Combining Incomplete Information From Independent Assessment Surveys for Estimating Masonry Deterioration

William B. FAIRLEY, Alan J. IZENMAN, and Steven M. CRUNK

Construction materials used in building structures, such as masonry, wood, and reinforced concrete, deteriorate over time because of many factors including poor design, defective materials or manufacture, and poor workmanship. This article is concerned with estimates of masonry deterioration and the effects of covariates on the damage to bricks on the walls of five multiple-story buildings of a residential complex located in the Bronx, New York. In this case study, the damage of primary interest was a "spall," a physical separation of a portion of the brick face from the body of the brick. Eventually, the face becomes so damaged that portions fall off. The result is an unattractive appearance and a hazard to passersby. In this study, spall damage was assessed by means of three different and independent condition assessment surveys: an expensive, precise, and hence very limited scaffold drop survey and two additional inexpensive, but more detailed photographic and visual surveys. The photographic survey was obtained by photographing the walls of the entire residential complex, and the visual survey was done by individuals walking around the periphery of each building and making a visual assessment of the damage to each wall. In the photographic survey, a large amount of incomplete data was unavoidable because of either poor photo angles or various physical obstructions. A binomial regression model using four categorical explanatory variables or factors was fitted to the observed photographic spall data. Sparseness of the data, the presence of outliers, and overdispersion were major problems encountered in selecting and fitting a suitable model. A small pilot survey, in which the relevant portions of the photographic and visual surveys were matched to 11 drop locations of the scaffold survey, recorded spall counts using each survey method. From this pilot survey, photographic and visual spall data were calibrated to the scaffold drop survey data. It was determined that of the two surveys, only the photographic spall survey was needed to predict scaffold spalls. The estimate of total damage from the photographic survey was then adjusted using the calibration results. Finally, a multiple imputation procedure was used to impute values for the missing data and obtain an estimate (and its standard error) of the true spall rate over the entire residential complex. Sources of uncertainty to factfinders in legal trials are discussed and illustrated through the present case.

KEY WORDS: Binomial regression; Calibration; Condition assessment survey; Generalized linear model; Incomplete data; Interaction; Litigation; Model selection; Multiple imputation; Outlier; Overdispersion; Sparseness.

All in all, it's just another brick in the wall.
Pink Floyd (1979)

1. INTRODUCTION

An extremely important problem for civil engineers is how to assess the state of health of constructed facilities, such as building structures. Guidelines for assessing deterioration or damage to building structures (i.e., *condition assessment*) have been published by the American Concrete Institute (1984) and the American Society of Civil Engineers (1991). Condition assessment can be important in litigation; for example, when settlement amounts depend on the strength of the legal case and the expected evidence as to the extent of damage. Precise condition assessments for large, complex building structures are often too difficult to obtain, but they can be estimated by sampling.

We focus on a specific case study involving deterioration of the exterior brick masonry walls and parapets of a residential complex of five multiple-story buildings located in the Bronx, New York. Since 1972, when the buildings were constructed with walls made of specially designed jumbo-sized bricks, a number of engineering surveys had studied various types of brick damage, including cracking, efflorescence, spalling, and weathering, throughout the complex. The damage of primary interest in this particular study is spalling. A *spall* is a physical separation of a portion of the brick face from the body of the brick. Eventually, the face becomes so damaged that portions fall off. The result is an unattractive appearance and a hazard to passersby.

In the case that we analyze here, it was agreed that the prima facie mechanism for the spalling of these bricks was winter freeze–thaw weather cycles that occur when the ambient temperature changes from below freezing to above freezing and vice versa. In any given winter (December–March in New York) about 30–50 freeze–thaw cycles occur. There was no agreement, however, over why water was present in the exterior brick surfaces in the first place. Some attributed it to wind-driven rain that entered the surface of bricks through pores and then subsequently froze; others, to water entering through cracks in the mortar, leading to interior water retention due to blocked weep (or drainage) holes and channels clogged with mortar. The proponents of the wind-driven rain theory contended that faulty brick manufacture was the primary cause

William B. Fairley is President, Analysis and Inference, Inc., Swarthmore, PA 19081 (E-mail: *wfairley@analysis-and-inference.com*). Alan J. Izenman is Professor of Statistics and Senior Research Fellow, Department of Statistics, The Fox School of Business and Management, Temple University, Philadelphia, PA 19122 (E-mail: *alan@sbm.temple.edu*). Steven M. Crunk is Assistant Professor, Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102. The three types of condition assessment surveys described in this article were performed by KCI Technologies, assisted and counselled on statistical methods and procedures by Analysis and Inference, Inc. The authors thank Robert M. Blum, Jonathan J. Fink, Ronald J. Hunsicker, H. W. H. West, and A. Rhett Whitlock for allowing them to use the data and for helpful discussions and communications on issues of damage to building structures; Dirk F. Moore, Paul R. Rosenbaum, Donald B. Rubin, and John Jackson for discussions on statistical issues; Richard M. Heiberger and Liansheng (Nathan) Wang for computational assistance; and the many individuals, including Roderick J. A. Little, Diane Lambert, associate editors, and anonymous referees, who commented on previous versions of this article.

of spalling, whereas those favoring the blocked drainage holes theory argued that factors other than brick manufacture, such as poor masonry work, poor design, and poor building maintenance, were to blame for the spalling.

The owners of the residential complex sued the brick manufacturer for supplying defective bricks. If the brick manufacturer were to be found liable at trial for having caused the damage, then the dollar damages awarded to the building owner would be based on estimates given at the trial as to the total extent of spalling throughout the complex.

The analysis that we describe began with data on spalling from a photographic survey and from a visual survey. This analysis is similar to that presented at the trial by one of the authors. However, "an analysis is only as good as the data on which it rests" (Kaye and Freedman 1994, p. 341). In the present case, the data on spall counts in specially selected areas of walls that were originally made on behalf of the building owners were demonstrated to be grossly unreliable. These areas were referred to as "scaffold drops," because they were examined using scaffolds dropped from the rooftops of buildings. Counts had been made from the same areas of walls at 5-year intervals. Examination of the detailed maps of the wall areas showed that numerous spalls that had been found in an earlier survey were nowhere to be found in a subsequent survey. For example, of 101 spalls found in a single survey, only 48 were found in the survey conducted 5 years later on the same set of scaffold drops. At trial, this unreliability was conveyed effectively by comparing the reliability of using a coin toss to determine whether a spall would be found 5 years later. The coin toss would have a reliability of 50%, which is higher than the rate that was actually achieved.

The estimates of total damage made by engineers at the trial on behalf of the building owners were problematic because of selection bias. Scaffold drops selected for examination by both building owners and brick manufacturers were known by both to have been selected in part on the basis of a high prevalence of spalls (which were clearly visible from the ground). For example, in the visual survey conducted on behalf of the brick manufacturer, the median spalling rate of a set of 18 drops was found to be 12.1 spalls per 1,000 bricks. (These 18 rates were determined by counting spalls from photographs taken of just the portion of the wall on which the scaffold drop had been hung.) In contrast, the median spalling rate determined from the photographic survey for all 83 wall segments (see Sec. 2 for definitions) was 2.5 spalls per 1,000 bricks. Despite the obvious unrepresentativeness of the locations of the scaffold drops in terms of their spalling rates, a spalling rate of 14.2 spalls per 1,000 bricks was calculated on behalf of the building owners as an average from the 18 drops. This value was then applied, without adjustment, to the numbers of bricks in the entire building complex to get an estimated total number of spalls. The rate of 14.2 should be compared to our much smaller estimate of 6.44 spalls per 1,000 bricks (see Sec. 6.2) for the spalling rate in the entire complex.

Standard errors of estimators derived from the random sampling or error structure of models provide useful information for statisticians and engineers involved in a condition assessment survey, as well as to factfinders (in the present case, the jury) for quantifying uncertainty from these sources. But from the jury's point of view, the greatest source of uncertainty in estimates typically lies (as in the present case) in the variation between estimates provided by the different parties. These differences can be traced to differences in the underlying data, in models, in estimators, in bias sources such as selection or nonresponse, or in the larger analytical framework. (See the useful discussion in Kaye and Freedman 1994, pp. 376–377, that contrasts standard errors and other sources of uncertainty.) In the present case, the difference (7.76 spalls per 1,000 bricks) between the estimates made on behalf of the building owner (14.2 spalls per 1,000 bricks) and the brick manufacturer (6.44 spalls per 1,000 bricks) arose primarily from selection bias.

The question for the jury was which estimator was more credible. In fact, the outcome of the trial was an eleventh-hour settlement in the evening before the morning that the jury was due to deliberate and after all of the testimony about spall (and other) damage had been heard at trial. The settlement substituted a compromise amount of damages to be paid by the brick manufacturer to the building owner, thereby eliminating the risk to both parties of a large real or opportunity loss. According to attorneys for the building owners, the statistical testimony concerning the unreliability of their spall (and other) counts and the unrepresentative nature of the samples from which they made estimates of damage without adjustment was a major factor in their decision to settle rather than let the jury decide.

In this article we estimate the proportion of the almost three-quarters of a million bricks in the complex with water damage severe enough to cause spalling of the brick surface. Specifically, the true spalling rate, θ, is the spalling rate that would be determined by a (prohibitively expensive) 100% scaffold drop survey of all wall surfaces. In Section 2 we describe the three surveys. The photographic and visual surveys measure quantities related to θ. The photographic survey measures a "photo survey true spalling rate," denoted by θ^{photo}, and the visual survey measures a "visual survey true spalling rate," denoted by θ^{vis}. In Section 3 we model the photographic survey data, imputing extensive amounts of missing data via a binomial regression model, which leads to an estimate of θ^{photo}. A photographic survey is known to provide only a crude lower bound to the actual number of spalls, because a photograph cannot pick up spalls that can be determined on a scaffold drop right at the face of the brick. In Section 4 we use the visual survey data to compute an estimate of θ^{vis}, and in Section 5 we use the limited scaffold drop survey data as a gold standard to calibrate both the photographic survey data and the visual survey data, leading to an adjusted estimate of θ. In Section 6 we use multiple imputation to obtain a different estimate of θ and its estimated standard error by taking into account the sampling variability of the imputed missing values, and in Section 7 give conclusions and perspectives. A more detailed version of this article can be obtained by request from the authors.

2. THE CONDITION ASSESSMENT SURVEYS

We consider estimating the extent of masonry deterioration in the residential building complex by combining the results from three independent condition assessment surveys of the

complex carried out at different times by the brick manufacturer. Combining information obtained from several independent surveys should lead to better estimates of the parameters of interest.

The walls of the complex were first divided into 83 wall segments that were defined by natural boundaries of varying size, ranging from roughly 200 to 500 feet wide. Figure 1 shows the five buildings and the locations of the 83 wall segments. Each survey was then carried out on those 83 wall segments.

The Scaffold Spall Survey. The scaffold survey was intended to be an independent examination of the findings obtained from previous scaffold surveys carried out by the owners of the complex. The name comes from the procedure of lowering individuals in a scaffold from the roof parapet of the building. These individuals carefully checked every brick of every row on that part of the building wall bounded by the scaffold for spalling damage. The two ends of the scaffold determined the boundaries of a 20-foot–wide strip down the wall; this strip, which we call a "scaffold drop," was a small part of the entire wall segment. Although the information on examined bricks was very precise, this survey was expensive and time-consuming, so only 11 drops were examined. Figure 1 shows the drop locations. These drop locations were not selected randomly but rather were chosen to overrepresent surfaces with large numbers of visible spalls. Therefore, the average estimated spall rate (per 1,000 bricks) in the 11 drop locations was considerably higher than that for the entire complex using data obtained from the photographic and visual surveys. Variability among examiners on the scaffold was assessed by having different individuals examine the same wall portions and comparing rates. But we did not model measurement error of the spall counts from the scaffold survey. The engineers on the scaffold used a specific definition and operational protocol to detect the spalls. We use "scaffold spalling rate" as a shorthand way to refer to the rate derived from spalls identified by those engineers.

The Photographic Spall Survey. A census of wall spalls was conducted by photographing the entire residential complex. This "photographic spall survey" provided large photographs of all 83 wall segments. The photographic survey was inexpensive, but many bricks were not observable in the photographs because of poor photo angles and obstruction of walls by trees, pipes, vehicles, and other objects.

The Visual Spall Survey. This census was carried out to quantify the extent of spall damage that could be seen from the ground. Spall damage was visually assessed at each wall segment by walking around the periphery of each building. This survey was performed by three individuals in a single day and thus was inexpensive. Each individual surveyed a different portion of the complex. To improve comparability

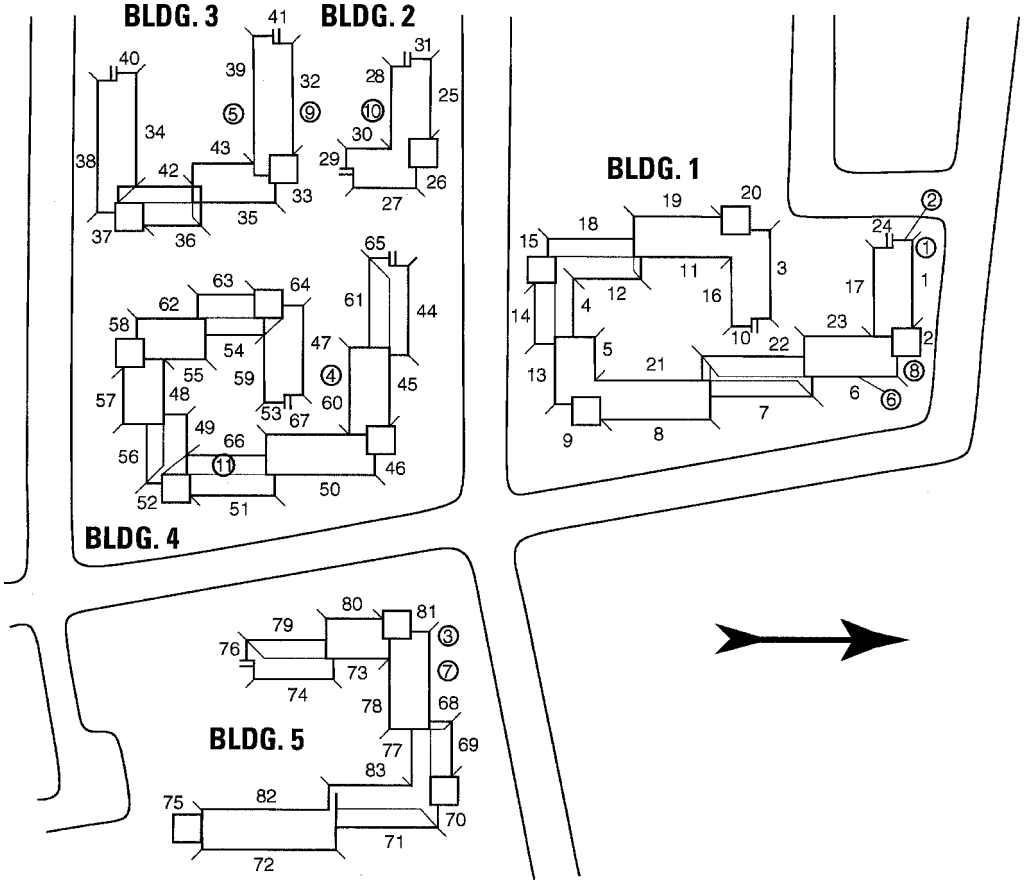

Figure 1. Building Location Map. The map shows the 5 buildings, the 83 wall segments, and the 11 scaffold drop locations (in circles). North is indicated by the direction of the arrow. The north and east orientations receive little sun in winter, and interior exposures such as courtyards provide some shielding from the weather.

between observers, several portions of the walls were surveyed in a preliminary study by all three individuals and the results compared. Measurement error of spall determination from the visual survey was not modeled, however.

It is important to understand that there are really three spall rates, which measure different quantities: a photographic spall rate, a visual spall rate, and a scaffold spall rate.

To compare the relative expenses of these surveys, A.R. Whitlock, Senior Vice-President of KCI Technologies, Inc., explained that for every $1 in total cost of the visual survey, the photographic survey cost $10, and the scaffold survey (if it were carried out throughout the whole complex) cost $100. All three types of surveys are common in assessing masonry damage, although it is quite rare for all three to be done on the same project.

3. THE INCOMPLETE PHOTOGRAPHIC SPALL SURVEY

The photographic survey contained a large amount of missing data. We can estimate θ^{photo} either by ignoring the missing data or by using imputation methods for missing data that provide better estimates of θ^{photo} than would be obtained by ignoring the missing data. In this section we discuss estimates obtained from these two strategies.

The photographic spall survey recorded the number of spalled bricks out of the number of observed bricks and the number of unobserved bricks for each of the $W = 83$ wall segments. Let n_w and n_w^* denote the numbers of photo-observed and (known) photo-unobserved bricks, at the wth wall segment, so that the total number of bricks at the wth wall segment is $N_w = n_w + n_w^*$. Also, let s_w and s_w^* denote the numbers of photo-observed and (unknown) photo-unobserved spalled bricks at the wth wall segment and let $S_w = s_w + s_w^*$. Over the entire residential complex, the total number of photo-observed spalled bricks was

$$S_{\text{obs}} = \sum_{w=1}^{W} s_w = 669$$

out of

$$n = \sum_{w=1}^{W} n_w = 429{,}422$$

photo-observed bricks. The total number of photo-unobserved bricks was

$$n^* = \sum_{w=1}^{W} n_w^* = 263{,}756,$$

of which 148,407 (56.3%) were unobservable in the photographs because of poor photographic angles and 115,349 (43.7%) were unobservable because of obstructions. Thus the total number of bricks included in the photographic survey was $N = n + n^* = 693{,}178$.

3.1 Preliminary Analysis Based on Oversimplified Independence Assumptions

Initial estimates of θ^{photo} were derived by discarding those bricks that were unobserved and analyzing only those that were observed. Justification for these estimates came from assuming that bricks spall at random locations throughout the complex and also a missing completely at random (MCAR) assumption (Little and Rubin 1987). That is, in thinking of spall rates as a function of a number of features of the wall segments, whether or not the spall rates were missing did not depend on either their values or the features of the wall segments.

If we regard S_{obs} as binomial, bin$(n, \theta^{\text{photo}})$, then a "naive" unbiased estimate is $\tilde{\theta}_{\text{naive}} = S_{\text{obs}}/n = 669/429{,}422 = .001558$, or 1.56 spalls per 1,000 bricks, with an estimated binomial standard error of .000061, or .061 spalls per 1,000 bricks. Table 4 records these and subsequent spall rate estimates. If we use wall segment information, and if θ_w^{photo} represents the probability that a given spalled brick on wall segment w shows up on the photographic survey, then $\theta^{\text{photo}} = N^{-1}\Sigma_w N_w \theta_w^{\text{photo}}$. Because $s_w + s_w^*$ is distributed as binomial, bin$(N_w, \theta_w^{\text{photo}})$, θ^{photo} can be estimated unbiasedly by the stratified estimator $\tilde{\theta}_{\text{st}} = N^{-1}\Sigma_w N_w \tilde{\theta}_w$, where $\tilde{\theta}_w = s_w/n_w$. Cochran (1977, chap. 5) compared naive and stratified estimators. From the photographic spall survey data, $\tilde{\theta}_{\text{st}} = 816.2/693{,}178 = .001177$, or 1.18 spalls per 1,000 bricks (see Table 4). Because W is large and we are assuming that the $\{s_w\}$ are independent, $\tilde{\theta}_{\text{st}}$ is approximately normal with mean θ^{photo} and variance

$$\text{var}(\tilde{\theta}_{\text{st}}) = \sum_w \left(\frac{N_w}{N}\right)^2 \left(\frac{N_w - n_w}{N_w - 1}\right) \frac{\theta_w^{\text{photo}}(1 - \theta_w^{\text{photo}})}{n_w}, \qquad (1)$$

where N and N_w are as defined earlier (see Cochran 1977, thm. 5.9). Replacing θ_w^{photo} by $\tilde{\theta}_w$ in (1), an estimated standard error of $\tilde{\theta}_{\text{st}}$ is given by .000029, or .029 spalls per 1,000 bricks.

The stratified estimate is smaller than the naive estimate, implying that the locations where bricks were missing had lower rates of spalling than the mean rate in the entire complex. But spalling may not actually occur at random locations, but rather may be produced by a spatial clustering process along and down the walls. In that case, it is more likely that the $s_w + s_w^*$ would be distributed as correlated or overdispersed binomial. Thus s_w would be distributed as binomial, bin$(n_w, \theta_w^{\text{photo}})$, with overdispersion.

3.2 A Modeling Approach to Imputation of Missing Data in the Photographic Spall Survey

Clearly, spalling is a more complex process than that described earlier. The missing data are more likely to be missing because of their specific locations. Given the values of the four factors that describe wall segments (explained later in this section), spalling rates for missing bricks are assumed to be missing at random (MAR) in the sense of Little and Rubin (1987). That is, whereas the values of missing spalling rates depend on the four factors, given values of those factors, it is not the values of the spalling rates themselves that determine whether or not they are missing.

A weighting adjustment method for analyzing these data (Oh and Scheuren 1983) would use strata constructed from the visual survey spall rates, which are available without missing values for all 83 wall segments. But this method would require that all wall segments with even a few missing bricks be eliminated from the analysis.

We preferred instead a modeling approach to imputing missing data, because, in using all available spalling information,

we expected it to be more precise. We assumed that the photo-observed spall counts, $\{s_w\}$, as a function of a number of qualitative factors, followed a possibly overdispersed binomial distribution, and then used the fitted model to estimate the expected values of the missing data.

The four qualitative factors were:

- *Building number (B).* This factor had five levels corresponding to the five buildings. It was known from observation that buildings 1 and 5 were more heavily spalled than the other buildings.
- *Orientation (O).* This factor had four levels corresponding to whether the wall segment faced north (1), east (2), south (3), or west (4). It was known that the north and east orientations were more heavily spalled.
- *Exposure (E).* This factor had two levels corresponding to whether the wall segment was exterior facing (1) or interior facing (2). Because of weather conditions, it was expected that exterior-facing wall segments would be more heavily spalled.
- *Level of building (L).* The levels were coded in the reverse order than which floors of a building are usually given; that is, starting at the roof parapet (with level = 1) and ending at the ground (with level = 5, 6, 7, or 8 depending on the height of the building). Of the 83 wall segments, 1 had 5 levels, 12 had 6 levels, 47 had 7 levels, and 23 had 8 levels. From observation, it was known that the higher floors (lower levels) of the wall segments were more heavily spalled than the lower floors (higher levels).

Introduction of the factor L, level of the building, required that we expand the number of units of observation from the 83 wall segments to a total of $(1)(5)+(12)(6)+(47)(7)+(23)(8)=590$ "sites." We refer to the wth wall segment, ith building, jth orientation, kth exposure, and lth level as *site* (w,i,j,k,l), $w=1,2,\ldots,W$, $i=1,2,3,4,5$, $j=1,2,3,4$, $k=1,2$, $l=1,2,\ldots,L_w$, where $L_w=5,6,7,$ or 8 as appropriate. For example, if n_{wijkl} and n^*_{wijkl} were the number of unobstructed and obstructed bricks at site (w,i,j,k,l), then the total number of bricks at site (w,i,j,k,l) would be $N_{wijkl}=n_{wijkl}+n^*_{wijkl}$. Of the 590 sites, 38 sites were totally obstructed ($n_{wijkl}=0$) and 21 sites had no obstructions ($n^*_{wijkl}=0$). The extent of the missing-value problem is shown in Table 1, which gives the distributions of $\sum n_{wijkl}$ and $\sum n^*_{wijkl}$ by each of the four factors B, O, E, and L. Figure 2 displays parallel boxplots of the photo-observed spalling rates corresponding to each of the four factors considered separately over the 590 sites.

We performed model fitting on the 552 sites that had at least one photographically observed brick. Let s_{wijkl} be the photographic spall count (out of the n_{wijkl} unobstructed bricks) at site (w,i,j,k,l). Assume that s_{wijkl} follows a possibly overdispersed binomial distribution with mean and variance

$$E(s_{wijkl})=\mu_{wijkl}=n_{wijkl}\theta_{wijkl} \tag{2}$$

and

$$\operatorname{var}(s_{wijkl})=\sigma^2_{wijkl}(\phi)n_{wijkl}\theta_{wijkl}(1-\theta_{wijkl}), \tag{3}$$

where $\theta_{wijkl}(=\theta^{\text{photo}}_{wijkl})$ is the probability that a given spalled brick at site (w,i,j,k,l) shows up on the photographic survey,

$$\sigma^2_{wijkl}(\phi)=1+\phi(n_{wijkl}-1) \tag{4}$$

yields the extrabinomial component of variation (Moore 1987), and $\phi\geq 0$ is an unknown overdispersion parameter (Williams 1982, model II). Note that the value of the overdispersion component (4) depends on the site in question. The canonical link function is given by

$$\eta_{wijkl}=\operatorname{logit}(\theta_{wijkl})=\log\left(\frac{\theta_{wijkl}}{1-\theta_{wijkl}}\right), \tag{5}$$

Table 1. Numbers and Proportions of Unobserved Bricks by Factor From the Photographic Survey for the Entire Residential Complex

Building number (*B*)					
	1	2	3	4	5
obs	136,639	32,612	70,619	116,097	73,455
unobs	74,820	17,305	35,575	60,834	75,222
%unobs	35.4	34.7	33.5	34.4	51
Orientation (*O*)					
	North (1)	East (2)	South (3)	West (4)	
obs	110,654	98,346	102,288	118,134	
unobs	39,820	93,944	49,946	80,046	
%unobs	26.5	48.9	32.8	40	
Exposure (*E*)					
	Exterior (1)			Interior (2)	
obs	251,374			178,048	
unobs	158,138			105,618	
%unobs	38.6			37	
Level (*L*)					

	1	2	3	4	5	6	7	8
obs	39,914	69,316	76,531	77,499	65,021	56,338	36,277	8,526
unobs	15,942	23,269	27,427	32,537	39,033	51,972	52,033	21,543
%unobs	28.6	25.1	26.4	29.6	37.5	48.0	58.9	72

NOTE: "obs" means the number of observed bricks; "unobs", the number of unobserved bricks; and "%unobs", unobs/(obs + unobs) × 100.

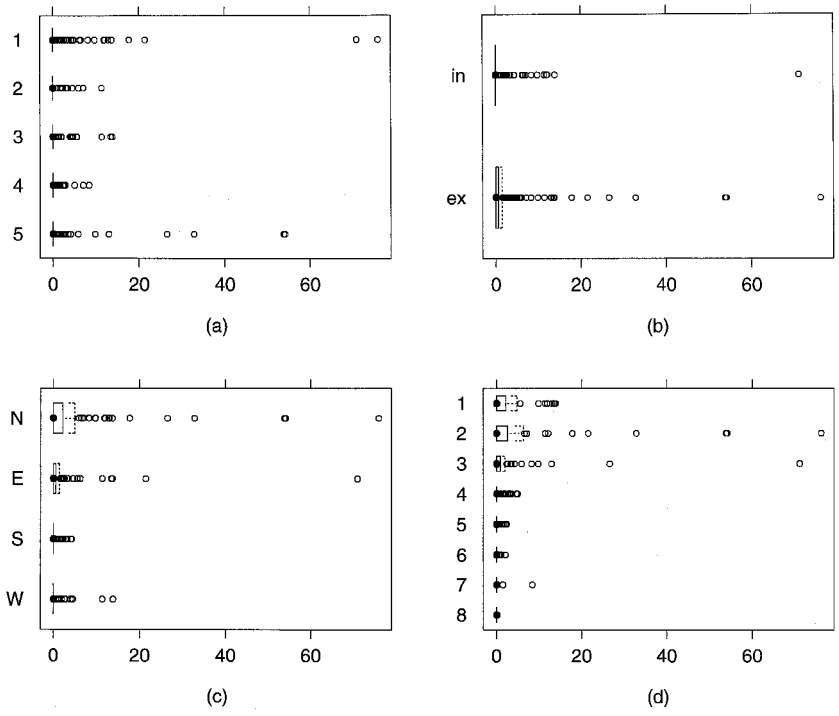

Figure 2. Parallel Boxplots of the Photo-Observed Spalling Rates at the 590 Sites for Each of the Four Factors (a) Building Number B, (b) Orientation O of Wall Segment, (c) Exposure E of Wall Segment, and (d) Level L of Wall Segment. For each plot, the horizontal axis is the spalling rate per 1,000 bricks and the vertical axis represents the levels of each factor. These boxplots confirm the expected locations of high spall counts obtained from visual observations of the buildings prior to carrying out the surveys.

where η_{wijkl} is modeled by an appropriate linear model $\eta = \mathbf{X}\beta$ in the four factors B, O, E, and L. In that model, $\eta = [\eta_{wijkl}]$ is an n-vector, $\mathbf{X}$ is an $(n \times p)$ design matrix, β is the p-vector of unknown regression parameters representing main effects and possible interaction effects, n is the total number of sites to be used as input (initially 552), and p is the total number of parameters that defines the model.

3.3 Model Selection: Sparseness, Outliers, and Overdispersion

To account for overdispersion in generalized linear models (see, e.g., Lambert and Roeder 1995 and references therein), an overdispersed model is fitted only when all other reasons for a poor fit have been eliminated. Thus our starting point in model fitting assumed no overdispersion, so that $\phi = 0$ in (4). Given $\mathbf{s} = \{s_{wijkl}\}$, we first used the S-PLUS (Becker, Chambers, and Wilks 1988; Chambers and Hastie 1992) `glim` function, which uses an iteratively reweighted least squares (IRLS) algorithm to compute maximum likelihood estimates, for model selection assuming no overdispersion.

Sparseness in the observed counts proved to be a major obstacle in modeling the data. Estimated coefficients that corresponded to terms in the regression whose marginal totals were 0 had unusually large estimated standard errors. Every wall segment with eight levels had no observed spalls at level 8, several two-factor interactions also had marginal totals of 0, and one other two-factor interaction had 0 observed spalls.

We tried to fit various partial two-factor interaction models to these data by leaving out certain troublesome interactions without violating the "hierarchy principle" (Bishop, Fienberg, and Holland 1975, pp. 34, 67–68). Because our focus was on predicting spall counts in missing areas of wall segments, and not to explain spalling causally, we favored models that were simple and parsimonious over ones that were overly complicated. None of the models that included interaction terms was found to be appealing in this sense.

This left us with the main effects–only model. To deal with the sparseness problem, we combined levels 7 and 8 into a new expanded level 7^+. The resulting model fit had a deviance of 1,050.4 on 537 degrees of freedom with no unusually large estimated standard errors. All main effects except B_5 and L_3 were statistically significant. A scatterplot (not given here) of the photographic spall counts s_{wijkl} for the 552 sites against their fitted values $\hat{\mu}_{wijkl}$ showed most points congregating near the origin and the remaining points scattered loosely around the 45-degree line. Increasing variability in the plot suggested the presence of overdispersion or outliers in the data.

Residual diagnostics (Pearson and absolute deviance residuals plotted against fitted values) from the main effects–only model fit with levels 7 and 8 renamed 7^+ showed definite indications of three outliers in the data (sites 2, 97, and 296). Fitting the model without those three sites yielded a significant reduction in the deviance to 704.9 based on 534 degrees of freedom. Figure 3 shows an index plot (Pregibon 1981) of the *deletion residuals* (Lloyd, 1999, p. 202; McCullagh and Nelder 1989, sec. 12.5; Williams 1987). Although there were many sites with large reductions in the deviance, the three sites 2, 97, and 296 again clearly stood apart from the rest, with the largest individual reductions in the deviance. Although the spall counts at these three sites could not be explained by the model, it was clear from the results below that these sites had a negligible effect on the final estimate of θ.

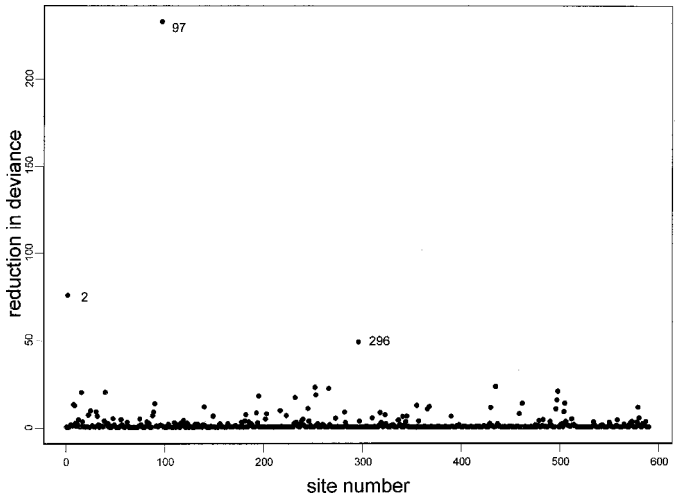

Figure 3. Index Plot of the Reduction in Deviance. The procedure is to delete a site from the data, fit the main-effects-only model with levels 7 and 8 combined, obtain the reduction in deviance from deleting that site, and then plot the reduction in deviance for each of the 552 sites against their respective site number.

We next used the `od.binomial` function of Lloyd (1999, p. 234) to compute estimates of the parameters, their standard errors, and the overdispersion parameter when overdispersion was added as a parameter to the main effects–only model of Section 3.3 with levels 7 and 8 combined. This methodology is considered both as a method-of-moments procedure and as a direct generalization of quasi-likelihood (Moore and Tsiatis 1991). Table 2 presents the model parameter estimates before and after the three outlying sites were removed.

Table 2. Parameter Estimates and Standard Errors for the Factor Components and Overdispersion Parameter (phi) of the Linear Additive Model From the Photographic Spall Data

		All data ($n = 552$)		*Three outliers deleted (sites 2,97,296)*	
Parameter		*Estimate*	*Standard error*	*Estimate*	*Standard error*
Intercept		−4.545	.400	−4.718	0.344
B	1	0		0	
	2	−1.211	.627	−.930	.499
	3	−1.167	.499	−1.011	.431
	4	−2.204	.635	−1.802	.501
	5	.047	.345	.412	.302
O	North	0		0	
	East	−.782	.341	−.895	.313
	South	−2.840	.753	−2.627	.573
	West	−2.508	.594	−2.695	.538
E	Exterior	0		0	
	Interior	−.632	.342	−1.018	.330
L	1	0		0	
	2	.977	.386	.893	.316
	3	.137	.441	−.268	.388
	4	−1.231	.668	−1.129	.508
	5	−2.815	1.336	−2.709	1.001
	6	−3.630	2.001	−3.507	1.486
	7+	−2.802	1.262	−2.711	.989
phi		.0134		.0069	

3.4 Single Imputation for Missing Values of Photographic Spalls

Next, we imputed the obstructed photographic spall counts. The essential idea of single imputation was to provide a single reasonable proxy for each missing value, then substitute those proxies into the incomplete dataset and analyze the resulting completed dataset using standard statistical procedures. (For a general discussion of single imputation, see, e.g., Little and Rubin 1987, sec. 3.4; Rubin 1987, sec. 1.4.) The estimated spalling rate from both the photo-observed spalls and the imputed values is

$$\hat{\theta}^{\text{photo}} = \frac{\widehat{S}^{\text{photo}}}{N} = \frac{S_{\text{obs}} + S_{\text{imp}}}{n + n^*} = \frac{937.39}{693{,}178} = .00135, \qquad (6)$$

or 1.35 spalls per 1,000 bricks.

4. THE VISUAL SPALL SURVEY

There were 83 sites for the visual survey, because no information about factor L was recorded. The number of bricks recorded by the visual survey at the wth site was n_w^{vis}, and the total number of bricks encountered in the visual survey over the entire residential complex was $n^{\text{vis}} = \sum_w n_w^{\text{vis}} = 694{,}756$. Assuming that bricks spall at random throughout the complex, a naive but unbiased estimate of the visual survey spall rate, θ^{vis}, is given by

$$\hat{\theta}^{\text{vis}} = \frac{s^{\text{vis}}}{n^{\text{vis}}} = \frac{1{,}532}{694{,}756} = .00221, \qquad (7)$$

or 2.21 spalls per 1,000 bricks, with an estimated binomial standard error of .00006, or .06 spalls per 1,000 bricks; see Table 4. There is no need to compute a separate stratified estimator as we did in Section 3.1, because without missing bricks in the visual survey, the stratified estimator is identical to the naive estimator. Although the visual survey estimate of 2.21 visual spalls per 1,000 bricks is higher than the (stratified) photographic survey estimate of 1.18 photographic spalls per 1,000 bricks, the difference reflects the greater ease with which spalls are visible to the eye of an observer on the ground than from a photograph. Thus two different spalling rates are being estimated.

5. CALIBRATING THE ASSESSMENT SURVEYS TO THE SCAFFOLD SPALLING RATE

We next calibrated the data from the photographic and visual surveys against the data from the more precise, but quite limited scaffold survey. The motivation behind this calibration was to provide an estimate of θ that might have been obtained had the scaffold survey examined each and every brick in the entire complex. We obtained a calibration sample in which we matched relevant portions of the photographic and visual surveys to 11 scaffold drop locations. We transformed the scaffold spall counts and the photographic and visual spall counts into spalling rates by dividing the appropriate spall counts by the number of bricks examined at that drop location.

5.1 Calibration Adjustments of Photographic and Visual Spall Rates

For this calibration application, suppose that $\mathcal{U}, \mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_s$ are random variables that all measure the same quantity, where $\mathcal{U}$ is determined with great precision and is costly and the $\mathcal{V}_j, j = 1, 2, \ldots, s$, are less precise but inexpensive. Calibration methodology comprises two steps. First, a "calibration phase" starts with a training sample of q measurements, $(u_d, \mathbf{v}_d), d = 1, 2, \ldots, q$, where $\mathbf{v}_d = (v_{d1}, \ldots, v_{ds})$, recorded on variables $(\mathcal{U}, \mathcal{V}_1, \ldots, \mathcal{V}_s)$, and the relationship between $\mathcal{U}$ and the $\{\mathcal{V}_j\}$ is estimated from those data. Then the estimated relationship is used in a "prediction phase" to determine the values of precise, but costly $\mathcal{U}$ measurements from a fresh sample of $\mathcal{V}_j, j = 1, 2, \ldots, s$, measurements.

The preferred method for the "uncontrolled" or "random" calibration problem (which is the situation at issue here) is the *inverse calibration* method (see, e.g., Brown 1993; Martens and Naes 1989; Osborne 1991 and the references therein). The inverse approach obtains an "inverse calibration curve," $\widehat{\mathcal{U}} = g(\mathcal{V}_1, \ldots, \mathcal{V}_s)$, usually by regressing the $\mathcal{U}$ measurements on the $\{\mathcal{V}_j\}$ measurements.

5.2 Results From the Calibration Sample

In the calibration sample from the residential complex, spalling rates were recorded at $q = 11$ scaffold drops; see Table 3 for data and Figure 1 for drop locations. At each drop location, $\mathcal{U}$, $\mathcal{V}_1$, and $\mathcal{V}_2$ were the spalling rates per 1,000 bricks from the scaffold, photographic, and visual surveys. At the dth drop location, a measurement (u_d, v_{1d}, v_{2d}) on $(\mathcal{U}, \mathcal{V}_1, \mathcal{V}_2)$ was made, $d = 1, 2, \ldots, 11$. Figure 4 displays a scatterplot matrix of variables. There is reason to believe that the relationship between $\mathcal{U}$ and $\mathcal{V}$ is independent of the building (B), orientation (O), and exposure (E) factors. The imprecision of the measurements $\mathcal{V}_1$ and $\mathcal{V}_2$ is because of the distance between the observer and the walls. Given $\mathcal{V}_1$ and $\mathcal{V}_2$, it was an engineering judgment that none of the B, O, and E factors would affect $\mathcal{U}$. Although L might be a factor that would influence the conditional expectation of $\mathcal{U}$ given $\mathcal{V}_1$, the calibration was determined using scaffold drops that covered all levels.

Table 3. Calibration Data From the Photographic, Visual, and Scaffold Surveys

Scaffold drop	Wall segment	B	O	E	Scaffold U	Photo V1	Visual V2
1	1	1	1	1	40.07	15.03	24.21
2	2	1	1	1	40.32	13.98	15.05
3	6	1	2	1	11.48	1.81	2.42
4	24	1	4	1	6.71	0.00	0.00
5	28	2	3	1	1.93	.97	3.87
6	32	3	1	1	3.94	.99	2.96
7	39	3	3	2	5.12	0.00	3.41
8	60	4	3	2	.87	.87	0.00
9	66	4	4	2	0.00	0.00	0.00
10	68	5	1	1	43.42	12.02	14.03
11	68	5	1	1	22.16	7.71	19.27

NOTE: Observations are spalling rates per 1,000 bricks.

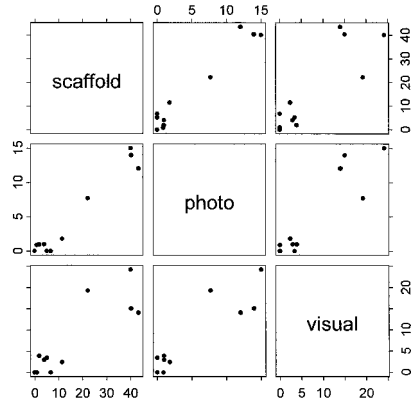

Figure 4. Scatterplot Matrix of Three Pairwise Variables for Calibration. The observations are spalling rates per 1,000 bricks from the scaffold, photographic, and visual spall surveys.

The calibration results were as follows. The calibration curve with $s = 2$ is given by

$$\widehat{\mathcal{U}} = \mathcal{V}\hat{\mathbf{b}} = \hat{b}_0 + \hat{b}_1\mathcal{V}_1 + \hat{b}_2\mathcal{V}_2 = 3.031 + 3.233\mathcal{V}_1 - .350\mathcal{V}_2, \quad (8)$$

the estimated covariance matrix of $\hat{\mathbf{b}}$ is given by

$$\widehat{\mathrm{cov}}(\hat{\mathbf{b}}) = \begin{bmatrix} 2.940 & .036 & -.199 \\ .036 & .297 & -.190 \\ -.199 & -.190 & .145 \end{bmatrix},$$

the R^2 value is .954, and the residual variance is $\hat{\sigma}^2 = (4.165)^2 = 17.35$. If we calibrate $\mathcal{U}$ only on $\mathcal{V}_1$, then the estimated intercept is 2.5501 (with an estimated standard error of 1.6196), and the estimated regression coefficient of $\mathcal{V}_1$ is 2.7726 (.2145). The R^2 value is .949, and the residual variance is $(4.13)^2 = 17.06$. If we calibrate $\mathcal{U}$ only on $\mathcal{V}_2$, then the estimated intercept is 2.644 (3.7563), the estimated regression coefficient of $\mathcal{V}_2$ is 1.724 (.3302), the corresponding R^2 value is .752, and the residual variance is $(9.12)^2 = 83.17$.

We chose the final calibration model based on the following considerations. First, the squared correlation coefficient between $\mathcal{V}_1$ and $\mathcal{V}_2$ was .918; second, using a sequential analysis of variance table (Weisberg 1985, p. 51), the partial F statistic for testing the coefficient of $\mathcal{V}_2$ after fitting the highly significant $\mathcal{V}_1$ was not significant ($F = .85$ with 1 and 8 degrees of freedom), whereas both predictors were highly significant when entered in the reverse order; and third (and most important), the residual variance actually *increased* when going from the model with $\mathcal{V}_1$ as the only predictor to the model with both $\mathcal{V}_1$ and $\mathcal{V}_2$ as predictors. We thus chose to proceed by calibrating $\mathcal{U}$ for the entire complex only on $\mathcal{V}_1$.

5.3 A Calibrated Estimate of the Scaffold Spalling Rate

The scaffold spalling rate per 1,000 bricks at site (w, i, j, k, l) is estimated by calibration as

$$u_{wijkl} = \hat{b}_0 + \hat{b}_1\left(\frac{1{,}000(s_{wijkl} + \hat{s}^*_{wijkl})}{n_{wijkl} + n^*_{wijkl}}\right), \quad (9)$$

whence the estimated number of scaffold spalls at that site is

$$S_{wijkl} = (n_{wijkl} + n^*_{wijkl}) \times \frac{u_{wijkl}}{1{,}000}. \tag{10}$$

Thus, the estimated total number of scaffold spalls over the entire complex is given by

$$S_{\text{cal}} = \sum_{w,i,j,k,l} S_{wijkl} = \hat{b}_0 \frac{N}{1000} + \hat{b}_1 \widehat{S}^{\text{photo}}. \tag{11}$$

From (6), it follows that a calibrated estimate of θ is

$$\hat{\theta}_{\text{cal}} = \frac{S_{\text{cal}}}{N} = \frac{\hat{b}_0}{1{,}000} + \hat{b}_1 \hat{\theta}^{\text{photo}}. \tag{12}$$

From Section 5.2 and the final model fit of Section 3, we have that $S_{\text{cal}} = 4{,}366.67$ spalls. Dividing by $N = 693{,}178$ yields $\hat{\theta}_{\text{cal}} = .00630$, or 6.30 spalls per 1,000 bricks; see Table 4.

The scaffold spall rate was regarded as the quantity of most interest in the litigation process and, hence this estimate of 6.30 spalls per 1,000 bricks was a possible culmination of the surveys. The rate of spalls that are actually visible to the eye also was of some interest, reflecting an aesthetic problem.

The standard error of $\hat{\theta}_{\text{cal}}$, which may be obtained as the usual standard error of a prediction from a linear model estimator, was estimated to be $\widehat{\text{s.e.}}(\hat{\theta}_{\text{cal}}) = .00146$, or 1.46 spalls per 1,000 bricks. We ignored the fact that the independent regressor had been estimated by a binomial regression in Section 3; in Section 6.2 we give for reasons why this cannot be very important. Thus the estimated standard error of $\hat{\theta}_{\text{cal}}$ is somewhat smaller than it would have been if we had taken into account missing value estimation.

5.4 Relationship of the Calibration Estimator to the Regression Estimator

Under the MAR condition for monotone missing data, which we believe holds here, the maximum likelihood estimate, $\widehat{\overline{Y}}$, of the mean scaffold spall rate over the entire complex is simply the classical regression estimator, $\widehat{\overline{Y}} = \bar{y} + b(\overline{X} - \bar{x})$, where $\overline{X}$ is the estimate of the mean photographic spalling rate for the entire complex, $\bar{x}$ and $\bar{y}$ are the sample mean photographic spalling rates and sample mean scaffold spalling rates at the 11 drop locations, and b is the slope of the regression of Y on X (see Little and Rubin 1987, eq. 6.9). Approaching calibration directly from sampling theory, Cochran (1977, pp. 189–190) noted that the regression estimator was the appropriate tool to use when trying to adjust the sample mean of more costly measurements (see also Särndal, Swensson, and Wretman 1992, chap. 6). That the calibrated estimator (12) is a type of regression estimator used in survey sampling contexts can be seen by noting that (12) is a spalling rate per brick, so that multiplying (12) by 1,000, setting $\overline{X} = 1{,}000\ \hat{\theta}^{\text{photo}}$, and using the fact that $\hat{b}_0 = \bar{y} - \hat{b}\bar{x}$, where $b = \hat{b}_1$, we have that (12) reduces to the classical regression estimator. From a survey sampling perspective, the regression estimator is known to be an efficient tool for using auxilliary information related to the variable in question. For example, Fairley, Izenman, and Bagchi (1990, sec. 2.2) used regression in a two-phase (or double) sampling scheme as a calibration between federal government estimates of error rates in the administration of welfare caseloads (the presumed gold standard measure) and state government estimates of the same error rates (the presumed inexpensive measurement).

Although the calibration equation is a type of regression estimate, the sample of 11 drops was not chosen randomly from the population of all drops in the complex. As a result, the usual finite-sampling basis for inference (e.g., Särndal et al. 1992, chap. 6) is not appropriate for this situation. But use of the regression estimator can be viewed within the usual model-based framework for regression analysis. That is, the sample calibration of the scaffold spalling rate against the photographic spalling rate should be judged by its goodness of fit to the specification of the linear model in question and by the appropriateness of the assumptions of independence and common variance of the error terms in the equation. Because these assumptions appear satisfactory for the sample data, the regression equation may be applied to the prediction of scaffold spalling rates from new values of the photographic spalling rates; that is, to the prediction of scaffold spalling rates for the 83 wall segments. Moreover, for prediction purposes, it is important that the regression equation be determined for values of the independent variable that span the range of values from which predictions are made. This condition is met in the present case, as the photographic spalling rates at the 11 scaffold drops span a range from near 0 to values higher than those from any of the 83 wall segments.

Table 4. Estimated Spalling Rates per 1,000 Bricks

	Spalling rate parameter		
Survey and method	θ^{photo}	θ^{vis}	θ
Photographic survey			
Naive	1.56 (.06)		
Stratified	1.18 (.03)		
Single imputation	1.35 (−)		
Visual survey			
Naive-stratified		2.21 (.06)	
Photographic and scaffold surveys			
Single imputation and calibration			6.30 (1.46)
Multiple imputation and calibration			6.44 (1.46)

6. MULTIPLE IMPUTATION FOR MISSING VALUES OF PHOTOGRAPHIC SPALLS

The single imputations reported in Section 3.4 did not provide any indication of the sampling variability incurred by imputing the missing values. In this section we develop a different estimate of θ with its associated standard error using Rubin's method of multiple imputation (Rubin 1987, 1996). In this application we are not interested in estimating the parameters (and their standard errors) of the regression model in the face of missing data. Rather, because our primary interest is in estimating θ and its standard error, we constructed and fitted the model solely for the purpose of imputation.

6.1 Multiple Datasets of Photographic Spalls Through Imputation

We obtained M imputed datasets of photographic spall counts as follows. In Section 3.2 we developed a binomial regression model for the proportion of spalls at a given site, which depended on some properties of the site including building number, orientation, exposure, and level. Using a Bayesian setting, based on a vague and uninformative multivariate prior distribution on the p-vector β in Section 3.2, the posterior distribution of the parameter vector β was multivariate normal with mean vector equal to the parameter estimates $\hat{\beta}$ and a covariance matrix $\mathbf{C}$ given by the covariance matrix of $\hat{\beta}$ (Box and Tiao 1973, p. 48).

We randomly generated a p-dimensional vector of logit-model regression coefficients (a β-vector) from this posterior distribution. To do this, we first used a Cholesky decomposition to compute $\mathbf{C}^{1/2}$, the matrix square root of the covariance matrix $\mathbf{C}$, and then drew p independent $\mathcal{N}(0, 1)$ random variates to form a vector $\mathbf{z}$. The random drawing of a β-vector from the posterior distribution was given by $\hat{\beta} + [\mathbf{C}^{1/2}]^{\tau}\mathbf{z}$ (see, e.g., Selvin 1998, pp. 190–192). We used the randomly drawn β-vector to create a vector of predicted spalling rates $\tilde{\theta}_{wijkl}$ at each of the 590 sites in the usual way from the fitted model. We then used the $\{\hat{\theta}_{wijkl}\}$ together with the (known) numbers $\{n^*_{wijkl}\}$ of photo-unobserved bricks to simulate a vector $[\tilde{s}^*_{wijkl}]$, say, from the binomial distribution that represented spall counts for the photo-unobserved bricks. Note that the distribution of these simulated spall counts followed an overdispersed binomial distribution, because our posterior distribution of coefficients included an overdispersion parameter. Adding these (simulated) imputed spalls from the photo-unobserved bricks at each site to the photo-observed spalls at each site, we obtained a vector of estimated total photographic spalls, $\tilde{\mathbf{t}} = [s_{wijkl} + \tilde{s}^*_{wijkl}] = [\widetilde{S}_{wijkl}]$, say.

Repeating this procedure M times yielded M imputed datasets, $\{\tilde{\mathbf{t}}_m, \quad m = 1, 2, \ldots, M\}$, of total (observed + imputed) photo spall counts at each of the 590 sites. According to Rubin, $M = 5$ would be large enough to allow us to continue with this multiple imputation procedure. From each of the M imputed datasets, we formed $\widetilde{S} = \sum_{w,i,j,k,l} \widetilde{S}_{wijkl}$, a quantity similar to $\widehat{S}^{\text{photo}}$ in Section 3.4. Call them $\{\widetilde{S}_m, m = 1, 2, \ldots, M\}$. Next, we computed M versions of $\hat{\theta}^{\text{photo}}$ in (6), say $\{\tilde{\theta}_m, m = 1, 2, \ldots, M\}$, where $\tilde{\theta}_m = \widetilde{S}_m/N$. The multiple imputation estimate of the spall rate θ was then the average of the values $\tilde{\theta}_m$ over all M imputations,

$$\hat{\theta}_{\text{mi}} = \frac{1}{M}\sum_{m=1}^{M} \tilde{\theta}_m = \frac{\widehat{S}_{\text{mi}}}{N}, \tag{13}$$

where

$$\widehat{S}_{\text{mi}} = \frac{1}{M}\sum_{m=1}^{M} \widetilde{S}_m \tag{14}$$

was the multiple imputation estimate of the total spall count. The standard error of $\hat{\theta}_{\text{mi}}$ was then estimated in the usual way (Little and Rubin 1987, sec. 12.4; Schafer 1997, sec. 4.3) by

$$\widehat{\text{s.e.}}(\hat{\theta}_{\text{mi}}) = \left(V_W + \left(1 + \frac{1}{M}\right)V_B\right)^{1/2}, \tag{15}$$

where V_W is the within-imputation variance and V_B is the between-imputation variance.

6.2 Parameter Estimate and Standard Error of Multiply Imputed Calibrated Scaffold Spalling Rate

We obtained a multiply imputed version of $\hat{\theta}$ by feeding each imputed dataset through the calibration adjustment derived in Section 5. The subscript "cal" indicates that the results refer to the calibrated estimates. From (14), $\widehat{S}_{\text{cal,mi}} = 4461.522$, where, from (13), $\hat{\theta}_{\text{cal, mi}} = .00644$, or a scaffold spalling rate of 6.44 per 1,000 bricks. Furthermore, $V_W = 2.1234 \times 10^{-6}$, whereas $V_B = 1.4188 \times 10^{-8}$. These results indicated that the variability of the estimate had very little to do with the missing data, and was due almost entirely to the uncertainty in the calibration. This was not unexpected given that the calibration estimates were formed based on only 11 points. Thus from (15), the estimated standard error of $\hat{\theta}_{\text{cal, mi}}$ was .00146, or 1.46 per 1,000 bricks. These results are summarized in Table 4.

The estimates of each of the true values of the photographic spalling rate, visual spalling rate, and scaffold spalling rate had small standard errors compared to the differences between them; see Table 4. The differences between estimated spalling rates were principally due to differences in the quantities estimated. Thus Table 4 should not be read as pointing up big differences in estimated spalling rates.

Multiple imputation gave an estimated scaffold spalling rate that was quite close to the estimate from single imputation given in Section 3.4. Estimates of the scaffold spalling rate made from different calibrations (i.e., calibrated photographic survey, calibrated visual survey, and jointly calibrated photographic and visual survey) were also similar. In the end, after considering the photographic, visual, and scaffold surveys, we found that all three can be reconciled through calibration. Thus good agreement was obtained on an estimate of the quantity that was of principal interest to the parties to the dispute—namely, the scaffold spalling rate.

7. CONCLUSIONS AND DISCUSSION

As far as we are aware, no method presented in this article has previously been applied to estimate total damage in building structures. Some statistical work addressed to (actual or potential) damage assessment questions has appeared. For example, reliability analysis has been used to study the deterioration over time of reinforced concrete structures (Crowder 1991), and time series analysis has been used to study vibrations of high-rise buildings caused by nuclear blasts, chemical explosions, or earthquakes (Kircher 1977). But very little has appeared on statistical techniques for assessing the total amount of damage, deterioration, or defects in existing building structures at a single point in time. We emphasize that the analysis presented here should not be considered as a typical example of actual practice, but rather as a possible model of how statistical analysis should be practiced. The data from the three surveys discussed in this article will be made publicly available via STATLIB.

This case study demonstrates how statistical methods can make important contributions to assessment surveys, and thus

how statisticians and engineers have an opportunity for useful collaboration in designing and analyzing such surveys. First, there currently are no rigorous guidelines for sampling in assessment surveys to investigate masonry deterioration. Although the American Society of Civil Engineers and the American Concrete Institute have published guidelines on assessment surveys, there is no comparable guidance about sampling. This is not an unusual situation according to Dr. H. W. H. West of British Ceramic Research Ltd. and Honorary Secretary of the British Masonry Society, who explained the typical course of such assessment surveys in the United Kingdom:

> The very detailed surveys such as were done [in this case] with full photographic recording are rare in U.K. It is usual for the various experts to inspect the site and report on the whole from a detailed examination of (probably the worst!) parts in order to assess the cause of damage, apportion responsibility and seek to minimize their client's own liability (West 1991).

In large structures, designed sampling can save costs and increase precision. For example, in the present case, walls might have been stratified by the visible spalling rate, each wall divided into squares, and then certain of those squares selected at random to be surveyed in detail. If so, the total cost would have been lower and/or the precision of the estimated total damage greater.

Second, without sampling guidance, assessment surveys are almost always "convenience" or "judgment" surveys of masonry units (such as bricks) rather than true random samples. These assessment surveys generally do not rise above the level of casual observation, and may embody very large biases. In the present case, the scaffold drops were carried out at unrepresentative locations. Estimates of total damage could not be made without substantial selection bias. Hahn and Meeker (1993, sec. 9) noted that "studies have shown that what might be called "judgment" can lead, even unintentionally, to seriously biased samples and, therefore, invalid or misleading results." In Section 1 we reported that the selection bias in using only the original scaffold surveys to estimate total damage was more than 100% of the new estimate based on calibration; see Section 5.

Third, the use of calibration allowed us to make statistically valid estimates of the quantities of most economic interest—the actual spalling rate θ and actual total number of spalls S—at a tiny fraction of the (prohibitive) cost of a complete census of the complex using only scaffold drops. With careful statistical design at the outset of an assessment survey and the use of calibration, the overall cost could be reduced to below that of the surveys reported here, or the precision of the estimate of total damage could be increased.

Fourth, note that in the context of combining information, the scaffold survey had a serious weakness in selection bias, whereas the photographic and visual surveys had a serious weakness in measurement bias. Through calibration, the strengths of the surveys were combined in such a way that neutralized their respective weaknesses. On the one hand, the photographic and visual surveys were censuses, and thus selection bias was not an issue, because no selection was made; on the other hand, the scaffold survey was the definitive measurement standard for the presence of spalling.

Fifth, a binomial regression model for the number of spalls in the 590 different wall segment–level units of the complex circumvented the important fact that spalls were unobservable in a good fraction of the bricks in the photographic survey. We used the model to estimate the numbers of spalls in the missing areas, which let us use the photographic survey to estimate total spalls in the entire complex.

Sixth, deriving an explicit model for the data made it possible for us in the present analysis to estimate standard errors for the calibration estimate of total damage; see Table 4. These standard errors inform survey users that the uncertainty in the estimate is tolerable even though the calibration used only 11 scaffold drops. In this case the sampling uncertainty—including that introduced by missing values—turned out to be tolerably small, whereas the selection bias was large and not tolerable.

The problem studied here is not unusual. To give one recent example, the brick facade of the 22-story tower of the 25-year-old James A. Byrne Federal Courthouse in Philadelphia's Independence Mall had been falling onto the surrounding sidewalks over a period of 4 years (Slobodzian 1998). A government investigation revealed that the problem was caused by a combination of bad design and poor construction. Replacement of the entire brick facade cost the government $25 million, because the legal time limit for filing suit against the builder had passed.

In conclusion, statistical methods can make an important and useful contribution to the design and analysis of condition assessment surveys of buildings and other structures. Costs can be minimized, and reasonable estimates of the extent of damage can be made. We strongly recommend the use of statistical design and analysis in any large assessment survey whenever an estimate of total damage is required.

[Received May 2000. Revised October 2000.]

REFERENCES

American Concrete Institute (1984), *Guide for Making a Condition Survey of Concrete in Service, Revised*. Detroit: Author.

American Society of Civil Engineers (1991), *Guidelines for Structural Condition Assessment of Existing Buildings*, New York: Author.

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The New S Language: A Programming Environment for Data Analysis*, Pacific Grove, CA: Wadsworth & Brooks/Cole.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis*, Cambridge, MA: MIT Press.

Box, G. E. P., and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, New York: Wiley.

Brown, P. J. (1993), *Measurement, Regression, and Calibration*, Oxford, U.K.: Oxford University Press.

Chambers, J. M., and Hastie, T. (eds.) (1992), *Statistical Models in S*, Pacific Grove, CA: Wadsworth & Brooks/Cole.

Cochran, W. G. (1977), *Sampling Techniques* (3rd ed.), New York: Wiley.

Crowder, M. (1991), "A Statistical Approach to a Deterioration Process in Reinforced Concrete," *Applied Statistics*, 40, 95–103.

Fairley, W. B., Izenman, A. J., and Bagchi, P. (1990), "Inference for Welfare Quality Control Programs," *Journal of the American Statistical Association*, 85, 874–890.

Hahn, G. J., and Meeker, W. Q. (1993), "Assumptions for Statistical Inference," *The American Statistician*, 47, 1–11.

Kaye, D. H., and Freedman, D. A. (1994), "Reference Guide on Statistics," in *Reference Manual on Scientific Evidence*, Washington, D.C.: U.S. Government Printing Office, pp. 331–414.

Kircher, C. A. (1977), "Ambient and Forced Vibration Analysis of Full-Scale Structures," doctoral dissertation, Stanford University, Dept. of Civil Engineering.

Lambert, D., and Roeder, K. (1995), "Overdispersion Diagnostics for Generalized Linear Models," *Journal of the American Statistical Association*, 90, 1225–1236.

Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis With Missing Data*, New York: Wiley.
Lloyd, C. J. (1999), *Statistical Analysis of Categorical Data*, New York: Wiley.
Martens, H., and Naes, T. (1989), *Multivariate Calibration*, Chichester, U.K.: Wiley.
McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), New York: Chapman and Hall.
Moore, D. F. (1987), "Modelling the Extraneous Variance in the Presence of Extra-Binomial Variation," *Applied Statistics*, 36, 8–14.
Moore, D. F., and Tsiatis, A. (1991), "Robust Estimation of the Variance in Moment Methods for Extra-Binomial and Extra-Poisson Variation," *Biometrics*, 47, 383–401.
Oh, H. L., and Scheuren, F. J. (1983), "Weighting Adjustment for Unit Nonresponse," in *Incomplete Data in Sample Surveys*, Vol. 2, eds. W. Madow, I. Olkin, and D. B. Rubin, New York: Academic Press, pp. 143–184.
Osborne, C. (1991), "Statistical Calibration: A Review," *International Statistical Review*, 59, 309–336.
Pink Floyd (1979), *The Wall*, Columbia Records, CBS, Inc.
Pregibon, D. (1981), "Logistic Regression Diagnostics," *The Annals of Statistics*, 9, 705–724.
Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
———(1996), "Multiple Imputation After 18+ Years" (with discussion), *Journal of the American Statistical Association*, 91, 473–520.
Särndal, C.-E., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag.
Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman and Hall.
Selvin, S. (1998), *Modern Applied Biostatistical Methods Using S-Plus*, New York: Oxford University Press.
Slobodzian, J. A. (1998), "U.S. Courthouse's Brick Facade to be Replaced," *Philadelphia Inquirer*, February 17, 1998, p. B1.
Weisberg, S. (1985), *Applied Linear Regression* (2nd ed.), New York: Wiley.
West, H. W. H. (1991), Letter to William B. Fairley, September 24, 1991.
Williams, D. A. (1982), "Extra-Binomial Variation in Logistic Linear Models," *Applied Statistics*, 31, 144–148.
———(1987), "Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions," *Applied Statistics*, 36, 181–191.