Research Article

Performance and Robustness of the Monte Carlo Importance Sampling Algorithm Using Parallelized S-ADAPT for Basic and Complex Mechanistic Models

Jurgen B. Bulitta^{1,2,3} and Cornelia B. Landersdorfer^{1,2}

Received 22 December 2010; accepted 25 January 2011; published online 4 March 2011

Abstract. The Monte Carlo Parametric Expectation Maximization (MC-PEM) algorithm can approximate the true log-likelihood as precisely as needed and is efficiently parallelizable. Our objectives were to evaluate an importance sampling version of the MC-PEM algorithm for mechanistic models and to qualify the default estimation settings in SADAPT-TRAN. We assessed bias, imprecision and robustness of this algorithm in S-ADAPT for mechanistic models with up to 45 simultaneously estimated structural parameters, 14 differential equations, and 10 dependent variables (one drug concentration and nine pharmacodynamic effects). Simpler models comprising 15 parameters were estimated using three of the ten dependent variables. We set initial estimates to 0.1 or 10 times the true value and evaluated 30 bootstrap replicates with frequent or sparse sampling. Datasets comprised three dose levels with 16 subjects each. For simultaneous estimation of the full model, the ratio of estimated to true values for structural model parameters (median [5–95% percentile] over 45 parameters) was 1.01 [0.94–1.13] for means and 0.99 [0.68–1.39] for between-subject variances for frequent sampling and 1.02 [0.81–1.47] for means and 1.02 [0.47–2.56] for variances for sparse sampling. Imprecision was $\leq 25\%$ for 43 of 45 means for frequent sampling. Bias and imprecision was well comparable for the full and simpler models. Parallelized estimation was 23-fold (6.9-fold) faster using 48 threads (eight threads) relative to one thread. The MC-PEM algorithm was robust and provided unbiased and adequately precise means and variances during simultaneous estimation of complex, mechanistic models in a 45 dimensional parameter space with rich or sparse data using poor initial estimates.

KEY WORDS: mechanistic population pharmacokinetic; Monte Carlo importance sampling; Monte Carlo parametric expectation maximization (MC-PEM) algorithm; nonlinear mixed-effects modeling in S-ADAPT; parallel computing; pharmacodynamic modeling.

INTRODUCTION

Substantial progress in experimental disciplines over the last decades has enabled measurement of many biological variables to characterize the time-course of drug concentrations, drug effects, and disease states. Mathematical models to describe such datasets are often complex and may require many estimated model parameters. As almost every biological process contains true variability, nonlinear mixed-effects modeling that can estimate this variability became a method of choice for data analysis.

In the past, nonlinear mixed-effects modeling algorithms that approximate the equation for the true log-likelihood were commonly applied, since these methods are computationally fast. However, estimation times of these approximate

methods such as the first-order conditional estimation (FOCE) method increase substantially with model complexity and with the number of dependent variables. The FOCE method employs a gradient search algorithm that may become notoriously unstable for estimation of complex mechanistic models with multiple dependent variables and many model parameters with biological variability.

Estimation of complex mechanistic models greatly benefits from robust algorithms that can handle many parameters with biological variability. Expectation maximization (EM) algorithms are robust, as they use integration instead of gradient search methods to optimize (update) parameter estimates. State-of-the-art EM algorithms (1–6) provide the additional advantage that they can approximate the true log-likelihood as precisely as needed by increasing the number of Monte Carlo samples used to approximate the integrals for calculation of the true log-likelihood. Importantly, algorithms, such as the FOCE method, which calculate the exact solution of a formula that approximates the log-likelihood can only improve the quality of the approximation by changing the algorithm (i.e., by using a more complex formula that approximates the log-likelihood more closely). Such methods that are based on

¹Ordway Research Institute, 150 New Scotland Avenue, Albany, New York 12208, USA.

²Department of Pharmaceutical Sciences, School of Pharmacy and Pharmaceutical Sciences, State University of New York at Buffalo, Buffalo, New York, USA.

³To whom correspondence should be addressed. (e-mail: j@bulitta.com)

approximate formulas for the log-likelihood usually provide no estimate for the quality of the approximation.

Robust and powerful nonparametric algorithms for population modeling, such as the nonparametric expectation maximization (NPEM) and nonparametric adaptive grid (NPAG) algorithms, which calculate the true log-likelihood, formed the core of the MM-USCPACK since more than a decade (7–11). The NPEM and the NPAG algorithm describe the between-subject variability (BSV) by a nonparametric distribution that can take any shape. During the last 5–10 years, parametric EM algorithms became available in software packages such as S-ADAPT (12), ADAPT V (13,14), Monolix (15), and recently also in NONMEM® 7 (16).

The original Monte Carlo Parametric Expectation Maximization (MC-PEM) algorithm employed direct sampling to approximate the multidimensional integrals (3,17). A large number of random model parameters requires integration in a high dimensional parameter space which makes direct sampling less efficient. Therefore, Bauer and Guzy (18–20) proposed and evaluated importance sampling versions of the MC-PEM algorithm (also called the importance sampling EM algorithm in ADAPT 5 and NONMEM® 7) that approximate the multidimensional integrals more efficiently than direct sampling by drawing Monte Carlo samples at more informative positions.

Importance sampling MC-PEM methods are now available in S-ADAPT (12), ADAPT V (13,14), and NONMEM® 7 (16). These methods provide a significant advantage when working with mechanistic models that may have 20 or more parameters to estimate along with their respective biological variability (21–38) and Landersdorfer *et al.* (Abstract at the American Conference on Pharmacometrics, Mashantucket, CT, USA, 2009). These problems present an at least 20-dimensional parameter space for integration. Given a hypothetical case with 20 parameters and a grid of 10 initial estimates for each parameter, this would lead to having to compute 10^{20} grid points for every subject during the first iteration. Even if the number of initial values per parameter was reduced to three, there are still 3^{20} (approximately 3.5×10^9) grid points per subject. Since mechanistic models are typically specified as nonlinear differential equations, the “curse of dimensionality” prevents evaluation of so many grid points.

A comprehensive review of the MC-PEM algorithm as published by Schumitzky (2) and Bauer *et al.* (4) is beyond the scope of this paper. The MC-PEM algorithm and its implementation in S-ADAPT are described in detail in the S-ADAPT manuals (12,19). The MC-PEM algorithm is particularly amenable to efficient parallelization, since the expectation step (E-step) contains the multidimensional integration for each subject and requires the vast majority of computation time, especially for complex models. Therefore, EM algorithms are well suited to solve mechanistic models on parallelized computer clusters and super-computers. There are no literature reports of a systematic evaluation of the performance and robustness of parametric EM algorithms for large numbers of random model parameters.

The primary hypothesis is that the importance sampling version of the MC-PEM algorithm is robust and efficient for estimation of basic and complex mechanistic models. The first objective was to systematically evaluate bias, precision, and

robustness of an importance sampling version of the MC-PEM algorithm. Secondly, we sought to qualify the default estimator settings implemented in the SADAPT-TRAN¹ preprocessing tool.

MATERIALS AND METHODS

The following features of the MC-PEM algorithm were systematically evaluated: (1) ability to optimize parameter estimates in an intermediate and high-dimensional parameter space, (2) robustness toward poor initial estimates, (3) performance for estimation of mechanistic models with rich or sparse data, and (4) parallelization efficiency for basic and complex models. These features were evaluated using basic and complex mechanistic models and the default estimation settings in SADAPT-TRAN (see companion paper for a description of SADAPT-TRAN).

Simulations. The time course of plasma concentrations and effect profiles were simulated for 1,000 patients each for three doses of 500, 2,000, or 8,000 mg given as a 30-min infusion at 0, 24, and 48 h (Figs. 1 and 2) using Berkeley Madonna (Version 8.3.14, University of California). All models were simulated using a major diagonal variance-covariance (var-cov) matrix. The pharmacokinetic (PK) model contained a parallel first-order and mixed-order elimination. The Beal M3 method (39) was implemented for PK and pharmacodynamic (PD) variables as described in the S-ADAPT manual (12).

Effect profiles from eight indirect response models and one direct effect model (Table I, Fig. 1) were simulated. The full model contained 14 differential equations, 45 random (P type) parameters with a population mean and estimated BSV, and 20 residual error variances (V type parameters). To evaluate models with a high-dimensional parameter space, 30 raw datasets with frequent sampling were randomly created: blood samples were obtained at 0.25, 0.5, 1, 2, 4, 8, 16, 24, 28, 32, 36, 48, 52, 56, 60, 72, 84, and 96 h and effects were measured at 0, 0.5, 1, 2, 4, 8, 12, 24, 28, 32, 36, 48, 52, 56, 60, 72, 84, 96, 120, 144, 168, and 192 h.

Then 30 datasets with randomly selected sparse samples were created. Each patient had a blood sample for PK measurements at 0.5 h and at six times randomly chosen from the following list: 1.5, 5, 12, 24, 48, 49, 60, 72, 84, or 96 h. For indirect response models 1–4 and the direct effect model, effects were observed at 0 h and at two randomly drawn time points for each patient. For precursor indirect response models, effects were observed at 0 h and at three randomly drawn time points for each patient. Sampling times for PD observations were randomly selected from the following list: 1.5, 5, 12, 24, 29, 48, 72, 96, 120, or 168 h.

Parameterization of Precursor Indirect Response Models. To make the precursor indirect response models with inhibition or stimulation of input of precursor uniquely identifiable (parts 7 and 9 in Table I), we estimated the half-life of loss from the response compartment as well as the difference between half-life of loss from the precursor and half-life of loss from the response compartment. This choice of parameterization must be

¹ The SADAPT-TRAN software is freely available via <http://bmsr.usc.edu/>.

Fig. 1. Structure of the full model containing a two-compartment PK model, four indirect response (IDR) models, a direct effect model, and four precursor indirect response models

manually specified by the user in the SADAPT-TRAN model code (Fig. 3, part \$OUTPUT_GLB). Similar to a flip-flop situation for a linear one-compartment model with first order absorption, this parameterization guarantees that the half-life of loss from the response compartment is faster than the half-life of loss from the precursor compartment for every patient. This parameterization was employed both during simulation and estimation. This choice also tended to improve the estimability of IC₅₀, SC₅₀, I_{max} and S_{max} for the other two precursor indirect response models (parts 8 and 10 in Table I).

Estimation. For the bootstrap analysis, all models were estimated using the true model structure, true parameter variability model with a major diagonal var-cov matrix, and true additive plus proportional residual error model for every dependent variable. In a side analysis, a model with a full var-cov matrix was estimated to evaluate the increase in estimation time relative to a model with major diagonal var-cov matrix.

The importance sampling MC-PEM algorithm (pmethod=4 in S-ADAPT) was employed. During each iteration, this method started with 100 random samples (NPOP=100 was chosen in S-ADAPT) to obtain initial estimates for a maximum a posteriori (MAP) Bayesian analysis for each subject. After the MAP Bayesian step, the envelope function was centered at the MAP Bayesian solution and 1,000 Monte Carlo samples (NPOPC=1,000, Table II) were randomly drawn to obtain each subject's conditional means and conditional var-cov matrix. A multivariate normal envelope function was used which is the default in S-ADAPT. While the MAP Bayesian step is not a part

of the core EM algorithm, this step helps to place the envelope function at a good position for each subject.

The envelope function was allowed to cover an increased search area by setting gefficiency to 0.6 (Table II) to cover potentially more extreme parameter values. This setting causes S-ADAPT to “waste” approximately 40% of the Monte Carlo samples during each iteration. To efficiently adjust the search area covered by the envelope function, S-ADAPT was allowed to multiply the envelope function by up to a factor of 20 in either direction (gamma_min=0.05 and gamma_max=20 setting in S-ADAPT; Table II).

The conditional means and conditional var-cov matrices were deleted after every 15th iteration (ndelpar=15 in S-ADAPT) to prevent S-ADAPT from getting stuck in local minima. Standard errors were computed using the default poperr_type=8 setting in S-ADAPT which uses the full second-order derivative matrix for error assessment (see S-ADAPT technical guide, appendix D, for details (19)).

All estimation settings not described here were kept at the default values in S-ADAPT. The estimation settings described above present the default settings for SADAPT-TRAN that were to be qualified for estimation of mechanistic models in the present analysis. These settings are summarized under RUN_SETTINGS in Table II.

All estimations were performed on six dual central processing unit (CPU) servers hosting a total of 12 Intel® Xeon® X5570 quad core CPUs connected by a standard 1 Gbit Ethernet network. The hyper-threading functionality of these CPUs was used and all estimations were run in

Fig. 2. Illustration of one dataset with frequent sampling for the full model including plasma concentrations and nine different pharmacodynamic effects at three dose levels (the dependent variable identifiers [DVIDs] refer to the respective model part number shown in Table I; doses were given at 0, 24, and 48 h)

hyper-threaded mode with two simultaneously executed threads per CPU core.

Evaluation of Parallelization Efficiency. The estimation times were compared for 150 iterations of the full model (Fig. 1) under different scenarios. Estimation was performed

(a) on a single thread (i.e., half of a hyper-threaded CPU core) on the master node, (b) on eight threads on the master node, (c) on one thread on the master node and seven threads on a worker node connected via the network, or (d) on eight threads on the master node and 40 threads on five

Table I. Components of the Full Population Pharmacokinetic/Pharmacodynamic Model

Part #	No. of Cmt	Model description	Dependent variable	Slow onset of PD response	Rebound of PD response
1	2	2-Cmt. model with parallel first-order and mixed-order (Michaelis-Menten) elimination	Drug concentration in central Cmt		
2	1	IDR with inhibition of input	PD effect 1	No	No
3	1	IDR with inhibition of loss	PD effect 2	No	No
4	1	IDR with stimulation of input	PD effect 3	No	No
5	1	IDR with stimulation of loss	PD effect 4	No	No
6	0	Stimulatory direct effect model	PD effect 5	No	No
7	2	Pre-IDR with inhibition of input into precursor Cmt	PD effect 6 in response Cmt	Yes	No
8	2	Pre-IDR with inhibition of transfer from precursor to response Cmt	PD effect 7 in response Cmt	Yes	Yes
9	2	Pre-IDR with stimulation of input into precursor Cmt	PD effect 8 in response Cmt	Yes	No
10	2	Pre-IDR with stimulation of transfer from precursor to response Cmt	PD effect 9 in response Cmt	Yes	Yes

Part # corresponds to DVID # in Fig. 2, Cmt compartment(s), IDR indirect response model, PD pharmacodynamic, Pre-IDR precursor indirect response model comprised of a precursor and a response compartment

different worker nodes. As estimation times for computation on the master node were very consistent (within 3%), two independent replicates were run for cases (a) and (b). Estimation times for cases (c) and (d) are based on at least six independent analyses using different datasets, unless stated otherwise.

Calculation of Bias and Imprecision. The 30 bootstrap datasets yielded 30 estimates for each of the 45 structural model parameters and their BSV as well as for the 20 residual error parameters. The ratio of the individual parameter estimates to the true parameter value was calculated for each replicate and the bias was denoted as a deviation of the average ratio from the ideal value 1.0. Imprecision is described as the standard deviation of the 30 ratios for the estimated to true parameter value.

RESULTS

Bias and Precision of Population Means. The 30 bootstrap replicates of the full model with frequent sampling ("rich data") showed average ratios of estimated divided by true population means between 0.9 and 1.13 for 41 of 45 (91%) population means indicating a bias of 13% or less (Table III and its summary by parameter type in Table IV). The standard deviation of this ratio was 0.25 or less for 43 of 45 (96%) population mean parameter estimates of the full model (Fig. 1) suggesting good precision (Table III). These results apply to an estimation scenario with initial estimates being set to 0.1 or 10 times the true values. Virtually identical ($\pm 1\%$ for 43 of 45 population means) bias and precision were obtained for the full model with frequent sampling, if initial estimates were set to 0.5 or two times instead of 0.1 or 10 times the true values (results not shown). Therefore, initial estimates for all further testing were set to 0.1 or 10 times the true value.

Performance for Low- and High-Dimensional Parameter Spaces. The importance sampling MC-PEM algorithm for models with a lower dimensional parameter space was evaluated via estimation of models with 15 random parameters as opposed to 45 random parameters for the full model. Individual model parts are listed in Table I. The full estimation problem was split into five separate problems and the model parts were grouped as follows: (a) parts 1, 2, and 7, (b) parts 1, 3, and 8, (c) parts 1, 4, and 9, (d) parts 1, 5, and 10, or (e) parts 1 and 6.

Thirty bootstrap replicates with the same data as for the full model were estimated for the five sets of separate models (i.e., 150 bootstrap runs with 48 subjects each in total). Bias and precision (Table III) for all model parameters were very similar for estimation of the full model compared to estimation of the five separate models suggesting that the importance sampling MC-PEM algorithm performed similarly well in a 15- and 45-dimensional parameter space for datasets with frequent sampling.

When data with sparse sampling were used for estimation of the full model, average ratios of estimated divided by true parameter values ranged from 0.7 to 1.3 for 39 of 45 (87%) population means (Tables III and V). The standard deviation of this ratio across 30 bootstrap replicates was 0.45 or less for 40 of 45 (89%) population means (Table III). As expected, ratios of estimated to true parameter values were more variable for sparse compared to frequently sampled datasets.

Bias and Precision for Between Subject Variability and Residual Error. The average ratios of between-subject variance estimates for the scenarios with frequent sampling ranged from 0.75 to 1.25 for 35 of 45 (78%) parameters. As expected, the estimates for BSV were less precise for datasets with sparse sampling and 48 subjects.

Bias and precision for proportional residual error was reasonable and (slightly) better for models with frequent compared to sparse sampling (Table III). As the true additive

residual error was set to very small values (standard deviation of 0.01 or 0.5; Table III) in comparison to the size of the respective concentrations and effects (Fig. 2), it was expected that estimates for the additive residual error terms were less precise.

Robust Choice of Initial Estimates. There was no termination of any run in this simulation–estimation study due to

numerical instability or any other reason. Relative standard errors were successfully computed for all models by S-ADAPT using the default Poperr_type=8 option. The following choice for initial estimates was found robust for the MC-PEM algorithm. Elimination clearance was set to small values and volumes of distribution to large values to prevent negligible

```

1  $PROJECT THIS WILL BE A TERRIFIC PROJECT
2
3  $DIFFEQ_DIF
4
5  ;#####
6  ; Calculation of time-dependent variables
7  DC1 = X(1)/V1
8  DC2 = X(2)/V2
9  CLMM = VMAX / (KM+DC1)
10
11  AINH = 1 - AIMAX * DC1 / (DC1 + AIC50)
12  BINH = 1 - BIMAX * DC1 / (DC1 + BIC50)
13  CSTIM = 1 + CSMAX * DC1 / (DC1 + CSC50)
14  DSTIM = 1 + DSMAX * DC1 / (DC1 + DSC50)
15
16  FINH = 1 - FIMAX * DC1 / (DC1 + FIC50)
17  GINH = 1 - GIMAX * DC1 / (DC1 + GIC50)
18  HSTIM = 1 + HSMAX * DC1 / (DC1 + HSC50)
19  ISTIM = 1 + ISMAX * DC1 / (DC1 + ISC50)
20
21  ;#####
22  ; Differential equations
23  XP(1) = R(1) - (CL+CLMM+CLd)*DC1 + CLd*DC2
24  XP(2) = CLd * DC1 - CLd*DC2
25
26  XP(3) = AKIN * AINH - AKOUT * X(3)
27  XP(4) = BKIN - BKOUT * BINH * X(4)
28  XP(5) = CKIN * CSTIM - CKOUT * X(5)
29  XP(6) = DKIN - DKOUT * DSTIM * X(6)
30
31  XP(7) = FKIN * FINH - FKTR * X(7)
32  XP(8) = GKin - GKTR * GINH * X(8)
33  XP(9) = HKIN * HSTIM - HKTR * X(9)
34  XP(10) = IKIN - IKTR * ISTIM * X(10)
35
36  XP(11) = + FKTR * X(7) - FKOUT * X(11)
37  XP(12) = + GKTR * GINH * X(8) - GKOUT * X(12)
38  XP(13) = + HKTR * X(9) - HKOUT * X(13)
39  XP(14) = + IKTR * ISTIM * X(10) - IKOUT * X(14)
40
41  $OUTPUT_GLB ; defining time independent global variables
42
43  VMAX = CLIC * KM
44
45  ;#####
46  ; Indirect response models
47  AKOUT = LOG(2) / AT12OUT
48  BKOUT = LOG(2) / BT12OUT
49  CKOUT = LOG(2) / CT12OUT
50  DKOUT = LOG(2) / DT12OUT
51
52  AKIN = ABASE * AKOUT
53  BKIN = BBASE * BKOUT
54  CKIN = CBASE * CKOUT
55  DKIN = DBASE * DKOUT
56
57  ;#####
58  ; Precursor indirect response models
59  FKOUT = LOG(2) / FT12OUT
60  GKOUT = LOG(2) / GT12OUT
61  HKOUT = LOG(2) / HT12OUT
62  IKOUT = LOG(2) / IT12OUT
63  FKTR = LOG(2) / (FT12TR+FT12OUT)
64  GKTR = LOG(2) / (GT12TR+GT12OUT)

```

Fig. 3. SADAPT-TRAN model code to define the full mechanistic model for simultaneous estimation

```

65 HKTR = LOG(2) / (HT12TR+HT12OUT)
66 IKTR = LOG(2) / (IT12TR+IT12OUT)
67 FKIN = FBASE * FKOUT
68 GKIN = GBASE * GKOUT
69 HKIN = HBASE * HKOUT
70 IKIN = IBASE * IKOUT
71
72 $OUTPUT_ICS ; setting initial conditions
73
74 X(3) = X(3) + ABASE
75 X(4) = X(4) + BBASE
76 X(5) = X(5) + CBASE
77 X(6) = X(6) + DBASE
78
79 X(7) = X(7) + FBASE * FKOUT / FKTR
80 X(8) = X(8) + GBASE * GKOUT / GKTR
81 X(9) = X(9) + HBASE * HKOUT / HKTR
82 X(10)= X(10)+ IBASE * IKOUT / IKTR
83
84 X(11)= X(11)+ FBASE
85 X(12)= X(12)+ GBASE
86 X(13)= X(13)+ HBASE
87 X(14)= X(14)+ IBASE
88
89 $OUTPUT_EQN ; specification of system outputs [Y(n)]
90
91 if (X(1).LT.0) X(1) = 0
92 if (X(3).LT.0) X(3) = 0
93 if (X(4).LT.0) X(4) = 0
94 if (X(5).LT.0) X(5) = 0
95 if (X(6).LT.0) X(6) = 0
96 if (X(11).LT.0) X(11) = 0
97 if (X(12).LT.0) X(12) = 0
98 if (X(13).LT.0) X(13) = 0
99 if (X(14).LT.0) X(14) = 0
100
101 Y(1) = X(1)/V1
102 Y(2) = X(3)
103 Y(3) = X(4)
104 Y(4) = X(5)
105 Y(5) = X(6)
106
107 C1 = X(1)/V1
108 Y(6) = EBASE * (1 + EEMAX * C1 / ( C1 + EEC50 ) )
109
110 Y(7) = X(11)
111 Y(8) = X(12)
112 Y(9) = X(13)
113 Y(10)= X(14)
114
115 $VARMOD_EQN ; specifying residual error variances
116
117 V(1) = ( SDin + SDsl*Y(1) ) * ( SDin + SDsl*Y(1) )
118 V(2) = ( ADin + ADsl*Y(2) ) * ( ADin + ADsl*Y(2) )
119 V(3) = ( BDin + BDsl*Y(3) ) * ( BDin + BDsl*Y(3) )
120 V(4) = ( CDin + CDsl*Y(4) ) * ( CDin + CDsl*Y(4) )
121 V(5) = ( DDin + DDsl*Y(5) ) * ( DDin + DDsl*Y(5) )
122 V(6) = ( EDin + EDsl*Y(6) ) * ( EDin + EDsl*Y(6) )
123 V(7) = ( FDin + FDsl*Y(7) ) * ( FDin + FDsl*Y(7) )
124 V(8) = ( GDin + GDsl*Y(8) ) * ( GDin + GDsl*Y(8) )
125 V(9) = ( HDin + HDsl*Y(9) ) * ( HDin + HDsl*Y(9) )
126 V(10)= ( IDin + IDsl*Y(10) ) * ( IDin + IDsl*Y(10) )
127
128 $POPMOD_EQN ; covariate effects - none for this model

```

Fig. 3. (continued)

Table II. Content of the Parameter Settings File to Specify the Full Model (Empty Columns with No Information for this Model were Omitted; Initial Means and Variances for Logistically Transformed Parameter are on Transformed Scale)

#PNAME	PMEAN	PCOV	PTYPE	PBLOCK	PTRANSF	PLOW	PHIGH	VARINI	VARBURN
CL	40	1	P	1	L			1	30
CLIC	2.6	1	P	2	L			1	30
KM	0.3	1	P	3	L			1	30
CLD	100	1	P	4	L			1	30
V1	100	1	P	5	L			1	30
V2	400	1	P	6	L			1	30
AT12OUT	20	1	P	7	L			1	30
ABASE	100	1	P	8	L			1	30
AIMAX	0.01	4	P	9	O	0	1	4	30
AIC50	30	1	P	10	L			1	30
BT12OUT	40	1	P	11	L			1	30
BBASE	100	1	P	12	L			1	30
BIMAX	0.01	4	P	13	O	0	1	4	30
BIC50	60	1	P	14	L			1	30
CT12OUT	60	1	P	15	L			1	30
CBASE	10	1	P	16	L			1	30
CSMAX	2	1	P	17	L			1	30
CSC50	90	1	P	18	L			1	30
DT12OUT	80	1	P	19	L			1	30
DBASE	100	1	P	20	L			1	30
DSMAX	0.5	1	P	21	L			1	30
DSC50	150	1	P	22	L			1	30
EBASE	100	1	P	23	L			1	30
EEMAX	0.3	1	P	24	L			1	30
EEC50	150	1	P	25	L			1	30
FT12OUT	20	1	P	26	L			1	30
FT12TR	100	1	P	27	L			1	30
FBASE	100	1	P	28	L			1	30
FIMAX	0.01	4	P	29	O	0	1	4	30
FIC50	15	1	P	30	L			1	30
GT12OUT	40	1	P	31	L			1	30
GT12TR	80	1	P	32	L			1	30
GBASE	100	1	P	33	L			1	30
GIMAX	0.01	4	P	34	O	0	1	4	30
GIC50	30	1	P	35	L			1	30
HT12OUT	60	1	P	36	L			1	30
HT12TR	60	1	P	37	L			1	30
HBASE	10	1	P	38	L			1	30
HSMAX	2	1	P	39	L			1	30
HSC50	45	1	P	40	L			1	30
IT12OUT	80	1	P	41	L			1	30
IT12TR	40	1	P	42	L			1	30
IBASE	100	1	P	43	L			1	30
ISMAX	0.5	1	P	44	L			1	30
ISC50	120	1	P	45	L			1	30
SDin	2		V						
SDsl	1		V						
ADin	5		V						
ADsl	1		V						
BDin	5		V						
BDsl	1		V						
CDin	5		V						
CDsl	1		V						
DDin	5		V						
DDsl	1		V						
EDin	5		V						
EDsl	1		V						
FDin	5		V						
FDsl	1		V						
GDin	5		V						
GDsl	1		V						
HDin	5		V						

Table II. (continued)

#PNAME	PMEAN	PCOV	PTYPE	PBLOCK	PTRANSF	PLOW	PHIGH	VARINI	VARBURN
HDsl	1		V						
IDin	5		V						
IDS1	1		V						
COVARIATES									
RUN_SETTINGS									
NFILE	Simultaneous_Estimation_of_Full_Model__10x_off_initials								
VERS	1								
mdelete <NFILE>_*									
DATAFILE	../Data/SADAPT_Data_Ful_Model_BS_00001.csv								
COVFILE	../Data/Covariates.csv								
BEOFILE	beo.txt								
NPOPITER	150								
GEFFICIENCY	0.6								
PMETHOD	4								
NPOP	100								
NPOPC	1000								
NDELPAR	15								
ATOL	5								
RTOL	5								
Gamma_min	0.05								
Gamma_max	20								
beosetup									
RUN_COMMANDS									
piteraten									
?finish.txt									
stop									

External scripts like the commands in the finish.txt file can be included via a “?” sign

PNAME parameter name, *PMEAN* initial mean, *PCOV* initial variance, *PTYPE* parameter type (*P* mean with variability, *V* residual error parameter), *PBLOCK* number of block in the variance covariance matrix, *PTRANSF* parameter transformation type (*L* log-normal, *O* logistic, *N* normal), *PLOW* lower bound of logistically transformed variable (default=0), upper bound of logistically transformed variable (default=1), *VARINI* initial variance during the variance burn phase, *VARBURN* number of iterations of the variance burn phase, *COVARIATES* none for this model, *RUN_SETTINGS* estimation settings that are called once at the beginning of the estimation (please see S-ADAPT manual for explanation of variables), *RUN_COMMANDS* sequence of commands to be executed by the S-ADAPT command window

predicted concentrations during the first iteration(s). Distribution clearance was set to large values to preserve the beta half-life of the two-compartment model. For PD parameters, baselines, turnover half-lives, and concentrations associated with 50% of maximal effect were set to large values. Maximum extents of inhibition or of stimulation were set to small values. This choice assured that a stimulated rate constant of loss, for example, would not take very large values (e.g., rate constants in excess of 500 h⁻¹). In our experience, this improved the numerical stability of the differential equation solver for estimation via the MC-PEM algorithm.

Variance Burn Phase. All variance terms (and residual error terms) were set to large initial estimates to cover a wide search space. A variance burn period (see companion manuscript for details) of 30 iterations with variances of 1 for log-normally distributed parameters and variances of 4 for logistically transformed parameters was used. The initial estimate for all logistically transformed parameters was set to a mean of 0.01 on transformed scale (equivalent to about 0.50 on untransformed scale). An initial variance of 4 on transformed scale assured that S-ADAPT searched virtually the entire range of possible values during the first 30 iterations for logistically transformed parameters such as *I_{max}*. The variance burn phase is primarily intended to obtain better initial estimates for the population means using a very wide search space for all or for selected model

parameters. As the variance burn phase is switched off after a certain number of iterations, the MC-PEM algorithm can proceed as usual and provide maximum likelihood estimates.

This variance burn phase allows one to perform an initial search from an overview perspective and during this phase the algorithm can improve the initial population means. After iteration 30, the estimator can shrink the large initial variances as needed and focus on the specific portion of the parameter space that contains the maximum likelihood solution. To some degree, the variance burn phase is comparable to the burn-in phase of a full Bayesian analysis.

The present analysis contained a variance burn phase with 30 iterations. For datasets with frequent sampling, 150 iterations (in total) were used, and 200 iterations were used for datasets with sparse sampling. Visual inspection of convergence plots showed that this number of iterations yielded stable estimates for population means and for essentially all cases also for variances.

Shorter Estimation Time via Parallelization. Estimation of the full model with frequent sampling took approximately 129 h for estimation on a single CPU core (in hyper-threaded mode). Parallelized estimation on eight CPU threads in hyper-threaded mode on the same server took approximately 18.7 h. Therefore, parallelization accelerated the estimation 6.9-fold (18.7 vs. 129 h) for eight vs. one thread(s) (86% efficiency) without the limitation of data transfer through the network.

Table III. Average (SD) of Estimated/True Population PKPD Parameter Means and Variances from 30 Bootstrap Replicates for Simultaneous Estimation of the Full Model or of Separate Model Components with Frequent Sampling (“rich data”) and for Simultaneous Estimation of the Full Model with Sparse Data (see “[MATERIALS and METHODS](#)” section)

	Symbol (unit)	True mean or true CV for BSV	Full model “rich data”	Separate models “rich data”	Full model “sparse data”
Two compartment pharmacokinetic model (part 1)					
First-order clearance	CL (L/h)	4	1.08 (0.09)	1.10 (0.11)	1.05 (0.12)
		0.3	0.71 (0.36) ^b	0.69 (0.24) ^b	0.76 (0.56) ^b
Intrinsic clearance of mixed-order elimination	CL _i (L/h)	26	0.83 (0.07)	0.77 (0.07)	0.77 (0.08)
		0.3	0.68 (0.47)	0.79 (0.54)	0.50 (0.42)
Michaelis-Menten constant	K _m (mg/L)	3	1.13 (0.24)	1.21 (0.30)	1.35 (0.31)
		0.3	3.82 (2.10)	4.92 (2.57)	2.61 (2.54)
Distribution clearance	CL _d (L/h)	10	0.96 (0.04)	0.94 (0.05)	0.98 (0.06)
		0.3	1.11 (0.28)	1.17 (0.28)	0.71 (0.38)
Volume of central compartment	V ₁ (L)	10	1.00 (0.05)	1.01 (0.05)	1.02 (0.05)
		0.3	0.89 (0.26)	0.89 (0.26)	0.37 (0.27)
Volume of peripheral compartment	V ₂ (L)	40	0.94 (0.06)	0.92 (0.06)	1.03 (0.06)
		0.3	0.96 (0.27)	0.94 (0.26)	0.89 (0.36)
Indirect response model with inhibition of input (part 2)					
Half-life of loss of response	AT _{12,out} (h)	2	0.99 (0.06)	1.01 (0.06)	1.06 (0.28)
		0.3	0.79 (0.32)	0.82 (0.32)	1.03 (1.02)
Baseline response	A _{Base}	10	1.00 (0.05)	1.00 (0.05)	1.00 (0.05)
		0.3	1.02 (0.23)	1.02 (0.23)	1.02 (0.27)
Maximum extent of inhibition	A _I max	0.731	1.02 (0.05)	1.02 (0.05)	1.07 (0.14)
	A _I max _{Logistic}	1 ^a	0.98 (0.22)	1.00 (0.20)	0.98 (0.45)
IC ₅₀ for inhibition of input	AIC ₅₀ (mg/L)	3	1.11 (0.09)	1.13 (0.09)	1.51 (0.84)
		0.3	0.94 (0.60)	1.01 (0.62)	1.62 (2.15)
Indirect response model with inhibition of loss (part 3)					
Half-life of loss of response	BT _{12,out} (h)	4	1.00 (0.08)	0.99 (0.08)	1.08 (0.31)
		0.3	1.25 (0.63)	1.28 (0.61)	1.14 (1.35)
Baseline response	B _{Base}	10	1.01 (0.05)	1.01 (0.05)	1.00 (0.05)
		0.3	0.94 (0.19)	0.94 (0.18)	0.94 (0.21)
Maximum extent of inhibition	B _I max	0.622	1.02 (0.10)	1.00 (0.10)	1.23 (0.32)
	B _I max _{Logistic}	1 ^a	1.02 (0.33)	1.00 (0.34)	0.91 (0.55)
IC ₅₀ for inhibition of loss	BIC ₅₀ (mg/L)	6	1.12 (0.15)	1.09 (0.15)	2.25 (1.28)
		0.3	0.58 (0.60)	0.53 (0.53)	4.56 (6.98)
Indirect response model with stimulation of input (part 4)					
Half-life of loss of response	CT _{12,out} (h)	6	1.00 (0.05)	1.00 (0.05)	0.94 (0.10)
		0.3	1.18 (0.32)	1.17 (0.32)	0.98 (0.72)
Baseline response	C _{Base}	1	0.98 (0.05)	0.98 (0.05)	1.00 (0.09)
		0.3	0.99 (0.28)	0.97 (0.27)	0.95 (0.29)
Maximum extent of stimulation of input	C _S max	20	1.06 (0.09)	1.05 (0.09)	1.05 (0.15)
		0.3	0.93 (0.31)	0.94 (0.28)	0.69 (0.57)
SC ₅₀ for stimulation of input	CSC ₅₀ (mg/L)	9	1.11 (0.13)	1.11 (0.12)	1.26 (0.43)
		0.3	1.02 (0.76)	0.98 (0.74)	2.38 (2.08)
Indirect response model with stimulation of loss (part 5)					
Half-life of loss of response	DT _{12,out} (h)	8	1.02 (0.05)	1.01 (0.06)	0.99 (0.16)
		0.3	0.99 (0.26)	1.06 (0.27)	1.22 (0.90)
Baseline response	D _{Base}	10	1.01 (0.04)	1.01 (0.04)	1.00 (0.05)
		0.3	1.05 (0.22)	1.02 (0.20)	1.07 (0.24)
Maximum extent of stimulation of loss	D _S max	5	1.05 (0.08)	1.05 (0.07)	1.12 (0.23)
		0.3	0.84 (0.29)	0.89 (0.24)	0.46 (0.43)
SC ₅₀ for stimulation of loss	DSC ₅₀ (mg/L)	15	1.13 (0.11)	1.12 (0.10)	1.32 (0.52)
		0.3	1.06 (0.77)	1.14 (0.72)	3.34 (2.90)
Direct effect model with stimulation of effect (part 6)					
Baseline effect	E _{Base}	10	1.00 (0.05)	1.00 (0.05)	0.99 (0.05)
		0.3	0.95 (0.19)	0.95 (0.19)	0.96 (0.22)
Maximum extent of stimulation	E _E max	3	0.99 (0.05)	0.99 (0.05)	1.02 (0.13)
		0.3	1.08 (0.27)	1.09 (0.28)	0.94 (0.63)
SC ₅₀ for stimulation of effect	EC ₅₀ (mg/L)	15	1.01 (0.07)	1.00 (0.07)	1.09 (0.27)
		0.3	0.94 (0.42)	0.89 (0.39)	1.50 (1.62)
Indirect response model with inhibition of input into precursor compartment (part 7)					
Half-life of loss of response	FT _{12,out} (h)	2	0.95 (0.28)	0.99 (0.27)	0.64 (0.36)
		0.3	1.39 (0.98)	1.39 (1.03)	1.97 (1.48)

Table III. (continued)

	Symbol (unit)	True mean or true CV for BSV	Full model “rich data”	Separate models “rich data”	Full model “sparse data”
Difference in half-life of loss of precursor and of response	$FT_{12,tr}-FT_{12,out}$ (h)	10 0.3	1.01 (0.13) 0.78 (0.51) ^b	0.99 (0.13) 0.79 (0.50) ^b	1.20 (0.26) 1.26 (1.58) ^b
Baseline response	FBase	10 0.3	1.00 (0.05) 1.01 (0.20)	1.00 (0.05) 1.01 (0.20)	1.01 (0.05) 1.02 (0.20)
Maximum extent of inhibition	Flmax Flmax _{Logistic}	0.731 1 ^a	0.99 (0.06) 1.04 (0.33)	0.99 (0.06) 1.07 (0.32)	1.01 (0.15) 1.18 (0.59)
IC ₅₀ for inhibition of input	FIC ₅₀ (mg/L)	1.5 0.3	1.05 (0.18) 0.64 (0.54)	1.07 (0.17) 0.72 (0.63)	1.22 (0.82) 1.58 (2.41)
Indirect response model with inhibition of transfer between precursor and response compartment (part 8)					
Half-life of loss of response	GT _{12,out} (h)	4 0.3	1.01 (0.09) 0.83 (0.42)	1.01 (0.09) 0.76 (0.40)	0.92 (0.19) 0.71 (0.72)
Difference in half-life of loss of precursor and of response	GT _{12,tr} -GT _{12,out} (h)	8 0.3	0.97 (0.16) 1.18 (1.24)	0.96 (0.16) 1.53 (1.33)	1.05 (0.44) 1.25 (1.23)
Baseline response	GBase	10 0.3	1.01 (0.04) 0.99 (0.20)	1.01 (0.05) 0.99 (0.20)	1.01 (0.05) 0.99 (0.20)
Maximum extent of inhibition	GImax GImax _{Logistic}	0.622 1 ^a	1.01 (0.05) 1.12 (0.35)	1.02 (0.05) 1.22 (0.40)	0.97 (0.12) 0.91 (0.66)
IC ₅₀ for inhibition of transfer	GIC ₅₀ (mg/L)	3 0.3	1.08 (0.13) 0.79 (0.67)	1.12 (0.14) 0.83 (0.71)	1.04 (0.33) 1.56 (1.91)
Indirect response model with stimulation of input into precursor compartment (part 9)					
Half-life of loss of response	HT _{12,out} (h)	6 0.3	1.07 (0.22) 0.86 (0.36)	1.06 (0.18) 0.87 (0.36)	0.90 (0.25) 0.67 (0.61)
Difference in half-life of loss of precursor and of response	HT _{12,tr} -HT _{12,out} (h)	6 0.3	0.85 (0.43) 1.39 (1.35)	0.86 (0.34) 1.50 (1.24)	1.14 (0.57) 1.28 (1.16)
Baseline response	HBase	1 0.3	1.00 (0.06) 0.93 (0.22)	1.00 (0.06) 0.93 (0.22)	0.96 (0.07) 1.04 (0.42)
Maximum extent of stimulation	HSmx	20 0.3	1.05 (0.10) 0.89 (0.32)	1.05 (0.10) 0.87 (0.33)	1.04 (0.16) 0.55 (0.39)
SC ₅₀ for stimulation	HSC ₅₀ (mg/L)	4.5 0.3	1.22 (0.24) 1.17 (1.20)	1.24 (0.23) 1.58 (1.30)	1.15 (0.37) 1.94 (1.75)
Indirect response model with stimulation of transfer from precursor to response compartment (part 10)					
Half-life of loss of response	IT _{12,out} (h)	8 0.3	0.96 (0.08) 1.00 (0.33)	0.96 (0.07) 0.96 (0.36)	0.82 (0.13) 0.66 (0.62)
Difference in half-life of loss of precursor and of response	IT _{12,tr} -IT _{12,out} (h)	4 0.3	1.18 (0.18) 0.72 (0.64)	1.21 (0.18) 1.03 (0.65)	1.49 (0.44) 0.58 (0.54)
Baseline response	IBase	10 0.3	0.99 (0.04) 1.04 (0.21)	0.99 (0.04) 1.04 (0.20)	0.99 (0.04) 1.02 (0.20)
Maximum extent of stimulation of transfer	ISmax	5 0.3	0.96 (0.14) 0.78 (0.36)	0.98 (0.17) 0.84 (0.38)	0.81 (0.23) 0.46 (0.86)
SC ₅₀ for stimulation of transfer	ISC ₅₀ (mg/L)	12 0.3	1.03 (0.25) 1.42 (1.08)	1.16 (0.26) 1.53 (0.98)	0.89 (0.44) 1.55 (1.79)
Standard deviations of additive and proportional residual error parameters					
Plasma concentration (DVID 1)	SD _{PK} (mg/L)	0.01	15.69 (2.85)	14.48 (2.19)	8.62 (2.33)
	CV _{PK}	0.1	1.12 (0.11)	1.16 (0.12)	1.95 (0.36)
Indirect response model 1 (DVID 2)	SD _A	0.01	2.53 (1.97)	2.47 (1.99)	7.83 (12.0)
	CV _A	0.1	0.99 (0.03)	0.99 (0.03)	0.95 (0.21)
Indirect response model 2 (DVID 3)	SD _B	0.5	0.43 (0.15)	0.43 (0.15)	0.82 (0.53)
	CV _B	0.1	0.92 (0.06)	0.92 (0.06)	0.86 (0.47)
Indirect response model 3 (DVID 4)	SD _C	0.5	0.83 (0.03)	0.83 (0.03)	0.87 (0.13)
	CV _C	0.1	0.72 (0.04)	0.73 (0.04)	0.54 (0.35)
Indirect response model 4 (DVID 5)	SD _D	0.01	1.70 (1.54)	1.42 (1.60)	2.52 (3.67)
	CV _D	0.1	0.99 (0.04)	0.99 (0.04)	0.98 (0.12)
Direct effect model (DVID 6)	SD _E	0.5	0.37 (0.14)	0.35 (0.14)	0.70 (0.58)
	CV _E	0.1	0.96 (0.06)	0.96 (0.06)	1.04 (0.50)
Precursor indirect response model (DVID 7)	SD _F	0.01	3.69 (2.92)	3.83 (3.10)	6.53 (7.32)
	CV _F	0.1	1.00 (0.05)	1.00 (0.05)	0.93 (0.13)
Precursor indirect response model (DVID 8)	SD _G	0.01	6.44 (4.13)	6.70 (4.05)	12.0 (12.9)
	CV _G	0.1	0.97 (0.05)	0.97 (0.06)	0.93 (0.21)
Precursor indirect response model (DVID 9)	SD _H	0.5	0.83 (0.03)	0.83 (0.03)	0.89 (0.10)
	CV _H	0.1	0.72 (0.06)	0.73 (0.06)	0.61 (0.18)

Table III. (continued)

	Symbol (unit)	True mean or true CV for BSV	Full model “rich data”	Separate models “rich data”	Full model “sparse data”
Precursor indirect response model (DVID 10)	SD _I	0.01	8.79 (6.84)	8.69 (6.60)	16.9 (9.55)
	CV _I	0.1	0.98 (0.06)	1.00 (0.06)	1.07 (0.22)

Initial estimates were different from the true values by a factor of 10

CV Coefficient of variation, BSV Between subject variability

^a This value is the variance on logistically transformed scale. The transformation from $\text{Imax}_{\text{Logistic}}$ to Imax on the untransformed scale was: $\text{Imax} = 1/[1 + \exp(-\text{Imax}_{\text{Logistic}})]$

^b Average and SD of the estimated/true ratios were calculated based on BSV expressed as variances

Parallelization over the network took 20.4 h on eight CPU threads. This yields a 6.3-fold acceleration for eight vs. one thread(s) (79% efficiency). Therefore, transfer of files through the network for each iteration slowed down the estimation by approximately 9% (20.4 vs. 18.7 h). For parallelization on 48 threads through the network, estimation took 5.5 h. This presents an approximately 23-fold acceleration for 48 vs. 1 thread(s) (49% efficiency). For this case, each thread had to calculate the expectation step of one subject in the 48 subject datasets. As the expectation step for one subject took roughly the same time as distributing 48 jobs through the network, parallelization on 48 threads was less efficient.

Estimation of a model with a full var-cov matrix took 7.2 h on 48 threads parallelized through the network (estimation time based on three replicates). This model contained 990 estimated covariances for the 45 random model parameters. Therefore, estimation of the full model with a full var-cov matrix increased estimation times by 30% (7.2 vs. 5.5 h) compared to a model with major diagonal var-cov matrix.

DISCUSSION

During the last 30 years, the field of PK/PD modeling has grown continuously and has become increasingly valuable to

describe, predict and understand biological systems. The availability of more powerful estimation algorithms, software packages, and computers enabled scientists to incorporate more sophisticated mechanisms in PK/PD models for clinical and experimental studies (21–38). Whenever true biological variability between individuals is present and significant, population PK/PD modeling is a powerful concept to describe and predict the behavior of such systems.

Several studies (4,7,20,40–47), Leary *et al.* (Abstract 491, 13th Annual Meeting of the Population Approach Group in Europe, 2004), and Girard and Mentré (Abstract 834, 14th Annual Meeting of the Population Approach Group in Europe, 2005) yielded valuable insights into the performance and robustness of various estimation algorithms for nonlinear mixed-effects modeling of common population PK or population PK/PD models. Most of these studies focused on models with up to moderate complexity, one or two dependent variables, and typically less than 10 structural model parameters with BSV. Therefore, these results may not be directly applicable to mechanistic modeling of complex models with multiple dependent variables, complex differential equations, and a high dimensional space of random model parameters.

Only a few studies (48,49) evaluated the estimability of model parameters for mechanistic models by using a pooled fitting approach. As systematic simulation estimation studies

Table IV. Averages of the Ratio of Estimated Divided by True Population PD Parameters for Simultaneous Estimation of the Full Model with Frequent Sampling

Model type	DVID	Half-life of loss of response		Difference in half-life of loss for precursor and for response		Baseline effect		Maximum extent of inhibition or stimulation		Concentration associated with 50% of maximum effect	
		Mean	BSV	Mean	BSV	Mean	BSV	Mean	BSV	Mean	BSV
Indirect response models	2	0.99	0.79			1.00	1.02	1.02 ^a	0.98	1.11	0.94
	3	1.00	1.25			1.01	0.94	1.02 ^a	1.02	1.12	0.58
	4	1.00	1.18			0.98	0.99	1.06	0.93	1.11	1.02
	5	1.02	0.99			1.01	1.05	1.05	0.84	1.13	1.06
Direct effect model	6					1.00	0.95	0.99	1.08	1.01	0.94
Precursor indirect response models	7	0.95	1.39	1.01	0.78	1.00	1.01	0.99 ^a	1.04	1.05	0.64
	8	1.01	0.83	0.97	1.18	1.01	0.99	1.01 ^a	1.12	1.08	0.79
	9	1.07	0.86	0.85	1.39	1.00	0.93	1.05	0.89	1.22	1.17
	10	0.96	1.00	1.18	0.72	0.99	1.04	0.96	0.78	1.03	1.42

Initial estimates were different from the true values by a factor of 10

BSV variance for between-subject variability, DVID dependent variable type identifier

The individual ratios (estimated/true) for all population PD model parameters included values below and above 1.

^a Refers to the ratio of Imax on linear scale (i.e., not on logistic scale)

Table V. Averages of the Ratio of Estimated Divided by True Population PD Parameters for Simultaneous Estimation of the Full Model with Sparse Sampling

Model type	DVID	Half-life of loss of response		Difference in half-life of loss for precursor and for response		Baseline effect		Maximum extent of inhibition or stimulation		Concentration associated with 50% of maximum effect	
		Mean	BSV	Mean	BSV	Mean	BSV	Mean	BSV	Mean	BSV
Indirect response models	2	1.06	1.03			1.00	1.02	1.07 ^a	0.98	1.51 (median: 1.24)	1.62 (median: 0.53)
	3	1.08	1.14			1.00	0.94	1.23 ^a	0.91	2.25 (median: 1.73)	4.56 (median: 1.81)
	4	0.94	0.98			1.00	0.95	1.05	0.69	1.26	2.38 (median: 1.57)
	5	0.99	1.22			1.00	1.07	1.12	0.46	1.32	3.34 (median: 2.34)
Direct effect model	6					0.99	0.96	1.02	0.94	1.09	1.50
Precursor	7	0.64	1.97	1.20	1.26	1.01	1.02	1.01 ^a	1.18	1.22	1.58
Indirect response models	8	0.92	0.71	1.05	1.25	1.01	0.99	0.97 ^a	0.91	1.04	1.56
	9	0.90	0.67	1.14	1.28	0.96	1.04	1.04	0.55	1.15	1.94
	10	0.82	0.66	1.49	0.58	0.99	1.02	0.81	0.46	0.89	1.55

Initial estimates were different from the true values by a factor of 10

The individual ratios (estimated/true) for all population PD model parameters included values below and above 1

BSV variance for between-subject variability, DVID dependent variable type identifier

^a Refers to the ratio of I_{max} on linear scale (i.e., not on logistic scale)

for mechanistic models using nonlinear mixed-effects modeling were not available in the literature, we systematically assessed the performance of population PK/PD modeling algorithms for models with increased complexity.

Due to a lack of performance data on nonlinear mixed-effects modeling algorithms for complex mechanistic models, one may ask whether parameters for such complex systems can be estimated with small bias and good precision. The “curse of dimensionality” prevents the direct application of the NPEM algorithm, since computation of 10^{45} or even 3^{45} grid points per patient during the first iteration for a model with 45 random parameters is computationally not feasible. Mechanistic models usually need to be specified as a set of nonlinear differential equations which require long computation times. Therefore, a highly parallelizable importance sampling version of the MC-PEM algorithm was systematically evaluated for complex mechanistic models using the robust and efficient differential equation solver (50) implemented in the ADAPT/S-ADAPT package (12,51).

The MC-PEM algorithm contains a parallelizable E-step that usually requires >99% of the computation time (<1% of time required for the maximization step). Gradient search algorithms such as the FOCE method can also be parallelized; however, to our knowledge the fraction of the computations of the FOCE method that can be parallelized (and therefore accelerated) is notably less than 99%. Among the parametric EM algorithms (1–6,13,15,19,52,53), the MC-PEM method requires the fewest number of iterations (often 80–300, depending on the model complexity and data), since the MC-PEM algorithm spends most of its computation time on exploring potential parameter values for each subject per iteration. The MC-PEM algorithm typically uses 1,000–3,000 random samples per subject and iteration for the multi-dimensional integration to compute the conditional means and conditional var–cov matrices.

We sought to explore the applicability of the MC-PEM algorithm for mechanistic models up to a complexity that may become relevant in the not too distant future and evaluated

both the MC-PEM algorithm in S-ADAPT and the standard settings of the SADAPT-TRAN package. A rather low number of 1,000 random samples per subject and iteration (NPOPC=1,000 in S-ADAPT) yielded unbiased and fairly precise estimates for population means and BSV for datasets with frequent sampling (Tables III and IV) despite integration in a 45-dimensional parameter space. Estimates were largely unbiased and reasonably precise for sparse data, but not as good as for frequent sampling (Tables III and V). With only 48 subjects per dataset estimates for BSV terms were less precise but still reasonable for sparse sampling. Evaluation of the bias and precision for sparse data is useful for mechanistic modeling to solve biomedical problems.

For complex models, the importance sampling MC-PEM method with the MAP step during each iteration (pmethod=4 in S-ADAPT; method IMPMAP in NONMEM® 7) was more robust than the importance sampling MC-PEM method (pmethod=8 in S-ADAPT) which employs the MAP step only after the first iteration and then every 15th iteration (depending on the ndelpar setting), for example. Premature termination of estimation with the latter method occurred more often for real data and could not be addressed by increasing the number of Monte Carlo samples to 10,000 (NPOPC) for pmethod 8. In our experience, the MAP Bayesian step at the beginning of each iteration of pmethod 4 improves the robustness of this algorithm, since it assures that the envelope function used to calculate the conditional means and conditional var–cov matrix for every subject is placed at an informative position.

There is an interesting similarity between the importance sampling MC-PEM algorithm as applied here and the NPAG algorithm by Leary *et al.* (14th IEEE Symposium on Computer Based Medical Systems, Bethesda, MD: IEEE Computer Society, 2001: 389–94). The variance burn phase at the beginning of the MC-PEM algorithm seeks to obtain a global overview and optimize the initial guesses for the population means. In its first iteration, the NPAG algorithm obtains such a global perspective by evaluating a full, but

rather coarse grid (depending on the model complexity) over the entire space of potential parameter values (10). In subsequent iterations, the NPAG algorithm employs an adaptive grid strategy by randomly spreading out new daughter points located around the best support points from the previous iteration. The adaptive grid provides a refined resolution compared to the original grid and thus seeks to optimize the parameter values by sampling at more informative positions. This aspect of the NPAG algorithm shares notable similarity with the importance sampling MC-PEM method.

The NPEM and NPAG algorithm use discrete support points to characterize the BSV. Each support point is a set of model parameter values and an associated non-zero probability. The implementation of the NPEM and NPAG algorithm in the MM-USCPACK (7–10) assumes that the residual error is normally distributed for calculation of the log-likelihood. However, the distribution of support points can take any shape (i.e., the BSV is described by a nonparametric distribution). The NPAG algorithm can estimate the residual error variance in a separate step at the end of each iteration of the population analysis.

A potential limitation of the present bootstrap analysis is the relatively low number of 30 replicates per scenario. These replicates yielded very similar results for the means and medians for frequent sampling and reasonably consistent means and medians for sparse data. Only the standard deviation is reported as a measure of dispersion, as reporting of 90% confidence intervals would require several hundred bootstrap replicates. Nevertheless, it could be shown that computation of a complex mechanistic PK/PD model with 48 subjects, 10 dependent variables, 45 population means, 45 BSV terms, and 20 residual error parameters was robust and feasible. One benefit of EM algorithms is that estimation time and robustness is only slightly affected by the complexity of the var-cov matrix. Estimation of a model with full var-cov matrix including 990 covariances was only 30% slower than estimation of the final model with major diagonal var-cov matrix.

The MC-PEM algorithm always yielded standard errors and had no numerical problems for any estimation in this study. This makes the importance sampling MC-PEM method an ideal algorithm to efficiently obtain robust initial estimates for subsequent estimation of mechanistic models via other population methods such as NPAG, stochastic approximation expectation maximization, or a full Bayesian approach via Markov Chain Monte Carlo (4). We propose to derive an initial density matrix for the NPAG algorithm by the importance sampling MC-PEM method in order to support estimation of complex mechanistic models by nonparametric population algorithms.

CONCLUSION

The importance sampling version of the MC-PEM algorithm (pmethod=4 in S-ADAPT) yielded unbiased and adequately precise estimates for a complex mechanistic model with 10 dependent variables, 45 parameter means, and 45 BSV terms for datasets with frequent or sparse sampling despite poor initial estimates (0.1- or 10-fold the true value). One thousand Monte Carlo samples

centered on the individual MAP estimates during each iteration yielded precise estimates despite integration in a 45-dimensional space of random parameters. This simulation estimation study qualified the default settings of the SADAPT-TRAN package for estimation of mechanistic models. The excellent robustness and efficient parallelizability suggest that mechanistic modeling of complex biological systems with variability is feasible and efficient via the importance sampling MC-PEM algorithm.

ACKNOWLEDGMENT

We greatly thank Drs. David Z. D'Argenio, Robert J. Bauer, Nicholas H. G. Holford, Alan Schumitzky, and Chee M. Ng for insightful discussions on EM algorithms and valuable comments for the design and functionality of SADAPT-TRAN. We thank Dr. William J. Jusko for providing a stimulating environment during our post-doctoral fellowships at the University at Buffalo. We thank the three reviewers of this manuscript for very helpful comments. The Perl scripts included in the SADAPT-TRAN package were completely written by JBB since 2003 and do not contain any code from any other software. The work presented here was not supported by a funding source.

Conflicts of Interest None for all authors.

REFERENCES

1. Dempster AP, Laird N, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B*. 1977;39:1–38.
2. Schumitzky A. EM algorithms and two-stage methods in pharmacokinetic population analysis. In: D'Argenio DZ, editor. *Advanced methods of pharmacokinetic and pharmacodynamic systems analysis*. New York: Plenum Press; 1995. p. 145–60.
3. Walker S. An EM algorithm for nonlinear random effects models. *Biometrics*. 1996;52:934–44.
4. Bauer RJ, Guzy S, Ng C. A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. *AAPS J*. 2007;9: E60–83.
5. Wang X, Schumitzky A, D'Argenio DZ. Nonlinear random effects mixture models: maximum likelihood estimation via the EM algorithm. *Comput Stat Data Anal*. 2007;51:6614–23.
6. Wu C-FJ. On the convergence properties of the EM algorithm. *Ann Stat*. 1983;11:95–103.
7. Bustad A, Terziyanov D, Leary R, Port R, Schumitzky A, Jelliffe R. Parametric and nonparametric population methods: their comparative performance in analysing a clinical dataset and two Monte Carlo simulation studies. *Clin Pharmacokinet*. 2006;45:365–83.
8. Jelliffe RW. The USC*PACK PC programs for population pharmacokinetic modeling, modeling of large kinetic/dynamic systems, and adaptive control of drug dosage regimens. *Proc Annu Symp Comput Appl Med Care*. 1991: 922–4.
9. Jelliffe RW, Schumitzky A, Bayard D, Milman M, Van Guilder M, Wang X, *et al*. Model-based, goal-oriented, individualised drug therapy. Linkage of population modelling, new 'multiple model' dosage design, Bayesian feedback and individualised target goals. *Clin Pharmacokinet*. 1998;34:57–77.
10. Schumitzky A. Nonparametric EM algorithms for estimating prior distributions. *Appl Math Comput*. 1991;45:143–57.
11. Laird N. Nonparametric maximum likelihood estimation of a mixing distribution. *J Amer Stat Assn*. 1978;73:805–11.

12. Bauer RJ. S-ADAPT/MCPEM user's guide (Version 1.56). Software for Pharmacokinetic, Pharmacodynamic and Population Data Analysis. Berkeley, CA; 2008
13. D'Argenio DZ, Schumitzky A, Wang X. ADAPT 5 User's Guide: Pharmacokinetic/Pharmacodynamic Systems Analysis Software. Biomedical Simulations Resource. Los Angeles, CA; 2009
14. D'Argenio DZ, Schumitzky A. A program package for simulation and parameter estimation in pharmacokinetic systems. *Comput Programs Biomed*. 1979;9:115–34.
15. Lavielle M. Monolix (Version 3.1)—A free software for the analysis of nonlinear mixed effects models. The Monolix Group. Paris; 2009
16. Beal SL, Sheiner LB, Boeckmann AJ, Bauer RJ. NONMEM user's guides (1989–2009). Ellicott City, MD, USA: Icon Development Solutions; 2009.
17. Smith AFM, Gelfand AE. Bayesian statistics without tears: a sampling-resampling perspective. *Am Stat*. 1992;46:84–8.
18. Bauer RJ, Guzy S. Monte Carlo parametric expectation maximization (MC-PEM) method for analyzing population pharmacokinetic/pharmacodynamic data. In: D'Argenio DZ, editor. *Advanced methods of pharmacokinetic and pharmacodynamic systems analysis*. Boston, MA: Kluwer Academic Publishers; 2003. p. 135–63.
19. Bauer RJ. Technical guide on the population analysis methods in the S-ADAPT program (Version 1.56). Berkeley, CA; 2008
20. Bauer RJ. S-ADAPT/MCPEM Validation of Population Analysis of PK/PD Models. Berkeley, CA; 2006
21. Earp JC, Dubois DC, Molano DS, Pyszczynski NA, Keller CE, Almon RR, *et al*. Modeling corticosteroid effects in a rat model of rheumatoid arthritis I: mechanistic disease progression model for the time course of collagen-induced arthritis in Lewis rats. *J Pharmacol Exp Ther*. 2008;326:532–45.
22. Landersdorfer CB, DuBois DC, Almon RR, Jusko WJ. Mechanism-based modeling of nutritional and leptin influences on growth in normal and type 2 diabetic rats. *J Pharmacol Exp Ther*. 2009;328:644–51.
23. Ait-Oudhia S, Scherrmann JM, Krzyzanski W. Simultaneous pharmacokinetics/pharmacodynamics modeling of recombinant human erythropoietin upon multiple intravenous dosing in rats. *J Pharmacol Exp Ther*. 2010;334:897–910.
24. Bulitta JB, Yang JC, Yohonn L, Ly NS, Brown SV, D'Hondt RE, *et al*. Attenuation of colistin bactericidal activity by high inoculum of *Pseudomonas aeruginosa* characterized by a new mechanism-based population pharmacodynamic model. *Antimicrob Agents Chemother*. 2010;54:2051–62.
25. Mager DE, Jusko WJ. Development of translational pharmacokinetic–pharmacodynamic models. *Clin Pharmacol Ther*. 2008;83:909–12.
26. Bulitta JB, Ly NS, Yang JC, Forrest A, Jusko WJ, Tsuji BT. Development and qualification of a pharmacodynamic model for the pronounced inoculum effect of ceftazidime against *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother*. 2009;53:46–56.
27. Ng CM, Joshi A, Dedrick RL, Garovoy MR, Bauer RJ. Pharmacokinetic–pharmacodynamic efficacy analysis of efalizumab in patients with moderate to severe psoriasis. *Pharm Res*. 2005;22:1088–100.
28. Ng CM, Patnaik A, Beeram M, Lin CC, Takimoto CH. Mechanism-based pharmacokinetic/pharmacodynamic model for troxacitabine-induced neutropenia in cancer patients. *Cancer Chemother Pharmacol*. 2010. doi:10.1007/s00280-010-1393-y.
29. Ng CM, Bai S, Takimoto CH, Tang MT, Tolcher AW. Mechanism-based receptor-binding model to describe the pharmacokinetic and pharmacodynamic of an anti- $\alpha(5)\beta(1)$ integrin monoclonal antibody (volociximab) in cancer patients. *Cancer Chemother Pharmacol*. 2009;65:207–17.
30. Friberg LE, Karlsson MO. Mechanistic models for myelosuppression. *Investig New Drugs*. 2003;21:183–94.
31. Hooker AC, Ten Tije AJ, Carducci MA, Weber J, Garrett-Mayer E, Gelderblom H, *et al*. Population pharmacokinetic model for docetaxel in patients with varying degrees of liver function: incorporating cytochrome P4503A activity measurements. *Clin Pharmacol Ther*. 2008;84:111–8.
32. Bulitta JB, Zhao P, Arnold RD, Kessler DR, Daifuku R, Pratt J, *et al*. Mechanistic population pharmacokinetics of total and unbound paclitaxel for a new nanodroplet formulation *versus* Taxol in cancer patients. *Cancer Chemother Pharmacol*. 2009;63:1049–63.
33. Wiczling P, Lowe P, Pigeolet E, Ludicke F, Balser S, Krzyzanski W. Population pharmacokinetic modelling of filgrastim in healthy adults following intravenous and subcutaneous administrations. *Clin Pharmacokinet*. 2009;48:817–26.
34. Tanos PP, Isbister GK, Lalloo DG, Kirkpatrick CM, Duffull SB. A model for venom-induced consumptive coagulopathy in snake bite. *Toxicol*. 2008;52:769–80.
35. Wajima T, Isbister GK, Duffull SB. A comprehensive model for the humoral coagulation network in humans. *Clin Pharmacol Ther*. 2009;86:290–8.
36. Fung HL, Tsou PS, Bulitta JB, Tran DC, Page NA, Soda D, *et al*. Pharmacokinetics of 1, 4-butanediol in rats: bioactivation to gamma-hydroxybutyric acid, interaction with ethanol, and oral bioavailability. *AAPS J*. 2008;10:56–69.
37. Xu L, Eiseman JL, Egorin MJ, D'Argenio DZ. Physiologically-based pharmacokinetics and molecular pharmacodynamics of 17-(allylamino)-17-demethoxygeldanamycin and its active metabolite in tumor-bearing mice. *J Pharmacokinet Pharmacodyn*. 2003;30:185–219.
38. Marathe A, Peterson MC, Mager DE. Integrated cellular bone homeostasis model for denosumab pharmacodynamics in multiple myeloma patients. *J Pharmacol Exp Ther*. 2008;326:555–62.
39. Beal SL. Ways to fit a PK model with some data below the quantification limit. *J Pharmacokinet Pharmacodyn*. 2001;28:481–504.
40. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetics parameters. I. Michaelis–Menten model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm*. 1980;8:553–71.
41. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters II. Biexponential model and experimental pharmacokinetic data. *J Pharmacokinet Biopharm*. 1981;9:635–51.
42. Steimer JL, Mallet A, Golmard JL, Boisvieux JF. Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the nonlinear mixed-effect model. *Drug Metab Rev*. 1984;15:265–92.
43. Roe DJ. Comparison of population pharmacokinetic modeling methods using simulated data: results from the Population Modeling Workgroup. *Stat Med*. 1997;16:1241–57. discussion 57–62.
44. Duffull SB, Kirkpatrick CM, Green B, Holford NH. Analysis of population pharmacokinetic data using NONMEM and WinBUGS. *J Biopharm Stat*. 2005;15:53–73.
45. Jonsson S, Kjellsson MC, Karlsson MO. Estimating bias in population parameters for some models for repeated measures ordinal data using NONMEM and NL MIXED. *J Pharmacokinet Pharmacodyn*. 2004;31:299–320.
46. Plan EL, Maloney A, Troconiz IF, Karlsson MO. Performance in population models for count data, part I: maximum likelihood approximations. *J Pharmacokinet Pharmacodyn*. 2009;36:353–66.
47. Jonsson EN, Wade JR, Karlsson MO. Nonlinearity detection: advantages of nonlinear mixed-effects modeling. *AAPS PharmSci*. 2000;2:E32.
48. Krzyzanski W, Jusko WJ. Note: caution in use of empirical equations for pharmacodynamic indirect response models. *J Pharmacokinet Biopharm*. 1998;26:735–41.
49. Krzyzanski W, Jusko WJ. Indirect pharmacodynamic models for responses with multicompartmental distribution or polyexponential disposition. *J Pharmacokinet Pharmacodyn*. 2001;28:57–78.
50. Petzold LR. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *Siam J Sci Stat Comput*. 1983;4:136–48.
51. D'Argenio DZ, Schumitzky A. ADAPT II user's guide: pharmacokinetic/pharmacodynamic systems analysis software. Biomedical Simulations Resource. Los Angeles; 2006
52. Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *Ann Stat*. 1999;27:94–128.
53. Kuhn E, Lavielle M. Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM Probab Stat*. 2004;8:115–31.