REVIEW OF STATISTICAL METHODS FOR ANALYSING HEALTHCARE RESOURCES AND COSTS, APPLICABLE TO CLINICAL TRIAL DATA

Borislava Mihaylova, Andrew Briggs*, Anthony O'Hagan**, Simon Thompson***

Health Economics Research Centre, University of Oxford, Old Road Campus, Headington, Oxford OX3 7LF, email: boby.mihaylova@dphpc.ox.ac.uk
* Public Health and Health Policy, University of Glasgow
** Department of Probability and Statistics, University of Sheffield
*** MRC Biostatistics Unit, University of Cambridge

ABSTRACT

Aims: Healthcare resource use and cost data often exhibit significant positive skewness, multimodality and heavy right tails. The objective of this review was to examine the state-of-the-art of statistical analysis of these data, by identifying the approaches employed, their ability to address the specific characteristics of the data, and their ease for general use. Thus we aimed to provide guidance to analysts on the appropriate strategy to consider when analysing resource use and costs in clinical trials where sample sizes are often limited.

Methods and data: Electronic searching of bibliographic databases was supplemented by a request to key individuals working in the area to identify further published research and hand searching of references from the papers included in the review and papers published in key journals in recent months. In total, 90 manuscripts were included in the review, and review templates were completed to summarise details of statistical model and estimation method(s), ease of implementation, characteristics of data used (including sample size), and the authors' conclusions with respect to the analytical method used.

Results: The review identified ten broad categories of analytical methods currently employed to evaluate mean healthcare resource use and costs: (1) models based on the normal distribution, (2) models based on normality following a transformation of the data, (3) single-distribution generalized linear models (GLMs), (4) other parametric models based on skewed distributions outside the GLM family of distributions, (5) models based on a mixture of parametric distributions, (6) two-part models (also hurdle models for count data), (7) semi-parametric and non-parametric methods, (8) methods based on data trimming, (9) data components models, (10) methods based on averaging across a number of models, and (11) Markov chain methods. Some of the more flexible approaches identified in this review are not readily implemented and require expertise in statistical modelling and computation. Most of the identified methods were applied to sample sizes of few hundred to thousands and their performance in more limited samples is unclear.

Conclusions: Based on this review, the recommendations are that, firstly, simple methods are preferred in large sample sizes (in the thousands) where the Central Limit Theorem guarantees normality of sample means. Secondly, in somewhat smaller sample sizes (in the hundreds), relatively simple methods, able to deal with one or two of the data characteristics studied, may be preferable but checking sensitivity of results is necessary. Finally, some more complex methods, identified in the current review, hold some promise for the future, but are relatively untried in practice; their application requires substantial expertise and, therefore, these are not currently recommended for wider applied work.

1. Introduction

It is well recognized that statistical analyses of healthcare resource use and cost data pose a number of difficulties. These data often exhibit substantial positive skewness, can have heavy tails and could be multimodal due to different categories of users (including non-users). The standard statistical approach for handling such non-normal data has been to use non-parametric methods, such as rank order statistics. Nevertheless, it is widely accepted in health economics that it is the estimated population mean cost that is the statistic of interest to decision/policy makers. Two broad areas have emerged in which the statistical modelling of cost data is undertaken.

The 'risk adjustment' field is characterised by the use of large quantities of observational data to model individual health care expenditures, with a view to understanding how the characteristics of the individual, including their health status or recent medical experience, influence overall costs. In the 'evaluation' field health care costs are collected, often alongside randomised controlled trials, in order to study the impact of interventions on individuals' healthcare resource use and costs, while minimising different biases. These studies are used to evaluate the effectiveness and cost-effectiveness of healthcare interventions and guide treatment decisions. Healthcare resource use and cost data generated in clinical trials exhibit all the features identified earlier, but with the added problem of often limited sample sizes. Approaches for the analysis of mean resource use and cost in clinical trials that take into consideration the specific features of the data are expected to lead to significant gains in precision and to more informative estimates.

The analytical literature on evaluating costs for the purpose of cost-effectiveness analysis and that on risk adjustment have developed largely independently. This paper reviews the analytical approaches employed in both areas. The objective of this review is to examine the state-of-the-art of statistical analysis of healthcare resource use and cost data, by identifying the methods employed, their ability to address the challenges of the data and their ease for general use, although our particular interest is in their applicability to mean cost differences estimated in clinical trials. Studies that propose approaches to evaluate mean resource use or costs as well as those that focus on covariate adjustment for improving efficiency of the estimates are considered for inclusion. Furthermore, this review proposes a framework to guide analysts in the appropriate approaches to consider when analysing resource use and costs in clinical trials.

2. Review methods

2.1 Identification of papers for inclusion in the review

The review aims to identify the analytical methods currently employed in evaluating health care resource use and costs that are likely to be applicable to clinical trial data. A number of exclusion

criteria were employed in order to limit the scope of what might potentially include a very large literature.

1. Since the focus of the review was methodological, we did not aim to incorporate all applications of these methods.
2. Analytical methods developed to account for selection bias or unobserved heterogeneity are not generally directly relevant to randomised trials data and were therefore considered beyond the scope of this review.
3. Methods that do not allow (or do not focus on) the estimation of the mean were also excluded.

Analytical methods that account for administrative censoring that is usually present in clinical trials are not explicitly reviewed but are briefly summarised in the discussion.

The three stage process employed to identify the key publications to include in this review is outlined in **Figure 1**. Firstly, a broad search of Medline, EconLit and MathSci bibliographic databases generated 250 publications of possible methodological interest. Applying the criteria above reduced these to 48 key publications. In the second stage, a brief outline of the review objectives together with the list of these key publications were sent to 41 individuals working in the area (identified through known contacts and posting to relevant email discussion lists) with a request to suggest further published research for inclusion in the review. 23 researchers responded and suggested a further 38 publications of potential interest of which 19 met the inclusion criteria (9 papers by one author were replaced with a subsequent review paper by the same author). In the final stage, a review of the citations from the studies included in the review yielded a further 23 studies for inclusion. In total 90 manuscripts were included in the review and 66 review templates were developed, of which 10 covered more than one paper.

2.2 Review process

The identified papers were reviewed by two of the authors (BM initially reviewed all papers and each of the remaining authors reviewed a proportion of the papers). The review aimed to present a structured factual review of the papers focusing on the method(s) employed; data used and overall methodological findings and conclusions. The summarised study details include (a) the parameters of interest and the representation of the estimation error; (b) statistical model and estimation method(s), (c) ease of implementation, (d) characteristics of the applied or simulated data used (including sample size), and (e) the authors conclusions with respect to the analytical method used. In addition, the review authors systematically tried to judge the ability of the methods to (1) address skewness, multimodality and kurtosis in data; (2) incorporate adjustment for covariates and extend to cost-effectiveness analysis, and (3) be applicable to datasets of small to

moderate sample sizes usually available in clinical trials, where these issues were not explicitly considered by the authors themselves.

3. Review results

3.1 Categorising the analytical approaches

The review identified eleven categories of analytical methods currently employed to evaluate mean healthcare resource use and costs. These methods are outlined below together with a brief description.

Normal distribution based methods

Methods based on the normal distribution are widely employed in the estimation of mean healthcare resource use and costs. These methods include inference concerning the sample mean (broadly defined to include the t-test), and the more general ordinary least squares (OLS) regression based approach. These approaches are often used in comparative studies (Austin et al. 2003;Basu 2005;Briggs et al. 2005;Briggs & Gray 1998;Chen & Zhou 2006;Deb & Burgess 2003;Dinh & Zhou 2006;Dudley et al. 1993;Gilleskie & Mroz 2004;O'Hagan & Stevens 2003;Thompson & Barber 2000;Zhou 2002;Zhou & Dinh 2005) but, although they provide unbiased estimates if the underlying distribution is not normal, this category of method is likely to lead to inefficient estimators in small to medium sample sizes.

Methods based on normality of the dependent variable following a transformation

For these methods the individual healthcare resource use or cost data is transformed in order to achieve normality and comparison of means on the transformed scale (Briggs & Gray 1998) or an OLS approach to model the transformed dependent variable (Ai & Norton 2000;Duan et al. 1983;Manning & Mullahy 2001;Veazie et al. 2003) is employed. These comparisons of means on the transformed scale do not directly inform the comparison of means on the original scale; back transformation of the results to the original scale is needed to allow for the evaluation of the mean resource use and costs (Ai & Norton 2000;Duan 1983). Although this approach is shown to provide potentially more efficient estimates (Manning & Mullahy 2001;O'Hagan & Stevens 2003), it could perform badly if an inappropriate transformation is used (Briggs et al 2005). The approaches to find the appropriate transformation have centred predominantly on the Box-Cox transformations (Hollenbeak 2005), while the approaches for back transformation to the original scale are dominated by the Duan's non-parametric smearing approach (Duan 1983) and its variants.

Single distribution generalized linear models

In generalized linear models (GLMs) mean and variance functions for the observed variable, conditional on covariates, are specified and the parameters are estimated given these structural assumptions. This approach adds flexibility due to choice of link function and mean-variance relationship relevant for the distribution (Barber & Thompson 2004). Furthermore, as the estimation is directly on the scale of the data, unlike the transformation based approaches, there is no need for back transformation. Extensions of the standard GLM approach are provided based on extending the family of distribution (Manning et al. 2005), more complex specification of the mean-variance relationships (Basu et al. 2006;Basu 2005;Blough et al. 1999) and GLM estimators that are more robust to outliers (Cantoni & Ronchetti 2006). GLMs are also used in two-part models to model the positive resource use and cost data (references in the two-part models section below).

Other parametric models based on skewed distributions outside the GLM family of distributions

Methods based on distributions outside the GLM family have been used to improve flexibility while modelling healthcare costs or resource use. Log-normal and log-logistic distributions are used by Nixon and Thomson (Nixon & Thompson 2004;Thompson & Nixon 2005) to model costs, and extended approaches based on the Poisson distribution by Cameron & Johansson (Cameron & Johansson 1997), Cameron & Trivedi (Cameron & Trivedi 1986) and Marazzi et al (Marazzi et al. 1998) have been proposed to model resource use.

Models based on mixture of parametric distributions

Mixtures of Poisson distributions (Mullahy 1997) and Negative binomial distributions (Deb & Holmes 2000;Deb & Trivedi 1997;Jimenez Martin et al. 2002) have been suggested for resource use data, and mixtures of gamma distributions for cost data (Deb & Burgess 2003) . A mixture of distributions from different families has been shown to improve on a mixtures of distributions from the same family (Atienza et al. 2008). Mixture models often perform better than model alternatives based on single distributions for total resource use or costs.

Two-part models (also hurdle model in count data models)

The two-part model is a special case of a mixture model in which only two components are allowed and one of these is degenerate. These models have been widely employed in situations where, due to large numbers of non-users of health services, there is an excessive number of zeros in the resource use or cost data. Usually, a logit or probit model in the first part estimates the probability of incurring any resource use or costs, while the mean resource use or costs, conditional on having incurred any, are evaluated in the second part. Log-linear (Duan et al 1983;Leung & Yu 1996), GLM or OLS models have been employed in the second part to evaluate mean costs, and

truncated at zero Poisson, Negative Binomial (Grootendorst 1995;Pohlmeier & Ulrich 1995) or truncated Poisson-log-normal models (Winkelmann 2004) to evaluate resource use.

Extensions to the standard two-part model have been identified in the review. Santos-Silva and Windmeijer suggest more appropriate modelling of spells of care and number of visits per spell through a Poisson model for the first part (number of spells) and a logarithmic model for the second part (number of visits per spell) (Santos-Silva & Windmeijer 2001). A semi-parametric two-part model for count regression (including Poisson or Negative-Binomial hurdle models as a special case) is suggested by Gurmu (Gurmu 1997), and can be extended through a semi-parametric single-index two-part regression model (Zhou & Liang 2006). A generalised hurdle model with a generalised logistic model in the first part is proposed by Gurmu (Gurmu 1998) and a modified second part in the two-part model is suggested by Mullahy (Mullahy 1998). A Bayesian implementation of a two-part model in WinBUGS is illustrated by Cooper et al (Cooper et al. 2003). An extension of Duan's smearing estimate to back-transform the log-linear model in the second part of the two-part model with explicit accounting for heteroscedasticity is developed by Welsh (Welsh & Zhou 2006). Another extension uses a Poisson-Lognormal model to improve flexibility when modeling positive outcomes (Winkelmann 2004).

A number of papers compare the two-part and selectivity models (Dow & Norton 2003;Duan et al. 1984;Jimenez Martin et al 2002;Maddala 1985;Manning et al. 1987). The selectivity models describe both the decision to consume healthcare and the decision on how much to consume in an attempt to account for the possibility that these decisions are made jointly and therefore are likely to be correlated. These models are excluded from this review as the aim is to inform methods useful in the context of clinical trials where the selectivity problem is less of a concern. A modified two-part model, that explicitly models and estimates the correlation between the logistic and lognormal part is suggested by Tooze et al. (Tooze et al. 2002).

Semi-parametric and non-parametric methods

This category includes the bootstrap approach (Barber & Thompson 2000;O'Hagan & Stevens 2003;Thompson & Barber 2000), modified t-test methods based on a generalized pivotal statistic (Chen & Zhou 2006) or Edgeworth expansion (Dinh & Zhou 2006;Zhou & Dinh 2005), discrete approximation of the density function of the outcome using a sequence of conditional probability density functions (Gilleskie & Mroz 2004), survival data based approaches such as the Cox proportional hazards model (Austin et al 2003;Basu et al. 2004;Dudley et al 1993;Lipscomb et al. 1998), an Aalen regression model with additive hazard function (Pagano et al. 2008) and methods based on smooth quantile estimation (Dominici et al. 2005;Dominici & Zeger 2005). A model based on piecewise constant densities (for non-extreme values) and a Generalised Pareto distribution

(GPD) (for extreme values) is also suggested (Conigliani & Tancredi 2005). Different bootstrap procedures are compared by Barber and Thompson (Barber & Thompson 2000) but O'Hagan and Stevens (O'Hagan & Stevens 2003) criticize these methods as inefficient for use with skewed healthcare cost data and caution against their use with small datasets. Although approaches based on modifications of the t-test add flexibility to account for skewness, multimodality and kurtosis, these modifications are subject to a degree of subjectivity in the choice of transformation (Dinh & Zhou 2006) and do not generally allow for adjustment for covariates (Chen & Zhou 2006;Dinh & Zhou 2006;Zhou & Dinh 2005). The non-parametric approach based on a discrete approximation of the density function of the outcome is likely to perform well when large datasets are available. Survival data based approaches perform well when the underlying assumptions are met but produce biased estimates otherwise.

Methods based on trimming of data

A series of papers by Marazzi (Marazzi 2002;Marazzi & Barbati 2003;Marazzi & Ruffieux 1999;Marazzi & Yohai 2004) illustrates the use of truncation to provide more robust estimates of the mean. Data are modelled using parametric distribution(s), that are subsequently truncated from both ends (to discard contaminants) in a way to preserve the mean of an underlying uncontaminated distribution. A degree of robustness is claimed under moderate contamination. The approach is based on the important assumption that the data are contaminated (in particular with high values) that is likely to be inappropriate in the case of healthcare resource use and costs and to lead to substantial underestimation of the mean.

Data components models

An emerging area of research is analysis of data components in which separate components of resource use and costs are modelled and the results are combined under a common analytical framework. The applications published to date refer to components of costs or resource use modelled as normally or log-normally distributed data (Hahn & Whitehead 2003;Lambert et al. 2008) and use bivariate or multivariate distributions to simultaneously model resource use components. Lambert et al. also illustrate the possibility to increase the flexibility using a two-part model for cost components with excess zeros (Lambert et al 2008).

Analytical extensions to model two or more healthcare resource use variables simultaneously have been suggested in the literature. A Bivariate Poisson (Cameron & Johansson 1998), Negative Binomial marginal distribution and copula functions (Cameron et al. 2004), independent Poisson distributions with conditional mean functions that depend on correlated latent effects (Chib & Winkelmann 2001), a multivariate over-dispersed Poisson mixture model (Gurmu & Elder 2000), Bivariate Poisson-Lognormal mixture and Bivariate Negative Binomial regression models (Munkin

& Trivedi 1999), and a bivariate zero-inflated binomial regression for count data with excess zeros (Wang 2003) have been suggested.

Methods based on averaging across a number of models

Another recent area of research explored whether averaging results across a number of models (the model averaging approach) could lead to better performance when modelling resource use and cost data. Conigliani and Tancredi (Conigliani & Tancredi 2006) showed that the performance of the Bayesian model averaging depends on whether there was a model in the model set that fits the data well; otherwise an approach based on mixtures of parametric distributions seemed more appropriate.

Markov chain methods

A new approach based on a finite Markov chain is suggested to estimate resource use over different phases of health care (Coxian phase-type distribution) and evaluate total cost by attaching unit costs to these phases (Marshall et al. 2007). A way to implement adjustment for covariates through Bayesian belief networks is also proposed (Marshall & McClean 2003). This approach could be very flexible and well tailored to the data but relies on sufficient data to allow robust modelling and estimation.

3.2 Ability of the identified analytical approaches to address the characteristics of healthcare resource use and cost data

Table 1 summarises our judgement with respect to the ability of the identified analytical approaches to account for the skewness, heavy tails, excess zeros and multimodality in data, under the adopted analytical categorisation.

Skewness and heavy tails

The use of initial transformation of the data or explicit use of parametric distributions allowing for skewness (GLM family and other distributions) can allow appropriately for skewness in data. Right tails, heavier than those of the normal distribution, are often observed with resource use and cost data. Manning and Mullahy (Manning & Mullahy 2001) report better performance of OLS models on log-transformed data in the case of heavy tail log-scale residuals, and of GLM models in the case of light tail residuals. Extensions of the GLM to Generalised Gamma or Extended Estimating Equations are shown to perform well on heavy tailed data.

The approaches based on mixture of distributions, data components or model averaging seem to add further flexibility in acknowledging difficulties in explicitly modelling the skewness in the data. Cox proportional hazards semi-parametric model is shown to perform well, at least when the

proportional hazards assumption is met, presumably due to the non-parametric evaluation of the underlying baseline hazard.

Excess zeros and multimodality

Two-part models seem to outperform other methods when excess zeros are present in data, although models based on mixtures of distributions, non-parametric density approximation and proportional hazards methods are shown to perform well in some datasets. It is generally unclear what number or proportion of zeros in data would deem the use of these approaches desirable. When more general multimodality is present, methods based on mixture of distributions, modelling the whole distribution of the data through a non-parametric density approximation, or modelling different data components seem promising approaches.

3.3 Other characteristics of the analytical approaches for modelling healthcare resource use and cost data

Adjustment for covariates and need to back transform to the original scale

In randomised trials, it is important to be able to adjust for covariates to gain precision in estimating the mean cost. Most of the identified approaches allow incorporating adjustments for covariates. The analytical approach based on initial transformation of the data has an added issue that transformation to the original scale is needed. Non-parametric approaches (Duan 1983;Welsh & Zhou 2006) for back transformations to the original scale are usually employed.

Sample size implications

Most of the analytical approaches identified have been applied to samples of a few hundreds or thousands of observations, and their performance in more limited samples is unclear.
Some of the more flexible approaches are clearly tailored for the situations when sufficient data are available to inform them. The extended GLM approach proposed by Basu (Basu 2005), for example, is not recommended for samples smaller that 5000, which would preclude its use in the majority of clinical trials. Further research is needed to study the performance of approaches based on mixtures of distributions, model averaging and data components in small samples.

Ease of implementation

The ease of implementation of different analytical approaches was judged. Approaches were judged “easy to implement” if available in standard statistical software. These include the standard OLS, methods based on alternative distributions with and without initial data transformations, the two-part models and the proportional hazards approaches. Most of the more flexible analytical methods identified in this review are not readily available and require expertise in statistical

modelling and computation. These include approaches based on mixtures of distributions, non-parametric density approximation, data components, and model averaging and Markov chain methods.

3.4 Comparing the performance of the analytical approaches

Many of the studies reviewed included an informal comparison of different methods and approaches applied to the same data (Austin et al 2003;Basu et al 2004;Briggs & Gray 1998;Buntin & Zaslavsky 2004;Chen & Zhou 2006;Cooper et al. 2006;Lipscomb et al 1998;Manning & Mullahy 2001;O'Hagan & Stevens 2003;Zhou 2002). O'Hagan (2003) suggests that methods based on asymptotic normality or simple bootstrap procedures could be misleading in the case of skewed data and approaches that explicitly account for skewness (eg. a lognormal distribution for data) could be more appropriate. The proportional hazards method was used as one of compared models in (Austin et al 2003;Basu et al 2004;Dudley et al 1993;Lipscomb et al 1998). This model shows good performance when the proportional hazards assumption holds. Test for proportional hazards is suggested in (Basu & Manning 2006). Adjusted bootstrap approaches (to account for skewness) are suggested in (Barber & Thompson 2000)(bootstrap t-methods), and (Tu & Zhou 2000). A number of publications compare different models (Austin et al 2003;Basu et al 2004;Briggs & Gray 1998;Buntin & Zaslavsky 2004) without recommending a particular approach.

However, twenty of the studies identified included a more formal evaluation of the performance of the approaches studied and comparison of different approaches in the controlled environment of simulated data. These studies have shown that methods based on initial transformation perform well only when this transformation was the appropriate for the data (Briggs et al 2005) and that extended GLM methods are likely to outperform the standard GLM and approaches based on initial data transformation (Basu et al 2006;Basu 2005;Manning et al 2005). A few studies have shown that approaches based on mixtures of parametric distributions outperform single distribution alternatives in the case of heavy tailed data (Conigliani & Tancredi 2006;Deb & Burgess 2003;Deb & Trivedi 1997).

4. Discussion

Scope of the review

We aimed to review the methods currently employed to evaluate mean healthcare resource use and costs, likely to be relevant to such data in clinical trials. Clearly, the process of identification of relevant studies to include is not straightforward. While many studies report some analysis of resource use and cost data, these analyses rarely aim to contribute to developing analytical methods. We have employed a combination of literature searches and key researcher contacts in

order to try and ensure that we have not missed important contributions to the analytical approaches for analysis of resource use and cost data.

We have excluded some statistical methods judged beyond the scope of the review. These include the panel data methods that are often employed in econometrics to control for unobservable (longitudinal) individual effects constant over time, instrumental variable approaches that are used to model the selection bias when evaluating treatment effects based on non-experimental data, selectivity models as these aim to adjust for selection biases that are usually resolved by the randomisation process in clinical trials, and parametric, semi-parametric and non-parametric methods not focusing on mean inferences (such as ordered probit/logit, grouped data regression, multinomial logit, nested logit/probit, kernel-based estimators, quantile regression not aimed at mean estimation) (Jones 2000).

Extensions to cost differences and cost-effectiveness in trials

Two main approaches towards comparing means are often employed. Firstly, direct comparison of means and their uncertainty when no adjustment for covariates is present is widely used. Secondly, allocation to treatment as a covariate in a regression model is considered. In this case, it should be noted that a direct estimation of mean cost difference is allowed only in the case of linear model wrere treatment allocation is represented as additive effect on the original scale of data.

Extension to cost-effectiveness analysis implies the ability to set the analysis of costs and health benefits in a correlated bivariate framework. Few of the reviewed papers explicitly considered extensions to cost-effectiveness. Approaches from other published work suggest that this is done either by considering net benefit, by non-parametric bootstrapping, or by setting up an explicit statistical model that links costs and effects. As the net-benefit could have very different properties over ranges of willingness to pay values, representation of cost-effectiveness over the range of acceptable willingness to pay values is needed. The bootstrap approach involved estimating the mean incremental cost (by what ever method is chosen) and the mean incremental effect for each bootstrap sample; the succession of estimates provide a distribution of points on the cost-effectiveness plane from which cost-effectiveness summaries such as the ICER and CEAC can be estimated. Setting up an explicit statistical model that links costs and effects has proved challenging beyond the bivariate normal situation (in which the incremental costs and effects follow bivariate normal distribution). Any extensions of this approach, so far (i.e. two-part or mixtures approaches), have been implementation in Markov chain Monte Carlo (MCMC), since ML solutions may be difficult to obtain (Lambert et al 2008).

Censored cost data

Although the methods accounting for censoring are not central to the review, we briefly summarise the key approaches in acknowledgement of the fact that the resource use and cost data in clinical trials are likely to be subject to a degree of administrative censoring.

The two principal methods are the Kaplan Meier Sample Average (KMSA) estimator, proposed by Lin et al. (Lin et al. 1997), and the Inverse Proportional Weighting (IPW) estimator, proposed by Bang and Tsiatis (Bang & Tsiatis 2000). Both methods are non-parametric and assume that at any follow-up time the probability of censoring is independent of the future outcomes of individuals (eg. resource use and costs). Similarly, both approaches can be applied over the overall study period or by using the study period partitioned into intervals. The KMSA estimator weights the mean cost, accumulated during a time interval by the uncensored individuals at the beginning of the interval, by the probability of surviving at the beginning of the respective interval; these are summed across all time intervals to estimate the censoring adjusted mean cost. The IPW estimator, weights costs, accumulated during time intervals by the uncensored individuals at the end of the interval, by the inverse of the probability of being censored in the time interval; these are summed across the time intervals and participants. Asymptotic standard errors or variances are provided in the single sample case. The theoretical connections between these approaches are explored (O'Hagan & Stevens 2004) and both methods are shown to perform well over different levels of censoring (Raikou & McGuire 2004). Further methods allow for adjustments for covariates together with censoring adjustment (Carides et al. 2000;Lin 2000) and extensions to cost-effectiveness (Pullenayegum & Willan 2007;Willan et al. 2005).. Extensions to better adjustment for survival time (Liu et al. 2007) and two-part models for costs (Tian & Huang 2007) are also proposed,

5. Guidance to analysts based on this review

In making practical recommendations, we have not advocated many methods suggested in the literature. We have dismissed the truncation methods of Marrazzi since their main aim is to be robust to contaminant outliers, while we regard ‘outliers’ as being part of the true distribution of costs. Other models that have been used in the papers we reviewed appear too complicated for randomised trial data of usually fairly small sample size, have only been shown to be of (sometimes rather slight) benefit in particular data sets, and present formidable problems of implementation to applied analysts. We have outlined three categories of methods.

The Green Orbit

In the green orbit, we can apply simple methods because we have enough data. Analysis can be based on assuming normal distributions for costs. Their underlying true distribution will of course not be normal, but the analysis will depend only on sample means and variances. How do we

know when we are in the green orbit? The sample size must be big enough for a number of possible problems to disappear.

- Despite skewness, excess zeroes, multimodality and/or heavy tails, the samples should be big enough for the Central Limit Theorem to guarantee near-normality of sample means. How big this is depends on the degree of skewness and also on the complexity of the covariate adjustment or subgroup analysis that is to be performed.
- The number of large costs should be sufficient for the answers not to be unduly influenced by a few very large outlying costs. This is related to the last point, but merits separate checking.

Whether we use frequentist of Bayesian analyses does not matter (because any prior information should be weak enough to be overwhelmed by the data), and so is a matter of personal preference.

Although it would be useful to have some more concrete guidance on the necessary sample sizes, this is really a matter of judgement. In broad terms, to be sure we are in the green orbit, we should certainly have hundreds of observations in each treatment group (or sub-group), and possibly thousands.

The Amber Orbit

In the amber orbit, we can still use relatively simple methods, but checking the sensitivity of the results is recommended. These methods will generally be able to deal with one or two of the possible complications with cost data, but any increase in complexity rapidly takes us into the red orbit.

Alternative distributions. Where the data are skewed and/or heavy-tailed, we can model the costs using appropriate alternative distributions instead of assuming normality. Gamma distributions are not recommended because they are sufficiently light tailed that the answers will often be similar to using normal distributions (Manning & Mullahy 2001). Inverse gamma or lognormal distributions may often be appropriate, but particularly in the case of the lognormal distribution the results may be non-robust to outliers in the data. The log-logistic distribution is probably too heavy-tailed to often be realistic in practice. Where there are enough data, or background knowledge, to suggest a particular form of distribution, then analysis using this distribution can be recommended, but sensitivity to alternative choices of distribution should be assessed (Nixon & Thompson 2004).

Transformations. We can also consider transforming the costs to reduce skewness, so that normality can be assumed on the log scale. Using the log transform is equivalent to assuming log-normality, and comes with the same caveat about non-robustness to outliers. (It also cannot be used if there are zero costs in the data; the device of replacing zero by a small number is not recommended.) Other power transformations may be considered. However, it is essential that an appropriate ‘back transformation’ is used to produce inferences on the original cost scale, rather

than on the transformed scale. Checking sensitivity to the choice of transformation is recommended.

Generalised linear models. GLMs are an attractive approach when we have covariates, because they offer some of the benefits of alternative distributions and/or transformation without the need to back transform. Limitations of GLMs are that they are based implicitly on assuming a particular distributional form (and so there is again a recommendation to check for sensitivity to this choice), and that the frequentist inferences involve approximation. Also, unless the identity link function is used (which may not always be realistic) there is still a back transformation issue.

Two-stage or hurdle models. These models specifically address zero costs, since the presence of zero values in the data is usually incompatible with any assumed continuous distribution of costs. (When modelling resource use with discrete distributions, zeroes may be expected, but two-stage modelling may still be indicated to deal with large numbers of zeroes.)

The Red Orbit

Methods in the red orbit are more complex and generally require substantial expertise, both in statistical modelling and in computation. Many of these approaches are relatively (or even completely) untried in practice. For highly complex models, the only practical computational tool to obtain results is Markov chain Monte Carlo (MCMC). This means operating in the Bayesian framework (although it is quite common for researchers to use MCMC simply for reasons of convenience, without any philosophical commitment to Bayesian statistics). The lack of user-friendly software for Bayesian analysis and MCMC means that these methods are inevitably in the red orbit. We indicate here some other typical combinations of circumstances that characterise being in the red orbit.

- When addressing skewness through alternative distributions or transformations, the presence of covariates or censoring will typically mean that suitable analyses are complex and not available in standard software.
- Multimodality may suggest either decomposing costs into resource use components, or else using mixture models so as to link the prevalence of different mixture components to covariates (in the same way as is usually done for the prevalence of zeroes in two-stage models). Another use for mixture models is to allow the tail thickness of the cost distribution to be fitted separately from the main body of data. Both resource use components models and mixture models are sufficiently complex to belong in the red orbit.
- Where the analysis involves comparing different treatment groups or sub-groups, there will typically be information to suggest some similarity between parameters of the different groups. This kind of structural prior information may be modelled naturally in a Bayesian framework, and can be influential in the analysis even when more quantitative prior information is lacking. The resulting Bayesian hierarchical models are in the red orbit.

Any suggested practical strategy comes with its own health warnings. It is based on our interpretation and opinion of the current literature, and others may disagree. The recommendations may be modified, or even overturned, by future research. Moreover, for any particular data set, it is likely that special methods may be found which out-perform those suggested here; our intention is only to give a strategy which we believe should have wide applicability.

6. Future research

The literature review has suggested that mixture models could have significant advantages in modelling skewed, heavy tailed, multimodal data. Although a number of applications exemplify models of health care resource use employing this method, further applications are needed to support its wider use. There are methods for analysing cost data that we believe hold some promise for the future, but cannot at present be recommended for applied work. One is model averaging approaches which make some allowance for the uncertainty in choosing an appropriate statistical model (Conigliani & Tancredi 2006). A second is models which allow for potentially heavy upper tails, while allowing flexible (or even non-parametric) distributions for the bulk of the data (Conigliani & Tancredi 2005). A third is Bayesian approaches which incorporate informative priors about distributional shape, particularly about the upper tails (O'Hagan & Stevens 2003). The latter is in contrast to models, implemented in MCMC simply for computational convenience, and that in general use priors intended to be non-informative. So far, data component models have not shown to lead to any improvement in efficiency and face significant technical difficulties but the research in this area is limited (Hahn & Whitehead 2003;Lambert et al 2008).

A major limitation of the implementation of more complicated models in the field of clinical trials is the need for the analytical framework to accommodate both costs and health effects and evaluate the summary cost-effectiveness measures. In doing so, the analysis should allow for the correlation structure of different outcomes. The future development of such approaches in different situations is recommended, perhaps especially for two-part models or mixture models.

Figure 1 Flow chart of selection of papers into the review

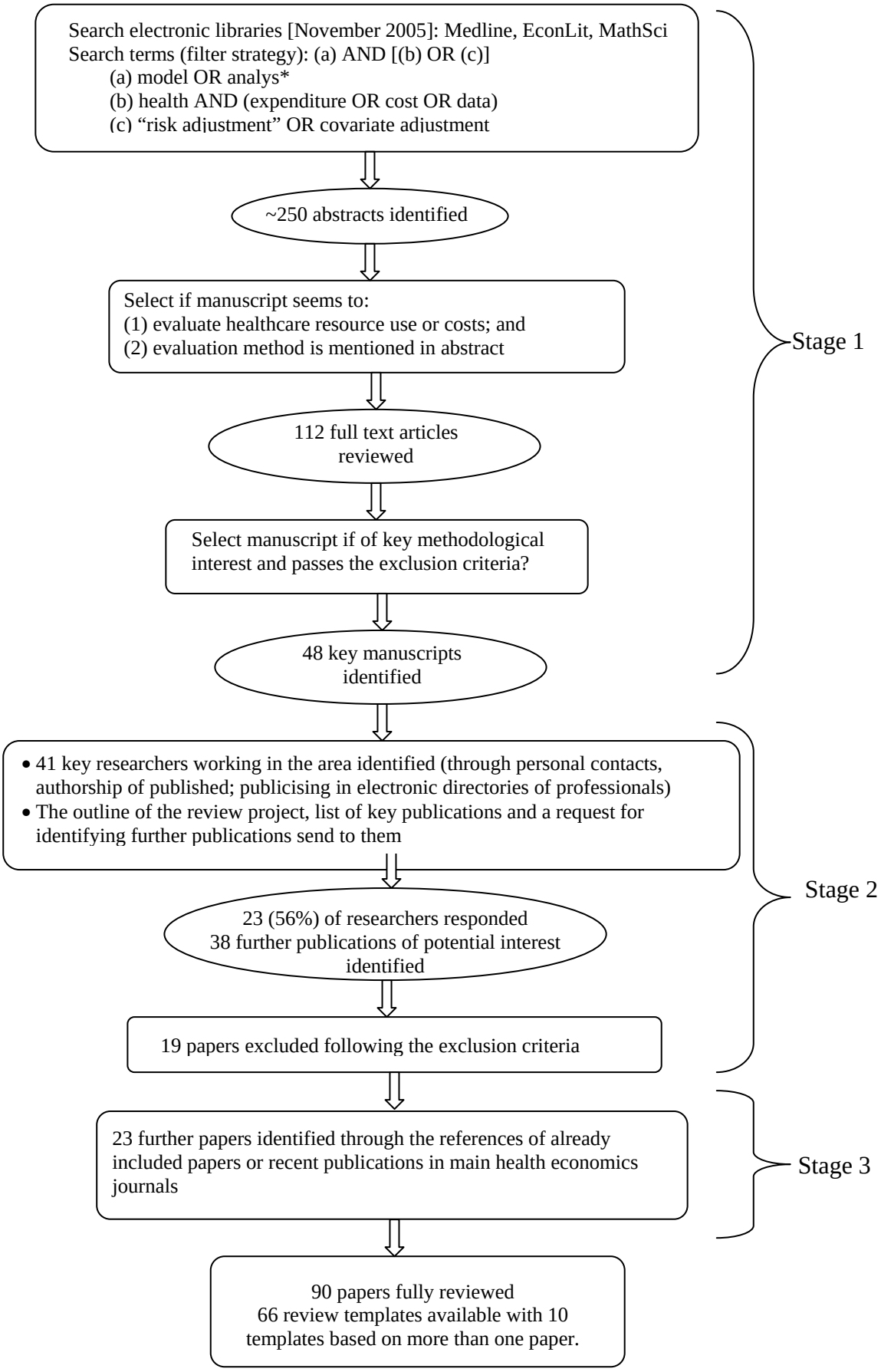

Table 1: Summary characteristics of the reviewed methods for modelling cost and resource use data

Analytical approach	Features of data: Skewness	Heavy tails	Excess zeros	Multimo-dality	Features of method: Testing for cost difference	Covariate adjustment	Analysis on original scale/ No need to back transform	Works with small samples*	Ease of implemen-tation
Methods based on normality	☹	☹	☹	☹	☺	☺	☺	😐	☺
Methods based on transformation of the dependent variable and normality on the transformed scale	☺	☺	☹	☹	☹	☺	☹	😐	☺
Single distribution generalized linear models (GLM)	☺	☹	☹	☹	☺, ☹	☺	☺	😐☹**	☺
Methods based on parametric distributions outside GLM	☺	😐	☹	☹	☺, ☹	☺	☺	😐 ☹	😐
Methods based on mixtures of parametric distributions	☺	☺	😐	☺		☺	☺	☹	☹
Two-part and selection models	☺	😐 ☺**	☺	☹	😐	☺	☹ ☺**	😐 ☹	😐 ☺
Semi-parametric and non-parametric approaches:									
Cox proportional hazards (Basu et al 2004;Lipscomb et al 1998)	☺	☺	☺	😐	☹	☺	☺	😐	☺
Aalen additive hazard model(Pagano et al 2008)	☺	☺	☹	😐	☹	😐	☺	☹	☹
Non-parametric adjusted tests	☺		☹	☹	☹	☹		😐	☺
Non-parametric density approximation (Gilleskie & Mroz 2004)	☺	☺	☺	☺	😐	☺	☺	☹	😐 ☹
Quantile based smoothing(Dominici et al 2005;Dominici & Zeger 2005)	☺	☺	☺	☺	☺	☺	😐	☹	☹
Trimming of data	☹	☹	☹	☹	😐	☹	☺	😐 ☹	☺
Methods based on data components	☺		😐	☺	😐	☺	☺	😐 ☹	☹
Model averaging	☺	☺	😐	☺	😐	☺	☺	😐 ☹	☹
Markov chain methods	☺	☺	😐	☺	😐	😐	☺	☹	☹

☺ Yes 😐 May be ☹ Not

*Small sample refers to few tens to few hundreds of participants

** Depends on exact implementation

1. Ai, C. & Norton, E. C. 2000, "Standard Errors for the Retransformation Problem with Heteroscedasticity", *J.Health.Econ.*, vol. 19, no. 5, pp. 697-718.
2. Atienza, N., Garcia-Heras, J., Munoz-Pichardo, J. M., & Villa, R. 2008, "An application of mixture distributions in modelization of length of hospital stay", *Statistics in Medicine*, vol. 27, no. 9, pp. 1403-1420.
3. Austin, P. C., Ghali, W. A., & Tu, J. V. 2003, "A comparison of several regression models for analysing cost of CABG surgery", *Stat.Med.*, vol. 22, no. 17, pp. 2799-2815.
4. Bang, H. & Tsiatis, A. A. 2000, "Estimating medical costs with censored data", *Biometrika*, vol. 87, no. 2, pp. 329-343.
5. Barber, J. A. & Thompson, S. G. 2004, "Multiple regression of cost data: use of generalised linear models", *J.Health.Serv.Res.Policy*, vol. 9, no. 4, pp. 197-204.
6. Barber, J. A. & Thompson, S. G. 2000, "Analysis of cost data in randomized trials: an application of the non-parametric bootstrap", *Stat.Med.*, vol. 19, no. 23, pp. 3219-3236.
7. Basu, A. 2005, "Extended generalised Linear Models: Simultaneous Estimation of Flexible Link and Variance Functions", *The Stata Journal*, vol. 5, no. 4, pp. 501-516.
8. Basu, A., Arondekar, B. V., & Rathouz, P. J. 2006, "Scale of interest versus scale of estimation: Comparing alternative estimators for the incremental costs of a comorbidity", *Health.Econ.*, vol. 15, pp. 1091-1107.
9. Basu, A. & Manning, W. G. 2006, "A test for proportional hazards assumption within the exponential conditional mean framework", *Health Services and Outcomes Research Methodology*, vol. 6, pp. 81-100.
10. Basu, A., Manning, W. G., & Mullahy, J. 2004, "Comparing Alternative Models: Log vs Cox Proportional Hazard?", *Health.Economics.*, vol. 13, no. 8, pp. 749-765.
11. Blough, D. K., Madden, C. W., & Hornbrook, M. C. 1999, "Modeling risk using generalized linear models", *J.Health.Econ.*, vol. 18, no. 2, pp. 153-171.
12. Briggs, A. & Gray, A. 1998, "The distribution of health care costs and their statistical analysis for economic evaluation", *J.Health.Serv.Res.Policy*, vol. 3, no. 4, pp. 233-245.
13. Briggs, A., Nixon, R., Dixon, S., & Thompson, S. 2005, "Parametric modelling of cost data: some simulation evidence", *Health.Econ.*, vol. 14, no. 4, pp. 421-428.
14. Buntin, M. B. & Zaslavsky, A. M. 2004, "Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures", *J.Health.Econ.*, vol. 23, no. 3, pp. 525-542.
15. Cameron, A. C. & Johansson, P. Bivariate count data regression using series expansions: with applications. 1998.
    Ref Type: Unpublished Work
16. Cameron, A. C. & Johansson, P. 1997, "Count Data Regression Using Series Expansions: With Applications", *Journal of Applied Econometrics*, vol. 12, no. 3, pp. 203-223.
17. Cameron, A. C., Li, T., Trivedi, P. K., & Zimmer, D. M. 2004, "Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismeasured Counts", *Econometrics Journal*, vol. 7, no. 2, pp. 566-584.
18. Cameron, A. C. & Trivedi, P. K. 1986, "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests", *Journal of Applied Econometrics*, vol. 1, no. 1, pp. 29-53.
19. Cantoni, E. & Ronchetti, E. 2006, "A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures", *Journal of Health Economics*, vol. 25, no. 2, pp. 198-213.
20. Carides, G. W., Heyse, J. F., & Iglewicz, B. 2000, "A regression-based method for estimating mean treatment cost in the presence of right-censoring", *Biostatistics*, vol. 1, no. 3, pp. 299-313.
21. Chen, Y. H. & Zhou, X. H. 2006, "Interval estimates for the ratio and difference of two lognormal means", *Stat.Med*, vol. 25, pp. 4099-4113.
22. Chib, S. & Winkelmann, R. 2001, "Markov Chain Monte Carlo Analysis of Correlated Count Data", *Journal of Business and Economic Statistics*, vol. 19, no. 4, pp. 428-435.
23. Conigliani, C. & Tancredi, A. 2005, "Semi-parametric modelling for costs of health care technologies", *Stat.Med.*, vol. 24, no. 20, pp. 3171-3184.
24. Conigliani, C. & Tancredi, A. Comparing parametric and semi-parametric approaches for Bayesian cost-effectiveness analyses. Working Paper del Dipartimento di Economia, Universita Roma Tre n. 64. 2006.
    Ref Type: Unpublished Work
25. Cooper, N. J., Lambert, P. C., Abrams, K. R., & Sutton, A. J. Predicting costs over time using Bayesian Markov Chain Monte Carlo methods: an application to early inflamatory polyartritis.

Health Economics . 2006.
Ref Type: In Press

26. Cooper, N. J., Sutton, A. J., Mugford, M., & Abrams, K. R. 2003, "Use of Bayesian Markov Chain Monte Carlo Methods to Model Cost-of-Illness Data", *Medical Decision Making*, vol. 23, no. 1, pp. 38-53.
27. Deb, P. & Burgess, J. 2003, *A Quasi-experimental Comparison of Econometric Models for Health Care Expenditures*, Hunter College Department of Economics.
28. Deb, P. & Holmes, A. M. 2000, "Estimates of use and costs of behavioural health care: a comparison of standard and finite mixture models", *Health.Econ.*, vol. 9, no. 6, pp. 475-489.
29. Deb, P. & Trivedi, P. K. 1997, "Demand for Medical Care by the Elderly: A Finite Mixture Approach", *Journal of Applied Econometrics*, vol. 12, no. 3, pp. 313-336.
30. Dinh, P. & Zhou, X. H. 2006, "Nonparametric statistical methods for cost-effectiveness analyses", *Biometrics*, vol. 62, no. 2, pp. 576-588.
31. Dominici, F., Cope, L., Naiman, D. Q., & Zeger, S. L. 2005, "Smooth quantile ratio estimation", *Biometrika*, vol. 92, no. 3, pp. 543-557.
32. Dominici, F. & Zeger, S. L. 2005, "Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable diseases", *Biostatistics*, vol. 6, no. 4, pp. 505-519.
33. Dow, W. & Norton, E. 2003, "Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions", *Health Services and Outcomes Research Methodology*, vol. 4, no. 1, pp. 5-18.
34. Duan, N. 1983, "Smearing Estimate: A Nonparametric Retransformation Method", *Journal of the American Statistical Association*, vol. 78, no. 383, pp. 605-610.
35. Duan, N., Manning, W. G., Jr., Morris, C. N., & Newhouse, J. P. 1984, "Choosing between the Sample-Selection Model and the Multi-Part Model", *Journal of Business & Economic Statistics*, vol. 2, no. 3, pp. 283-289.
36. Duan, N., Manning, W. G., Jr., Morris, C. N., & Newhouse, J. P. 1983, "A Comparison of Alternative Models for the Demand for Medical Care", *Journal of Business & Economic Statistics*, vol. 1, no. 2, pp. 115-126.
37. Dudley, R. A., Harrell, F. E., Jr., Smith, L. R., Mark, D. B., Califf, R. M., Pryor, D. B., Glower, D., Lipscomb, J., & Hlatky, M. 1993, "Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery", *Journal of Clinical Epidemiology*, vol. 46, no. 3, pp. 261-271.
38. Gilleskie, D. B. & Mroz, T. A. 2004, "A Flexible Approach for Estimating the Effects of Covariates on Health Expenditures", *J.Health.Econ.*, vol. 23, no. 2, pp. 391-418.
39. Grootendorst, P. V. 1995, "A comparison of alternative models of prescription drug utilization", *Health Economics*, vol. 4, no. 3, pp. 183-198.
40. Gurmu, S. 1998, "Generalized Hurdle Count Data Regression Models", *Economics Letters*, vol. 58, no. 3, pp. 263-268.
41. Gurmu, S. 1997, "Semi-Parametric Estimation of Hurdle Regression Models With an Application to Medicaid Utilization", *Journal of Applied Econometrics Special Issue: Econometric Models of Event Counts*, vol. 12, no. 3, Special Issue: Econometric Models of Event Counts, pp. 225-242.
42. Gurmu, S. & Elder, J. 2000, "Generalized Bivariate Count Data Regression Models", *Economics Letters*, vol. 68, no. 1, pp. 31-36.
43. Hahn, S. & Whitehead, A. 2003, "An illustration of the modelling of cost and efficacy data from a clinical trial", *Statistics in Medicine*, vol. 22, no. 6, pp. 1009-1024.
44. Hollenbeak, C. S. 2005, "Functional form and risk adjustment of hospital costs: Bayesian analysis of a Box-Cox random coefficients model", *Stat.Med*, vol. 24, no. 19, pp. 3005-3018.
45. Jimenez Martin, S., Labeaga, J. M., & Martinez Granado, M. 2002, "Latent class versus two-part models in the demand for physician services across the European Union", *Health.Econ.*, vol. 11, no. 4, pp. 301-321.
46. Jones, A. M. 2000, "Health econometrics," in *Handbook of Health Economics*, A. J. Culyer & J. P. Newhouse, eds., Elsevier.
47. Lambert, P. C., Billingham, L. J., Cooper, N. J., Sutton, A. J., & Abrams, K. R. 2008, "Estimating the cost-effectiveness of an intervention in a clinical trial when partial cost information is available: A Bayesian approach", *Health.Econ.*, vol. 17, no. 1, pp. 67-81.
48. Leung, S. F. & Yu, S. 1996, "On the Choice between Sample Selection and Two-Part Models", *Journal of Econometrics*, vol. 72, no. 1-2, pp. 197-229.
49. Lin, D. Y. 2000, "Proportional means regression for censored medical costs", *Biometrics*, vol. 56, no. 3, pp. 775-778.

50. Lin, D. Y., Feuer, E. J., Etzioni, R., & Wax, Y. 1997, "Estimating medical costs from incomplete follow-up data", *Biometrics*, vol. 53, no. 2, pp. 419-434.
51. Lipscomb, J., Ancukiewicz, M., Parmigiani, G., Hasselblad, V., Samsa, G., & Matchar, D. B. 1998, "Predicting the cost of illness: a comparison of alternative models applied to stroke", *Med.Decis.Making*, vol. 18, no. 2 Suppl, p. S39-S56.
52. Liu, L., Wolfe, R. A., & Kalbfleisch, J. D. 2007, "A shared random effects model for censored medical costs and mortality", *Statistics in Medicine*, vol. 26, no. 1, pp. 139-155.
53. Maddala, G. S. 1985, "A Survey of the Literature on Selectivity Bias as it Pertains to Health Care Markets," in *Advances in Health Economics and Health Services Research*, R. Scheffler & L. Rossiter, eds., JAI Press, Greenwich, CT.
54. Manning, W. G., Basu, A., & Mullahy, J. 2005, "Generalized modeling approaches to risk adjustment of skewed outcomes data", *J.Health.Econ.*, vol. 24, no. 3, pp. 465-488.
55. Manning, W. G., Duan, N., & Rogers, W. H. 1987, "Monte Carlo Evidence on the Choice between Sample Selection and Two-Part Models", *Journal.of.Econometrics.*, vol. 35, no. 1, pp. 59-82.
56. Manning, W. G. & Mullahy, J. 2001, "Estimating log models: to transform or not to transform?", *J.Health.Econ.*, vol. 20, no. 4, pp. 461-494.
57. Marazzi, A. 2002, "Bootstrap tests for robust means of asymmetric distributions with unequal shapes. Computational Statistics & Data Analysis", *Computational Statistics & Data Analysis*, vol. 39, pp. 503-528.
58. Marazzi, A. & Barbati, G. 2003, "Robust parametric means of asymmetric distributions: estimation and testing", *Estadistica*, vol. 54, pp. 47-72.
59. Marazzi, A., Paccaud, F., Ruffieux, C., & Beguin, C. 1998, "Fitting the distributions of length of stay by parametric models", *Med.Care*, vol. 36, no. 6, pp. 915-927.
60. Marazzi, A. & Ruffieux, C. 1999, "The truncated mean of an asymmetric distribution", *Computational Statistics & Data Analysis*, vol. 32, no. 1, pp. 79-100.
61. Marazzi, A. & Yohai, V. J. 2004, "Adaptively truncated maximum likelihood regression with asymmetric errors", *Journal of Statistical Planning and Inference*, vol. 122, no. 1-2, pp. 271-291.
62. Marshall, A. H. & McClean, S. I. 2003, "Conditional phase-type distributions for modelling patient length of stay in hospital", *International Transactions in Operational Research*, vol. 10, no. 6, p. 565.
63. Marshall, A. H., Shaw, B., & McClean, S. I. 2007, "Estimating the costs for a group of geriatric patients using the Coxian phase-type distribution", *Statistics in Medicine*, vol. 26, no. 13, pp. 2716-2729.
64. Mullahy, J. 1998, "Much ado about two: reconsidering retransformation and the two-part model in health econometrics", *J.Health.Econ.*, vol. 17, no. 3, pp. 247-281.
65. Mullahy, J. 1997, "Heterogeneity, Excess Zeros, and the Structure of Count Data Models", *Journal of Applied Econometrics*, vol. 12, no. 3, pp. 337-350.
66. Munkin, M. K. & Trivedi, P. K. 1999, "Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models, with Application", *Econometrics Journal*, vol. 2, no. 1, pp. 29-48.
67. Nixon, R. M. & Thompson, S. G. 2004, "Parametric modelling of cost data in medical studies", *Stat.Med.*, vol. 23, no. 8, pp. 1311-1331.
68. O'Hagan, A. & Stevens, J. W. 2003, "Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality?", *Health.Econ.*, vol. 12, no. 1, pp. 33-49.
69. O'Hagan, A. & Stevens, J. W. 2004, "On estimators of medical costs with censored data", *Journal of Health Economics*, vol. 23, no. 3, pp. 615-625.
70. Pagano, E., Petrinco, M., Desideri, A., Bigi, R., Merletti, F., & Gregori, D. Survival models for cost data: the forgotten additive approach. Stat.Med . 2008.
Ref Type: In Press
71. Pohlmeier, W. & Ulrich, V. 1995, "An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care", *Journal of Human Resources*, vol. 30, no. 2, pp. 339-361.
72. Pullenayegum, E. M. & Willan, A. R. 2007, "Semi-parametric regression models for cost-effectiveness analysis: improving the efficiency of estimation from censored data", *Statistics in Medicine*, vol. 26, no. 17, pp. 3274-3299.
73. Raikou, M. & McGuire, A. 2004, "Estimating medical care costs under conditions of censoring", *Journal of Health Economics*, vol. 23, no. 3, pp. 443-470.
74. Santos-Silva, J. M. C. & Windmeijer, F. 2001, "Two-Part Multiple Spell Models for Health Care Demand", *Journal of Econometrics*, vol. 104, no. 1, pp. 67-89.

75. Thompson, S. G. & Barber, J. A. 2000, "How should cost data in pragmatic randomised trials be analysed?", *BMJ*, vol. 320, no. 7243, pp. 1197-1200.
76. Thompson, S. G. & Nixon, R. M. 2005, "How sensitive are cost-effectiveness analyses to choice of parametric distributions?", *Med.Decis.Making*, vol. 25, no. 4, pp. 416-423.
77. Tian, L. & Huang, J. 2007, "A two-part model for censored medical cost data", *Statistics in Medicine*, vol. 26, no. 23, pp. 4273-4292.
78. Tooze, J. A., Grunwald, G. K., & Jones, R. H. 2002, "Analysis of repeated measures data with clumping at zero", *Statistical Methods in Medical Research*, vol. 11, no. 4, pp. 341-355.
79. Tu, W. & Zhou, X. H. 2000, "Pairwise comparisons of the means of skewed data", *Journal of Statistical Planning and Inference*, vol. 88, no. 1, pp. 59-74.
80. Veazie, P. J., Manning, W. G., & Kane, R. L. 2003, "Improving risk adjustment for Medicare capitated reimbursement using nonlinear models", *Med.Care*, vol. 41, no. 6, pp. 741-752.
81. Wang, P. 2003, "A Bivariate Zero-Inflated Negative Binomial Regression Model for Count Data with Excess Zeros", *Economics Letters*, vol. 78, no. 3, pp. 373-378.
82. Welsh, A. H. & Zhou, X. H. 2006, "Estimating the retransformed mean in a heteroscedastic two-part model", *Journal of Statistical Planning and Inference*, vol. 136, no. 3, pp. 860-881.
83. Willan, A. R., Lin, D. Y., & Manca, A. 2005, "Regression methods for cost-effectiveness analysis with censored data", *Statistics in Medicine*, vol. 24, no. 1, pp. 131-145.
84. Winkelmann, R. 2004, "Health Care Reform and the Number of Doctor Visits--An Econometric Analysis", *Journal of Applied Econometrics*, vol. 19, pp. 455-472.
85. Zhou, X. H. 2002, "Inferences about population means of health care costs", *Stat.Methods.Med.Res*, vol. 11, no. 4, pp. 327-339.
86. Zhou, X. H. & Dinh, P. 2005, "Nonparametric confidence intervals for the one- and two-sample problems", *Biostatistics.*, vol. 6, no. 2, pp. 187-200.
87. Zhou, X. H. & Liang, H. 2006, "Semi-parametric single-index two-part regression models", *Computational Statistics & Data Analysis*, vol. 50, no. 5, pp. 1378-1390.