REVIEW OF STATISTICAL METHODS FOR ANALYSING HEALTHCARE RESOURCES AND COSTS

ADDITIONAL MATERIAL

CONTENT

GLOSSARY OF CATEGORIES OF METHODS

I. **Methods based on the normal distribution**- Methods rely on the Central Limit Theorem that the mean of a sufficiently large sample will be approximately normally distributed, independently of the population distribution. The category includes the one-sample normal confidence intervals for population mean cost, two-sample normal confidence intervals for the difference in mean costs between two populations, the t-test and ordinary least squares regression (OLS) regression.

II. **Methods following transformation of data** - Methods that usually employ OLS regression following data transformation to ensure that it has an approximately normal distribution.

III. **Single-distribution generalized linear models (GLMs)**- Methods that relate the random distribution of the dependent variable to the systematic (non-random) part (the *linear predictor*) through a function called the link function. Each outcome of the dependent variables is assumed to be generated from a particular distribution function in the exponential family.

IV. **Parametric models based on skewed distributions outside the GLM family** – Models that are based on distributions outside the exponential family in order to allow for larger skewness in data. These usually do not have estimating functions linear in the response variable, and are usually fitted by numerical iteratively techniques.

V. **Models based on mixtures of parametric distributions**- Finite mixture models decompose a density function into component density functions, where each component describes a particular class. However, the current estimation of finite mixture models assumes the mixture component density functions have a classical parametric form (i.e. Gaussian, Poisson, etc.). Algorithms estimate both the parameters of these distributions, as well as the contribution of each class in the mixture.

VI. **Two (or multi)-part models and Generalised Tobit models**-In these models, the dependent variable is decomposed into two components which are modelled as separate random variables. Most frequently, one of the components consists of the “0” values of the dependent variable, and the second component consists of the positive values of the dependent variable modelled through separate regressions. Generalized Tobit model explicitly models the correlation between the two components through a bivariate normal error term.

VII. **Survival (or duration) methods**- methods to model cost at "time" variable that relax the parametric assumptions of the exponential conditional mean models. Examples include Weibull and Cox proportional hazards models and the Aalen regression model with additive hazard function.

VIII. **Non-parametric methods**- a broad category of methods including

(i) Bootstrap approach- a method that estimates properties of a statistic by measuring those properties when sampling from an approximating distribution, often the empirical distribution of the statistic based on the data.

(ii) Modified t-test methods based on a generalized pivotal statistic or Edgeworth expansion,

(iii) Discrete approximation of the density function of the outcome using a sequence of conditional probability density functions or piecewise constant densities

(iv) Quantile based smoothing- a method that smoothes, over percentiles, the ratio of the quantiles of the two distributions.

IX. **Methods based on truncation or trimming of data**- In these methods, data are modelled using parametric distribution(s), which are subsequently truncated from both ends (to discard contaminants) in a way to preserve the mean of an underlying uncontaminated distribution.

X. **Data components models**- In these models components of resource use or costs are modelled separate and the results are combined under a common analytical framework to allow for the correlation between components.

XI. **Methods based on averaging across a number of models**-Methods in this category weight the estimates obtained from individual models from a set of models by some measure of model fit to provide an overall estimate.

XII. **Markov chain methods**- The dependent variable is modelled through a small number of compartments usually corresponding to stages of clinical care. The papers reviewed include the use of Coxian phase-type distribution.

ABBREVIATIONS

ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
ATML	Adaptively truncated maximum likelihood
BIVARNB	Bivariate Negative Binomial
BMA	Bayesian model averaging
BVNB	Bivariate Negative Binomial
BVPLN	Bivariate Poisson-Lognormal mixture
BZINB	Bivariate zero-inflated negative binomial
ECM	Exponential conditional mean
EM	Expectation maximisation
FIML	Full-information maximum likelihood
FMM	Finite mixture model
FMNB	Finite mixture negative binomial
GBIVARNB	Generalized Bivariate Negative Binomial
GLM	Generalized linear model
GMM	Generalized method of moments
GPD	Generalized Pareto distribution
HNB	Hurdle Negative Binomial
LCM	Latent class models
LIML	Limited-information maximum likelihood
LR	Likelihood ratio
MAE	Mean absolute error
MAPE	Mean absolute prediction error
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood
MRSE	Mean relative squared error
MSE	Mean squared error
MSPE	Mean squared prediction error
MTPM	Modified two-part model
NB	Negative Binomial
NLLS	Non-linear least squares
OLS	Ordinary Least Squares
RAND HIS	RAND Health Insurance Experiment
RMSE	Root mean squared error
SSM	Sample selection model

SQUARE	Smooth quantile ratio estimation
TML	Truncated maximum likelihood
TPM	Two-part model

STATISTICAL PROGRAMS (specified in papers)

GAUSS, GLIM, LIMDEP, MATLAB, R, SAS, S-Plus, Stata, BUGS/WinBUGS, XploRe

TEMPLATES OF REVIEWED PAPERS

Reference: **Ai2000 (Ai & Norton 2000)**
Category: Models based on normality following a transformation of the data
Parameter(s): Standard error of mean costs in retransformation problems with heteroscedasticity
Data: Positive cost data *Model*: Variances of conditional mean of *y* in problems with retransformation from **log-transformation** with heteroscedasticity within normal linear model, normal nonlinear model, nonparametric model with homoscedasticity, nonparametric model with heteroscedasticity. *Statistical method(s):* Standard methods (ordinary least squares, nonlinear least squares) *Implementation*: STATA; program code could be requested from the authors *Representation of estimation error*: Delta method, bootstrap method
Applied data: 1 dataset (Total cost over 2 years on 9863 individuals with severe mental illness and at least one healthcare claim); *Sample size*: 9863 *Comparisons with other models/methods:* Models with and without heteroscedasticity were compared.
Authors' suggestions: Controlling for heteroscedasticity is needed when evaluating the point estimates and the standard errors. *Sample size implications as discussed by the authors*: In large datasets the assumption for normality is justified and the suggested equations could be used directly, but in small datasets the bootstrap approach is recommended.
Connections to other papers: Duan1983, Duan1984, Manning 1998, Mullahy1998
Review Comment *Skewness, heavy tails, multimodality:* This paper discusses the variance estimation in log-transformed model. Therefore skewness and heavy tails are addressed appropriately only if data is lognormally distributed. The approach could not accommodate multimodality. *Adjustment for covariates*: The methods allow for adjustment for covariates. *Extensions to cost-effectiveness*: Extensions to model both costs and health outcomes does not seem straightforward. *Sample size implications:* The notion that the bootstrap will perform better in small samples is misleading as the bootstrap method provides approximations for confidence intervals and coefficients valid only for sufficiently large samples.

Reference: **Atienza2008 (Atienza et al. 2008)**
Category: Models based on mixtures of parametric distributions
Parameter(s): Mean resource use
Data: Skewed, heavy tailed, multimodal resource use data *Model*: A finite mixture of distributions in the union of Gamma, Weibull and Lognormal families. *Statistical method:* Maximum likelihood method and expectation maximisation (EM) algorithm. ML estimators of each family of distributions is used for initial values and conjugate gradient acceleration method is used to improve speed of convergence. *Implementation*: Not specified. *Representation of estimation error*: Variance of the distribution mixture.
Applied data: Length of stay data for five Diagnostic Related Groups in a Spanish hospital; *Sample size*: 1579, 1156, 1086, 1639 and 1528. *Simulated data*: Mixtures of Lognormal-Gamma-Weibull for 128 combinations of parameter values; *Sample size*: 100, 500 *Comparisons with other models/methods:*
Authors' suggestions: A finite mixture of distributions in the union of Gamma, Weibull and Lognormal families performed well in a series of simulation and improved on mixtures of

distributions from the same family. *Sample size implications as discussed by authors*: Not discussed but method employed on sample sizes of 100.
Connections to other papers: Austin2003a, Deb1997, Deb2000, Deb2003, Marazzi1998, Mullahy1997
Review Comment *Skewness, heavy tails, multimodality:* The use of mixture of distributions from different families is likely to increase the flexibility in modelling such data. *Adjustment for covariates/ Extensions to cost-effectiveness*: Use of covariates or joint modelling (e.g. of resource use, costs and health outcomes) are not discussed, but the approach should extend to these cases although this does not seem straightforward. *Sample size implications:* The approach seems not to need large sample sizes. *General comments:* Method seems suited to modelling length-of-stay data, but further work required to look at more complex mixtures of resource use items.

Reference: **Austin2003 (Austin et al. 2003)**
Category: Comparative study of methods
Parameter(s): Mean cost
Data: Positive cost data *Models*: Compares linear regression; linear regression on log-transformed cost; generalised linear models with log link and Poisson, negative binomial and gamma distributions; median regression; and proportional hazards models. Mean squared prediction error (MSPE), bias, mean relative squared error (MRSE), mean absolute prediction error (MAPE) are employed for model selection. *Statistical method:* Standard methods in the software. *Implementation*: SAS (S-Plus for median regression). *Representation of estimation error*: Model-based standard errors; bootstrap approach for median regression.
Applied data: 1 dataset; *Sample size*: 1959 patients; 18% data used for validation.
Authors' suggestions: Different selection criteria lead to different preferred models. Variety of models should be considered and the final model selected based on careful assessment of predictive ability and tailored to the particular dataset. Using MSPE, the proportional hazards regression and the GLMs studied predicted best the costs in an independent sample. *Sample size implications as discussed by authors*: Not discussed
Connections to other papers: Blough2000, Duan1983, Dudley1993, Lipscomb1998, Manning1998, Manning2001
Review Comment *Skewness, heavy tails, multimodality:* There is not explicit modelling of skewness, heavy tails, multimodality. The proportional hazards model performs well under the MSPE and bias criteria. The log model and GLM could accommodate a degree of skewness in the data that corresponds to the employed distribution. *Adjustment for covariates*: The models have the ability to adjust for covariates with different degree of flexibility. *Extensions to cost-effectiveness*: Extensions to joint modelling of costs and health outcomes does not look straightforward beyond the normal linear regression model. *Sample size implications:* The methods seem applicable for medium sample sizes. For small datasets the ability to model covariates could be restricted.

Reference: **Austin2003a (Austin & Escobar 2003)**
Category: Models based on mixtures of parametric distributions
Parameter(s): Distribution of health utility index (HUI)
Data: Health utilities of individuals

Model: Finite mixture of normal distributions (mixtures of up to 6 components considered). *Statistical method*: Bayesian inference. *Implementation*: MCMC Gibbs sampler in BUGS; Bayes factors to select number of mixture components; program code not available. *Representation of estimation error:* Posterior distribution.
Applied data: 1 dataset; *Sample size*: a random sample of 2,000 from a dataset of 17,530; *Comparisons with other models/methods*: None
Authors' suggestions: Finite mixture models allow describing complex distributions using a small number of simple components. *Sample size implications*: Subsample of 2000 used because fitting mixture models to censored data is a computationally intensive process.
Connections to other reviewed papers:
Review Comment *Skewness, heavy tails, multimodality:* Mixtures could be used to model outcomes and costs, allowing for skewness, heavy-tails and multi-modality. *Adjustment for covariates/ Extensions to cost-effectiveness*: Use of covariates or joint modelling (e.g. of costs and utilities) are not discussed, but the approach should extend to these cases. *Sample size implications:* In many cases, fitting mixtures would not involve censoring, so the computational problems of large samples would be reduced.

Reference: **Barber2000 (Barber & Thompson 2000)**
Category: Non-parametric methods
Parameter(s): Mean cost difference
Data: Cost data that is potentially skewed and with excess zeros. *Models*: Non-parametric bootstrap methods (bootstrap-t, variance stabilized bootstrap-t, percentile, bias corrected percentile). *Statistical method:* Standard methods in the statistical software. *Implementation*: STATA *Representation of estimation error*: Implemented
Applied data: 2 datasets; *Sample size*: 144; 32 *Comparisons with other models/methods:* Normal theory (t-tests on original and log-transformed scale), Mann-Witney U-test and permutation test.
Authors' suggestions: t-test is shown to be robust to departures from normality. Among the methods for obtaining bootstrap confidence intervals, the most reliable are the bootstrap-t and the bias-corrected percentile approaches. In smaller datasets the bootstrap approach is recommended for highly skewed data. *Sample size implications as discussed by authors*: Larger sample size is better but in moderate sized samples bootstrap is likely to perform well for parameters such as means or mean differences.
Connections to other papers: Briggs1998, Duan1983, O'Hagan2003, Thompson2000, Tu1999, Zhou1997B
Review Comment *Skewness, heavy tails, multimodality:* There is not explicit modelling of skewness, heavy tails, multimodality. *Adjustment for covariates*: Normal theory and bootstrap methods extend to modelling covariates through multiple regressions. *Extensions to cost-effectiveness*: Bootstrap approaches are readily extendible to cost-effectiveness. *Sample size implications:* The bootstrap methods seem to perform well in estimating means in medium size datasets. *General comments:* Models based on transformation of the dependent variable should present

the results on the raw scale.

Reference: **Barber2004 (Barber & Thompson 2004)**
Category: Single-distribution generalized linear models
Parameter(s): Mean cost difference
Data: Positively skewed cost data *Model*: Generalised Linear methods (GLM) with identity link (additive effects) and log link (multiplicative effects); and with Gaussian, Overdispersed Poisson, Gamma, or Inverse Gaussian distribution; AIC criteria used to compare models. *Statistical method:* Standard methods in the statistical software- extended quasi maximum likelihood. *Implementation*: STATA, SAS *Representation of estimation error*: Confidence intervals based on asymptotic standard errors.
Applied data: 2 datasets; *Sample size*: 667, 149
Authors suggestions: GLM attractive for modelling costs with direct inference about the mean cost. Akaike information criteria could guide the choice of distribution but not helpful for the choice of the link function. *Sample size implications as discussed by authors*: Likelihood based confidence intervals could be more appropriate for small datasets.
Connections to other papers: Barber2000, Blough1999, Blough2000, Duan1983, Manning2001, Mullahy1998
Review Comment *Skewness, heavy tails, multimodality:* GLM do not explicitly model skewness, heavy tails, and multimodality in the data. The choice of distribution in GLM could accommodate a degree of skewness in data. *Adjustment for covariates*: GLM are flexible in modelling covariates. *Extensions to cost-effectiveness*: GLM are not easily extendible to modelling costs and utilities together. Extension through net benefit is suggested by the authors. *Sample size implications:* More complex GLMs would generally require larger sample sizes. *General comments:*

Reference: **Basu2004 (Basu & Manning 2006;Basu et al. 2004)**
Category: Comparative study of methods
Parameter(s): Mean length of stay, mean cost
Data: Positive length of stay, positive total costs. *Model*: OLS on log-transformed stay or costs; GLM with log link and gamma distribution, GLM with log link and Weibull distribution and Cox proportional hazards model. Test for proportional hazards assumption within exponential conditional mean models proposed. *Statistical method(s):* Standard methods *Implementation*: STATA *Representation of estimation error*: Robust standard errors (using the appropriate analogue of the Huber/White correction to the variance/covariance matrix).
Applied data: 1 dataset; *Sample size*: 6500 *Simulated data*: 3 data generating mechanisms (1) log normal model with constant error variance on log scale; (2) gamma model with shapes corresponding to either monotonically declining probability density function or a distribution that is skewed to the right; (3) model meeting the proportional hazard assumption under Gompertz distribution; *Sample size*: 10000
Authors' suggestions: Proportional hazards model acceptable only when the proportional hazards assumption holds. Gamma model with a log link performs satisfactorily under different data generating mechanisms and applied examples. The proposed test for proportional hazards within the class of exponential conditional mean models (within generalised gamma regression) performs as well as traditional test based on Cox regression. This general method for selecting between parametric models and the semi-

parametric Cox model increases the flexibility of use of the Generalised Gamma Regression framework. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Blough1999, Duan1983, Dudley1993, Lipscomb1998, Manning1998, Manning2001, Manning2003, Manning2005, Mullahy1998
Review Comment *Skewness, heavy tails, multimodality:* GLM do not explicitly model skewness, heavy tails, and multimodality in the data. Proportional hazards model are likely to provide advantages in this type of data in the case when the proportional hazards assumption holds. *Adjustment for covariates*: All compared methods allow for adjustment for covariates. *Extensions to cost-effectiveness*: Extensions of the models to model both costs and utilities do not seem straightforward. *Sample size implications:* Larger samples would improve model choice.

Reference: **Basu2006, Basu2005a, Basu2005b (Basu 2005;Basu et al. 2006;Basu & Rathouz 2005)**
Category: Single-distribution generalized linear models
Parameter(s): Mean cost, mean cost difference
Data: Positive cost data *Model*: Extended Generalised Linear Model (GLM) in which the link and the variance are defined as functions of the linear predictor and estimated together with the covariates. *Statistical method*: Extended estimating equations *Implementation*: Implemented in STATA; program code available. *Representation of estimation error*: Variance estimated using sandwich estimator to account for clustering.
Applied data: 2 datasets; *Sample size*: 6500; 7428 *Simulated data*: Gamma, Inverse Gaussian, Log-normal, Poisson and negative binomial data. *Sample size*: 2000; 10000 (500 replications) *Comparison with other models/methods*: Compared with standard OLS; OLS on log-transformed costs with homoscedastic or heteroscedastic smearing estimate; Gamma, Poisson, and Inverse Gaussian GLMs with log link.
Authors' suggestions: This extended GLM appropriately represents the scale of estimation and is less susceptible to over fitting. Some convergence problems with log normal data are reported. *Sample size implications*: Samples above 5000 recommended.
Connections to other papers: Basu2004, Basu2005a, Blough1999, Duan1983, Manning1998, Mullahy1998, Manning2001
Review Comment *Skewness, heavy tails, multimodality:* Extension of widely-advocated GLM approach. GLMs do not explicitly model skewness, heavy tails or multi-modality of data. The extensions increase the GLM's flexibility without really addressing those limitations. *Adjustment for covariates:* It has the advantages of the GLM approach generally, primarily the ability to model covariates flexibly. *Extensions to cost-effectiveness*: Extension to joint modelling of costs and utilities does not look straightforward. *Sample size implications:* Large sample sizes needed to fit well this model. These sample sizes are unlikely to be available in clinical trials. *General comments:* This extended GLM has the GLM's disadvantages that the inference (using estimating equations or quasi-likelihood) is approximate and there is no explicit modelling of skewness, heavy tails or multi-modality.

Reference: **Blough1999 (Blough et al. 1999)**

Category: Single-distribution generalized linear models
Parameter(s): Mean cost
Data: Positive total costs *Model*: Extended GLM: use of quasi-likelihood to model response without specifying distribution and use of extended quasi-likelihood methods to estimate models with different variance functions or different dispersion parameters. *Statistical method(s):* Extended quasi likelihood (McCullagh and Nelder, 1989). Adjustment for over-fitting and calibration based on Copas1983. *Implementation*: Not stated *Representation of estimation error*: Standard errors estimated through extended quasi likelihood.
Applied data: 1 dataset; *Sample size*: 125 109 for estimation, 1284 for validation
Authors' suggestions: Flexible approach for modelling skewed data. This approach does not require distributional assumptions and the choice of link and variance functions can be formally tested by embedding the model in parametric classes of each. Authors also suggest GEE can be used for correlated observations/repeated measures. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Duan1983, Duan1984, Manning1987
Review Comment *Skewness, heavy tails, multimodality:* This approach (quasi and extended quasi-likelihood) relaxes the distributional assumptions of the standard GLM. *Adjustment for covariates*: The methods allow for adjustment for covariates. *Extensions to cost-effectiveness*: Extensions of the models to model both costs and utilities do not seem straightforward.

Reference: **Briggs1998 (Briggs & Gray 1998)**
Category: Comparative study of methods
Parameter(s): Mean cost
Data: Positive skewed cost data *Model*: Parametric methods on the transformed (natural log, square root, reciprocal) scale and non-parametric methods (Mann-Whitney U test, Wilcoxon rank sum test, Kendall's S test; non-parametric bootstrapping). *Statistical method(s):* Standard methods in most statistical software *Implementation*: Software not specified *Representation of estimation error*: Yes
Applied data: 5 datasets; *Sample size*: 63, 148, 148, 187, 224 *Comparisons with other models/methods:* None
Authors' suggestions: The variability in costs is often large and large samples are needed to detect moderate differences in costs. Parametric methods on transformed data should back transform results on the original scale. Non-parametric rank sum methods have limited use in economic analyses as they do not provide inference for the mean. Non-parametric bootstrap is a useful test of the appropriateness of parametric assumptions and an estimation method where the parametric assumptions do not hold. *Sample size implications as discussed by the authors*: Sufficiently large samples justify the assumption of normal sampling distribution for the mean (Central Limit Theorem)
Connections to other papers: Barber2000, O'Hagan2002
Review Comment *Skewness, heavy tails, multimodality:* Non-parametric rank sum methods can not inform mean estimation. The methods based on transformation would be suitable if the transformation suitably accounts for skewness and the back transformation to original scale is well performed. None of the methods account for multimodality in data. *Adjustment for covariates:* Parametric methods easily adjust for covariates. The non-

parametric methods studied do not allow adjusting for covariates
Extended to cost-effectiveness: Extensions of the methods studied to cost-effectiveness analysis are not straightforward beyond the normal linear regression model.
Sample size implications: Model choice could be difficult in small samples.

Reference: **Briggs2005 (Briggs et al. 2005)**

Category: Models based on normality following a transformation of the data

Parameter(s): Mean cost

Data: Positive skewed cost data; Lognormal and gamma-distributed data
Model: Lognormal estimators in the context of lognormal and gamma distributed data and 3 empirical datasets
Statistical method(s): Standard methods
Implementation: Software not specified
Representation of estimation error: Yes

Applied data: 3 datasets; *Sample sizes*: 972, 1191, 1852
Simulated data: Lognormal and Gamma distributed data with coefficients of variation ranging from 0.25 to 2; *Sample sizes*: 20, 50, 200, 500, 2000
Comparisons with other models/methods: Sample mean

Authors' suggestions: Failure to appropriately model the distributional form of cost data could lead to misleading estimates. If the appropriate distribution is modelled considerable gains in efficiency can be realised. Sample mean performs well and is unlikely to lead to inappropriate inference. Central Limit Theorem seems to apply for sample sizes much larger that the rule of thumb of 30 (more likely about 200-500) in the presence of marked skewness.
Sample size implications as discussed by the authors: Parametric modelling of costs difficult with small sample sizes

Connections to other papers: Briggs1998, Deb2003, Duan1983, Manning1998, Manning2001, Mullahy1998, O'Hagan2003, Thompson2000, Zhou1997, Zhou1997b

Review Comment
Skewness, heavy tails, multimodality: The primary aim of this paper was to demonstrate that **blindly** following the lognormal distribution can lead to very bad results. Criticism could easily be levelled towards this as there was no accounting for the fit of the models.
Adjustment for covariates: No adjustment for covariates was performed but from either of the two models could be done easily.
Extended to cost-effectiveness: Extensions to cost-effectiveness are not straightforward.
Sample size implications: As acknowledged by the authors large sample sizes are needed to parametrically model data well.

Reference: **Buntin2004 (Buntin & Zaslavsky 2004)**

Category: Comparative study of methods

Parameter(s): Mean cost

Data: Total cost
Model: Comparison of OLS, two-part OLS with lognormal retransformation, two-part OLS with smearing retransformation, two-part OLS with two smearing factors, two-part OLS with square root smeared retransformation, two-part GLM with constant variance, GLM with constant variance, GLM with variance proportional to mean squared, GLM with variance proportional to mean.
Statistical method(s): OLS, quasi-likelihood
Implementation: Not stated.
Representation of estimation error: Yes

Applied data: 1 dataset (Medicare Current beneficiary Survey); *Sample size*: 10134 (about 9% had zero total cost)
Comparisons with other models/methods: None

Authors' suggestions: The authors recommend that exploration should start with the one-part GLM models and choose link (check log link) and variance (Park test); then continue with 2-part version of the best fitting GLM and then two-part OLS models with potentially different smearing factors for different parts of the distribution. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Basu2005a, Blough1999, Duan1983, Duan1984, Manning1998, Manning2001, Mullahy1998
Review Comment *Skewness, heavy tails, multimodality:* The two (and more)–part models are useful in presence of multimodality. Transformations could be used to normalise the data but special attention is needed to back transform the estimates while acknowledging the possible heteroscedasticity. *Adjustment for covariates:* All methods incorporate adjustments for covariates. *Extended to cost-effectiveness:* Extensions to cost-effectiveness are not straightforward. *Sample size implications:* Large sample sizes are needed to well model the data. *General comments*: Main message that there is no easy answer and researchers should go through a series of models to develop an appropriate model for the given data.

Reference: **Cameron1986 (Cameron & Trivedi 1986)**
Category: Parametric models based on skewed distributions outside the GLM family
Parameter(s): Mean resource use
Data: Non-negative resource use data (number of consultations with a doctor or a specialist). *Model*: Count data models- Normal, Normal with log transformation, Poisson, Negative Binomial with constant variance-mean ratio, Negative Binomial with variance-mean ratio linear in the mean, Compound Poisson models and Ordinal Probit model. Tests for specification (score test, overdispersion, Wald and Likelihood ratio test) are proposed. *Statistical method(s):* (Quasi-generalised) pseudo maximum likelihood; *Implementation*: LIMDEP & GLIM *Representation of estimation error*: Yes
Applied data: 1 dataset: Australian Health Survey; *Sample size*: 5190 *Comparisons with other models/methods:* None
Authors' suggestions: Quasi-generalized pseudo-maximum likelihood is a general method of evaluating flexible count data models. The power of different tests needs further evaluation. A sequential modelling strategy recommended. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Cameron1984, Cameron1985
Review Comment *Skewness, heavy tails, multimodality:* Some of the proposed models could to a degree adjust for skewness and heavy tails if the distribution employed represents data well. These models could not accommodate well multimodal data. *Adjustment for covariates:* Implemented but interpretation of model parameters not straightforward *Extended to cost-effectiveness:* The models could not easily extend to modelling both cost and health outcomes data. *Sample size implications:* Reasonable sample size is needed to model well data parametrically.

Reference: **Cameron1997 (Cameron & Johansson 1997;Cameron & Johansson 1998)**
Category (Univariate case): Parametric models based on skewed distributions outside the GLM family **Category (Multivariate case): Data components models**
Parameter(s): Mean of univariate or bivariate count data
Data: Univariate or bivariate count data

Model: Univariate case: baseline density Poisson polynomial model of order p (PPp model); polynomial expansion for the count variable density that fits both over- and under-dispersed data. Bivariate case: Parametric models based on a squared polynomial expansion around a given joint count density (PPp is a polynomial Poisson expansions of order p). Bivariate Poisson used in model presentation, simulations and application. *Statistical method(s)*: Maximum likelihood; a hybrid fast simulated annealing/ gradient method *Implementation*: Univariate case: GAUSS programs at http://www.econ.queensu.ca/jae/1997-v12.3/cameron-johansson *Representation of estimation error:* Standard errors for parameter estimates
Applied data: 2 datasets; *Sample size*: 126; 5190 *Simulated data*: Univariate case: Poisson polynomial of order 1 data; Bivariate case: Independent Poissons, bivariate Poisson data; *Sample size*: 200 *Comparisons with other models/methods:* Univariate data: Poisson, Hurdle, double Poisson, GECk (underdispersed data); Poisson, Negative Binomial (overdispersed data). Bivariate data: independent Poissons, independent negative binomials, independent PP1/PP2/PP3, bivariate Poisson, bivariate PP1/PP2
Authors' suggestions: The new class of models achieves good distributional fit. The use of fast simulated annealing could provide computational advantages, but it is advisable to try a range of starting values. The PPp model was very useful in the case of underdispersed data where it outperformed the double Poisson and GECk models and was more parsimonious than the hurdle model. In the case of overdispersed data, fitted well by negative binomial model, the PPp outperformed slightly the negative binomial model but was not as parsimonious. Use of negative binomial instead of Poisson densities could lead to more parsimonious model in the case of overdispersed data. In the bivariate case the approach allows simultaneously to model overdispersion and correlation between the two variables of interest. *Sample size implications as discussed by the authors*: Not discussed.
Connections to other papers: Cameron1986, Cameron1998, Gurmu1997, Gurmu1998, Gurmu2000
Review Comment *Skewness, heavy tails, multimodality*: This extension to the Poisson count model could fit better skewed, heavy tailed data than Poisson model but its advantages were not demonstrated in the case of data that was fit well by the negative binomial model. Modelling overdispersion and the correlation between the two variables seems to lead to gains in efficiency compared to models that do not account for these features in data. *Adjustment for covariates/ Extensions to cost-effectiveness*: The models implements adjustment for covariates through the parameter in the (Poisson) model. Further adjustments for covariates, for example of the parameters of the polynomial expansion, are suggested. Extensions to cost-effectiveness do not look straightforward. *Sample size implications*: It is likely that the models proposed would need larger sample sizes. *General comments*: Computational requirements are increased. Implicit in the argument that the approach is better suited to under dispersion is that this unlikely to replace the NegBin model for health care data that are more usually over dispersed. Although implementation method is not stated, it is likely that this is not straightforward. The paper suggesting the application in the bivariate case is not published.

Reference: **Cameron2004 (Cameron et al. 2004)**
Category: Data components models
Parameter(s): Mean resource use (counts)
Data: Bivariate resource use data *Model*: Copula functions to generate bivariate parametric models to model differences in

counts. Negative binomial-2 marginal distributions (NB2) and Frank copula are used. *Statistical method*: ML *Implementation*: Quasi-Newton procedure (requiring only first derivatives); straightforward and computationally efficient; code not available *Representation of estimation error*: Approximate variance based on robust sandwich method.
Applied data: 1 dataset; *Sample size*: 502 *Comparisons with other models/methods:* Marshall-Olkin bivariate negative binomial with univariate negative binomial marginals, generated as a shared frailty; unobserved heterogeneity approach (Munkin1999)
Authors' suggestions: The copula model outperforms the Marshall-Olkin and the unobserved heterogeneity models possibly due to being the least restrictive. The approach can be applied to model any function of two outcomes when data on the two outcomes is available but where no convenient expression for the joint distribution of the two outcomes is available. *Sample size implications*: Not discussed
Connections to other papers: Cameron1998, Chib2001, Munkin1999
Review Comment *Skewness, heavy tails, multimodality:* The approach could accommodate a degree of skewness through choice of marginal distributions. *Adjustment for covariates:* Implemented *Extensions to cost-effectiveness*: It is hard to see how this could be used in cost-effectiveness analysis. In principle, it could be applied to resource use for a single cost component, but only if both treatments were given to the same patient, which is hardly ever the case in practice. Extension to more than one cost component will not be straightforward.

Reference: **Cantoni2006 (Cantoni & Ronchetti 2006)**
Category: Single-distribution generalized linear models
Parameter(s): Mean cost
Data: Skewed and heavy tailed cost data *Model*: Robust GLM approach for evaluation of skewed and heavy-tailed outcomes. *Statistical method*: Robust estimation equations *Implementation*: Weighted least squares algorithm; S-PLUS programs could be requested from the authors. *Representation of estimation error*: M-estimator of variance
Applied data: 1 dataset; *Sample size*: 100 *Simulated data*: Gamma model with log link (non-contaminated and 10% contaminated); *Sample size*: 1000 *Comparisons with other models/methods:* Standard GLM.
Authors' suggestions: This approach to estimation is more robust to outliers. The approach provides a class of robust test statistics for variable selection by comparison of two nested models. Future extensions will account for zero inflated data as wells as robust simultaneous equations for GLM. *Sample size implications*: Not discussed
Connections to other papers: Blough1999, Cantoni2006, Duan1983, Gilleskie2004, Manning1987, Manning2001, Marazzi2003, Marazzi2004
Review Comment *Skewness, heavy tails, multimodality:* The extension should increase the GLM's ability to deal with skewness and heavy tails, but it is not clear whether down-weighting of extreme costs is consistent with the need to estimate population mean cost, or whether it biases the estimation analogously to trimming. It also has the GLM's disadvantages that the inference (using estimating equations or quasi-likelihood) is approximate and there is no explicit modelling of skewness, heavy tails or multi-modality. *Adjustment for covariates:* It has the advantages of the GLM approach generally, primarily

the ability to model covariates flexibly. *Extensions to cost-effectiveness*: Extension to joint modelling of costs and utilities does not look straightforward. *General comments:* Extension of widely-advocated GLM approach. Robustness and consistency of mean estimates need to be further investigated.

Reference: **Chen2006 (Chen & Zhou 2006)**
Category: Comparative study of methods
Parameter(s): Difference of two lognormal means
Data: Two-sample positive skewed data *Model*: Maximum likelihood approach based on back transformation from log transformed data, bootstrap approach, signed log-likelihood ratio approach, generalized pivotal approach. *Statistical method(s)*: Maximum likelihood, Sampling methods, generalized pivotal *Implementation*: R *Representation of estimation error:* Confidence interval of difference of means
Applied data: Medical costs of patients with Type I diabetes and patients treated for diabetic ketoacidosis; *Sample size*: Not stated. *Simulated data*: Log-normally distributed data ; *Sample size*: 5/5; 25/25; 50/50; 5/25; 25/50
Authors' suggestions: The generalized pivotal method provided the most accurate coverage frequencies. The maximum likelihood and bootstrap approaches showed extremely poor coverage. The signed log-likelihood ratio approach resulted in fairly accurate coverage, but the frequencies were noticeably worse in small samples. The generalized pivotal approach is recommended. These methods are inappropriate for data that includes zeros. Quantile plots of the log-transformed data and Shapiro-Wilk test are recommended to be used to check whether the data follow lognormal distribution. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Briggs2005, Zhou2000
Review Comment *Skewness, heavy tails, multimodality*: The methods performed well for log-normally distributed data and are unlikely to perform similarly for other data distributions. *Adjustment for covariates/ Extensions to cost-effectiveness*: The methods do not adjust for covariates. Extensions to multiple outcomes and cost-effectiveness are likely not to be straightforward. *Sample size implications*: In small samples the checks for log-normality of the data are likely to be of limited power. *General comments*: Suggested method computationally very expensive. No assessment of methods in data that isn't lognormal but is not rejected in normality tests on log scale.

Reference: **Chib2001 (Chib & Winkelmann 2001)**
Category: Data components models
Parameter(s): Joint posterior distribution of correlated count data
Data: Correlated count data *Model*: Conditional on correlated latent effects (one for each subject and outcome), the counts are assumed to be independent Poisson with a conditional mean function that depends on the latent effects and a set of covariates. Extensions with distribution of latent effects being multivariate-*t* instead of the multivariate normal or allowing the mean of the latent effects to depend on a set of covariates. *Statistical method(s)*: Bayesian methods *Implementation*: MCMC methods *Representation of estimation error:* Posterior distribution
Applied data: 2 datasets; *Sample size*: 4406; 16 *Comparisons with other models/methods:* Independent count models.

Authors' suggestions: This is a general simulation-based approach for the analysis of multivariate count data. Adding parametric distributional assumptions for the latent effects, allows fitting high-dimensional models and provides a clean interpretation for the correlation structure. The approach allows for computation of the predictive distributions based on the output from the MCMC algorithm.
Sample size implications as discussed by the authors: The MCMC method is practical even in high dimensional problems.

Connections to other papers: Deb1997, Gurmu2000, Munkin1999

Review Comment
Skewness, heavy tails, multimodality: Limited by the ability of the Poisson models to model separate counts.
Adjustment for covariates/ Extensions to cost-effectiveness: Adjustments for covariates implemented. Extensions to cost-effectiveness possible but with unclear overall performance.

Reference: **Conigliani2005 (Conigliani & Tancredi 2005)**

Category: Non-parametric methods

Parameter(s): Mean cost

Model: A Bayesian approach to model cost data with a distribution composed of a piecewise constant density up to an unknown endpoint, and a generalized Pareto distribution (GPD) for the remaining tail.
Implementation: MCMC: Gibbs kernel and Metropolis Hastings for parameter updating; program code not available
Representation of estimation error: Posterior distribution

Applied data: 1 dataset; *Sample size*: 905
Comparisons with other models/methods: Compared with normal theory, bootstrap, lognormal models

Authors' suggestions: This is a very flexible model able to fit dataset with very different shapes both in the bulk of data and in the tail. This model is suitable for cost data that exhibit high skew, heavy tail and multi-modality. Limitations related to extensions for covariate adjustment, informative priors on GPD.
Sample size implications: not discussed

Connections to other papers: Ai2000, Briggs2000, Nixon2004, O'Hagan2001, Thompson2000

Review Comment
Skewness, heavy tails, multimodality: Model designed to deal with skewness, heavy tails and multi-modality.
Adjustment for covariates/ Extensions to cost-effectiveness: Differences in mean costs are considered and extension to joint modelling of costs and utilities is mentioned, although it is not obvious how that would be done or how covariates could be incorporated.
General comments: A complex method that will not be accessible to health economists until software is available.

Reference: **Conigliani2006 (Conigliani & Tancredi 2006) and Conigliani2009 (Conigliani & Tancredi 2009)**

Category: Methods based on averaging across a number of models

Parameter(s): Mean cost, cost-effectiveness

Data: Cost data typical in clinical trials
Model: A Bayesian model averaging (BMA) approach that weights the inferences obtained with individual models from a specified sensible set of models for cost data by their posterior probability. The employed models includes Log-normal, Gamma, Generalized Pareto distribution (GPD), Weibull, and Log-logistic.
Implementation: MCMC; program code not available

Representation of estimation error: Posterior distribution
Applied data: 1 dataset; *Sample size*: 328 *Simulated data: D*ata generated from a log-normal distribution, a Weibull distribution, a mixture of a gamma distribution and a GPD, and a mixture of three log-normal distributions; *Sample size:* 50, 200, 500 *Comparisons with other models/methods:* Compared with the semi-parametric mixture approach combining a piecewise constant density up to an unknown endpoint, and a generalized Pareto distribution for the remaining tail in Connigliani2005 (Conigliani & Tancredi 2005).
Authors' suggestions: The simulation experiments showed that the posterior credible intervals obtained with the mixture model are wider that those obtained with the BMA but have better coverage. This is due to the fact the mixture model does not make assumptions about the distribution of costs, while the performance of the BMA model depends on whether there is a model within the set that is a good approximate for the data. In the BMA approach particular care should be devoted in specifying the set of models to ensure that the models are appropriate for the data. This requirement leads to difficulties in assigning and interpreting prior model probabilities especially due to models from the set being nested within each other. The flexibility of the semi-parametric mixture model also comes at the price of in terms of efficiency of the achieved inference. *Sample size implications as discussed by the authors*: With sample size of 50 it was considered inappropriate to apply the semi-parametric mixture model as 5-10% of data should be used for estimation of the tail.
Connections to other papers: Briggs2005, Conigliani2005, Nixon2004, Thompson2005
Review Comment *Skewness, heavy tails, multimodality:* The BMA increases the flexibility to model the skewness, heavy tails, and multimodality in the cost data but poses challenges in the establishment of all possible models to be included and fully defining the Bayesian model in terms of sensible priors. *Adjustment for covariates/ Extensions to cost-effectiveness*: Extension to cost-effectiveness is implemented in the applied data example. Extensions to incorporate covariates do not seem straightforward. *General comments:* The BMA method was not shown to outperform the semi-parametric mixture model and its performance should be explored in more applications. BMA is considerably more straightforward than the mixture model, however there are substantive issues over the choice of models to consider and use of appropriate prior distributions.

Reference: **Cooper2003 (Cooper et al. 2003)**
Category: Two-part models
Parameter(s): Mean cost
Data: Total cost (19% zeros) *Model*: A two-part model with first part a logistic model and second part a normal linear regression of log-transformed cost from a Bayesian perspective. The non-parametric smearing factor of Duan1983A is employed within the back transformation. *Statistical method(s)*: Bayesian methods: Markov Chain Monte Carlo (MCMC). *Implementation*: WinBUGS *Representation of estimation error*: Posterior distribution.
Applied data: 1 dataset; *Sample size*: 187 *Comparisons with other models/methods:* None
Authors' suggestions: Flexible approach to simultaneous evaluation/validation of cost models in WINBUGS. *Sample size implications as discussed by the authors*: Cross-validation methods useful when small sample sizes are available.

Connections to other papers: Barber1998, Duan1983, Dudley1993, Lipscomb1998, Manning1998, Manning2001, Mullahy1998, Thompson2000
Review Comment *Skewness, heavy tails, multimodality*: This is a Bayesian application of the standard two-part model (in Duan1983). The two-part model and log transform can address these issues in a limited way. *Adjustment for covariates*: The method allows for adjustment for covariates. *Extensions to cost-effectiveness*: Extensions of the model to model both costs and utilities does not seem straightforward.

Reference: **Cooper2007 (Cooper et al. 2007)**
Category: Comparative study of methods
Parameter(s): Mean cost
Data: Annual cost data for inflammatory polyarthritis patients over 5 years *Model*: Multilevel/hierarchical linear regression and two-part models: (1) lognormal regression with a random effect on the intercept; (2) lognormal regression with random effects on intercept and year; (3) two-part model with 2nd part lognormal regression with random effect on intercept and year; and (4) two-part model with 2nd part gamma regression with a log link and random effects on intercept and year. Two different transformation factors applied for back transformation: (a) a global smearing estimator across all years; and (2) a separate smearing estimator for each year. *Statistical method(s)*: Bayesian hierarchical methods *Implementation*: WinBUGS; implementation code could be requested from the lead author *Representation of estimation error:* Posterior distribution
Applied data: Costs of inflammatory polyarthritis patients over 5 years; *Sample size*: 431 (split into a learning sample of 331 and a test sample of 102 patients).
Authors' suggestions: Two-part models provide better fit to the data than the single equation models and avoid the arbitrary addition of a constant to the single equation models to enable transformation. The incorporation of correlation between the random effects of a model is important, as they may not be independent. Future work will address the incorporation of the smearing estimators fully into the Bayesian framework (currently the frequentist's non-parametric estimators are employed) and incorporation of partial cost information. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Cooper2003, Lipscomb1998, Tooze2002
Review Comment *Skewness, heavy tails, multimodality*: The paper presents applications of the Bayesian hierarchical single-equation and two-part models and demonstrate their superiority in the case of skewed data with excess zeros. The predictive performance of the models is evaluated based on split sample cross-validation. The flexibility of the models to model the skewness, heavy tails and multimodality is fully determined by the statistical model (independently of whether implemented in frequentist or Bayesian framework) but the Bayesian framework is likely to facilitate modelling. *Adjustment for covariates/ Extensions to cost-effectiveness*: Adjustment for covariates is incorporated and extensions to modelling cost-effectiveness are likely to be straightforward. *Sample size implications*: Difficulties with convergence, likely to be more difficult with small sample sizes. Short runs for results suggests that running time of the model could be long so that larger sample sizes could lead to computational difficulties. *General comments*: Availability of code makes using the model straightforward however it is important that users fully understand the model before using it. For example, random effects for each patient induce positive correlation across time points.

Reference: **Deb1997 (Deb & Trivedi 1997)**

Category: Models based on mixtures of parametric distributions
Parameter(s): Mean resource use
Data: Resource use count data *Model*: A finite mixture negative binomial (FMNB) count model. *Statistical method(s)*: Quasi maximum likelihood (Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm) *Implementation*: SAS *Representation of estimation error*: Robust standard errors (sandwich type)
Applied data: 1 dataset; *Sample size*: 4406 *Simulated data*: Negative Binomial (NB), Hurdle Negative Binomial (HNB), Finite Mixture Negative Binomial FMNB2 (2 components) with low means and high proportion of zeros; *Sample size*: 500 *Comparisons with other models/methods:* Standard NB model and the hurdle extension of the negative binomial model (HNB).
Authors' suggestions: FMNB with two components best within each density class and mixture models based on NB1 (variance proportional to the mean) specification performs better than NB2 (variance proportional to the squared mean) specification. The performance of the models on the real data supports the consumer-theoretic framework, in which there is no need to assume that the decision making for the demand for health care has a two-part structure. Further research could explore whether "low user" and "high user" categories across different type of usage are correlated. If so, this will further support such dichotomy. *Sample size implications as discussed by the authors*: Not discussed.
Connections to other papers: Cameron1988, Pohlmeier1995, Deb2000, Deb2003
Review Comment *Skewness, heavy tails, multimodality*: Mixture models give great flexibility in modelling skewed, heavy-tail and potentially multimodal data. A general finite mixture model, even with only two components, is shown to be more flexible than a two-part model. *Adjustment for covariates*: Mixture models allow adjustment for covariates to be implemented. A separate covariate model within each component of the mixture and a covariate model between components (to describe how their weights depend on covariates) are built. *Extensions to cost-effectiveness*: The extension of the mixture model to model multiple response categories, simultaneously correlated health outcomes, and to extend to cost-effectiveness does not seem straightforward.

Reference: **Deb2000 (Deb & Holmes 2000)**
Category: Models based on mixtures of parametric distributions
Parameter(s): Mean number of visits/ Mean cost
Data: Number of outpatient visits (zero data present); positive cost data *Model:* Finite mixture model (FMM) to estimate the utilisation of (mixture of two negative binomial densities) and expenditures (mixture of 2 lognormal densities) of behavioural health care. *Statistical method:* ML (Broyden-Fletcher-Goldfarb-Shanno algorithm) *Implementation*: SAS/IML; program code not available. *Representation of estimation error*: Robust (sandwich-type) standard errors.
Applied data: 1 dataset; *Sample size*: 1594 *Comparisons with other models/methods:* Hurdle (two-part) model for the count data (negative binomial density for the first part (probability of being non-user) and truncated binomial distribution for the second part); lognormal regression model for the continuous data.
Authors' suggestions: This method leads to better model fit and superior estimates of mean costs and utilization compared to traditional two-part hurdle models. This model allows

distinguishing between distinct classes of users of behavioural health care (i.e. "worried well", "severely mentally ill"). Using FMM, the authors modelled only positive values in the continuous data (expenditures) case as no continuous density function can model the large numbers of non-users. *Sample size implications as discussed by the authors*: Larger number of components could be considered with larger sample sizes but care should be taken against the potential for overfitting the data.
Connections to other papers: Deb1997, Deb2003, Duan1983, Pohlmeier1995
Review Comment *Skewness, heavy tails, multimodality:* Finite mixture models have flexibility to explicitly model skewness, heavy tails and multimodality. *Adjustment for covariates/ Extensions to cost-effectiveness*: Finite mixture models have the ability to adjust for covariates. The extension to joint modelling of costs and effects for the purpose of cost-effectiveness analysis does not look straightforward. *General comments:* Mixture models appear to be a useful way of allowing for flexible distributional shape.

Reference: **Deb2003 (Deb & Burgess 2003)**
Category: Models based on mixtures of parametric distributions
Parameter(s): Mean cost
Data: Positive cost data *Model*: Finite mixture model (FMM) with two/three gamma densities and linear mean specifications. *Statistical method:* Maximum likelihood *Implementation*: Not stated *Representation of estimation error*: Mean (absolute) prediction error in validation samples.
Applied data: 1 dataset; *Sample size*:~2,500,000 *Simulated data*: Quasi-experimental (samples from original data); *Sample size*: 10000, 50000, 100000, 200000, 500000. *Comparisons with other models/methods:* OLS model (on untransformed, log and square root transformed expenditures with prediction by smearing); GLM (Gamma density; linear and squared mean specification)
Authors' suggestions: FMM with two gamma densities perform better than OLS and GLM models investigated. Adding a third component reduces bias at the cost of poorer individual predictions. A small number of components is likely to be sufficient if one starts with reasonable first approximation to the true data density. *Sample size implications as discussed by the authors*: Large sample sizes needed for good convergence (above 15000-20000). Given the advances in computing, the computational burden of finite mixture models should not be a problem.
Connections to other papers: Blough1999, Deb1997, Deb2000, Manning2001
Review Comment *Skewness, heavy tails, multimodality:* Finite mixture models have flexibility to explicitly model skewness, heavy tails and multimodality. *Adjustment for covariates/ Extensions to cost-effectiveness*: Finite mixture models have the ability to adjust for covariates. The extension to joint modelling of costs and effects for the purpose of cost-effectiveness analysis does not look straightforward. *Sample size implications:* Such mixture models appear to require medium to large sample sizes for reliable estimation.

Reference: **Dinh2006 (Dinh & Zhou 2006)**
Category: Non-parametric methods
Parameter(s): Mean cost, mean outcome

Data: Skewed cost, outcome data *Model*: Edgeworth expansions for the studentised t-statistic to reduce the effect of skewness and estimation of confidence intervals for incremental cost-effectiveness ratio, net health benefit. Three forms of expansion (based on three functions) are studied. *Statistical method(s)*: Algebraic approach based on means, variances, covariances and sample sizes. *Implementation*: Not stated *Representation of estimation error*: 95% confidence intervals.
Applied data: 1 dataset; *Sample size*: 107 *Simulated data*: Bivariate normal, bivariate mixture, bivariate lognormal; *Sample size*: 100 *Comparisons with other models/methods*: Taylor's confidence interval, Fieller's confidence interval, the bootstrap percentile confidence interval and the bootstrap bias-corrected acceleration confidence interval
Authors' suggestions: The CIs based on Edgeworth expansions with the third function have comparable coverage and are narrower than the previously recommended methods for evaluation of confidence intervals. *Sample size implications as discussed by the authors*: Not discussed.
Connections to other papers: Zhou2005
Review Comment *Skewness, heavy tails, multimodality*: The approach aims to adjust for skewness and improve the precision in the estimation of cost-effectiveness. It is unclear how to select the appropriate form of expansion (three methods were employed and one of them systematically performed worse than comparative methods). *Adjustment for covariates*: The proposed approach does not allow for adjustment for covariates. *Extensions to cost-effectiveness*: The approach works with cost-effectiveness statistic and its components and reports uncertainty related to cost-effectiveness. *General comments*: Any improvements over other methods appear to be slight, particularly for inference about net benefits.

Reference: **Dominici2005A (Dominici et al. 2005), Dominici2005B (Dominici & Zeger 2005)**
Category: Non-parametric methods
Parameter(s): Mean cost difference
Data: Skewed cost data with likely imbalance of sample sizes between the two populations. *Model*: Smooth quantile ratio estimation (SQUARE)- a nonparametric model of mean cost difference that employs smoothing of the log-transformed ratio of the two quantile functions through basis functions. Two-part SQUARE- an extension of the smooth quantile ratio estimation (SQUARE) method to account for zero costs and covariates. Propensity score approach is used to establish matched propensity score strata (for each case), estimate the fractions of non-zero costs and evaluate SQUARE for positive costs for each stratum. The overall mean cost difference is estimated by averaging across the strata. *Statistical method:* Ordinary least squares for the estimation of contributions of the basis functions. The overall mean cost difference is a linear combination of order statistics with weights estimated from data. A modified nearest neighbour matching algorithm for propensity scores. *Implementation*: R software, program code could be requested from the authors *Representation of estimation error*: Asymptotic variance
Applied data: 1 dataset; *Sample size*: Positive cost: 118 vs 2262; 112 vs 980; All costs: 188 vs 9228; 165 vs 4682 *Simulated data*: Lognormal/Lognormal, Smooth function of Lognormal/Lognormal, Gamma/Gamma, Sampling from empirical distribution of cost data, sampling from a two-part

linear regression model of log-transformed costs; *Sample size*: total 100 vs 100; 100 vs 1000; and within strata 50 vs 125; 25 vs 125; 25 vs 50; 50 vs 50. *Comparisons with other models/methods:* Smooth quantile ratio estimation with estimated degrees of freedom; with 2 and 4 degrees of freedom; Smooth quantile ratio estimation assuming lognormal populations; maximum likelihood estimator under log-normal model; sample mean difference; For two-part variant: weighted sample mean difference within each stratum; Maximum likelihood estimation based on two-part lognormal model within propensity score matched strata and averaging; Maximum likelihood estimation based on two-part linear regression model for the log transformed costs to account for covariates.
Authors' suggestions: Smooth quantile ratio estimation has lower mean squared error than sample mean difference and the log-normal parametric estimator. The efficiency in SQUARE is achieved by borrowing strength across the two samples to learn about the shape of distribution but this could lead to small biases. Two-part regression SQUARE produces most efficient estimates of mean conditional cost differences compared to non-parametric and parametric alternatives studied. The estimation of overall cost difference based on SQUARE is more variable than the MLE obtained under a two-part log-linear model. *Sample size implications as discussed by authors*: Results are qualitatively similar between samples with smaller and larger sizes in simulations.
Connections to other papers: Duan1983, Duan1983a, Lipscomb1998, Manning2001, Mullahy1998, Zhou1997, Zhou1997, Zhou1997a
Review Comment *Skewness, heavy tails, multimodality*: SQUARE is a potentially flexible approach to account for skewness and heavy tails when comparing mean costs across two samples but further research is needed to address possible biases in real data applications. *Adjustment for covariates/ Extensions to cost-effectiveness*: Extensions to account for covariates (by propensity score matching) and zero costs available in Dominici2005B but extensions to cost-effectiveness that acknowledge correlations of costs and outcomes do not seem straightforward. *Sample size implications:* The method is likely to need at least moderate sample sizes in total, although numbers within each matching strata can be quite small. The use of propensity score matching to covariate adjustment entails the need for large samples overall. Also sufficient data within each strata is needed for the estimation of cost differences conditional on covariates. *General comments:* The approach seems a useful approach for moderate to large datasets but its use in usual samples in clinical trials may be more limited.

Reference: **Dow2003 (Dow & Norton 2003)**
Category: Two-part models
Parameter(s): Mean cost, mean resource use
Data: Total cost, resource use with zeros (25%, 50%, 75% zeros) *Model*: Two-part model (TPM) and limited information sample selection (Heckit) model to account for excess zeros. *Statistical method(s)*: ML. Tests for choosing between TPM and Heckit: the inverse Mill's coefficient, an adaptation of the Toro-Vizcarrondo and Wallace empirical mean square error test. *Implementation*: Not stated *Representation of estimation error*: Monte Carlo simulation based SD of estimators over 200 samples
Simulated data: Heckman sample selection model; *Sample size*: 1000 *Comparisons with other models/methods*: Full information sample selection model
Authors' suggestions: TPM is more appropriate than Heckit for modelling actual costs. The Toro-Vizcarrondo and Wallace empirical mean square error are better than the inverse Mill's

coefficient. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Ai2000, Duan1983, Duan1984, Gilleskie2000, Leung1996, Manning1987, Manning1998, Mullahy1998, Mullahy2001
Review Comment *Skewness, heavy tails, multimodality*: Not addressed (assumes normality for positive costs). Two-part and sample selection models explicitly model the large proportion of zeros. *Adjustment for covariates*: The methods allow for adjustment for covariates. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and health outcomes does not seem straightforward. *General comments*: Useful paper setting out the differences between Heckit and other two part models.

Reference: **Duan1983 (Duan et al. 1983) Duan1983A (Duan 1983)**
Category: Two-part models
Parameter(s): Mean cost
Data: Healthcare costs (inpatient, other healthcare costs) *Model*: Two-part model: probit (probability of any healthcare cost) and linear model on log scale (to model cost conditional on having incurred any), four-part model: probability of any healthcare expenses; probability of hospital expenses given any expenses; healthcare expenses given any expenses but not hospital expenses, healthcare expenses given any expenses and hospital expenses. A nonparametric smearing estimate of the expected cost on the untransformed scale after fitting a linear regression model on the transformed scale (homoscedastic and heteroscedastic variants suggested). *Statistical method(s)*: Maximum likelihood, non-parametric smearing estimate, GLS and Newton-Raphson algorithm to account for within-family correlation. *Implementation*: Not stated *Representation of estimation error*: Mean squared prediction error in validation sample.
Applied data: 1 dataset; RAND Health Insurance Study; *Sample size*: 8,765 person-years of data *Comparisons with other models/methods*: ANOVA, ANCOVA (OLS), OLS on log-transformed expenditure+$5.
Authors' suggestions: Two-part model and four-part model better than other models. The simulation experiment did not lend support for one of them but the authors believe that the four-part model will fare better when more data becomes available as it is consistent with respect to both non-users and users of inpatient services. Two-part model is shown to be inconsistent (due to the proportion of inpatient services users) and it is argued that more data would not change that but will improve the precision of the four-part model. *Sample size implications as discussed by the authors*: The availability of more data would allow better modelling of the right tail of cost distribution.
Connections to other papers:
Review Comment *Skewness, heavy tails, multimodality*: The multi-part models are likely to fit better multimodal data with high proportion of zeros. *Adjustment for covariates*: The multi-part models have the capability to adjust for covariates. The smearing estimate need to account for heteroscedasticity if such present. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and outcomes does not seem straightforward.

Reference: **Duan1984 (Duan et al. 1984)**
Category: Two-part and Tobit models
Parameter(s): Mean cost

Data: Total cost *Model*: Sample selection (SSM: adjusted Tobit model) and two-part model (TPM) *Statistical method(s)*: SSM (Mill's ratio limited-information method, Full-information maximum likelihood method); TPM (normal theory or smearing estimate) *Implementation*: Not stated *Representation of estimation error*: Mean squared prediction error in cross-validation sample.
Applied data: 1 dataset; *Sample size*: Not stated *Comparisons with other models/methods*: One-part model, ANCOVA, ANOVA
Authors' suggestions: TPM has better performance in RAND HIS data; and generally for modelling actual (vs. potential) outcomes. SSM has poor statistical and numerical properties. Argues that TPM for actual expenditures can have correlated parts. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Duan1983
Review Comment *Skewness, heavy tails, multimodality*: The two-part models are likely to fit better data with high proportion of zeros. *Adjustment for covariates*: The SSM and multi-part models have the capability to adjust for covariates. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and outcomes does not seem straightforward.

Reference: **Dudley1993 (Dudley et al. 1993)**
Category: Comparative study of methods
Parameter(s): Mean cost
Data: Positive cost data *Model*: (1) ordinary least squares multiple linear regression, (2) OLS after logarithmic transformation of the dependent variable, (3) Weibull proportional hazards model, and (4) Cox semi-parametric proportional hazards model. Binary logistic regression, model also included but not applicable for evaluation of mean costs. *Statistical method(s)*: Standard methods *Implementation*: SAS *Representation of estimation error:* Predictive accuracy of the mean defined as the percentage deviation of predicted from actual cost.
Applied data: 1 dataset; *Sample size:* 155 patients undergoing coronary artery bypass grafting surgery. *Comparisons with other models/methods:* None
Authors' suggestions: The study demonstrates that the Cox proportional hazards model is valuable in the analysis of skewed censored data. The proportional hazards assumption was met. The percentage deviation of predicted from actual cost was lowest for the Cox model 14.3%, followed by the log-transformed OLS model 16.6%. *Sample size implications as discussed by the authors*: Not discussed.
Connections to other papers: Austin2003, Basu2004, Cooper2003, Lipscomb1998
Review Comment *Skewness, heavy tails, multimodality*: The authors have shown that accounting for the skewness and the tail in data by more flexible models leads to more precise results in their dataset. The superior performance of the Cox proportional hazards model is likely to be due to the fact that the proportional hazards assumption was met. The authors acknowledge that simulation studies are needed to establish what models are suitable for different circumstances. The comparison of the methods could be limited by the use of predictive accuracy only. *Adjustment for covariates/ Extensions to cost-effectiveness*: The suggested methods adjust for covariates. Extensions to cost-effectiveness are not likely to be straightforward for models

beyond the standard OLS. *Sample size implications*: All methods are likely to be applicable to reasonably large datasets. The tests for the underlying assumptions could have limited power in small datasets.

Reference: **Faddy2008 (Faddy et al. 2009)**
Category: Markov chain methods
Parameter(s): Mean resource use
Data: Skewed resource use data without zeros (length of stay in hospital) *Model*: Phase-type distribution or a finite Markov chain in continuous time. The model has quadratic variance-mean relationship. The effects of covariates on mean length of stay is multiplicative. *Statistical method:* Maximum likelihood estimation. Maximum likelihood , Bayesian Information Criteria, generalised Person criteria, residual quantile- quantile plots and coverage properties were used to assess model fit. *Implementation*: A software tool developed using MATLAB. *Representation of estimation error*: Standard errors of estimates of covariate effect provided.
Applied data: A dataset of patients' lengths of stay in hospital and covariates ; *Sample size*: 1901 *Comparisons with other models/methods:* Gamma and log-normal models.
Authors' suggestions: Phase-type distributions are shown to provide better fit to data than gamma and log-normal models, because they are better able to accommodate extreme values. *Sample size implications as discussed by authors*: Not discussed.
Connections to other papers: Basu2004, Manning2001, Marshall2003 & Marshall 2007
Review Comment *Skewness, heavy tails, multimodality:* The phase-type distribution provide increased flexibility in modelling skewed, heavy tailed data. Within sample performance of the model is superior to lognormal and gamma models. It is not clear whether overfitting is likely to present substantial problems for the model. *Adjustment for covariates/ Extensions to cost-effectiveness*: The method allows for adjustment for covariates. Further research is needed in modelling more than one resource use component, overall costs and cost-effectiveness is needed. *Sample size implications:* The robustness of the approach will depend strongly on having sufficient data to estimate well model parameters. *General comments:* A flexible approach to modelling separate episodes of care in hospital. Further work is needed to see whether the framework is practically extendable to more complex healthcare resource use patterns.

Reference: **Gilleskie2004 (Gilleskie & Mroz 2004)**
Category: Non-parametric methods
Parameter(s): Mean cost
Data: Total healthcare cost *Model*: Use of a sequence of conditional probability functions to construct a discrete approximation to the density function of an outcome conditional on exogenous explanatory variables. *Statistical method:* ML; Simultaneous optimisation for the number of intervals and the polynomial specification of conditional densities. *Implementation*: STATA; program code could be requested from the authors. *Representation of estimation error*: Standard errors evaluated through bootstrapping the whole procedure.
Applied data: 1 dataset; *Sample size*: Not stated *Simulated data*: 5 data generating processes; *Sample size*: 1000-5000 *Comparisons with other models/methods:* OLS, GLM (two-part: first part probit and gamma

density with log link).

Authors' suggestions: This approach estimates the entire distribution and allows effects of covariates to be different at different points of support in the distribution.
Sample size implications as discussed by the authors: Larger sample size better; the method probably is unsuitable for small to average sample sizes.

Connections to other papers: Blough1999, Manning1998, Manning2001, Mullahy1998

Review Comment
Skewness, heavy tails, multimodality: In theory the method could model explicitly skewness, heavy tails, multimodality as it estimates the entire distribution. In practice these advantages of method are likely to be realised only when large datasets are available.
Adjustment for covariates/ Extensions to cost-effectiveness: The method allows for different covariate effects over particular ranges of the outcome. The extension to joint modelling of costs and effects for the purpose of cost-effectiveness analysis does not look straightforward.
Sample size implications: It is likely that large samples are needed to model well the distribution.

Reference: **Grootendorst1995 (Grootendorst 1995)**

Category: Two-part models (count data)

Parameter(s): Mean count

Data: Count data with zeros
Model: Two-part model: probit model for the first part and negative binomial model for the second part.
Statistical method(s): Maximum likelihood
Implementation: Not stated
Representation of estimation error: Standard errors implied

Applied data: 1 dataset; *Sample size*: male 2644 (1959 with any cost; 26% non-users), female 3099 (2495 with any cost, 19% non-users)
Comparisons with other models/methods: Poisson, Negative binomial, Zero inflated probit and negative binomial.

Authors' suggestions: The two-part model dominated the Poisson, negative binomial and the "zero altered" negative binomial models.
Sample size implications as discussed by the authors: Not discussed

Connections to other papers: Duan1983, Gurmu1997

Review Comment
Skewness, heavy tails, multimodality: The suggested two part model accommodates better the high proportion of zeros and the heavy right tail present in the data.
Adjustment for covariates/ Extensions to cost-effectiveness: Adjustment for covariates implemented. Extensions to cost-effectiveness not straightforward.

Reference: **Gurmu1997 (Gurmu 1997)**

Category: Two-part models (count data)

Parameter(s): Mean resource use

Data: Healthcare resource use (visits to doctors, health centre visits)
Model: Semi-parametric estimation of hurdle (two-part) count regression models (nests Poisson and Negative binomial hurdle models).
Statistical method(s): Quasi maximum likelihood estimator: Laguerre series expansion estimators within the maximum likelihood framework.
Implementation: GAUSS, http://www2.gsu.edu/~ecosgg/spcount.html
Representation of estimation error: Asymptotic standard errors implied

Applied data: 2 datasets; *Sample size*: 485 (49.7% non-users), 511 (33.1% non-users)
Comparisons with other models/methods: Standard Poisson and negative binomial models.

Authors' suggestions: The author recommends the semi-parametric hurdle (two-part) count

regression models for analysis of overdispersed data with high proportion of non-users and highly skewed distribution of counts for users. Semi-parametric estimation avoids the misspecification in the distribution of unobserved heterogeneity within the hurdle framework. Hurdle models provide better fit for count data with large proportion of zeros and relatively heavy tails. *Sample size implications as discussed by the authors*: In limited samples with large proportion of zeros, it is not practical to estimate the second part of the hurdle model. Conclusions should be treated with caution given the small sample size. The number of Laguerre polynomials is chosen to minimise the AIC, and hence will tend to be fewer for smaller sample sizes.
Connections to other papers: Cameron1988, Grootendorst1995, Pohlmeier1995
Review Comment *Skewness, heavy tails, multimodality*: The two-part hurdle models are likely to fit better data with high proportion of zeros. *Adjustment for covariates*: The models have the capability to adjust for covariates. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and outcomes does not seem straightforward.

Reference: **Gurmu1998 (Gurmu 1998)**
Category: Two-part models (count data)
Parameter(s): Mean counts
Data: Overdispersed or underdispersed count data with zeros *Model*: Generalised hurdle model with first binary choice part generalized logistic model, allowing for asymmetric departures from the binary logit model, and second part either truncated-at-zero-double Poisson regression model or Negative binomial model. *Statistical method(s)*: Maximum likelihood *Implementation*: Not stated *Representation of estimation error:* Asymptotic standard errors implied
Applied data: 1 dataset; *Sample size*: 485 (41% non-users) *Comparisons with other models/methods:* None reported
Authors' suggestions: Generalized hurdle models perform well in terms of the fitted frequencies. The generalized Negbin hurdle provides the most satisfactory fit among the models considered for the application of interest. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Gurmu1997, Gurmu2000, Mullahy1997, Mullahy1998, Pohlmeier1995
Review Comment *Skewness, heavy tails, multimodality*: *Adjustment for covariates/ Extensions to cost-effectiveness*: Adjustment for covariates allowed for but extensions to modelling cost-effectiveness are not straightforward.

Reference: **Gurmu2000 (Gurmu & Elder 2000a;Gurmu & Elder 2000b)**
Category: Data component models
Parameter(s): Mean count
Data: Correlated count data *Model*: Generalised Bivariate Negative Binomial model with dependence between count variables introduced by means of stochastically related unobserved heterogeneity components. Squared polynomial series expansions are used for the unknown joint density of the unobserved heterogeneity components to allow for both positive and negative correlation between the counts. This semiparametric model nests parametric models such as the multivariate negative binomial model. *Statistical method(s)*: Expansions based on Laguerre polynomials, maximum likelihood. *Implementation*: GAUSS program at http://www2.gsu.edu/~ecosgg/spcount.html

Representation of estimation error: Asymptotic standard errors implied
Applied data: 2 datasets; *Sample size*: 485; 5190 *Comparisons with other models/methods:* Poisson, Negative Binomial, Bivariate Poisson, Bivariate Negative Binomial
Authors' suggestions: Bivariate semiparametric estimation methods for multivariate count regression models are shown to outperform univariate and other bivariate models in the empirical applications considered. Monte Carlo simulation to evaluate model performance compared to other models is needed. Extensions to truncated or censored multivariate count data given. *Sample size implications as discussed by the authors*: Models with more than two components are computationally demanding and a simplified model of the unobserved heterogeneity could be needed.
Connections to other papers: Cameron1997, Cameron1998
Review Comment *Skewness, heavy tails, multimodality*: *Adjustment for covariates/ Extensions to cost-effectiveness:* Adjustment for covariates allowed for but extensions to modelling cost-effectiveness are not straightforward. *Sample size implications*: *General comments*: The review is based on 2000a reference (unpublished paper); 2000b reference seems to be a simplified version of 2000a.

Reference: **Hahn2003 (Hahn & Whitehead 2003)**
Category: Data component models
Parameter(s): Mean cost, mean resource use
Data: Cost and resource use components of trial data, total cost and total resource use. *Model*: Four models are employed: 1) total cost is assumed to be normally distributed; 2) total costs is modelled through 3 normally distributed independent cost components, 3) total costs is modelled through 3 normally distributed correlated cost components (multivariate normal), 4) model 3 with relaxed assumption on normality (some log-normally distributed components, proportional odds model for resource use component). *Statistical method(s)*: Bayesian methods; MCMC Gibbs sampler; non-informative priors used *Implementation*: BUGS, program code not available *Representation of estimation error*: Posterior distribution, Cost-effectiveness acceptability curve
Applied data: 1 dataset; *Sample size*: 373 *Comparisons with other models/methods*: None
Authors' suggestions: Bayesian approach has a natural link with decision making and facilitates modelling the combination of diverse information that accounts for the uncertainty in the parameters. Methods which are robust against model misspecification need to be explored. The main disadvantage is that it requires specification of prior distributions for unknown parameters that could influence the posterior distribution. *Sample size implications as discussed by the authors*: Choosing suitable models for costs and effects is restricted by small sample sizes.
Connections to other papers: Lambert2008, O'Hagan2003
Review Comment *Skewness, heavy tails, multimodality*: The methods are likely to facilitate modelling of skewness, heavy tails and multimodality in data if extended beyond the multinormal distribution. *Adjustment for covariates*: Although all methods were illustrated without adjustment for covariates they are extendible to adjust for covariates. *Extensions to cost-effectiveness*: All models are designed to evaluate simultaneously costs and effects and cost-effectiveness.

General comments: In the non-Normal model, the correlations between cost components are apparently ignored.

Reference: **Hill2009 (Hill & Miller 2009)**
Category: Single-distribution generalized linear models
Parameter(s): Mean cost, marginal cost
Data: Positive cost data with heteroscedasticity, skewness and kurtosis *Model*: Generalized Gamma and Extended Estimating Equations models *Statistical method*: Full Information Maximum likelihood (Generalized Gamma); Extended Quasi-likelihood methods (Extended Estimated Equations models) *Implementation*: Not stated *Representation of estimation error*: Cross validation- bias, predictive accuracy, bias in marginal effect
Applied data: 4 datasets; *Sample size*: 28 579, 12 547, 22 011, 11 671 *Comparison with other models/methods*: Compared with linear OLS, log OLS with homoscedastic smearing factor, Gamma and Poisson.
Authors' suggestions: The EEE model is a robust estimator for health expenditures that minimises bias and prediction errors, while minimising data overfitting. The method performs as well, or better, than the other models in each distribution; and unlike linear OLS fits well across the distribution. This makes this model particularly suitable when the aim is the calculation of mean marginal effects. Poisson and linear OLS also fit data well in terms of bias and prediction errors both overall and for people with specific chronic conditions. *Sample size implications*: Larger samples needed for fitting the Extended estimating Equations.
Connections to other papers:, Basu2005a, Basu2006, Blough1999, Buntin2004, Deb2003, Duan1983, Manning1998, Manning2001, Manning2005, Mullahy1998
Review Comment *Skewness, heavy tails, multimodality:* Generalized Gamma and Extended Estimating Equation models both present flexibility in modelling skewed, heavy tailed data. The latter is more robust to misspecifications related to modelling unknown distributions. Both methods are used for strictly positive costs. *Adjustment for covariates:* It has the advantages of the GLM approach generally, primarily the ability to model covariates flexibly. *Extensions to cost-effectiveness*: Extensions to joint modelling of costs and utilities are not straightforward for both methods. *Sample size implications:* Large sample sizes needed to fit these models well (particularly the Extended Estimating Equations). These sample sizes are unlikely to be available in clinical trials. *General comments:* Both methods do not accommodate explicit modelling of mass at zero. Convergence problems with the Extended Estimating Equations are observed even with large datasets.

Reference: **Hollenbeak2005 (Hollenbeak 2005)**
Category: Models based on normality following a transformation of the data
Parameter(s): Hospital ranking by expected cost
Data: Skewed positive cost data *Model*: Explores the functional form of hospital cost data by fitting Box-Cox model with random coefficients using MCMC methods. Box-Cox transformation of -1 seems to be the most appropriate. The natural log transformation is rejected. *Statistical method(s)*: Bayesian methods; MCMC: Metropolis within Gibbs framework *Implementation*: GAUSS, program code not available *Representation of estimation error*: Posterior distribution

Applied data: 1 dataset, patients undergoing a coronary artery bypass graft surgery; *Sample size*: 496 patients from 4 hospitals *Comparisons with other models/methods*: Linear, log-linear models
Authors' suggestions: Flexible method that takes into account clustering. For the data use Box-Cox transformation of "-1" was shown to be more appropriate than natural log transformation. The method is more complex than simpler methods and further exploration is needed of benefits in the presence of more clusters. *Sample size implications as discussed by the authors*:
Connections to other papers:
Review Comment *Skewness, heavy tails, multimodality*: Box-Cox transformation increases the flexibility of modelling skewness in data compared to the log transformation but can not accommodate multimodality and could not be sufficiently flexible in presence of heavy tails. *Adjustment for covariates*: The methods allows for adjustment for covariates. *Extensions to cost-effectiveness*: By using Bayesian methods the approach is potentially extendible to modelling simultaneously costs and effects in order to evaluate cost-effectiveness.

Reference: JimenezMartin2002 (Jimenez Martin et al. 2002)
Category: Two-part models and models based on mixtures of parametric distributions
Parameter(s): Mean resource use
Data: Healthcare resource use- visits to general practitioners and visits to specialists. *Model*: Two-part models (TPM: probit model for the probability of incurring visit(s) and a truncated Negative Binomial model for the number of visits conditional of incurring any) and latent class models (LCM or finite mixture model of Negative binomials) *Statistical method(s)*: Maximum likelihood *Implementation*: Not stated *Representation of estimation error*: Standard error
Applied data: 1 dataset (2 variables; 12 countries; separate analyses by variable and country); *Sample size*: 1613 to 13764 (visits to General practitioner; 15-62% non-users); 1613 to 13759 (visits to specialists; 16-83% non-users) *Comparisons with other models/methods*: None
Authors' suggestions: Latent class models more appropriate for estimating general practitioners utilisation while two-part models are more appropriate for visits to the specialists. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Cameron1986, Cameron1998, Deb1999, Deb2000, Pohlmeier1995
Review Comment *Skewness, heavy tails, multimodality*: Latent class models are likely to be more flexible in presence of multimodality and heavy tails. *Adjustment for covariates*: Both methods allow for covariate adjustment. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and outcomes does not seem straightforward.

Reference: **Keeler1988 (Keeler et al. 1988)**
Category: Two-part models
Parameter(s): Mean resource use and mean costs
Data: Skewed data with zeros on use and costs of mental health services (RAND Health Insurance Experiment) *Model*: A Weibull survival model evaluating hazard of starting outpatient mental health treatment within the year and a log linear model of costs, conditional on incurring any with

explanatory variables including hazard index, remaining time to the end of the year, personal characteristics, insurance plan and deductibles, etc.
Statistical method: Maximum likelihood estimation, adjusted for family and site clustering.
Implementation: Not stated
Representation of estimation error: Standard errors or t-ratios presented

Applied data: 1 dataset; *Sample size*: Data on 5,809 individuals over up to 3 years; 555 person-years with outpatient mental health spending
Comparisons with other models/methods: None

Authors' suggestions: A two-part specification of a survival time model for starting treatment and loglinear model on costs of treatment conditional on incurring any is suitable in situations where treatments are persistent and span over large period of time in distinct contacts with providers (such as mental treatment episodes).
Sample size implications as discussed by authors: Not explicitly discussed but issues with overlapping parameter estimates likely due to large standard errors identified.

Connections to other papers:

Review Comment
Skewness, heavy tails, multimodality: The model specification is particularly suitable for the situation intended (rare use with somewhat high related costs) when quantity of episodes is less informative. Lognormal costs model is suggested to not model well the occurrence of very large costs in a related manuscript by Keeler and Rolph.
Adjustment for covariates/ Extensions to cost-effectiveness: Adjustments for covariates are incorporated and are integral part of model specification. Extensions to cost-effectiveness (relating both costs and health outcomes) is likely to present a major further complication.
Sample size implications: Reasonable sample sizes are needed to provide efficient estimates.
General comments: A type of two-part model that is likely useful in applications.

Reference: **Lambert2008 (Lambert et al. 2008)**

Category: Data components models

Parameter(s): Mean cost, mean cost of a resource use component, mean survival and cost-effectiveness

Data: Individual patient data on total costs consisting of four cost components (cost data with possible zero observations) and survival times.
Model: 1) Modelling cost components assuming multivariate normality; 2) Incorporation of a two-part model: a logistic regression plus a linear regression for one of the cost components. Conditional univariate distributions were used to model the multivariate nature of the relationship between cost components.
Statistical method(s): Bayesian methods: Markov Chain Monte Carlo (MCMC).
Implementation: WinBUGS
Representation of estimation error: Posterior distribution.

Applied data: 1 dataset; *Sample size*: 115
Comparisons with other models/methods: None

Authors' suggestions: A general and flexible approach that allows better modelling of cost components and their relationship. The multivariate normality assumption is likely to be rejected in the case of skewed cost data and other distributions should be investigated. Another approach would be to model units of resource use but this would increase the computational complexity of the model substantially.
Sample size implications as discussed by the authors: not discussed.

Connections to other papers: Hahn2003

Review Comment
Skewness, heavy tails, multimodality: Modelling separate cost components could provide with a better overall model of the mean total cost but assuming multivariate normality is likely to ignore the skewness, heavy tails and multimodality in the components.

Adjustment for covariates: The model allows adjustment for covariates to be implemented. *Extensions to cost-effectiveness*: The model evaluates outcomes as well as costs and provides inference for cost-effectiveness. *Sample size implications*: *General comments*: The authors suggest using alternative conditional distributions instead of the normal, to overcome the very limiting assumption of multivariate normality. However, unlike the multivariate normal, such structures are not invariant to the sequence in which the conditional distributions are constructed. Hence such models would have restrictive new assumptions.

Reference: **Leung1996 (Leung & Yu 1996)**
Category: Two-part models
Parameter(s): Mean cost
Data: Healthcare cost data (25%, 50%, 75% zeros) *Model*: Sample-selection model (SSM: adjusted Tobit model) and two-part models (TPM: probability of incurring cost is modelled as a probit model; the cost, conditional on incurring any, is modelled as a linear model on the log-transformed costs). *Statistical method(s)*: SSM-limited-information maximum likelihood (LIML) or full-information maximum likelihood (FIML);TPM-Maximum likelihood and least squares. *Implementation*: GAUSS, program code not available *Representation of estimation error*: Mean prediction bias; mean squared prediction error; parameter bias, parameter squared errors; elasticity bias; elasticity squared errors
Simulated data: Sample selection or two-part model with single regressor; *Sample size*:1000 *Comparisons with other models/methods*: None
Authors' suggestions: The two models perform well under different conditions; SSM susceptible to colinearity problems; t-test can differentiate between both methods when no colinearity problems persist. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Duan1984, Maddala1985, Manning1987
Review Comment *Skewness, heavy tails, multimodality*: The two models are likely to perform similarly under the conditions of skewness, heavy tails, and multimodality. *Adjustment for covariates*: Both methods allow for covariate adjustment. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and outcomes does not seem straightforward.

Reference: **Lipscomb1998 (Lipscomb et al. 1998)**
Category: Comparative study of models
Parameter(s): Mean cost
Data: 20% of Medicare enrolees hospitalised for ischemic stroke in 1991. *Model*: Single-equation linear models of untransformed cost, single-equation linear model with log transformation of cost and Duan's smearing estimator to adjust for bias, two-part (mixture) models without log transformation, two-part (mixture) models with log transformation and Duan's smearing estimator to adjust for bias, Cox proportional-hazards model stratified by time intervals. *Statistical method(s)*: Standard methods; Split sample comparison of predictive abilities of different models. *Implementation*: Not specified. *Representation of estimation error:* Bootstrap standard errors.
Applied data: 1 dataset; *Sample size*: 41823 separated in model training (20939) and model validation (20884) samples. *Comparisons with other models/methods:* None

Authors' suggestions: The log-transformed linear model, log-transformed two-part model and the Cox proportional hazards model perform similarly in terms of mean prediction and outperform the untransformed models. The quality of cost prediction for a given dataset depends on the selected model, and therefore, the predictive validity of the models should be the criteria for choosing among models. *Sample size implications as discussed by the authors*: Large sample sizes are needed to be able to use the split sample analysis of predictive ability.
Connections to other papers: Dudley1993
Review Comment *Skewness, heavy tails, multimodality*: It is shown that the log transformed single and two-equation models and the Cox proportional-hazards model outperform the untransformed models. While the better performance of the log-transformed models is likely to be due to partial accountance for skewness and mass at zero, the Cox proportional hazards model is likely to perform well because the underlying assumptions of proportional hazards are met. *Adjustment for covariates/ Extensions to cost-effectiveness*: The suggested methods adjust for covariates. Extensions to cost-effectiveness are not likely to be straightforward. *Sample size implications*: Split sample cross validation does need a large initial sample size, and therefore, could be of limited application in usually small to moderate clinical trials. *General comments*: While authors claim to use Duan's smearing factor they in fact just use the appropriate estimate for the mean of a lognormal distribution. The use of a series of validity measures RMSE, MAE and log score shows that some utility measures can distinguish between models and some can not.

Reference: **Maddala1985 (Maddala 1985)**
Category: Two-part models
Parameter(s): Mean cost
Data: Total cost *Model*: Sample selection and self-selection models (SSM) and two-part models (TPM) *Statistical method(s)*: Not stated *Implementation*: Not stated *Representation of estimation error*: Not stated
A discussion of results of published studies with a focus on selectivity models. *Comparisons with other models/methods*: none
Authors' suggestions: A-more careful structural formulation of the sample-selection models is likely to lead to more useful policy results; "otherwise there is no need to depart from mother OLS". *Sample size implications as discussed by the authors*: not discussed
Connections to other papers: Duan1984
Review Comment *Skewness, heavy tails, multimodality*: The two models are likely to perform similarly under the conditions of skewness, heavy tails, and multimodality. *Adjustment for covariates*: Both methods allow for covariate adjustment. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and outcomes does not seem straightforward.

Reference: **Manning1987 (Manning et al. 1987)**
Category: Two-part models
Parameter(s): Mean
Data: Data with a cluster at zero (simulated with bivariate normal correlation between parts) *Model*: Sample selection models (SSM: the adjusted Tobit model) and two-part models (TPM: probability of incurring cost is modelled as a probit model; the cost, conditional on incurring any, is modelled as a linear model on the log-transformed costs).

Statistical method(s): SSM –limited information maximum likelihood (LIML), full-information maximum likelihood (FIML) estimators, TPM-naïve (true specification, omitting correlation coefficient) and a data-analytic variant (the best specification of the conditional equation is searched for during the analysis). *Implementation*: Not stated *Representation of estimation error*: Bias, mean squared error
Simulated data: Data with one or two uncorrelated regressors drawn randomly from a uniform distribution and errors drawn from a bivariate normal distribution; *Sample size*: 1000 *Comparisons with other models/methods*: none
Authors' suggestions: TPM are better in the absence of exclusion restrictions (i.e. exclusions or the exact specification of the right-hand side variables) but can perform poorly for extreme values of independent variable. With exclusion restrictions SSM are better than naïve TPM, but negligibly better than the data-analytic version. The data-analytic version of the two-part model is robust if the interest is in the response surface. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Duan1983, Duan1984
Review Comment *Skewness, heavy tails, multimodality*: The two models are likely to perform similarly under the conditions of skewness, heavy tails, and multimodality. *Adjustment for covariates*: Both methods allow for covariate adjustment. *Extensions to cost-effectiveness*: The extension of the models to model simultaneously costs and outcomes while acknowledging the likely correlation between them does not seem straightforward.

Reference: **Manning2001b (Manning & Mullahy 2001)**
Category: Comparative study of models
Parameter(s): Mean cost; Mean resource use
Data: Positive costs or resource use data *Model*: (1) OLS on ln(y) with homoscedastic or heteroscedastic retransformation; (2) GLM with additive error term and variance independent of x; (3) GLM with a log link and variance proportional to E(y/x) (Poisson with overdispersion); (4) GLM with a log link and standard deviation proportional to E(y/x) (Gamma); (5) Non-linear least squares by GLM with a log link, and an additive homoscedastic error term. *Statistical method(s):* Standard OLS and GLM methods. *Implementation*: Implementation code unavailable but standard in most statistical software. *Representation of estimation error*: Standard errors.
Applied data: 2 datasets; *Sample size*: 27598, 3618 *Simulated data*: lognormal, gamma, heavy-tailed distribution on log-scale, heteroscedastic in x on log-scale; *Sample size*: 10000
Authors' suggestions: No single model best under all alternatives. If bias is only concern use GLM. If precision is of concern, OLS provides the most precise estimates. A procedure to choose between OLS on ln(y) and GLM (Poisson, Gamma, Gaussian with log link) for skewed data, with heavy tails, heteroscedasticity, etc is proposed. In general: if the log-scale residuals are heavy-tailed (coefficient of kurtosis>3) OLS on ln(y) is recommended; if the coefficient of kurtosis is about 3 or less use Park test on raw residuals to choose a GLM. *Sample size implications as discussed by the authors*: not discussed.
Connections to other papers: Blough1999, Duan1983, Manning1998, Mullahy1998
Review Comment *Skewness, heavy tails, multimodality:* Neither method could accommodate multimodality. The inference of GLM (using estimating equations or quasi-likelihood) is approximate and GLMs do not explicitly model skewness, heavy tails or multi-modality. *Adjustment for covariates*: All methods allow for adjustment for covariates.

Extensions to cost-effectiveness: Extensions of the models to model both costs and utilities does not seem straightforward. *General comments*: An algorithm for deciding which model to use is provided to guide analysts.

Reference: **Manning2005 (Manning et al. 2005)**
Category: Single-distribution generalized linear models
Parameter(s): Marginal effect on mean cost, mean length of stay
Data: Positive cost or resource use data. *Model*: Models using the three parameter Generalised Gamma distribution: a model with constant variance, a model with variance a linear function of x; and a model for heteroscedastic and heavy tailed data; includes OLS, OLS for log normal, gamma and exponential models with log link, and Weibull as special cases. *Statistical method(s):* Full information maximum likelihood method. *Implementation*: STATA; program code could be requested from the authors. *Representation of estimation error*: Robust standard errors using analog of Huber/White correction of the variance/covariance matrix.
Applied data: 1 dataset; *Sample size*: 6500 *Simulated data*: Homoscedastic and heteroscedastic log-normal data, heavy tailed, log normal, Gamma and Weibull data; *Sample size*: 10000 *Comparisons with other models/methods:* OLS regression of ln(y) with homoscedastic smearing factor for the retransformation, Gamma GLM model for y with a log link function, Weibull model for y.
Authors' suggestions: The three parameter Generalised Gamma model is a very flexible general model. The tests for identifying distributions are robust. Standard Generalised Gamma is not consistent when data generating mechanism is the heteroscedastic model for ln(y) and should be adapted to handle heteroscedasticity; careful examination for linearity, functional form, and the link function needed. *Sample size implications as discussed by the authors*: Not discussed.
Connections to other papers: Basu2004, Basu2005a, Manning1998, Manning2001
Review Comment *Skewness, heavy tails, multimodality:* The model extends the gamma model and the two additional parameters of the Generalized Gamma model partially determine the skewness and kurtosis of the distribution. By extending the set of included models, the Generalised Gamma model improves the ability to improve precision in presence of heavy tails compared to standard GLM. The inference of GLM (using estimating equations or quasi-likelihood) is approximate and if the true distribution is not a generalized gamma, the estimated parameters will be inconsistent. The method is suitable for positive response data. *Adjustment for covariates*: The method allows for adjustment for covariates. *Extensions to cost-effectiveness*: Extensions of the model to model both costs and utilities does not seem straightforward.

Reference: **Marazzi1998 (Marazzi et al. 1998), Marazzi1999 (Marazzi & Ruffieux 1999), Marazzi2002 (Marazzi 2002), Marazzi2003 (Marazzi & Barbati 2003), Marazzi2004 (Marazzi & Yohai 2004)**
Category: Methods based on data trimming
Parameter(s): Mean hospital length of stay mean cost
Data: Length of stay data for hospitalisations classified in Diagnostic Related Groups; Weibull, Gamma and Lognormal data contaminated with Uniform data; costs of hospitalisations *Model*: A robust estimation based on parametric distributions (Lognormal, Weibull, Pareto and Gamma) and truncation or trimming of data. Adaptive truncation procedure is suggested

to assure data is not rejected when generated according the model. *Statistical method(s)*: M-estimators, location-dispersion estimators, constrained maximum-likelihood estimation, non-parametric exponential tilting and semi-parametric approaches, bootstrap. *Implementation*: S-Plus; http://www.iumsp.ch/Unites/us/Alfio/msp_programmes.htm *Representation of estimation error:* Available.
Applied data: A number of datasets on length of stay and hospitalisation costs; *Sample size*: from few tens to hundreds *Comparisons with other models/methods:* Classical ML estimators
Authors' suggestions: Truncated means are computationally and statistically attractive estimators. Estimators based on models with moderate tails (e.g. Weibull) offer better protection than those based on long tailed models (e.g. Lognormal). Estimates with high breakdown points for setting trimming limits are to be preferred to the interquartile range. Nonparametric models provide less stable null distributions (w.r.t. outliers) than parametric models but parametric models tend to underestimate real fluctuations by ignoring contamination. Extensions to more complex parametric models (e.g. generalised Gamma and F distributions) could be straightforward as soon as corresponding robust estimators are available. *Authors' suggestions*: Both the truncated ML (TML) and adaptively truncated ML (ATML)-estimates attain a much higher efficiency than the initial S-estimate and the variances of the ATML are closer to the maximum likelihood values when sample size increases. However, the efficiency gain provided by adaptive versus fixed truncation does not appear to be appreciable unless sample size is large (more than 200). *Sample size implications as discussed by the authors*: A third of the samples did not associate with any of the models due to excessive sample size or early peak in distribution (at 1-3 days). The ATML seems to outperform TML-estimates in large sample sizes only.
Connections to other papers:
Review Comment *Skewness, heavy tails, multimodality:* The performance of the models is fully determined by the robustly fitted parametric distributions to the data. Different estimators within the class are likely to lead to moderate differences in applied examples. The ATML estimators seem to provide few advantages over the TML-estimates and to do so they would reject data that does not conform to the model. By ignoring parts of the data within the model selection it is unclear how the model selection could inform on the most appropriate model for the mean. *Adjustment for covariates/ Extensions to cost-effectiveness*: Earlier models did not extend to covariate adjustment and cost-effectiveness but covariate adjustment is implemented in Marazzi2004. Generally, the extendibility to covariate adjustment and joint modelling of costs and effects will depend on extendibility of the robust estimators to adjust for covariates and joint modelling. *Sample size implications:* Some of the methods are used in small samples but it is unclear how robust estimates could be achieved in small samples. *General comments:* It is unclear how ignoring data could lead to better inference for the mean. The method discards data considered contaminated which is inappropriate. The general work in this area would be good in applications where contamination is a realistic assumption.

Reference: **Marshall2003 & Marshall 2007 (Marshall et al. 2007;Marshall & McClean 2003)**
Category: Markov chain method
Parameter(s): Mean resource use, mean total cost
Data: Skewed resource use data without zeros (length of stay in hospital) *Model*: Coxian phase-type distribution or a finite Markov chain in continuous time and a cost

model. Costs are attached to different phases and total costs derived from the overall model. Bayesian belief network suggested as a way to incorporate adjustment for covariates through conditional phase-type distribution. *Statistical method:* Maximum likelihood estimation. Bayesian information criteria used for model selection (eg. number of phases). *Implementation*: A software tool developed using MATLAB. *Representation of estimation error*: Variance is estimated for the total costs.
Applied data: Two datasets of length of stay in hospital; *Sample size*: 1392, 4722 *Comparisons with other models/methods:* None
Authors' suggestions: Coxian phase-type distribution is good for modelling skewed continuous data and the extention of this model to estimate overall cost provides an opportunity to evaluate effects of policy changes before their implementation. *Sample size implications as discussed by authors*: Not discussed.
Connections to other papers: Lipscomb1998, Nixon2004
Review Comment *Skewness, heavy tails, multimodality:* The Coxian phase-type distribution seems to provide flexibility in modelling skewed, heavy tailed data. Clear application to length-of-stay type studies with extension to cost through 'expert opinion' of the unit cost of each phase. *Adjustment for covariates/ Extensions to cost-effectiveness*: Conditional (on covariates) phase-type distribution is proposed. Extension to cost-effectiveness seems possible but further research is needed in this area. *Sample size implications:* The robustness of the approach will depend strongly on having sufficient data to estimate well model parameters. *General comments:* A flexible approach to modelling separate episodes of care in hospital. Further work is needed to see whether the framework is practically extendable to more complex healthcare resource use patterns.

Reference: **Mullahy1997 (Mullahy 1997)**
Category: Models based on mixtures of parametric distributions
Parameter(s): Mean count data
Data: Resource use- number of consultations with a doctor *Model*: Poisson mixture model. *Statistical method(s)*: Goodness-to-fit tests for mixture alternatives to null probability models are suggested. *Implementation*: Not stated; program code not available *Representation of estimation error:* Not stated
Applied data: 1 dataset; *Sample size*: 5190 *Comparisons with other models/methods:* Poisson, Negative Binomial, Zero inflated Poisson
Authors' suggestions: Mixture models outperform Poisson model in presence of excess zeros due to unobserved heterogeneity and are recommended. *Sample size implications as discussed by the authors*: Not discussed.
Connections to other papers: Cameron1986, Deb1997, Duan1983, Gurmu1996
Review Comment *Skewness, heavy tails, multimodality:* Mixture models are likely to perform well in the case of heavy tails, skewness and multimodality. *Adjustment for covariates/ Extensions to cost-effectiveness*: The models adjust for covariates. *Sample size implications*: Larger sample sizes are needed to evaluate mixture models. *General comments*: Overdispersion and excess zeroes are related phenomena that arise commonly with resource use data, in which case a Poisson distribution is not appropriate for modelling such data. This paper argues that a mixture of Poisson distributions may be more suitable than a zero-inflated Poisson distribution or two-part (hurdle) model, and provides tests that are claimed to distinguish better the presence of a mixture model than existing tests

of overdispersion.

Reference: **Mullahy1998 (Mullahy 1998)**
Category: Two-part models
Parameter(s): Mean resource use
Data: Resource use- number of doctor visits *Model: A* modified two-part model (MTPM). To model the positive variable y (e.g. resource use), the first part models the probability that y>0 (as in conventional two-part models). The second part models the conditional (on both covariates x and the fact that y is positive) expectation of y, whereas two-part models have more often modelled that of log y. *Statistical method(s)*: One-step non-linear least squares (NLLS) and two-step (logit/probit and NLLS) estimators, exponential conditional mean (ECM): GMM-type estimator. *Implementation*: Not stated *Representation of estimation error:* Standard heteroscedasticity-robust covariance estimators.
Applied data: 1 dataset; *Sample size*: 36111 *Simulated data*: Lognormal data with heteroscedastic error in the covariate; *Sample size*: 5000 *Comparisons with other models/methods:* Two-part models (TPM) using Duan's homoscedastic smearing estimator to back transform, One part models which model the expectation of y given x (not conditioning on y>0), such as the ECM model.
Authors' suggestions: Modified TPM performed better than ECM and TPM and the two-step estimator performed better than the one-step. ECM could be a reasonable choice when interest is in the mean. Some bias-robustness trade-offs are likely due to thick upper tails and/or outliers but these are beyond the scope of this paper. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Duan1983, Duan1983a, Manning1987, Mullahy1997
Review Comment *Skewness, heavy tails, multimodality:* ECM is likely to fail in the presence of a large number of zeros, heavy tails and/or multimodality. Two-part models generally are attractive in dealing with the extra skewness associated with excess zeroes but do not address heavy tails (and the kinds of skewness that this induces) or multi-modality. The MTPM is a useful addition to the available TPMs. *Adjustment for covariates/ Extensions to cost-effectiveness*: The extensions of the MTPM to multiple dependent variables and cost-effectiveness do not seem straightforward. *Sample size implications*: *General comments*: Mullahy points out technical difficulties with the conventional TPM when the smearing factor depends on x. Rather than assume a functional form for this and have to estimate it separately, Mullahy assumes a functional form for the expectation of y that implicitly subsumes this smearing function. In practice, the conventional TPM with a (possibly) non-constant smearing function is more flexible but more complicated to use (with the result that most applications have ignored the possibility of non-constant smearing).

Reference: **Munkin1999 (Munkin & Trivedi 1999)**
Category: Data components models
Parameter(s): Mean resource use (counts)
Data: Correlated resource use data *Model*: Bivariate Poisson-Lognormal mixture (BVPLN) and Bivariate Negative Binomial (BVNB) regression models. Correlated multiplicative unobserved heterogeneity terms are introduced individually into each conditional mean function. The resulting joint density of two count variables is a bivariate Poisson-lognormal or bivariate Poisson-gamma mixture. *Statistical method*: Simulated ML estimator; first-order correction of the asymptotic bias introduced; antithetic sampling used to reduce variance. *Implementation*: Not stated; code for fitting not available

Representation of estimation error: Asymptotic variance based on robust sandwich method.

Applied data: 1 dataset; *Sample size*: 4406
Simulated Data: Yes two correlated Poisson variables ; *Sample size*: 1000
Comparisons with other models/methods: None

Authors' suggestions: The proposed model and simulated maximum likelihood estimation are flexible with respect to distributional assumptions on the bivariate unobserved heterogeneity. Other distributional assumptions for heterogeneity could be used, such as the inverse Gaussian, and even cases when exact ML estimator exists, provided that negligible errors due to numerical approximation are assured. Antithetic sampling and adjustment for first-order simulation bias improve the accuracy of the model. BVNB model fits the data better than the BVPLN in the current application.
Sample size implications: Not discussed

Connections to other papers: Deb1997

Review Comment
Skewness, heavy tails, multimodality: The methods could accommodate a degree of skewness in data.
Adjustment for covariates: Implemented
Extensions to cost-effectiveness: Authors claim that the model can be easily extended to more count variables by mixing conditional Poisson densities.

Reference: **Nixon2004 (Nixon & Thompson 2004)**

Category: Single-distribution generalized linear models, models based on skewed distributions outside the GLM family

Parameter(s): Mean costs

Data: Total cost
Model: Empirical comparison of Normal, Gamma (2- and 3-parameter versions), Log-normal (2- and 3-parameter versions), Log-logistic (2- and 3-parameter versions) models.
Statistical method(s): Maximum likelihood and MCMC to fit models. Frequentist and Bayesian deviance measures to compare models.
Implementation: WinBUGS
Representation of estimation error: Parametric bias corrected and accelerated (BCa) bootstrap, and Bayesian credible intervals.

Applied data: 4 datasets; *Sample size*: 335, 80, 58, 43
Comparisons with other models/methods: The paper is an empirical comparison of above models.

Authors' suggestions: The assumption of a single parametric form could be misguided. Models that fit badly can give similar inferences to those that fit well. Conversely, parametric models that fit data equally well can lead to substantially different inferences. Sensitivity of results to model choice needs to be investigated. Bootstrapping could be deficient because the asymptotic assumptions are not fulfilled.
Sample size implications as discussed by the authors: Sample size needs to be sufficient to model tail of distribution. The authors suggest the use of rules to determine whether the sample size is sufficient to rely on asymptotic normality.

Connections to other papers: Blough1999, Manning2001, O'Hagan2003

Review Comment
Skewness, heavy tails, multimodality: While some distributions are better in fitting the skewness, heavy tails, and multimodality in data this does not guarantee improved estimation of the population mean.
Adjustment for covariates/ Extensions to cost-effectiveness: The authors suggest extensions based on joint cost-effectiveness distribution but without developing the idea. These extensions are likely to be of limited flexibility.
Sample size implications: Larger sample sizes would lead to better fit of data with parametric

models but it is unclear whether the gained precision is sufficient to lead to better estimators than the sample mean.

Reference: **Nixon2009 (Nixon et al. 2009)**
Category: Non-parametric methods
Parameter(s): Mean Incremental Net Benefit
Data: Cost data that is potentially skewed. *Models*: Central Limit Theorem, Non-parametric bootstrap methods (percentile bootstrap, normal bootstrap) *Statistical method:* Standard methods. *Implementation*: Excel *Representation of estimation error*: Standard error and coverage
Applied data: 1 datasets; *Sample size*: 136/ 119; Simulated data; *Sample size*: 10 to 1000 *Comparisons with other models/methods:* None
Authors' suggestions: In moderate to large samples (>50) and in small samples with low skewness both methods accurately estimated the standard errors. In small datasets with high skewness, the Central Limit Theorem produced more accurate SEs. *Sample size implications as discussed by authors*: As above
Connections to other papers: Briggs2005, O'Hagan2003
Review Comment *Skewness, heavy tails, multimodality:* Impact of skewness on estimates of standard error and skewness explored but no interest in efficient estimates is pursued. *Adjustment for covariates*: Not studied. The authors admit that this will entail adding parametric assumptions. *Extensions to cost-effectiveness*: the study focuses directly on estimating incremental net benefits (a measure of cost-effectiveness). *Sample size implications:* Both methods perform well in moderate to large datasets. The Central limit Theorem slightly outperformed the bootstrap approaches in small datasets with large skewness. *General comments:* Both methods are shown to produce valid estimates in varied situations of different sample sizes and skewness but do not address approaches to improve efficiency in the estimates. It is unclear what the performance of the methods is in the presence of large proportion of zeros, multimodality. Heavier tails than allowed in the simulations are likely to occur in practice more often than suggested by the authors.

Reference: **O'Hagan2003 (O'Hagan & Stevens 2003)**
Category: Models based on the normal distribution, non-parametric methods and parametric models based on skewed distributions outside the GLM family
Parameter(s): Mean costs
Data: Single- and two-sample skewed total cost data *Model*: Sample mean and bootstrap mean to analyse skewed cost data are compared to alternative parametric models (i.e. employing a log-normal distribution) that explicitly model skewness in data. *Statistical method(s)*: Normal-theory and bootstrap methods, Bayesian inference *Implementation*: Frequentist and Bayesian approaches *Representation of estimation error:* Based on Student *t*-distribution, standard bootstrap, Bayesian bootstrap, Bayesian parametric (log-normal distribution)
Applied data: 1 dataset; *Sample size*: 26 (one arm), 120 (whole study) *Comparisons with other models/methods:* None
Authors' suggestions: Bootstrap and methods based on asymptotic normality may lead to inefficient and even misleading inferences. Methods that recognise the skewness in data need to be applied. Bayesian non-parametric methods that incorporate prior knowledge on the

underlying true distribution could have potential.
Sample size implications as discussed by the authors: Modest sample sizes exacerbate concerns with employing methods based on normal distribution (normal-theory methods, bootstrap approaches).

Connections to other papers: Barber2000, Briggs1998, Thompson2000

Review Comment
Skewness, heavy tails, multimodality: Parametric models model better skewness, heavy tails and multimodality in the sample data but their use for estimation of population mean needs to be well justified.
Adjustment for covariates/ Extensions to cost-effectiveness: Adjustment for covariates is not implemented but is straightforward. Extensions to cost-effectiveness are straightforward based on normal-theory and bootstrap methods but could be with limited flexibility in the case of parametric models.
General comments: Although the authors clearly feel that using lognormal distributions is generally better, Briggs2005 showed that the log-normal estimator could perform badly when the underlying distribution is not log-normal. Nixon2004 also did not show clearly much advantage in parametric modelling. Neither nonparametric approaches (e.g. asymptotic normal theory, bootstrap) nor parametric models will reliably produce good results in all situations. Probably parametric models should be used when there is good reason to choose a particular parametric form, or where the data are extensive enough to identify the distribution adequately.

Reference: **Pagano2008 (Pagano et al. 2008)**

Category: Non-parametric methods

Parameter(s): Mean costs

Data: Cost data
Model: Aalen regression model with additive hazard function.
Statistical method: Not stated
Implementation: R software
Representation of estimation error: Variance

Applied data: 1 dataset; *Sample size*: 458
Simulated data: 2 lognormal and 2 gamma distributed datasets with one continuous or one continuous and one binary covariates; with and without 5% and 20% contamination or censoring; *Sample size*: 20-5000.
Comparisons with other models/methods: OLS model; Gamma regression model with log link.

Authors' suggestions: The Aalen model is performing well and can be a reasonable alternative to the standard Gamma regression model. It also provides additional information on the relationships between costs and covariates.
Sample size implications as discussed by authors: Aalen model had substantially larger biases in small datasets with contamination compared to Gamma regression model and substantially smaller biases in datasets with heavy censoring.

Connections to other papers: Basu2004, Dudley1993

Review Comment
Skewness, heavy tails, multimodality: The Aalen approach cannot accommodate zero cost data. When near-zero contamination is present and in highly skewed data the Aalen approach has unacceptable bias in small samples.
Adjustment for covariates/ Extensions to cost-effectiveness: Adjustment for covariates incorporated but seems to worsen the overall performance of the method. Increased flexibility in studying relationship between covariate and cost is reported. Extensions to cost-effectiveness possible but unlikely to be easily implemented.
Sample size implications: Bias can remain substantial in highly skewed data even in large

datasets (for example n=1000).
General comments: The Aalen method is not shown to outperform greatly the Gamma regression model in different circumstances (even in presence of censoring).

Reference: **Pohlmeier1995 (Pohlmeier & Ulrich 1995)**
Category: Single-distribution generalized linear models, two-part models
Parameter(s): Mean count
Data: Count data with zeros *Model*: Negative binomial distributed (Negbin) hurdle (two-part) model: with negative binomial distribution (with variance proportional to the mean) for both the binary and frequency parts. This specification allows for over-dispersion at the individual level. *Statistical method(s)*: Maximum likelihood *Implementation*: Not stated *Representation of estimation error:* Asymptotic standard errors implied
Applied data: 1 dataset; *Sample size*: 5096 *Comparisons with other models/methods:* Poisson model, Negbin model, Poisson hurdle model
Authors' suggestions: Contact and frequency decisions in the application studied are governed by different stochastic processes and need to be modelled separately. Ignoring these differences leads to inconsistent parameter estimates and to economic misinterpretations. Future research should target development and application of appropriate hurdle models for panel data and enable separation of individual differences in health endowments from physician-induced demands for healthcare visits. *Sample size implications as discussed by the authors*: The coefficients for the first part are more precisely determined due to larger sample size and better descriptors.
Connections to other papers: Cameron1986
Review Comment *Skewness, heavy tails, multimodality*: *Adjustment for covariates/ Extensions to cost-effectiveness*: Adjustment for covariates allowed for but extensions to modelling cost-effectiveness are not straightforward. *General comments*: The negative binomial model for the first part is not identifiable, and so reduces to a Poisson model.

Reference: **SantosSilva2001 (Santos-Silva & Windmeijer 2001)**
Category: Two-part models
Parameter(s): Mean resource use
Data: Skewed data with zeros *Model*: Multi-episode two-part model: a Stopped-sum negative binomial distribution is used to model two-part decision process when multiple spells are likely. The number of spells is modelled as Poisson with variance proportional to mean and number of visits per spell as logarithmic distribution. A further specification allows for unobserved and correlated heterogeneity in both the Poisson and logarithmic distributions. Test of single spell hypothesis proposed. *Statistical method:* Non-parametric maximum likelihood estimator employing expectation maximisation algorithm. Maximum likelihood can be used if the stopped-sum distribution is fully specified. *Implementation*: Not specified *Representation of estimation error*: Standard errors for parameter estimates of the two stages.
Applied data: 1 dataset; *Sample size*: 5096 *Comparisons with other models/methods:* Two-part (hurdle model) with binary and volume parts based on Negative binomial distributions with variances proportional to mean.
Authors' suggestions: The two-part decision process governing implies that the total demand

for health care follows a stop-sum distribution. This new approach requires only a minimal set of assumptions and is robust. *Sample size implications as discussed by authors*: Rich datasets are needed so the gain in robustness is not at a high price in terms of efficiency.
Connections to other papers: Deb1997, Cameron1986, Grootendorst1995, Gurmu1998, Gurmu2000a, Gurmu2000b, Mullahy1998, Phlmeier1995
Review Comment *Skewness, heavy tails, multimodality:* The approach have similar flexibility to hurdle models and in current application outperforms the hurdle models in terms of Schwartz criterion. *Adjustment for covariates/ Extensions to cost-effectiveness*: Adjustments for covariates are incorporated. Extensions to cost-effectiveness could present a major further complication to be resolved in terms of identification. *Sample size implications:* Large sample sizes are needed in order not to affect efficiency. *General comments:* As authors note this stop-sum distribution approach is useful if the interest is in studying the two stages in the decision process but less so if the interest is in studying the overall demand as such.

Reference: **Thompson2000 (Thompson & Barber 2000)**
Category: Methods based on the normal distribution, non-parametric methods
Parameter(s): Mean costs
Data: Two-sample skewed cost data in clinical trials *Model*: *t*-test, non-parametric bootstrap *Statistical method(s)*: Standard methods *Implementation*: Not stated *Representation of estimation error:* 95% confidence intervals
Applied data: 2 datasets; *Sample size*: 148, not stated *Comparisons with other methods:* *t*-test on log-transformed data, Mann-Whitney U test
Authors' suggestions: *t*-test and non-parametric bootstrap recommended. Validity of *t*-test in small samples or extremely skewed data should be checked by bootstrap techniques. Non-parametric tests or those based on transformed data do not address the arithmetic mean and should be avoided. *Sample size implications as discussed by the authors*: Small samples could limit performance of methods based on the normal distribution. Moderately large sample size needed so that *t*-test is valid.
Connections to other papers: Barber2000, Briggs1998, O'Hagan2003
Review Comment *Skewness, heavy tails, multimodality:* Both the t-test and the non-parametric bootstrap are likely to underperform in small datasets and/or highly skewed data. *Adjustment for covariates/ Extensions to cost-effectiveness*: Extensions to covariate adjustment and cost-effectiveness are relatively straightforward but do not resolve the limitations of the methods in terms of accounting for skewness, heavy tails, multimodality. *Sample size implications*: Larger sample sizes are needed to secure the validity of methods based on the normal distribution in the case of highly skewed data. *General comments*: O'Hagan and Stevens challenged the claim that bootstrapping could provide any check on the validity of the t-test.

Reference: **Thompson2005a (Thompson & Nixon 2005)**
Second review: Yes T4
Category: Single-distribution generalized linear models, Parametric models based on skewed distributions outside the GLM family
Parameter(s): Mean costs, mean cost difference, cost-effectiveness
Data: Cost and outcome data from a clinical trial

Model: Normal, gamma and lognormal distributions for cost data; Bivariate normal, Normal-gamma and Normal-lognormal distributions for effectiveness and costs in cost-effectiveness analyses with allowed correlations between effects and costs. *Statistical method(s)*: Bayesian methods *Implementation*: MCMC in BUGS *Representation of estimation error:* Posterior distributions
Applied data: 1 dataset; *Sample size*: 380 *Comparisons with other models/methods:* None
Authors' suggestions: Conclusions of analysis are sensitive to the choice of distribution (in particular model of the upper tail). Sensitivity analyses with alternative distributions should address uncertainty in the choice of distribution. *Sample size implications as discussed by the authors*: Large datasets or external data could inform on the shape of the right tail of distribution.
Connections to other papers: Barber2000, O'Hagan2001, Nixon2004, Deb2003, Zhou2002
Review Comment *Skewness, heavy tails, multimodality:* Parametric models of cost data highly increase flexibility of modelling skewness, heavy tails, multimodality in data. The distributions considered here can account for skewness and heavy tails (although not multimodality). They could fit the data better than the normal distribution, but the impact of the shape of the tail of distribution beyond the observed data could be large and needs to be further investigated. *Adjustment for covariates/ Extensions to cost-effectiveness*: Adjustments for covariates were not implemented but are straightforward to implement. The model was extended to cost-effectiveness with costs and outcomes at individual level correlated. The way this correlation is modelled though could also have different impact on cost-effectiveness and should be investigated. *General comments*: Larger datasets facilitate parametric modelling but these models are less needed.

Reference: **Tooze2002 (Tooze et al. 2002)**
Category: Two-part models
Parameter(s): Mean cost
Data: Repeated measures data with clumping at zero. *Model*: Mixed-distribution model for repeated measure data with clumping at zero and correlated random effects. Logistic model for the occurrence of positive value, lognormal model for the size of the positive value and bivariate normal model for the random effects in the two preceding models. *Statistical method(s)*: Generalised linear and nonlinear mixed-effects models, mixed models. *Implementation*: PROC GENMOD and PROC NLMIXED in SAS (a new SAS macro MIXCORR could be requested from the authors). *Representation of estimation error:* Only standard errors of parameter estimates are reported.
Applied data: Medical Expenditure Panel Survey from 1996; *Sample size*: 22601 individuals in 10596 households *Simulated data*: Logistic-lognormal mixed distribution data with correlated random effects; *Sample size*: 100 units with cluster size of 7. *Comparisons with other models/methods:* None
Authors' suggestions: The model and methods of fitting gives unbiased results for both fixed and random effects. The model allows for the estimation of the mean amount, including the probability of zeros. Extensions of the model to different error distributions for the two parts, different methods (non-parametric, semi-parametric) for fitting this type of data and additional random effects are possible subjects for future research. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Manning1987

Review Comment *Skewness, heavy tails, multimodality*: The model allows for the modelling of correlated data with excess zeros and possibly skewed distribution of the nonzero values and allows for correlation between the probability of positive and the size of the positive values. *Adjustment for covariates/ Extensions to cost-effectiveness*: The model incorporates adjustment for covariates. Extensions to modelling multiple outcomes and cost-effectiveness are possible but would further complicate the model. *General comments*: Assumes cost distributions similar across time points. Extensions to other transformations in addition to log suggested by authors as possible.

Reference: **van de Ven & van Praag1981 (van de Ven & Van Praag 1981)**
Category: Tobit models
Parameter(s): Mean costs
Data: Skewed data on costs of hospitalisations and treatment by specialists with zero values. *Model*: Adjusted Tobit model: if unobserved index of demand for health care less than zero then the observed annual expenditures are zero; if non-negative then the logarithm of the expenditureis modelled as a linear function of covariates. The index is modelled as a separate linear function of covariates. Different explanatory variables could contribute to these two models. The error terms are modelled as bivariate normal model with no need for the error in the non-zero expenditure part to be truncated as in classical Tobit model. *Statistical method:* A procedure proposed by Heckman (1979) (Heckit/LIML). *Representation of estimation error*: T-ratios for parameter estimates
Applied data: 1 dataset; *Sample size*: Data on 6,068 families; *Comparisons with other models/methods:* None
Authors' suggestions: This adjusted Tobit model has a number of advantages over the standard Tobit model including possibility for different explanatory variables and parameter estimates in the two parts and untruncated error terms. *Sample size implications as discussed by authors*: Not discussed
Connections to other papers: Leung1996, Maddala1985, Manning1987
Review Comment *Skewness, heavy tails, multimodality:* The adjusted Tobit model might present advantages over the two-part model specification by explicitly modelling correlation between the model equations (the two decisions (a) to consume healthcare and (b) how much to consume). Whether this leads to more efficient estimators for mean costs is not clear given its susceptibility to collinearity problems (Heckman's correction term is not statistically significant in the applied example). *Adjustment for covariates/ Extensions to cost-effectiveness*: Adjustments for covariates are incorporated and different covariates could be specified for each equation (similar to two-part model and different from standard Tobit model). Extensions to cost-effectiveness (relating both costs and health outcomes) in a common analytical framework is likely to present a problem.

Reference: **Veazie2003 (Veazie et al. 2003)**
Category: Models based on normality following a transformation of the data
Parameter(s): Mean costs
Data: Skewed cost data with zeros *Model*: Linear regressions on expenditure (untransformed model) and square root of expenditure (square root model). The least squares residual on the square root scale a linear model of the covariates to enable estimation of the variance in the error term on the square root scale for each case. *Statistical method(s)*: Weighted least squares using sampling weights *Implementation*: Not stated

Representation of estimation error: Robust estimation of variance-covariance matrix
Applied data: 1 dataset; *Sample size*:~7000-8000 *Comparisons with other models/methods:* Log-transformed and two-part models proved problematic and were not reported.
Authors' suggestions: The model of the square root of expenditures performs better than a model of untransformed expenditures in terms of both mean forecast squared error (precision) and overfit (generalisability). Mixed results regarding bias: the untransformed model performs better in most categories, whereas the square root model performs better in the most severe category. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Blough1999, Buntin2004, Duan1983, Manning1998, Manning2001, Mullahy1998
Review Comment *Skewness, heavy tails, multimodality:* For the dataset used in this paper the square root transformation of the dependent variable has improved the model performance for the heavier users of health care and the overall precision and generalisability. The performance of the square root model in general is likely to be largely data dependent. *Adjustment for covariates/ Extensions to cost-effectiveness*: Both models adjust for covariates. Extension of the square root model to modelling multiple outcomes and cost effectiveness should allow for appropriate presentation of the error terms and is likely not to be straightforward.

Reference: **Wang2003 (Wang 2003)**
Category: Data components models
Parameter(s): Mean resource use
Data: Overdispersed resource use data with zeros-the number of doctor visits and the number of non-doctor health professional visits *Model*: Bivariate zero-inflated negative binomial regression model (BZINB) *Statistical method(s)*: Expected maximisation and quasi-Newton algorithms *Implementation*: Not stated *Representation of estimation error:* Standard errors (t-ratio presented)
Applied data: 1 dataset; *Sample size*: 5190 *Comparisons with other models/methods:* Bivariate Negative Binomial (BIVARNB); Generalized Bivariate Negative Binomial (GBIVARNB)
Authors' suggestions: BZINB model can be used to model and lend insight into the source of excess zeros and overdispersion for two dependent variables of event counts. In the current application, the BZINB model fits the data better than either the BIVARNB or the GBIVARNB. Further research will address the development of a score test for the proposed model against alternative of the BIVARNB. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Cameron1988, Gurmu1998, Gurmu2000
Review Comment *Skewness, heavy tails, multimodality:* The proposed model extends the bivariate negative binomial model to account for the excess zeros in the count data and is likely to suit more the healthcare utilisation data. Extensions to modelling cost data across more than two categories is likely to be restricted. *Adjustment for covariates/ Extensions to cost-effectiveness*: The method includes adjustment for covariates. Extensions to cost-effectiveness are possible in some restricted cases but do not appear straightforward in the general case.

Reference: **Wang2010 (Wang & Zhou 2010)**
Category: Non-parametric methods

Parameter(s): Conditional mean cost given covariates and mean cost difference
Data: Cost data with zeros and possibly heteroscedasticity *Model*: Regressin of quantiles of cost data for the poitive cost, logistic regerssion for probability of incurring cost. Monotone transformation of positive costs. *Statistical method(s)*: Interior point algorithm *Implementation*: Rq function in R package *Representation of estimation error:* Bootstrap standard errors
Applied data: Inpatient medical costs included by Department of Veterans Affairs for the 1999; *Sample size*: 3902 *Simulated data:* Cost data with zeros and lognormal positive costs with //without different forms of heteroscedasticity*;* Sample size: 100, 500 *Comparisons with other models/methods:* Maximum likelihood estimator, the generalized linear model estimator (Blough& Ramsay), Welsh&Zhou internal weighted estimator, smooth quantile ratio estimator (Dominici, 2005).
Authors' suggestions: Simulation studies indicate that the regression of the quantiles of costs has competitive (better or competitive efficiency in different situations) and more robust performance than existing estimators in various heteroscedastic models. *Sample size implications as discussed by the authors*: Smaller mean square errors achieved for both sample sizes of 100 and 500 compared to the other methods in the paper.
Connections to other papers: Blough1999 **,**Buntin2004, Dominici2005A, Dominici2005B, Duan1983, Duan1983A, Mullahy1998, Welsh2006, Zhou2001
Review Comment *Skewness, heavy tails, multimodality:* This distribution-free approach should cope well with skewness and heavy tails. *Adjustment for covariates/ Extensions to cost-effectiveness*: The approach allows for adjustment for covariates within quantiles. *Sample size implications:* As the approach has been compared to models with similar data requirements, its performance was competitive even in limited sample sizes. *General comments:* Wider comparative work of the performance of this approach with the other methods identified in the current review might be informative.

Reference: **Welsh2006 (Welsh & Zhou 2006;Zhou & Cheng 2008)**
Category: Two-part models
Parameter(s): Mean costs
Data: Total cost- skewed, heteroscedastic data with zeros *Model*: Two-part model. A transformation is applied to positive responses to achieve linearity but probably not normality and homoscedasticity. Heteroscedasticity and non-normality were handled explicitly. Bias adjusted back-transformation for the positive part was developed through extensions of Duan's smearing estimator: the "externally" weighted estimator and the "internally" weighted estimator. *Statistical method(s)*: Estimating equations and non-parametric adjustment for non-normality and/or heteroscedasticity on the original scale. *Implementation*: Not stated *Representation of estimation error:* Standard error
Applied data: 1 dataset; *Sample size*: 1785 *Simulated data:* Data with zeros and heteroscedasticity in covariates ; *Sample size:* 130, 150, 200, 500, 1000 *Comparisons with other models/methods:* None
Authors' suggestions: The two nonparametric estimators of the mean on the original scale accommodate the zeros, the heteroscedasticity and the possible non-normality of the positive part of the two-part model. They are consistent and asymptotically normal. If the transformation used in the second part is estimated from data the asymptotic variance of

estimators would increase. *Sample size implications as discussed by the authors*: Not discussed
Connections to other papers: Zhou2002
Review Comment *Skewness, heavy tails, multimodality*: The flexibility to accommodate skewness, heavy tails and multimodality is fully determined by the flexibility modelled in the two-part model. *Adjustment for covariates/ Extensions to cost-effectiveness*: The methods proposed adjust for covariates. Extensions to cost-effectiveness are not straightforward.

Reference: **Winkelmann2004 (Winkelmann 2004)**
Category: Two-part models
Parameter(s): Mean difference as percentage reduction
Data: Skewed resource use data with zeros *Model*: A hurdle model with a probit model for the hurdle and a truncated Poisson- lognormal model for positive outcomes. *Statistical method:* Numerical quadrature. *Implementation*: Not stated *Representation of estimation error*: Standard errors.
Applied data: 1 dataset ; *Sample size*: 32,837 (5 occasions for c.6500 people) *Comparisons with other models/methods:* Poisson, Negative Binomial, Poisson-log-normal, Hurdle negative binomial, two-component negative binomial finite mixture, Poisson-logarithmic multi-episode model.
Authors' suggestions: The Probit-Poisson-lognormal model outperforms the other models in the data studied in terms of better Schwartz model fit statistic and Vuong's test for model selection. *Sample size implications as discussed by authors*: Not discussed.
Connections to other papers: Deb1997, SantosSilva2001
Skewness, heavy tails, multimodality: The Probit-Poisson-lognormal model is a two-part model that accounts for excess zero and has increased flexibility in fitting skewed heavy tailed data. *Adjustment for covariates/ Extensions to cost-effectiveness*: Adjustment for covariates is incorporated but extensions to cost-effectiveness could be difficult. *Sample size implications:* Similar to other two-part models. *General comments:* Successful fitting of such a complex model in this case study may be down to the very large sample size available. Also the specialised context (longitudinal data with latent clustering) is unlikely to be relevant to clinical trial data.

Reference: **Zhou2002 (Zhou 2002), a review paper of** Zhou1997 (Zhou et al. 1997b), Zhou1997A (Zhou & Gao 1997), Zhou1997B (Zhou et al. 1997a), Zhou1998 (Zhou 1998), Zhou1999 (Zhou & Tu 1999), Zhou2000 (Zhou & Tu 2000), Zhou2000A (Zhou & Gao 2000), Zhou2001 (Zhou et al. 2001a), Zhou2001A (Zhou et al. 2001b), Tu1999 (Tu & Zhou 1999), Tu2000 (Tu & Zhou 2000)
Category: Comparative study of methods
Parameter(s): Mean costs, Mean cost' difference
Data: Costs data with heteroscedasticity and possibly additional zero values *Models*: Methods based on log transformed data. This article reviews methods for inference of population means of cost data with heteroscedasticity and possibly additional zero values in the follow situations: 1) One population without zeros: point estimation (Zhou1998: sample mean, ML estimator, uniformly minimum variance unbiased estimator (UMVU), conditionally minimal MSE estimator), interval estimation (Zhou1997A, Zhou2000A: bias-adjusted method following an initial log-transformation of outcome; interval based on

Edgeworth expansion of standard t-statistic) 2) One population with extra zeros: Nonzero values lognormally distributed and the number of zero observations follows a Binomial distribution; point estimate and confidence intervals (Zhou2000: a bootstrap method, a likelihood ratio method) 3) Two populations without zeros: hypothesis tests (Zhou1997: t-test on original and log transformed data, bootstrap, Z-score, Wilcoxon test; Zhou2000B: bootstrap-based and ML-based approaches, Zhou2001: Z-score, LR test, Z-score with jack-knife, bootstrap approaches) 4) Three or more populations with extra zeros: hypothesis testing (Zhou1999): LR test; multiple comparisons (Tu2000): bootstrap approach 5) Regression models: two-part method: probit and linear regression on log-transformed positive costs (Duan1983); linear regression model with heteroscedasticity for positive costs (Zhou2001A) *Statistical method(s):* ML, MSE, bootstrap-based approaches *Implementation*: Not stated *Representation of estimation error*: population(s) without additional zeros: bias-adjusted after log-transformation, Edgeworth expansion, ML-based approach, bootstrap,etc.; population(s) with additional zeros: bootstrap, likelihood ratio method
Applied data: 5 datasets employed to illustrate different situations; *Sample size*: 355, 40, 225, 364, 125
Authors' suggestions: The author recommends the methods reviewed for application according their performance in certain situations. Bootstrap methods are shown to have better coverage in multi-sample situations with different degree of skewness. *Sample size implications as discussed by authors*: ML estimators recommended when skewness is not high, MSE estimators otherwise. Bootstrap method recommended for small samples and likelihood ratio for moderate to large samples. Generally, methods perform well with large samples.
Connections to other papers: Duan1983, Zhou1997, Zhou1997A, Zhou1997B, Zhou1998, Zhou1999, Zhou2000, Zhou2000A, Zhou2001, Zhou2001A, Zhou2001B, Tu1999, Tu2000
Review Comment *Skewness, heavy tails, multimodality:* Log transformation where appropriate has been shown to normalise skewed data but it does generate difficulties with the need to present results on the raw scale (retransformation). *Adjustment for covariates*: Implemented in the regression models only. *Extensions to cost-effectiveness*: Methods are not extended to model both costs and utilities and this does not seem straightforward.

Reference: **Zhou2005 (Zhou & Dinh 2005)**
Category: Non-parametric methods
Parameter(s): Mean cost, mean cost difference
Data: One- and two-sample skewed data *Model*: Confidence intervals for the one- and two-sample case based on Edgeworth expansion to modify the *t*-statistic and remove effect of skewness. Three specifications of the transformation are explored. *Statistical method(s)*: Standard methods *Implementation*: Not specified. *Representation of estimation error:* Nonparametric 95% confidence intervals.
Applied data: 1 dataset; *Sample size*: 211 *Simulated data*: Pairs of Log-normal and Gamma distributed data; *Sample size*: 25, 50, 100 *Comparisons with other models/methods:* Ordinary-t, the bootstrap-t, the bias-corrected acceleration
Authors' suggestions: The bootstrap-*t* and the new (Edgeworth) transformation confidence

intervals showed good coverage and are recommended over the ordinary-*t* interval. The transformation intervals were shorter and require less computing than bootstrap-*t* intervals. In the one sample case high skewness and relatively small sample sizes worsens the performance of the ordinary-t interval. The relative skewness of both samples in the two-sample case affects the ordinary-*t* interval. *Sample size implications as discussed by the authors*: Standard recommendation of sample size of 30 for the good performance of normality based methods is inadequate for highly skewed data. With large sample sizes the effect of skewness on the ordinary-t interval diminishes.
Connections to other papers: Zhou2002
Review Comment *Skewness, heavy tails, multimodality*: Confidence intervals based on Edgeworth transformations of the *t*-statistic to adjust for skewness seem to outperform the ordinary-*t* intervals and perform similarly to the bootstrap-*t*. The flexibility of the transformation approaches seems to be limited and based on measures of skewness and sample sizes. *Adjustment for covariates/ Extensions to cost-effectiveness*: The suggested models do not incorporate adjustments for covariates. I am not sure how easy and practical it would be to introduce adjustment for covariates or extensions to cost-effectiveness. *Sample size implications*: *General comments*: General advantage is simplicity of method. However performance of intervals really not that great – coverage for sample of 500 from lognormal distribution still less than 90. Interesting observation that t-intervals fine if samples similarly skewed.

Reference: **Zhou2006 (Zhou & Liang 2006)**
Category: Two-part models
Parameter(s): Mean cost
Data: Skewed data with additional zero values, with potentially heteroscedastic variance. *Model*: A semi-parametric single-index two-part regression model (first part is a parametric logit model; second part is a semi-parametric single-index) *Statistical method:* Average derivative estimation; local constant kernel smoothing *Implementation*: Implemented in XploRe available from www.xplore-stat.de; program code for the implementation not available *Representation of estimation error*: Asymptotic standard error, bootstrap standard error
Applied data: 1 dataset; *Sample size*: Not available *Simulated data*: Yes; *Sample size*: 100, 200, 500 *Comparisons with other models/methods:* Parametric two-part model
Authors' suggestions: The approach handles skewed data with excess zeros without adding distributional assumptions. The proposed estimators are consistent and have asymptotically normal distributions under some regularity conditions. The first stage zero versus non-zero model follows a parametric logit model (the real data follow this distribution) but it could be generalised to a semi-parametric single-index model as well. The focus is on prediction of costs given covariates. *Sample size implications as discussed by authors*: Not discussed
Connections to other papers: Duan1983
Review Comment *Skewness, heavy tails, multimodality:* Suitable for skewed data with excess zeros and potentially heteroscedastic variance. *Adjustment for covariates/ Extensions to cost-effectiveness*: The approach allows for adjustment for covariates to be implemented. The extension to joint modelling of costs and effects for the purpose of cost-effectiveness analysis does not look straightforward. *Sample size implications:* Unclear *General comments:* The approach is not straightforward to implement.

CENSORED AND MISSING RESOURCE USE AND COST DATA

Censored data

Although the methods accounting for censoring, a feature of the data collection, are not central to the review, we briefly summarise the key approaches since resource use and cost data in randomised (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, (Etzioni et al. 1996;Lin et al. 1997), and the Inverse Probability Weighting (IPW) estimator (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 to 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 have been explored (O'Hagan & Stevens 2004) and both methods are shown to perform well over different levels of censoring (Raikou & McGuire 2004). A more general class of estimators, encompassing both the KMSA and the IPW estimators, is proposed by Strawderman (Strawderman 2000). Further methods allow for adjustments for covariates together with censoring adjustment (Carides et al. 2000;Lin 2000a) proportional means regression (Lin 2000b) and median regression of lifetime costs(Bang & Tsiatis 2002) (not of interest in the current review). Jain and Strawderman(Jain & Strawderman 2002) propose a more general approach for covariate adjustment that adapts the hazard regression framework by employing weighting using inverse probability of censoring to model lifetime costs that does not rely on restrictive assumptions such as proportional hazards or a linear relationship between survival time and covariates. Extensions to model censored skewed data better through generalised linear methods (Lin 2003), adjust for survival time (Liu et al. 2007) and incorporate two-part models for costs (Tian & Huang

2007) are also proposed. Extensions of some of the methods to cost-effectiveness are also proposed (Pullenayegum & Willan 2007;Willan et al. 2002;Willan et al. 2005). Further research is needed to develop extensions of some of the more complex methods identified in our review to account for administrative censoring. For a recent review of methods, including approaches for inference on lifetime costs using right-censored data, see Huang (Huang 2009).

Missing data

The paper does not review methods for dealing with missing resource use and cost data but missingness is often encountered in trial data, despite the efforts to fully capture the information. Therefore, we briefly summarise the key approaches. In situations where the missing data can be assumed to depend on observed variables, data are missing at random and imputing these missing values is usually recommended. Since single imputation methods overstate precision, multiple imputation techniques, which allow for the uncertainty in the imputations, have been widely advocated (Rubin 1987;Schafer 1997). Different approaches for imputation are available, including regression based, expectation maximisation, propensity method and MCMC methods (Briggs et al. 2003). Some issues arise with respect to dealing with missing resource use and cost data. Firstly, there appears to be not much research into exploring the pros and cons of imputing missing resource use rather than higher-level cost data. Secondly, current methods for imputation seem little studied in the case of multimodal, heavy tailed data. Further research is needed to extend most of the more complex methods identified in our review to account for missing data. In situations where data are missing due to unobserved factors, these data are missing not at random and the potential reasons should be carefully studied and appropriate methods used (Diggle et al. 2007;Little 1995). However, in such cases imputation relies on untestable assumptions and extensive sensitivity analyses and cautious conclusions are recommended.

REFERENCES

Ai, C. & Norton, E. C. 2000, "Standard Errors for the Retransformation Problem with Heteroscedasticity", *Journal of Health Economics.September 2000*, vol. 19, no. 5): 697-718.

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.

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.

Austin, P. C. & Escobar, M. D. 2003, "The Use of Finite Mixture Models to Estimate the Distribution of the Health Utilities Index in the Presence of a Ceiling Effect", *Journal.of.Applied.Statistics*, vol. 30, no. 8, pp. 909-923.

Bang, H. & Tsiatis, A. A. 2000, "Estimating medical costs with censored data", *Biometrika*, vol. 87, no. 2, pp. 329-343.

Bang, H. & Tsiatis, A. A. 2002, "Median regression with censored cost data", *Biometrics*, vol. 58, pp. 643-649.

Barber, J. & Thompson, S. 2004, "Multiple regression of cost data: use of generalised linear models", *J.Health.Serv.Res.Policy*, vol. 9, no. 4, pp. 197-204.

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.

Basu, A. Extended generalised Linear Models: Simultaneous Estimation of Flexible Link and Variance Functions. 2005.
Ref Type: Unpublished Work

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.

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.

Basu, A. & Rathouz, P. J. 2005, "Estimating marginal and incremental effects on health outcomes using flexible link and variance function models", *Biostatistics.*, vol. 6, no. 1, pp. 93-109.

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.

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.

Briggs, A., Clark, T., Wolstenholme, J., & Clarke, P. 2003, "Missing... presumed at random: cost-analysis of incomplete data", *Health Economics*, vol. 12, no. 5, pp. 377-392.

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.

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.

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.

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.

Cameron, A. C. & Johansson, P. Bivariate count data regression using series expansions: with applications. Department of Economics Discussion Paper[15]. 1998. Davis, University of California.
Ref Type: Serial (Book,Monograph)

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.2004*, vol. 7, no. 2): 566-84.

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.January 1986*, vol. 1, no. 1): 29-53.

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.

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.

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.

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.

Conigliani, C. & Tancredi, A. 2005, "Semi-parametric modelling for costs of health care technologies", *Stat.Med.*, vol. 24, no. 20, pp. 3171-3184.

Conigliani, C. & Tancredi, A. Comparing parametric and semi-parametric approaches for Bayesian cost-effectiveness analyses. Working Paper del Dipartimento di Economia[64]. 2006. Università Roma Tre.
Ref Type: Serial (Book,Monograph)

Conigliani, C. & Tancredi, A. 2009, "A Bayesian model averaging approach for cost-effectiveness analyses", *Health Economics*, vol. 18, no. 7, pp. 807-821.

Cooper, N. J., Lambert, P. C., Abrams, K. R., & Sutton, A. J. 2007, "Predicting costs over time using Bayesian Markov chain Monte Carlo methods: an application to early inflammatory polyarthritis", *Health Economics*, vol. 16, no. 1, pp. 37-56.

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.

Deb, P. & Burgess, J. 2003, *A Quasi-experimental Comparison of Econometric Models for Health Care Expenditures*, Hunter College Department of Economics.

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.

Deb, P. & Trivedi, P. K. 1997, "Demand for Medical Care by the Elderly: A Finite Mixture Approach", *Journal of Applied Econometrics.May June 1997*, vol. 12, no. 3): 313-36.

Diggle, P., Farewell, D., & Henderson, R. 2007, "Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal", *Applied Statistics*, vol. 56, no. 5, pp. 499-550.

Dinh, P. & Zhou, X. H. 2006, "Nonparametric statistical methods for cost-effectiveness analyses", *Biometrics*, vol. 62, no. 2, pp. 576-588.

Dominici, F., Cope, L., Naiman, D. Q., & Zeger, S. L. 2005, "Smooth quantile ratio estimation", *Biometrika*, vol. 92, no. 3, pp. 543-557.

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.

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.

Duan, N. 1983, "Smearing Estimate: A Nonparametric Retransformation Method", *Journal of the American Statistical Association*, vol. 78, no. 383, pp. 605-610.

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.

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.

Dudley, R. A., Harrell, F. E. J., 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", *J.Clin.Epidemiol.*, vol. 46, no. 3, pp. 261-271.

Etzioni, R., Urban, N., & Baker, M. 1996, "Estimating the costs attributable to a disease with application to ovarian cancer", *Journal of Clinical Epidemiology*, vol. 49, no. 1, pp. 95-103.

Faddy, M., Graves, N., & Pettitt, A. 2009, "Modelling Length of Stay in Hospital and Other Right Skewed Data: Comparison of Phase-Type, Gamma and Log-Normal Distributions", *Value in Health*, vol. 12, no. 2, pp. 309-314.

Gilleskie, D. B. & Mroz, T. A. 2004, "A Flexible Approach for Estimating the Effects of Covariates on Health Expenditures", *Journal of Health Economics.March 2004*, vol. 23, no. 2): 391-418.

Grootendorst, P. V. 1995, "A comparison of alternative models of prescription drug utilization", *Health Economics*, vol. 4, no. 3, pp. 183-198.

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.

Gurmu, S. 1998, "Generalized Hurdle Count Data Regression Models", *Economics Letters*, vol. 58, no. 3, pp. 263-268.

Gurmu, S. & Elder, J. Estimation of multivariate count regression models with application to health care utilization. 2000a.
Ref Type: Unpublished Work

Gurmu, S. & Elder, J. 2000b, "Generalized Bivariate Count Data Regression Models", *Economics Letters*, vol. 68, no. 1, pp. 31-36.

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.

Hill, S. C. & Miller, G. E. Health expenditure estimation and functional form: applications of the generalized Gamma and extended estimating equations models. Health Economics . 2009.
Ref Type: In Press

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.

Huang, Y. 2009, "Cost analysis with censored data", *Medical Care*, vol. 47, no. 7:Suppl 1, p. Suppl-9.

Jain, A. K. & Strawderman, R. L. 2002, "Flexible hazard regression modelling for medical cost data", *Biostatistics*, vol. 3, pp. 101-118.

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.

Keeler, E. B., Manning, W. G., & Wells, K. B. 1988, "The demand for episodes of mental health services", *Journal of Health Economics*, vol. 7, no. 4, pp. 369-392.

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.

Leung, S. F. & Yu, S. 1996, "On the Choice between Sample Selection and Two-Part Models", *Journal of Econometrics.May 1996*, vol. 72, no. 1-2): 197-229.

Lin, D. 2000a, "Linear regression analysis of censored medical costs", *Biostatistics*, vol. 1, pp. 35-47.

Lin, D. Y. 2000b, "Proportional means regression for censored medical costs", *Biometrics*, vol. 56, no. 3, pp. 775-778.

Lin, D. Y. 2003, "Regression analysis of incomplete medical cost data", *Statistics in Medicine*, vol. 22, no. 7, pp. 1181-1200.

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.

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.

Little, R. J. A. 1995, "Modeling the Drop-Out Mechanism in Repeated-Measures Studies", *Journal of the American Statistical Association*, vol. 90, no. 431, pp. 1112-1121.

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.

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.

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.

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.

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.

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.

Marazzi, A. & Barbati, G. 2003, "Robust parametric means of asymmetric distributions: estimation and testing", *Estadistica*, vol. 54, pp. 47-72.

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.

Marazzi, A. & Ruffieux, C. 1999, "The truncated mean of an asymmetric distribution", *Computational Statistics & Data Analysis*, vol. 32, no. 1, pp. 79-100.

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.

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.

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.

Mullahy, J. 1998, "Much ado about two: reconsidering retransformation and the two-part model in health econometrics", *Journal of Health Economics*, vol. 17, pp. 247-281.

Mullahy, J. 1997, "Heterogeneity, Excess Zeros, and the Structure of Count Data Models", *Journal of Applied Econometrics.May June 1997*, vol. 12, no. 3): 337-50.

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.

Nixon, R. M. & Thompson, S. G. 2004, "Parametric modelling of cost data in medical studies", *Stat.Med.*, vol. 23, no. 8, pp. 1311-1331.

Nixon, R. M., Wonderling, D., & Grieve, R. D. Non-parametric methods for cost-effectiveness analysis: the Central Limit theorem and the bootstrap compared. Health Economics . 2009. Ref Type: In Press

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.

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.

Pagano, E., Petrinco, M., Desideri, A., Bigi, R., Merletti, F., & Gregori, D. 2008, "Survival models for cost data: the forgotten additive approach", *Statistics in Medicine*, vol. 27, no. 18, pp. 3585-3597.

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.

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.

Raikou, M. & McGuire, A. 2004, "Estimating medical care costs under conditions of censoring", *Journal of Health Economics*, vol. 23, no. 3, pp. 443-470.

Rubin, D. B. 1987, *Multiple Imputation for Nonresponse in Surveys* Wiley, New York.

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.

Schafer, J. L. 1997, *Analysis of incomplete multivariate data* Chapman and Hall, London.

Strawderman, R. L. 2000, "Estimating the mean of an increasingly stochastic process at a censored stopping time", *Journal of the American Statistical Association*, vol. 95, pp. 1192-1208.

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.

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.

Tian, L. & Huang, J. 2007, "A two-part model for censored medical cost data", *Statistics in Medicine*, vol. 26, no. 23, pp. 4273-4292.

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.

Tu, W. & Zhou, X. H. 1999, "A Wald test comparing medical costs based on log-normal distributions with zero valued costs", *Statistics in Medicine*, vol. 18, no. 20, pp. 2749-2761.

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.

van de Ven, W. P. M. M. & Van Praag, B. M. S. 1981, "Risk aversion and deductibles in private health insurance: application of an adjusted tobit model to family health care expenditures," in *Health, economics, and health economics*, J. van der Gaag & M. Perlman, eds., North-Holland Publishing Company.

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.

Wang, H. J. & Zhou, X. H. 2010, "Estimation of the retransformed conditional mean in health care cost studies", *Biometrika*, vol. 97, no. 1, pp. 147-158.

Wang, P. 2003, "A Bivariate Zero-Inflated Negative Binomial Regression Model for Count Data with Excess Zeros", *Economics Letters.March 2003*, vol. 78, no. 3): 373-78.

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.

Willan, A. R., Lin, D. Y., Cook, R. J., & Chen, E. B. 2002, "Using inverse-weighting in cost-effectiveness analysis with censored data", *Stat.Methods.Med.Res.*, vol. 11, no. 6, pp. 539-551.

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.

Winkelmann, R. 2004, "Health Care Reform and the Number of Doctor Visits--An Econometric Analysis", *Journal of Applied Econometrics*, vol. 19, pp. 455-472.

Zhou, X. H. 1998, "Estimation of the log-normal mean", *Stat.Med*, vol. 17, no. 19, pp. 2251-2264.

Zhou, X. H. 2002, "Inferences about population means of health care costs", *Stat.Methods.Med.Res*, vol. 11, no. 4, pp. 327-339.

Zhou, X. H. & Cheng, H. 2008, "A computer program for estimating the re-transformed mean in heteroscedastic two-part models", *Computer Methods & Programs in Biomedicine*, vol. 90, no. 3, pp. 210-216.

Zhou, X. H. & Dinh, P. 2005, "Nonparametric confidence intervals for the one- and two-sample problems", *Biostatistics.*, vol. 6, no. 2, pp. 187-200.

Zhou, X. H. & Gao, S. 1997, "Confidence intervals for the log-normal mean", *Stat.Med*, vol. 16, no. 7, pp. 783-790.

Zhou, X. H. & Gao, S. J. 2000, "One-sided confidence intervals for means of positively skewed distributions", *American Statistician*, vol. 54, no. 2, pp. 100-104.

Zhou, X. H., Li, C., Gao, S., & Tierney, W. M. 2001a, "Methods for testing equality of means of health care costs in a paired design study", *Statistics in Medicine*, vol. 20, no. 11, pp. 1703-1720.

Zhou, X. H., Melfi, C. A., & Hui, S. L. 1997a, "Methods for comparison of cost data", *Ann.Intern.Med*, vol. 127, no. 8 Pt 2, pp. 752-756.

Zhou, X. H., Stroupe, K. T., & Tierney, W. M. 2001b, "Regression analysis of health care charges with heteroscedasticity", *JRSS, Ser.C*, vol. 50, no. 3, pp. 303-312.

Zhou, X. H. & Tu, W. 2000, "Confidence intervals for the mean of diagnostic test charge data containing zeros", *Biometrics*, vol. 56, no. 4, pp. 1118-1125.

Zhou, X. H., Gao, S., & Hui, S. L. 1997b, "Methods for Comparing the Means of Two Independent Log-Normal Samples", *Biometrics*, vol. 53, no. 3, pp. 1129-1135.

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.

Zhou, X. H. & Tu, W. 1999, "Comparison of Several Independent Population Means When Their Samples Contain Log-Normal and Possibly Zero Observations", *Biometrics*, vol. 55, no. 2, pp. 645-651.