**Does transfer to intensive care units reduce mortality? A comparison of an instrumental variables design to risk adjustment.**

Authors: Luke Keele, Steve Harris, Richard Grieve

Total words: X words.

Key words: critical care, instrumental variables, risk adjustment

Acknowledgements: We thank Kathryn Rowan, Mervyn Singer, Colin Sanderson and David Harrison for their guidance in the design of the SPOTlight study. We also acknowledge Anirban Basu and Stephen O'Neill for comments and discussion. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health. This report is independent research supported by the National Institute for Health Research (Senior Research Fellowship, Richard Grieve; SRF-2013-06-016). Conflict of interest: None.

**Author Information Page**

Luke Keele, Asociate Professor, University of Pennsylvania, 3400 Spruce St, Philadelphia, PA 19104, Email: luke.keele@gmail.com, corresponding author.

Steve Harris, Clinical Lecturer in Anaesthesia and Critical Care, University College London, Hospital, Email: doc@steveharris.me.

Richard Grieve, Professor of Health Economics Methodology, Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH, Email: richard.grieve@lshtm.ac.uk.

## Abstract:

**Background:** Instrumental variable (IV) analysis can estimate treatment effects in the presence of residual or unmeasured confounding. In settings where measures of baseline risk severity are unavailable, IV designs are therefore particularly appealing, but where established measures of risk severity are available, it is unclear whether IV methods are preferable. **Objective:** We compared regression to an IV design to estimate the effect of ICU transfer on mortality in a study with well-established measures of risk-severity.

**Research Design:** We use ICU bed availability at the time of assessment for ICU transfer as an instrument. Bed availability increases the chance of ICU admission, contains little information about patient characteristics, and it is unlikely that bed availability has any direct effect on in-hospital mortality.

**Subjects:** We used a cohort study of deteriorating ward patients assessed for critical care unit admission, in 49 UK National Health Service hospitals between November 1, 2010, and December 31, 2011.

**Measures:** Detailed demographic, physiological, and comorbidity data were collected for all patients.

**Results:** The risk adjustment methods reported that after controlling for all measured covariates including measures of risk severity, ICU transfer was associated with higher 28-day mortality with a risk difference of 7.2%, (95% CI=5.3% to 9.1%). The IV estimate of ICU transfer was -5.4% (95% CI=-47.1% to 36.3%) and applies to the subsample of patients whose transferwas 'encouraged' by bed availability.

**Conclusions:** IV estimates indicate ICU care is beneficial but are imprecisely estimated. Risk adjusted estimates are more precise, but even with a rich set of covariates report that ICU care is harmful.

Obtaining reliable evidence about whether transferring critically ill patients to Intensive Care Units (ICU) reduces mortality is difficult, since randomization of patients faces both ethical and practical barriers. Observational studies have tried to establish whether ICU transfer is effective for deteriorating ward patients[1,2]. In an attempt to address confounding by indication, one study used risk adjustment (RA) via multivariate regression analyses[3]. Confounding by indication is an important threat in any study of comparative effectiveness, and estimating treatment effects under an RA approach requires investigators to assume that all potential confounders have been observed. This assumption may be implausible if the reasons for transfer to the ICU are not fully recorded.

Recent studies have applied the IV design to evaluate the effectiveness of ICU transfer[4–9]. Subject to a set of assumptions, the IV study design can provide consistent effectiveness estimates even when there are unmeasured prognostic differences between the comparison groups[10–12]. Brookhart[12] advocates the use of IV designs given that measures of patient physiology or risk severity are often unrecorded in many medical data sources. However, others have strongly criticized the IV designs as imposing assumptions that may be unrealistic in many clinical settings[13–15].

In this study, we extend the (SPOT)light prospective cohort study in which investigators identified an IV for ICU transfer[7]. Based on ICU census data, they measured the number of available ICU beds at the specific time that a patient was assessed for ICU admission. They proposed that bed availability at the time of assessment was a valid IV for ICU transfer. An IV design replaces the assumption of no unmeasured confounding with the assumption that the number of ICU beds available at the time of assessment only has an effect on the outcome (mortality) through receipt of care in the ICU. An important feature of (SPOT)light is that the data collection included detailed measures of the patient's physiological status which allowed detailed, validated measures of risk severity to be calculated. The aim of this paper is to compare estimates from the IV design with those from risk adjustment in a setting where detailed measures of baseline risk are available.

# Methods

## Data

This study uses data from the (SPOT)light study which was a prospective cohort study of deteriorating ward patient referred for assessment by critical care[7]. The (SPOT)light investigators collected data from a population of deteriorating ward patients referred to critical care in 48 National Health Service (NHS) hospitals between 1 November 2010 and 31 December 2011. The full study protocol is available on the ICNARC website[16]. Repeat visits, re-admissions, cardiac arrests, and deaths during the assessment were excluded, as were admissions following surgery (where delay may be due to the process of care), and patients with pre-existing treatment limitations. Fact and time of admission to critical care were obtained from the Intensive Care National Audit and Research Case Mix Programme (ICNARC CMP). Physiology measurements at the time of the ward assessment were abstracted, and used to generate physiological severity of illness scores for case mix adjustment.

In our study, the initial population is comprised of 15,158 patients on general hospital wards that were assessed for admission to the ICU. From this population, we excluded 2,141 patients from the study due to the presence of a treatment limitation order which prevented the possibility of more aggressive care. Of the remaining 13,017 patients, six are excluded since data was missing on the availability of beds in the ICU at the time of assessment, which implies a study population of 13,011 patients. Figure 1 details both the exclusion criteria and status by the exposure.

Our primary outcomes were death 7 and 28 days after admission to the ICU. We defined the exposure of interest as transfer to and receipt of ICU care. Covariates included data on age, gender, septic diagnosis (0/1), and peri-arrest (0/1). Physiology data were also collected to create three measures of risk severity. These validated risk severity measures include the Intensive Care National Audit & Research Centre (ICNARC) physiology score[17], the NHS National Early Warning (NEWS) score which measures whether respiratory rate, oxygen saturations, temperature, systolic blood pressure, pulse rate, and level of consciousness varied from the norm[18], and the Sequential Organ Failure Assessment (SOFA) score which ranges from 0 to 24, with higher scores indicating a greater degree of organ failure[19]. In addition, the data also record the patient's existing level of care at assessment and recommended level of care after assessment were defined using the UK Critical Care Minimum Dataset (CCMDS) levels of care[20]. These levels are 0 and 1 for general ward care, 2 for care within a high dependency unit, and 3 for care within the ICU. As such, the data include both objective and subjective measures of risk severity that would be absent in many health care databases.

## Instrument: ICU Bed Availability

We use ICU bed availability at the time of assessment for ICU as an instrument for ICU transfer. That is, if there are few ICU beds available when the patient is assessed, this discourages ICU transfer. Bed availability ranges from 0 to 19 with a median of 4 and an IQR of 4. For the analysis that follows, we use a binary measure of the instrument where a 1

indicates that fewer than 4 ICU beds were available at the time of assessment. We made this decision based on IV diagnostics. See Section 1.2 in the online supplementary materials. The Section 2.2 in the online supplementary materials contains results using the multi-valued version of the instrument. When ICU bed availability was above the median, 42.4% of patients assessed were admitted to the ICU; when there were fewer than the median number of ICU beds available, patients were admitted to the ICU 33.8% of the time. For patients admitted to the ICU, the median time to the ICU was 4 hours (IQR: 11) if there were 4 or fewer beds available, and the median time to the ICU was 3 hours (IQR: 8). See[21] for a study that uses the SPOTlight data to study whether timing to the ICU affected mortality.

For a measure to be a valid IV, several assumptions must hold. Moreover, a number of diagnostic tests have been developed to probe the IV assumptions. Here, we provide a brief review of our assessment of bed availability as an IV. The online supplementary materials contain a full review of those conditions following recent guidelines for reporting IV estimates in clinical settings[22,23]. Here, we provide a brief review of those results. First, results from a weak instrument test validated that the IV was sufficiently strong. We also assessed whether instrument status appears as 'if random' and does not have any common causes that jointly affect the outcome. Using balance tests, bias ratios, and a graphical test, we found that bed availability at time of assessment tells us nothing of clinical relevance about the patients[24–26]. Table 1 contains descriptive statistics stratified by both the IV and the expsoure. Notably, patients admitted to the ICU had higher risk severity scores, while the IV tells us little about patient frailty. See Section 1.3 in the online supplementary materials for details on this analysis.

Next, an IV cannot have a direct effect on 7- or 28-day mortality–this is referred to as the exclusion restriction. One way this assumption would be violated is if, when the ICU is relatively full, the quality of care deteriorated leading to higher mortality. However, the instrument is the number of ICU beds free at the time of assessment not transfer. ICU bed occupancy varies by the time of day, and day of the week, and also randomly. Therefore the number of ICU beds free at assessment, may be only weakly related to the ICU occupancy during the ICU stay for those actually transferred. Moreover, the empirical evidence on the association between ICU occupancy and outcomes provides ambiguous results[1,27]. In sum, it appears plausible, that the number of ICU beds at assessment only affects mortality by inducing some patients to be admitted to the ICU.

While this assumption cannot be tested, analysts can often apply a faslfication test to probe the exclusion restriction[28,29]. One falsfication test consists of identifying a subgroup that is always or never treated[30]. In this subgroup, the IV should have no effect on the outcome if the exclusion restriction holds. We found that 96% of patients that scored a 3 on CCMDS recommended levels of care variable were transferred to the ICU. While 91% of patients that scored a zero on the scale were not transferred to the ICU. To conduct a falsification test, we tested whether the IV was associated with either outcome in these subgroups using linear regression models. First, we present results for patients with 3 on the CCMDS variable. For the 7-day mortality outcome, the estimated risk difference is $-$0.632% with a p-value of 0.744. For the 28-day mortality outcome, the risk difference is $-$1.0% with a p-value of 0.686. Next, we present results for patients with 0 on the CCMDS variable. For

the 7-day mortality outcome, the estimated coefficient is 0.40% with a p-value of 0.875. For the 28-day mortality outcome, the estimated coefficient is -0.01% with a p-value of 0.733. As such, in all cases, we find that the IV is not associaed with outcome, which implies that we do not find evidence against the exclusion restriction. See Section 1.4 in the online supplementary materials for details including full model results.

Finally, it is important to note that RA and IV estimate different causal quantities. Estimates from RA may be interpreted as the average treatment effect. Under the IV design, we estimate a complier average causal effect, that is the treatment effect for the subpopulation of patients who receive ICU care because they were encouraged to by ICU bed availability. Table 2 contains descriptive statistics for the compliers as well as the overall cohort for select covariates. The online supplementary materials contains summary statistics for all covariates.

### Statistical Methods

We estimated the effect of transfer to the ICU using two-stage least squares (2SLS) for the IV analyses. We used models based on 2SLS, since they impose relatively weak assumptions and generate consistent IV estimates, even when the treatment and outcome are binary variables[31–33]. See Section 2.1 in the online appendix for results based on logistic regression and a more flexible fit based on an ensemble of machine learning algorithms. Under the RA design, we adjusted for observed confounders using multivariate regression to ensure the scale of the estimates are comparable. For both approaches, we used three different model specifications according to the covariates included. Under the first specification, to provide unadjusted estimates of the effects of ICU transfer on hospital mortality, no covariates were included. The second specification was intended to mimic the 'reduced list' of covariates, generally available in routine administrative databases. For the reduced specification, we adjusted for age, gender, septic diagnosis, peri-arrest, and hospital fixed effects – indicators for each hospital. The final specification included all observed covariates, including the validated measures of risk severity, together with hospital fixed effects. Section 2.4 in the online supplement contains complete model results based on the 2SLS estimator. See Section 3 in the online supplementary materials for details on a secondary matching analysis that we also performed. Analyses were conducted in R 3.3.2 (*R Foundation*) and Stata software version 14.

## Results

Table 3 contains the RA versus IV estimates of the effects of ICU transfer on 7-day hospital mortality. The unadjusted RA model reported that ICU transfer was associated with higher mortality, with a risk difference (RD) of 9.3% (95% CI= 8.0% to 10.5%). The unadjusted IV estimate indicated that on average ICU transfer reduced mortality with a RD of -5.7% (95% CI= -19.7% to 8.3%). The RA approach that included a 'reduced list' of potential confounders reported estimated effects that were somewhat smaller (RD: 8.2%, 95% CI= 6.9% to 9.5%). While including the measures of risk severity reduced the magnitude of the estimated treatment effect using RA, they still suggested that ICU transfer caused an increase in 7-day mortality (RD: 3.5%, 95% CI from 2.0% to 5.1%). The IV models that used

the 'reduced list' of covariates, and the full specification both found that ICU transfer was beneficial on average (RD: -6.2% 95% CI= -40.7% to 28.3%).

Table 2 contains the estimates for the effect of ICU transfer on 28-day mortality. The pattern of effects corresponded to those for 7-day mortality. Estimates based on RA indicate that mortality increases following ICU transfer, and including measures of risk severity reduced the magnitude of the estimated treatment effect. For example, using the full specification, the estimated based on RA suggested that ICU transfer causes an increase in 28-day mortality (RD: 7.2%, 95% CI from 5.3% to 9.1%). Estimates based on the IV design indicate that on average ICU transfer reduces mortality across all specifications. When we use the full specification with the IV method, we find that ICU transfer causes a decrease in mortality (RD: -5.4% 95% CI= -19.7% to 8.3%) though the IV estimates contains considerable statistical uncertainty. While the IV estimates are imprecisely estimated, they are comparable with other IV estimates of ICU effectiveness[5,6,8].

## Discussion

This paper contrasts an IV design with conventional RA; in a setting with a rich set of baseline covariates from a high-quality clinical database. The IV design estimates that, on average, ICU transfer reduces 7- and 28- day mortality for deteriorating ward patients. By contrast, conventional RA estimates that ICU transfer for deteriorating ward patients is associated with an increase in hospital mortality, even when detailed measures of baseline risk are included in the adjustment. In supplementary analyses we undertook matching prior to both RA and IV estimation, and reported similar results (see supplement). While all the IV estimates are imprecise, this does suggest that the IV design, even in its unadjusted form offers protection against bias from confounding by indication. However, even when the RA includes case-mix measures from a high-quality clinical database that have previously been validated in relevant settings, they provide estimates that lack clinical plausibility.

The method of instrumental variables has become more widely used in clinical applications. The obvious attraction of the IV approach is that it offers some protection against unobserved confounding outside of randomization. Of late, the use of IV methods in clinical settings have been strongly critiqued[13–15,34]. Soumerai and Koppel argue that,``the majority of IVs are not a reliable way to control for bias in medical effectiveness research.''[14] One response is to rely on RA methods and seek to overcome confounding by indication with richer specifications. The major limitation to this approach is that due to unobserved prognostic variables, such as the patients' past medical history that may influence treatment assignment, the estimated treatment effects are likely to be biased whichever measured covariates are included in the RA models. Here, those who are ICU transferred are likely to have the a high risk of death versus those who remain on general wards, according to prognostic measures that are not available to the study. We would maintain that the results presented here suggest careful application of IV methods should remain a key part of outcomes resarch. While our IV estimates are imprecisely estimated, they are comparable with other IV estimates of ICU effectiveness[5,6,8].

Moreover our study adds to the knowledge based through the use of an additional IV for ICU admissions in a new setting. Rosenbaum and Imbens recommend that evidence be built across a series of studies based on different instruments: ''…if each instrument is plausible but not certain, there may be no reason why these different instruments should be biased in the same direction. In replicating observational studies, the goal is to replicate whatever treatment effects may exist without replicating whatever biases may coexist and this goal is sometimes achievable by using a variety of instruments . . . .''[35] Thus it is worth noting the congruence between our study and the results in Valley et al used distance to hospital as an alternative, but also plausible IV.[8].

Our study can be viewed as a new addition to the use of instruments to study the effectiveness of ICU transfer. Although the (SPOT)light study included multiple measures of risk severity, the estimates based on RA methods appear to be confounded with clinical severity as they indicate that ICU transfer causes higher levels of mortality. Our results based on RA methods suggest that confounding by indication cannot be overcome even in a context with validated measures of risk severity. It is conceivable that the RA estimates are unbiased and the IV estimates are biased, due for example to violation of the exclusion restriction. This would seem unlikely for the following reasons: first, we followed recent methodological guidance and assessed relative bias for each prognostic measure. The results suggest that the RA results were more likely to be biased according to the prognostic measures of greatest importance. Second, while it is plausible that ICU transfer can lead to harm, due for example to more aggressive therapy for older or frailer patients, this seems implausible for the overall cohort of patients included in the (SPOT)light study. Moreover, the IV design passed a falsification test for this key assumption. Third, as additional measures of baseline risk were included in the RA, the magnitude of the estimated risk increase following ICU transfer was reduced. Clinical opinion indicates that patients recommended for ICU transfer would have relatively poor prognosis according to unmeasured covariates, versus those who were assessed but remain on general wards. Hence, it is plausible that this and previous studies that use RA underestimate the benefits of ICU care.

This study inevitably has limitations, and provokes areas for further research. First, IV study designs are inefficient, and the study was not powered to formally test hypotheses as to whether the IV versus RA approaches provided different estimates of the effect of ICU transfer. While power in an IV design depends on several factors, the most straightforward way to increase the precision of the estimates would be more extensive data collection. (SPOT)light is based on one year of admissions across 48 hospitals. A longer time span and additional hospitals are obvious ways to expand the data.

Second, no case study can provide formal assessment of the relative bias and precision of alternative study designs. A complementary approach would be to design formal simulation studies characterized by 'real' examples such as the (SPOT)light study. Third, we contrasted the IV and RA designs within a case study of prime policy relevance, with the aim of providing insights to guide the future design of observational studies that use routine data. That said, definitive conclusions about the relative merits of these complementary approaches cannot be drawn from a single study, and further research that contrasts these approaches across a range of clinical areas, would be useful. Finally, while

bed availability appears to be a valid IV, we cannot rule the possibility of unobserved IV and outcome confounders. Critical confounders would be factors that cause both bed availability and mortality to move in tandem. One possible factor might be some time trend that increased admissions thus reducing ICU space but is also associated with admissions with higher mortality.

## Conclusion

In conclusion, we contrasted IV and RA methods in evaluating the effectiveness of ICU transfer for deteriorating ward patients. We find that the IV design appears to reduce bias from unobserved confounders relative to estimates based on risk adjustment.

### References


1. Gabler N, Ratcliffe S, Wagner J, et al. Mortality among patients admitted to strained intensive care units. *Am J Respir Crit Care Med*. 2013;188(7):800-806.

2. Renaud B, Santin A, Coma E, et al. Association between timing of intensive care unit admission and outcomes for emergency department patients with community-acquired pneumonia*. *Critical care medicine*. 2009;37(11):2867-2874.

3. Simchen E, Sprung CL, Galai N, et al. Survival of critically ill patients hospitalized in and out of intensive care. *Crit Care Med*. 2007;35(2):449-457.

4. Hu W, Chan CW, Zubizarreta JR, Escobar GJ. An examination of early transfers to the icu based on a physiologic risk score. *Manufacturing & Service Operations Management*. 2018.

5. Kc DS, Terwiesch C. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management*. 2012;14(1):50-65.

6. Shmueli A, Baras M, Sprung CL. The effect of intensive care on in-hospital survival. *Health Services and Outcomes Research Methodology*. 2004;5(3-4):163-174.

7. Harris S, Singer M, C CS, Grieve R, Harrison D, Rowan K. Impact on mortality of prompt admission to critical care for deteriorating ward patients: An instrumental variable analysis using critical care bed strain. *Intensive Care Medicine*. 2018;in press.

8. Valley T, Sjoding M, Ryan A, Iwashyna T, Cooke C. Association of intensive care unit admission with mortality among older patients with pneumonia. *JAMA*. 2015;314(12):1272-1279. doi:10.1001/jama.2015.11068

9. Pirracchio R, Sprung C, Payen D, Chevret S. Benefits of icu admission in critically ill patients: Whether instrumental variable methods or propensity scores should be used. *BMC medical research methodology*. 2011;11(1):1.

10. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Statistics in medicine*. 2014;33(13):2297-2340.

11. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 1996;91(434):444-455.

12. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *The international journal of biostatistics*. 2007;3(1):14.

13. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Annals of internal medicine*. 2014;161(2):131-138.

14. Soumerai SB, Koppel R. The reliability of instrumental variables in health care effectiveness research: Less is more. *Health services research*. 2017;52(1):9-15.

15. Swanson SA, Hernán MA. The challenging interpretation of instrumental variable estimates under monotonicity. *International Journal of Epidemiology*. 2017:dyx038.

16. (SPOT)light Investigators. *(SPOT)light Study Website*.; 2010.

17. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: The intensive care national audit & research centre (icnarc) model. *Critical care medicine*. 2007;35(4):1091-1098.

18. Williams B, Alberti G, Ball C, et al. *National Early Warning Score (News): Standardising the Assessment of Acute-Illness Severity in the Nhs*. London: The Royal College of Physicians; 2012.

19. Vincent J-L, Moreno R, Takala J, et al. The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*. 1996;22:707-710.

20. Danbury C, Gould T, Baudouin S. *Guidelines for the Provision of Intensive Care Services*. London: The Intensive Care Society; 2015.

21. Kennedy EH, Harris S, Keele LJ. Survivor-complier effects in the presence of selection on treatment, with application to a study of prompt icu admission. *Journal of the American Statistical Association*. 2018;(just-accepted):1-40.

22. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology*. 2013;24(3):370-374.

23. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: A systematic review. *Epidemiology*. 2013;24(3):363-369.

24. Davies NM. Commentary: An even clearer portrait of bias in observational studies? *Epidemiology (Cambridge, Mass)*. 2015;26(4):505.

25. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology*. 2015;26(4):498-504.

26. Zhao Q, Small DS. Graphical diagnosis of confounding bias in instrumental variables analysis. *Epidemiology*. 2018;In Press.

27. Kahn J, Ten Have T, Iwashyna T. The relationship between hospital volume and mortality in mechanical ventilation: An instrumental variable analysis. *Health Services Research*. 2009;44(3):862-879.

28. Pizer SD. Falsification testing of instrumental variables methods for comparative effectiveness research. *Health services research*. 2016;51(2):790-811.

29. Yang F, Zubizaretta J, Small DS, Lorch S, Rosenbaum P. Dissonant conclusions when testing the validity of an instrumental variable. *The American Statistician*. 2014;68(4):253-263.

30. Keele LJ, Zhao Q, Kelz RR, Small DS. Falsification tests for instrumental variable desings with an application to tendency to operate. *Medical Care*. 2019;In press.

31. Vansteelandt S, Didelez V. Robustness and efficiency of covariate adjusted linear instrumental variable estimators. *The Scandinavian Journal of Statistics*. 2018;In press.

32. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *American journal of epidemiology*. 2008;169(3):273-284.

33. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business & Economic Statistics*. 2001;19(1):2-28.

34. Swanson SA, Hernán MA. Think globally, act globally: An epidemiologist's perspective on instrumental variable estimation. *Statistical science: a review journal of the Institute of Mathematical Statistics*. 2014;29(3):371.

35. Imbens GW, Rosenbaum P. Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *Journal of The Royal Statistical Society Series A*. 2005;168(1):109-126.