MULTI-CRITERIA INFERENCE FOR PROCESS MODELS: STRUCTURAL AND PARAMETRIC INFERENCE FOR A STOCHASTIC MODEL OF FELINE HEMATOPOEISIS.

Joel Howard Reynolds[1] & Daniela Golinelli[2]
1 – U. S. Fish & Wildlife Service, 1011 E. Tudor Rd, MS 221, Anchorage, AK 99502
2- RAND Statistics Group, 1700 Main St, Santa Monica, CA 90407

ABSTRACT: Stochastic process models intersect statistical and mechanistic modeling. Their assessment and calibration raise important questions regarding the appropriateness of statistical inference methods when one cannot implicitly assume a correct model specification. The early stages of process modeling focus primarily on structural inference, with parameter inference secondary, exactly the opposite sequencing of traditional statistical inference. A stochastic model of feline hematoepiesis is used to illustrate a new structural inference method based on simultaneous performance on multiple goodness-of-fit criteria, the Pareto Optimal Model Assessment Cycle (POMAC). Multi-criteria optimization allows flexible direct model structure assessment, with parameter inference a byproduct upon achieving model structure adequacy. The example model appears adequate with regard to the selected assessment criteria, in contrast to conclusions from a more classical statistical inference (Catlin 2001). Simulations with the POMAC-based parameter estimates more closely mimic the experimental observations. Differences in parameter inferences from the two approaches are discussed, including biological implications.

KEYWORDS: goodness-of-fit, mechanistic model, model assessment, model specification, model selection, multi-criteria optimization, Pareto optimization, POMAC, stem cells.

INTRODUCTION

Stochastic process models intersect probabilistic and mechanistic modeling. Their mechanistic basis places primacy on structural inference, or model assessment, over parameter inference (Mallows 1998; Reynolds 1999). However, their probabilistic structure supports use of standard statistical inference methods which generally reverse this inference sequence.

Statistical inference methods primarily focus on parameter estimation, followed, if possible, by formal or informal model structure assessment at the selected parameterization (Figure 1). For process models, these methods may severely restrict the information used in the inferences: the proposed model structure $f(y,\theta)$, rather than the modeler, dictates both (i) the relevant parameter inference information in the observations y, via the minimal sufficient statistics t, as well as (ii) what, if any, information remains available for goodness-of-fit assessment via $f(y|t)$ (Sprott 2000). Though specific process characteristics might be identified as key features an acceptable model must be able to reproduce, if they are not formally 'identified' by the proposed model structure they aren't incorporated into either inference stage. In contrast, mechanistic modeling often uses such process characteristics for ad hoc structural assessment.

Given a set of goodness-of-fit criteria, based on the identified process characteristics, structural inference can be viewed as a problem in multi-criteria optimization (Reynolds 1999): does there exist a parameterization allowing the model structure to simultaneously satisfy all criteria? The Pareto Optimal Model Assessment Cycle (POMAC) utilizes a model's joint performance across a vector of criteria to dynamically search out the vector optimum, i.e., Pareto Frontier, in the objective ('criteria') space. This can both reveal the existence of structural deficiencies and provide insight into their sources. Its effectiveness derives from its focus on joint performance across criteria. The underlying evolutionary optimization algorithm allows for broad selection of goodness-of-fit criteria (Reynolds 1997).

POMAC was used for structural and parametric inference of a stochastic simulation

model of feline hematopoeisis. A previous simulation study revealed feasible parameter ranges but did not formally assess the model structure (Abkowitz 1996).

The model structure is adequate. The POMAC-based parameter estimates differ from those in both the original study (Abkowitz 1996) and a more recent estimating equations analysis (Catlin 2001); simulations from the current estimates more closely mimic the experimental observations. Implications for the task of model specification (Mallows 1998) are discussed.

METHODS:

Multi-criteria Structural Inference

Structural inference proceeds by defining multiple goodness-of-fit criteria, each focusing on a different characteristic of the process under investigation (Reynolds 1999). A multi-criteria optimization search is conducted to detect if any parameterizations allow the model structure to simultaneously '*satisfy*' all of the goodness-of-fit criteria. The optimization search explores and returns the model's Pareto Optimal Frontier or Tradeoff Surface.

Pareto or vector optimization is based on the Pareto dominance relation between two vectors: let X = (1, 3, 2), Y = (2, 3, 3), and Z = (2, 4, 1). If the objective is to minimize each of the three component elements, then X *dominates* Y, '$X >_p Y$', all of its elements are at least equal to Y's respective elements and some are smaller ('some better and all no worse'). Neither X nor Z dominates the other as each does better on some components but not all three. If the feasible space consisted of just these three vectors, the Pareto Frontier would be {X, Z}; larger spaces require a dynamic search algorithm to reveal the Pareto Frontier (i.e., Reynolds 1999). Defining the Frontier requires neither commensurable components, specific objective target values, nor quantitative distance measures from those targets (Reynolds 1999), though different optimization algorithms may require these.

For structural assessment, a model's Pareto Frontier is defined as follows. Let {x, y, z} denote a set of, in this case three, observation features, i.e., descriptive statistics, for goodness-of-fit assessment. For each define a distance measure $d_x(x_{Observed}, x_{Simulated})$. Search over the parameter space to find the Pareto Frontier minimizing the objective vector (d_x, d_y, d_z).

Adding thresholds of *acceptance* for each distance or goodness-of-fit measure provides a structural assessment (Figure 2): if no objective vectors in the Pareto Frontier have component distances all within the acceptance thresholds, i.e., no parameterizations allow the model structure to simultaneously satisfy all the criteria, then a deficiency exists, perhaps structural or of another type (Reynolds 1999). Parameter inference makes no sense since the model is inadequate.

If the Pareto Frontier contains vectors that simultaneously satisfy all criteria, then their associated parameterizations are clear candidates for parameter estimates. There may not be a unique 'Pareto optimal' parameterization. No sampling distributions have been specified, so no direct estimates of standard errors are provided. These might be available from bootstrapping the observations and repeating the structural inference, but this has not been investigated.

The POMAC_EVOLVE software (http://faculty.washington.edu/edford/research/software.html) was used to reveal the model's Pareto Frontier. The search software utilized an evolutionary computation algorithm, similar to genetic algorithms, to explore the Pareto Frontier of a general multi-criteria optimization problem (Reynolds 1997).

Application Details

Hematopoiesis

Hematopoiesis is the multistage process by which stem cells specialize into mature blood cells (Golde 1991). At the first stage, hematopoietic stem cells (HSC) can self-renew or differentiate into *progenitor* cells, moving to a second stage of development. Progenitor cells eventually differentiate into mature white or red blood cells or platelets.

Little is known of HSC behavior as they cannot be observed *in vivo*, unlike progenitor cells, so researchers have developed simulation models to investigate hematopoiesis. Insight into hematopoiesis has direct bearing on clinical therapies, e.g., stem cell transplants or gene therapies, especially for cancer (Abkowitz 1997).

Experimental Observations

An experimental method was developed to monitor the progenitor cell population in Safari cats (Abkowitz 1990). Safari cats are offspring of mating between domestic (Eurasian origin) and wild Geoffroy (South African origin) cats, species which evolved independently for 12 million years (O-Brien 1986). The two species have electrophoretically distinct phenotypes of the X chromosome-linked enzyme G6PD, denoted as d G6PD and G G6PD, respectively. Safari cats are generally balanced heterozygotes,

having equal numbers of progenitor cells of each parental genotype (Abkowitz 1995). A stem cell's phenotype is retained during self-renewal and differentiation, so the proportion of progenitor cells of type d G6PD tracks migration of stem cells to the second stage. This proportion can be estimated by sampling n progenitor cells and counting the proportion that are d G6PD cells.

Safari cats were irradiated to kill existing bone marrow cells, then a limited amount of their own marrow cells, harvested prior to radiation, were transplanted back into the animals. Samples were taken every 2-6 weeks, starting 10 weeks after transplant, and % d G6PD monitored. Six subjects were monitored for 2-6 years on each (Figure 3).

A two-compartment stochastic process model

A two-compartment stochastic Markov model was proposed for hematopoiesis (Figure 4) (Abkowitz 1996; Catlin 2001; Golinelli 2001). All stem cell 'decisions' - HSC renewal or differentiation, or progenitor cell development, are modeled as stochastic outcomes. As the first compartment cannot be observed, it is a 'hidden' Markov model (Catlin 2001). Structural and parametric inferences for the model are based on comparisons to experimental observations of progenitor cells in the second compartment (described above).

The model is fully described in (Abkowitz 1996; Catlin 2001). Denote the domestic and Geoffrey-type populations of HSC in the first compartment by $Z(t)=(Z_d(t),Z_G(t))$ and of progenitor cells in the second compartment by $X(t)= (X_d(t),X_G(t))$ at time t. The transition probabilities for a short time interval (t, t+h) are: $P(Z(t+h)=z+1|Z(t)=z) = \lambda zh+o(h)$; $P(Z(t+h)=z-1,X(t+h)=x+1|Z(t)=z,X(t)=x)= \nu zh+ o(h)$; and $P(X(t+h) =x-1|X(t)=s) = \mu xh+o(h)$. The waiting time to the next event is exponentially distributed with rate $r=Z(t)(\lambda+\nu)+X(t)\mu$; the probability the next event is: a birth = $Z(t)(\lambda)/r$, an emigration = $Z(t)(\nu)/r$, a 'death' from the second compartment= $X(t)(\mu)/r$. Apoptosis, death of an HSC, is not incorporated to avoid identification problems with λ. The number of domestic progenitor cells in a simple random sample of size n at time t is given by $C_d(t)\sim Bin(n, X_d(t)/(X_d(t)+ X_G(t))$.

Goodness-of-fit Features & Distance Measures

Consideration of the experimental process and observations (Figure 3) suggested five distinct observation features, each motivating a specific descriptive statistic and goodness of fit measure (Abkowitz 1996): (i) an early phase of substantial variability in % Domestic, i.e., % of clones of d G6PD type, (ii) a fairly wide range in % Domestic across the observation period, (iii) relative variation in % Domestic during first 15 observations following week 10, (iv) relative variation in % Domestic during the last 15 observations, and (v) maintenance of both cell lineages throughout the observation period (328 weeks). Features 1 and 3 begin on week 10 after transplantation 'to assure that contribution of stem cells, rather than more differentiated cells also present in the marrow inoculum, were assayed' (Abkowitz 1996). Each features' summary statistic and goodness-of-fit measure are defined below.

Duration of early variability: time after transplant when variation in % Domestic subsides. A score test of homogeneity in the observed percent of d G6PD clones from week j to the end of observation is calculated using a ratio estimator for the mean percent d G6PD over the period. Initially all observations are used (j = 10), then the first observation is dropped (j = 14), then the first two observations are dropped, etc., until the test returns a p-value > 0.05. The associated week is considered the week of subsidence. $d_x(x_{Observed}, x_{Simulated})$ = the p-value from a Kolmogorov-Smirnov ('KS') two-sample test of distribution equality (Conover 1999) comparing the six observed duration values, $x_{Observed}$, to 200 simulated values from the current model parameterization, $x_{Simulated}$.

Range of % Domestic: range of % Domestic across the observation period. $d_x(x_{Observed}, x_{Simulated})$ = the p-value from a KS two-sample test of distribution equality comparing the six observed ranges to 200 simulated values from the current model parameterization.

Relative variation in early phase: Pearson's binomial goodness of fit statistic, a score test of homogeneity in observed % Domestic, was calculated for the first 15 observations following week 10. $d_x(x_{Observed}, x_{Simulated})$ = the p-value from a KS two-sample test of distribution equality comparing the six observed values to 200 simulated values from the current model parameterization.

Relative variation in late phase: same as *Relative variation in early phase* but over the last 15 observations.

Extinction of cell lineages: The percentage of 200 realizations in which one or both cell lineages went extinct. No experimental animals had lineage extinctions (Abkowitz 1996), so $d_x(x_{Observed}, x_{Simulated})$ = % $Extinct_{Simulated}$.

Acceptance Thresholds for Structural Inference

Each distance measure has an associated target value of 0; i.e., the goal of the Pareto search is to minimize all components of the distance vector ($d_{VarDuration}$, d_{Range}, $d_{EarlyVar}$, $d_{LateVar}$, $d_{Extinction}$). Acceptance thresholds were defined for each goodness-of-fit distance measure to judge model structure adequacy. For distance measures using KS tests, i.e., features 1 – 4, the acceptance threshold was p-value Ţ [0.1, 1.0]; for Extinction, the acceptance threshold was % $Extinct_{Simulated}$ ħ 10%.

Pareto Optimization Search details

Pareto optimization was conducted over the feasible parameter space (Table 1) defined by (Abkowitz 1996). The initial number of HSC was treated as a parameter, assuming equal numbers of both phenotypes. The evolutionary computation algorithm searched for 1000 generations, using a population size of 100 parameterizations each generation. Algorithm details are in Ford et al. (2000). The search took 5 days on a DellŞ Latitude C840 laptop with 2 GHz PentiumŞ 4 running Windows XPŞ.

RESULTS

The evolutionary computation search revealed 1850 parameterizations on the Pareto Frontier, trading off in there performance among the five goodness-of-fit features (Figure 5). Nineteen parameterizations allowed the model structure to simultaneously 'satisfy' all five acceptance thresholds (Table 2). Thus the model structure is acceptable with regards to this set of features, distance measures, and thresholds.

DISCUSSION

Structural inference using POMAC showed that the model structure was acceptable with regards to the selected features. Following more traditional statistical inference sequencing by assessing the model structure strictly at the parameterization identified using estimating equations (Catlin 2001) suggests the structure is inadequate (Table 2). In the context of inference for process models, this highlights that (i) parameter inference methods which condition on an acceptable model structure may not be utilizing all the relevant process information in the data, and (ii) restricting structural inference to performance at these particular parameterizations may be misleading.

While developed independently, POMAC formalizes the ad hoc parameter inference approach of Abkowitz (1996). Not surprisingly, POMAC's more rigorous search improved on their parameter estimates (Table 2) with regards to the selected goodness-of-fit measures. The POMAC parameter inferences more closely mimic the observed process (Table 2) than the inference using estimating equations (Catlin 2001), even given that only six experimental units were available for the KS tests used as distance measures. Importantly, the POMAC parameter inferences imply rather different biological rates (Table 3).

Statistical methodology has focused primarily on parametric inference, with structural inference receiving secondary attention, if at all, only after parameter inference (Lehmann 1990; Mallows 1998). This may be due, in part, to the methodological power and influence of Fisher (1922), which acknowledged the problem of structural inference, or model specification, only briefly while passing on to parameter inference (Mallows 1998). Fundamentally, they are very different problems: conditioning on a proposed model structure makes parameter inference deductive while assessing the adequacy of that model structure is inductive. "... [I]nductive reasoning is more strict than is deductive reasoning, since in the latter any item of the data may be ignored, and valid inferences may be drawn from the rest; … whereas in inductive inference the whole of the data must be taken into account," (Fisher 1955 p 77).

Structural inference is increasingly acknowledged in the context of empirical models (Chatfield 1995) in terms of model specification (e.g., Burnham 1998) or predictive methods of handling model structure uncertainty (e.g., Raftery 1997). However, inferential tools for mechanistic models remain limited in both availability and usage.

POMAC allows broader usage of process information not readily captured by the sufficient statistics. Its central focus on joint performance across multiple goodness-of-fit criteria provides an effective, formalized means of assessing a proposed model structure or comparing competing model structures (Reynolds 1999). It provides more appropriate inference for process models.

REFERENCES

Abkowitz, J. L. (1995). "The behavior of hematopoietic stem cells in a large animal." Proceedings of the National Academy of Science **92**: 2031-2035.

Abkowitz, J. L., Catlin, Sandra N., Guttorp, Peter (1996). "Evidence that hematopoiesis may

be a stochastic process *in vivo*." Nature medicine **2**(2): 190-197.

Abkowitz, J. L., Catlin, Sandra N., Guttorp, Peter (1997). "Strategies for hematopoietic stem cell gene therapy: insights from computer simulation studies." Blood **89**(9): 3192-3198.

Abkowitz, J. L., Linenberger, M., Newton, M. A., Shelton, G. H., Ott, R. L., Guttorp, Peter (1990). "Evidence for maintenance of hematopoiesis in a large animal by the sequential activiation of stem cell clones." Proceedings of the National Academy of Science **87**: 9062-9066.

Burnham, K. P., Anderson, D. R. (1998). Model selection and inference: a practical information-theoretic approach. New York, Springer.

Catlin, S. N., Abkowitz, Janis L., Guttorp, Peter (2001). "Statistical Inference in a two-compartment model for hematopoiesis." Biometrics **57**: 546-553.

Chatfield, C. (1995). "Model uncertainty, data mining and statistical inference (with discussion)." Journal of the Royal Statistical Society, Series A **158**: 419-466.

Conover, W. J. (1999). Practical Nonparametric Statistics. New York City, Wiley & Sons.

Diggle, P. J., Gratton, R. J. (1984). "Monte Carlo methods of inference for implicit statistical models (with discussion)." Journal of the Royal Statistical Society, Series B **46**: 193-227.

Fisher, R. A. (1922). "On the mathematical foundations of theoreticl statistics." Philosophical Transaction of the Royal Society, London, Series A **222**: 309-368.

Fisher, R. A. (1955). "Statistical methods and scientific induction."

Golde, D. W. (1991). "The stem cell." Scientific American **265**(6): 86-93.

Golinelli, D. (2001). Bayesian Inference in Hidden Stochastic Population Processes. Statistics. Seattle, WA, University of Washington**:** 154.

Lehmann, E. (1990). "Model Specification: the views of Fisher and Neyman, and later developments." Statistical Science **5**(2): 160-168.

Mallows, C. (1998). "The Zeroth Problem." The American Statistician **52**: 1-9.

O-Brien, S. J. (1986). "Molecular genetics in the domestic cat and its relatives." Trends in Genetics **2**: 137-142.

Raftery, A. E., Madigan, D., Hoeting, J. A. (1997). "Bayesian model averaging for linear regression models." Journal of the American Statistical Association **92**: 179-191.

Reynolds, J. (1997). Multi-criteria assessment for ecological process models using pareto optimization. Quantitative Ecology & Resource Management. Seattle, University of Washington**:** 234.

Reynolds, J., Ford, E. David (1999). "Multi-criteria assessment of ecological process models." Ecology **80**: 538-553.

Sprott, D. A. (2000). Statistical Inference in Science. New York, Springer Verlag.

FIGURES

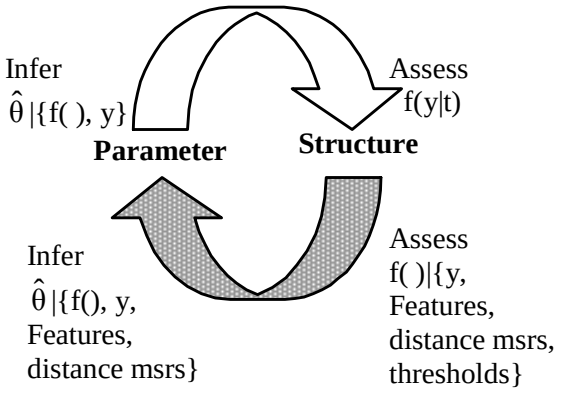

Figure 1. Structural and parametric inference sequences and information sources. Top - classical statistics assumes the model structure, f(y, θ) is correct, deriving estimators for the unknown parameters θ based on minimal sufficient statistics t summarizing the relevant information in the observations y. Remaining sample information may allow for formal goodness-of-fit testing of the proposed model structure via investigation of f(y|t) (Sprott 2000). Bottom – POMAC first assesses the adequacy of model structure f() with respect to {y, the features chosen as goodness-of-fit criteria, their

associated distance measures and thresholds}. If f() is adequate, then parameter inference occurs.

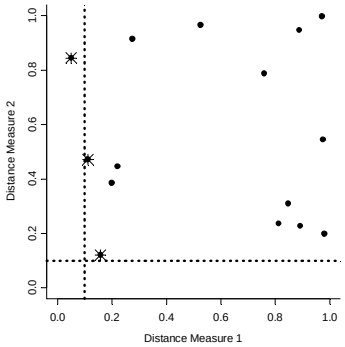

Figure 2. If the dots are the two-component goodness-of-fit vectors (d_x, d_y) for 15 different model parameterizations assessed with regards to their ability to reproduce process features x and y, and the objective is to simultaneously minimize both distance measures dx, dy, then the Pareto Frontier consists the parameterizations associated with the points denoted '*'. If each distance's acceptance threshold is d_i Ɽ [0,0.1] (dashed lines), then the model structure is inadequate as no parameterization allows the model to simultaneously satisfy both criteria.

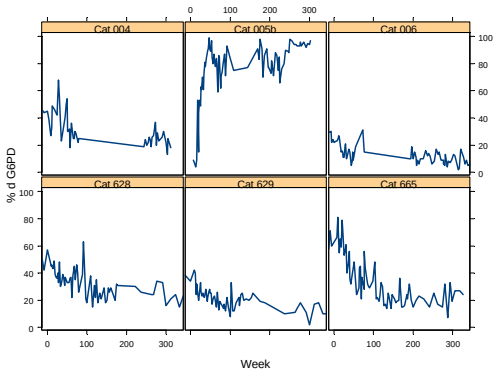

Figure 3. Observed sequences of % progenitor cells with d G6PD phenotypes ('% Domestic') following experimental treatment for 6 feline subjects (Abkowitz et al 1996; for treatment details see ???).

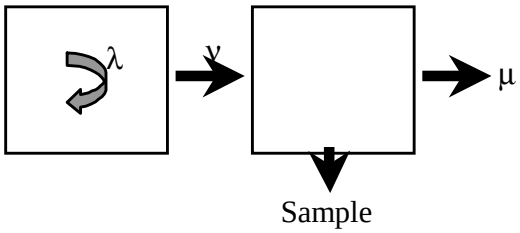

Figure 4. Schematic of two-compartment hematopoeisis stochastic model, with HSC (left compartment) renewal rate λ, rate of differentiation to progenitor compartment ν, and progenitor (right compartment) differentiation rate μ. Model details are summarized in the text.

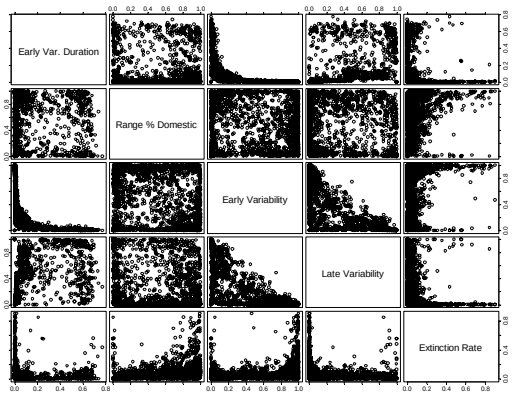

Figure 5. Pairwise projection scatterplots of the Pareto Frontier revealed by the evolutionary computation optimization. The objective was to maximize the first four components (KS p values) while minimizing the last (extinction rate). 'Acceptable' simulations have values of the first four components Í 0.10 and ħ 0.10 for the last. See Table 2.

Figure 6. Realizations of Pareto Optimal parameterization (row X of Table 2) (top) versus realizations using estimating equation-derived parameterization (row Z, Table 2) (Catlin 2001).

Table 1. Parameter search ranges (Abkowitz 1996) and minimum step size for evolutionary optimization search algorithm (Reynolds 1997). See Figure 4 and text for model and parameter description.

Parameter	$Z_d(0)=Z_G(0)$	ν	λ/ν	ν/μ
Minimum	5	0.0167	1.1	0.1
Maximum	50	0.35	2.0	2.0
Minimum Step Size	2	0.02	0.05	0.1

Table 2. Pareto Frontier parameterizations that meet the acceptance thresholds (rows 1 – 19) compared to the parameterization identified in (Abkowitz 1996), 'NM', and the closest parameterization estimate from (Catlin 2001). Parameters (columns 1 – 6, see Figure 4): $Z_d(0)$ – initial number of each type of HSC, ν – HSC emigration rate, λ – HSC renewal rate, μ – progenitor commitment rate. Goodness-of-fit distance measures (columns 7 – 11, see text): Duration – p-value from KS test of duration of early variation phase, Range – p-value from KS test of range of % Domestic, Early Variance – p-value from KS test of early phase variance of % Domestic, Late Variance – p-value from KS test of late phase variance of % Domestic, Extinct – number of 200 simulations that had one or both HSC types go extinct. The acceptance thresholds were p-value Ɍ [0.10, 1.00] and ħ 20 extinctions.

$Z_d(0)$	ν	λ/ν	λ/μ	λ	μ	Duration	Range	Early Variance	Late Variance	Extinct
5	0.057	1.85	0.5	0.1054	0.1140	0.1324	0.4998	0.1814	0.384	9
5	0.057	1.85	0.6	0.1054	0.0950	0.1067	0.4954	0.1849	0.622	17
5	0.057	1.8	0.7	0.1026	0.0814	0.1113	0.7793	0.1192	0.6855	20
5	0.057	1.95	0.5	0.1112	0.1140	0.1381	0.6360	0.1352	0.3926	12
5	0.057	1.95	0.6	0.1112	0.0950	0.1296	0.7204	0.1189	0.4937	12
5	0.057	2.00	0.6	0.1140	0.0950	0.1485	0.7513	0.1262	0.5518	17
7	0.057	1.65	1.0	0.0940	0.0570	0.1010	0.6429	0.1211	0.6589	18
7	0.057	1.70	1.5	0.0969	0.0380	0.1243	0.2813	0.1942	0.5504	9
7	0.057	1.85	0.3	0.1054	0.1900	0.2992	0.2144	0.1246	0.3466	13
7	0.057	1.90	0.5	0.1083	0.1140	0.1091	0.6865	0.1114	0.6200	4
7	0.057	1.90	0.7	0.1083	0.0814	0.1037	0.8178	0.2211	0.4316	10
7	0.057	1.95	0.5	0.1112	0.1140	0.1465	0.8672	0.1159	0.4412	3
7	0.077	1.90	0.2	0.1463	0.3850	0.1986	0.1162	0.1114	0.1913	4
9	0.057	1.55	0.7	0.0884	0.0814	0.1306	0.9468	0.1306	0.4316	8
9	0.057	1.85	0.5	0.1054	0.1140	0.1512	0.8276	0.1277	0.4222	1
9	0.057	1.85	0.8	0.1054	0.0712	0.1205	0.4906	0.1930	0.8101	2
9	0.057	1.95	0.9	0.1112	0.0633	0.1189	0.2012	0.1885	0.7576	1
11	0.057	1.90	0.5	0.1083	0.1140	0.2126	0.6826	0.1061	0.6935	1
11	0.057	1.95	0.5	0.1112	0.1140	0.1325	0.4571	0.1352	0.7244	0
NM 15	0.08	1.25	0.533	0.1000	0.1493	0.02	0.83	0.11	0.34	6
EE 15	0.1676	1.15	0.565	0.1927	0.3411	0.006	0.74	0.29	0.003	39

Table 3. Hematopoiesis process rates (weeks per event) according to parameter inference using POMAC, an ad hoc simulation study (Abkowitz 1996), and estimating equations (Catlin 2001). The simulation-based inferences, which more closely mimic the observations (Table 1) suggest much slower biological processes.

Source	$Z_d(0)$	HSC Renewal	HSC Emigration	Progenitor Committment
POMAC	5	9.49	17.54	8.77
POMAC	11	8.99	17.54	8.77
Abkowitz (1996)	15	10	12.5	6.70
Catlin (2001)	15	5.19	5.96	2.93