Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Joint Models
[parts adapted from Henderson (2000) and Kolamunnage-Dona (2017)]

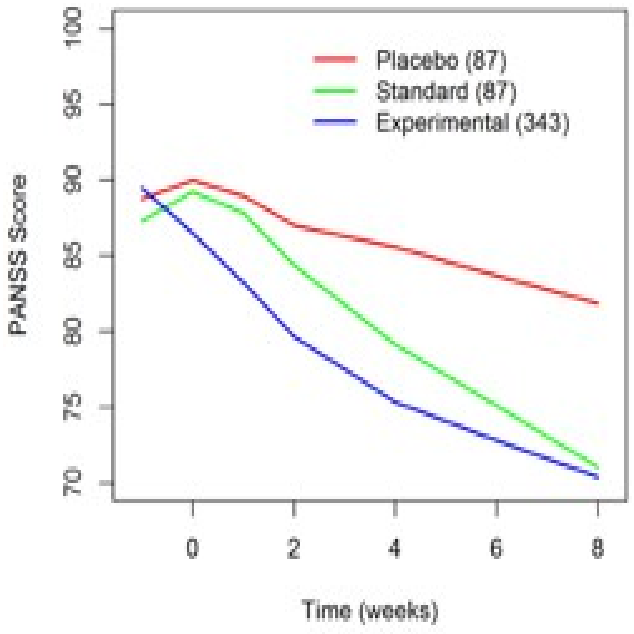
Many scientific investigations generate both : a) Longitudinal outcomes - Repeated measurements of response variables at a number of time points b) Event time or Survival outcomes - Times to recurrent or terminating events

Examples : a) Quality of life and survival time in cancer studies b) CD4 count and time to AIDS-defining condition in HIV research c) Cognitive decline and time to diagnosis of Alzheimer's disease in suspected cases

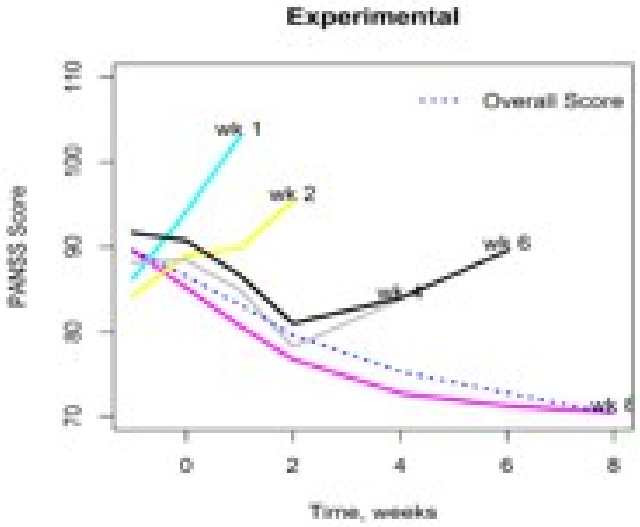
Lets consider a rather famous example : A longitudinal study into the treatment of schizophrenia A placebo-controlled randomized clinical trial of drug treatments for schizophrenia (Henderson et al. 2000) The primary response variable was the total score obtained on PANSS (positive and negative symptom scale), a measure of psychiatric disorder There were three treatment arms: placebo, standard, and experimental

Schizophrenia trial data Protocol: PANSS assessments at baseline followed by 1, 2, 4, 6, and 8 weeks Patients were withdrawn from the study due to 'inadequate response'(36% of patients). Some were failed to complete the trial protocol for other reasons, unrelated to the patient's mental state (17% of patients)

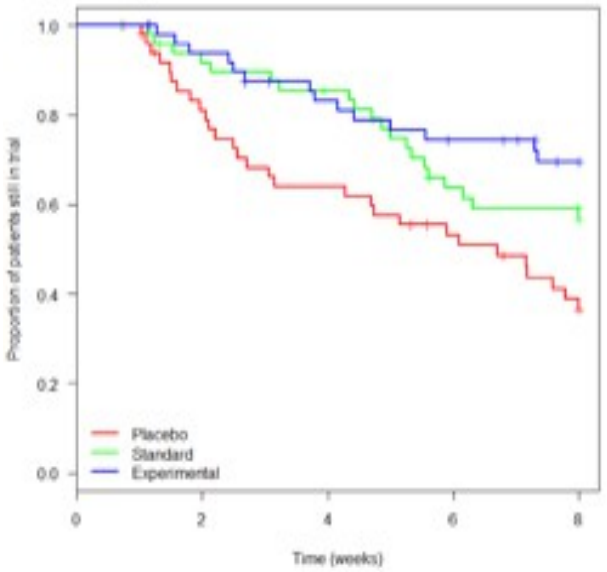
Schizophrenia trial outcomes Longitudinal outcome : PANSS assessments at 0, 1, 2, 4, 6, 8 weeks Event-time outcome : Time to withdrawn from the study (dropout)



It is not clear whether the apparent decrease in PANSS profiles reflects a genuine change over time or an artefact caused by selective dropout.
Patients with high PANSS values (worse prognosis) being less likely to complete the trial?



Joint modelling (of longitudinal and event time data) combines the longitudinal and event time data simultaneously.
As of 2015: > 200 methodological papers and > 60 clinical applications of joint models.



Joint modelling may : a) improve inference for a longitudinal outcome b) provide an adjustment of inferences about longitudinal measurements to allow for possibly outcome-dependent dropout c) improve inference for a time-to-event outcome
We consider the distribution of time to a terminating or recurrent event conditional on longitudinal measurements Evaluate the underlying relationship between longitudinal and event time processes

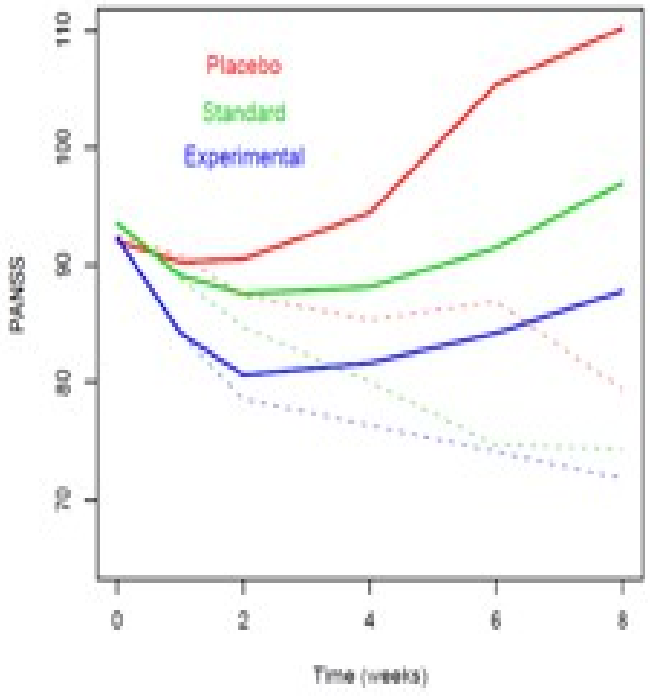
There are a number of implementations of this model in R and Stata [for example]

Perhaps the most well developed [currently ?] uses the R package joineR,

with the general recipe:
a) Create a joint data object using the joineR with the jointdata() function
b) Fit a joint model using the joineR with the joint() function
c) Calculate SEs using joineR with the jointSE() function

The joint model for y = PANSS and for dropout with $\log h(t)$ has [for example]
a) subject specific terms in the [sub] model for the longitudinal data : $u_0 + u_1 t$
b) then this term is included in the [sub] model for dropout : $\gamma_1 u_1 + \gamma_2 (u_0 + u_1 t)$
which has estimates : 0.349 u_1 + 0.042 ($u_0 + u_1 t$)

Many details are included in the paper by Henderson et al (2000). See Table 3
There is evidence against $\gamma_1 = 0$ and against $\gamma_2 = 0$ and so there is evidence of a crucial association between the two parts of the joint model. The results are not the same as the two separate [marginal ?] models.



[from Henderson et al (2000) section 5]
[The complete dataset does not appear to be available.]

We now return to the illustrative example introduced briefly in Section 1. Data are available from 523 patients, randomly allocated amongst the following six treatments: placebo, haloperidol 20 mg and risperidone at dose levels 2 mg, 6 mg, 10 mg and 16 mg. Haloperidol is regarded as a standard therapy. Risperidone is described as ‘a novel chemical compound with useful pharmacological characteristics, as has been demonstrated in in vitro and in vivo experiments’. The primary response variable was the total score obtained on the PANSS, a measure of psychiatric disorder. In an earlier analysis of these data, Diggle (1998) combined the four risperidone groups into one, and for comparability we do the

same. The resulting numbers of patients randomized to placebo, haloperidol and risperidone treatments were 88, 87 and 348.

The study design specified that the PANSS score should be taken at weeks −1, 0, 1, 2, 4, 6 and 8, where −1 refers to selection into the trial and 0 to baseline. The week between selection and baseline was used to establish a stable regime of medication for each patient, and in the analysis of the data we shall exclude the week −1 measurements. Of the 523 patients, 270 were identified as drop-outs and 183 of these gave as the reason for drop-out 'inadequate response'. In our analysis we shall treat drop-out due to inadequate response as a potentially informative event, and drop-out for other reasons as a censored follow-up time. Exact drop-out time was not recorded during the trial, the only information being on the first missed observation time. For this analysis we imputed each drop-out time from a uniform distribution over the appropriate interval between last observed and first missed measurement times. Results were not sensitive to imputation.

All three groups had a decreasing mean response, perhaps at a slower rate towards the end of the study. The overall reduction in mean response within each active treatment group is very roughly from between 90 and 95 to around 75. This appears close to the criterion for clinical improvement, which was stated, in advance of the trial, to be 'a reduction of 20% in the mean PANSS scores'. The decrease in the placebo group was smaller overall. However, at each time-point these observed means are, necessarily, calculated only from those subjects who have not yet dropped out of the study. Figure 1(b) shows estimated survival curves for time to drop-out due to inadequate response: large differences between groups are evident, with the highest drop-out rate in the placebo group and the lowest in the risperidone group.

In the analyses reported here we assume a "saturated" model for the mean PANSS response, with a distinct element of β for each treatment and measurement time combination. We have also analyzed the data assuming a quadratic time trend for each treatment, which gave very similar results and is not reported here. For the drop-out model we assume time constant treatment effects, with the standard treatment haloperidol group as baseline and thus one entry in α for each of the placebo and risperidone groups. We used the method of Section 4 for estimation for all models, with an initial analysis using M = 500 simulations for all expectations, then once the approximate estimates were obtained a more refined analysis with M = 1000 simulations at the E-step within the inner EM algorithm but M = 2000 simulations for likelihood evaluation. This led to a standard deviation of around 0.02 for Monte Carlo error in the log expected likelihoods.

Table 2. *Log maximized likelihoods for schizophrenia data*

	$W_1(t)$	$W_2(t)$	$\log L_Y$	$\log L_{N\|Y}$	$\log L$
Intercept only					
I	U_1	0	−10251.85	−1228.55	−11480.40
II	U_1	$\gamma W_1(t)$	−10252.58	−1181.13	−11433.72
Intercept +SGP					
III	$U_1 + V(t)$	0	−10126.66	−1228.55	−11355.21
IV	$U_1 + V(t)$	$\gamma W_1(t)$	−10132.55	−1146.14	−11278.69
V	$U_1 + V(t)$	$\gamma_1 U_1 + \gamma_2 V(t)$	−10139.79	−1107.61	−11247.40
Intercept +slope					
VI	$U_1 + U_t t$	0	−10127.31	−1228.55	−11355.86
VII	$U_1 + U_t t$	$\gamma W_1(t)$	−10133.76	−1137.41	−11271.17
VIII	$U_1 + U_t t$	$\gamma W_1(t) + U_3$	−10135.99	−1132.90	−11268.88
IX	$U_1 + U_t t$	$\gamma_1 U_1 + \gamma_2 U_2 + \gamma_3 W_1(t)$	−10147.75	−1096.05	−11243.80
X	$U_1 + U_t t$	$\gamma_2 U_2 + \gamma_3 W_1(t)$	−10148.42	−1095.60	−11244.03

Table 2 summarizes the results of fitting a variety of models for the latent processes W_1 and W_2. We begin (Model I) with a simple random intercept model for W_1, no random effects allowed in the drop-out model and hence no association. There is no improvement in fit (measured by the likelihood value) when frailty orthogonal to W_1 is allowed (not shown) with the estimated frailty variance lying on the boundary of zero. However, once latent association is allowed there is a substantial improvement in combined likelihood (Model II). The frailty result here is worth a comment: the initial analysis without latent association indicated no frailty in the drop-out component, which might be taken to suggest that there are no unmeasured covariates which could influence drop-out. Yet from Model II there is a clear association with the PANSS score. We suspect this result is caused by the known difficulty of estimating frailty effects with a semi-parametric baseline when there are few covariates.

A standard Laird–Ware random slope and intercept model apparently fits the marginal PANSS distribution almost as well as the Diggle intercept and Gaussian process model (Model VI), with once more strong evidence of association between the longitudinal and drop-out components (Model VII). Inclusion of a frailty term U_3 leads to improved likelihood (Model VIII). Of the models considered this is the only instance where we found frailty to have a non-negligible effect. When the model is extended to a full linear random effects model for $W_2(t)$ (Model IX) there is another substantial increase in likelihood, with a large drop in L_Y more than compensated by an increase in the cond likelihood component $L_{\mathrm{N|Y}}$. There is almost no loss in likelihood in removing from this model the separate effect on drop-out of the random intercept term U_1 (Model X). On the basis of this likelihood analysis, this is our "final model" for these data.

Thus, under Model X drop-out appears to be affected separately by two latent factors: the current value of $W_1(t)$ and the steepness of the trajectory. Since high PANSS indicates poor condition, both of

these conclusions are clinically reasonable: patients with either poor or rapidly declining mental health have increased risk of drop-out due to inadequate response. Parameter values for this model are given in Table 3 together with the corresponding estimates when no association between PANSS and drop-out is assumed, the two components being analysed separately. Note the reduction in the random slope variance σ_2^2 under separate analyses, and the attenuation of estimated treatment effects on the drop-out process, both consistent with the simulation results in Table 1. Standard errors in Table 3 were obtained by a Monte Carlo method, refitting Model X to 100 simulated data sets generated using parameter values taken from the original analysis. In order to complete the re-estimations within a reasonable time we used the smaller value of M = 500 in the Monte Carlo likelihood evaluation and stopped each inner EM algorithm as soon as an iteration caused a decrease in estimated likelihood. We accepted the value of $\xi = (\sigma_{12}, \sigma_2, \rho)$ which gave the maximum estimated likelihood after 30 iterations of the outer simplex. Our experience is that the estimated ξ is quickly very stable but Monte Carlo noise in the likelihood function evaluation prevents the simplex procedure from indicating convergence when M is small. Estimated values of β under Model X are shown in Figure 2(a), with ± two standard errors. These estimate the hypothetical drop-out-free population PANSS profiles, and are higher than the observed profiles (also shown) as a result of the tendency for patients with high scores to drop out due to inadequate response. Mean values for haloperidol-treated patients are lower (and thus better) than for the placebo group, as expected, but higher than the mean values for the resperidone group. A similar though less pronounced pattern is seen if the same model is assumed for the measurements but there is no latent association with drop-out (Model VI, Figure 2(b)). The general estimated drop-out-free pattern in Figures 2(a) and 2(b), of a late increase in PANSS after an inital fall, occurs to some extent for all the Laird–Ware intercept plus slope models we considered (Models VI–X) but for none of the other models. To illustrate, Figures 2(c) and 2(d) show estimated values of β under the apparently best fitting intercept-only and intercept plus SGP models (Models II and V).

We investigated model adequacy by comparing data simulated under the final Model X with that observed in the trial. Figure 3 illustrates some of our findings, based on 100 simulations of samples of 88, 87 and 348 subjects in the three treatment groups, as in the trial itself. The upper plot in Figure 3 compares the observed mean PANSS scores amongst patients still involved in the study with the corresponding simulation means, and the lower plot compares observed and simulated survival curves. The simulated values are close to those observed in all cases. A similar plot (not shown) based on the intercept- only model with latent association (Model II) also shows very good agreement, and there is reasonable agreement also (but larger standard errors) for the best intercept plus SGP model (V), except for some overestimation of drop-out rate for the haloperidol group (again not shown). Other plots based only on PANSS profiles for subjects who complete the trial also show good agreement between observed and simulated data for all models.

Table 3. *Parameter estimates and standard errors for final joint model with and without latent association*

	Var(U_1)	Var(U_2)	Corr(U_1, U_2)	Var(Z)	Placebo	Resp.	U_2	$W_1(t)$
	σ_1^2	σ_2^2	ρ	σ_z^2	β_{21}	β_{21}	γ_2	γ_3
Joint	283.37	12.59	0.06	100.24	0.779	−0.884	0.349	0.042
(X)	(18.17)	(1.31)	(0.02)	(3.95)	(0.344)	(0.322)	(0.051)	(0.007)
Separate	275.92	7.12	0.01	106.10	0.480	−0.508	0	0
(VI)	(21.42)	(0.81)	(0.08)	(4.18)	(0.218)	(0.196)	–	–

Notation and choices of symbols remains quite varied in this literature. To stay closer to choices made in these sessions, we might get:

$$\begin{aligned} y = &\beta_0 + \beta_1 \delta_1 + \beta_2 \delta_2 \\ &+ \beta_3 w_1 + \beta_4 w_2 + \beta_5 w_4 + \beta_6 w_6 + \beta_7 w_8 \\ &+ \beta_8 \delta_1 w_1 + \beta_9 \delta_1 w_2 + \beta_{10} \delta_1 w_4 + \beta_{11} \delta_1 w_6 + \beta_{12} \delta_1 w_8 \\ &+ \beta_{13} \delta_2 w_1 + \beta_{14} \delta_2 w_2 + \beta_{15} \delta_2 w_4 + \beta_{16} \delta_2 w_6 + \beta_{17} \delta_2 w_8 \\ &+ u_0 + u_1 t + \epsilon \end{aligned}$$

where δ_1 is the indicator for placebo and δ_2 is the indicator for risperidone.

$$\log h(t) = \log h_0(t) + \alpha_1 \delta_1 + \alpha_2 \delta_2 + \gamma_0 u_0 + \gamma_1 u_1 + \gamma_2 (u_0 + u_1 t)$$

where u_0 and u_1 are common to both sub-models.

This model pair would be Model IX in Henderson. Their Model X sets $\gamma_0 = 0$

Marchenko (2016) considers Henderson's Model II [no subject specific slope u_1] but with only a portion of the dataset [150 subjects] available to Henderson [523 subjects]

Some authors call u_1 the steepness and they call $u_0 + u_1 t$ the 'current value'.

Notice, with the dropout sub-model, if $\gamma_2 \neq 0$, then the model is not a proportional hazards model.