Bayesian Modelling of Resource Use Data Collected Alongside Randomised Controlled Trials[§]

Andrea Manca[1,**], Stefano Conti[1], Paul Lambert[2], Keith Abrams[2]

1. Centre for Health Economics, University of York, UK
2. Centre for Biostatistics & Genetic Epidemiology, Department of Health Sciences, School of Medicine, University of Leicester, Leicester, LE1 6TP, UK

Abstract

Aims to explore and contrast alternative multivariate modelling strategies for individual patient-level resource use data collected alongside randomised controlled trials (RCTs).
Methods Bayesian multivariate statistical models estimated using Markov chain Monte Carlo techniques. This paper explores the use of more realistic distributional assumptions to represent discrete and continuous resource use data, other than relying on Normal approximations. Formal model criticism and comparison are implemented on the grounds of predictive performance and goodness-of-fit measures.
Data individual patient-level data (N=3164) collected alongside a randomised clinical trial comparing low- versus high-dose ACE-inhibitor lisinopril in chronic heart failure.
Results multivariate statistical modelling explicitly accounts for the nature of, as well as the relationships between, various resource use components, thereby providing more accurate insights into the main cost drivers. Replacement of inappropriate Normal approximations with more meaningful distributional assumptions seems to lead to improved predictive accuracy, albeit at the cost of reduced model simplicity and interpretation.
Conclusions Bayesian multivariate modelling lends itself as a flexible and sound strategy for the analysis of resource use and cost data in RCTs. Overall the reliability of inferences seems to benefit from assigning realistic distributions to the quantities involved in a cost-effectiveness analysis.

Keywords: Bayesian statistics; cost-effectiveness analysis; multivariate modelling; RCTs

Acknowledgments
A. Manca is recipient of a Wellcome Trust funded post-doctoral Training Fellowship in Health Services Research (grant number GR071304MA). The views and opinions expressed therein are the authors' and do not necessarily reflect those of the funding institutions.

[§] Paper presented at the *69th Health Economists' Study Group Meeting*, The University of York, York, UK, 26-28 July 2006
[**] Corresponding author at: Centre for Health Economics, Alcuin 'A' Block, The University of York, Heslington, York, YO10 5DD, UK. Email: am126@york.ac.uk

Figure Captions

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

List of Tables

Table 1:

1 Introduction

Cost-effectiveness analysis (CEA) is an established input to support healthcare decision-making in the UK [1]. Most of the clinical and cost-effectiveness evidence used (either directly or indirectly) to inform healthcare resource allocation decisions comes primarily from individual patient-level data (IPD) collected alongside randomised controlled clinical trials (RCTs). The analysis of IPD is therefore central to the methodological development of healthcare economic evaluation techniques.

In the cost-effectiveness equation, the analysis of costs typically focuses on the estimation of the *treatment effect* on total costs (i.e. difference in mean costs), only rarely examining the impact of the intervention on the use of the various healthcare resources required in the management of patients enrolled in the trial. Analyses quantifying the treatment effect on resource use consumption are even less frequently encountered. However, it would seem reasonable to take advantage of this wealth of refined information to disentangle total costs into their relevant drivers and identify the dynamics of underlying related components.

In order to account for the relationships featured by resource uses components recorded in the trial this paper proposes setting the analysis within a multivariate statistical modelling framework. It is argued that the proposed strategy would allow to structure, analyse and interpret the information provided by IPD more efficiently and transparently than with a more standard total cost analysis.

The main purpose of the paper is to investigate the pros and cons of jointly modelling IPD on resource use collected alongside RCTs by means of meaningful statistical distributions. This is attained by specifying multivariate, not necessarily Normal, models for resource use components rather than by modelling aggregate costs (as typically done in CEA). Different modelling assumptions are formulated to appropriately reflect the intrinsic nature (i.e. continuous or discrete) of the examined data. As to the distributions considered in this paper, the choice is essentially driven by common sense. A more formal discussion on methods and criteria for model selection falls outside the remit of the paper and can be found in Kadane and Lazare [2]. The reviewed models are compared and contrasted with the conventional multivariate Normal specification, in order to assess the consequences of applying simpler (and sometimes inappropriate) assumptions. In the context of resource use components analysis, it is shown that the Normal model may lead to inadequate inferences even in the presence of large sample sizes. This is particularly so when dealing with data like resource use IPD, which often exhibit excess zeros and strong skewness, model mis-specifications may lead to poor fit and reliability. Candidate models are fitted and critically assessed within a Bayesian perspective, which offers a suite of powerful and flexible modelling tools within a sound and coherent methodological framework.

The paper is organised as follows. Section 2 outlines the rationale for the paper, introduces and sets up the proposed statistical modelling framework, and discusses its advantages and deficiencies for alternative distributional choices on individual components. Section 3 illustrates the model fitting and comparison stages on practical grounds, with reference to a real-world example. Concluding remarks are summarised in Section 4.

1 Modelling resource uses

The main objective of a RCT-based CEA is broadly to estimate differential mean costs and effects between the treatment arms of the study, together with relevant measures of sampling variation. In its most standard formulation (i.e. no missing or censored observations) the building block of the cost analysis consists of quantifying the monetary value of each resource deployed during the study period, followed by aggregating the costs of the different healthcare resources used by the patient.

The conceptual simplicity and practical appeal of this approach explain its widespread popularity. However such methodology suffers from a potentially significant drawback: it neglects the intrinsic nature of count and categorical resource components (e.g. hospital in- and out-patients,

GP visits). Corresponding statistical models may fall short of fully capturing the linkage between the different resource components relating to individual patients, thereby leading to potential inaccuracies in the analysis. The simplest approach to directly model healthcare resources would be to rely upon Normal distributional assumptions, which is typically justified by the Central Limit Theorem, according to which means from sufficiently large (typically 50 or more) samples will approximately be Normally distributed. While economic evaluation is indeed concerned with population means, their sample counterparts are known to be susceptible to features like large frequencies at zero and highly skew distributions, as is often the case with count and cost data [3, 4]. This paper therefore argues in favour of carefully modelling the underlying distributions of patient-level healthcare resource data, rather than simply relying upon Normal-based approximations.

2.1 A General Framework

Let R_{ijt} indicate direct measurements on resource use $j = 1,...., p$ for each patient $i = 1,...., n_t$ allocated to either the treatment (t=1) or control (t=0) arm of the trial. These can refer for instance to total day cases, in-patient stays in hospital (measured in days), out-patient visits (measured in number of visits) and corresponding costs. As already discussed in the introduction, much of the work in RCT-based CEA is centred on modelling the distribution of patient level total cost, $C_{it} = \sum_{j=1}^{p} c_{jt}$. Here interest instead lies in directly examining the distribution of individual resource uses R_{ijt} at patient level, the cost of which being simply the product between the resource use and its unit cost. Within the framework proposed in the paper, records of any resource use j arise from some population model whose distribution under treatment t is characterised by some unknown parameter(s) θ_{it}.

As explained in section 1, the aim here is to account for the relationships between the examined resource uses and their aggregate costs by fitting them to multivariate, rather than independent, distributions. When dealing with both continuous and discrete data, as it is often the case in CEA, full-scale joint models usually turn out to be intractable and difficult to interpret whereas specification of the corresponding sequence of conditional distributions offers more viable and useful insights [5-7].

Although the conditioning order in principle is irrelevant‡, it is convenient for practical purposes to draw attention on the full conditional distribution of the aggregate costs given all resource use components, which is denoted by $Dist(\cdot|\theta_t)$. Indicating with $Dist_{k|1,.....,k-1}$ the conditional model for resource use R_{kt} given $R_{1t},....., R_{(k-1)t}$, the generic model formulation

$$\begin{aligned} &R_{i1t} \sim Dist_1(\cdot|\theta_{1t}) \\ &R_{i2t}|R_{i1t} \sim Dist_{2|1}(\cdot|\theta_{1t},\theta_{2t}) \\ &\text{M} \\ &R_{ipt}|R_{i1t},....,R_{i(p-1)t} \sim Dist_{p|1,...,(p-1)}(\cdot|\theta_{1t},.....,\theta_{pt}) \end{aligned} \qquad (1)$$

thus is obtained.

2.2 Model Estimation

‡ The distribution of any pair of random variables X and Y is equivalently obtained as Pr(X,Y)=Pr(X) Pr(Y|X) or Pr(X,Y)=Pr(Y) Pr(X| Y).

Model (1) outlines in a very general fashion the dependency relations between the various resource use components at patient-level. It obviously features a greater level of complexity (as well as of detail) than the conventional univariate Normal specifications $C_{it} \sim N(\mu_t, \sigma_t^2)$ for describing aggregate cost data.

As to the problem of which parametric distributions in (1) actually fit to IPD, the choice appears to be mostly driven by empirical considerations and analytical tractability. Popular choices in the health economics literature comprise the Poisson (henceforth Poi) and Negative Binomial (NBin) models for count data, together with their hurdle (e.g. HPoi) or zero-inflated extensions to accommodate for over-dispersion [8, 9]. The Normal (N), LogNormal (LN) and Gamma (G) distributions instead are commonly used to represent costs [5, 6, 10, 11].

Regardless of which distributions are specified in (1), Bayesian simulation-based methods like Markov chain Monte Carlo (McMC) offer a feasible and convenient solution to the problem of estimating the indexing parameters θ_{it} and θ_t. Once parameters have been inferred and model (1) has been fitted to the data, its predictive accuracy can be assessed through various goodness-of-fit measures, again readily available at little computing cost from the Bayesian toolbox. Careful consideration of such diagnostic checks may provide valuable insights as to whether the proposed model validates reasonably well, or revision of some key features are required to obtain a reliable statistical model.

3 The ATLAS trial: a test bed

For the purpose of illustrating the methodology outlined in Section 2.1 various instances of model (1) are fitted to a subset of patient-level data collected alongside the Assessment of Treatment with Lisinopril and Survival (ATLAS) RCT. This multinational trial recruited 3164 patients from 19 countries, and compared low- (n_1=1596) versus high-dose (n_2=1568) ACE-inhibitor lisinopril in the management of patients with chronic heart failure. The trial collected data on four main healthcare resources: number of day cases, in-patient hospital stay, study drug use, and open label ACE-inhibitors. For simplicity the analysis here focuses on records of total day cases, total amount of days in hospital and study drug use. The latter is analysed here on its cost scale, rather than on its physical unit (e.g. milligram), but being both continuous variables the validity of the analysis remain unaffected. Consistently with notation introduced in Section 2.1, after dropping the no longer required arm allocation subscript t variables being modelled are indicated respectively by $R_{i1,}R_{i2}$, and R_{i3}.

A total of 39 cases of unrealistically large total days in hospital were discarded from the examined data set. Accordingly the results hereby reported should not be considered informative for policy making, and interested readers are invited to refer to the original reports of the clinical and CEA results [12, 13].

TABLE 1. Summary statistics of selected healthcare resources and costs from the treatment arm of the ATLAS data-set

	Day Cases	Days in Hospital	Drug Cost
Min	0	0	2.46
1st Qu	0	0	502.75
Median	0	6	1238.74
Mean	0.381	16.94	1101.43
3rd Qu	0	23	1633.88
Max	30	165	2153.93
Var	1.4053	6.5378e+2	395883

A straightforward inspection of summary statistics (Table 1) and plots (Figure 1) from the data highlights a substantial degree of positive skewness for all variables, the presence of several large

costs and a strong concentration of zero counts for both day cases (77.03% of records) and days in hospital (33.01%).

Such features are clearly responsible for over-dispersion in count data and asymmetry in costs, which represent modelling issues currently at the heart of much applied statistical literature. It unsurprising that standard distributions as the Poisson and the Normal, often misrepresent heavily skewed, over-dispersed, data such as those displayed in Figure 1.

In the paper the conditional distributions for $R_{i1,}R_{i2}$, and R_{i3} were chosen out of a number of parametric models. The objective of model selection was twofold. It primarily aimed at (a) sensibly addressing any patterns exhibited by the data-set at hand, while (b) contrasting such specifications with more tractable, albeit less defensible, alternatives. In the light of these endeavours and to what seems to emerge from the specialised literature, the NBin as well as Poi and HPoi were utilised to model the count variables R_{i1} and R_{i2}, whereas observed continuous data (e.g. study drug use) were dealt with G, N and LN distributions. Histograms of the empirical distributions in Figure 1 seem to indicate that the aforementioned may be plausible choices, especially for modelling the healthcare components.

As already mentioned in Section 2.2, no attempt is made to propose any formal criterion of model selection (extensively reviewed elsewhere [2]). Nevertheless various diagnostic checks, available off-the-shelf from the Bayesian model criticism toolbox (see Gelman *et al* [14]Chapter 6 for an extensive discussion on the topic), were implemented to assess and compare the goodness of fit of the competing models and draw conclusions about their predictive reliability.

3.1 Fitting the Models to the Data

The first combination of statistical models here explored for the data at hand is represented by a joint N-N-N. This instance of model (1) is possibly the least meaningful among those assessed, in the light of features of the ATLAS data as illustrated e.g. in Figure 1. On the other hand the Normal model still represents a cornerstone in broad statistical modelling, hence its inclusion in the analysis on the grounds of tractability and usefulness as a benchmark for alternative (possibly more meaningful) distributions.

Unlike any of the other reviewed cases, mathematical properties of the Normal distribution conveniently enable a compact joint specification

N-N-N

$$\begin{bmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \end{bmatrix} \sim N_3 \left(\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}, \begin{bmatrix} \theta_4 & \theta_7 & \theta_8 \\ \theta_7 & \theta_5 & \theta_9 \\ \theta_8 & \theta_9 & \theta_6 \end{bmatrix} \right)$$

Other examined model combinations were

Poi-Poi-LN

$$R_{i1} \sim Poi(\lambda_1)$$
$$R_{i2} \mid R_{i2} \sim Poi(\lambda_{2|1})$$
$$R_{i3} \mid R_{i1}, R_{i2} \sim LN(\mu_{3|1,2}, \sigma^2_{3|1,2})$$

where $\lambda_1 = \theta_1$ and $\lambda_{2|1} = \exp\{\theta_2 + \theta_3[R_{i1} - E(R_{i1})]\}$ denote the rates for the healthcare resource components $\mu_{3|1,2} = \theta_4 + \theta_5[R_{i1} - E(R_{i1})] + \theta_6[R_{i2} - E(R_{i2})]$ and $\sigma^2_{3|1,2} = \theta_7$ the mean and variance (on the logarithmic scale) for the mean patient level drug use.

HPoi-HPoi-LN

$$R_{i1} \sim HPoi(\lambda_1, \theta_2)$$
$$R_{i2} \mid R_{i1} \sim HPoi(\lambda_{2|1}, \theta_5)$$
$$R_{i3} \mid R_{i1}, R_{i2} \sim LN(\mu_{3|1,2}, \sigma^2_{3|1,2})$$

where $\lambda_1 = \theta_1$ and $\lambda_{2|1} = \exp\{\theta_3 + \theta_4[R_{i1} - E(R_{i1})]\}$ indicated the rates for the healthcare resource components and θ_2 and θ_5 their corresponding zero-count probabilities and $\lambda_{3|1,2} = \theta_6 + \theta_7[R_{i1} - E(R_{i1})] + \theta_8[R_{i2} - E(R_{i2})]$ and $\sigma^2_{3|1,2} = \theta_9$ the mean and variance (on the logarithmic scale) for the mean patient level drug use.

NBin-NBin-LN

$$R_{i1} \sim NBin(r_1, \frac{r_1}{r_1 + \mu_1})$$
$$R_{i2} \mid R_{i1} \sim NBin(r_{2|1}, \frac{r_{2|1}}{r_{2|1} + \mu_{2|1}})$$
$$R_{i3} \mid R_{i1}, R_{i2} \sim LN(\mu_{3|1,2}, \sigma^2_{3|1,2})$$

where $r_1 = \theta_1$ and $r_{2|1} = \theta_3$ denote the over-dispersion parameters for the healthcare resource components, $\mu_1 = \theta_2$ and $\mu_{2|1} = \exp\{\theta_4 + \theta_5[R_{i1} - E(R_{i1})]\}$ their corresponding means and $\mu_{3|1,2} = \exp\{\theta_6 + \theta_7[R_{i1} - E(R_{i1})] + \theta_8[R_{i2} - E(R_{i2})]\}$ and $\sigma^2_{3|1,2} = \theta_9$ the mean and variance (on the logarithmic scale) for the mean patient level drug use.

NBin-NBin-G

$$R_{i1} \sim NBin(r_1, \frac{r_1}{r_1 + \mu_1})$$
$$R_{i2} \mid R_{i1} \sim NBin(r_{2|1}, \frac{r_{2|1}}{r_{2|1} + \mu_{2|1}})$$
$$R_{i3} \mid R_{i1}, R_{i2} \sim G(\frac{\mu^2_{3|1,2}}{\sigma^2_{3|1,2}}, \frac{\mu_{3|1,2}}{\sigma^2_{3|1,2}})$$

where $r_1, \mu_1, r_{2|1}$, and $\mu_{2|1}$ have the same interpretation as in the case NBin-NBin-LN, and $\mu_{3|1,2} = \exp\{\theta_6 + \theta_7[R_{i1} - E(R_{i1})] + \theta_8[R_{i2} - E(R_{i2})]\}$ and $\sigma^2_{3|1,2} = \theta_9$ are the mean and variance for the mean patient level drug use.

Poi-Poi-G

$$R_{i1} \sim Poi(\lambda_1)$$
$$R_{i2} \mid R_{i2} \sim Poi(\lambda_{2|1})$$
$$R_{i3} \mid R_{i1}, R_{i2} \sim G(\frac{\mu^2_{3|1,2}}{\sigma^2_{3|1,2}}, \frac{\mu_{3|1,2}}{\sigma^2_{3|1,2}})$$

where λ_1 and $\lambda_{2|1}$ are the same as in Poi-Poi-LN, while $\lambda_{3|1,2} = \theta_6 + \theta_7[R_{i1} - E(R_{i1})] + \theta_8[R_{i2} - E(R_{i2})]$ and $\sigma^2_{3|1,2} = \theta_9$ are the mean and variance for the mean patient level drug use.

A few comments are in order as regards the parameterisation adopted for the above listed model structures. One of the main appeals of the N-N-N model lies in enabling immediate derivation of marginal distributions from their conditional counterparts (as detailed in Gelman *et al* [14]). This is a prerogative of the Normal model which unfortunately is not entertained by any of the proposed statistical models. The adopted parameterisation is aimed at expressing the above conditional means as linear functions of the corresponding conditioning variables, in analogy to what is automatically obtained under Normality assumptions. An additional complication here is represented by the examined resource use data being non-negative, which in turn requires corresponding mean values (which depend on the chosen parameterisation) to be non-negative too. By specifying the expected values to be linear on the logarithmic scale (e.g. the NBin-NBin-G case) it is argued that the resulting model structure, even if more laborious than the N-N-N instance, still retains the ability to yield meaningful and interpretable parameter estimates. Further comments as to alternative parameterisations that could be considered are deferred to the discussion section.

TABLE 2. Predictive summaries from joint models of selected healthcare resources and costs from the treatment arm of the ATLAS data-set

Model	Mean (std. dev.)		
	R_1	R_2	$\bar{C}$
ATLAS	0.381	16.9363	1101.4328
	(1.1855)	(25.5692)	(629.1923)
N-N-N	-0.0635	15.654	1102.9552
	(2.2303)	(53.1165)	(1487.2441)
Poi-Poi-LN	0.4074	20.3822	950.4481
	(0.6396)	(4.5382)	(2082.7655)
HPoi-HPoi-LN	0.6627	6.4131	263025.7
	(0.6187)	(10.365)	(19435180)
NBin-NBin-LN	0.4730	17.8932	1170.3979
	(1.1042)	(31.6392)	(3411.938)
NBin-NBin-G	0.6729	7.7986	0.0165
	(1.1537)	(59.0529)	(2.732)
Poi-Poi-G	0.4101	17.9614	701.7828
	(0.6411)	(4.2655)	(697.313)

The ultimate purpose of fitting the above listed models lies in inferring the corresponding joint predictive distributions of the resource use and cost components of interest. Standard probability results allow immediate derivation of posterior predictive inferences from the N-N-N specification. Conversely all remaining models require McMC simulation to fit the above models, adding an additional layer of computational complexity to the estimation problem. Vague, albeit proper, prior distributions were assigned to all model parameters to ensure proper posterior distributions. On the implementation side, estimation was found to be difficult especially when

dealing with models structured via Hurdle Poisson or Gamma components, due to the occurrence of numerical singularities. Even estimates for the N-N-N model structure were found to be numerically unstable, presumably due to the obviously poor fit exhibited in this case by the underlying data (Figure 2).

Once parameters have been inferred from the McMC output, corresponding modal values were utilised to finally estimate the conditional distributions defining the models in Section 2.2. Each of these was subsequently validated by means of Bayesian residual analysis [14], that is by obtaining standardised residuals from which root mean squared prediction errors (RMSPEs) and squared Mahalanobis distances (SMDs) were derived to detect model deficiencies. When the model fit is perfect RMSPEs and SMDs are distributed respectively like a χ_1^2 and χ_3^2, therefore are expected to average in turn around 1 and 3. RMSPEs are a measure of goodness of fit of marginal (that is, univariate) models, whereas SMDs carry a genuinely multivariate flavour and indicate how accurately the observed covariance structure has been captured by the model. More detailed information as to these, as well as alternative, diagnostic tools for model criticism can be found in Bayarri and Berger [15].

Table 3: Model criticism and comparison for selected healthcare resources and costs from the treatment arm of the ATLAS data-set

	E(RMSPE)			E(SMD)
	R_1	R_2	R_3	
N-N-N	1.1532	1.11	1.0881	3.7435
Poi-Poi-LN	2.1118	5.7736	1.0573	38.9735
HPoi-HPoi-LN	2.2079	2.8431	1.0026	14.3355
NBin-NBin-LN	1.4695	1.2915	1.0223	4.9426
NBin-NBin-G	1.4327	1.9981	1.0532	7.6467
Poi-Poi-G	2.1035	6.0914	1.4705	43.7602

Together with plots in Figure 2 and predictions in Table 2, average figures in Table 3 offer some valuable insights as regards the predictive performance and limitations of each proposed model. From a first inspection it seems that in particular model NBin-NBin-LN provides an overall acceptable fit: the bulk and structure of the modelled distributions roughly overlap with those of the empirical data represented in Figure 1, which also explains the fairly accurate inferences on their location and spread. Despite the presence of a number of extreme values, its overestimation of the spread of aggregate costs and the somewhat large expected SMD it produces, suggest that the NBin-NBin-LN model fails to predict some features of the data. On the other hand none of the reviewed models seems to improve upon these specific deficiencies, as they appear to generate too thin-tailed predictive distributions as underpinned by corresponding large RMSPEs. In addition to this, the magnitude of computed SMDs indicates that they fall somewhat short from adequately conveying the relationships between the examined resource uses and costs. This criticism seems to apply especially to models adopting Poi of even HPoi components. Despite the latter being designed to deal more efficiently with excess zeros than the former, both models for resource uses in ATLAS appear too poorly parameterised to satisfactorily account for their large degree of over-dispersion, which the Negative Binomial instead seems to more flexibly deal with. As to the costs side, it was briefly mentioned in Section 3.1 that significant numerical issues occurred while fitting G components. This partly explains their overall unsatisfactory validation, as highlighted by figures in Table 3, and predictions as from corresponding plots in Figure 2. Conversely inferences drawn from LN components turned out to be more stable and overall accurate than those obtained under G assumptions. However the presence of several outlying predictions again triggers some unresolved model deficiencies, which call for additional research to obtain a satisfactory representation of the underlying distribution of costs.

A separate mention about the N-N-N model is in order. Despite appearing to validate well, plots in Figure 2 expose its serious mismatch with the underlying IPD distributions. The extent of the

lack of fit is particularly noticeable with both resource use components, to which the model assigns a non-negligible probability of taking negative values. These pathological features clearly originate from the inability of the Normal distribution to account for patterns like discrete records, skewness and `spikes' at zero. It is not surprising that when applied to resource use data-sets the Normality assumption leads to generally poor predictive performance.

4 Discussion

Utilisation of multivariate statistical models in healthcare resource and cost analysis of a RCT entertains a number of desirable features. First, they explicitly take advantage of the relationships between modelled variables to corroborate inferences about relevant estimands. Univariate models such as those utilised in conventional analysis of total costs do not acknowledge the dependencies between separate healthcare resources. As a consequence, they would be of limited assistance to a policy maker interested in addressing specific issues. Transparent modelling of healthcare IPD in a multivariate setting instead helps dealing for instance with extrapolation beyond the trial follow-up period (by estimating the treatment effect on particular resources, or the impact of specific events on resource use), handling of missing values [7] understanding how resource components could be combined to treat patients in different settings and examine substitution effects between healthcare resource when changes in relative prices occur.

In addition to this, multivariate representations allow to evaluate the potentially relevant information raw ILD may convey by offering alternative (that is, non-Normal) modelling options. The present paper has the merit to draw attention to the importance of assigning meaningful distributions to the data at hand in a CEA. It is often the case that a Normal set-up lends itself as a convenient, flexible and often realistic modelling tool. However, in circumstances where its adoption is not recommended in the light of specific aspects of the underlying data, it can also yield inaccurate representations and point to misleading conclusions.

Several instances of cost analysis are documented in the specialised literature to date. Much of this work revolves around applications of results from Normal approximations [16-18], exploration of the distributions of aggregate costs and extensions of established analytical methodologies to the Bayesian framework [4-6, 10, 19].

The originality of the discussed approach is twofold: on one hand it aims at exploiting the flexibility offered by multivariate statistical models to account for the relationships between the main cost-effectiveness drivers. On the other it calls upon the capabilities of the Bayesian modelling machinery to fit alternative statistical models to available IPD and comparatively evaluate their predictive ability.

As to the application carried out on the ATLAS data-set, a significant finding was that the reliability of inferences was sensitive to the distributions assigned to healthcare resources use. The process of selecting and fitting alternative models was to various degrees laborious and technically challenging, depending upon the complexity of the chosen model instances. This seems to indicate the existence of a trade-off between model complexity and stability, which in application-specific contexts may lean in favour of more convenient lower-level modelling. This contrast may be exacerbated when dealing with IPD from large RCTs where many healthcare resources are utilised. In this case, structuring a high-dimensional statistical model is a task unlikely to be accomplished without compromise or simplifications.

A related issue is the choice of which parameterisation to adopt once a model is selected. When analysing the ATLAS data-set this was found to be a practically relevant matter for both model estimation and interpretation. When compared with more pragmatic alternatives, some sophisticated formulations may still fail to provide significantly deeper insights as to features of the distributions they attempt to characterise. Notwithstanding the wealth of model selection tools available from the statistical literature, it should be emphasised that none of them is intended for 'default' use but rather as guidance throughout the model discrimination process. Even in the absence of a compelling rationale in favour of some particular model (or even model

parameterisation), once more, a compromise between tractability and simplicity should ultimately drive modelling choices. The main advantages of the parameterisation adopted throughout the paper lie in its ease of interpretation and ability to meet the non-negativity constraints implied by the examined resource uses.

Although the analysis implemented in Section 3 on the ATLAS data-set yielded overall promising results, still there is scope for additional research. For instance it is unclear to what extent the illustrated methodology can be generalised to incorporate covariate adjustments for subgroup analysis, or hierarchical assumptions for introducing multi-centre effects. Another topic of investigation relates to the bimodality of the distribution of aggregate IPD costs in ATLAS, which calls for more sophisticated possibly non-parametric models other than those proposed in Section 3.

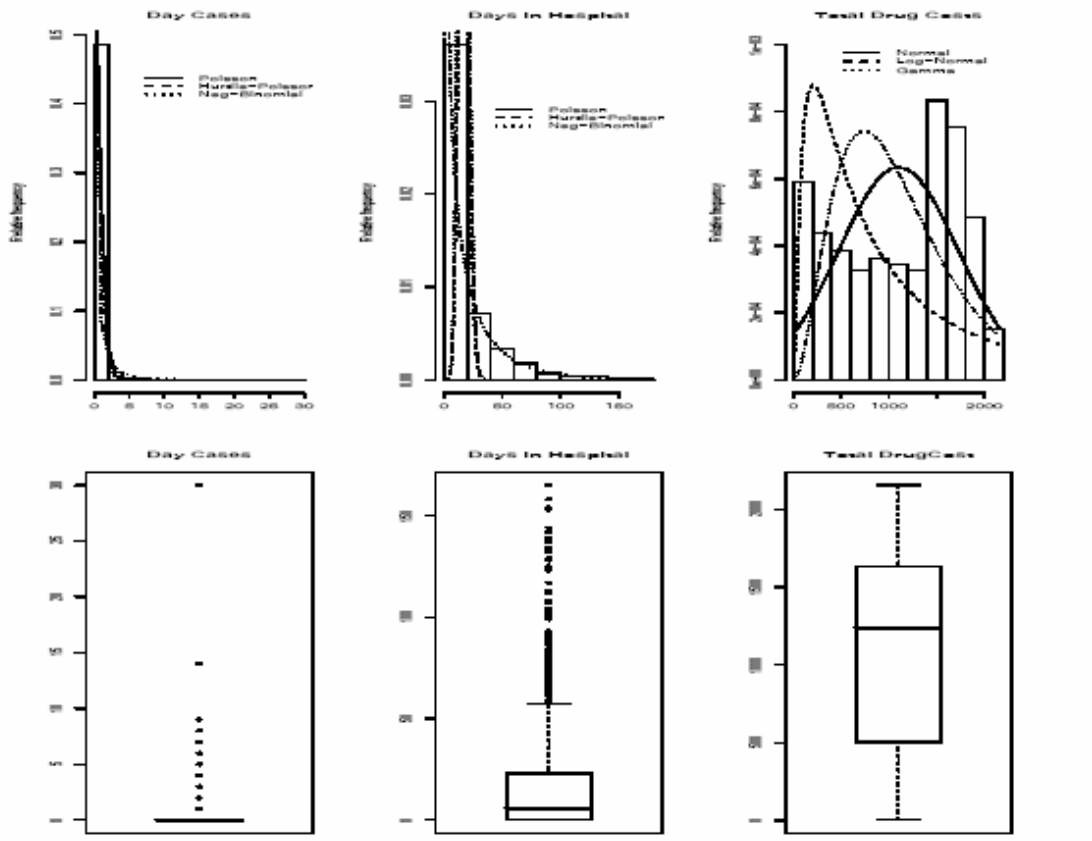

FIGURE 1. Descriptive plots (with superimposed mean- and variance-matching models) of selected healthcare resources and costs from the treatment arm of the ATLAS data-set

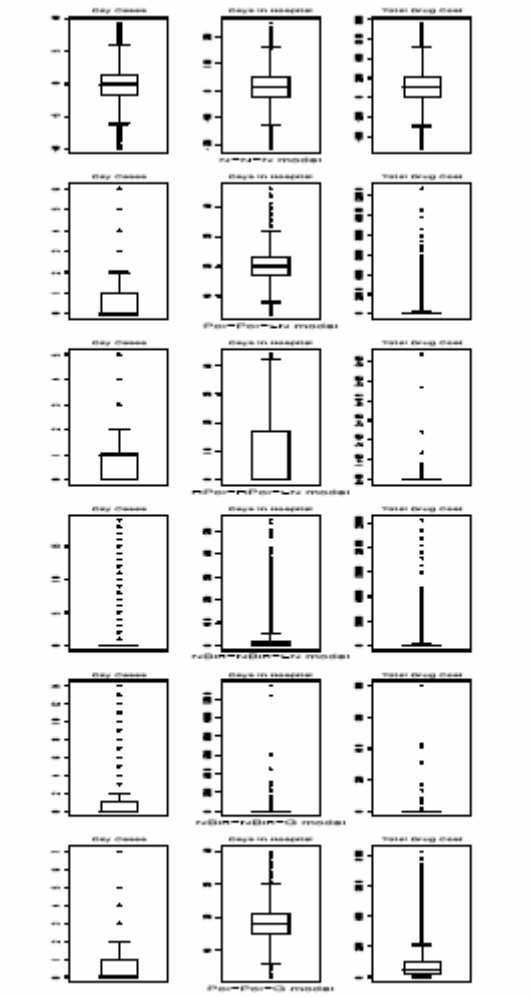

FIGURE 2. Predictive distributions from models of selected health-care resources and costs from the treatment arm of the ATLAS data-set

References

1. NICE. Guide to the Methods of Technology Appraisal. London: National Institute for Clinical Excellence; 2004 April 2004.
2. Kadane JB, Lazare NA. Methods and criteria for model selection. Journal of the American Statistical Association 2004;99(465):279-290.
3. O'Hagan A, Stevens JW. Assessing and comparing cost: How robust are the bootstrap and methods based on asymptotic normality? Health Economics 2003;12(1):33-49.
4. Briggs AH, Nixon R, Dixon S, Thompson SG. Parametric modelling of cost data: some simulation evidence. Health Economics 2005;14(4):421-8.
5. Thompson SG, Nixon RM. How sensitive are cost-effectiveness analyses to choice of parametric distributions? Medical Decision Making 2005;25(4):416-423.
6. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness analysis. Health Econ 2005;14(12):1217-29.
7. Lambert PC, Billingham C, Cooper NJ, Sutton A, Abrams KR. Estimating the cost-effectiveness of an intervention in a clinical trial when partial cost information is available: a Bayesian approach. In: Developing Economic Evaluation Methods Meeting; 2003; Aberdeen; 2003.
8. Congdon P. Bayesian Models for Categorical Data. Chicester: John Wiley and Sons Ltd.; 2005.
9. Cameron AC, Trivedi PK. Microeconometrics: Methods and Applications. New York: Cambridge University Press; 2005.
10. O'Hagan A, Stevens JW, Montmartin J. Bayesian cost-effectiveness analysis from clinical trials data. Statistics in Medicine 2001;20:733-753.
11. Nixon RM, Thompson SG. Parametric modelling of cost data in medical studies. Statistics in Medicine 2004;23:1311-31.
12. Paker M, Poole-Wilson PA, Armstrong PW, Cleland J, Horowitz JD, Massie BM, et al. Comparative effects of low and high doses of the angiotensin converting-enzyme inhibitor, lisinopril, on morbidity and mortality in chronic heart failure. Circulation 1999;100:2312-2318.
13. Sculpher MJ, Poole L, Cleland J, Drummond MF, Armstrong PW, Horowitz JD, et al. Low doses vs. high doses of the angiotensin converting-enzyme inhibitor lisinopril in chronic heart failure: a cost-effectiveness analysis based on the Assessment of Treatment with Lisinopril and Survival (ATLAS) study. European Journal of Heart Failure 2000;2:447-454.
14. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. New York: Chapman & Hall/CRC; 2004.
15. Bayarri MJ, Berger JO. Quantifying Surprise in the Data and Model Verification. In: Bayesian statistics. New York: Oxford University Press; 1999. p. 53–82.
16. Van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and c/e-ratios alongside a clinical trial. Health Economics 1994;3:309-319.
17. Laska EM, Meisner M, Siegel C. Statistical inference for cost-effectiveness ratios. Health Economic 1997;6:229-242.
18. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making 1998;18:S68-S80.

19. O'Hagan A, Stevens JW. A framework for cost-effectiveness analysis from clinical trials data. In; 2000. p. 27.