



Published in final edited form as:

Stat Med. 2009 February 15; 28(4): 659–678. doi:10.1002/sim.3484.

Subjective Prior Distributions for Modeling Longitudinal Continuous Outcomes with Non-Ignorable Dropout

Susan M. Paddock and Patricia Ebener
RAND Corporation

Abstract

Substance abuse treatment research is complicated by the pervasive problem of non-ignorable missing data – i.e., the occurrence of the missing data is related to the unobserved outcomes. Missing data frequently arise due to early client departure from treatment. Pattern-mixture models (PMMs) are often employed in such situations to jointly model the outcome and the missing data mechanism. PMMs require non-testable assumptions to identify model parameters. Several approaches to parameter identification have therefore been explored for longitudinal modeling of continuous outcomes, and informative priors have been developed in other contexts. In this paper, we describe an expert interview conducted with five substance abuse treatment clinical experts who have familiarity with the Therapeutic Community modality of substance abuse treatment and with treatment process scores collected using the Dimensions of Change Instrument. The goal of the interviews was to obtain expert opinion about the rate of change in continuous client-level treatment process scores for clients who leave before completing two assessments and whose rate of change (slope) in treatment process scores is unidentified by the data. We find that the experts' opinions differed dramatically from widely-utilized assumptions used to identify parameters in the PMM. Further, subjective prior assessment allows one to properly address the uncertainty inherent in the subjective decisions required to identify parameters in the PMM and to measure their effect on conclusions drawn from the analysis.

Keywords

Bayesian methods; expert opinion; longitudinal data; pattern-mixture model; prior elicitation; substance abuse treatment

1. Introduction

Missing data are pervasive in longitudinal substance abuse treatment studies, which often have treatment dropout as high as 50–70% [1,2,3,4]. The length of stay in treatment is a crucial consideration in effectiveness; longer lengths of stay (usually at least three months) have consistently been associated with better post-treatment outcomes [5,6]. This suggests that clients who leave treatment before completing a full course of treatment could have non-ignorable missing data [7] – namely, the probability an observation is missing is related to its unobserved value, which is in contrast to scenarios in which the probability of missingness is predicted by observed outcomes or covariates. This type of missing data is problematic when the goal is to measure the treatment process in substance abuse treatment. Improving the quality of care requires first measuring the treatment process and then testing whether treatment process predicts post-treatment outcomes [8]. Thus, it is important to calculate the rate of

change in a repeatedly measured treatment process score for all clients, rather than just for those clients with complete data.

Non-ignorable missing longitudinal data must be accommodated by jointly modeling the outcome of interest, y , and the occurrence of missing data, r , using a selection model (e.g., $f(y) = \int f(r|y)f(y)dr$) or a pattern-mixture model (PMM) (e.g., $f(y) = \int f(y|r)f(r)dr$) [9]. Linear random coefficient PMMs are widely used for modeling longitudinal data in the presence of non-ignorable dropout [10–15] and provide a way to assess the rate of change (slope) in treatment process measures over time. However, a limitation to PMMs is that many model parameters may not be identified. A prime example of this is the lack of identification of a slope parameter for persons who drop out of a study before at least two responses are observed.

Two approaches that address PMM parameter non-identification include *identifying restrictions* and *model simplification* [16]. Examples of commonly used identifying restrictions include assuming data for drop-outs follow the same distribution as that for study completers (a complete case missing variable restriction [17]) or assuming that the data distribution for drop-outs is equal to that for persons who have slightly more observed data (a neighboring case missing variable (NCMV) restriction [18]). Model simplification provides an additional benefit by decreasing the total number of parameters in the model, which can be helpful if there are many candidate patterns (and associated parameters) relative to observations. A frequently employed model simplification in standard PMMs [e.g., 9,11,13] or latent-class PMMs [19, 20] is to assume a linear time trend within each pattern and to extrapolate the trend beyond the point of last observation. Other forms of model simplification can result by assuming fixed parameter values for non-identified parameters – for example, by assuming the slope is zero for persons who drop-out of the study; this is equivalent to carrying the last observed value forward (LOCF) [21], since this approach effectively imputes later (missing) observations to equal an earlier (observed) one. Yet another model simplification is to assume a smooth functional form for the relationship between the slope parameter and censoring time [22,23, 24].

Identifying restrictions and model simplification require the analyst to make subjective and non-testable assumptions. The effect of these assumptions is often evaluated using sensitivity analyses [10,18] or Bayesian prior specification to incorporate subject matter expert judgment directly into the model. The Bayesian approach to non-ignorable non-response has witnessed growing popularity in recent years due to its incorporation of uncertainties about a range of plausible scenarios into posterior inferences of a target quantity of interest in both non-longitudinal [25,26,27,28] and longitudinal [29,30] data analyses. Despite the widespread use of linear random-coefficient pattern-mixture models for longitudinal data analysis, no attention has yet been devoted to eliciting prior distributions from subject-matter experts about the identification of the rate of change (slope) parameter for persons who drop out of the study after completing just one assessment. This paper addresses this gap for a study of the quality of care in the therapeutic community modality of substance abuse treatment, for which we elicit subjective prior distributions for the slope of a repeatedly measured continuous outcome when non-ignorable non-response is a concern.

It is unclear whether parameter identification strategies align with expert opinion and therefore how credible they are in practice for addressing experts' concerns about the effect of attrition on results. If expert opinion and identification strategies were to agree in our study, then this would strengthen conclusions drawn from similar quality of care studies that use PMM parameter identification strategies. In the absence of such a comparison, it is not possible to assess how realistic are parameter identification strategies. A contribution of this paper is to make such a comparison in the context of the quality of care in substance abuse treatment.

One challenge we faced in eliciting expert opinion in our study is that the clinical experts from whom we desire to elicit prior information come from a population that is less familiar with probability statements than those from whom elicitation is almost uniformly reported in the Bayesian literature. Most of the experts in our study are substance abuse treatment counselors in the therapeutic community who typically have not had training in statistics or probability. This is in contrast to physicians, who have provided prior information in other studies by reporting probability intervals [31,32] or through using graphical methods to specify prior distributions of proportional hazards model parameters [33]. Like physicians, engineers typically have had quantitative training and have provided expert opinion about means, variances, covariances, and other key parameters of interest [34,35,36]. We therefore combined model simplification techniques with a prior elicitation strategy that yielded expert opinions that were sufficiently detailed about important, subjective pattern-mixture model-building issues but did not require the experts to have knowledge about probability theory itself [37].

2. Motivation: Quality of Care in the Therapeutic Community

The motivating study involved characterizing the treatment process in the therapeutic community modality of substance abuse treatment [38], with the ultimate goal to improve the quality of care in the therapeutic community. The therapeutic community is based on the view that substance abuse is a disorder that involves the whole person; *i.e.*, substance abuse is a symptom of larger problems such as the abuser having problems with socialization, cognitive and emotional skills, and overall psychological development. The goal of the therapeutic community is to change the client's lifestyle and for the client to rehabilitate through mutual- and self-help [39,40,41]. The treatment in the therapeutic community consists of all interactions among staff and clients in formal and informal settings. Examples of such interactions include confrontation and sharing in groups; clients serving as role models for other clients; and clients providing mutual support in daily interactions [5,41]. The therapeutic community modality has been shown to improve long-term outcomes, such as reducing post-treatment drug use and criminal behavior, increasing employment and other social functioning [5,42–46].

To understand the process by which positive outcomes are achieved, the Dimensions of Change in Therapeutic Community Treatment Instrument [47,48] was developed. This survey is repeatedly administered to therapeutic community clients during their residential stay and consists of 54 items assessed with a 5-point scale indicating respondents' extent of agreement with each item (1=not at all; 5=completely). These 54 items are summarized into eight scales representing domains that describe aspects of the therapeutic community treatment process: clarity and safety; community responsibility; group process; introspection and self-management; problem recognition; resident sharing, support, & enthusiasm; positive self-attitude and commitment to abstinence; and social network.

The Dimensions of Change Instrument was used to collect treatment process data from adults who are in the therapeutic community at five time points – 0, 1, 3, 6, and 9 months post-treatment entry – who were enrolled in 6 community residential drug treatment programs that utilize the therapeutic community approach. All programs are part of the Phoenix House Foundation, a therapeutic community in the United States that operated 90 programs at the time of the study in New York, California, Texas, Florida, and New England. Data were collected from 519 adult clients. The participants in the study are drawn from the population of persons undergoing substance abuse treatment in the Phoenix House network of treatment programs during the data collection period. The RAND Corporation and Phoenix House Institutional Review Boards approved the study. All adults provided written consent to participate. Table 1 shows the distribution of the time of last treatment process assessment among clients. The most frequent reason for missing assessments was client departure from the program (82%).

The Dimensions of Change Instrument is motivated by the Donabedian quality of care framework [8], which posits that treatment process predicts post-treatment outcomes. This paper focuses on issues relevant for modeling the treatment process in the therapeutic community with the longer-term goal of improving its quality of care. Given this motivation, the question of central importance for the analyses presented here is to measure the expected change in treatment process scores for all clients who enter treatment. It is imperative to quantify the overall change expected by all clients rather than just those clients who remain in the TC relatively long lengths of time, since focusing on the latter could result in a biased assessment of the relationship between process and outcomes. Thus, missing outcomes must be modeled in such a way that accounts for potential differences between clients who stay versus those who drop-out early with respect to treatment process.

3. Methods

Model

Because we are interested in measuring client-level change over time in a continuous outcome, we specified a linear growth curve modeling approach for analysis. Let y_{ijk} be a treatment process score for person i in program j at wave k , where larger scores indicate better treatment process. Let r_{ij} be the missing data pattern membership indicator for person i in program j ($r_{ij} = 1, \dots, R$), where $R=5$ missing data patterns are assumed, constructed from the last time of treatment process assessment (Table 1). We jointly model (y_{ijk}, r_{ij}) via the conditional specification, $f(y_{ijk} | r_{ij}) f(r_{ij})$, using a linear random coefficient pattern mixture model [9] as follows:

$$r_{ij} \sim \text{Multinomial}(\pi_1, \dots, \pi_R)$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + \sum_{m=1}^R I_{\{r_{ij}=m\}} \beta_1^{(m)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma^2), \tag{1}$$

where (b_{0ij}, b_{1ij}) are person-level random effects for the intercept and slope that are assumed to be multivariately normally distributed with mean vector $\mathbf{0}$ and variance-covariance matrix \mathbf{D} . Time, t , is measured in the square root of months since treatment entry, so the change over time in y_{ijk} is assumed to be linear (in square root time) within pattern. The treatment program random effects, φ_j , are assumed to follow a normal distribution with mean 0 with variance η . One fixed intercept effect was assumed across all patterns, β_0 . An analysis of variance on baseline scores confirmed that the variation in mean baseline scores by pattern was not statistically significant. Also, the deviance information criterion [49] favored models that had a common intercept across patterns rather than including pattern-specific intercepts in the model. The fixed slope term for pattern m is $\beta_1^{(m)}$ ($m = 1, \dots, R=5$ possible study drop-out times), which measures change in the treatment process score over 1 unit time; since t represents the square root of time, $\beta_1^{(m)}$ represents the expected amount of change in y occurring in each of the following windows: 0–1 months ($t=1$), 1–4 months ($t=2$), 4–9 months ($t=3$). $I_{\{r_{ij}=m\}}$ is an indicator function of whether person i of program j is in pattern m . Three demographic variables – age (x_1), race (x_2), gender (x_3) – were included as covariates with regression coefficients α_l ($l=1,2,3$). Inference will ultimately focus on the change in treatment process scores over time, marginalized over the missing data distribution, where the marginal slope, $\bar{\beta}_1$, is:

$$\bar{\beta}_1 = \sum_{m=1}^R \beta_1^{(m)} \pi_m.$$

The following prior distributions for model parameters were assumed:

- $\beta_0 \sim N(3,10)$
- $\beta_1^{(m)} \sim N(0, 10)$ ($m=1, \dots, 5$),
- $\alpha_l \sim N(0,10000)$ ($l=1,2,3$),
- $\sigma^{-2} \sim \text{Gamma}(1,1)$ so that $E(\sigma^{-2})=1$,
- $\eta^{-2} \sim \text{Gamma}(5,5)$ so that $E(\eta^{-2})=1$,
- $\mathbf{D}^{-1} \sim \text{Wishart}(4 * \mathbf{I},4)$ so that $E(\mathbf{D}^{-1})=\mathbf{I}$, and
- $(\pi_1, \dots, \pi_R) \sim \text{Dirichlet}(1, \dots, 1)$.

Note that $\beta_1^{(1)}$ is identified only by its prior distribution, since only a single treatment process assessment at treatment entry is available for clients in Pattern 1. We examine several prior distributions for $\beta_1^{(1)}$ below that are informed by sensitivity analysis ideas from the literature and expert opinion. The BRugs package [50] in R was used to run Markov Chain Monte Carlo as implemented in OpenBUGS [51] to simulate all marginal posterior distributions for model parameters.

Scope of the elicitation

A strategy for elicitation of model parameters in complex models is to elicit prior distributions for those parameters for which informative prior distributions are critical to achieving inferential goals [35]. For example, data will always be available to update the prior distribution of the intercept term, since everyone is observed at treatment entry. Thus, it is less critical to elicit an informative prior for this parameter than for a non-identified slope, since in this case we would expect to have sufficient data about the intercept that would swamp any reasonable prior. Despite this, we considered eliciting a joint prior distribution for the intercept and slope for Pattern 1. In this case it would have been possible to only partially update the joint prior, since data would only have been available about the intercept but not the slope. If the intercept and slope were correlated *a priori*, this could distort the experts’ opinions about the slope. We circumvented this difficulty by assuming prior independence between the intercept and the slope. The plausibility of this assumption is supported by finding near-zero posterior correlations between the intercepts and slopes for those with at least 2 treatment process scores observed.

In order to ensure that we focused the elicitation on prior distributions for which informative priors were necessary, we conducted an extensive sensitivity analysis to identify treatment process domains for which there was sensitivity to modeling assumptions under plausible missing data assumptions. Our sensitivity analyses reflect a mix of parameter identification and model simplification strategies – for example, we use a common simplification of all of our models (e.g., [9,11,13]) by assuming a linear trend can be estimated for all clients who have at least two treatment process scores observed, regardless of when those scores were observed. We fit the following models, noting that $\epsilon_{ijk} \sim N(0, \sigma^2)$ and $(b_{0ij}, b_{1ij}) \sim \text{MVN}(\mathbf{0}, \mathbf{D})$ in all models below:

Model 1 is a standard longitudinal random effects model (REM), under which missing data are assumed to be ignorable [52] and thus missing data pattern, r , need not be modeled. One slope term, β_1 , is common to all clients in the study:

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + \beta_1 t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \epsilon_{ijk}$$

Model 2 is a PMM with two missing data patterns, one for study non-completers – *i.e.*, those clients who do not complete the 9-month assessment and are assumed to share a common slope term, $\beta_1^{(0,1,3,6)}$ -- versus the study completers, whose slope is given by $\beta_1^{(9)}$ [13]. The two missing data patterns, r , occur with probabilities $\pi_{\{0,1,3,6\}}$ and $\pi_{\{9\}}$, respectively. The model is:

$$r_{ij} \sim \text{Multinomial}(\pi_{\{0,1,3,6\}}, \pi_{\{9\}})$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + I_{\{r_{ij} \in \{0,1,3,6\}\}} \beta_1^{(0,1,3,6)} t + I_{\{r_{ij} = 9\}} \beta_1^{(9)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}$$

Model 3: Since study completers remain in treatment longer (9 months) than the 3-month minimum length of stay that has been associated with better post-treatment outcomes [5,6], the 9-month ‘completer’ designation might not be appropriate for this population. Thus, Model 3 defines ‘completion’ by focusing on the 3-month mark, by defining ‘completion’ in this case to imply the client’s final assessment was at least 3-months post-treatment entry. Model 3 identifies the slope for those having their only assessment at 0 months by assuming their slope is equal to that of clients whose last assessment is at 1 month. This corresponds to a type of NCMV identification (which will be called NCMV-1):

$$r_{ij} \sim \text{Multinomial}(\pi_{\{0,1\}}, \pi_{\{3,6,9\}})$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + I_{\{r_{ij} \in \{0,1\}\}} \beta_1^{(0,1)} t + I_{\{r_{ij} \in \{3,6,9\}\}} \beta_1^{(3,6,9)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}$$

Model 4: It is not entirely clear from the literature whether it is more important for a client to stay exactly 3 months versus *more than* 3 months in treatment in terms of post-treatment outcomes. We therefore also examined the effect of including those whose last treatment process assessment was at 3 months in a shorter-stayer pattern, which is in contrast to Model 3 that places these clients in the longer-stayer pattern. Thus, in Model 4 there are two patterns defined by those whose last assessments were at 0, 1, or 3 months versus 6 or 9 months. Model 4 employs a different type of identification than does Model 3 (which we call NCMV-2), since it treats clients who stay 3 or fewer months as not achieving a reasonably long length of stay to benefit from treatment:

$$r_{ij} \sim \text{Multinomial}(\pi_{\{0,1,3\}}, \pi_{\{6,9\}})$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + I_{\{r_{ij} \in \{0,1,3\}\}} \beta_1^{(0,1,3)} t + I_{\{r_{ij} \in \{6,9\}\}} \beta_1^{(6,9)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}$$

Model 5 has four patterns that are defined by the last assessments occurring during 0–1, 3, 6, 9 months. Model 5 identifies the slope for clients who have their only assessment at 0 months by assuming their slope is equal to that of clients whose last assessment is at 1 month – *i.e.*, the NCMV-1 identification. This is just like Model 3, except more patterns are specified for last assessment times beyond 3 months:

$$r_{ij} \sim \text{Multinomial}(\pi_{\{0,1\}}, \pi_{\{3\}}, \pi_{\{6\}}, \pi_{\{9\}})$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + I_{\{r_{ij} \in \{0,1\}\}} \beta_1^{(0,1)} t + I_{\{r_{ij} = 3\}} \beta_1^{(3)} t + I_{\{r_{ij} = 6\}} \beta_1^{(6)} t + I_{\{r_{ij} = 9\}} \beta_1^{(9)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}$$

Model 6 is a PMM with patterns for those whose last assessments were during 0–3, 6, & 9 months. Model 6 employs the NCMV-2 identification strategy. This is just like Model 4, except more patterns are specified for last assessment times beyond 3 months:

$$r_{ij} \sim \text{Multinomial}(\pi_{\{0,1,3\}}, \pi_{\{6\}}, \pi_{\{9\}})$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + I_{\{r_{ij} \in \{0,1,3\}\}} \beta_1^{(0,1,3)} t + I_{\{r_{ij}=6\}} \beta_1^{(6)} t + I_{\{r_{ij}=9\}} \beta_1^{(9)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}$$

Model 7 is a PMM that is fit to data for only those clients with at least two assessments. Model 7 deletes the 0 month-only clients from the analysis:

$$r_{ij} \sim \text{Multinomial}(\pi_{\{1\}}, \pi_{\{3\}}, \pi_{\{6\}}, \pi_{\{9\}})$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + \sum_{m \in \{1,3,6,9\}} I_{\{r_{ij}=m\}} \beta_1^{(m)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}$$

Model 8 is a PMM with 5 patterns, each of which correspond to a possible time of last assessment (0, 1, 3, 6, or 9 months). The slope for the missing data pattern of those having their final assessments at 0 months is assumed to be zero ($\beta_1^{(0)}=0$) – i.e., no change occurs. Model 8 can be thought of as a ‘last observation carried forward’ (LOCF) scenario for Pattern 0, since it implies that no change occurs in the treatment process scores over time for those clients whose only assessment was at 0 months. This was suspected by us to reflect a ‘worst case’ scenario prior to our expert interviews:

$$r_{ij} \sim \text{Multinomial}(\pi_{\{0\}}, \pi_{\{1\}}, \pi_{\{3\}}, \pi_{\{6\}}, \pi_{\{9\}})$$

$$y_{ijk} = b_{0ij} + b_{1ij}t + \varphi_j + \beta_0 + \sum_{m \in \{1,3,6,9\}} I_{\{r_{ij}=m\}} \beta_1^{(m)} t + \sum_{l=1}^3 \alpha_l x_{l,ij} + \varepsilon_{ijk}$$

All but three treatment process domains demonstrated significant increases in all eight of the sensitivity analyses listed above. For the expert interviews, we therefore focused on the three domains for which there was some ambiguity across the sensitivity analyses with respect to whether the treatment process score increased over time: Clarity & Safety, Community Responsibility, and Resident Sharing, Support, & Enthusiasm.

Expert Recruitment and Interviews

The Director of Research at Phoenix House provided three recommendations for clinical experts to interview. These clinical experts oversee the training of substance abuse treatment counselors at Phoenix House in New York and southern California, were familiar with the Dimensions of Change Instrument, and were believed by the Director of Research to be comfortable with the planned exercise of asking for estimates and bounds on those estimates. All three agreed to participate and also provided names of two senior counselors to approach for additional clinical expert interviews who also agreed to participate. This number of experts should allow us to obtain a range of experience and expertise among our clinical experts [53].

The interview was conducted by the first author one-on-one with each expert. The interviews took about one to 1–1/2 hours to complete. Each expert was given three sample questions to acquaint them with providing estimates along with the associated lower and upper bounds [54]. For this, the experts were asked about everyday items, namely the cost of a gallon of milk and the price of a new car. Similarly, the experts completed an exercise that familiarized them with reporting change over time; for this exercise, the experts were told to imagine they started with \$100 on a weekend trip to Las Vegas and then they were asked how much money they thought they would have at the end of the weekend. Following these training exercises, the experts’ opinions about treatment process were discussed; an illustrative part of the interview script showing this is provided in Appendix A. Each expert was shown a piece of paper with two columns, an example of which is shown in Appendix B. The left column showed the range

of treatment process scores for each domain (either Clarity & Safety, Community Responsibility, or Resident Sharing, Support & Enthusiasm) possible at treatment entry, and a circle marked the average treatment process score at treatment entry for each domain. The right column showed the possible treatment process scores at 1 month post-treatment entry. The experts were told that, unlike the left-hand column, a score was not circled in the right-hand column because clients who left before 1 month could not complete the assessment and therefore we do not have data to make this estimate. The English translation of the response scale was provided alongside the numeric values, so experts would know whether their response represented, for example, ‘complete’ or ‘a great deal’ of agreement with the items making up the domain. The experts were also shown the individual items that made up the Clarity & Safety, Community Responsibility, and Resident Sharing, Support & Enthusiasm scores. After discussing differences between those who leave treatment early versus those who stay longer, the expert was asked to provide an estimate and its bounds for their belief of how clients who left treatment before 1 month would have done if those clients had actually stayed 1 month instead of leaving.

Validation

In order to validate the experts’ abilities to provide accurate assessments about the average slopes and their uncertainty about their estimates, we also asked about change in Resident Sharing, Support & Enthusiasm scores for clients who were in missing data patterns 2 through 5 (i.e., those whose last assessments were at 1, 3, 6 or 9 months) and for whom we had data against which we could compare the expert’s assessments of the slope. To avoid substantially increasing the length of the interview, we completed the validation exercise for only the Resident Sharing, Support & Enthusiasm score. The validation exercise is applicable to the other scores only to the extent that the experts had comparable abilities to report on all scores. Experts were knowledgeable about all scores and even differentiated among the scores (e.g., the experts expressed thoughts such as expecting one score to change a lot, another to be relatively steady, and yet another would be low initially). Therefore, we would not expect the experts’ abilities to report on Resident Sharing, Support & Enthusiasm to be differentially better or worse than that for Clarity & Safety or Community Responsibility.

Model specification and synthesis of opinions across experts

For each expert, we encoded his/her prior belief about $\beta_1^{(1)}$ into a triangle distribution having mode equal to the expert’s ‘best guess’ about the change in treatment process score and the triangle distribution’s bounds equaled the upper and lower bounds reported by the expert. We implemented the triangle prior distribution in a discretized form since BUGS software does not offer a triangle prior distribution option. We incorporated these priors into the model depicted by Equation (1). In addition to examining the variation in expert opinions, we sought to synthesize the expert opinions to derive a composite estimate. For this, we will implement linear opinion pooling [55]:

$$f(\beta_1^{(1)}) = \sum_{p=1}^E w_p f_p(\beta_1^{(1)}),$$

where f_p is expert p ’s prior distribution for $\beta_1^{(1)}$, E is the number of experts from whom subjective opinions about $\beta_1^{(1)}$ were collected, and w_p is the weight assigned to expert p such that the weights sum to 1. Setting $w_p = 1/E$ (for $p = 1, \dots, E$) implies the decision maker has equal confidence in the abilities of all experts [55]. Alternatively, scoring rules could also be used to develop expert weights that reflect how well the experts’ probability distributions of

uncertain quantities reflect actual values that are observable yet unknown to the experts. We examined several scoring rules for continuous outcomes [56]. The scores for expert p for the RS slope of clients in pattern m (m = 2, 3, 4, 5), are:

- Equal weight scoring rule: $s_{p,m} = 1$
- Logarithmic score: $s_{p,m} = \ln[f_{p,m}(\beta_1^{(m)*})]$
- Quadratic score: $s_{p,m} = 2f_{p,m}(\beta_1^{(m)*}) - \int_{-\infty}^{\infty} \{f_{p,m}(\beta_1^{(m)})\}^2 d\beta_1^{(m)}$
- Spherical score: $s_{p,m} = f_{p,m}(\beta_1^{(m)*}) / \sqrt{\int_{-\infty}^{\infty} \{f_{p,m}(\beta_1^{(m)})\}^2 d\beta_1^{(m)}}$

where $\beta_1^{(m)*}$ is the posterior mean of the RS slope based on analyzing data on clients in pattern m and $\beta_1^{(m)}$ is the RS slope arising from the expert's prior. Then, for a given scoring rule,

$$s_p = \frac{1}{4} \sum_{m=2}^5 s_{p,m} \text{ and the weights become } w_p = s_p / \sum s_p, \text{ summing to 1 across experts.}$$

4. Results

Validation

Figure 1 summarizes the prior probability distributions for the experts' elicited prior distributions for the Resident Sharing, Support & Enthusiasm slope. The prior distribution on the slope was derived from the changes reported in Resident Sharing, Support & Enthusiasm scores on the 1–5 scale (See Appendix B). The width of each box reflects the bounds of the assumed triangle distribution, which was parameterized by the expert's 'best guess,' or mode, along with their reported bounds. The solid lines in Figure 1 reflect the subjective prior modes, the dashed lines represent the posterior mean slopes based on data analysis. Expert 5 was not asked the validation question about the expected Resident Sharing, Support & Enthusiasm change for those who stay at least 9 months because this expert reported not having seen clients who stayed in the therapeutic community this long; the reason for this is that the treatment site at which this expert worked lowered its maximum length of stay to six months following the collection of the treatment process data and prior to this expert's involvement at the treatment site. None of Expert 2's prior probability distributions covered the posterior means, but the other four experts consistently reported prior distributions that contained the posterior mean slopes with one exception: three of the four experts who answered the question regarding 9-month length of stay were slightly overconfident about how well clients would do if they had stayed 9 months, which is indicated by the dashed horizontal lines falling just below the 9-month boxplots in Figure 1. Experts quite consistently reported greater uncertainty about clients whose last assessment was taken at 1 month, as reflected by the width of the priors. Interestingly, Expert 4 and Expert 5 worked in the same program, which suggests that Expert 4's prior distribution for the 9-month survey completers are wider than the 3- and 6- month distributions perhaps because of not having seen many recent clients with such a long length of stay following their program's change to a 6-month maximum length of stay.

Most prior probability distributions covered the estimated posterior mean Resident Sharing, Support & Enthusiasm slopes, suggesting that the experts were not overconfident in their uncertainty assessments. A possible exception to this is Expert 2, whose prior distributions did not cover these slopes. This could be due to overconfidence or lack of substantive knowledge about the Resident Sharing, Support & Enthusiasm score; though Figure 1 alone cannot address this distinction, it provides a summary of the overall ability of the experts to report Resident

Sharing, Support & Enthusiasm scores that encompasses both calibration and substantive expertise. Wide prior distributions might suggest underconfidence on the part of the experts; most often the prior mode was not very close to the estimated posterior mean, suggesting that most experts were able to appropriately express their uncertainty about the slope.

Elicited priors

Figure 2 shows the elicited prior distributions for the slope parameter among clients who only had treatment process data observed at treatment entry. The clinical experts were generally consistent across the three dimensions - Clarity & Safety (CS), Community Responsibility (CR) & Resident Sharing, Support & Enthusiasm (RS) – in terms of the degree of uncertainty expressed – for example, Expert 1 was always more certain than Experts 3 & 4 overall – with Expert 2 showing the most within-expert variation in degree of uncertainty in the three elicited priors. Expert 5 expressed prior opinions that were equivalent to the LOCF assumption, and uncertainty relative to the baseline values was therefore not provided.

Analysis of treatment process scores

Figure 3 shows the 95% posterior probability intervals for β_1 , the slope after marginalizing with respect to the distribution of drop-out time. The posterior summaries of β_1 are shown when assuming data are missing at random under the standard random effects model (Model 1). Posterior probability intervals are also provided when $\beta_1^{(1)}$ is identified by simplifying the model to examine study completers versus non-completers as well as employing NCMV-like identifying restrictions. For each of the three treatment process domains presented in Figure 3, the results are mixed with respect to whether each domain is positively changing over time: the Model 5 analysis shows the 95% posterior probability intervals fall partially below zero (the vertical dashed line on Figure 3), while the intervals under other parameter identification strategies fall above zero. The results for Model 3 are practically identical to those of Model 5 and are thus not shown in Figure 3. Similarly, the results from Models 4 and 6 are similar and so only results from the more general model (Model 6) are shown. Model 7 is omitted because it requires one to delete clients from the analysis solely on the basis of having missing data. The results for Model 8 are identical to those for Expert 5. Across these analyses, the posterior distributions are relatively narrow.

The 95% posterior probability intervals are also provided in Figure 3 when conducting separate analyses under each expert's prior distribution for $\beta_1^{(1)}$. Expert 5 did not provide prior modes and bounds when asked about clients who left before 1 month; instead, Expert 5 stated that treatment process scores for all three domains would not change for those who were in the therapeutic community less than 1 month because this would not be a long enough period of time to witness change. Expert 5's marginal posterior distribution is therefore identical to assuming a zero slope term, which is Model 8. The posterior distributions are generally considerably under elicited priors versus selecting a single identification strategy. The direction of the slope consistently is positive under the identification strategies, but not under the experts' priors.

We used the Resident Sharing, Support & Enthusiasm scores reported by the clinical experts as inputs to scoring rules as previously described for analyses of scores from all three domains for the missing 1 month scores. Differences among the 95% posterior probability intervals using the various weighting strategies proposed above were negligible, so only results under equal weight linear pooling are presented. Histograms of 30,000 draws from the marginal posterior distributions of β_1 when all five expert's priors were pooled (Figure 4). The pooled results had wider 95% posterior probability intervals than those for any single expert, reflecting heterogeneity in expert opinion about clients who failed to take the survey after treatment entry.

However, these marginal posterior distributions are non-symmetric and multimodal, with the experts' opinions resulting in modes of β_1 , being near zero as well as being less than zero. We computed goodness-of-fit measures of the growth curve models conditional on missing data class for each model in Figure 3–Figure 4 (Table 2). A reviewer asked us to compare model fit while accounting for model complexity. The Deviance Information Criterion (DIC) [49] incorporates model fit (posterior mean deviance) and model complexity (effective number of model parameters), with the latter term penalizing unduly complex models. A model with relatively low DIC is considered to be better than one with a higher DIC, and models that differ by 1–2 DIC units are considered to have equivalent fits to the observed data given model complexity. A limitation with using DIC in this context is that its penalty term can be affected by reparameterization; priors that convey essentially identical information, such as Gaussian and discretized triangle priors used in our analyses, can result in very different DICs. To make a fair comparison of penalized model fit, Table 2 contains DICs for all models under Gaussian priors for β_1 . Accounting for model complexity by using DIC leads one to identify numerous models as equivalent for each outcome examined – eight of the 10 models have DICs of 2645–2646 for the Clarity & Safety outcome, seven of 10 have DICs of 1755–1756 for Community Responsibility, and six of 10 have DICs of 2177–2178. Overall, the information in the observed data alone fails to select a 'best' model. Regardless of how much evidence exists for any of these models in the observed data, it is important to also bear in mind that the suitability of any of these models depends on non-testable assumptions about unobserved data as well. Thus, the importance of expert opinion is paramount to the analysis.

In summary, our validation exercise confirmed that the experts' opinions properly reported the uncertainty of their assessments for the Resident Sharing domain. We found substantial variation in our posterior distributions for β_1 across experts, reflecting heterogeneity in opinions about the effect of clients who left our study after a very short period of time. Two of the five experts had beliefs about these very short stay clients that allowed for the marginal posterior distribution of β_1 to sit above zero, thus implying positive change in treatment process scores over time for all clients who enter the therapeutic community. Three experts were so negative in their assessments of how very short-stay clients would do that it implied no positive change could be expected for the total group of those entering treatment, had everyone stayed in treatment for the full duration of the study.

5. Discussion

We have presented an approach for eliciting subjective prior distributions from experts about the expected rate of change of continuous treatment process measures. This has potential for wide application for analyses of repeatedly observed measures where non-ignorable drop-out is a concern. Of central importance to this analysis was the question, 'What is the expected change in treatment process scores for persons who enter treatment?' This question assists with understanding the overall change expected by clients, not just the change experienced by clients who remain in treatment for relatively long lengths of time. This differs from addressing the issue of validating the survey by understanding whether the treatment process score is sensitive to measuring changes in treatment process over time, a question best answered by examining data from persons having at least two assessments [48]. The use of subjective prior distributions for the non-identified slope parameter as shown here will also be important for future modeling of the entire therapeutic quality of care process, in which client-level treatment process could be captured by client-specific growth parameters which would be used as predictors of treatment outcomes [8], as well as for accommodating differential drop-out in longitudinal randomized trials.

We succeeded in structuring our elicitation interview to be appropriate for therapeutic community clinical experts. Some of the clinical experts reported seeing neither probability

intervals nor confidence intervals in practice, so the training exercises conducted prior to the interview were critical for communicating to the experts the type of information asked of them. We therefore asked experts for best guesses (modes) along with bounds and encoded this information into triangle prior distributions, which is a popular choice for encoding expert judgments since the most crucial aspect to obtaining expert opinions is to formulate the questions in ways the experts can understand [37]. We evaluated approaches that required some statements of probabilities (e.g., the variable interval method [55]) but found this to be too complicated for our purposes. We also ruled out graphical approaches [33] to eliciting the prior for the slope parameter because we believed these experts would be more familiar with a survey-type format (such as provided in Appendix B). Further, our elicitation strategy was informed by the constraint on the amount of time we felt we could reasonably maintain the expert's attention to the task, considering its difficulty and the competing demands on the experts' time. If more time had been available, further sensitivity analyses could have been pursued in which, for example, prior information could have been elicited about unobserved treatment process scores for clients who were observed more than once, as opposed to linearly extrapolating within pattern which is typical in PMM applications [e.g., 9,11,13].

The validation exercise suggested that four of the five experts were well-calibrated with respect to their elicited prior distributions for change in treatment process scores aligning with known values for the Resident Sharing, Support & Enthusiasm score, suggesting that overconfidence was not a major concern; however, Expert 2 consistently over- or under-estimated this score for four missing data patterns. Though some elicitation experts suggest combating overconfidence by using an elicited prior distribution that is actually wider than what is reported by the expert - for instance, by assuming the intervals reported here actually reflect 90%, rather than 100%, of the expert's probability distribution and transforming the elicited intervals accordingly [57] - this was not pursued here since most experts did not exhibit overconfidence. It would not have changed the main result of having great prior variability in the clinical experts' opinions, and there is a lack of guidance on exactly how much additional variation to lend to experts' subjective intervals in this scenario.

An important finding from this work is that the widely-used approaches such as model simplification, parameter identification strategies (e.g., NCMV-like identification) and missing data strategies that identify parameters (e.g., LOCF) were clearly inappropriate given the experts' judgments about very short stayers. Only one of the five experts had a prior belief that matched any of the commonly used missing data assumptions for analyzing data with non-ignorable dropout. Though we did not ask the experts explicitly whether any of these identification strategies were appropriate here, the elicited priors address that concern: only a few of the elicited intervals from Experts 1-4 were centered about 0, thus suggesting little direct support for LOCF as an overall rule to apply in this setting. The subjective priors elicited here demonstrate the inadequacy of relying on default identification strategies to accommodate non-ignorable censoring for the application at hand. None of these parameter identification strategies allowed for the experts' views that treatment process scores could decrease over time. These parameter identification strategies do not address the possibility of other legitimate expert opinions about early leavers.

In conclusion, we have demonstrated here that expert opinion and default parameter identification strategies can be misaligned. Those who plan to conduct sensitivity analyses using PMMs or selection models should make a serious effort to incorporate expert opinion into the model in order to address concerns about non-testable assumptions and to reflect the uncertainty in inferences attributable to these assumptions. The divergent expert opinions demonstrated here would make it difficult to draw conclusions from future modeling of post-treatment outcomes as functions of change in treatment process scores. These results emphasize the practical limitations of sensitivity analyses for exploring the effect of missing data on

inferences. Motivated by such concerns, some researchers are searching for alternatives to non-ignorable non-response modeling, such as measuring covariates that account for the missing data mechanism (e.g., repeated subject-level self-assessments on the intent to complete future assessments) in order to avoid the non-ignorability assumption [58]. This would provide an alternative if study subjects could provide reliable self-ratings about their future intentions.

Acknowledgments

This research was supported by the Agency for Healthcare Research and Quality (AHRQ) grant R03HS014805. Data collection of the treatment process scores was supported by the National Institute on Drug Abuse (NIDA) R01DA14969. The authors thank the RAND-Phoenix House research project team for helpful conversations, Lionel Galway and four peer reviewers for thoughtful comments on the paper. This work could not have been completed without the thoughtful participation and generous contribution of time from the Phoenix House clinical experts.

Appendix A

Example of the script used in the expert interview

So, now we are going to talk about the DCI items. The first item we are going to talk about is called ‘Clarity & Safety.’

Clients are asked to complete the items that that are on this sheet.

(Look at the sheet with the Clarity and Safety items and describe the items, the 1–5 scale, and that the responses are averaged to obtain the score.)

(Now, change to the sheet in which the experts will make their guesses.)

This sheet has two columns:

- The ‘Treatment entry’ column represents the average Clarity and Safety score at baseline
- The ‘Month 1’ column represents the average Clarity & Safety score at 1 month.

The circle in the ‘Treatment entry’ column shows the average Clarity & Safety scores observed for the group of clients who left the TC before 1 month has passed.

We have data at later time periods for people who stay in treatment and do more DCI surveys, but if they don’t stay, then we don’t have that data. Since these clients left before 1 month passed, we only have their Clarity & Safety score from their first day of treatment. You can see that no numbers are circled in the ‘Month 1’ column, because we do not have enough data to know what this should be.

Because some clients leave the TC early, we are sometimes forced to estimate how they would have done if they had stayed. For example, if they had stayed they may have done as well as the people who did stay, or they could have done worse, or perhaps not have changed at all.

You know much more about people who stay and people who don’t stay than I do, and while it may be a guess on your part, your guess will be better than mine. I’d like you to think about people you know who have dropped out, or been discharged or left for other reasons before being in treatment for 1 month.

- Can you give me a few examples of early leavers you’ve known? I do not want names – just think about these clients in general.

Now think about whether you believe these early leavers would have done differently in treatment if they stayed 1 month compared with the people you know who actually have stayed in treatment for 1 month.

- Would the treatment process for them be the same as for the people who did stay? Why? Why not?

Now the task I'd like you to do is to give me your BEST GUESS about how people who left before 1 month would have scored on the DCI at 1 month if they had instead stayed in treatment for 1 month. (Draw a circle in the right-hand column where your best guess falls.)

Among all of the possibilities you considered for your best guess,

- What is the LOWEST score they could have?
- What is the HIGHEST score they could have?

Appendix B

Example answer sheet for the expert interview

CLARITY & SAFETY score for those who leave the program before 1 month

<u>TREATMENT ENTRY</u>		<u>1 month</u>
5.0	COMPLETE agreement with the items	5.0
4.9		4.9
4.8		4.8
4.7		4.7
4.6		4.6
4.5		4.5
4.4		4.4
4.3		4.3
4.2		4.2
4.1		4.1
4.0	Agrees A GREAT DEAL with the items	4.0
3.9		3.9
3.8		3.8
3.7		3.7
3.6		3.6
3.5		3.5
3.4		3.4
3.3		3.3
3.2		3.2
3.1		3.1
3.0	Agrees SOMEWHAT with the items	3.0
2.9		2.9
2.8		2.8

TREATMENT ENTRY		1 month
2.7		2.7
2.6		2.6
2.5		2.5
2.4		2.4
2.3		2.3
2.2		2.2
2.1		2.1
2.0	Agrees A LITTLE with the items	2.0
1.9		1.9
1.8		1.8
1.7		1.7
1.6		1.6
1.5		1.5
1.4		1.4
1.3		1.3
1.2		1.2
1.1		1.1
1.0	Agrees NOT AT ALL with the items	1.0

References

1. Baekeland F, Lundwall L. Dropping out of treatment: a critical review. *Psychological Bulletin* 1975;82:738–783. [PubMed: 1103201]
2. Primm AB, Gomez MB, Tzolova-Iontchev I, Perry W, Vu HT, Crum RM. Severely mentally-ill patients with and without substance use disorders: characteristics associated with treatment attrition. *Community Mental Health Journal* 2000;36:235–246. [PubMed: 10933241]
3. Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* 2002;11:437–457.
4. Yang X, Shoptaw S. Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. *Drug and Alcohol Dependence* 2005;77:213–225. [PubMed: 15734221]
5. Hubbard, GM.; Mardsen, ME.; Rachal, JV.; Harwood, HJ.; Cavanaugh, ER.; Ginzburg, HM. *Drug Abuse Treatment: National Study of Effectiveness*. Chapel Hill, N.C.: University of North Carolina Press; 1989.
6. Simpson DD. Drug Treatment evaluation research in the United States. *Psychology of Addictive Behaviors* 1993;7:120–128.
7. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. New York: Wiley; 2002.
8. Donabedian A. The quality of care: how can it be assessed? *Journal of American Medical Association* 1988;260:1743–1748.
9. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of American Statistical Association* 1995;90:1112–1121.
10. Daniels MJ, Hogan JW. Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics* 2000;56:1241–1248. [PubMed: 11129486]
11. Pauler DK, McCoy S, Moinpour C. Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Statistics in Medicine* 2003;22:795–809. [PubMed: 12587106]

12. Fitzmaurice GM, Laird NM, Shneyer L. An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. *Statistics in Medicine* 2001;20:1009–1021. [PubMed: 11276032]
13. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods* 1997;2:64–78.
14. Hogan JW, Laird NM. Model-based approaches to analyzing incomplete longitudinal and failure-time data. *Statistics in Medicine* 1997;16:259–272. [PubMed: 9004396]
15. Demirtas H, Schafer JL. On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* 2003;22:2553–2575. [PubMed: 12898544]
16. Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-mixture models. *Biostatistics* 2001;3:245–265. [PubMed: 12933616]
17. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993;88:125–133.
18. Molenberghs G, Thijs H, Kenward MG, Verbeke G. Sensitivity analysis for continuous incomplete longitudinal outcomes. *Statistica Neerlandica* 2003;57:112–135.
19. Roy J. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* 2003;59:829–836. [PubMed: 14969461]
20. Roy J, Daniels M. A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*. 2008 to appear.
21. Fitzmaurice GM. Methods for handling dropouts in longitudinal clinical trials. *Statistica Neerlandica* 2003;57:75–99.
22. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* 1989;45:939–955. and 1990; correction, **46**:889. [PubMed: 2486189]
23. Hogan JW, Lin X, Herman B. Mixtures of varying coefficient models for longitudinal data with discrete or continuous non-ignorable dropout. *Biometrics* 2004;60:854–864. [PubMed: 15606405]
24. Paddock SM. Bayesian variable selection for longitudinal substance abuse treatment data subject to informative censoring. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 2007;56:293–311.
25. Rubin DB. Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* 1977;72:538–543.
26. Kadane JB. Subjective Bayesian analysis for surveys with missing data. *The Statistician* 1993;42:415–426.
27. Scharfstein DO, Daniels MJ, Robins JM. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* 2003;4:495–512. [PubMed: 14557107]
28. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. *Journal of the Society for Clinical Trials* 2007;4:125. [PubMed: 17456512]
29. Kaciroti NA, Raghunathan TE, Schork MA, Clark NM, Gong M. A Bayesian approach for clustered longitudinal ordinal outcome with nonignorable missing data: evaluation of an asthma education program. *Journal of the American Statistical Association* 2006;101:435–446.
30. Lee, JY.; Hogan, JW.; Hitsman, B. Sensitivity analysis and informative priors for longitudinal binary data with outcome-related dropout. Talk, Joint Statistical Meeting; Salt Lake City, Utah. 2007.
31. Freedman LS, Spiegelhalter DJ. The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *The Statistician* 1983;32:153–160.
32. Chaloner K, Rhame FS. Quantifying and documenting prior beliefs in clinical trials. *Statistics in Medicine* 2001;20:581–600. [PubMed: 11223902]
33. Chaloner K, Church T, Matts JP, Louis TA. Graphical elicitation of a prior distribution for an AIDS clinical trial. *The Statistician* 1993;42:341–353.
34. O'Hagan A. Eliciting expert beliefs in substantial practical applications. *The Statistician* 1998;47:21–25.

35. Craig PS, Goldstein M, Seheult AH, Smith JA. Constructing partial prior specifications for models of complex physical systems. *Journal of the Royal Statistical Society Series D (The Statistician)* 1998;47:37–53.
36. Oakley J. Eliciting Gaussian process priors for complex computer codes. *Journal of the Royal Statistical Society Series D (The Statistician)* 2002;51:81–97.
37. Kadane JB, Wolfson LJ. Experiences in elicitation. *The Statistician* 1998;47:3–19.
38. Wenzel SL. Quality of care in the therapeutic community. An NIH/NIDA-sponsored Study. 2001 5R01DA014969-03.
39. DeLeon, G.; Rosenthal, M. Therapeutic Community. In: DuPont, RL.; Goldstein, A.; O'Donnell, J., editors. *Handbook on drug abuse*. Washington, D.C.: U.S. GPO; 1979.
40. DeLeon, G.; Rosenthal, MS. Treatment in Residential Therapeutic Communities. In: Karasu, T., editor. *Treatments of Psychiatric Disorders*. Vol. 2. Washington, D.C.: American Psychiatric Press; 1989. p. 1379-1396.
41. DeLeon, G. *The Therapeutic Community: Theory, Model and Method*. New York, New York: Springer; 2000.
42. Hubbard RL, Craddock SG, Flynn PM, Anderson J, Etheridge RM. Overview of 1-year follow-up outcomes in the drug abuse treatment outcomes study (DATOS). *Psychology of Addictive Behaviors* 1997;11:261–278.
43. Simpson DD, Joe GW, Rowan-Szal GA, Greener JM. Drug abuse treatment process components that improve retention. *Journal of Substance Abuse Treatment* 1997;4:565–572. [PubMed: 9437628]
44. Gerstein, DR.; Harwood, HJ. *Treating Drug Problems*. Washington, D.C.: National Academy Press; 1990.
45. U.S. Department of Health and Human Services (SAMHSA, CSAT). Preliminary Report: The Persistent Effects of Substance Abuse Treatment One Year Later. Rockville: 1996. National Treatment Improvement Evaluation Study (NTIES).
46. Messina N, Wish E, Nemes S. Predictors of treatment outcomes in men and women admitted to a therapeutic community. *American Journal of Drug and Alcohol Abuse* 2000;26:207–227. [PubMed: 10852357]
47. Orlando M, Wenzel SL, Ebener P, Edwards MC, Mandell W, Becker K. The dimensions of change in therapeutic community treatment instrument. *Psychological Assessment* 2006;18:118–122. [PubMed: 16594821]
48. Paddock SM, Edelen MO, Wenzel SL, Ebener P, Mandell W, Dahl J. Measuring changes in client-level treatment process in the therapeutic community with the dimensions of change instrument. *American Journal of Drug and Alcohol Abuse* 2007;33:537–546. [PubMed: 17668339]
49. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 2002;64:583–639.
50. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. *R News* 2006;6:12–17.
51. Thomas, A. Open BUGS. 2004. Available: <http://mathstat.helsinki.fi/openbugs>
52. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–974. [PubMed: 7168798]
53. Meyer, MA.; Booker, JM. *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Philadelphia, PA: SIAM; 2001. p. 169
54. Morgan, MG.; Henrion, M.; Small, M. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge, United Kingdom: Cambridge University Press; 1990.
55. O'Hagan, A.; Buck, CE.; Daneshkhan, A.; Eiser, JE.; Garthwaite, PH.; Jenkinson, DJ.; Oakley, JE.; Rakow, T. *Uncertain Judgements: Eliciting Expert Probabilities*. New York: Wiley; 2007.
56. Matheson JE, Winkler RL. Scoring rules for continuous probability distributions. *Management Science* 1976;22:1087–1096.
57. Garvey, PR. *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*. New York: Marcel Dekker; 2000.

58. Leon AC, Demirtas H, Hedeker D. Bias reduction with an adjustment for participants' intent to dropout of a randomized controlled clinical trial. *Clinical Trials* 2007;4:540–547. [PubMed: 17942469]

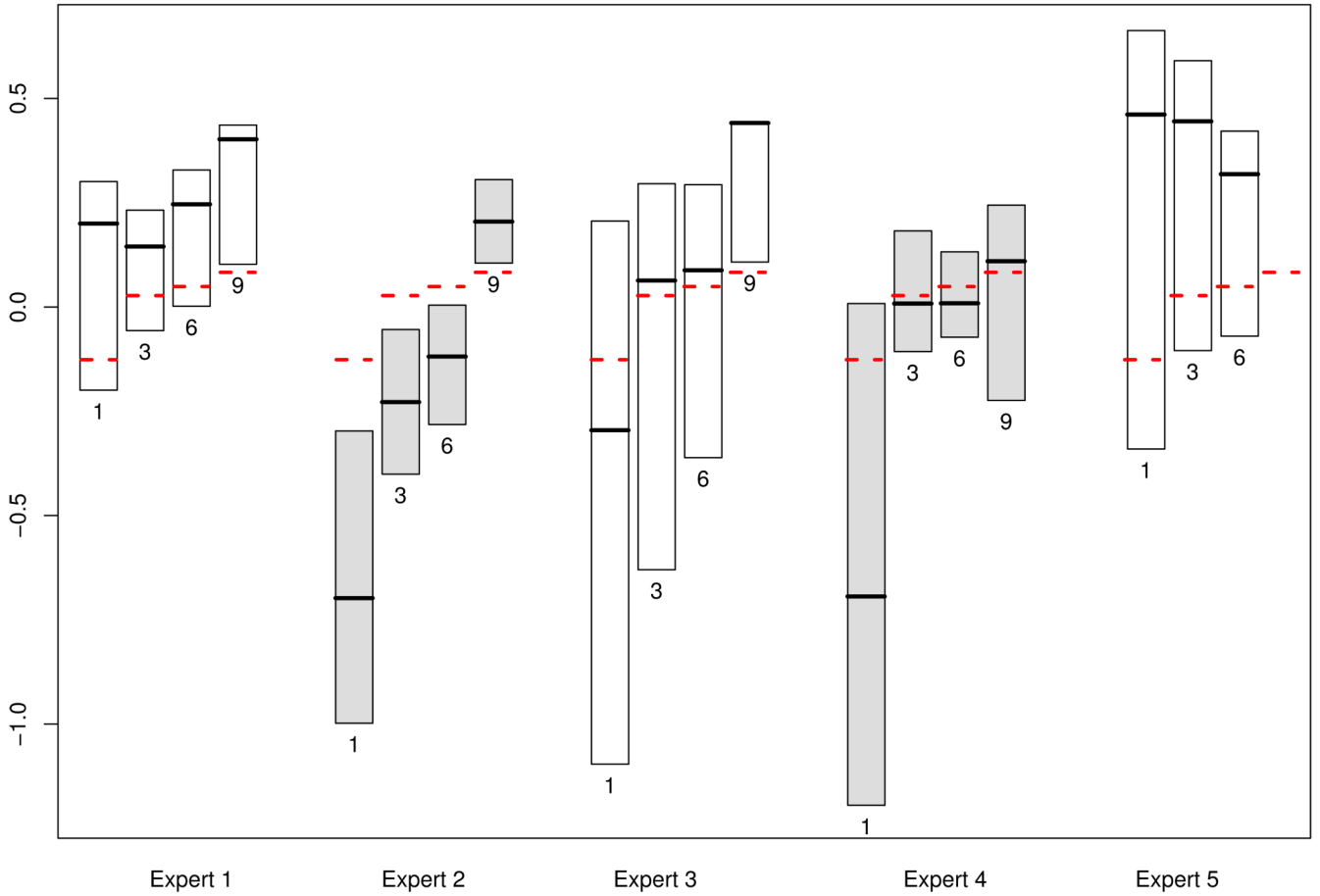


Figure 1. Prior distributions provided by Experts 1–5 for the slope of the Resident Sharing, Support & Enthusiasm score for those clients whose last assessments were at 1, 3, 6, and 9 months, as noted beneath each box. Prior mode is provided by the dark solid line, the box width reflects the difference between the upper and lower bounds, and the dashed lines are the posterior mean slopes from the data analysis. (Expert 5’s 9-month assessment (far-right hand side) is missing.)

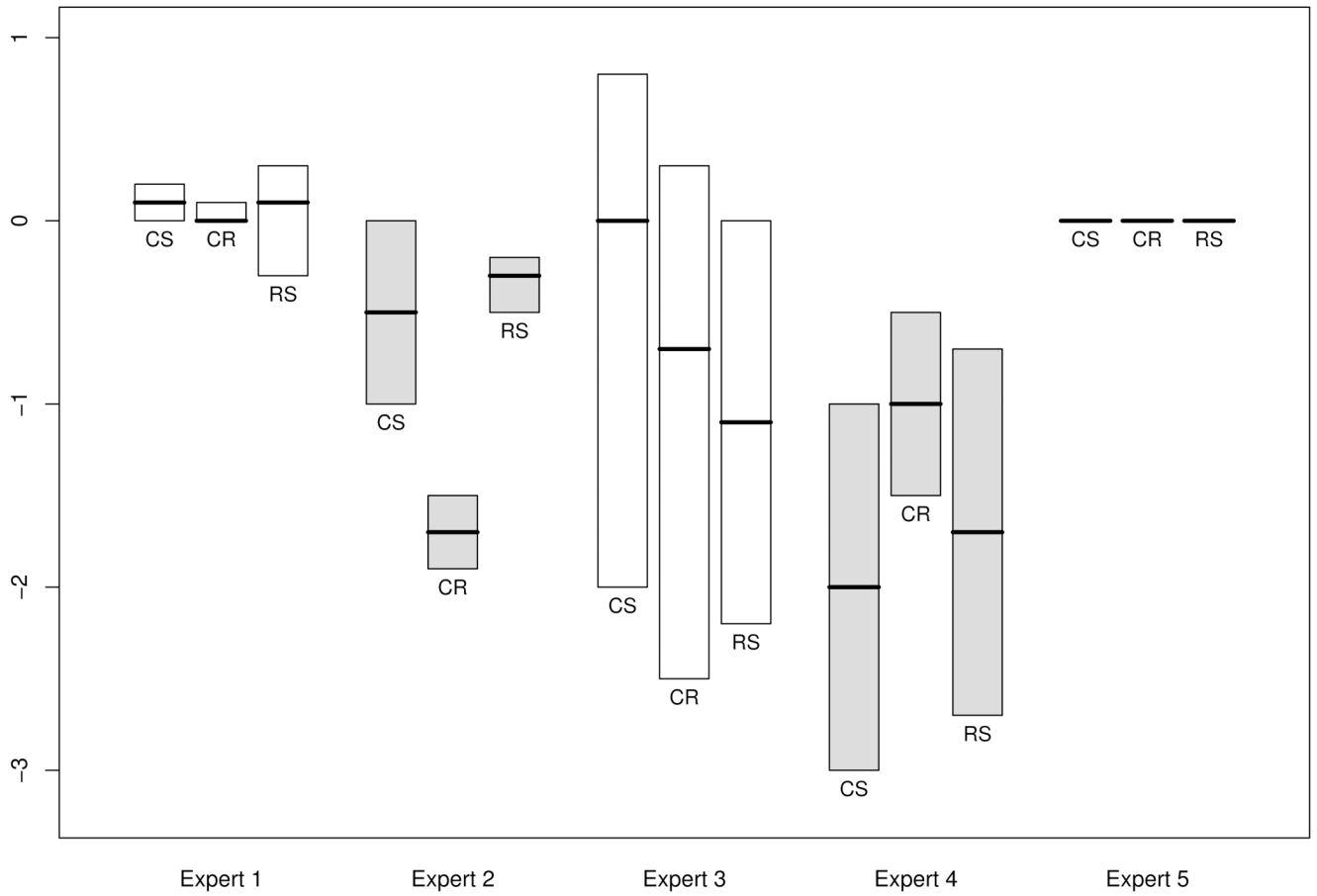


Figure 2. Prior distributions provided by clinical experts 1–5 for the slope of each Dimensions of Change domain for clients who only provided these data at treatment entry. CS=Clarity & Safety; CR=Community Responsibility; RS=Resident Sharing, Support & Enthusiasm. Note that Expert 5’s prior is equivalent to the last observation carried forward (LOCF) assumption for those with only one observation at time 0.

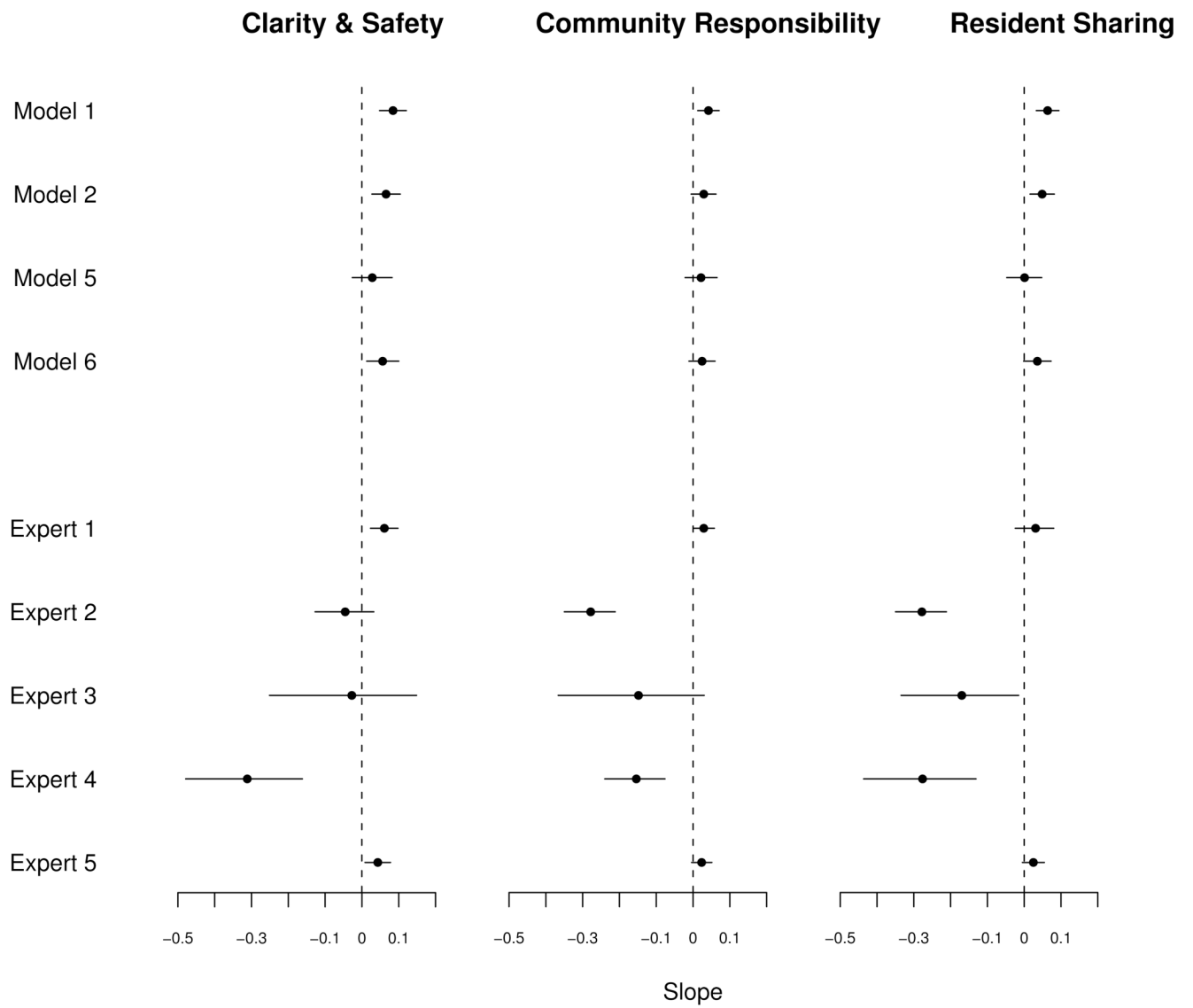


Figure 3. Marginal posterior distributions of the marginal slope (averaged over drop-out time) under sensitivity analyses and subjective prior distributions.

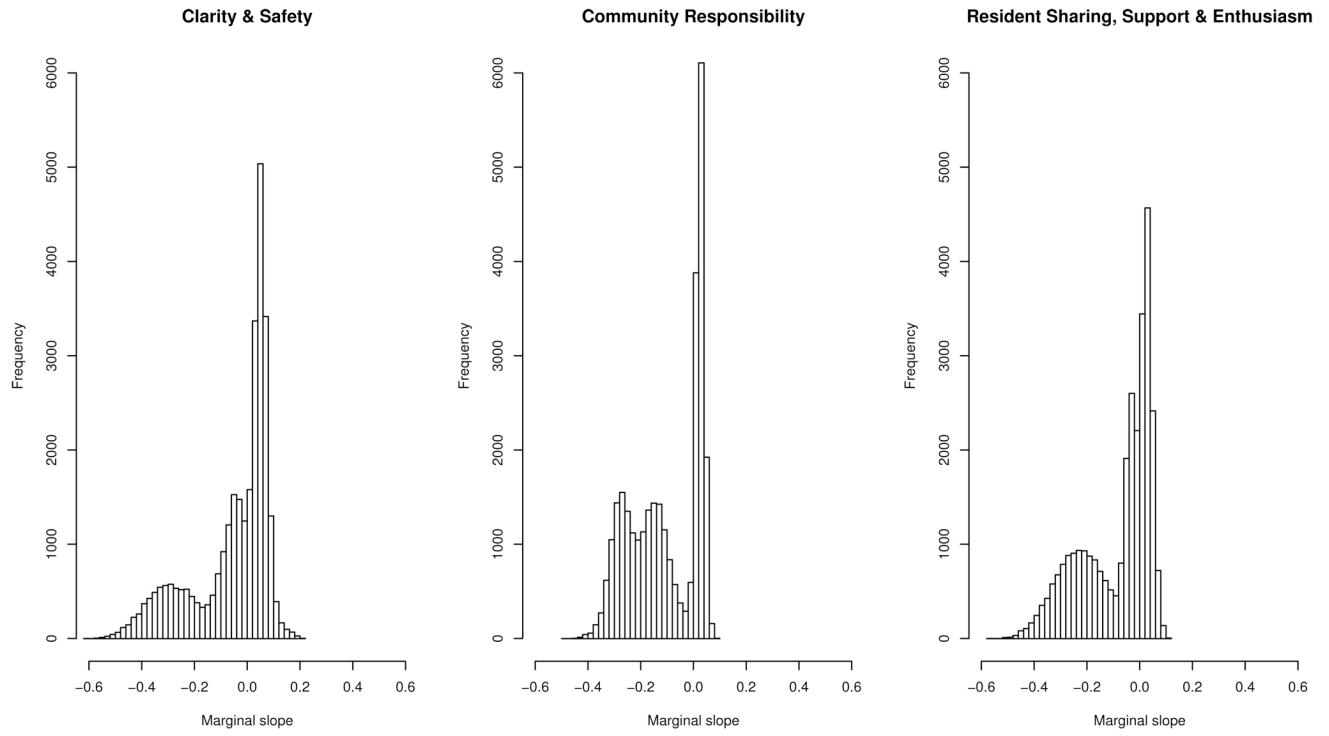


Figure 4. Marginal posterior distributions of β_1 under equal weight linear opinion pooling of the five experts' priors for $\beta_1^{(1)}$.

Table 1
 Sample Size for Each Missing Data Pattern of In-treatment Treatment Process Scores

Missing data pattern number	1	2	3	4	5	TOTAL
Time of last assessment (months post-treatment entry)	0	1	3	6	9	n/a
Sample size %	92 18%	86 17%	84 16%	85 16%	172 33%	519 100%

Table 2

Deviance Information Criteria (DICs) of growth curve models conditional on missing data patterns for analyses shown in Figures 3–4.

	Model 1	Model 2	Model 5	Model 6	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Pooled (Fig. 4)
Clarity & Safety	2650	2645	2698	2646	2645	2645	2645	2645	2646	2646
Community Responsibility	1755	1807	1843	1810	1756	1756	1756	1756	1756	1756
Resident Sharing	2188	2183	2239	2181	2177	2177	2177	2177	2177	2178