

Evaluation of information of prior relative to current data in analysis with prior

Satoshi Shoji, Yoshiro Tomono, Shinichi Tsuchiwata, Akiyuki Suzuki
Pharmacometrics, Clinical Pharmacology, Development Japan, Pfizer Japan Inc.

Introduction

When data in a population of interest is sparse due to practical constraints, it is occasionally difficult to estimate all the parameters of a model, specifically a complex model, with the sparse data alone. Use of priors for parameters which are difficult to estimate is one of the ways to analyze such limited data while borrowing information from the priors and stabilizing the estimation [1]. The analysis with priors can be performed by PRIOR functionality of NONMEM.

When analyzing a sparse data with priors, parameter estimates which incorporate corresponding priors depend on information of the priors as well as information of the data. Therefore, it is considered important to evaluate information of the priors relative to the data when interpreting the result. In addition, if use of priors for specific parameters is supposed in a future study, it is considered meaningful to evaluate optimal study designs which improve precision of parameter estimates in the analysis with priors.

In this work, 1) we present information of priors relative to sparse data based on variance of estimates from Fisher information with or without the priors. 2) We also show that the Fisher information with priors can be used for evaluating study designs for a future study as a similar approach to design optimization of experiments [2].

Methods

In this presentation, information of a prior relative to sparse data is defined as how many times the prior is as informative as the sparse data is. This idea is based on prior information of Bayes estimator [3]. To measure the relative strength of priors, we use variance of parameter estimates based on Fisher information matrix with or without priors. For example, as shown in Figure 1, when variance of an estimate ($\hat{\theta}$) obtained from maximum likelihood estimation of sparse data with the prior (V_{P+S}) is equal to variance of the estimate obtained from $N_p + 1$ times of the sparse data ($V_{N_p \times S + S}$), strength of the prior is considered N_p -time of the sparse data. Since the N_p depends on information of both the data and the prior, the N_p is considered as an index of information of the prior relative to the sparse data. We describe below how to obtain the index through Fisher information matrix with priors.

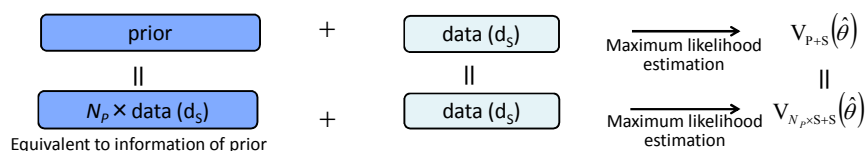


Figure 1. Information of a prior (N_p) is defined as the prior is N_p times as informative as sparse data (d_s).

The population Fisher information matrix (FIM) is derived from first-order approximation for random effect parameters (η s) of a population model [2]. The FIM for a sparse data (FIM_S) is a minus expectation of the second derivative of the log-likelihood (O_S). Hereafter, sparse data which consists of n observations with additive residual error for an individual i , $y_i \sim N(f(\text{dose}_i, \text{time}_i, \eta_i, \theta, \Omega), \sigma^2)$, is used as an example for explanation. The FIM_S is described in equation (1) and (2). The θ and $\omega^2_{(k)}$ represent a vector of fixed effect parameters and a variance for k^{th} random effect parameter, respectively. The Σ below is a variance-covariance matrix of the 1st order Taylor series approximation of the model for random effects η_i at zero [2]. The $\text{tr}(X)$ is a sum of diagonal elements of matrix.

$$O_S = \log(d_s)$$
$$O_S = -\frac{n}{2} \pi - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (y - f)^T \Sigma^{-1} (y - f)$$
$$FIM_S = -E \left[\frac{\partial^2 O_S}{\partial \theta \partial \theta^T} \right] = \frac{\partial f^T}{\partial \theta} \Sigma^{-1} \frac{\partial f}{\partial \theta} \dots (1)$$
$$FIM_{S(k,k')} = -E \left[\frac{\partial^2 O_S}{\partial (\omega_{(k)}^2) \partial (\omega_{(k')}^2)} \right] = \frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial (\omega_{(k)}^2)} \Sigma^{-1} \frac{\partial \Sigma}{\partial (\omega_{(k')}^2)} \right) \dots (2)$$

The Fisher information matrix for analysis with priors is introduced below. When a prior analysis is conducted with \$PRIOR NWPRI in NONMEM, prior distributions for θ and Ω are normal and inverse-Wishart distributions, respectively. Therefore, log-likelihood for the analysis with priors (O_{P+S}) can be described as follows (Gisleskog et al [1]).

$$O_{P+S} = O_P + O_S$$
$$O_P = \log(d_p), \quad O_S = \log(d_s)$$
$$O_{P+S} = O_S + C_1 - \frac{1}{2} (\theta - \hat{\theta}_p)^T \Gamma^{-1} (\theta - \hat{\theta}_p) + C_2 - \frac{\mu}{2} (\log |\Omega| + \text{tr}(\hat{\Omega}_p \Omega^{-1}))$$

where C_1 and C_2 represent constant for normal and inverse-Wishart distributions, respectively, $\text{tr}(X)$ is a sum of diagonal elements of matrix, $\mu = \nu$ (degree of freedom) + m (dimension of Ω) + 1. The $\hat{\theta}_p$ and $\hat{\Omega}_p$ represent mode of each prior distribution.

The second derivative of O_{P+S} with respect to θ and Ω (diagonal matrix for simplicity) is defined as FIM for analysis with priors. Since inverse of FIM is the variance-covariance matrix of estimates, inverse of FIM_{P+S} is considered variance-covariance matrix of estimates in the analysis with priors (V_{P+S}). The following equations represent an element of FIM_{P+S} for j^{th} fixed effect parameter and variance of k^{th} random effect parameter, respectively (the first term is zero if the parameter does not incorporate a prior).

$$FIM_{P+S(j,j)} = \Gamma^{-1}_{(j,j)} + FIM_{S(j,j)} \dots (3)$$

$$FIM_{P+S(k,k)} = \frac{\mu}{2} \cdot \left(-\frac{1}{(\omega_{(k)}^2)^2} + \frac{2(\hat{\omega}_{(k)}^2)}{(\omega_{(k)}^2)^3} \right) + FIM_{S(k,k)} \dots (4)$$

The proposed information of a prior related to the sparse data is as follows.

$$N_{P(j)} = V_{S(j,j)} / V_{P+S(j,j)} - 1 = FIM_{S(j,j)}^{-1} / FIM_{P+S(j,j)}^{-1} - 1 \dots (5)$$

In practice, the FIM_S in equations (1) and (2) are obtained by PFIM of R software tool [2]. Hence, the FIM_{P+S} is obtained by adding minus second derivative matrix of log of prior distributions (first term of each equation (3) and (4)) to the FIM_S from PFIM as below. The A and B represent minus second derivative of log of the prior distribution with respect to θ and Ω , respectively.

$$FIM_{P+S} = \begin{pmatrix} A_{(j \times j)} & 0 & 0 \\ 0 & B_{(k \times k)} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} FIM_{S(j \times j)} & 0 & 0 \\ 0 & FIM_{S(k \times k)} & FIM_{S(k \times l)} \\ 0 & FIM_{S(l \times k)} & FIM_{S(l \times l)} \end{pmatrix}$$

Minus second derivatives of log of prior distributions (elements without priors are zero) FIM for data obtained from PFIM

Results

FIM_{P+S} introduced in equations (3) and (4) was checked by using simulated PK data. As shown in Figure 2, although there were some trends to be checked further, variance of parameter estimates obtained from inverse of FIM_{P+S} was generally comparable with corresponding variance obtained from R matrix of NONMEM, suggesting use of the FIM_{P+S} is acceptable.

Using the FIM_{P+S} and FIM_S , the information of prior (equation 5) was calculated for analysis of simulated sparse data with different priors (Figure 3). As expected, parameter estimates of analysis with priors of large information (e.g., $N_p \geq 10$) were mostly governed by the priors. This suggests that it is necessary to make stricter assumption that those parameters are exchangeable between population of the sparse data and population behind the priors. Meanwhile, parameter estimates were little influenced by choice of the priors when analyzing the sparse data with priors of small information (e.g., $N_p \leq 1.0$), suggesting that the estimates might be around true parameter values even if wrong priors are specified, though it depends on extent of the discrepancy.

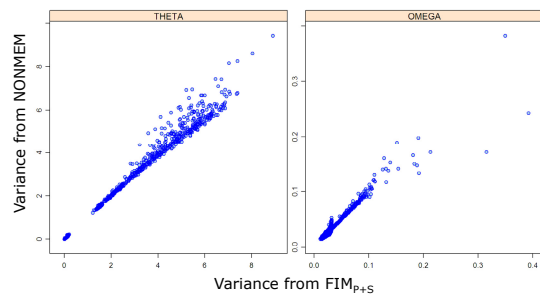


Figure 2. Variance of estimates from inverse of FIM_{P+S} of equation (3, 4) and those of NONMEM (from the second derivative matrix of the objective function with respect to the parameters in analysis with \$PRIOR) (o).

Note: Comparison of variance of Ω was conducted as $\mu = \nu$ (NONMEM degree of freedom) in stead of $\mu = \nu + m + 1$. Simulation condition: 1CMP, $\theta_{CL} = 1.5 \text{ h}^{-1}$, $\theta_{CL} = 5.0 \text{ L/h}$, $\theta_{CL} = 40 \text{ L}$, inter-individual variability = exponential error model with diagonal Ω ($\omega_{ka}^2, \omega_{km}^2, \omega_{Q}^2 = 0.1$), residual variability = proportional error with $\sigma^2 = 0.1$. Prior for θ_{CL} : $\theta_{CL} = 1.5 \text{ h}^{-1}$, $\theta_{CL} = 5.0 \text{ L/h}$, $\theta_{CL} = 40 \text{ L}$ with different information (CV from 1 to 20%). Prior for Ω (right panel): diagonal Ω ($\omega_{ka}^2, \omega_{km}^2, \omega_{Q}^2 = 0.1$) with μ from 3 to 50. Sampling design of sparse data: 1, 3, 6 h/subject ($n = 100$), dose = 150 mg. One outlier for Ω (1.4, 0.2) was removed.

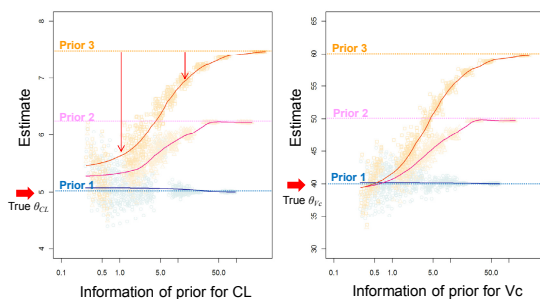


Figure 3. Relationships between parameter estimates and information of priors (N_p) when analyzing sparse data with different information and location (mode) of priors. The $\circ$, Δ , $\square$ represent estimates of analysis with priors, whose mode values are 0, 25, and 50% different from true parameters (prior 1, 2, and 3), respectively.

Simulation condition: Same as Figure 2 but for θ_{CL} and θ_{Vc} , the location (mode) of the priors are 0, 25, and 50% different from the true θ s. In this simulation, diagonal Ω was estimated without the priors.

The Fisher information matrix for analysis of sparse data with priors (FIM_{P+S}) is considered to associate with precision of the estimates. Therefore, it is considered possible to use the FIM_{P+S} for evaluating optimal study designs in advance to analysis of sparse data with priors. Figure 4 shows expected precision of estimates in different study designs with different priors obtained from FIM_{P+S} and NONMEM (\$PRIOR). As a whole, the precision of estimates based on FIM_{P+S} was consistent to that of NONMEM.

Table 1 shows an example of design evaluations based on FIM_{P+S} in analysis with priors (a two-compartment model with a first-order absorption/non-linear elimination). Parameters ka , km , Q which were assumed difficult to estimate incorporated the priors. Precision of all estimates (#3,4) was improved by adding priors (#1,2). The precision in the sparse design (#1) was comparable with the dense one (#2). In both designs, however, strict assumption of exchangeability of parameters between population of the data and population behind the priors is needed since information of the priors was large.

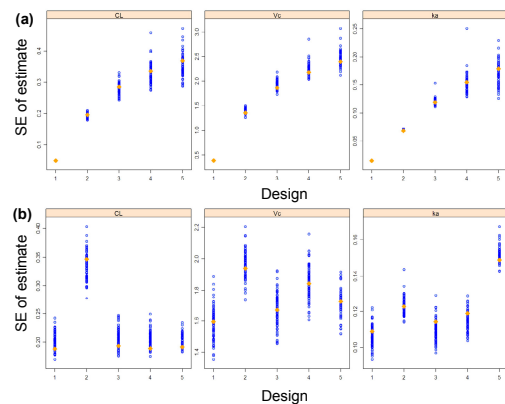


Figure 4. Precision of parameter estimates (SE) in analysis of sparse data with different priors (a) and different designs (b). The $\circ$ and Δ represent SE of estimate obtained from inverse of FIM_{P+S} and R matrix of NONMEM analysis with \$PRIOR, respectively.

Simulation condition: Same as Figure 2, but the priors for θ_{CL} and θ_{Vc} were used. The θ_{CL} and Ω were estimated without the priors. In addition, the following conditions were added. (a) Different information of priors for θ_{CL} and θ_{Vc} : Design #1 [CV = 1%], #2 [CV = 5%], #3 [CV = 10%], #4 [CV = 15%], and #5 [CV = 20%]. (b) Different sampling time points: Design #1 [1, 3, 6, 12 h], #2 [1, 3, 6 h], #3 [1, 3, 12 h], #4 [1, 6, 12 h], and #5 [3, 6, 12 h].

Table 1. Precision of parameter estimates (RSE%) by each design when analyzing data with priors (θ s and ω^2 s for ka , km , Q) and without priors. The N_p represents information of the priors.

Evaluation (design)		ka	km	Q	V_{max}	V_2	V_3	$\omega^2 ka$	$\omega^2 km$	$\omega^2 Q$	$\omega^2 V_{max}$	$\omega^2 V_2$	$\omega^2 V_3$
#1 (1)	RSE% N_p	16.9 >1000	23.9 >1000	4.1 >1000	17.6 --	16.2 --	8.5 --	47.0 3.8	45.1 2.5	50.4 8.6	32.9 --	49.7 --	27.3 --
#2 (2)	RSE% N_p	16.1 59.5	20.6 207.9	4.1 677.4	15.6 --	15.2 --	6.6 --	44.8 2.4	42.6 1.8	39.3 1.2	31.1 --	41.3 --	23.6 --
#3 (1)	RSE% No prior	2756.2 --	11297.9 --	3674.8 --	8320.2 --	2294.2 --	3571.6 --	103.5 --	84.1 --	156.3 --	49.0 --	69.7 --	36.6 --
#4 (2)	RSE% No prior	125.5 --	297.6 --	106.6 --	220.1 --	114.1 --	100.3 --	82.6 --	70.8 --	58.4 --	42.3 --	53.5 --	25.6 --

Design 1: $n = 100$, sampling time points at visit 1 [1,2,3 h], visit 2 [10,11,12 h], dose = 150 mg. Design 2: $n = 100$, sampling time points [1,1.5,2,3,4, 6, 8, 10, 12 h], dose = 150 mg. Model: a two compartment model with a 1st-order absorption/non-linear elimination (θ_{CL} , θ_{CL} , θ_{CL} , θ_{CL} , θ_{CL} = 1.2 h^{-1} , 1.15 ng/ml , 25 L/h , 110 mg/h , 80 L , 40 L). The inter-individual variability is an exponential model (all $\omega^2 = 0.1$). The residual variability is a proportional error model ($\sigma^2 = 0.08$). Mode (variance) of priors for θ_{CL} , θ_{CL} , θ_{CL} = 1.2 (0.05), 1.15 (0.100), 25 (1.10). The prior of Ω (a diagonal matrix of ka , km , Q) is all 0.1. The μ for prior of $\Omega = 3$ (ν) + 3 (m) + 1.

Conclusion

The information of priors based on variance of estimates from Fisher information with or without priors is considered a useful index for evaluating strength of the priors relative to current data. Use of the Fisher information matrix with priors is also considered as an option for evaluating study designs when the data is supposed to be analyzed with priors. To ensure this approach, further studies in different conditions are required

References

- [1] Gisleskog PO, Karlsson MO, Beal SL. Use of prior information to stabilize a population data analysis. J Pharmacokinet Pharmacodyn 2002; 29:473-505.
- [2] Retout S, Mentre F, Bruno R. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. Statist. Med. 2002; 21:2623-39
- [3] Mori H. Evaluation of Prior Information in Bayesian Inference. J. Japan Statist. Soc. 2010; 40 :1-22, 2010.