Comparison of methods to analyse imprecise faecal coliform count data from environmental samples

H. CARABIN[1]*, T. W. GYORKOS[1], L. JOSEPH[1], P. PAYMENT[2] AND J. C. SOTO[3]

[1] *Department of Epidemiology and Biostatistics, McGill University and Division of Clinical Epidemiology, Montréal General Hospital, Montréal, Québec, Canada*
[2] *Laboratoire de Virologie, Institut Armand-Frappier, Laval, Québec, Canada*
[3] *Unité des maladies infectieuses, Direction de la Santé Publique, Régie Régionale de la santé et des services sociaux de Laval, Laval, Québec, Canada*

SUMMARY

Imprecise values arise when bacterial colonies are too numerous to be counted or when no colonies grow at a specific dilution. Our objective was to show the usefulness of multiple imputation in analysing data containing imprecise values. We also indicate that interval censored regression, which is faster computationally in situations where it applies, can be used, providing similar estimates to imputation. We used bacteriological data from a large epidemiological study in daycare centres to illustrate this method and compared it to a standard method which uses single exact values for the imprecise data. The data consisted of numbers of FC on children's and educators' hands, from sandboxes and from playareas. In general, we found that multiple imputation and interval censored regression provided more conservative intervals than the standard method. The discrepancy in the results highlights both the importance of using a method that best captures the uncertainty in the data and how different conclusions might be drawn. This can be crucial for both researchers and those who are involved in formulating and regulating standards for bacteriological contamination.

INTRODUCTION

Bacteriological analysis of environmental samples often provides imprecise estimates of the number of colonies, especially when the number is small. The problem of imprecise values is commonly found in environmental microbiology [1–18]. For example, in water, colonies of bacteria tend to be distributed in clusters [19], so that laboratory results may include values less than the value of the lowest dilution used or values more than the value of the highest dilution used. Therefore, it is possible that no colony will grow

* Author for correspondence: Department of Infectious Disease Epidemiology, Imperial College School of Medicine, Norfolk Place, London W2 1PG.

on the Petri dish, for a given screening dilution. Similarly, it is possible that the Petri dish will be completely covered by bacterial growth. The laboratory will then report that the number of colonies was less than what would have been reported had one colony grown, or that the number of colonies was more than what could maximally have been reported at that screening dilution, respectively. These situations lead to the reporting of imprecise values by the laboratory because the bacteria in the samples cannot be precisely counted at the screening dilution used. If additional dilutions are performed, the number of samples which cannot be counted will decrease, but it is more than likely that some imprecise values will still be reported. Given the large variation

in faecal coliform (FC) concentrations in time and space [20] and, given the increasing consideration of laboratory costs, the choice of a screening dilution is usually set at one (or two) optimal levels. The screening dilution is that dilution of liquid volume used that can best measure the bacterial concentration present in the study samples. The screening dilution differs by type and quantity of sample examined (e.g. water, saliva; litre, millilitre) and type of outcome to be measured (e.g. virus, bacteria). The possibility of precisely determining the concentration by adding dilutions may not be possible. Thus, bacteriological analyses performed either routinely or for research purposes will include imprecise values and these will need to be taken into account in any data analysis.

Problems arise when data containing imprecise values are used in univariate and multivariate analyses; for example, when summarizing contamination levels or identifying risk factors associated with higher levels of contamination. Different approaches have been proposed to analyse data with imprecise results. Results have sometimes simply been dichotomized into positive and negative categories – resulting not only in the loss of potentially valuable information but also in introducing the possibility of important misclassification bias [21–24]. Other researchers have replaced imprecise counts with exact values representing half the value of the lowest screening dilution [25, 26]. Another technique, well known to statisticians but not yet applied to bacterial count data, is multiple imputation [27]. Here one takes the average results from several analyses, each analysis having used a data set constructed by randomly drawing (imputing) one value from the feasible range of each imprecisely known FC count. The imputation variances from both within and between iterations are included in making the final inferences, so that the uncertainty inherent in the imprecise data is fully accounted for. How imprecise results are represented in analyses may affect both the type of data analyses performed and the conclusions drawn.

Our objective is to show how techniques which account for the uncertainty inherent in imprecise values reported by the laboratory, including multiple imputation and interval censored regression, can be used to provide better point and interval estimates. We compare results from analyses of bacterial count data containing imprecise values using multiple imputation and interval censored regression to the standard method where the imprecise values are replaced with values representing half the minimal

detectable level. The data set originated from a baseline survey of the bacterial contamination of the indoor and outdoor environments of 52 daycare centres (DCCs) in Québec, Canada (28).

In theory, if one uses the 'correct' imputation model, one can rigorously prove that final inferences are then also statistically 'correct' [27]. In practice, one can never be certain that values which are imprecise have been imputed from a perfect model. Nevertheless, it is always the case that much uncertainty is left unaccounted for by either ignoring or assuming 'exact' values for imprecise results.

MATERIALS AND METHODS

Data set

Indoor and outdoor environmental samples from 52 DCCs in Québec, Canada had been evaluated for bacterial contamination as follows:

Indoor sampling

Children and educators in one or two toddler groups in each DCC were asked to wash their hands for 40 s in a 200 ml rinse bag (Ziploc®, Dow Brands, London, Canada) filled with a solution of saline and 0·1% Tween 80 (Anachema, Lachine, Canada). Three randomly selected toys from each group's classroom were rinsed for 45 s in similar bags. If two groups shared the same room and were using the same toys during the day, only three randomly selected toys were sampled for the two groups. Bags from each toddler group (for children's hands, educators' hands and toy samples, respectively) were pooled. All samples were stored at 4 °C until analysed.

Outdoor sampling

Both the sandbox and the play area of each DCC were sampled. The surfaces of the sandbox and play area of each DCC were measured and divided into 25 equal areas and each sampled to maximize recovery and minimize the important spatial variation in contamination [20]. From each area, for the sandbox and play area separately, 100 ml of sand was obtained and put into a 5 l bottle (Fisher, Montréal, Canada) containing a solution of 2 l of saline with 0·1% of Tween 80. Therefore, a total of 2·5 l each of sand and soil were sampled. Bottles were shaken for 1 min by hand and left to sediment for an additional 1 min.

Two litres of this liquid were collected in a plastic bottle (Fisher, Montréal, Canada) and stored at 4 °C until analysis.

Bacteriological analysis

All samples were kept at 4 °C until arrival at the laboratory (less than 6 h). Bacteriological analyses were performed at the Centre de Recherche en Virologie, Institut Armand–Frappier, Laval, Québec, Canada. A membrane filtration method for the identification of FC was used [29]. First, the liquid was shaken to homogenize the sample. The samples were left to sediment at 4 °C overnight. The samples were then filtered (Millipore Ltd, Toronto, Canada), placed on m-FC medium (Difco, Montreal, Canada) and incubated at 44 °C for 24 h. Blue colonies with metallic sheen were counted. One dilution (1:100 ml) was used for each sample except for sandboxes and play areas where two dilutions were used (1:0·1 ml and 1:1 ml). Bacteriological results were reported as number of FC colonies per ml.

Methods for replacing imprecise values in a data set

Routine statistical methods used to estimate the distribution of the average number of FCs or to identify factors associated with contamination cannot be used when the laboratory data contain imprecise values. Imprecise values were defined as any laboratory result where either no FC colonies had grown at the lowest screening dilution used or the number of colonies at the highest screening dilution were too dense to count (e.g. < 50, < 1, > 100 FC/ml, etc.). There was no specific number of colonies defined as too high to be counted. The laboratory technician subjectively assessed what could be counted. Usually, colonies could not be counted when they had entirely covered the petri dish.

Standard method

There are several types of methods commonly used to deal with imprecise microbiological data. These include assuming a zero value for the imprecise count, dichotomizing the results [21–24], adding a value of 1 and then $\log_{10}$ transforming the data (therefore attributing a value of 1 to every imprecise value) [9, 19], and replacing the imprecise value with a value equal to half the lowest screening dilution used [13, 19, 25]. This latter method is perhaps the most common and will be used as the comparison method in this paper. All imprecise values in the data set are replaced by single 'exact' numbers equal to half the lowest screening dilution used. The data are then $\log_{10}$ transformed and analysed. If the number of colonies were too numerous to be counted, we have used the average between the value of the highest screening dilution used and the sum of that highest screening dilution plus one $\log_{10}$.

Multiple imputation method

Each imprecise value is replaced by an imputed value obtained from a distribution of values randomly selected from within a feasible range. Since we are uncertain about the true value for the imprecise value, we reflect this uncertainty by selecting more than one value. We then average over these values, while incorporating the imputation-to-imputation variance. In this way we correctly account for all inherent uncertainty. More precisely, using the example of FC counts when no colonies have been observed, one assumes that the imprecision in the true value could be represented by a probability distribution covering a feasible range between the minimal detectable value to the number of FC that would have been reported had one colony grown at that dilution. We define the latter as the 'maximal possible $\log_{10}$ (FC)'. A uniform distribution over the possible range of values was selected to best represent the uncertainty in our data set, although in principle any distributional form can be used. We chose to use a uniform distribution rather than estimating values from the rest of the data because, as is common in environmental microbiology, values were not missing at random but rather from the lower tail of the distribution of FC counts.

The multiple imputation algorithm proceeds as follows: an independent uniform random number is selected for each imprecise value from its feasible range. These imputed values are then combined with that portion of the data set that is precisely known to form a 'complete' data set of 'exact' values. Statistical analyses then proceed as usual for this now 'complete' data set, and the results are stored. This procedure is then repeated a large number of times (we used 10 000 repetitions), providing slightly different results each time, since the randomly imputed values are different in each iteration. Final inferences are then computed which account for both usual sampling variability (within iteration variances of parameter estimates) and variability due to the imprecise data (between

iteration variances of parameter estimates). The final variance for each parameter is essentially the sum of these two components. Multiple imputation is attractive, since no specialized statistical methods are required to account for imprecise values, as in each iteration, one creates a 'complete' data set. This Bayesian approach requires the calculation of a joint posterior distribution over all unknown parameters, which in turn leads to a predictive distribution for future, missing or imprecise data. Therefore, once a model is created for the full data, imputations can be created with little extra effort, if a Bayesian approach is used. See Rubin for details [27].

Interval censored regression method

The procedure intreg of Stata release 6 was used to run a model with a dependent variable that could have any value between a defined interval [35]. The programme samples the value for the imprecise data from a normal distribution between the two limits of the interval, assuming the errors are normally distributed. In the current model, intreg samples values from the feasible range (log transformed), which is the equivalent of sampling from a log normal distribution. This method is computationally more efficient that the multiple imputation procedure described above, and could be preferred in cases where all of the assumptions are likely to hold.

Statistical analysis

As is common for this type of data, the FC counts were log-transformed in order to normalize their distribution [19]. Since the unit of analysis was the DCC and since the group sizes of children and number of educators per group varied, weighted averages were used to calculate mean $\log_{10}$ (FC counts) for children's and educators' groups within each DCC.

Minimum detectable values are required for the multiple imputation method. The minimal detectable number was assumed to be 1 colony per 100 ml for outdoor samples from the sandbox and the play area, and 1 colony per 200 ml was set as the minimum detectable number for indoor samples in the rinse bags. For the standard method, half the value of the minimal screening dilution used was assumed to be an exact value. For the multiple imputation method, the

outdoor imprecise values were sampled from a uniform distribution ranging from $\log_{10}$ (0·01) to the $\log_{10}$ of the number of FC that would have been found at the minimal screening dilution used (maximal possible $\log_{10}$ (FC counts)). Similarly, the indoor imprecise values were sampled from a uniform distribution ranging from $\log_{10}$ (0·005) to the maximal possible $\log_{10}$ (FC counts). When the number of colonies were too numerous to be counted, we used the minimum number of colonies reported by the laboratory as the lower tail of the distribution and the sum of that minimum number of colonies plus one $\log_{10}$ for the upper tail of the distribution. The imprecise values were then sampled from a uniform distribution ranging from that minimum (in $\log_{10}$ scale) to that maximum. We used non-informative priors so that our inferences are based solely (or almost) on the data. We used the method of Raftery and Lewis [30] to ensure accuracy of final estimates. Convergence was assured by running the model many times from widely divergent starting points and checking that all results were very close in each run.

For the two methods, descriptive statistics were first computed and reported as the median and Inter-Quartile Range (IQR), since the log-transformed data still remained quite skewed. SAS software was used to obtain summary statistics (quartiles, IQR, medians) [31].

While in theory multiple imputation could be used simultaneously with a model selection procedure, this would involve creating a large number of imputation data sets for each possible model under consideration. For ease of calculation, therefore, we carried out our model selection procedure using a single 'complete' data set where imprecise values were replaced by exact values corresponding to the midpoint in the $\log_{10}$ scale of the feasible range. We preferred the midpoint on the $\log_{10}$ scale because it is in fact the $\log_{10}$ scale that is used in the models. This method was used to identify the risk factors associated with bacterial contamination in the play area (with and without bacterial contamination in the sandbox as an independent variable) and bacterial contamination of children's hands (with bacterial contamination of educators' hands as an independent variable). The factors tested for their association with the outcome for the different models are described in Table 1. The Bayesian Information Criterion (BIC) was used to select the model best supported by the data using each imputation method. The BIC is designed to give, on average, the best out-of-sample predictions and is not

Table 1. *Candidate predictors for model selection using multiple linear regression for predicting log FC counts in the play area and on children's hands*

Dependent variable	Candidate predictors
FC in the play area	Sighting of cats in the backyard, sighting of dogs in the backyard, sighting of squirrels in the backyard, sighting of birds in the backyard, sighting of raccoons in the backyard, geographical region, presence of residential neighbours, outdoor temperature the afternoon of sampling (°C), outdoor humidity level the afternoon of sampling (percent), presence of grass in the backyard, presence of soil in the backyard, surface of the play area (m²)
FC on children's hands	Presence of a disease outbreak in the DCC the day of sampling, sampling of children's hands after outdoor play, sampling of children's hands after handwashing, average indoor humidity level the day of sampling (percent), average indoor temperature the day of sampling (°C), presence of the two groups of children in a same room, sharing of rooms with older age children, average classroom size, average space per child (m³), presence of an infants' room, geographical region, legal status, proportion of educators with a formal training, products used to disinfect toys, frequency of toys' cleaning, housekeeping done every day, participation in a government subsidy program, presence of a poster on handwashing, presence of a poster on infectious diseases, years of operation of the DCC, intervention by public health authorities because of an outbreak in the past 9 months, use of potties, log (FC) on educators' hands

subject to the overfitting tendencies of the usual backward and forward model selection techniques [32]. The BIC was computed using ProcReg ln SAS 6.11 as $n \times \ln(1-r^2)+[(p) \times \ln(n)]$ where r^2 is the coefficient of determination for the current model, n is the sample size and p the number of parameters in the model without the intercept [31]. Estimates of the linear regression coefficients and their 95% credible intervals (CI), which are Bayesian analogs of the usual confidence intervals, were computed. The statistical package BUGS was used to compute the estimates of linear regression coefficients and for the multiple imputation [33]. The interval regression method was applied to the play area using the procedure intreg in Stata. The BUGS programme we used for our analysis is available from the authors. The goodness-of-fit and extreme values of the models were checked for all models using jackknife residuals and Cook distances [34], using SAS software [31].

RESULTS

The percentage of imprecise values reported by the bacteriology laboratory for the daycare study ranged from 25 to 60% (Table 2). For educators' hands and both indoor and outdoor toys, more than 40% of the samples had imprecise values.

Table 3 reports the median values and IQRs across DCCs of the different samples using the standard and the multiple imputation methods. It is important to remember that the results are presented on the $\log_{10}$ scale and per ml of sample, which reduces the apparent size of any difference. Both methods had the samples ranked in the same order, in terms of amount of contamination. The most contaminated area was the play area, followed by the sandbox and children's hands. Outdoor toys had higher median counts than indoor toys. The standard method gave higher median values than the multiple imputation method except for educators' hands and the play area. The IQRs from the multiple imputation method were always larger than the ones calculated from the standard method, except for the outdoor toys where they were almost equal.

The performance of the two methods is contrasted and compared in Tables 4 (play area) and 5 (children's hands). In addition, the interval regression method results are reported in Table 4.

Model for bacterial contamination in the play area (Table 4)

The model selected includes the temperature on the day of sampling (°C), sighting of cats in the backyard

Table 2. *Number and levels of imprecise values for faecal coliform (FC) counts reported by the laboratory following bacteriological analysis of samples taken in 52 daycare centres in Québec, Canada*

Sample	Total no. of samples	<50	<10	<1	<0·9	<0·1	<0·09	<0·01	>3	>30	(%)
Sandbox	52	8	2		6						16 (31%)
Play area	52	4	4		5						13 (25%)
Outdoor toys	52			1		20				1	22 (42%)
Children's hands (group 1)	52					8	3	4	1		16 (31%)
Children's hands (group 2)	51*					11	3	3	1		18 (35%)
Educators' hands (group 1)	51†					9	3	8	2		22 (43%)
Educators' hands (group 2)	51*					10	3	9			22 (43%)
Indoor toys (group 1)	52					12	3	11			26 (50%)
Indoor toys (group 2)	45*‡					11	5	11			27 (60%)

* One daycare centre only had one group of toddlers.

† One educator's hands sample had leaked and could not be used for the analysis.

‡ The same toys were used by the two groups of children in six daycare centres.

Table 3. *Median logarithm (in base 10) of faecal coliform counts and interquartile ranges obtained using two different methods* for taking account of imprecise values†*

	Method used to account for imprecise values					
	Standard method			Multiple imputation		
Sample ($n = 52$)	m‡	QR	(Q1, Q3)§	m‡	QR	(Q1, Q3)§
Sandbox	1·389	1·331	(0·699, 2·030)	1·033	1·713	(0·317, 2·030)
Play area	1·515	1·651	(0·651, 2·302)	1·568	2·088	(0·214, 2·302)
Outdoor toys	−0·61	2·000	(−1·301, 0·699)	−0·7	2·001	(−1·302, 0·699)
Children's hands	−0·42	1·812	(−1·301, 0·511)	−0·48	1·910	(−1·429, 0·480)
Educators' hand‖	−1·15	1·085	(−1·347, −0·261)	−1·11	1·369	(−1·687, −0·318)
Indoor toys	−1·3	1·219	(−1·706, −0·487)	−1·43	1·377	(−1·875, −0·498)

* Standard method: Take half of the minimal detectable value and then $\log_{10}$ transform; Multiple imputation method: Impute value from a uniform distribution of $\log_{10}$ values over the feasible range.

† The data set containing imprecise values originated from a study of indoor and outdoor environmental contamination in 52 daycare centers in Québec, Canada.

‡ m, median logarithm (in base 10) of faecal coliform counts.

§ IQR (Q1, Q3): Interquartile range (first quartile, third quartile) of $\log_{10}$ of faecal coliform counts.

‖ n, 51 DCCs.

(by the DCC director), the presence of residential neighbours (neighbours could either be residential or industrial) and the presence of soil in the backyard. Linear regression coefficient estimates from the standard method were always lower than the estimates obtained using the multiple imputation and the interval regression methods. The two latter methods gave very similar results. Credible intervals were always narrower using the standard method. The presence of soil in the backyard was significantly associated with bacterial contamination in the play area when using the multiple imputation method ($\beta = 0·73$; 95% CI 0·05, 1·44) or the interval regression method ($\beta = 0·7200$; 95% confidence interval = 0·0912, 1·3489) but not when the standard method was used ($\beta = 0·45$; 95% CI −0·13, 1·09).

Model for bacterial contamination on children's hands (Table 5)

The model selected includes the use of a disinfectant (other than bleach) to clean the toys, the frequency of toy cleaning, the presence of a poster on infectious diseases at the DCC (usually affixed to the wall at the entrance of the DCC), the use of cloth toys by toddlers, housekeeping done every day and $\log_{10}$ (FC) on educators' hands. In this model, imprecise values had been reported in both the dependent variable and

Table 4. *Comparison of the multiple linear regression coefficient estimates for predicting the logarithm (in base 10) of faecal coliform counts in the play area using the standard, interval regression and multiple imputation methods for imprecise values*

	Multiple linear regression coefficients (95% Credible Interval)				Multiple linear regression coefficients (95% Confidence Interval)	
	Standard method		Multiple imputation method		Interval regression	
Constant	−0·3874	(−1·4420, 0·6456)	−0·9850	(−2·2870, 0·28760)	−0·9847	(−2·1517, 0·1824)
Temperature*	0·0599	(0·0139, 0·1083)	0·0683	(0·0133, 0·12500)	0·0693	(0·0176, 0·1210)
Cats†	0·7469	(0·1699, 1·3360)	1·0090	(0·3546, 1·6670)	1·0014	(0·3961, 1·6067)
Neighbors‡	0·3678	(−0·2013, 0·9135)	0·5458	(−0·0861, 1·1940)	0·5383	(−0·0603, 1·1369)
Soil§	0·4760	(−0·1289, 1·0870)	0·7275	(0·0466, 1·4420)	0·7200	(0·0912, 1·3489)

* Outdoor temperature the afternoon of sampling (°C).
† Sighting of cats in the backyard.
‡ Presence of residential neighbours.
§ Presence of soil in the play area.

Table 5. *Comparison of the multiple linear regression coefficient estimates for predicting the logarithm (in base 10) of faecal coliform counts on children's hands using the two methods for imprecise values*

	Multiple linear regression coefficients (95% Credible Interval)			
	Standard method		Multiple imputation method	
Constant	−0·8958	(−2·072, 0·257)	−0·0862	(−1·7220, 2·1060)
Disinfectant*	−0·7551	(−1·343, −0·189)	−0·8448	(−1·4420, −0·2380)
Toy cleaning 1†	1·5520	(0·698, 2·419)	1·8110	(0·9000, 2·7270)
Toy cleaning 2†	1·0579	(0·202, 1·909)	0·8851	(−0·0430, 1·8170)
Toy cleaning 3†	1·159	(0·046, 2·299)	1·1170	(−0·1160, 2·3460)
Poster‡	0·7882	(0·154, 1·427)	0·4641	(−0·2800, 1·1850)
Cloth§	0·7398	(0·119, 1·384)	0·9283	(0·2620, 1·5970)
Housekeeping‖	−1·0540	(−1·727, −0·358)	−0·9597	(−1·6570, −0·2520)
Educators' hands¶	0·6208	(0·311, 0·923)	1·3900	(0·3350, 2·5690)

* Use of a disinfectant other than bleach to clean the toys.
† Frequency of toy cleaning, with cleaning every 1 or 2 days as the reference group; toy cleaning 1: toys cleaned once a week; toy cleaning 2: toys cleaned at least once a month but less than once a week; toy cleaning 3: no systematic rules of cleaning.
‡ Presence of a poster on infectious diseases at the DCC.
§ Children play with cloth toys.
‖ Housekeeping done every day.
¶ Logarithm (in base 10) of FC counts on educators' hands.

one of the independent variables and therefore needed to be taken into account in the analysis. Important discrepancies in the results obtained were found between the two methods. In the multiple imputation method, the 95% CIs were always larger than those using the standard method, the most striking difference being in the 95% CI of the multiple linear coefficient of educators' hands. The point estimate of the effect of $\log_{10}$ (FC) on educators' hands is more than twice the point estimates using the standard method. However, the large 95% CI for this variable indicates that uncertainty in its measurement is large. It is of course not surprising that the multiple imputation method is usually associated with the largest CIs, since the uncertainty inherent in the imprecise values is included in the analyses, in contrast to the standard method which assumed 'exact' values. In addition, variables that seemed significant using the standard method became non-significant using the multiple imputation method. For example, using the standard method, the presence of a poster on infectious disease in the classroom was found to be associated with an increased risk of contamination on children's hands whereas it was not significant using the multiple imputation method. The biological and epidemiological interpretation of these results are not

only difficult but they can be different depending on which method was used to impute the imprecise values.

DISCUSSION

Both methods provided similar estimations of median bacterial contamination, except in one instance (the sandbox) where it was higher using the standard method. The small differences observed could be explained by the narrow range of possible values from which imputed values were selected in the multiple imputation method. If preliminary results had suggested a larger variation in the number of faecal coliform counts, we might have used a wider feasible range for imputing the imprecise values which would have reflected a greater variability in the data.

As the outcome is expressed in $\log_{10}$ scale, differences in the point estimates of the multiple linear regression coefficients and their 95% CI appear small. It is interesting to note that, even though the uncertainty around the estimates was larger using the multiple imputation method, the point estimates were also considerably larger, changing the interpretation one can make about the strength of association between risk factors and play area contamination.

Methods other than multiple imputation may sometimes be usable for adjusting for missing or censored data of the type discussed here. For example, the procedure intreg contained in the Stata statistical package [35] implements normal parameter estimation for interval censored data. We used the intreg procedure to run a multiple linear regression with the data on the $\log_{10}$ (FC) in the sandbox (data not shown) and play area. The coefficient estimates were very similar to those obtained with the multiple imputation method. This procedure does not directly apply to the indoor data because an average of two observations is used, with missing values occurring in one or both of the observations. In the case of a single missing value, a reasonable *ad hoc* procedure would be to consider the non-missing value as the average. If both are missing, one can consider the average as a single missing value, and again apply the intreg procedure.

The model selected to predict bacterial contamination on children's hands shows important discrepancies between the two methods. The model includes two variables with imprecise values: children's hands and educators' hands. The 95% CI of the effect of the $\log_{10}$ (FC) on educators' hands on the

$\log_{10}$ (FC) on children's hands is much larger using the multiple imputation method. This is because an outcome with some uncertainty is predicted using a variable which also has some uncertainty. The large 95% CIs obtained using multiple imputation demonstrate its more comprehensive accountability of variation in the data. Large 95% CIs will change the interpretation of the model and the resulting conclusion but they more appropriately reflect the increased uncertainty inherent in a data set which includes replacement of imprecise values by imputed values.

These results show that using multiple imputation for imprecise values yields more conservative estimates of credible intervals especially in models where one or more variables have imprecise values. In effect, multiple imputation would seem theoretically more valid, since it better incorporates the uncertainty inherent in the imprecise values. If the conclusion after applying multiple imputation is similar to the results obtained using the standard method, then one should feel confident that the conclusions are robust to this sort of uncertainty. On the other hand, if the conclusions change, then this paper illustrates the importance of including the extra source of uncertainty in the analysis. The exact form of the imputation distribution, however, can play a role. Using a model selected with values assumed on the $\log_{10}$ scale may often be the best choice, given that the range of values on the logarithmic scale is more realistic for bacterial data. FC counts follow a $\log_{10}$ normal distribution [19], consequently, using the $\log_{10}$ normal scale to impute values (for the imprecise values) is coherent with standard practice. A uniform distribution on the $\log_{10}$-transformed scale seems the most plausible choice, but other distributions could also be considered depending on the data source.

Bacteriological results often include imprecise values [2, 6, 11, 14–16, 19, 21–25]. How imprecise values are incorporated into the data analysis is critical to correctly interpret the results. Replacing imprecise values with singly imputed 'exact' values for a moderate or large proportion of the data would seem unwarranted because the standard errors of any results would be artificially too small and consequently lead to questions of data validity. By allowing the values of the imprecise results to vary within the feasible range, multiple imputation improves both the validity and reported precision of the statistical inferences made from the data. These results are important for researchers, health professionals, public

health authorities and others who are involved in the formulation and regulation of standards for bacteriological contamination and for related research.

ACKNOWLEDGEMENTS

REFERENCES

1. Clark DL, Milner BB, Stewart MH, Wolfe RL, Olson BH. Comparative study of commercial 4-Methyl-umbelliferyl-$\beta\phi$-D-Glucuronide preparations with the standard methods membrane filtration fecal coliform test for the detection of *Escherichia coli* in water samples. Appl Environ Microbiol 1991; **57**: 1528–34.

2. Schoeni JL, Wong ACL. Inhibition of *Campylobacter jejuni* colonization in chicks by defined competitive exclusion bacteria. Appl Environ Microbiol 1994; **60**: 1191–7.

3. Moyer NS, Martinetti G, Holcomb LA, Hall NH, Altwegg M. Application of ribotyping for differentiating aeromonads isolated from clinical and environmental sources. Appl Environ Microbiol 1992; **58**: 1940–4.

4. Le Guyader F, Apaire-Marchais V, Brilet J, Billaudel S. Use of genomic probes to detect hepatitis A virus and enterovirus RNAs in wild shellfish and relationship of viral contamination to bacterial contamination. Appl Environ Microbiol 1993; **59**: 3963–8.

5. Paul JH, Rose JB, Jiang S, Kellogg C, Shinn EA. Occurrence of fecal indicator bacteria in surface waters and the subsurface aquifer in Key Largo, Florida. Appl Environ Microbiol 1995; **61**: 2235–41.

6. Lamosa P, Martins LO, Da Costa MS, Santos H. Effects of temperature, salinity, and medium composition on compatible solute accumulation by *Thermococcus* spp. Appl Environ Microbiol 1998; **64**: 3591–8.

7. Nagata S, Adachi K, Sano H. Intracellular changes in ions and organic solutes in halotolerant *Brevibacterium* sp. Strain JCM 6894 after exposure to hyperosmotic shock. Appl Environ Microbiol 1998; **64**: 3641–7.

8. Potrawfke T, Timmis KN, Wittich R-M. Degradation of 1,2,3,4-tetrachlorobenzene by *Pseudomonas chloraphis* RW71. Appl Environ Microbiol 1998; **64**: 3798–806.

9. Termorshuizen AJ, Davis JR, Gort G et al. Interlaboratory comparison of methods to quantify microsclerotia of *Verticillium dahliae* in soil. Appl Environ Microbiol 1998; **64**: 3847–53.

10. Waterman SR, Small PLC. Acid-sensitive enteric pathogens are protected from killing under extremely acidic conditions of pH 2·5 when they are inoculated onto certain solid food sources. Appl Environ Microbiol 1998; **64**: 3882–6.

11. Munkvold G, Stahr HM, Logrieco A, Moretti A, Ritieni A. Occurrence of fusaproliferin and beauvericin in *Fusarium*-contaminated livestock in Iowa. Appl Environ Microbiol 1998; **64**: 3923–6.

12. Nakaune R, Adachi K, Nawata O, Tomiyama M, Akutsu K, Hibi T. A novel ATP-binding cassette transporter involved in multidrug resistance in the phytopathogenic fungus *Penicillium digitatum*. Appl Environ Microbiol 1998; **64**: 3983–8.

13. Hielm S, Bjorkroth J, Hyytia E, Korkeala H. Prevalence of *Clostridium botulinum* in Finnish trout farms: pulsed-field gel electrophoresis typing reveals extensive genetic diversity among type E isolates. Appl Environ Microbiol 1998; **64**: 4161–7.

14. Roelof J, Werlen C, Nishino SF, Spain JC. Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. Appl Environ Microbiol 1998; **64**: 4185–93.

15. Adams CA, Austin B, Meaden PG, McIntosh D. Molecular characterization of plasmid-mediated oxytetracycline resistance in *Aeromonas salmonicida*. Appl Environ Microbiol 1998; **64**: 4194–201.

16. Federici BA, Bauer LS. Cyt1Aa protein of *Bacillus thuringiensis* is toxic to the cottonwood leaf beetle, *Chrysomela scripta*, and suppresses high levels of resistance to Cry3Aa. Appl Environ Microbiol 1998; **64**: 4368–71.

17. Munoz D, Castillejo JI, Caballero P. Naturally occurring deletion mutants are parasitic genotypes in a wild-type nucleopolyhedrovirus population of *Spodoptera exiga*. Appl Environ Microbiol 1998; **64**: 4372–7.

18. Bukhari Z, McCUin RM, Fricker CR, Clancy JL. Immunomagnetic separation of *Cryptosporidium parvum* from source water samples of various turbidities. Appl Environ Microbiol 1998; **64**: 4495–9.

19. Gale P. Coliforms in the drinking-water supply; what information do the 0/100 ml samples provide? J Water SRT-Aqua 1996; **45**: 155–61.

20. Carabin H, Gyorkos TW, Payment P, Kokoskin E, Joseph L, Soto JC. Comparison of methods of sampling for the recovery of *Toxocara* spp. and fecal coliforms in an outdoor day care environment. Can J Infect Dis 1998; **9**: 149–56.

21. Ekanem EE, Dupont HL, Pickering LK, Selwyn BJ, Hawkins CM. Transmission dynamics of enteric bacteria in the day-care centers. Am J Epidemiol 1983; **118**: 562–72.

22. Holaday B, Pantell R, Lewis C, Gillis CL. Patterns of fecal coliform contamination in day-care centers. Publ Hlth Nursing 1990; **7**: 224–8.

23. Van R, Morrow AL, Reves RR, Pickering LK. Environmental contamination in child day-care centers. Am J Epidemiol 1991; **133**: 460–70.

24. Weniger BG, Ruttenberg AJ, Goodman RA, Juranek DD, Wahlquist SP, Smith JD. Fecal coliforms on environmental surfaces in two day care centers. Appl Environ Microbiol 1983; **45**: 733–5.

25. Laborde DJ, Weigle KA, Weber DJ, Kotch JB. Effect of fecal contamination on diarrheal illness rates in day-care centers. AM J Epidemiol 1993; **138**: 243–55.

26. Laborde DJ, Weigle KA, Weber DJ, Sobsey MD, Kotch JB. The frequency, level, and distribution of fecal contamination in day-care center classrooms. Pediatrics 1994; **94** (Suppl 6): 1008–11.

27. Rubin D. Introduction. In: Rubin D, ed. Multiple imputation for non response in surveys. New York: John Wiley & Sons; 1987: 21.

28. Carabin H, Gyorkos TW, Soto JC, Joseph L, Payment P, Collet J-P. Effectiveness of a training programme in reducing infections in toddlers attending daycare centres. Epidemiol 1999; **10**: 219–27.

29. American Public Health Association. Standard methods for the examination of water and wastewater, 18th ed. Washington D.C.: American Public Health Association 1992.

30. Raftery AE, Lewis S. How many iterations in the Gibbs sampler? In: Bernado JM, Bergerm JO, Dawid AO, Smith AFM, eds. Bayesian Statistics 4. Oxford: Oxford University Press, 1992: 763–73.

31. SAS Institute, Inc. SAS/STAT user's guide, version 6.11, vol 2. Cary (NC): SAS Institute, Inc, 1990.

32. Kass RE, Raftery AE. Bayes' factors. J Am Stat Assoc 1995; **90**: 773–95.

33. Spiegelhalter DJ, Thomas A, Best NG, Gilks WR. BUGS: Bayesian Inference Using Gibbs Sampling. version 0.5. Cambridge: MRC Biostatistics Unit, 1995.

34. Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariable methods, 2nd ed. Belmont (CA): Duxbury Press, 1988.

35. StataCorp. Stata Statistical Software: release 6.0. College Station, Texas: Stata Corporation 1999.