸ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p. 20.

¹⁹ Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings. Skumatz Economic Research Associates, Inc., December 28, 2009, p. 20.

²⁰ See, for example, Wang, M-C. and Chang, S-H., Nonparametric estimation of a recurrent survival function, *Journal of the American Statistical Association*, 1999, **94**, 146-153.

missing data on lifetimes that is more and more likely dates recede into history (as discussed in further detail below); the likelihood of increased missing data and misreporting further back in time is one reason that a statistician might prefer (ii) to (i) here. Observed differences (between these two estimating approaches) provides an indirect way to examine these concerns.

Second, repeated lifetimes from the same building (either commercial or residential) are likely to be correlated and this also introduces another form of dependent censoring (and thus bias) and also distorts the standard error (uncertainty) calculations as reported in Appendix B²¹. For example, long prior lifetimes are likely to lead to subsequent lifetimes that are automatically censored. Methods are available to address this issue adequately but are not used.

Third, recall data on *prior thermostat lifetimes* is likely to suffer from far more systematic and random measurement error than affects reported installation dates of current thermostats (and thus their censored lifetimes). For example, the longer back in time recall is required, the more likely it is that ownership has changed hands and appropriate knowledge and records are lost. It is unacceptable that the report does not contain information on the *number* of censored and uncensored observation underlying Figure 4.1, for example. This information would at least allow for consideration of how much bias might be introduced here by recall bias that may be differential across censored and uncensored observations. Systematic error may be further compounded through missing data on prior thermostat lifetimes, for example, when prior short lifetimes are more likely remembered and reported as compared to longer prior lifetimes that are either unknown and/or not reported. Unfortunately, the most accurate data is presumably information on existing thermostats, but this information suffers from length bias as noted above and does not yield any completed lifetime observations.

As the noted references and related literature shows, there are existing methods that statisticians have devised to deal with each of these issues, but these techniques are ignored entirely. Given the possible offsetting nature of various biases induced, it is impossible to speculate exactly how large the bias in estimates (and related standard errors) might be. In addition, some of these biases are influenced by the *birth process* that describes the rate of introduction of mercury-containing thermostats over time. The assessment of bias and uncertainty measures might usefully be considered through simulation methods based on existing rough estimates of information regarding input rates and lifetimes.

Other important, but more minor issue concern the fact that the lifetime distribution analysis has ignored the weighted nature of the sampling (why are estimates not

²¹ See, for example, Schaubel, D. E. and Cai, J., Non-parametric estimation of gap time survival functions for ordered multivariate failure time data, 2004, **23**, 1885-1900, or Luo, X. and Huang, C-Y., Analysis of recurrent gap time data using the weighted risk-set method and the modified within-cluster resampling method.

also weighted? The absence of weighting makes an unnecessary, and perhaps implausible, assumption that thermostat lifetimes do not vary in character depending on the size of the associated businesses that use the relevant buildings. Similarly, there is no discussion as to whether the thermostat lifetime distribution varies between commercial and residential properties. Again, it would be straightforward to avoid an assumption of no relationship between lifetimes and the nature of the building by implementing appropriately stratified estimators.

Estimate Ranges

Ultimately, much of the report and discussion refers to all thermostats, not just those containing mercury. These numbers need to thus be adjusted by using the type information under various scenarios, leading to the “lower” and “high” estimates, contained in Table 1.1, for example. For example, the Executive Summary of the report ultimately indicates a point estimate of 5.8 million mercury-containing thermostats in California with a range of 5.1—9.3 million, for example, or 5.1 to 10.5 million if 13.5% undercounting is accommodated (as illustrated in the validation sample; it is unclear why the lower range is not also raised due to the undercounting). This subsequently leads to a range of 237,000 – 490,000 mercury-containing thermostats flowing out of buildings in “Year 1”.

However, these limits do not reflect any estimation uncertainty whatsoever; such uncertainty arises from sampling variation, uncertainty in estimating type, uncertainty in the validation sample, and uncertainty regarding estimation of the lifetime distribution. That is, there is no analysis of the propagation of errors that will necessarily lead to much broader ranges than those reported and quoted above, leaving aside the very substantial issues of bias already raised earlier. It is therefore entirely plausible that the number of extracted mercury-containing thermostats could be significantly higher than 500,000 in the first year. Whether addressing the impact of large biases, coupled with an appropriate assessment of full uncertainty, would suggest ranges that are not sufficiently narrow to be of value to the California Department of Toxic Substances is outside of the range of my expertise. There is no discussion in either report regarding the required level of accuracy demanded to provide effective reporting regulations.