Journal of Industrial Microbiology, 12 (1993) 273–276

Published by The Macmillan Press Ltd

# Protocols for data generation for predictive modeling

S.J. Walker and J.E. Jones

*Campden Food and Drink Research Association, Chipping Campden, Glos. GL55 6LD, UK*

(Received 31 March 1993)

*Key words:* Modeling; Data; Protocols; Database; Microbiology

## SUMMARY

Essential to the development of useful predictive models for foods is the collection of appropriate data. Within the UK Predictive Food Microbiology Program, a series of documents have been produced to aid the standardization of data collection by a number of laboratories. Documents include a protocols form, notes on experimental design, notes on the accuracy of counts by plating, the preferred method of calculating a plate count value, preferred MPN tables and the preferred methods of recording data values. These documents have proven useful in aiding collaboration between laboratories.

## INTRODUCTION

Within the UK, the Government (Ministry of Agriculture, Fisheries and Food) commissioned a large program of work on predictive microbiology. One of the aims of this program was to develop a set of mathematical models to predict the growth, survival and death of food-borne pathogens under conditions relevant to foods. Development of the models for ten pathogenic bacteria (*Aeromonas hydrophila, Bacillus cereus, B. subtilis, Campylobacter, Clostridium botulinum, Escherichia coli, Listeria monocytogenes, Salmonella, Staphylococcus aureus* and *Yersinia enterocolitica*) require the generation of large quantities of data relating to the kinetic changes of the bacteria. This activity has been central to the UK Predictive Food Microbiology Program (UKPFMP). Early in the UKPFMP it was recognized that a single laboratory could not produce all the necessary data within a reasonable time-scale. Therefore a number of laboratories were involved in data collection. The main contributing laboratories were the AFRC Institute of Food Research (Reading and Norwich Laboratories), Campden Food and Drink Research Association, Flour Milling and Baking Research Association, Leatherhead Food Research Association and Torry Research Station. Other contributors included universities and food companies. As a number of data generation sites were involved in this collaborative effort, it was considered important that there was a degree of consistency between the laboratories. Furthermore, the production of data of high quality is essential to the development of good quality models. Therefore, one of the first tasks of the UKPFMP was to appoint a Protocols Group to address these issues. The purpose of this paper is to present the recommendations of this group. This may help other laboratories involved in data generation for predictive microbiology. Complete copies of the documents can be obtained from the authors.

Correspondence to: S.J. Walker, Campden Food and Drink Research Association, Chipping Campden, Glos. GL55 6LD, UK.

## PROTOCOLS GROUP

The Protocols Group was comprised of representatives from the main participating laboratories outlined previously and was chaired by Unilever Research Laboratory (Colworth House, Sharnbrook, Beds., UK). Composition of the group included microbiologists, mathematicians and statisticians.

To help improve the quality of data and standardize the production and submission of data to the database, the Protocols Group developed a series of documents. The documents are not mandatory for the acceptance of data to the database, but are intended to help contributors. Within the program, it was considered necessary to have a degree of flexibility in order to accommodate as much data as possible. It was recognized that data might be available from laboratories commissioned to do work for other purposes. The following documents may help them plan the practical work so it is also suitable for modeling or model validation. The documents are as follows:

### *Protocols form*

The primary objective of this was to ensure that, as far as is possible, the proposed work will be suitable for modeling. Therefore, this form should be completed prior to the commencement of practical work. As discussed previously, data production is costly and it would be unfortunate if the data collected were not appropriate for modeling. Perhaps the laboratory need only consider relatively little extra effort to ensure that data are suitable.

The document may also be used as an 'aide memoire' such that by completing this document, even laboratories with little previous experience can produce better data and collect the relevant ancillary data.

The first section of the document identifies the general and specific objectives of the proposed work. This may indicate that the primary aim of the study was not specifically for modeling which may affect subsequent recommendations.

The second section poses a series of questions about the microorganism(s) of concern (i.e. the strains to be used, reason(s) for choice of particular strain(s), method of maintenance of cultures, inoculation level, inoculation procedure and inoculum composition).

The next series of questions concerns the growth substrate. For example, will the substrate be a microbiological medium or a food? If a food is to be used, it is important to know if it is sterile. If not, what other microorganisms are likely to be present? The site of inoculation of a food might also be relevant.

Finally, information relating to the specific factors (or variables) to be investigated and ancillary information is recorded (e.g. temperature, pH, salt, preservatives, atmosphere etc.) Questions include the range and values of factors used, the methods used to adjust the factors (e.g. acidulant or humectant used), the method of measurement and the frequency of measurement of the factor.

Overall, completing the Protocols Form will help ensure that the primary issues relevant to data collection are considered. In addition, when submitted to one of the participating laboratories in the UKPFMP, this gives the opportunity for discussion with workers experienced in data collection for modeling.

*Notes on experimental design*

The experimental design for data collection is very important to maximize the efficiency of the proposed work. Use of an inappropriate design may result in the user collecting too much data or even limit its use for modeling. Like the Protocols Form, the notes on Experimental Design, provide guidance for the user, but also permit a degree of flexibility.

For the development of kinetic models, it recommends that for a particular combination of conditions, a minimum of 10 data points should normally be collected. It was recognized that curve fitting can be done with fewer data points, but the position of such points is critical for accurate fitting. The points must be positioned at the areas of inflection where the rate of change of the microbial kinetics is maximal, e.g. at the end of the lag phase and end of the logarithmic phase of growth. Failure to position these points correctly may result in a fitted curve that poorly reflects the true growth of the microorganism.

The selection of the factor values to be investigated is also important and must encompass the range of values of interest to the potential user. For example, in some cases the factor values may be equally spaced (e.g. pH, 7.0, 6.5, 6.0, 5.5 and 5.0) whilst in others it is often appropriate that they should be concentrated near the boundaries for growth (e.g. pH, 7.0, 6.0, 5.5, 5.2, 5.0). With this latter distribution, the factor values are concentrated in the areas where the maximum rates of change of the microbial response are expected. This distribution is generally preferable, but the experimenter needs prior knowledge of the boundary for growth. Furthermore, it is important to identify if there are other variables dependent or linked to the factor under consideration i.e. the concentration of undissociated acid when organic acids are used to adjust the pH of a growth substrate.

The inoculation of the growth substrate is important. Use of a single strain inoculum may be easier to model and shows how a single strain behaves over a wide range of environmental conditions. Within the UKPFMP it is generally recommended that a cocktail of defined strains are used, as different strains in the cocktail may grow preferentially at different combinations of conditions. Typically, the cocktail will contain between three and five strains. Therefore, the data (and model) reflect the 'leading edge' of the microbial response. The level of inoculation to study the growth of a microorganism must be high enough to ensure that all portions of the substrate contain the relevant microorganisms at a level which can be enumerated, but low enough to allow growth to be measured before the microorganisms change the environment. Typically, a level of 10–1000 CFU per g is recommended. If the growth of a microorganism is likely to significantly change the environment, then the change must be recorded or appropriate control measures adopted to minimize this. For example, buffers may be added to prevent a change in pH. For death and survival studies, a higher level of inoculation will generally be required and it is recommended that this should allow a reduction of 6 log cycles to be followed.

Correct incubation of samples is critical and it is generally recommended that growth media are pre-equilibrated at the required temperature prior to inoculation. For example, inadvertent storage of a sample at room temperature for several hours when 0 °C was the required temperature will greatly affect the observed behavior. Ideally, the temperatures of storage should be continuously monitored throughout a study. If this cannot be done, then some measure of the temperature range experienced (e.g. minimum–maximum thermometers) should be collected along with routine temperature checking.

With both growth and death studies, the phase of growth and any pre-treatment of the microorganisms can affect the observed kinetics. Pre-treatments may include pre-conditioning of the microorganism to low temperature or low pH values before inoculation of the growth medium, or exposure of the microorganisms to sub-lethal heating temperatures prior to thermal inactivation studies. Similarly when investigating the heat resistance of spore-forming bacteria, the method of preparation of a sporecrop can greatly affect the destruction kinetics. Therefore, any such treatments should be recorded and must be consistent during a particular study.

The acquistion of data on the survival and death of microorganisms under conditions of stress (e.g. low pH or high salt) or by thermal treatments have some additional issues. Failure to allow the recovery of injured cells (e.g. by use of selective media) will overestimate the lethality of the conditions studied. The use of open heating systems

may underestimate lethality and so closed heating systems (e.g. capillary tubes or the submerged heating coil apparatus) are generally preferable. Heating kinetics of non-homogeneous substrates, particularly foods, may cause particular problems in determining the lethality of a treatment. Therefore the site of inoculation, use of a consistent substrate and temperature measurements within a substrate are all important.

*Preferred method for calculating a bacterial plate count value*

To ensure consistency, the number of microorganisms obtained from replicate agar plates from a single sample should be calculated using the procedure recommended by the International Standards Organization [1] and Bacteriological Analytical Manual [2]. The values obtained from replicate samples should not be averaged, but submitted to the database as separate values.

*Notes on the accuracy of counts of viable bacteria by plating*

Within the UK initiative, and that from other countries, most of the data used for kinetic modeling have been collected using plate counting techniques. While other systems have occasionally been used (e.g. optical density or impedance), there is not a general consensus on the application of the data for modeling. Therefore the UKPFMP have not yet developed detailed procedures for such data, other than a method to submit these to a database.

The precision of a bacterial count estimate is affected by the number of colonies counted. The notes indicate the maximum precision attainable under ideal conditions. As the size of the confidence intervals for a count value is dependent on the number of bacteria counted, when low numbers are present, the confidence intervals represent a large proportion of the value. The number of bacteria counted in a sample may be increased by using a more concentrated dilution, increasing the volume of solution plated or increasing the number of replicate plates at a dilution. With all of these, the confidence intervals become proportionally less significant. Consequently, the data may be of greater use to a modeler. It is not however, a requirement of the UKPFMP that confidence intervals are submitted with data. The confidence intervals (95%) for low count values using 2, 3, 4 or 5 replicate plates to obtain a mean value are shown in Table 1. Should the mean colony count exceed 30 (for any number of replicates) then the ratio between the highest and lowest individual values should not exceed 2 to 1. This rule was derived pragmatically.

It was intended that all data produced by laboratories be submitted to the database of the UKPFMP and should not have been censored by the laboratory. Therefore, the use of 'flags' to indicate that a data value may be limited, is recommended. A 'flag' is a notation which may be linked to a particular data value. This is of interest in collaborative programs of work, as the person modeling the data may not necessarily be the originator of the data. Therefore the modeler has an indication of the confidence the laboratory generating the data has in individual values. The presence of a 'flag' does not automatically result in the exclusion of the data value from modeling activities.

TABLE 1

Limits for individual replicates from plate count techniques

| Mean count per replicate | 95% limits on individual counts from: (no. of replicates) | | | |
|---|---|---|---|---|
| | 2 | 3 | 4 | 5 |
| 1 | 0–4 | 0–5 | 0–5 | 0–5 |
| 2 | 0–6 | 0–7 | 0–7 | 0–7 |
| 3 | 0–8 | 0–8 | 0–9 | 0–9 |
| 4 | 1–10 | 1–10 | 1–11 | 1–11 |
| 5 | 1–11 | 1–12 | 1–12 | 1–12 |
| 6 | 2–13 | 2–13 | 2–14 | 1–14 |
| 7 | 2–14 | 2–15 | 2–15 | 2–16 |
| 8 | 3–16 | 3–16 | 3–17 | 3–17 |
| 9 | 4–17 | 3–18 | 3–18 | 3–18 |
| 10 | 4–18 | 4–19 | 4–20 | 4–20 |
| 11 | 5–20 | 5–20 | 4–21 | 4–21 |
| 12 | 6–21 | 5–22 | 5–22 | 5–22 |
| 13 | 6–22 | 6–23 | 6–24 | 5–24 |
| 14 | 7–24 | 7–24 | 6–25 | 6–25 |
| 15 | 8–25 | 7–26 | 7–26 | 7–27 |
| 16 | 8–26 | 8–27 | 8–28 | 8–28 |
| 17 | 9–28 | 9–28 | 8–29 | 8–29 |
| 18 | 10–29 | 9–30 | 9–30 | 9–31 |
| 19 | 11–30 | 10–31 | 10–32 | 10–32 |
| 20 | 11–31 | 11–32 | 11–33 | 10–33 |
| 21 | 12–33 | 12–33 | 11–34 | 11–35 |
| 22 | 12–34 | 12–35 | 12–36 | 12–36 |
| 23 | 14–36 | 13–36 | 13–37 | 13–37 |
| 24 | 14–37 | 14–37 | 13–38 | 13–38 |
| 25 | 15–38 | 15–38 | 14–38 | 14–40 |
| 26 | 16–39 | 15–40 | 15–41 | 15–41 |
| 27 | 17–40 | 16–41 | 16–42 | 16–42 |
| 28 | 17–41 | 17–42 | 17–43 | 16–43 |
| 29 | 18–42 | 18–43 | 17–44 | 17–45 |
| 30 | 19–44 | 18–45 | 18–46 | 18–46 |

'Flags' should be attached to values which were obtained from replicate plates with a low mean count (<15), a high mean count (>300), poor replication as described above or when the method for calculating a count was not that outlined previously.

In addition to the above, the accuracy of a population estimate may be affected by random or systematic errors. For example, random errors may be due to errors in the volumes plated or preparation of dilutions. These can be minimized by careful operator control. Systematic errors may have several causes, e.g. the use of selective or nutritionally deficient media which fail to recover all the microorganisms. Care is therefore needed to ensure that appropriate media and procedures are used.

*Preferred MPN tables*

When the number of bacteria present in a substrate are at a low concentration, users may wish to enumerate using

TABLE 2

Methods for the recording of data values

| Datum type | Units | Significant figures | Examples |
|---|---|---|---|
| 1. Non-microbological values | | | |
| Temperature | °C | 2 (3 if > 100 °C) | 37 °C, 4.5 °C |
| pH | — | 2 (3 if > 10) | 4.9, 10.1 |
| $a_w$ | — | 2 | 0.97 |
| Preservatives | mg $kg^{-1}$ or mg $L^{-1}$ [a] | 3 | 125, 1.27 |
| Atmosphere | atm absolute | 2 (3 if over pressured used) | 0.75 atm $CO_2$, 0.25 atm $N_2$ |
| 2. Microbiological values | | | |
| Plate counts | CFU per g or ml[b] | 2 plus exponent | 1.1 E2 |
| Optical density[c] | untransformed absorbance units | 4 | 1.317 |

[a] Where relevant, this is the concentration in the aqueous phase.
[b] Count values for microorganisms should not be transformed i.e. should be CFU per ml or g, but not $\log_{10}$ CFU per ml or g.
[c] The wavelength (nm), substrate for the blank, volume of media used and calibration graph are also required.

MPN procedures. To aid consistency, preferred tables for 3-tube and 5-tube MPN analyses are supplied to potential contributors.

Alternative MPN tables (e.g. asymmetric) may be used, but must then be submitted with the data sets.

*Preferred methods for recording data values*

It is a requirement of the UKPFMP that data from the main contributors are supplied in a common spreadsheet format. To aid flexibility, data may be submitted using a wide variety of spreadsheet packages conforming to the format.

The first section of the spreadsheet is for the registration of the experiment. This provides brief details about the laboratory producing the data, the microorganism(s) used, substrate used and the enumeration method. A reference to an associated text file will also be given. The text file contains fuller details about the work, why the particular matrix was chosen and explains the 'flags' used.

Section two records the details of the experimental matrix and the response variables. The experimental matrix (or target values) is the combination of analyses the experimeter wished to estimate (e.g. pH 6.0, 15 °C, 1% salt). Each combination of factors is called an analysis and is given a unique analysis number. This number is then used elsewhere in the spreadsheet. For each analysis, the time is also recorded. This will permit the experimenter to record when conditions were deliberately altered during an experiment (e.g. temperature of incubation changed). The response variables (or experiment values) are the actual measured values for each combination (e.g. pH 5.9, 15.2 °C and 1.05% salt). These are the conditions the microorganism experienced and are the values that should be modeled. As before, the system is flexible enough to record how a value changed with time (e.g. a deliberate or unintentional change in pH during storage).

The third section contains the counts, or other measure of microbial behavior, that were observed. For each set of values, the time of analysis, count value and 'flag' are recorded. When microbial counts are obtained the value recorded should not be transformed e.g. to $\log_{10}$ CFU per ml or g. By storing such data, any transformation required may be done by the modeler. With data obtained using optical density or impedance methods, the calibration data are also stored.

It is essential in a collaborative trial that all the stored data be recorded in a consistent and standardized format. Failure to do this may cause confusion or prevent the successful amalgamation of data sets from separate laboratories. The methods of recording data within the UKPFMP are shown in Table 2.

The documents discussed in this paper address the methods used in the UK to ensure a successful collaborative project between a number of laboratories. Adherence to these, has permitted data sets from different laboratories to be combined and modeled together. The wider adoption of such procedures would be extremely useful and potentially permit much greater collaboration in predictive microbiology throughout the world, as data could be easily and rapidly exchanged.

## REFERENCES

1 Anon. 1991. Methods for microbiological examination of food and animal feeding stuffs. Part 1: Enumeration of microorganisms. International Standards Organization 4833: 1991.

2 Anon. 1992. Food and Drug Administration Bacteriological Analytical Manual, 7th Edn. AOAC International, Arlington VA.