Green, Martin J. and Medley, Graham F. and Browne, William J. (2009) Use of posterior predictive assessments to evaluate model fit in multilevel logistic regression. Veterinary Research, 40 (4). Article 30. ISSN 0928-4249

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/1273/1/Posterior_prediction_paper_in_Vet_Res.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see:
http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
Use of posterior predictive assessments to evaluate model fit in multilevel logistic regression

Martin J. Green1,2*, Graham F. Medley3, William J. Browne4

1 School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, LE12 5RD, United Kingdom
2 School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, United Kingdom
3 Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom
4 Department of Clinical Veterinary Science, University of Bristol, Langford House, Langford, Bristol, BS40 5DT, United Kingdom

(Received 5 November 2008; accepted 24 March 2009)

Abstract – Assessing the fit of a model is an important final step in any statistical analysis, but this is not straightforward when complex discrete response models are used. Cross validation and posterior predictions have been suggested as methods to aid model criticism. In this paper a comparison is made between four methods of model predictive assessment in the context of a three level logistic regression model for clinical mastitis in dairy cattle; cross validation, a prediction using the full posterior predictive distribution and two ‘mixed’ predictive methods that incorporate higher level random effects simulated from the underlying model distribution. Cross validation is considered a gold standard method but is computationally intensive and thus a comparison is made between posterior predictive assessments and cross validation. The analyses revealed that mixed prediction methods produced results close to cross validation whilst the full posterior predictive assessment gave predictions that were over-optimistic (closer to the observed disease rates) compared with cross validation. A mixed prediction method that simulated random effects from both higher levels was best at identifying the outlying level two (farm-year) units of interest. It is concluded that this mixed prediction method, simulating random effects from both higher levels, is straightforward and may be of value in model criticism of multilevel logistic regression, a technique commonly used for animal health data with a hierarchical structure.

model fit / posterior predictive assessment / mixed predictive assessment / cross validation / Bayesian multilevel model

1. INTRODUCTION

Random effect statistical models are being increasingly used in veterinary sciences within both frequentist and Bayesian frameworks. Models are commonly specified with a binary outcome to represent, for example, ‘diseased’ or ‘non-diseased’ states and therefore take the form of multilevel logistic regression [5]. An important element of constructing and finalising a statistical model is to critically assess the fit and performance of the model [8]. However, model checking with discrete data regressions is problematic because usual methods, such as residual plots, have complicated reference distributions that depend on the parameters in the model [7, 4]. Thus, these traditional methods...
are considered to be of limited value in discrete
outcome, random effects models [2]. It may be
because of this that, in the applied literature,
particularly when complex discrete response
models are specified, attention to model fit is
often cursory.

In this research, a recently reported method
of mixed predictive model assessment [10] is
examined and illustrated in the context of an
example from veterinary epidemiology. The
concept is extended from the two level Poisson
regression originally reported, to a three logistic
regression setting with the focus of interest on
prediction of bovine clinical mastitis on dairy
farms in a specific year [6].

Posterior prediction is a general term used
when data are generated under a proposed
model, often so that comparisons can be made
between specific features of the observed and
generated data [3]. The approach provides a
useful means for model assessment and cross
validatory posterior predictive distributions are
generally considered a ‘gold standard’ [10,
13]. Using cross validation, the data are parti-
tioned ‘k’ times into subsets and an analysis is
initially performed on the ‘training’ subset.
The other ‘testing’ subset(s) are retained to val-
itate the initial analysis by making predictions
from the data. Data predictions are compared
with the observed data. The procedure is
repeated k times and k may equal the total num-
ber of data points in the dataset or may repre-
sent groups of data within the full set. An
important element of cross validation is that
predictions made on each subset of testing data
are independent of the observed outcome for
that subset. The comparisons are used to iden-
tify discrepancies between model and data.

There is an important difference between
conventional residual analysis and cross valida-
tion as a means of assessing outlying data
regions in the context of model assessment. In
conventional residual analysis, all data points
are included in the model fit and thus will have
a direct effect on model parameters and fitted
values, and hence the difference between
observed and fitted values. This is not the case
with cross validation when the data points or
groups have no influence at all on their cross
validatory predicted values, because they are
omitted during estimation, and in this respect,
classical residual plots are likely to be over-
optimistic in the assessment of model fit (i.e.
they may not identify all of the true outlying
regions) compared with cross validation. Outly-
ing units from cross validation are those for
which the other units do not provide sufficient
information for the model to fit; outliers from
residual analysis are those for which their
own influence is insufficient to provide a fit.
Therefore, regions of poor fit identified by cross
validation will not necessarily be identified by
d residual analysis indicating the importance of
the former method.

A significant disadvantage of cross valida-
tion is that it is computationally intensive and
thus time consuming. A model has to be re-
estimated for each of k subsets and this may
include hundreds or thousands of data points
or regions. If Markov chain Monte Carlo
(MCMC) procedures are being used (as has
been recommended for random effects logistic
regression models [1]), and particularly with
large data sets, the timescale required means
that cross-validation may often become imprac-
tical (depending on the choice of k).

Alternative methods to cross-validatory pre-
dictions have been suggested that have the
advantage of being more straightforward to
compute and less computationally intensive.
Gelman et al. [3] proposed use of the full model
predictive distribution to make predictions on
any required aspect of the data. This method
may be over-optimistic in the context of model
checking (i.e. it may fail to identify true outly-
ing regions) compared to cross-validation
because, as for residual analysis, the prediction
of any data region tends to be strongly influ-
enced by the equivalent observed data for the
region. Marshall and Spiegelhalter [10] pro-
posed a method termed the ‘mixed’ predictive
check which they have illustrated in the context
of disease mapping, and which appeared to per-
form in a similar manner to cross validation.
The mixed predictive check incorporates simu-
lated random effects, generated from their
underlying distribution which is characterised
from fitting the initial model, rather than the
random effects estimated directly from the data.
Use of the mixed predictive distribution has
also been reported in the context of differential
gene expression [9]. In that study, mixed pre-
dictive Markov chain P values were used to
evaluate hierarchical models [3, 10] but compa-
risons were not made between different meth-
ods of posterior predictions as a means to assess
model fit. In this context, Markov chain P val-
ues are an indicator of the probability that a pre-
dicted data region is numerically higher
(or lower) than the observed equivalent. If the
probability is high (typically greater than 95% or
97.5%) or low (typically less than 5% or 2.5%) then it suggests that the model is per-
forming poorly in the data region.

The purpose of this paper is to illustrate and
close four methods of model predictive
assessment in the context of a multilevel logis-
tic regression model, in which the specific clin-
ic interest was the prediction of disease in a
higher level unit (in this example a farm-year).
The methods are cross validation, a full posteri-
or predictive assessment and two mixed predic-
tive methods based on the approach proposed
by Marshall and Spiegelhalter [10]. An exten-
sion to the concept of the mixed prediction is
described that is generalisable to three level
hierarchical models.

2. MATERIALS AND METHODS

2.1. The data and initial model

The data for this analysis comprises clinical mas-
titis and farm management information from fifty two
commercial dairy herds, located throughout England
and Wales, with a mean herd size of approximately
150 cows and has been described in detail previously
[6]. Data were collected over a two year period. The
aim of the original research was to investigate the
influence of cow characteristics, farm facilities and
herd management strategies during the dry period,
on the rate of clinical mastitis after calving. Interest
was focussed on identifying determinants for clinical
mastitis occurrence and to assess the extent to which
these determinants could be used to predict the occur-
rence of clinical mastitis in each year on each farm.
The response variable was at the cow level; a cow
either got a case of clinical mastitis ( = 1) or not
( = 0) within 30 days of calving and a cow could be
at risk in both years of the study. Predictor variables
were included at the cow, year and farm levels. The
model hierarchical structure was cows within farm-
years within farms, and can be summarised as:

\[ CM_{ijk} \sim \text{Bernoulli}(\pi_{ijk}) \]

\[ \text{Logit}(\pi_{ijk}) = \beta_0 + \beta_1 X^{(1)}_{ijk} + \beta_2 X^{(2)}_{ijk} + \beta_3 X^{(3)}_{k} + u_{jk} + v_{ik} + v_{1k} P_{ijk} \]

\[ u_{jk} \sim N(0, \sigma_u^2), v_{ik} = \begin{pmatrix} v_{ik} \\ v_{1k} \end{pmatrix} \sim \text{MVN}(0, \Omega) \]

(1)

where the subscripts i, j and k denote the three
model levels, \( \pi_{ijk} \) the fitted probability of clinical
mastitis (CM) for cow i in year j on farm k, \( \beta_0 \)
the regression intercept, \( X^{(1)}_{ijk} \) the vector of covari-
ates at cow level, \( \beta_1 \) the coefficients for covariates
\( X^{(1)}_{ijk}, X^{(2)}_{ijk} \) the vector of farm-year level covariates,
\( \beta_2 \) the coefficients for covariates \( X^{(1)}_{jk}, X^{(2)}_{jk} \) the vec-
tor of farm level covariates, \( \beta_3 \) the coefficients for
covariates \( X^{(3)}_{k} \), \( P_{ijk} \) is a covariate (within \( X^{(3)}_{jk} \)) that
identifies cows of parity one (after first calf), \( u_{jk} \) is a
random effect to reflect residual variation between
years within farms, and \( v_{ik} \) and \( v_{1k} \) are random
effects to reflect residual variation between farms,
and for the difference in rates for parity 1 cows
between farms respectively.

Model selection was made from a rich dataset of
more than 350 covariates. Model building has been
described in detail previously [6] but briefly pro-
cedeed as follows. Each of the covariates was exam-
ined individually, within the specified model
framework, to investigate individual associations with
clinical mastitis whilst accounting for the data struc-
ture. Initial covariate assessment was carried out using
penalised quasi-likelihood for parameter estimation
(MLwiN, [11]) and final models were selected using
MCMC for parameter estimation in WinBUGS [12].
A burn-in of at least 2 000 iterations was used for
all MCMC runs during which time model conver-
gence had occurred. Parameter estimates were based
on a further 8 000 iterations. The final model included
the following predictor variables; cow parity, cow his-
toric infection status, whether the farm maintained a
cow standing time of 30 min after administration of
treatments at drying off (the end of the previous lacta-
tion), whether farms reduced the milk yield of high
yielding cows before drying off, whether cow bedding
was disinfected during the early dry period, type of
cow bedding during the late dry period, the time peri-
d between sequential cleaning out of the calving
pens, and the time between calving and the cows
being first milked after calving.
2.2. Predictive assessments

Of particular clinical interest in the research was the prediction of the incidence rate of clinical mastitis (number of cases per cow at risk) for each of the $j = 1...103$ farm-years and thus the predictions of these rates were used to investigate methods of model assessment. Four methods of predictive assessment were compared; cross validation, a full posterior predictive check and two ‘mixed’ predictive assessments similar to that suggested by Marshall and Spiegelhalter [10].

After final model selection, each method of prediction was incorporated into the MCMC process. At each iteration after model convergence, a prediction was made for the occurrence of mastitis for each individual cow ($y_{ijk}$) by drawing from the appropriate conditional probability distribution (see below). Similarly, at each iteration, the number of predicted cases of clinical mastitis were summed over all cows in each farm-year and divided by the total cows at risk in each farm-year, to provide a Monte Carlo estimate of the farm-year incidence rate of clinical mastitis. Predictions were made from 8000 MCMC iterations after model convergence.

To describe the four methods of predictive assessment, we condense the model terms, such that the disease status for each cow ($y_{ijk}$) is conditional on a set of model fixed effect parameters $\beta$, covariates ($X_{ijk}$), and random effects $v_k$ and $u_{jk}$:

$$y_{ijk} \sim p(y_{ijk} | \beta, X_{ijk}, V_k, U_{jk})$$

The random effects have parameters represented by $\sigma_u^2$, and $\Omega_v$:

$$U_{jk} \sim p(U_{jk} | \sigma_u^2)$$
$$V_k \sim p(V_k | \Omega_v)$$

The four methods of predictive assessment employed were:

A. Cross validation (“xval”). Each of the 103 farm-years was removed from the analysis in turn and the model fitted to a reduced data set excluding the $j$th farm-year (denoted $(-j)^k$), from which new model parameters were estimated ($\beta(-j)^k$, $\gamma(-j)^k$, $u(-j)^k$, $\sigma_u^2(-j)^k$, $\Omega(-j)^k$). A replicate observation for the omitted data, $y_{ijk}^{xval}$ was simulated from the conditional distribution:

$$y_{ijk}^{xval} \sim p(y_{ijk}^{xval} | \beta(-j)^k, X_{ijk}, U_{jk}^{xval}, V_k^{xval})$$
$$u_{jk}^{xval} \sim p(u_{jk}^{xval} | \sigma_u^2(-j)^k)$$
$$v_k^{xval} \sim p(v_k^{xval} | \Omega_v(-j)^k)$$

B. Posterior predictive assessment from the full data (“full”). The predictive distribution was conditional on all fixed effect and random effect parameters estimated in the final model and a replicate observation $y_{ijk}^{full}$ generated from the conditional distribution:

$$y_{ijk}^{full} \sim p(y_{ijk}^{full} | \beta, X_{ijk}, V_k, U_{jk})$$

C. Mixed prediction 1 (“mix1”). This predictive distribution was conditional on the fixed effect parameters and the random effect distributions from which new random effects, $u_{jk}^{mix1}$ and $v_k^{mix1}$, were simulated to make the prediction. Thus a replicate observation $y_{ijk}^{mix1}$ was generated from the conditional distribution:

$$y_{ijk}^{mix1} \sim p(y_{ijk}^{mix1} | \beta, X_{ijk}, U_{jk}^{mix1}, V_k^{mix1})$$
$$u_{jk}^{mix1} \sim p(u_{jk}^{mix1} | \sigma_u^2)$$
$$v_k^{mix1} \sim p(v_k^{mix1} | \Omega_v)$$

D. Mixed prediction 2 (“mix2”). This predictive distribution was conditional on the fixed effect parameters, the random effects distribution at level 2, (from which new random effects, $u_{jk}^{mix2}$ were simulated), and the level 3 random effects from the model, $v_k$. Thus a replicate observation $y_{ijk}^{mix2}$ was simulated from the conditional distribution:

$$y_{ijk}^{mix2} \sim p(y_{ijk}^{mix2} | \beta, X_{ijk}, U_{jk}^{mix2}, V_k^{mix2})$$
$$u_{jk}^{mix2} \sim p(u_{jk}^{mix2} | \sigma_u^2)$$

2.3. Comparisons between methods of predictive assessments

In each case, predictions of farm-year incidence rates of clinical mastitis were compared with observed rates. Predictions from cross validation (taken as a gold standard) were also compared to the other methods of prediction to assess which best mimicked this procedure. To assess the degree of discrepancy between observed and predicted farm-year incidence rate of mastitis, the predicted distributions, $y_{jk}^{pred}$ were compared to the observed values using Monte Carlo predictive $P$ values. At each iteration of the MCMC procedure, an indicator variable was set to 1 when $y_{jk}^{pred} > y_{jk}$, to 0.5 if $y_{jk}^{pred} = y_{jk}$ and to 0 if $y_{jk}^{pred} < y_{jk}$; the Monte Carlo $P$ value was estimated as the mean of this indicator variable.
Therefore predictive \( P \) values > 0.975 or < 0.025 indicated that the probability of the observed incidence rate of clinical mastitis being within the predicted distribution was less than 5% and represented a relatively extreme result.

3. RESULTS

Figure 1 (A–D) illustrates the mean predicted incidence rate of clinical mastitis for each method of posterior prediction, plotted against the observed incidence of clinical mastitis. The graphs illustrate that the full posterior predictive method most closely resembled the observed data and cross validation and the “mix1” method displayed considerably more variability. The “mix2” method provided an intermediate result. Figure 2 illustrates the comparison between mixed and full predictive methods and cross validation. Both mixed predictive methods yielded better estimates of the cross validatory prediction than the full posterior predictive method, and the “mix2” method produced estimates most similar to cross validation.

The median error for each predictive method was calculated as the median of the unsigned differences between predicted and cross validatory farm-year incidence rates of clinical mastitis, as
a percentage of the cross validatory farm-year incidence rate of clinical mastitis. The median errors were 13.7%, 11.5% and 9.4% for the full posterior prediction, the mixed prediction 1, and for mixed prediction 2 respectively.

Figure 2. Plots of cross validatory predictions of farm-year clinical mastitis incidence against full and mixed predictive methods of farm-year clinical mastitis incidence (cases per cow at risk per year).

Figure 3. Comparison of MCMC $P$ values from cross validation (for values > 0.80 and < 0.20) and from different methods of predictive assessment for farm-year incidence of clinical mastitis.
identified with cross validation, these being the most divergent regions eligible for identification and further investigation. At large and small $P$ values ($P < 0.20$ or $> 0.80$) the mixed predictive methods performed more similarly to cross validation than the full posterior prediction with the “mix1” method most closely representing cross validatory MCMC $P$ values. This is confirmed in Table I that provide the sensitivity and specificity for each predictive method, taking cross validation MCMC $P$ values as the “gold standard”, and different $P$ value thresholds. The “mix1” method had the highest sensitivity indicating that this method identified the largest proportion of “true” extreme values as determined by cross validation. The “mix1” method identified 82.4% (14 out of 17) of extreme values when a threshold of $< 0.10$ or $> 0.90$ was used and 60% (3 out of 5) of extreme values with a threshold set at $< 0.025$ or $> 0.975$.

The computing times to complete 10 000 iterations (using an Intel Centrino 2.0 GHz Processor, 1.5GB RAM) for 103 cross validation predictions and the “mix1” method were 334

| Table I. Sensitivity and specificity of MCMC $P$ values for each prediction method (full = full posterior predictive method, mix 1 and mix 2 = mixed predictive methods 1 and 2 respectively) compared to MCMC $P$ values for cross validation, at different $P$ value thresholds (as specified). |
|---------------------------------------------|------------------|------------------|------------------|
| $P$ value > 0.90 or < 0.10               | Cross validation | Total            | Sens (%)         | Spec (%)         |
|                                           | 0                | 1                |                  |                  |
| full                                       | 0                | 86               | 14               | 100              | 17.6            | 100.0 |
|                                           | 1                | 0                | 3                | 3                |
| Total                                      | 86               | 17               | 103              |                  |                  |
| mix 1                                      | 0                | 84               | 3                | 87               | 82.4            | 97.7   |
|                                           | 1                | 2                | 14               | 16               |
| Total                                      | 86               | 17               | 103              |                  |                  |
| mix 2                                      | 0                | 86               | 10               | 96               | 41.2            | 100.0  |
|                                           | 1                | 0                | 7                | 7                |
| Total                                      | 86               | 17               | 103              |                  |                  |
| $P$ value > 0.95 or < 0.05                 | Cross validation | Total            | Sens (%)         | Spec (%)         |
|                                           | 0                | 1                |                  |                  |
| full                                       | 0                | 93               | 8                | 101              | 20.0            | 100.0  |
|                                           | 1                | 0                | 2                | 2                |
| Total                                      | 93               | 10               | 103              |                  |                  |
| mix 1                                      | 0                | 90               | 5                | 95               | 50.0            | 96.8   |
|                                           | 1                | 3                | 5                | 8                |
| Total                                      | 93               | 10               | 103              |                  |                  |
| mix 2                                      | 0                | 93               | 7                | 100              | 30.0            | 100.0  |
|                                           | 1                | 0                | 3                | 3                |
| Total                                      | 93               | 10               | 103              |                  |                  |
| $P$ value > 0.975 or < 0.025               | Cross validation | Total            | Sens (%)         | Spec (%)         |
|                                           | 0                | 1                |                  |                  |
| full                                       | 0                | 98               | 5                | 103              | 0.0             | 100.0  |
|                                           | 1                | 0                | 0                | 0                |
| Total                                      | 98               | 5                | 103              |                  |                  |
| mix 1                                      | 0                | 98               | 2                | 100              | 60.0            | 100.0  |
|                                           | 1                | 0                | 3                | 3                |
| Total                                      | 98               | 5                | 100              |                  |                  |
| mix 2                                      | 0                | 98               | 4                | 102              | 20.0            | 100.0  |
|                                           | 1                | 0                | 1                | 1                |
| Total                                      | 98               | 5                | 103              |                  |                  |
h and 3.6 h respectively. This did not include the
time required to format the data and set up each
model and this took approximately the same
time per model. Thus it took approximately
103 times longer for the cross validatory predic-
tions than the “mix1” method.

4. DISCUSSION

Identifying divergent data regions in statisti-
cal modelling is important for two reasons. Firstly,
numerous divergent regions could indi-
cate that underlying statistical assumptions are
incorrect, for example the model does not cap-
ture the true data structure. Secondly, individual
divergent units could represent those that are
fundamentally different from other units in the
dataset after accounting for predictor variables,
and the possible absence of unknown but
important explanatory covariates. In either case,
further investigations would be warranted.

Cross validation provides a useful method of
accurately identifying divergent units in com-
plex statistical models, but faster methods
would be of practical value in model assess-
ment and it was for this reason that the alterna-
tive strategies were investigated in this research.

The predictions of clinical mastitis incidence
rates obtained from the different methods show
clear differences in results obtained, as shown
in Figure 1. The full predictive method pro-
vided predicted incidence rates of clinical mas-
titis that most closely resembled the observed
incidence rates, but these appeared to be over-
optimistic in terms of model performance in
comparison to cross validatory predictions. This
is not surprising since the random effects from
the initial model are directly incorporated into
the prediction steps but it does highlight the dif-
fERENCE between this method and cross
validation.

For the three level logistic regression models
in this example, the mixed predictive methods
provided a better approximation to cross-
validation than the full posterior predictive
assessment. This is concordant with the first
study that used a mixed prediction for approxi-
mating cross validation in a two level Poisson
model for disease mapping [10]. In the current
study using a three level logistic regression
model, the “mix2” method provided the closest
overall approximation to cross validatory pre-
dictions of farm-year incidence of clinical
mastitis. However, the “mix1” method per-
formed best for the more extreme outlying val-
ues identified by cross validation and thus this
method was more useful for identifying the
most divergent higher level units in these data.
The mixed predictive methods look promising
as a means of practical model assessment for
the relatively common statistical approach of
multilevel logistic regression and as such, war-
rant further investigations.

Importantly, the mixed predictive methods
take considerably less time to implement
(in this example approximately one hundredth
of the time of cross validation) and therefore pro-
vide a clear advantage in terms of practical use.
The “mix2” method is essentially a compromise
between the “mix1” method and a full posterior
prediction. The method simulates a new random
effect at level 2 but uses the estimated random
effects from the model at level 3. In the current
example there were only two level 2 units for
each level 3 unit and it may be that if more level
two units existed for each level 3 units, mixed
prediction method 2 would tend to become sim-
ilar to mixed method 1 (the higher level unit hav-
ing less influence on the predicted data). Similary, the relative performance of the two
mixed predictive methods may depend on the
relative sizes of the higher level variances and
more research into the importance of the relative
size of higher level variances when using mixed
predictive methods would be beneficial. In this
example the variance at level two (farm-year)
was 0.06 and at level three (farm) was 0.10
(for cows greater than parity one) and 0.64 (for
cows of parity one). If the level three variances
had been very small in comparison to the level
2 variance, it is possible that both mixed predic-
tive methods used in this study would have
yielded similar results. Further investigations
of mixed predictive methods using different
types of models, numbers of levels, units per
level and relative sizes of higher unit variances
would be worthwhile.

From our results, it would appear that, out of
the methods examined, the “mix1” method is

Page 8 of 9 (page number not for citation purpose)
likely to provide the closest representation of cross validation for potentially divergent data regions in multilevel logistic regression. However, it is important to note that these results apply only to one dataset and whilst in agreement with a previous study [10], need to be viewed with this perspective. It may be possible to generalise this approach to logistic regression and other multilevel models, but more research in this area is required.

Our results indicate that whilst mixed predictions provide a reasonable approximation to cross validation, they do not provide precise replication of the results. Therefore, a pragmatic approach for implementation of mixed predictive assessments may be for an initial highlighting of possible divergent data regions on which to undertake further model checking using cross validation. Thus, instead of undertaking cross validation on all possible regions an intermediate step could be to first use a mixed prediction approach and then to use cross validation for data regions that are potentially divergent based on the mixed prediction. A reduced mixed prediction MCMC $P$ value threshold could be used to improve the likelihood that all ‘true’ outliers are identified, possibly the central 80 percentile region and cross validation then carried out on regions that fall outside this interval. This would increase the sensitivity of identifying “true” divergent regions using the mixed methods but would reduce the computing time required compared to using cross validation for all regions.

Assessment of model performance is important and problematic particularly when large datasets and complex model structures are used. Posterior predictions are recognised as a useful method to investigate model fit and more research on mixed posterior predictions may be useful to facilitate straightforward, fast assessments for these types of model.

Acknowledgements. Martin Green is funded by a Wellcome Trust Intermediate Clinical Fellowship.

REFERENCES


