Statistics in review
Part 2: Generalised linear models, time-to-event and time-series analysis, evidence synthesis and clinical trials

John L Moran and Patricia J Solomon

ABSTRACT

In Part I, we reviewed graphical display and data summary, followed by a consideration of linear regression models. Generalised linear models, structured in terms of an exponential response distribution and link function, are now introduced, subsuming logistic and Poisson regression. Time-to-event ("survival") analysis is developed from basic principles of hazard rate, and survival, cumulative distribution and density functions. Semi-parametric (Cox) and parametric (accelerated failure time) regression models are contrasted. Time-series analysis is explicated in terms of trend, seasonal, and other cyclical and irregular components, and further illustrated by development of a classical Box–Jenkins ARMA (autoregressive moving average) model for monthly ICU-patient hospital mortality rates recorded over 11 years. Multilevel (random-effects) models and principles of meta-analysis are outlined, and the review concludes with a brief consideration of important statistical aspects of clinical trials: sample size determination, interim analysis and "early stopping".

Crit Care Resusc 2007; 9: 187-197

In Part I of this review,[1] we established the basis of an approach to statistical analysis with a review of graphical display and data summary, followed by a consideration of linear regression models. Traditionally, an exposition of classical ordinary least squares (OLS) regression is followed by separate discussions of Poisson (count data) and logistic (binary data) regression models. However, we may subsume all these regression models within the generalised linear model framework.[2,3]

Generalised linear models

Generalised linear models (GLMs) are transformations of the classical linear regression model and are distinguished by model, rather than data, transformations: specifically, a response distribution of one of the exponential family of distributions (normal, Poisson, gamma, binomial or inverse Gaussian[4]) and a monotonic link function (identity, logarithmic, square root, logistic or power), which relates the mean of the response to a scale on which the model effects combine additively. The general techniques used to analyse both continuous and discrete data are synthesised into a unified conceptual framework; the properties of the response variable are matched by the particular distribution, the variance is a function of the mean (var $(y|x) = {}^{2} v(x)$), except for the normal distribution, where the mean and variance are independent. Thus, OLS can be formulated as a GLM with the normal (Gaussian) distribution and identity link. GLMs are fitted by either maximum likelihood or iteratively reweighted least squares, and a key parameter is the deviance, 2log (where log = likelihood [full or "saturated" model] likelihood [null or intercept only model]). For the normal distribution model, the deviance is the residual sum of squares, and hence the notion of R^2 (= 1 [residual sum of squares/total sum of squares]), may be interpreted as the familiar "percent variance explained". Although there are "pseudo-R^2" statistics for the GLM, the deviance for non-normal distributions is different from the residual sum of squares, and the scalar values of these various statistics are not monotone transformations, as would apply to the normal linear model.[5]

ICU length of stay data have been formally analysed using a GLM with Gaussian distribution and log-link[6] (that is, in OLS with a log transformation (g), the expectation (E) is $E(g(Y_i)) = + x$; for the GLM, the form of the expectation is $g(E(Y)) = + x$; the GLM log-links the predictor (x) rather than the response; and the predicted values are on the *same* scale as the response, Y) and ICU cost data using the inverse Gaussian family and log-link.[5]

The logistic regression model,[7] a model of binary responses ("successes and failures"; coded as 1, 0), uses the Bernoulli distribution (a "degenerate" case of the

Table 1. Mantel–Haenszel odds ratio estimates for age–mortality effect, over calendar year, controlling for strata (n = 36)

Year	Odds ratio (OR)	*P*	OR 95% confidence limits Lower	Upper
1993	1.020	0.0001	1.016	1.025
1994	1.025	0.0001	1.023	1.028
1995	1.025	0.0001	1.023	1.027
1996	1.025	0.0001	1.023	1.027
1997	1.028	0.0001	1.026	1.030
1998	1.026	0.0001	1.023	1.028
1999	1.027	0.0001	1.024	1.029
2000	1.026	0.0001	1.024	1.028
2001	1.028	0.0001	1.026	1.030
2002	1.027	0.0001	1.025	1.028
2003	1.027	0.0001	1.025	1.029

* Data for this table and for figures and other examples were obtained from the Australian and New Zealand Intensive Care Society (ANZICS) Adult Patient Database (1993–2003) with permission of the ANZICS Database Management Committee.

binomial distribution, where the number of trials = 1[8]) and the (canonical) logit (= log odds) link:

$$\log_e\left(\frac{p}{1-p}\right).$$

The Bernoulli variate (variate a set of all random variables that obey a given probabilistic law) is written $\boldsymbol{B}(1,p)$, whereas the general binomial variate is written $\boldsymbol{B}(n,p)$ (the number of successes in n independent Bernoulli trials; the random variable being the number of successes).[4] Thus,

$$P(Y=1)=\frac{1}{1+\exp\left[-(\alpha+\beta_1 X_1+....+\beta_k X_k)\right]}$$

where the denominator has a form similar to OLS (note that the estimates from the linear probability model, $_i$ = + x_i, [that is, OLS applied to a binary dependent variable] are inefficient with biased SE). This expression transforms to

$$\log\left(\frac{p}{1-p}\right)=\alpha+\beta_1 X_1+...\beta_k X_k$$

(which is approximately linear from p [ ] = 0.2 0.8). To generate probabilities from the coefficients of a given predictive algorithm, say APACHE II,[9] the equation would be:

$$P=\frac{1}{\left(1+2.718218^{-(\text{constant}+\beta_1\times\text{data}+....\beta_k\times\text{data})}\right)},$$

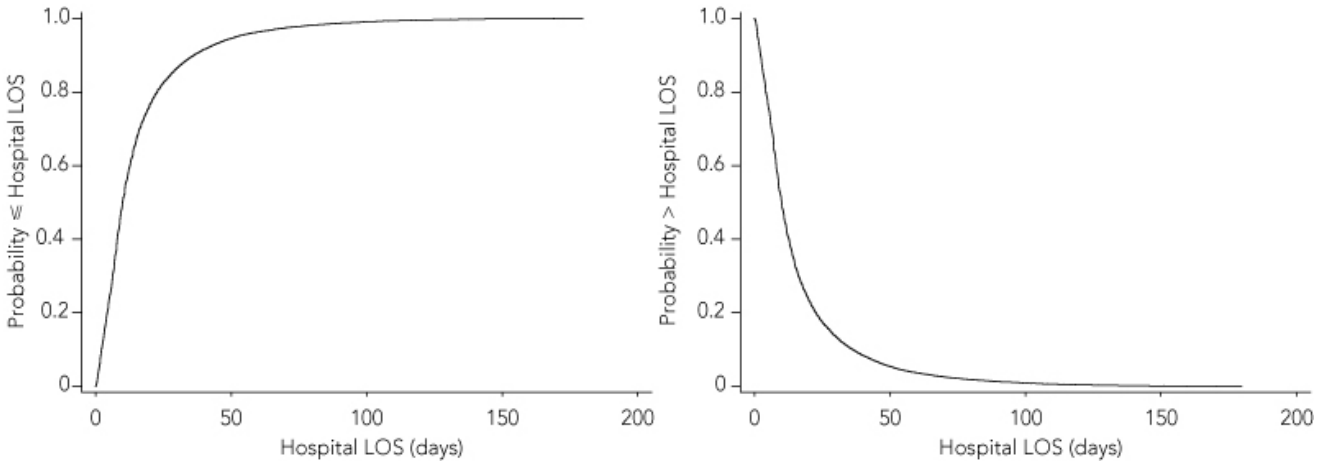
where 2.718218 = "e", with death coded 1, and survival coded 0; that is, for the $P(Y = 1)$. The predicted values for such a logistic regression are mortality probabilities, and predictions can be considered as conditional, at particular fixed values of covariates, and marginal, averaging conditional predictions over a distribution of values for the covariates.[10,11] Beta coefficients from a logistic regression are exponentiated as (log) odds ratio (OR), and have the conventional interpretation of increase in (log) OR per unit increase in predictor (reflecting a non-linear re-parameterisation of probabilities[11]). That is, logistic regression is a linear additive model for the log(odds), but multiplicative for the odds; an increase of x by 1 increases the logit by $_x$, but multiplies the odds by e . The slope of the effect of x on the probability scale is non-linear (the maximum slope of effect, at $P = 0.5$, is given by 0.25). When the incidence of outcome of interest is $>$ 10% (that is, not "rare"), the OR cannot be equated with risk ratio (RR). An estimate of the RR is

$$\frac{OR}{(1-P_0)+(P_0\times OR)},$$

where P_0 is the incidence of outcome of interest.[12] However, this adjustment has been criticised,[13] and RR may be estimated directly using a GLM (binomial family and log-link[14]). Model generation using logistic regression proceeds on the lines[7] described under *Model selection* in Part 1 of this review.[1] An important caveat is the question of number of events (say, deaths) per predictor variable; for fewer than 10 events per variable, estimation may be problematic and reflected in coefficient bias.[15]

Although logistic regression is a standard means of analysis, yielding parameter estimates of variable effect as OR, it is sometimes useful to use simpler techniques to illustrate effect measures with stratified analysis, in particular Mantel–Haenszel ORs.[16,17] For instance, this may be useful to understand the interaction of a continuous and a categorical variable(s). If, as part of an observational analysis in the Australian and New Zealand Intensive Care Society (ANZICS) database (years, 1993–2003), we were interested in the effect of, say, age on hospital outcome over time, across both ICU category (four levels) and locality (nine levels), we might compute the Mantel–Haenszel OR for age, over calendar years,

Figure 1. Cumulative distribution and "survival" functions of hospital length of stay (LOS) of ICU patients



controlling for strata ($n = 36$) formed by the cross-product of ICU category and locality, as shown in Table 1.

The overall Mantel–Haenszel OR estimate, controlling for ICU category and locality, was 1.026 (95% CI, 1.025–1.026; $P = 0.0001$). The overall score test for homogeneity of OR over calendar years was marginally significant at $P = 0.08$. For strata of ICU category and locality analysed separately, the score test for homogeneity of OR over calendar years yielded P values of 0.03 and 0.14, respectively, suggesting that the ICU levels "averaged out" for each locality.

When the dependent variable is a count (for instance, number of admissions per day to an emergency department), Poisson regression is appropriate, which in GLM terms incorporates the Poisson family (the Poisson variate is written $Po(\ )$, with range $0 \leqslant x <$, and the parameter mean is , the SD being equal to the mean) and the log-link.[18]

Time-to-event (survival) analysis

If we are interested in the time of an event (say, death), then we may chose to use survival or time-to-event analysis.[19] The time variable may be continuous (event occurs at any time) or discrete (event occurs within time units), and we may "observe" the event (coded, say, as 1) or we may not (for instance, the period of observation in a trial may end at a certain time), and the subject(s) observation history would be considered as right-censored (coded, say, as 0), and not discarded. If the time to the event of interest is, say, T then time will be a non-negative random variable from a homogeneous population and will have a particular distribution, characterised by the following functions, which, given that one is known, are completely determined:[20,21]

survival function — the probability of a subject surviving to time t: $S(t) = P(T > t)$; is equal to unity at $t = 0$, and decreases monotonically towards 0 with increase in t;

cumulative distribution function (cdf): $F(t) = P(T \leqslant t)$; thus $S(t) = 1 \quad F(t)$;

- the cdf and "survival" function for the hospital length of stay (days) of ICU patients are shown in Figure 1.

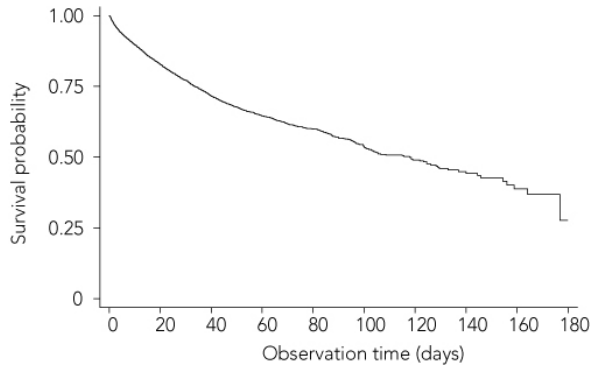
density function — the unconditional failure rate of the event occurring in the interval: $f(t) = dF(t)/dt$;

hazard (rate) or force of mortality — the instantaneous probability that the event occurs in a given interval (which is *conditional* on the subject surviving to the beginning of that interval) divided by the interval width:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t \mid T > t)}{\Delta t}.$$

The hazard rate has limits from 0 (no risk) to infinity (failure certainty at that instant), and may vary over time. Under a constant hazard assumption (for example, exponentially distributed failure times), the average hazard rate may be thought of as the (number of failures)/(sum of observed survival times), or the rate of

Figure 2. Kaplan–Meier estimate of survival of ICU patients observed until hospital discharge (≤ 180 days) or death



events per person-time,[22] and has similarities to the parameter of the Poisson distribution;[2] time-to-event analysis may be formulated as a Poisson process.[23-25]

The cumulative hazard function $H(t)$, the integral of (instantaneous) hazard rates from time zero to time t,

$$H(t) = \int_0^t h(x)dx,$$

is equal to $-\ln(S(t))$, and $S(t) = (-\exp(H(t))$. At $t = 0$, $S(t) = 1$ and $H(t) = 0$; at $t =$, $S(t) = 0$ and $H(t) =$. A cumulative hazard rate of "x"/day indicates an expectation of "x" events over the whole day (the probability of this $= 1 - \exp(-x)$), albeit the instantaneous hazard rate profile may be quite variable. Conversely, for a constant hazard rate of x/day, an event would occur after $1/x$ days (a Poisson counting process with $= x$).[21] For small numerical values of $H(t)$ (say, < 0.05), the cumulative risk (CR) is also approximated by $1 - S(t) = 1 - \exp(-H(t))$, under the condition of no competing risks,[26] albeit the cumulative hazard has no strict probabilistic interpretation. For a general proportional or constant (negative exponential) hazard model (see below), the hazard ratio can also be approximated to $p/(1-p)$ odds, where the interpretation of p is, say, treatment A healing faster than treatment B.[22] If the full event history of a patient is unobserved, then censoring is said to occur: right-censoring if the period of study observation ceases before the "event"; left-censoring if the event occurs before the study beginning; and left-truncation when history before the first observation point is unobserved. Information is contributed to failure times by uncensored cases and to survival by censored,[27] the hazard rate $h(t)$ being understood as the ratio of unconditional failure to survival.

The familiar Kaplan–Meier (product–limit) estimator[28] is a non-parametric estimate of $S(t)$; if the observed failure times are $t_1, \ldots, t_k$, there being k distinct failure times, n_j is the number of individuals at risk at time t_j, and d_j is the failure number at t_j; then

$$\hat{S} = \prod_{j|t_j \le t} \left(\frac{n_j - d_j}{n_j} \right),$$

where the product ( ) is over all failure times t. The Kaplan–Meier estimator is a step function with "jumps" at observed event times, with the size of the jumps depending on the number of events at each t_j and the censoring pattern prior to t_j.[20] Figure 2 shows the Kaplan–Meier survival estimate for ICU patients observed in hospital for up to 180 days or until death (patients discharged alive are censored).

A general regression approach, allowing predictors, may be formulated as

$$t_j = \beta_0 + \beta_1 x_j + \varepsilon_j \text{ (or } \ln(t_j) = \ldots\ldots) \text{ or}$$

$$h_j(t) = g\left(t, \beta_0 + x_j \beta_x\right),$$

where $g()$ is a particular function.[21] Proportional hazards models are of the form

$$h_j(t) = h_0(t) \exp\left(\beta_0 + x_j \beta_x\right),$$

in that the subject hazard is multiplicatively related to the common baseline hazard (h_0). The Cox model, the standard medical survival model,[29] is a semi-parametric proportional hazards model where the intercept ($_0$) has been incorporated into h_0, which is not parameterised, and is thus written

$$h_j(t) = h_0(t) \exp\left(x_j \beta_x\right).$$

It expresses the multiplicative effect of covariates on the hazard (force of mortality) via conditioning on the ordered failure times and, by so analytically "combining" individual binary-outcome analysis at these failure times,[21] is indifferent to any information between them. Covariate estimates as hazard ratios

$$\frac{h_i(t)}{h_0(t)} = \exp\left(\beta'\left(x_i - x_j\right)\right),$$

are interpreted as the increase or decrease in hazard per unit change in covariate. In the presence of non-

proportional hazards, the Cox model may be modified by covariate stratification or the incorporation of time-varying covariates.[30]

Survival time may be modelled by parametric analysis (accelerated failure time [AFT] models), where (log) time (t) is parameterised thus: $\ln(t_j) = x_j \beta + \ln(\varepsilon_j)$; x is the covariate vector and β the corresponding coefficient; and the random quantity $\ln(\varepsilon_j)$ has a specified distribution (exponential, Weibull, log-normal, log-logistic or gamma).[31] This is not to say that parametric proportional hazards formulations do not exist,[32] merely that they are not implemented in most statistical packages. The most intuitive manner in which to express AFT model coefficients is in the exponentiated form, as time ratios (TR = t_j^*/t_j) for a unit increment change in the covariate

$$t_j = \exp(\beta_1 x_1 + ...)\tau_j \text{ and}$$

$$t_j^* = \exp\{\beta_1 (x_1 + 1) + ...\}\tau_j .^{21}$$

Thus TR < 1 is associated with a decrease in (survival) time, and TR > 1 is associated with a prolonged (survival) time, or, more accurately, a contraction or expansion of time to failure. The Weibull and exponential AFT also have a proportional hazards interpretation. There is a comparability of representation between the forms (proportional hazard, semi-parametric and AFT) of the "general regression approach" above, and this may be specified as

$$\lambda(t \mid Z) = \lambda_0 \{t \exp(\beta_1 Z)\} \exp(\beta_2 Z),$$

where $\lambda(t|Z)$ is the hazard function at time t, Z is the p-vector covariate and λ_0 is the baseline hazard function. The model becomes proportional hazards if $\beta_1 = 0$, AFT when $\beta_1 = \beta_2$, and accelerated hazards (scale change in the hazard function without altering the overall shape of the baseline hazard function) when $\beta_2 = 0$.[33]

The choice between non-nested models in the current context — Cox regression versus AFT — is not facile, as the basis for estimating the Akaike and Bayesian information criteria (AIC and BIC) differs (partial versus maximum likelihood, respectively), rendering strict comparison somewhat tenuous. The proportional hazards assumption of the Cox model requires that covariate effect be constant on the relative hazard scale, which may be unrealistic; and, in pathophysiological states where initial (mortality) event rates are high, AFT models may possess advantage,[34] although the use of time-varying covariates with the Cox model has utility in these circumstances.[30] Logistic and Cox regression will generally produce similar results (that is, the effects of the covariates on failure/survival are proportional under the assumptions outlined) depending on the follow-up time and pattern of events (or failures) and censoring.[35] However, logistic regression is "wasteful of data", to the extent that the "distribution of survival times" (and, more importantly, the hazard) cannot be readily obtained with the logistic function[36] except through, say, a piece-wise survival approach (a discrete-time hazard model).[37]

Time series

Few examples of uni- or multivariate time series analysis[38] where specific account is taken of the auto-correlation among repeated responses over time have appeared in the critical care literature.[39] This is not the case in the general biomedical literature where, for instance, the relation between pollution levels and mortality has been extensively explored using this paradigm.[40] A time series (or stochastic) process is a sequence of random variables (Y_0, Y_1, Y_2, …) ordered by time, there being only one observation at each time point, and the assumption is that, given enough time (which from a "practical" viewpoint, may never occur), observations are independent. The set of all possible realisations of a time series process is equivalent to the concept of "population".[41] The approach to time series is analysis in the domain of time (based on the autocorrelation or autocovariance function) or frequency (based on the spectral density function).[42] Classically,[43] a time series is decomposed into trend, seasonal, and other cyclical and irregular components; trend and seasonal components are removed (for example, by differencing, $x_t = x_t - x_{t-1}$) to achieve a "stationary" time series, which is then modelled by, say, an autoregressive moving average (ARMA) process:

$$y_t = \alpha + \phi_1 y_{t-1} + ...\phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ...\theta_q \varepsilon_{t-q},$$

where y_t is the "differenced" series; ϕ_1, ϕ_2, …, ϕ_p are the "autoregressive" constants relating the value of y at time t to its past p values; θ_1, θ_2, …, θ_q are the "moving average" coefficients, relating the current "white-noise" ε_t to its past q values; and $\varepsilon_t \sim N(0, \sigma^2)$.[44] For such stationary processes, y_t (= $\mu + u_t$) is equal to some mean level (μ_0) plus a zero mean error process ($u_t = \phi_1 u_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$); and the specification "t 1" is

equivalent to "ARMA(1,1)". The general form of the ARMA (p,q) model, where p represents the number of autoregressive and q the number of moving average terms, may be represented using the lag operator notation as

$$\left(1-\phi_1 L-\ldots-\phi_p L^p\right)\left(y_t-\beta_0\right)=\left(1+\theta_1 L+\ldots+\theta_q L^q\right)\varepsilon_t .^{45}$$

This classical (Box–Jenkins) approach is based on prediction/forecasting, rather than "explanation".

Figure 3 illustrates trend and seasonality using the overall monthly hospital mortality rate of Australian and New Zealand ICU patients, 1993–2003: the top left panel (A) shows the characteristic time-series data display; the top right panel (B) shows a 12-month centred moving-average curve, revealing a definite trend for reduction in mortality; the bottom left panel (C) shows a 3-month centred moving-average curve, which is coincident and barely distinguishable from the data curve; and the bottom right panel (D) shows a realisation of the latter three-period centred moving-average via a pooled seasonal indicator for normalised mortality, with a peak in mortality in July (winter season). Figure 4 shows the first differenced series; two distinct seasonal ARMA models were fitted to the data (the number of terms and the indication for seasonality were determined by examination of partial autocorrelation and spectral density plots [periodogram and cumulative spectral distribution[46]]):

(i) an additive model, as ARMA(1,1) for monthly + ARMA(0,12) for seasonal variation:

$$y_t=\alpha+\phi_1 y_{t-1}+\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_{12}\varepsilon_{t-12} \quad \text{or}$$

$$\left(1-\phi_1 L\right)\Delta y_t=\left(1+\theta_1 L+\theta_{12}L^{12}\right)$$

using lag-operator notation, where represents first differencing and $_1 = 0.23$ (95% CI, 0.07–0.39; $P = 0.006$), $_1 =$ 0.97 (95% CI, 1.07 to 0.86; $P = 0.0001$), and $_{12} = 0.12$ (95% CI, 0.01–0.24; $P = 0.03$).

Figure 3. Time series analysis of mean hospital monthly mortality rates in ICU patients, 1993–2003

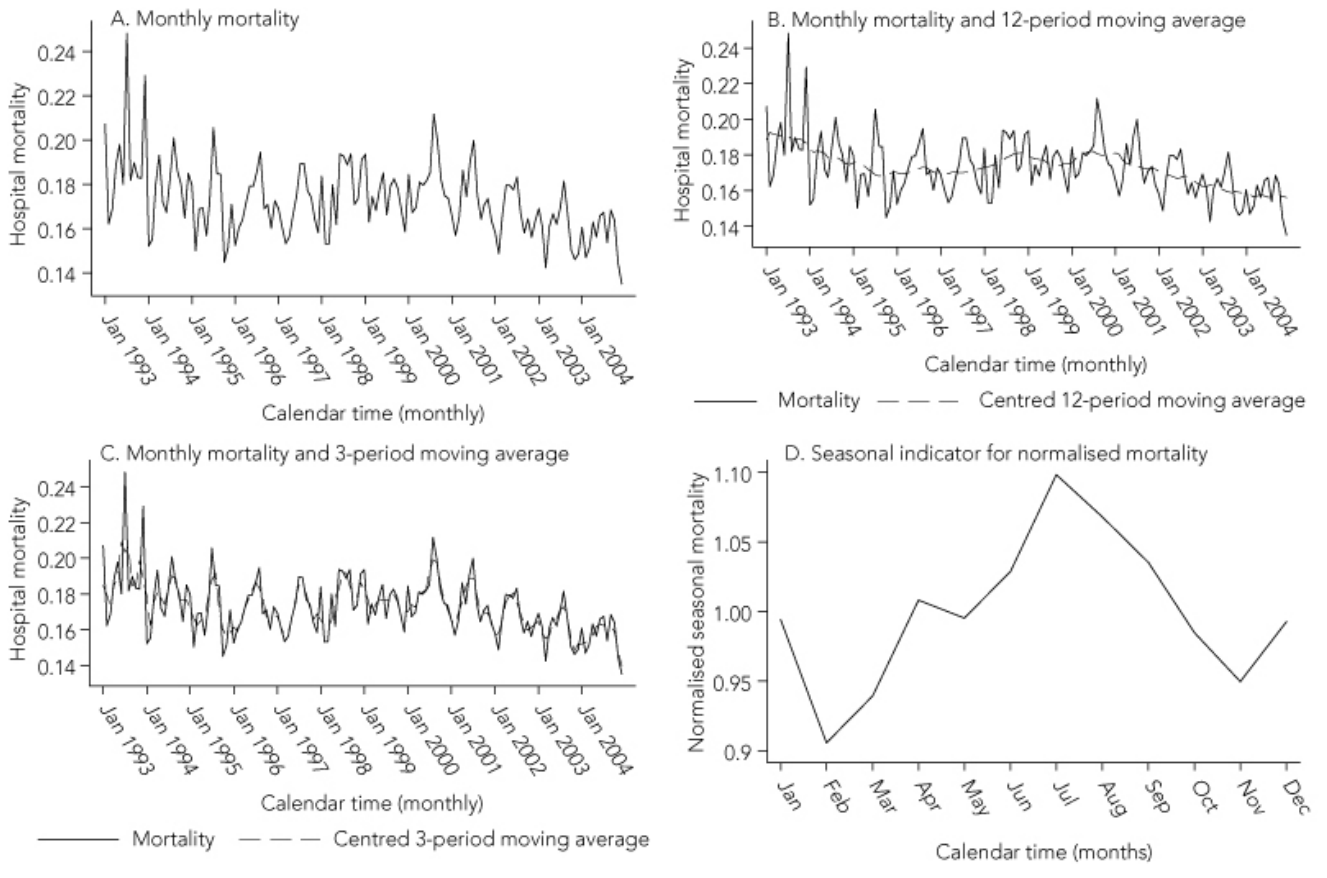
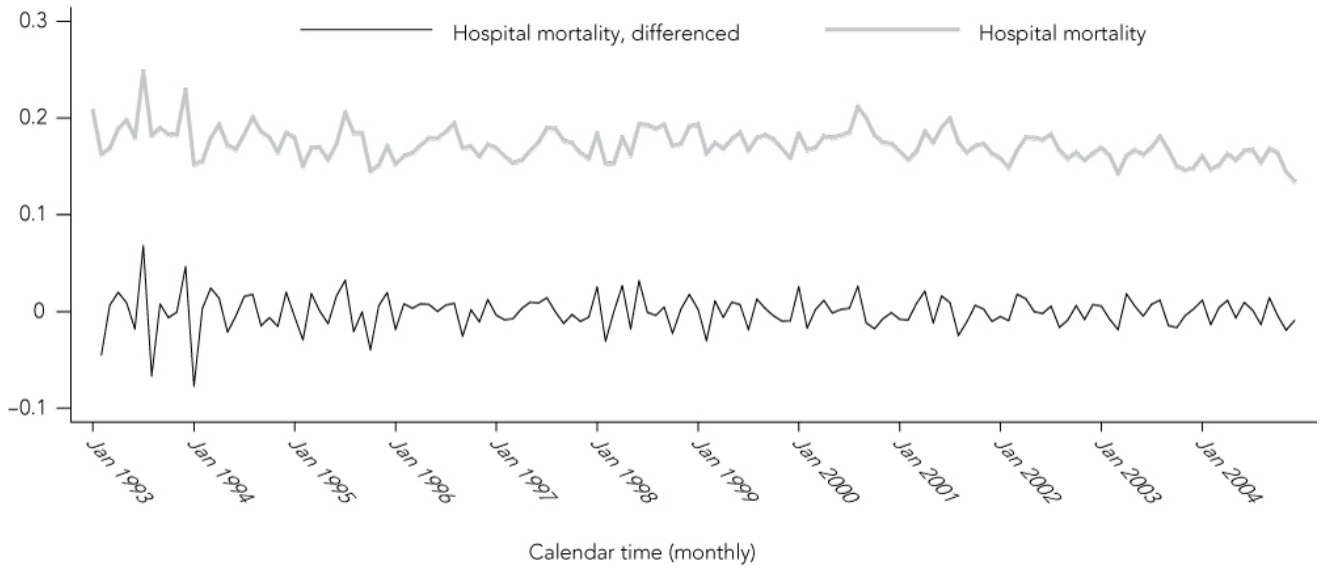


Figure 4. First differencing of hospital mortality rates



(ii) a multiplicative model, as ARIMA $(p,d,q) \times (P,D,Q)_s$, that is the multiplication of non-seasonal and seasonal factors, where d and D indicate the differencing, and s the seasonality (here 12 for monthly data), as ARIMA $(0,1,1) \times (1,0,1)_{12}$:

$$\left(1-\phi_{12,1}L^{12}\right)\Delta\Delta_{12}^{0}y_t = \left(1+\theta_1 L\right)\left(1+\theta_{12,1}L^{12}\right)\varepsilon_t ,$$

where $_{12,1} = 0.95$ (95% CI, 0.86–1.04; $P = 0.0001$); $_1 =$ 0.87 (95% CI, 0.97 to 0.77; $P = 0.0001$); and $_{12,1} =$ 0.76 (95% CI, 1.01 to 0.52; $P = 0.0001$).

The multiplicative model had some advantage over the additive model in terms of a lesser BIC (805.1 versus 783.7; note the sign of the BIC), but the additive model performed better in terms of forecast error (plotted as the difference between the raw data and two forecast estimates):[38] one-step-ahead (forecast at time t based on time t 1), and dynamic (forecast from time t based on forecasts at times $t + k$), as seen in Figure 5. The better performance of the additive approach is understandable in terms of the raw data plot (top panel of Figure 4), where the seasonal effect does not appear to be proportional to the mean of the data (a circumstance where the multiplicative approach would be preferred). The evidence for annual seasonality seems captured by both approaches. The MA(1) parameter indicates that any (latent) "shocks" to the system (determining mortality, that is) are transient, lasting no longer than 1 lag. The AR(1) parameter indicates that the current value is linearly related to its past value plus an additive (stochastic) shock. The scalar value of this AR parameter, as opposed to that of the MA parameter, determines the memory of the shocks of the system, which for the seasonal component of the multiplicative approach are substantial (0.95), reflecting a very slow cycle with a long period, corresponding to the observed gradual decline in mortality.

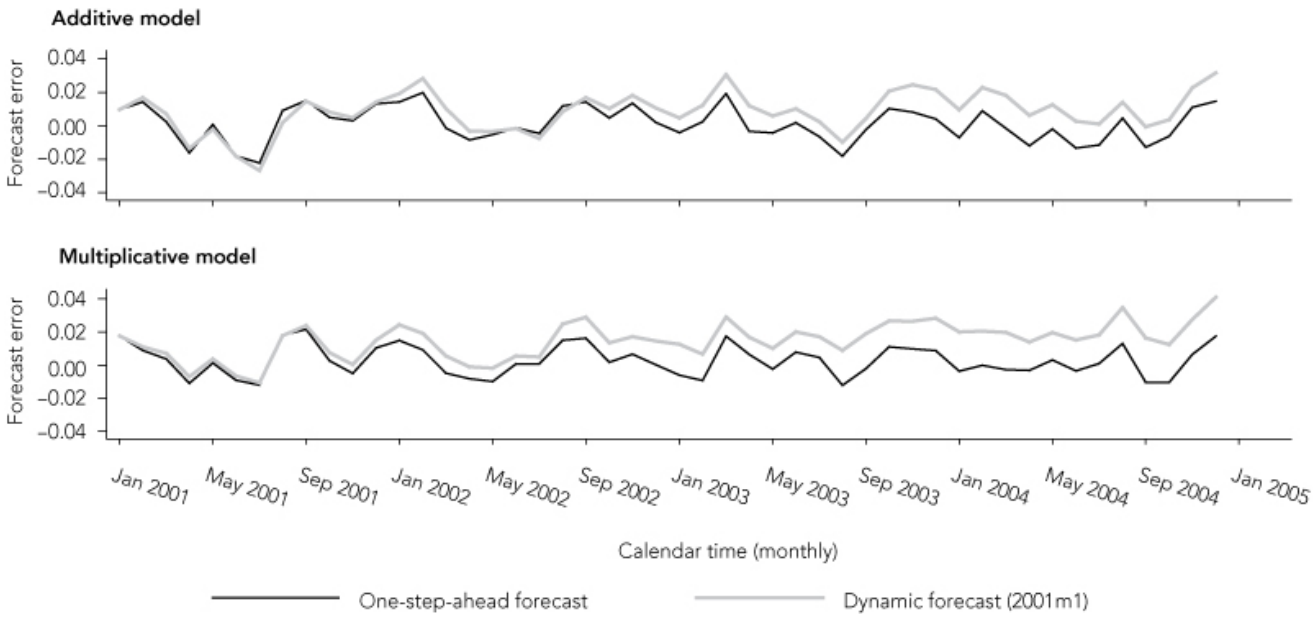
Random effects, mixed or multilevel models

In the regression context, the explanation of dependent variability in terms of covariate(s) variability may be conceptualised as being composed of "observed" and "unobserved" heterogeneity.[47] The latter may be accounted for by means of random effects modelling, a rapidly expanding field in the social and economic sciences and in the medical literature.[24,48] Thus, say, for subject i, the total residual is partitioned into a subject-specific component u_i (random intercept) plus residual $_i$:

$$y_i = \beta^T x_i + u_i + \varepsilon_i$$

where $^T x_i$ is the "linear predictor", $_0 +$ $x_{1i} +$ $x_{2i} + \ldots$.[49] GLM may be extended using this approach to generalised linear mixed models (GLMM).[50]

Figure 5. Forecast errors for additive and multiplicative (ARIMA) models*



* In an ARIMA(p,d,q) model, the data are differenced, d times, to obtain a stationary series, and then an ARMA(p,q) model is fitted to the differenced data.

Meta-analysis

Since its introduction in 1976 by Glass,[51] meta-analysis has become an established, albeit controversial,[52] method of review in the medical literature.[53,54] The areas of disputation may be classified as:

procedural — relating to the individual studies of the particular meta-analysis, with questions relating to search strategy, study quality, the impact of large versus small studies and the propriety of pooling the studies in the first place; and

statistical — relating to the choice of the metric of the treatment effect in the case of binary outcomes, the diagnosis of and adjustment for both heterogeneity of treatment effects and publication bias, and the relation of the treatment effect to underlying demographic and patient characteristics (metaregression).[55]

The particular techniques and processes of meta-analysis have been comprehensively reviewed;[56,57] pooled efficacy estimates are expressed, for binary outcomes, as RR, OR and RD (see under *Effect measures* in Part 1 of this review[1]) and, for continuous variables, as weighted or standardised mean difference. Two key concepts which may confound any pooled estimate of treatment effect[58] are:

heterogeneity, which reflects clinical, methodological and statistical features of the component studies (the search for "predictors of between-study heterogeneity" may be the "primary value" of meta-analysis[59]); and

publication bias, "the major problem in meta-analysis",[60] although it has been noted that meta-analyses may be "resistant" to publication bias.[55]

Sample size determination

The interpretation of the results of randomised clinical trials (RCTs) are bedevilled by questions of sample size; the classical study of Freiman et al in 1978[61] demonstrated inadequate sample size as one of the causes of "no difference from control", a finding that has been confirmed more recently.[62] Consideration of appropriate effect size (minimally clinically significant difference[63]) is important; unrealistic expectations of therapy derived from, for example, initial "pilot" studies, may compromise RCT outcomes.[64] In the critically ill, the average treatment effect of "positive" trials was an RD of 8.7% (range, 3.4%–16%).[65] Simple guidelines to sample size determination have been provided;[66-68] more complex derivations are available.[69]

Monitoring and conduct of clinical trials

Monitoring the conduct and the analysis of an RCT,[67,70,71] in particular the impact of interim analyses and early stopping,[72] is an important and controversial issue — more so as the four major trials reporting efficacy in the critically ill[73-76] have used interim analyses and/or early stopping. Recent reports have highlighted inadequacies in the analysis and interpretation of such trials,[77-79] and critical care practitioners need to be cognisant of these controversies. Further substantive statistical issues,[80] such as the centre effect in multicentre trials,[81] post-randomisation adjustment of covariates,[82] multiple end-points,[83] selection bias,[84] and techniques of randomised allocation,[85] are beyond the scope of this review.

Caveats

The current wealth of statistical analytic instruments and software packages is impressive, but cautions must apply to analytic endeavour, especially where observational studies are concerned: no amount of statistical "adjustment" will transform observational studies into "natural experiments" (the illusion of statistical control[86]), and the tendency to think of all regression coefficients as causal effects should be resisted.[87] What should be insisted upon is the Fisherian requirement of "rigorous uncertainty".[88]

Postscript

The seminal paper by Kaplan and Meier, *Non parametric estimation from incomplete observations* published in 1958,[28] is one of the most quoted papers in the biomedical and statistical literature. Edward Kaplan died in 2006 at Corvallis, Oregon, USA (http://www.imstat.org/bulletin/Bulletin35_10.pdf).

Acknowledgements

We thank the Australian and New Zealand Intensive Care Society (ANZICS) Database Management Committee for permission to use data from the ANZICS Adult Patient Database (1993–2003) (http://www.anzics.com.au).

Author details

John L Moran, Senior Consultant[1]
Patricia J Solomon, Associate Professor[2]
1 Department of Intensive Care Medicine, Queen Elizabeth Hospital, Adelaide, SA.
2 School of Mathematical Sciences, University of Adelaide, SA.
Correspondence: john.moran@adelaide.edu.au

References

1 Moran JL, Solomon PJ. Statistics in review. Part I: graphics, data summary and linear models. *Crit Care Resusc* 2007; 9:81-90,
2 Hoffman JP. Generalized linear models: an applied approach. Boston, Mass: Pearson Education, 2004.
3 McCullagh P, Nelder JA. Generalized linear models. 2nd ed. London: Chapman and Hall, 1989.
4 Evans M, Hastings N, Peacock B. Statistical distributions. 3rd ed. New York: John Wiley and Sons, 2000.
5 Angus DC, Linde-Zwirble WT, Sirio CA, et al. The effect of managed care on ICU length of stay: implications for medicare. *JAMA* 1996; 276: 1075-82.
6 Moran JL, Solomon PJ, Peisach AR, Martin J. New models for old questions: generalized linear models for cost prediction. *J Eval Clin Pract* 2007; 13: 381-9.
7 Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: John Wiley and Sons, 2000.
8 Hardin J, Hilbe J. Generalized linear models and extensions. College Station, Tex: Stata Press, 2001.
9 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13: 818-29.
10 Wilcosky TC, Chambless LE. A comparison of direct adjustment and regression adjustment of epidemiologic measures. *J Chronic Dis* 1985; 38: 849-56.
11 Brant R. Digesting logistic regression results. *Am Stat* 1996; 50: 117-9.
12 Zhang J, Yu K. What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 1998; 280: 1690-1.
13 McNutt A, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003; 157: 940-3.
14 Hardin J, Cleves M. sbe29: generalized linear models: extensions to the binomial family. *Stata Technical Bulletin Reprints* 1999; 9: 140-6.
15 Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49: 1373-9.
16 Juul S. Stratified analysis. In: Juul S. An introduction to Stata for health researchers. College Station, Tex: Stata Press, StataCorp LP, 2006: 171-84.
17 Mantel N, Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; 22:719-48.
18 Long JS. Regression models for categorical and limited dependent variables. Thousand Oaks, Cal: Sage Publications, 1997.
19 Kalbfleisch JD, Prentice RL. Statistical analysis of failure time data. 2nd ed. Hoboken, NJ: JD Wiley and Sons, 2002.
20 Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. 2nd ed. New York: Springer-Verlag, 2003.
21 Cleves MA, Gould WW, Gutierrez RG. An introduction to survival analysis using Stata. Revised ed. College Station, Tex: Stata Press, 2004.
22 Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother* 2004; 48: 2787-92.
23 Clayton D. The analysis of event history data: a review of progress and outstanding problems. *Stat Med* 1988; 7: 819-41.
24 Lindsey JK. Counts and times to events. *Stat Med* 1998; 17: 1745-51.
25 Moran JL, Solomon PJ, Williams PJ. Assessment of outcome over a 10-year period of patients admitted to a multidisciplinary adult intensive care unit with haematological and solid tumours. *Anaesth Intensive Care* 2005; 33: 26-35.
26 Moran JL, Solomon PJ. Therapeutic hypothermia after cardiac arrest – once again. *Crit Care Resusc* 2006; 8: 151-4.

27 Box-Steffensmeier JM, Jones BS. Event history modeling: a guide for social scientists. Cambridge: Cambridge University Press, 2004.
28 Kaplan EL, Meier, P. Non parametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457-81.
29 Tibshirani R. A plain man's guide to the proportional hazards model. *Clin Invest Med* 1982; 5: 63-8.
30 Moran JL, Bersten AD, Solomon PJ, et al, The Australian and New Zealand Intensive Care Society Clinical Trials Group. Modeling survival in acute severe illness: Cox versus accelerated failure time models. *J Eval Clin Pract* 2007. In press.
31 Hosmer DW Jr, Lemeshow S. Applied survival analysis: regression modeling of time to event data. New York: John Wiley and Sons, 1999.
32 Bruedel J, Diekmann A. The log-logistic rate model. Two generalisations with an application to demgraphic data. *Sociol Methods Res* 1995; 24: 158-86.
33 Chen YQ. Accelerated hazards regression model and its adequacy for censored survival data. *Biometrics* 2001; 57: 853-60.
34 Knaus WA, Harrell FE, Fisher CJ Jr, et al. The clinical evaluation of new drugs for sepsis. A prospective study design based on survival analysis. *JAMA* 1993; 270: 1233-41.
35 Green MS, Symons MJ. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *J Chronic Dis* 1983; 36: 715-23.
36 Hauck WW. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *J Chronic Dis* 1985; 38: 125-6.
37 Singer JB, Willett JB. Applied longitudinal analysis. New York: Oxford University Press, 2003.
38 Yaffee R, McGee M. Introduction to time series analysis and forecasting: with applications of SAS and SPSS. London: Academic Press, 2000.
39 Imhoff M, Bauer M, Gather U, Lohlein D. Statistical pattern detection in univariate time series of intensive care on-line monitoring data. *Intensive Care Med* 1998; 24: 1305-14.
40 Zeger SL, Irizarry R, Peng RD. On time series analysis of public health and biomedical data. *Annu Rev Public Health* 2006; 27: 57-79.
41 Wooldridge JM. Basic regression analysis with time series data. In: Wooldridge JM. Introductory econometrics: a modern approach. Cincinnati, Ohio: South-Western College Publishing, 2000.
42 Chatfield C. Some recent developments in time-series analysis. *J R Stat Soc [Ser A]* 1977; 140: 492-510.
43 Box GEP, Jenkins GM. Time-series analysis, forecasting and control. San Francisco, Calif: Holden-Day, 1970.
44 Ho M-HR, Shumway R, Ombao H. The state-space approach to modeling dynamic processes. In: Walls TA, Schafer JL. Models for intensive longitudinal data. New York: Oxford University Press, 2006: 148-75.
45 Diebold FX. Modeling cycles: MA, AR, and ARMA models. In: Diebold FX. Elements of forecasting. 2nd ed. Cincinnati, Ohio: South-Western College Publishing, 2001: 143-82.
46 Venables WN, Ripley BD. Time series analysis. In: Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer-Verlag, 2006: 387-418.
47 Skrondal A, Rabe-Hesketh S. Generalised latent variable modeling. Multilevel, longitudinal, and structural equation models. Boca Raton, Fla: Chapman and Hall/CRC, 2004.
48 Moran JL, Peter JV, Solomon PJ, et al. Tympanic temperature measurements – are they reliable in the critically ill? A clinical study of measures of agreement. *Crit Care Med* 2007; 35: 155-64.
49 Bingenheimer JB, Raudenbush SW. Statistical and substantive inferences in public health: issues in the application of multilevel models. *Annu Rev Public Health* 2004; 25: 53-77.
50 Hedeker D. Generalized linear mixed models. Encyclopedia of statistics in behavioral science. John Wiley and Sons, 2005. Available at: http://media.wiley.com/product_data/excerpt/04/04708608/0470860804-3.pdf (accessed Mar 2006).
51 Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res* 1976; 5: 3-8.
52 Mullen PD, Ramirez G. The promise and pitfalls of systematic reviews. *Annu Rev Public Health* 2006; 27: 81-102.
53 Cook DJ, Guyatt GH. Dissent. The professional meta-analyst: an evolutionary advantage. *J Clin Epidemiol* 1994; 47: 1327-9.
54 Rosendaal FR. Presentation. The emergence of a new species: the professional meta-analyst. *J Clin Epidemiol* 1994; 47: 1325-6.
55 Moran J, Solomon P, Warn D. Methodology in meta-analysis: a study from critical care meta-analytic practice. *Health Serv Outcomes Res Methodol* 2004; 5: 207-26.
56 Egger M, Smith GD, Altman DG. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Publishing Group, 2001.
57 Sutton AJ, Abrams KR, Jones DR, et al. Methods for meta-analysis in medical research. John Wiley and Sons, 2000.
58 Kaizar EE. Metaanalyses are observational studies: how lack of randomization impacts analysis. *Am J Gastroenterol* 2005; 100: 1233-6.
59 Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; 140: 290-6.
60 Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci* 1993; 703: 135-46.
61 Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978; 299: 690-4.
62 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials [see comments]. *JAMA* 1994; 272: 122-4.
63 Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *COPD* 2005; 2: 57-62.
64 Fitzmaurice GM. Sample size and power: how big is big enough. *Nutrition* 2002; 18: 289-90.
65 Moran JL, Peake SL, Solomon P. Hypothermia as therapy in cerebral injury. *Crit Care Resusc* 2002; 4: 93-102.
66 Glasziou P, Doll H. Was the study big enough? Two cafe rules. *Evid Based Med* 2006; 11: 69-70.
67 Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials* 1981; 2: 93-113.
68 Lenth RV. Some practical guidelines for effective sample size determination. *Am Stat* 2001; 55: 187-93.
69 Barthel FMS, Babiker A, Royston P, Parmar MK. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Stat Med* 2006; 25: 2521-42.
70 DeMets DL, Furberg CD, Friedman LM. Data monitoring in clinical trials: A case studies approach. New York: Springer Science +Business Media, 2006.
71 Ellenberg SS, Fleming TR, DeMets DL. Data monitoring committees in clinical trials: a practical perspective. Hoboken, NJ: John Wiley and Sons, 2002.
72 Moran JL, Solomon PJ. Some aspects of the design and monitoring of clinical trials. *Crit Care Resuscitation* 2003; 5: 137-46.
73 Bernard GR, Vincent JL, Laterre PF, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 2001; 344: 699-709.
74 Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001; 345: 1368-77.
75 Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med* 2000; 342: 1301-18.

76 van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in the critically ill patients. *N Engl J Med* 2001; 345: 1359-67.
77 Martin CM, Guyatt G, Montori VM. The sirens are singing: the perils of trusting trials stopped early and subgroup analyses. *Crit Care Med* 2005; 33: 1870-1.
78 Montori VM, Devereaux PJ, Neill KJA, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005; 294: 2203-9.
79 Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet* 2005; 365: 1657-61.
80 Senn S. Consensus and controversy in pharmaceutical statistics. *Statistician* 2000; 49: 135-76.
81 Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001; 135: 112-23.
82 Chen X, Liu M, Zhang J. A note on postrandomization adjustment of covariates. *Drug Inf J* 2005; 39: 373-83.
83 Sankoh AJ, D'Agostino RB, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Stat Med* 2003; 22: 3133-50.
84 Berger VW, Christophi CA. Randomization technique, allocation concealment, masking, and susceptibility of trials to selection bias. *J Mod Appl Stat Methods* 2003; 2: 80-6.
85 McEntegart DJ. The pursuit of balance using stratified and dynamic randomization techniques: an overview. *Drug Inf J* 2003; 37: 293-308.
86 Christenfeld NJ, Sloan RP, Carroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosom Med* 2004; 66: 868-75.
87 Schafer JL. Marginal modeling of intensive longitudinal data by generalized estimating equations. In: Walls TA, Schafer JL. Models for intensive longitudinal data. New York: Oxford University Press, 2006: 38-62.
88 Marks HM. Rigorous uncertainty: Why RA Fisher is important. *Int J Epidemiol* 2003; 32: 932-7.