

Population Analysis
Simulation
Automation
Study Design ...

Exploring the WinSAAM Population Modeling service
... the EMSA system

Ray Boston 1992 / 2012

A. Lyne a, R. Boston b, K. Pettigrew c and L. Zech d, 1992: EMSA: a SAAM service for the estimation of population parameters based on model fits to identically replicated experiments. *Computer Methods and Programs in Biomedicine*, 38 (1992) 117-151 117.

- Creating the EMSA System
- A Simulation-based Investigation of the WinSAAM Population Modeling Service

Creating the WinSAAM Population Modeling Service

The EMSA project

Given kinetic data from a **stratified study population** develop a method to:

1. automate the estimation of the **population's** kinetic parameters
2. automate the estimation of the population's parameter **covariances** for adjustables and for the dependencies **without** the need to exit WinSAAM and without **multiple** data handling

Solving this problem:

took, **9 NIH**, and allied faculty, spanning **4 institutions over 30 years** (1968 – 1998)
explored, **4 distinct** procedure **implementations** (Boston (2), Lyne, and Greif)
reviewed, **5 fundamentally different** estimation procedures

The estimation problem:

Solve the equations

$$\hat{\mu} = \left[\sum_{j=1}^N (V_j + \hat{\Sigma})^{-1} \right]^{-1} \sum_{j=1}^N (V_j + \hat{\Sigma})^{-1} x_j \quad (1)$$

$$\text{and} \quad \sum_{j=1}^N \zeta_j + \sum_{j=1}^N \zeta_j (x_j - \hat{\mu})(x_j - \hat{\mu}) \zeta_j = 0 \quad (2)$$

where $\zeta_j^{-1} = (V_j + \hat{\Sigma})$ for $\hat{\mu}$ and $\hat{\Sigma}$ respectively, **mean** and **sd**

The difficulties

Joint solutions to eqns (1) and (2) may **not** automatically qualify as either population estimates or errors of estimates

Errors must be **positive**, abs(correlations) must be $\leq '1'$

Physiologic parameters should be **positive**

Methods including: **Penalty functions**, **Barrier functions**, **Reduced gradient**, the **Augmented Lagrangian**, and the **Projected Lagrangian** were all considered and **rejected** due to their susceptibility to convergence failure

The finally accepted method was the **Two Stage Solution** Procedure similar to the EM algorithm of Nairn et al.

Variations included **smart initialization** of covariance matrix elements estimates and **vectorization** of the matrices to simplify the development, implementation and testing

A Simulation-based Investigation of the WinSAAM Population Modeling Service

A Current Trend in Statistical Methodologies

To understand the implications of:

different **degrees** of **variability** or
different **patterns** of **variability**

in our **observations** in regard to the **population size** needed for an investigation
our **best** and most **expedient** approach is with recourse to **simulation** methods

The steps here are as follows:

1. Set the degree and pattern of **data error***
2. Set the number and times of our **observations**#
3. Set the number of subject **responses** to be analyzed#
4. Create the **data** and **model files** for analysis
5. Analyze the **data** and **examine** the resolution of key study indices
6. Refine and **repeat** the analysis if important response features are not observed

* based on **literature** information or pilot data

susceptible to **manipulation** based on inadequacy of 'current structure'

This is the **commentary** at the Start of our 'do file' to generate the test data

```
// local nstudy=10 typically, but 10 to 50, say
// local a~N(0.1,0.05) ... y=A.exp(-a.t)+B.exp(-b.t)+e
// local b~N(0.01,0.005)
// local A~N(90,2)
// local B~N(10,0.5)
// local e~N(0,2.0) Observation error
// time values are stored in a file 'timevalues.dta'
/*
Steps are as follows:
1. Set number of studies `nstudy'
2. Generate the simulation parameters
2. Read the file containing the observation times
3. Generate the observations .. model+error
4. Output the observations and close the file
*/
```

~ => distributed as
 $A \sim N(x,y)$ means that
 $E(A) = 'x'$
 $var(A) = 'y'$
and that A is normally
distributed

Set project work area, initialize study variables, start the study automation loop

```
cd "c:\talks\winsaam meeting 2012"  
cd "Project"  
tempname fw  
local nstudy=30 // number of subjects  
local pct_err=20 // data and model error  
set seed 12347129
```

Setup some 'key' variables

```
forval i=1/\`nstudy' {  
  local name="St_"+string(`i')+ "_subs_"+string(`nstudy')  
  cap conf file "`name'.saam"  
  if _rc==0 erase "`name'.saam"
```

Start our data generation loop

```
file open `fw' using "`name'.saam", write text  
file write `fw' "A SAAM31" _col(27) "`name'" _n  
file write `fw' "2" _col(9) "12" _n  
file write `fw' "H PAR" _n
```

Define and start using data files

```
local sda=0.2*\`pct_err'/100  
local sdb=0.01*\`pct_err'/100  
local sdaA=90*\`pct_err'/100  
local sdbB=10*\`pct_err'/100
```

Define the spread of the params.

```
scalar a=rnormal(0.2, `sda')  
scalar b=rnormal(0.01, `sdb')  
scalar A=rnormal(90, `sdaA')  
scalar B=rnormal(10, `sdbB')  
scalar sdd=`pct_err'/100
```

Generate random parameter values

Write parameters, and functions. Retrieve observation times, generate and write data

```
file write `fw' _col(4) "P(1)" _col(13) "0.1" _col(42) "1000" _n
file write `fw' _col(4) "P(3)" _col(13) "90" _col(42) "1000" _n
file write `fw' _col(4) "P(2)" _col(13) ".01" _col(42) "1000" _n
file write `fw' _col(4) "P(4)" _col(13) "10" _col(42) "1000" _n
```

Write the H PAR contents

```
file write `fw' "H DAT" _n
file write `fw' "X G(1)=P(3)*exp(-P(1)*T)+P(4)*EXP(-P(2)*T)" _n
file write `fw' "101" _col(5) "G(1)" _col(42) "FSD=0.3" _n
```

Write items under H DAT

```
use time_points, clear
sort t
gen q1=A*exp(-a*t) + B*exp(-b*t) // +rnormal(0, sdd)
gen qo=q1*(1+sdd*rnormal(0,1))
di "A: `A' a: `a' B: `B' b: `b' " Retrieve observation times and generate data
list
```

```
local n=_N
forval n=1/`n' { Write the data to a WinSAAM model file
    local obs=qo[`n']
    local time=t[`n']
    file write `fw' _col(13) %6.1f (`time') _col(27) %7.3f (`obs') _n
}
```

```
file close `fw'
}
```

Close the file

The contents of the file 'time_points.dta' ... test set

```
. use time_points
```

```
. list
```

```
      +-----+  
      |     t  |  
      |-----|  
  1.  |     0  |  
  2.  |     1  |  
  3.  |     2  |  
  4.  |     4  |  
  5.  |     6  |  
      |-----|  
  6.  |     8  |  
  7.  |    16  |  
  8.  |    24  |  
  9.  |    32  |  
 10.  |    40  |  
      |-----|  
 11.  |    64  |  
 12.  |    80  |  
 13.  |   120  |  
 14.  |   160  |  
 15.  |   200  |  
      +-----+
```

A **typical** WinSAAM file created by our file management service

BLOCK: 1

```
A SAAM31                               St_1_Sub30
2      12
H PAR
```

BLOCK: 2

```
P(1)    0.1      1000
P(3)    90       1000
P(2)    .01      1000
P(4)    10       1000
```

BLOCK: 3

```
H DAT
X G(1)=P(3)*exp(-P(1)*T)+P(4)*EXP(-P(2)*T)
101 G(1)                                     FSD=0.3
```

BLOCK: 4

```
0.0      86.047
1.0      72.054
2.0      70.798
4.0      61.823
6.0      26.814
8.0      24.924
16.0     13.591
24.0     10.623
32.0      9.216
40.0     11.173
64.0      6.367
80.0      8.608
120.0     4.297
160.0     3.713
200.0     3.683
```

Part of the Population Analysis Processing ... Fitting our simulated data

```
> saam
* DECK BEING PROCESSED
*** L(0,1) ADDED AS A DUMMY.*3*

CONVERGENCE MEASURES
  IMPROVEMENT IN SUM OF SQUARES =  76.77(%)
  FINAL VALUE OF CONAB =  1.019E+00
  LARGEST CHANGE (  64.31 %) WAS IN PAR( 2, 0)

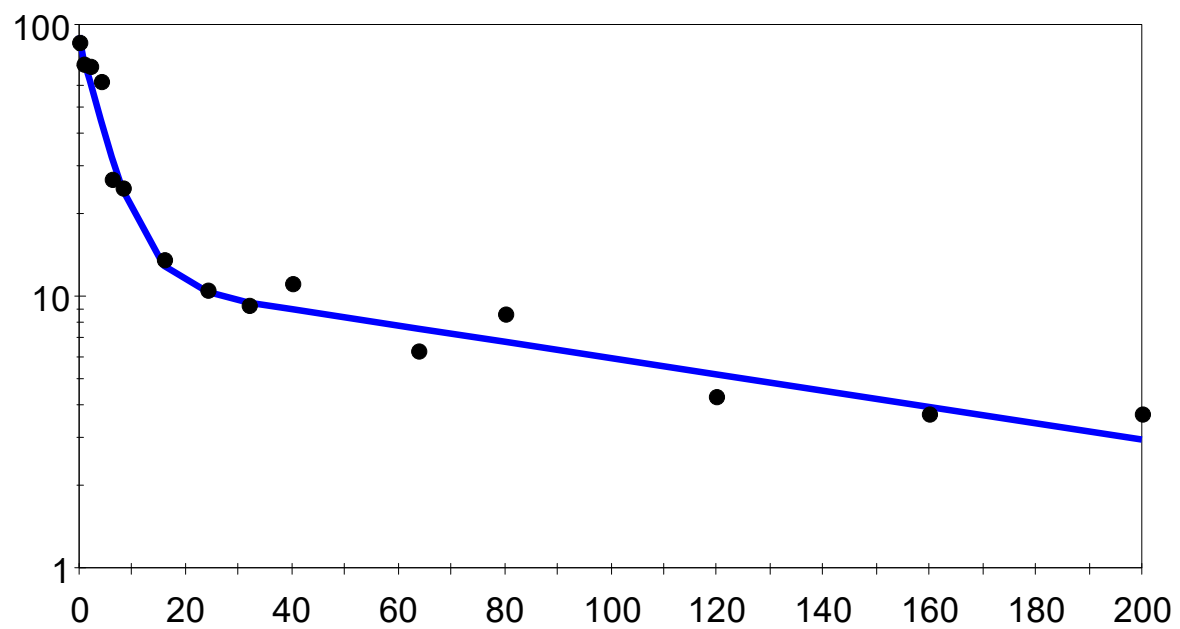
CONVERGENCE MEASURES
  IMPROVEMENT IN SUM OF SQUARES =  35.24(%)
  FINAL VALUE OF CONAB =  6.489E-01
  LARGEST CHANGE (  26.33 %) WAS IN PAR( 2, 0)

CONVERGENCE MEASURES
  IMPROVEMENT IN SUM OF SQUARES =  50.48(%)
  FINAL VALUE OF CONAB =  1.407E+00
  LARGEST CHANGE (  35.76 %) WAS IN PAR( 1, 0)

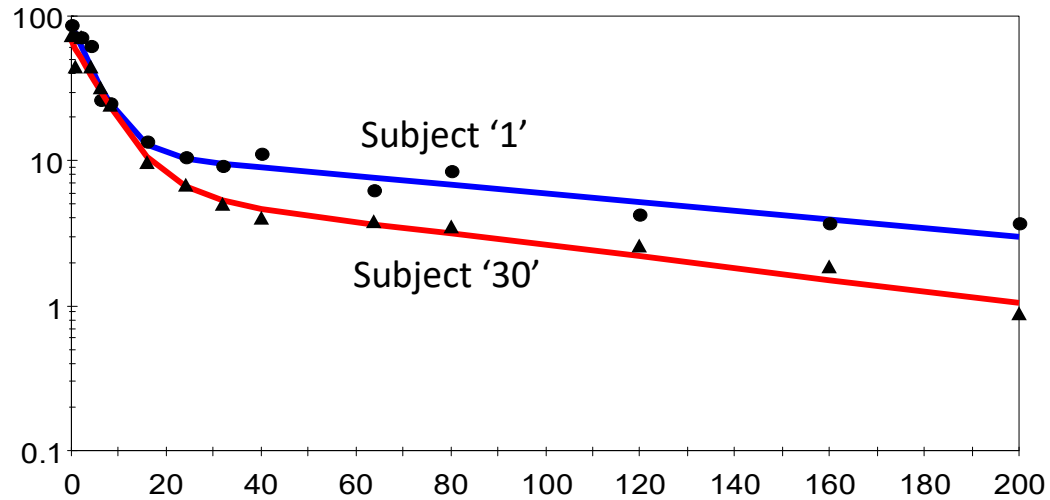
...

PARAMETER    VALUE      ERROR      FSD
P ( 1, 0)  2.185E-01  3.849E-02  1.761E-01
P ( 3, 0)  7.882E+01  1.076E+01  1.365E-01
P ( 2, 0)  6.877E-03  1.160E-03  1.687E-01
P ( 4, 0)  1.174E+01  1.436E+00  1.223E-01
CORRELATION MATRIX
  COLUMN  1    2    3    4
ROW 1  1.00  0.62  0.34  0.42
ROW 2  0.62  1.00  0.00  0.01
ROW 3  0.34  0.00  1.00  0.85
ROW 4  0.42  0.01  0.85  1.00
```

Final fit to Subject '1's observations .. 6 iterations



Analysis of simulated data for studies 1 & 30. The population analysis step invokes the fitting of all studies assembled into the population



Subject '1'

PARAMETER	VALUE	ERROR	FSD
P (1, 0)	2.185E-01	3.849E-02	1.761E-01
P (3, 0)	7.882E+01	1.076E+01	1.365E-01
P (2, 0)	6.877E-03	1.160E-03	1.687E-01
P (4, 0)	1.174E+01	1.436E+00	1.223E-01

CORRELATION MATRIX

COLUMN	1	2	3	4
ROW 1	1.00	0.62	0.34	0.42
ROW 2	0.62	1.00	0.00	0.01
ROW 3	0.34	0.00	1.00	0.85
ROW 4	0.42	0.01	0.85	1.00

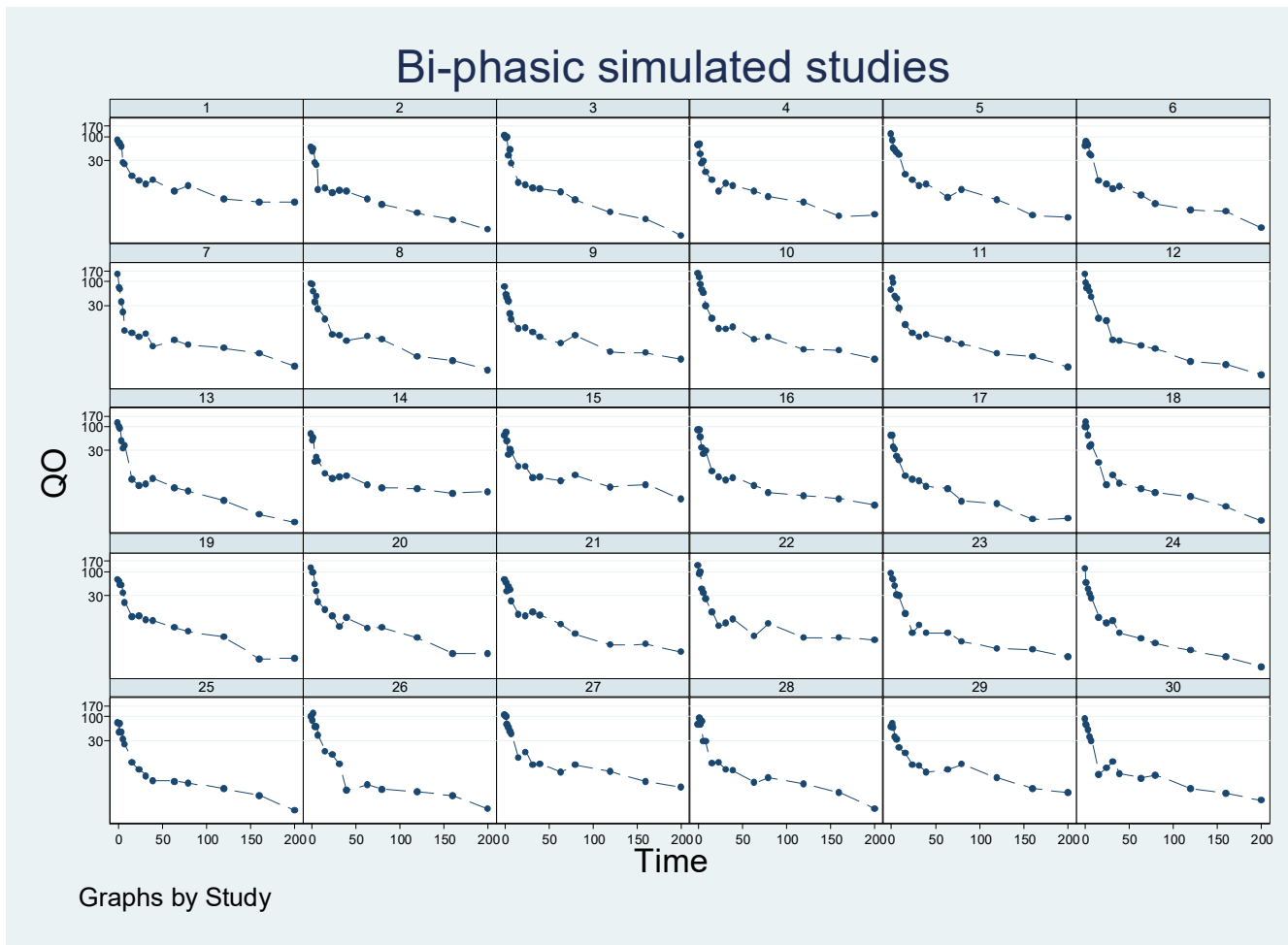
Subject '30'

PARAMETER	VALUE	ERROR	FSD
P (1, 0)	1.546E-01	2.146E-02	1.388E-01
P (3, 0)	6.113E+01	6.808E+00	1.114E-01
P (2, 0)	9.062E-03	1.159E-03	1.279E-01
P (4, 0)	6.480E+00	9.189E-01	1.418E-01

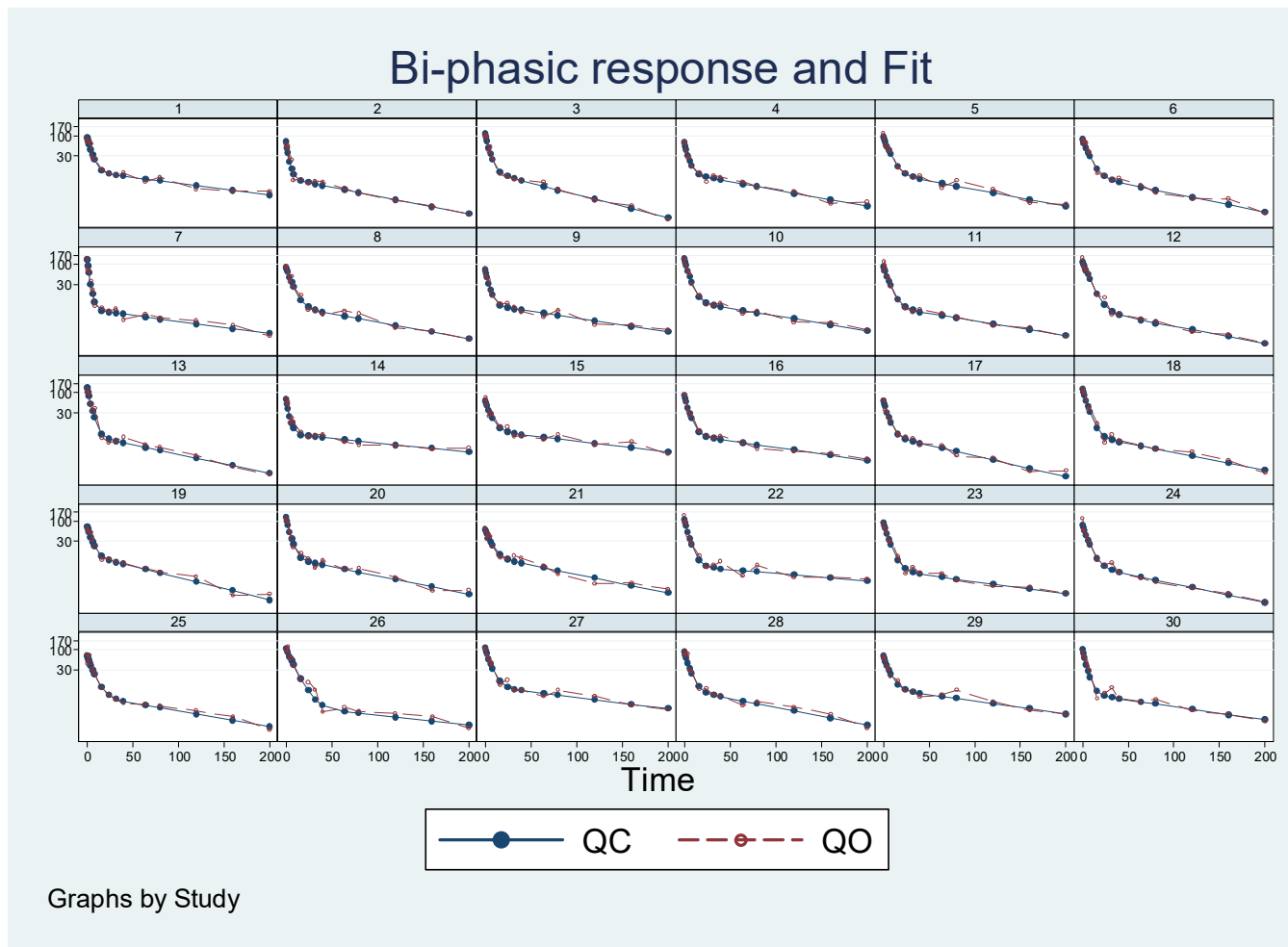
CORRELATION MATRIX

COLUMN	1	2	3	4
ROW 1	1.00	0.58	0.45	0.57
ROW 2	0.58	1.00	0.06	0.10
ROW 3	0.45	0.06	1.00	0.86
ROW 4	0.57	0.10	0.86	1.00

The raw data for the 30 simulated studies

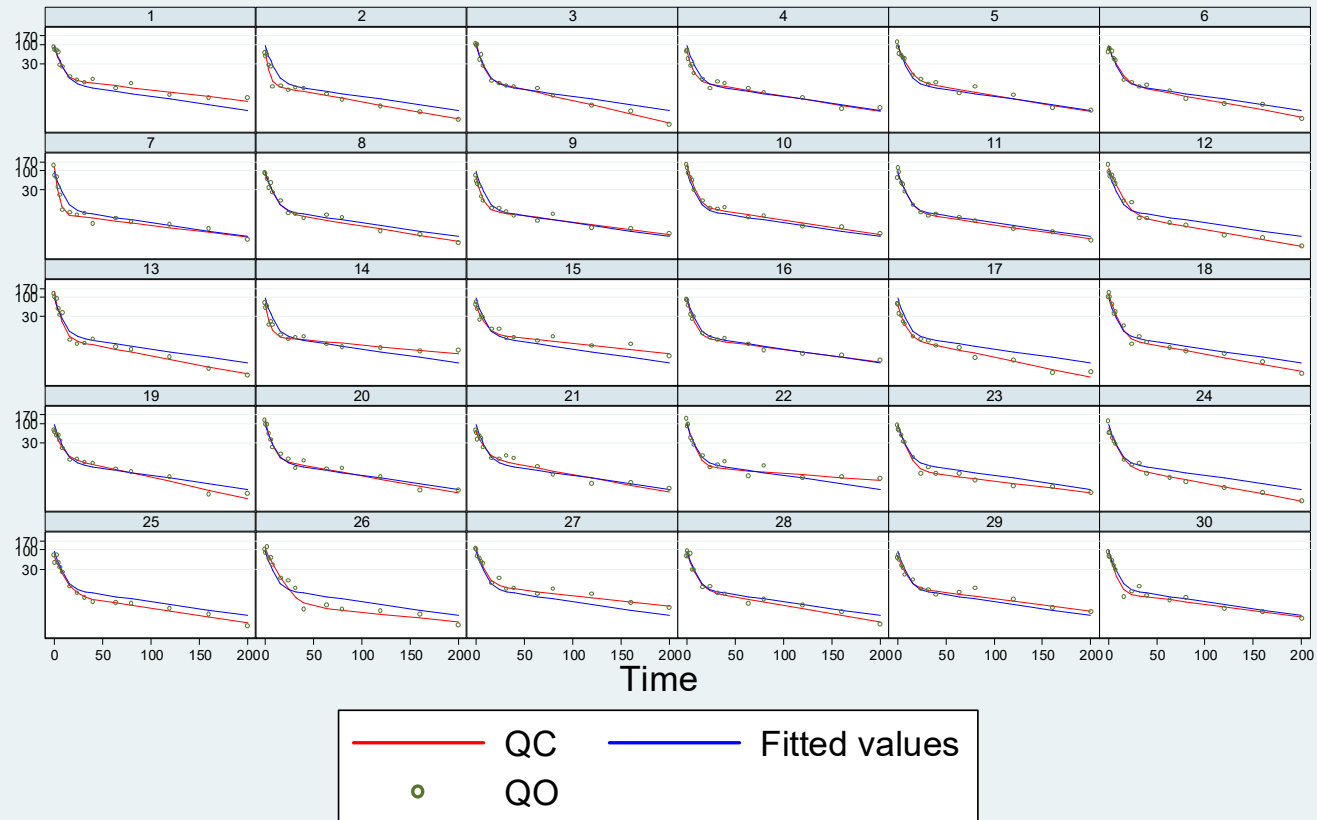


Using the population analysis service all (30) studies in the population are fitted and the results automatically compiled into a file for the second phase of the analysis



The **local** (red line) and **population** (blue line) fits to the **population** data ... local fitting information stored after the first phase and population fitting information after the second phase ... dots are the study level simulated data

Bi-phasic response and Fit



Graphs by Study

Reading the results of the population analysis into Stata and performing some additional statistical analysis

```
cd "c:\talks\winSAAM Meeting 2012\project"
insheet using popcom.stmp, clear
tabstat studyvalue popmean, by(parameter)
tabstat studyvalue, by(parameter) s(mean sd sem)
collapse (sd)   sparam = studyvalue ///
          (sem)  separam= studyvalue ///
          (mean) mparam = studyvalue, by(param)
sort param
save raw_values, replace

insheet using popparam.stmp, clear
list
sort param
merge param using raw_values
order popmean mparam esterror separam popstd sparam
list
```

How do our population statistics compare to the crude compilation of study level statistics?

Raw estimates of mean, error, and sd

```
. collapse ...
. save raw_values, replace
. list
```

	parame~r	sparam	separam	mparam
1.	P(01)	.068662	.0125359	.215809
2.	P(02)	.0023326	.0004259	.009119
3.	P(03)	24.54332	4.480977	82.79366
4.	P(04)	2.262309	.4130392	9.07728

```
. insheet using popparam.stmp, clear
. list
```

Population estimates of mean, error, and sd

	parame~r	popmean	esterror	popstd
1.	P(01)	.20278	.011105	.048273
2.	P(02)	.0091965	.00041	.0018104
3.	P(03)	79.981	4.1163	19.225
4.	P(04)	8.8133	.39974	1.7473

. list Combined illustration

	popmean*	mparam	esterror*	separam	popstd*	sparam	parame~r	_merge
1.	.20278	.215809	.011105	.0125359	.048273	.068662	P(01)	3
2.	.0091965	.009119	.00041	.0004259	.0018104	.0023326	P(02)	3
3.	79.981	82.79366	4.1163	4.480977	19.225	24.54332	P(03)	3
4.	8.8133	9.07728	.39974	.4130392	1.7473	2.262309	P(04)	3

Final population statistics forming the basis for our study design exploration
... posterior Bayesian analysis would be used here

```
. insheet using popparam.stmp, clear  
. list
```

```
+-----+  
| parameter      popmean    esterror|  
+-----+  
1. |      P(01)         .20278     .01111|  
2. |      P(03)        79.981     4.1163|  
3. |      P(02)         .00919     .00041|  
4. |      P(04)         8.8133     .39974|  
+-----+
```

Note 1: The ease with which the population analysis results are read into Stata

Note 2: The population estimates of the parameters based on the assumed distributions, assumed sampling schedule, and assumed study size are ALL well resolved

Procedure for capturing repeated activations of a specific design

```
cap conf file all_results_30.dta
if _rc==601 {
    save all_results_30, replace
}
else {
    save temp, replace
    use all_results_30, clear
    append using temp
    save all_results_30, replace
}
```

Companion Stata code for testing the principles of the investigation

```
cd "c:\talks\winsaam meeting 2012\project"
insheet using data.stmp, clear

scatter qo time, c(1) lp(dash) ms(. oh) by(study, style(compact) ///
    title(Bi-phasic simulated studies)) yscale(log) ylabel(,angle(0)) ///
    ylabel(30 100 170)
more

scatter qc qo time, c(1 1) lp(solid dash) ms(. oh) by(study, style(compact) ///
    title(Bi-phasic response and Fit)) yscale(log) ylabel(,angle(0)) ///
    ylabel(30 100 170)
more
nl (qo={p1}*exp(-{p2}*time)+{p3}*exp(-{p4}*time)), ///
    vce(cluster study) init(p1 90 p2 .1 p3 16 p4 .01) nolog
predict pqc

scatter qc pqc qo time, c(1 1 .) ms(i i oh) by(study, style(compact) ///
    title(Bi-phasic response and Fit)) yscale(log) lc(red blue .) ylabel(,angle(0)) ///
    ylabel(30 100 170)
more
byvar study, gen ret tab b(p1 p2 p3 p4): nl (qo={p1}*exp(-{p2}*time)+ ///
    {p3}*exp(-{p4}*time)) [awe=1/qo^2], init(p1 90 p2 .1 p3 16 p4 .01) nolog
rename B??_ B??
keep study Bp1 Bp2 Bp3 Bp4
contract B* study
drop _freq
sort study
save Stata_results, replace

forval i=1/4 {
    rreg Bp`i', nolog
}
```

Thank You

Comparison of WinSAAM and SAAM II

Our Method for Population Analysis has been **borrowed** by SAAM II. The Method for Simulating and Investigating Population Models and the terminology in each case has been independently outlined by the SAAM II development group

Whereas SAAM II offers Population Modeling support via its companion software POPKINETICS (Vicini et al), WinSAAM offers onboard Population Modeling support.

Whereas POPKINETICS offers instructional aids for Population Simulation via SAAM II, WinSAAM has built on Interoperability to implement extended services via Stata.

Stata offers a rich suite of statistical tools (e.g. mixed effects modeling), graphics tools, and data management services

Stata supports a command file operational approach, a matrix programming language, and facility for user development and refining commonly used functions

Rather than 'risking' costly extensions of WinSAAM's capabilities beyond its natural strengths we have elected to facilitate interoperability to augment the power and utility of its environment