Personalized medicine enrichment design for DHA supplementation clinical trial

CrossMark

Yang Lei [a], Matthew S. Mayo [a], Susan E. Carlson [b], Byron J. Gajewski [a, *]

[a] Department of Biostatistics, The University of Kansas Medical Center, School of Medicine, Mail Stop 1026, 3901 Rainbow Blvd., Kansas City, KS, 66160, USA
[b] Department of Dietetics and Nutrition, The University of Kansas Medical Center, Kansas City, KS, 66160, USA

ABSTRACT

Personalized medicine aims to match patient subpopulation to the most beneficial treatment. The purpose of this study is to design a prospective clinical trial in which we hope to achieve the highest level of confirmation in identifying and making treatment recommendations for subgroups, when the risk levels in the control arm can be ordered. This study was motivated by our goal to identify subgroups in a DHA (docosahexaenoic acid) supplementation trial to reduce preterm birth (gestational age<37 weeks) rate. We performed a meta-analysis to obtain informative prior distributions and simulated operating characteristics to ensure that overall Type I error rate was close to 0.05 in designs with three different models: independent, hierarchical, and dynamic linear models. We performed simulations and sensitivity analysis to examine the subgroup power of models and compared results to a chi-square test. We performed simulations under two hypotheses: a large overall treatment effect and a small overall treatment effect. Within each hypothesis, we designed three different subgroup effects scenarios where resulting subgroup rates are linear, flat, or nonlinear. When the resulting subgroup rates are linear or flat, dynamic linear model appeared to be the most powerful method to identify the subgroups with a treatment effect. It also outperformed other methods when resulting subgroup rates are nonlinear and the overall treatment effect is big. When the resulting subgroup rates are nonlinear and the overall treatment effect is small, hierarchical model and chi-square test did better. Compared to independent and hierarchical models, dynamic linear model tends to be relatively robust and powerful when the control arm has ordinal risk subgroups.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Background

An important trend in treatment paradigm is personalized medicine, which is aimed to match patients to the most beneficial treatments. Patient populations are heterogeneous even in the same study. Characteristics vary between individuals, such as demographics, life style, genetic variants, etc. These varied characteristics can potentially impact the treatment effects on different individuals or subsets of patient populations.

It is important to distinguish the subgroups that benefit from novel treatment relative to control (e.g. standard of care) and subgroups that don't benefit or might even unnecessarily be exposed to a hazardous side effect [1]. Our research was motivated by our goal to design a clinical trial to identify subgroups in a trial to supplement pregnant women with docosahexaenoic acid (DHA) to reduce the rate of preterm births (PTB, gestational age<37 weeks). This is the first step in an enrichment design where a certain subgroup will be identified and the succeeding steps will distinguish the treatment effect within the selected (enriched) subpopulation [2].

Berry et al. (2013) discussed three clinical trial models assuming four subgroups of patients under an adaptive framework: Simon's Optimal Two-Stage model, a Bayesian independent model, and a Bayesian hierarchical model [3]. They showed that the hierarchical model could provide additional power and reduction in sample size compared to other two methods but acknowledged that hierarchical modeling could make finding a single effective subgroup more difficult, if there was only one [3].

We followed their four-group design in this study. The four-group design is an example for illustration and can be generalized to different settings. The hierarchical model does not require

* Corresponding author.
E-mail addresses: leiyang07@gmail.com (Y. Lei), mmayo@kumc.edu (M.S. Mayo), scarlson@kumc.edu (S.E. Carlson), bgajewski@kumc.edu (B.J. Gajewski).

the entities to be related [3].

In practice it is common to classify subjects into different risk levels of subpopulations. We classified the four subgroups according to their risk levels assuming they are receiving placebo or standard of care (control arm). Our study extended Berry et al. (2013) and aimed to identify a more efficient design from a prospective perspective to achieve the highest level of confirmation in identifying and making recommendations for subgroups [4], given the fact that the risk subgroups can be ordered in the control arm.

Two major considerations on subgroup analysis in clinical trial designs are: preserving Type I error and improving power [1]. Testing each hypothesis in a multi-group study inflates the overall Type I error rate. Multiplicity adjustment is required to preserve the overall Type I error rate [5]. We calibrated [6] the operating characteristics in simulations to ensure the overall Type I error rate was close to 0.05 (one-sided) in all designs that used different statistical models.

Approaches to improve statistical power in subgroup analysis include: using available information from previous studies [5] and borrowing information across subgroups [3]. We did a meta-analysis that contained data from nine DHA supplementation trials across the world to obtain informative priors. Then informative priors were applied to three different models: independent model (IM), hierarchical model (HM), and dynamic linear model (DLM). All methods were compared to the chi-square test to see the benefit of each model in the trial design.

2. Methods

2.1. Prior distributions

In subgroup analysis, a prior distribution is assumed for subgroup-specific treatment effect [5]. There are advantages and disadvantages of using both non-informative and informative priors. When historical data for the control arm are consistent with current study data, using informative priors constructed from previous complete trials can improve testing power and generate robust results [7].

For the meta-analysis from nine completed DHA supplementation trials, five were included in a review study conducted by the Cochrane Collaboration: Denmark 1992, England 1995, Europe 2000, Netherlands 1994, and USA 2003 [8]. Besides these we included four other trials: KUDOS (Kansas University DHA Outcomes Study) 2013 [11], DOMInO (DHA to Optimize Mother Infant Outcome, Australia) 2010 [9], Mexico 2015 [12], and NICHD (Eunice Kennedy Shriver National Institute of Child Health and Human Development, USA) 2010 [13]. The data from these studies are summarized in Table 1.

We used a hierarchical model with relatively weak priors to obtain future prior distributions. Let P_{ij} denote proportion of preterm birth in the ith study ($i = 1, …9$) and j^{th} arm ($j = 0,1$;

0 = control, 1 = treatment). We modeled $\theta_{ij} = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) \sim N(\mu, \tau)$, where τ is the precision (1/variance), with relatively weak hyper priors: $\mu \sim N(-2, 0.5)$ and $\tau \sim Gamma\ (1, 1)$. Future priors for μ and τ were derived from the averaged methods of moment estimators in the posterior distributions from the treatment arm and control arm.

In this way we obtained informative priors $\mu \sim N(-1.91, 1.28)$ and $\tau \sim Gamma(4.6361, 3.622)$ (details in Appendix). Based on these prior distributions, the median of the proportion is 12.8%. This estimation is reasonable and consistent, as the current preterm birth rate in the US is about 11.4% [14]. We applied these informative priors on models for analyzing data simulated under several different treatment effect assumptions.

2.2. Statistical models in trial designs

We used four risk subgroups for illustration but the conclusion applies more generally. We assumed equal sample size of 250 subjects in each subgroup in both control and treatment arm. Two hypotheses were considered. The first represents large overall treatment effect. The overall PTB rates are 8% vs. 4% in control and treatment arm respectively. The second represents small overall treatment effect. The overall PTB rates are 8% vs. 6% in control and treatment arm respectively. These percentages are consistent with the results from our previous DHA supplementation trial [9]. The control arm in both hypotheses has the same structure with ordinal PTB rates in the four subgroups (4%, 6%, 10%, and 12% respectively, see Ref. [10]).

Within each hypothesis, we designed three different scenarios where the resulting rates in the treatment arm are linear, flat, or nonlinear across the risk-ordered subgroups (Table 2). Within each scenario we compared the power among three models.

2.3. Independent model

We followed Berry et al. (2013) and examined an independent model [3]. In this model, we presented a Bayesian model with no borrowing from subgroups but we did borrow information from previous studies by applying the informative priors obtained through meta-analysis. We modeled the rate in each subgroup within each arm separately through:

$$\theta_{ij} = \log\left(\frac{P_{ij}}{1 - P_{ij}}\right), i = 0, 1; j = 1, 2, 3, 4$$

The prior distribution for θ_{ij} is:

$$\theta_{ij} \sim N(-1.91, 1.28)$$

This informative prior results in mean proportion close to 12.8% but it can generate a proportion ranges from about 0.8% to 60%.

Table 1
Number of preterm babies and sample sizes in completed trials.

Study	Treatment		Control	
	Preterm birth	Total	Preterm birth	Total
Denmark 1992	9	266	15	267
England 1995	22	113	19	119
Europe 2000	152	394	167	403
Netherlands 1994	8	32	10	31
USA 2003	14	142	17	149
KUDOS 2013	12	154	13	147
DOMINO 2010	88	1202	67	1197
Mexico 2015	32	365	30	365
NICHD 2010	82	434	83	418

Table 2
Preterm birth rates in subgroups in simulated effects and scenarios.

Scenario	Subgroup 1	Subgroup 2	Subgroup 3	Subgroup 4
Control arm				
8% (control arm)	4%	6%	10%	12%
Treatment arm: effect is large				
4% (linear)	2%	3%	5%	6%
4% (flat)	4%	4%	4%	4%
4% (nonlinear)	1%	6%	3%	6%
Treatment arm: effect is small				
6% (linear)	4%	5%	7%	8%
6% (flat)	6%	6%	6%	6%
6% (nonlinear)	1%	6%	6%	11%

2.4. Hierarchical model

This is another model that was examined in Berry et al. (2013) [3]. This model integrates the heterogeneous information from each subgroup. The hierarchical model assumes the four subgroups are exchangeable and allowed borrowing information across the four subgroups. In this we model θ_{ij} with a normal distribution with unknown mean and precision

$$\theta_{ij} \sim N(\mu_i, \tau_i) \; i = 0, 1; j = 1, 2, 3, 4$$

By introducing a hierarchy to model the unknown mean and precision, the model borrows information from previous studies and the current data across the four subgroups.

$$\mu_i \sim N(-1.91, 1.28)$$

$$\tau_i \sim Gamma\,(4.6361, 3.622)$$

Bigger τ indicates more pooling and information borrowing across the subgroups and smaller τ represents less pooling, or more heterogeneity across the subgroups. Our priors from meta-analysis show significant heterogeneity from the data in previously conducted clinical trials and we think it is reasonable to keep this heterogeneity in our simulation analysis to apply to a general population.

2.5. Dynamic linear model

Dynamic linear model is another model that has an intrinsic hierarchical structure. Unlike the hierarchical model that we discussed in the previous section, the dynamic linear model does not assume exchangeability of the four subgroups and borrows more information from adjacent subgroups.

The motivation is that this model might be more efficient since the four groups have ordinal risk levels at baseline so the correlation between adjacent subgroups might be larger. Therefore the dynamic linear model might capture the trend better. In this model, the first subgroup has a prior we obtained through meta-analysis, and the other subgroups have a hierarchical structure, with a common precision and a mean related to the neighborhood means [15].

For the first subgroup, we have the observation equation [16]:

$$\theta_{i1} \sim N(-1.91, 1.28); i = 0, 1$$

For the other subgroups, we have the state equation [16]:

$$\theta_{ij} \sim N(\theta_{i, j-1}, \tau_i); \; i = 0, 1; \; j = 2, 3, 4$$
$$\text{and}$$
$$\tau_i \sim Gamma\,(4.6361, 3.622)$$

The second subgroup directly borrows information from the first and the third. The third subgroup directly borrows information from the second and the fourth. The first subgroup directly borrows information from previous studies and the second subgroup. However, since the borrowing process is dynamic, the subgroups not adjacent directly impact each other through an indirect borrowing mechanism. This structure allows subgroups to borrow more information from adjacent subgroups which locally smooths the trend.

2.6. Computation

In each scenario, 1000 simulated trials were used; that is we simulated Bernoulli draws under each of the scenarios and sample sizes. We assumed the treatment and control arms each had four subgroups and each subgroup had 250 patients. The PTB rates in the four subgroups under the control arm were ordered and represented the risk levels assuming subjects were receiving placebo or standard of care. The PTB rates in the four subgroups under the treatment arm represented the risk levels after treatment, e.g., DHA supplementation. Data were simulated from binomial distribution based upon the proportions in each scenario using R 3.2.2. All Bayesian computations were performed using Open Bayesian inference Using Gibbs Sampling (OpenBUGS) from within R 3.2.2.

3. Results

3.1. Calibrating cutoff for equal type I error across models

Trial success was defined as the posterior probability that the PTB rate in the control arm is bigger than that in the treatment arm is greater than a cutoff value δ: $\Pr(P_c > P_t | data) > \delta$. In simulations the power function is defined as the average trial success rate across simulations. In the null hypothesis both the control arm and the treatment arm have overall PTB rates of 8% and the average success rate is the Type I error rate. Since we have four subgroups, the overall Type I error rate is $1 - \prod_{i=1}^{4}(1 - \alpha_i)$, where α_i is the Type I error rate for the ith subgroup. In the null hypothesis, both the control and treatment arms have the same PTB in their subgroups.

The four risk subgroups have PTB rates of 4%, 6%, 10%, and 12% in both arms. Using iterations of simulation under the case of no treatment effect, we tuned the δ value in each method separately to ensure each method's overall Type I error rate is close to 0.05, which resulted in approximate Type I error rates of 0.0127 in each subgroup (therefore overall Type I error is $1-(1-0.0127)^4 = 0.05$. These iterations resulted in slightly different δ values for the independent model, hierarchical model, and dynamic model of 0.985, 0.985, and 0.98 respectively.

3.2. Power for large overall effect

First we examined the overall PTB rate is 8% in the control arm and 4% in the treatment arm, indicating a large overall treatment effect. Within this setting, we tried different structure of PTB rates in the treatment arm to mimic different subgroup effects (Table 2). First we simulated a scenario where the resulting rate in the treatment arm is quite linear all four subgroups: 2%, 3%, 5%, and 6% correspondingly. Accordingly, each subgroup experienced a 50% reduction in PTB rate.

Second we simulated the resulting rate in the treatment arm as flat in all four subgroups: 4%, 4%, 4%, and 4%. Here there is no effect in the first subgroup and the last subgroup has the biggest treatment effect. Since the first subgroup has no effect at all, the average success rate we obtained for this subgroup is a Type I error rate for this subgroup analysis.

Third we simulated resulting rates in subgroups in the treatment arm are non-linear: 1%, 6%, 3%, and 6% correspondingly. We let the second subgroup have no treatment effect while the other three subgroups have treatment effects. Therefore the average success rate we obtained from the second subgroup is the Type I error rate for this subgroup.

We noticed that the three models had reasonable Type I error rates. The Type I error rates in the second subgroup independent, hierarchical, and dynamic linear models (DLM) are 0.006, 0.011, and 0.055 respectively.

In the comparison of the three models, we focused on power in subgroup analysis. Dynamic linear model has the highest power in subgroup analysis in all scenarios and subgroups except for the first subgroup in the non-linear scenario (Table 3). In the non-linear scenario, the proportions in the second subgroup are 0.06 for

both control and treatment arm. In this case, dynamic linear model has a slightly higher Type I error rate than other models but the error rate is still acceptable (Table 3).

In the following paragraphs we focus on each model's power to captures the most affected subgroup. The criterion here is power to detect the treatment effect in the subgroup with the largest absolute risk reduction.

The dynamic linear model has the highest power in capturing the most affected subgroup in all three situations. This is important because it represents the capability to identify or confirm the most beneficial subgroup. In the situation where the rates in the four subgroups of the experimental arm are linear, the power to capture the most affected subgroup in independent, hierarchical, and dynamic linear models is 0.548, 0.578, and 0.595 respectively. A regular chi-square test with Bonferroni adjustment has a power of 0.541 to capture the most affected subgroup. In the situation where the rates in the four subgroups in the experimental arm are flat, the power to capture the most affected subgroup in independent, hierarchical, and dynamic linear models is 0.879, 0.891, and 0.917 respectively. The chi-square test has a power of 0.857 to capture the most affected subgroup. In the situation where the rates in the four subgroups in the experimental arm are non-linear, the power to capture the most affected subgroup in independent, hierarchical, and dynamic linear models is 0.848, 0.861, and 0.889 respectively. The chi-square test has a power of 0.827 in capturing the most affected subgroup. In a nutshell, the dynamic linear model increases power to capture the most affected subgroup compared to the other three methods where the overall treatment effect is large (Table 3). In addition to providing the power to capture the most affected subgroup, the dynamic linear model appears to be powerful and robust in other subgroup analysis.

3.3. Power for small overall effect

We also examined when the overall PTB rate is 8% in the control arm and 6% in the treatment, indicating a small overall treatment effect. Again, we tried different structure of PTB rates in the treatment arm to mimic different subgroup effects (Table 2). In the first scenario the rates in the four subgroups in the treatment arm are linear or ordinal: 4%, 5%, 7%, and 8% respectively.

In the second scenario the rates in the four subgroups in the treatment arm are flat: all 6%. In this situation, there is no effect in the second subgroup and a negative effect in the first subgroup. In the last situation the rates in the four subgroups in the treatment arm are non-linear: 1%, 6%, 6%, and 11% respectively.

When the second subgroup has no treatment effect while the other three subgroups had treatment effects, but the effects varied among the three subgroups. Therefore the average success rate we obtained from the second subgroup is the Type I error rate for this subgroup. The Type I error rates in the second subgroup for independent, hierarchical, and dynamic linear models are 0.009, 0.013, and 0.031 respectively. The three models had reasonable Type I error rate.

The results of subgroup power analysis were summarized in Table 4. In the scenario where the treatment arm subgroup rates are linear, the power to capture the most affected subgroup in independent, hierarchical, and dynamic linear models is 0.254, 0.253, and 0.258 respectively. A regular chi-square test with Bonferroni adjustment has a power of 0.226 to capture the most affected subgroup.

In the scenario where the treatment arm subgroup rates are flat, the power to capture the most affected subgroup in independent, hierarchical, and dynamic linear models is 0.555, 0.551, and 0.628 respectively. The chi-square test has a power of 0.541 to capture the most affected subgroup. In the second subgroup analysis in the flat scenario, the treatment effect is negative. Independent, hierarchical, and dynamic linear models perform well but the chi-square test cannot capture this trend unless we discern the results from comparing the control and treatment arms (Table 4). However, in practice we can look at the direction of association to assist in this limitation of the chi-square test.

In the scenario where the treatment arm subgroup rates are non-linear, the power to capture the most affected subgroup in independent, hierarchical, and dynamic linear models is 0.32, 0.461, and 0.304 respectively. The chi-square test has a power of 0.462 to capture the most effected subgroup. Therefore the chi-square test and the hierarchical model perform well to capture the most affected subgroup when the resulting subgroup rates are nonlinear (Table 4). When the resulting subgroup rates are quite linear or flat, the dynamic linear model still outperforms the other three methods. In other subgroup analysis, the dynamic linear model appears to be robust and powerful.

4. Discussion

We did a sensitivity analysis using 500 subjects per subgroup. The comparisons between the statistical models remained similar with increased capability to confirm futility or success of subgroups (Tables 3 and 4). We assumed ordinal risk subgroups in the control arm have equal sample sizes. The general population may have

Table 3
Power in subgroup analysis when the overall treatment effect is large (8% vs. 4%).

Scenarios	Subgroup	Control	Treatment	Power (250 subjects/group)				Power (500 subjects/group)			
		True P	True P	DLM	IM	HM	Chi-sq	DLM	IM	HM	Chi-sq
Linear	1	4%	2%	0.169	0.097	0.188	0.176	0.401	0.3	0.426	0.349
	2	6%	3%	0.462	0.211	0.283	0.266	0.681	0.496	0.596	0.519
	3	10%	5%	0.599	0.458	0.475	0.452	0.878	0.811	0.823	0.777
	4*	12%	6%	0.595	0.548	0.578	0.541	0.889	0.866	0.887	0.860
Flat	1	4%	4%	0.017	0.007	0.019	0.013	0.023	0.012	0.029	0.013
	2	6%	4%	0.227	0.095	0.132	0.112	0.323	0.188	0.272	0.214
	3	10%	4%	0.823	0.692	0.708	0.652	0.98	0.958	0.962	0.932
	4*	12%	4%	0.917	0.879	0.891	0.857	0.997	0.997	0.997	0.993
Non-linear	1	4%	1%	0.316	0.31	0.515	0.463	0.769	0.753	0.867	0.788
	2	6%	6%	0.055	0.006	0.011	0.013	0.039	0.015	0.026	0.013
	3*	10%	3%	0.889	0.848	0.861	0.827	0.998	0.995	0.995	0.988
	4	12%	6%	0.702	0.592	0.624	0.541	0.916	0.883	0.903	0.860

*Most affected subgroup as defined by absolute risk.
True P: assumed preterm birthrate.
DLM: dynamic linear model.
IM: independent model.
HM: hierarchical model.
Chi-sq: chi-squared test.

Table 4
Power in subgroup analysis when the overall treatment effect is small (8% vs. 6%).

	Subgroup	Control	Treatment	Power (250 subjects/group)				Power (500 subjects/group)			
		True P	True P	DLM	IM	HM	Chi-sq	DLM	IM	HM	Chi-sq
Linear	1	4%	4%	0.007	0.005	0.012	0.0125	0.019	0.014	0.024	0.013
	2	6%	5%	0.06	0.021	0.034	0.0399	0.085	0.055	0.08	0.061
	3	10%	7%	0.19	0.161	0.147	0.1491	0.385	0.342	0.337	0.294
	4	12%+	8%	0.258	0.254	0.253	0.2259	0.528	0.481	0.512	0.447
Flat	1	4%	6%	0	0	0	0.112[a]	0	0	0	0.214[a]
	2	6%	6%	0.025	0.017	0.023	0.0125	0.013	0.012	0.017	0.013
	3	10%	6%	0.354	0.291	0.266	0.276	0.635	0.565	0.559	0.536
	4	12%+	6%	0.628	0.555	0.551	0.541	0.902	0.876	0.886	0.86
Non-linear	1	4%	1%	0.304	0.32	0.461	0.462	0.764	0.748	0.84	0.788
	2	6%	6%	0.031	0.009	0.013	0.0125	0.031	0.009	0.019	0.013
	3	10%+	6%	0.307	0.296	0.281	0.2761	0.593	0.584	0.585	0.536
	4	12%	11%	0.052	0.039	0.037	0.0293	0.06	0.052	0.056	0.04

+Most affected subgroup as defined by absolute risk reduction.
True P: assumed preterm birthrate.
DLM: dynamic linear model.
IM: independent model.
HM: hierarchical model.
Chi-sq: chi-squared test.
[a] The power calculated from the one-sided Chi-square test cannot distinguish the direction of treatment effect. However, in practice we can look at the direction of association to assist in this limitation of the chi-square test.

unequal-sized strata for similar ordinal risk levels. However, in clinical trial designs, it is still possible to selectively include participants to create balanced numbers in each subgroup. T.

Predetermined subgroup sample sizes decrease the risk of insufficient statistical power at the end of study [17]. At the current stage of subgroup analysis, it is not necessary to meet the power requirement in general statistical analysis. Once the most affected subgroups are identified through efficient designs, we can "enrich" the interested subgroups, i.e., recruit more subjects from the interested subgroup populations and stop recruiting certain subgroups that are futile. The final statistical analysis will be based upon all stages' recruitment.

We used relatively weak priors in our statistical models. It is desirable to use subgroup specific priors if previous data are available. The choice of relatively weak priors instead of vague priors was so that the designed clinical trial would use all the information available currently. The results would not likely be sensitive unless a case occurred where we had a big treatment effect that reduced the PTB to 0.00. Then we would not have a proper posterior for IM, HM, and DLM. The cut-points would have to change if we used vague priors. If the subgroup data are not consistent with the informative priors, the result could result in decreased power, particularly if δ values are not adjusted [6].

Dynamic linear models provide us flexibility, increased power, and natural interpretation to model a trend if there is a real trend, even the trend is non-linear. One important assumption in dynamic linear models is normality in both the observation and the state equations. Extension of dynamic linear model to model discrete outcomes usually brings additional computational difficulties [16]. However, a Bayesian approach can embrace the flexibility of dynamic linear models without suffering the drawbacks through the MCMC techniques. With the computational advantages, a Bayesian dynamic linear model can accommodate time-varying parameters, outliers, and irregular changes [18]. This can be considered for future studies.

5. Conclusions

In clinical trial designs with subgroup analysis, it is important to preserve a low Type I error rate and improve power to capture the most affected subgroup. Informative priors are one way to increase power. When informative priors from historical data are consistent with current study data, they represent a more powerful mechanism.

Another way to increase power is through model selection. Models that used independent and hierarchical models have been discussed in previous studies [3]. Other efficient models may exist if we have ordinal risk subgroups. We compared the dynamic linear model, independent model, and hierarchical model. All these models can be tuned to have desirable operating characteristics in terms of overall Type I error rate. The main comparisons were executed among the three methods. The dynamic linear model outperformed the other models in most scenarios with various structures of subgroup effects. We conclude that the dynamic linear model is relatively robust and efficient. This study shows that when the subgroups have certain structure, more efficient models may exist in clinical trials.

Appendix 1. Prior calculation details

The posterior mean for μ is -1.872 in the control arm and -1.944 in the treatment arm. The posterior mean for τ is 1.273 in the control arm and 1.287 in the treatment arm. The standard deviation for τ is 0.5874 in the control arm, and 0.6015 in the treatment arm. The methods of moment estimator for μ is calculated as $\frac{(-1.872-1.944)}{2} = -1.91$. The estimator for τ is calculated as $\frac{1.273+1.287}{2} = 1.28$. The α and β estimators in the distribution of $\tau \sim Gamma\ (\alpha,\ \beta)$ are derived from equations: $\frac{\alpha}{\beta} = 1.28$ and. $\frac{\alpha}{\beta^2} = \frac{0.5874^2+0.6015^2}{2} = 0.3534$.

Appendix 2. Simulation code

```
#### generate data: two groups, 4 risk levels, binary outcome

datasim<-function(n0,n1,p01,p02,p03,p04,p11,p12,p13,p14){

set.seed(12345)

y0<-rep(NA,n0)

y1<-rep(NA,n1)

p0<-rep(NA,4)

p1<-rep(NA,4)

p0<-rbind(p01,p02,p03,p04)

p1<-rbind(p11,p12,p13,p14)

risk0=rep(1:4,each=n0/4)

for (i in 1:n0){

y0[i]<-rbinom(1,1,p0[risk0[i]])

}

risk1=rep(1:4,each=n1/4)

for (i in 1:n1){

y1[i]<-rbinom(1,1,p1[risk1[i]])

}

risk=c(risk0,risk1)

y=c(y0,y1)

trt=rep(0:1,c(n0,n1))

label(trt)<-"trt"

label(risk)<-"risk"

label(y)<-"y"

return(list(trt,risk,y))

}


###Dynamic Linear Model

model

{

for (j in 1:J)

{
        S0[j]~dbin(p0[j],n0[j])

        p0[j]<-exp(theta0[j])/(1+exp(theta0[j]))

    S1[j]~dbin(p1[j],n1[j])

        p1[j]<-exp(theta1[j])/(1+exp(theta1[j]))

}

for (j in 1:J-1)

{

theta0[j+1]~dnorm(theta0[j],invtau2)

theta1[j+1]~dnorm(theta1[j],invtau2)

}

theta0[1]~dnorm(-1.91,1.28)

theta1[1]~dnorm(-1.91,1.28)

invtau2~dgamma(4.6361,3.622)

###Probability different

for (j in 1:J) {

psig[j]<-step(p0[j]-p1[j])

pneg[j]<-step(p1[j]-p0[j])

}

}

###Independent Model

model

{

for (j in 1:J)

{

        S0[j]~dbin(p0[j],n0[j])

        p0[j]<-exp(theta0[j])/(1+exp(theta0[j]))

        theta0[j]~dnorm(-1.91,1.28)

        S1[j]~dbin(p1[j],n1[j])

        p1[j]<-exp(theta1[j])/(1+exp(theta1[j]))

        theta1[j]~dnorm(-1.91,1.28)
```

```
}

###Probability different

for (j in 1:J) {

psig[j]<-step(p0[j]-p1[j])

pneg[j]<-step(p1[j]-p0[j])

}

}

###Hierarchical Model

model

{

for (j in 1:J)

{

    S0[j]~dbin(p0[j],n0[j])

    p0[j]<-exp(theta0[j])/(1+exp(theta0[j]))

    theta0[j]~dnorm(mu0,prec0)

    S1[j]~dbin(p1[j],n1[j])

    p1[j]<-exp(theta1[j])/(1+exp(theta1[j]))

    theta1[j]~dnorm(mu1,prec1)

}

mu0~dnorm(-1.91,1.28)

var0<-1/prec0

prec0~dgamma(4.6361,3.622)

mu1~dnorm(-1.91,1.28)

var1<-1/prec1

prec1~dgamma(4.6361,3.622)

###Probability different

for (j in 1:J) {

psig[j]<-step(p0[j]-p1[j])

pneg[j]<-step(p1[j]-p0[j])

}

}
```

References

[1] N. Simon, R. Simon, Adaptive enrichment designs for clinical trials, Biostatistics 14 (4) (2013) 613—625.

[2] V.V. Fedorov, T. Liu, Enrichment Design. Wiley Encyclopedia of Clinical Trials, 2007, http://dx.doi.org/10.1002/9780471462422.eoct342/abstract (Accessed 30 November 2015).

[3] S.M. Berry, K.R. Broglio, S. Groshen, D.A. Berry, Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials, Clin. Trials 10 (2013) 720—734.

[4] S.J. Ruberg, L. Shen, Personalized medicine: four perspectives of tailored medicine, Stat. Biopharm. Res. 7 (3) (2015) 214—229.

[5] M. Alsoh, K. Fritsch, M. Huque, K. Mahjoob, G. Pennello, M. Rothmann, E. Russek-Cohen, F. Smith, S. Wilson, L. Yue, Statistical considerations on subgroup analysis in clinical trials, Stat. Biopharm. Res. 7 (2015) 286—303.

[6] A.P. Grieve, Idle thoughts of a 'well-calibrated' Bayesian in clinical drug development, Pharm. Stat. 15 (2) (2016) 96—108.

[7] Z. Chen, A. Liu, Y. Qu, L. Tang, N. Ting, Y. Tsong, Applied Statistics in Biomedicine and Clinical Trial Designs, Springer, New York, 2015, p. 14.

[8] M. Makrides, L. Duley, S.F. Olsen, Marine oil, and other prostaglandin precursor, supplementation for pregnancy uncomplicated by pre-eclampsia or intrauterine growth restriction, Cochrane Database Syst. Rev. 19 (3) (2006) CD003402.

[9] S. Carlson, J. Colombo, B.J. Gajewski, K. Gustafson, D. Mundy, J. Yeast, M. Georgieff, L. Markley, E. Kerling, D. Shaddy, Docosahexaenoic acid supplementation and pregnancy outcomes, Am. J. Clin. Nutr. 97 (4) (2013) 808—815.

[10] C. Zhang, L. Garrard, J. Keighley, S. Carlson, B.J. Gajewski, Subgroup identification of early preterm birth (ePTB): informing a future prospective enrichment clinical trial design, BMC Pregnancy Childbirth 17 (2017) 18.

[11] M. Makrides, R.A. Gibson, A.J. McPhee, L. Yelland, J. Quinlivan, P. Ryan, DOMInO Investigative Team, Effect of DHA supplementation during pregnancy on maternal depression and neurodevelopment of young children: a randomized controlled trial, JAMA 304 (15) (2010) 1675—1683.

[12] U. Ramakrishnan, A. Stinger, A.M. DiGirolamo, R. Martorell, L.M. Neufeld, J.A. Rivera, L. Schnaas, A.D. Stein, M. Wang, Prenatal docosahexaenoic acid supplementation and offspring development at 18 months: randomized controlled trial, PLoS One 10 (8) (2015) e0120065.

[13] M. Harper, E. Thom, M.A. Klebanoff, J. Thorp, Y. Sorokin, M.W. Varner, R.J. Wapner, S.T. Caritis, J.D. Iams, M.W. Carpenter, A.M. Peaceman, B.M. Mercer, A. Sciscione, D.J. Rouse, S.M. Ramin, G.D. Anderson, Eunice Kennedy, Shriver national institute of child health and human development maternal—fetal medicine units network, Obstet. Gynecol. 115 (201) (2010) 234—242.

[14] J.N. House, U.S. Preterm Birth Rate Still Too High: Q&A with Jennifer L. Howse, PhD, the March of Dimes, in: Culture of Health, NewPublicHealth, Robert Wood Johnson Foundation, 2014.

[15] T.J. Leininger, S.C. Reese, G.W. Fellingham, S.D. Grimshaw, An Adaptive Bayesian Approach to Continuous Dose-response Modeling, 2010 (Accessed 09 December 2015), https://stat.duke.edu/~tjl13/papers/TJL_MSthesis.pdf.

[16] G. Petris, S. Petrone, P. Campagnoli, Dynamic Linear Models with R, Springer, New York, 2009.

[17] S. Padmanabhan, Handbook of Pharmacogenomics and Stratified Medicine, first ed., Academic Press, Waltham, Massachusetts, 2014, p. p.317.

[18] M. West, P.J. Harrison, H.S. Migon, Dynamic generalized linear models and bayesian forecasting, J. Am. Stat. Assoc. 80 (389) (1985) 73—83.