Translational and Clinical Pharmacology
TCP

Tutorial

Check for updates

A simple time-to-event model with NONMEM featuring right-censoring

Quyen Thi Tran [1], **Jung-woo Chae** [1,2,*], **Kyun-Seop Bae** [3,*], and **Hwi-yeol Yun** [1,2,*]

[1]College of Pharmacy, Chungnam National University, Daejeon 34134, Korea
[2]Department of Bio-AI convergence, Chungnam National University, Daejeon 34134, Korea
[3]Department of Clinical Pharmacology and Therapeutics, Asan Medical Center, University of Ulsan, Seoul 05505, Korea

🔓 OPEN ACCESS

*Correspondence to
Jung-woo Chae
Clinical Pharmacy Lab, College of Pharmacy/
Bio-AI Convergence Research Center,
Chungnam National University, 99 Daehak-ro,
Yuseong-gu, Daejeon 34134, Korea.
Email: jwchae@cnu.ac.kr

Kyun-Seop Bae
Department of Clinical Pharmacology and
Therapeutics, Asan Medical Center, University
of Ulsan, 88 Olympic-ro 43-gil, Songpa-gu,
Seoul 05505, Korea.
Email: ksbae@amc.seoul.kr

Hwi-yeol Yun
Pharmacotherapy Lab, College of Pharmacy/
Bio-AI Convergence Research Center,
Chungnam National University, 99 Daehak-ro,
Yuseong-gu, Daejeon 34134, Korea.
Email: hyyun@cnu.ac.kr

ORCID iDs
Quyen Thi Tran 🔟
https://orcid.org/0000-0003-2218-0957
Jung-woo Chae 🔟
https://orcid.org/0000-0001-6026-7063
Kyun-Seop Bae 🔟
https://orcid.org/0000-0001-7399-5879
Hwi-yeol Yun 🔟
https://orcid.org/0000-0001-8793-2449

ABSTRACT

In healthcare situations, time-to-event (TTE) data are common outcomes. A parametric approach is often employed to handle TTE data because it is possible to easily visualize different scenarios via simulation. Not all pharmacometricians are familiar with the use of non-linear mixed effects models (NONMEMs) to deal with TTE data. Therefore, this tutorial simply explains how to analyze TTE data using NONMEM. We show how to write the code and evaluate the model. We also provide an example of a hands-on model for training.

Keywords: NONMEM; Right-Censoring; Time-to-Event; Tutorial

INTRODUCTION

Time-to-event (TTE) analysis considers whether an event of interest occurs and the time of occurrence. The TTE can be the time from diagnosis to death, or the time from enrollment in a study to the development of a disease of interest. Non-parametric, semi-parametric, and parametric approaches are used to analyze TTE outcomes. In this tutorial, we focus on a parametric approach featuring a non-linear mixed effects model (NONMEM). In addition, we also provide a short description of handling TTE data in R. Generally, with a simple TTE data, R can be a good option due to its simple and easy implementation. However, with a more complex situation such as the repeated TTE data or integrating with pharmacokinetics/pharmacodynamics model, R could not work well. In such cases, NONMEM is a common choice owing to its flexibility of modifying model for specific situations.

In a parametric survival model, the survival time (the outcome) is assumed to follow a known distribution. This is a well-recognized approach used to explore the relationship between survival and various explanatory variables. It yields an appropriate distribution of TTE data of interest. In TTE modeling, several mathematical terms must be understood, including the probability density function $f(t)$, the survival function $S(t)$, and the hazard function $h(t)$ [1]. The probability density function describes the likelihood of observing an event at a particular time t:

$$f(t) = S(t) \times h(t)$$

Generated by xmlinkpress

Conflict of Interest

- Authors: Nothing to declare
- Reviewers: Nothing to declare
- Editors: Nothing to declare

The hazard function is the instantaneous failure rate of event occurrence; this means that one has survived to time t but will experience the event in the next instant of time.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t_1 | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

The survival function reflects the probability that the event of interest has not yet occurred by time t.

$$S(t) = \Pr\ (T > t)$$

where T is the time to observation of an event.

The relationship between the survival and hazard functions is:

$$S(t) = e^{\left(-\int_0^t h(u)\,du\right)}$$

TTE DISTRIBUTIONS

Several survival models can be applied to describe TTEs, including exponential, Weibull, Gompertz, and log-logistic models. The equations of the baseline hazard and survival functions corresponding to each distributional form are listed in **Table 1**.

If any covariate (e.g., age, sex, or drug dose) is relevant, the covariate (COV) features in the hazard equation:

$$h(t) = h_0(t) \times COV$$

COVs can be diverse. For example, if a covariate is dichotomous, COV can take the form of $e^{\beta i x i}$. If a covariate is continuous, COV can be written as $e^{\beta i(xi - xi, mean)}$ [2].

TRANSLATION OF TRIAL DATA TO THE MODEL

In most survival analyses, censoring is a key analytical problem. Censoring is understood as when we lack information on exact survival times but have some data on individual survival times. Several types of censoring include right-censoring, left-censoring, and interval censoring [3]. For right-censored data, the true survival time is equal to or greater than the observed survival time. For left-censored data, the true survival time is less than or equal to the observed survival time; however, this is rarely encountered. For interval-censored data, the true survival time lies within a known interval. In this tutorial, we focus on how to handle right-censored data. Right-censoring can occur if the study ends without any event, or if patients are lost to follow-up or withdraw.

Table 1. The baseline hazard and survival functions of some common TTE distributions

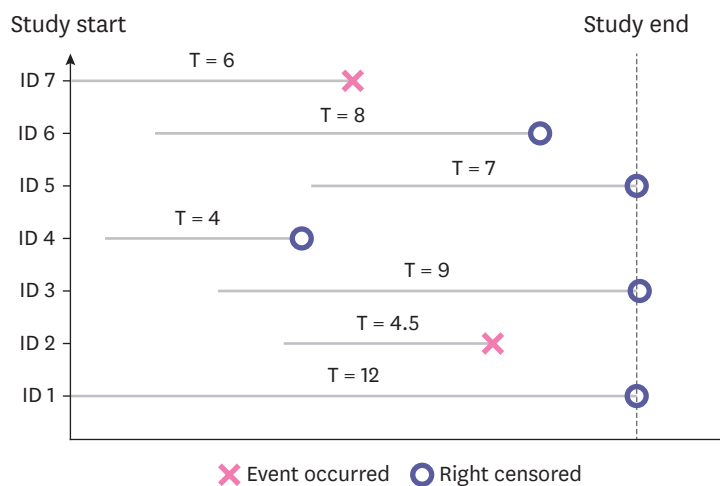| TTE distribution	Baseline hazard function $h_0(t)$	Survival function $S(t)$
Exponential distribution	λ	$e^{-\lambda t}$
Weibull distribution	$\lambda \gamma t^{\gamma-1}$	$e^{-\lambda t^\gamma}$
Gompertz distribution	$\lambda e^{\gamma t}$	$e^{-\frac{\lambda}{\gamma}(e^{\gamma t}-1)}$
Log-logistic distribution	$\dfrac{\lambda \gamma t^{\gamma-1}}{1 + \lambda t^\gamma}$	$\dfrac{1}{1 + \lambda t^\gamma}$

Figure 1. Example of raw time-to-event data.

ID	Time	DV	Sex	Note
1	0	.	0	
1	12	0	0	Right-censored
2	0	.	1	
2	4.5	1	1	Event occurred
3	0	.	0	
3	9	0	0	Right-censored
4	0	.	1	
4	4	0	1	Right-censored
5	0	.	0	
5	7	0	0	Right-censored
6	0	.	0	
6	8	0	0	Right-censored
7	0	.	1	
7	6	1	1	Event occurred

Figure 2. Translation of clinical data to non-linear mixed effects model data.

As an example, **Fig. 1** shows certain events occurring within a study. Event IDs 1, 3, 4, 5, and 6 are right-censored. Censoring of ID 1, 3, and 5 reflects the end of the study; censoring of ID 4 and 8 reflect possible loss to follow-up or withdrawal. For ID 2, the event occurred after 4.5 days; for ID 7, the event occurred at 8 days. To apply these data to NONMEM, it is necessary to translate and format the original data as shown in **Fig. 2** (the last column ["Note"] simply aids understanding; it is not part of the NONMEM data file). Each subject features two time points, of which one is the commencement of observation (time = 0) and the other the time of an event or censoring. If more subject information (e.g., age or gender) is available, one can add more columns to include the data.

BASIC STRUCTURE AND EXECUTION OF NONMEM

Creating a data file
The example dataset used in this tutorial features 50 individuals observed over 100 days. Under the TTE assumption, each individual would randomly experience only one event over

ID	Time	DV	MDV	Sex	EVID	Type
102	0	.	1	1	3	1
102	1	.	0	1	0	0
102	2	.	0	1	0	0
102	3	.	0	1	0	0
102	4	.	0	1	0	0
102	5	.	0	1	0	0
102	6	.	0	1	0	0
102	7	.	0	1	0	0
102	8	.	0	1	0	0
102	9	.	0	1	0	0
102	10	.	0	1	0	0
102	11	1	0	1	0	1
102	12	.	0	1	0	0
102	13	.	0	1	0	0

Figure 3. An example dataset used for time-to-event analysis by non-linear mixed effects model.
In the **Fig. 3**, ID: subject identification number; TIME, the times when observations commenced and those of events, including events that in fact occurred or were censored (these are required for the estimation step); and all times from zero to the end of the study (required for the simulation step). For example, for ID 102 in **Fig. 3**, the values in the TIME column consist of all time points from zero to the end of the study. However, for a specific purpose, TYPE is set to 1 when data are to be estimated (thus only at the times of zero and the events) and to 0 when data are to be simulated (at all time points); DV, discrete observation events (zero-0 or one-1) (zero means censored, one means occurred); MDV, missing DV; SEX, etc., individual characteristics (if available); EVID, event ID (optional) (EVID = 0 indicates an observation; EVID = 3 indicates the commencement of a new individual); TYPE, as mentioned above, TYPE is used for separating dataset as a specific purpose.

the 100 days (this may either occur or be censored). The data for simulation should include complete daily records. The full dataset is provided in the **Supplementary Data 1** (dataTTE_nm.csv file); the format is that of **Fig. 3**. It is tedious to format big data in Excel. Therefore, we provide an R code allowing simple creation of a NONMEM dataset from translated clinical data. All files are in the **Supplementary Data 2** and **3** (translated data, dataTTE.csv file; R code, part 1 in R_code.docx file).

The control file
The control stream features several blocks. Most terms were well-explained in the NONMEM Tutorial Part I [4], therefore, we will focus on terms and codes not mentioned there. An example control stream file of the TTE Gompertz model is provided in the **Supplementary Data 4** (run101.mod file).

In the $DATA block, the commands *;Sim_start* and *;Sim_end* allow the employment of user-defined simulation codes when running PsN with the option of *-flip_comments*. This inverts comments between the tags. For example, if the comments above is in the model, then the *MAXEVAL = 0*, model will run as such and the simulation model will instead have a line as in the right code:

```
$DATA dataTTE_nm.csv                 -flip_comments      $DATA dataTTE_nm.csv
IGNORE=@                                   ──────▶        IGNORE=@
;Sim_start                                                ;Sim_start
IGNORE=(TYPE.EQ.0)                                        ;IGNORE=(TYPE.EQ.0)
;Sim_end                                                  ;Sim_end
```

The $PK block defines the parameters of the hazard model. For example, the hazard model in this *.mod* file follows the Gompertz distribution; thus, two parameters are defined:

```
$PK
LAM = THETA(1)*EXP(ETA(1))
GAM = THETA(2)
DEL = 1E-8 ; to avoid value zero of time
```

The $DES block describes the hazard equation:

```
$DES
DADT(1) = LAM*EXP(GAM*(T + DEL)) ; hazard
```

The $ERROR block calculates the hazard, survival, probability of events, and commands for simulation.

```
$ERROR
CHAZ = A(1); cumulative hazard
SUR = EXP(-CHAZ) ; survival probability
HAZNOW = LAM*EXP(GAM*(TIME +DEL))


IF (DV.EQ.0) Y = SUR ; probability of survival
IF (DV.NE.0) Y = SUR*HAZNOW ; probability density function of event
```

The $ERROR block also contains the code for the simulation step (the right of **Fig. 4**). The explanation of simulation procedure is explained on the left of **Fig. 4**, where:


```
IF(ICALL.EQ.4) THEN ; for simulation
  IF(NEWIND.NE.2) THEN ; for new ID
  DV = 0
  TTE = 0
  OTTE = 0 ; count for previous event
  CALL RANDOM (2,R); 2nd distribution
  USUR = R
  ENDIF

IF(OTTE.EQ.0) THEN
  IF(USUR.GT.SUR) THEN
    IF(TIME.LE.100) THEN
    DV = 1
    TTE = 1
    OTTE = 1
    ENDIF
  ELSE
    IF(TIME.EQ.100) THEN
    DV = 0
    TTE = 1
    OTTE = 1
    ENDIF
  ENDIF
  ENDIF
ENDIF
```
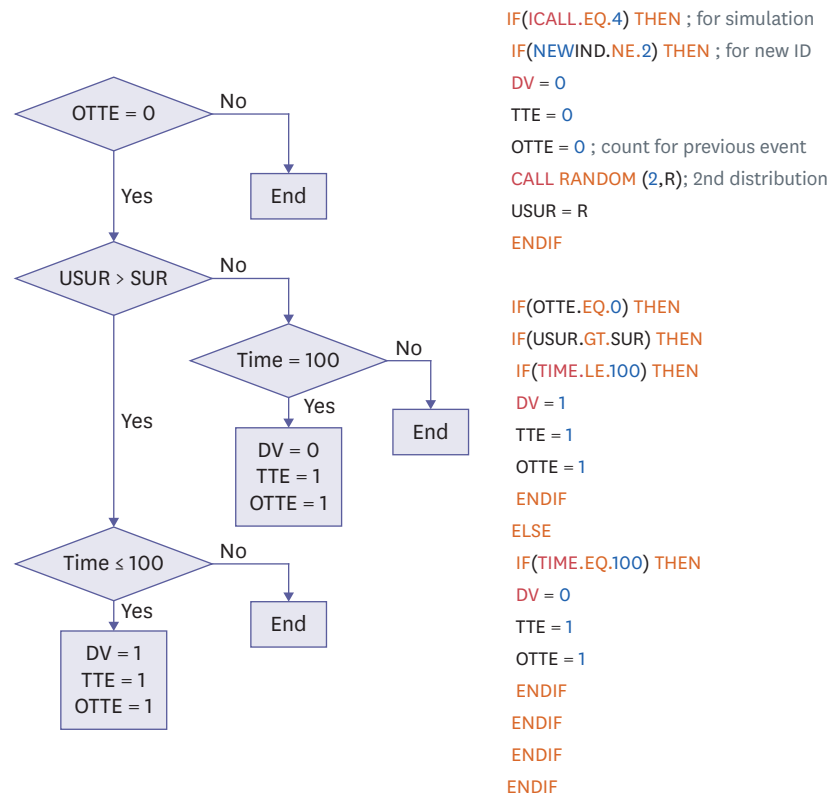
Figure 4. Simulation of a time-to-event model in non-linear mixed effects model.

DV, event; receives a value of *zero-0* when no event occurs (at time = 0) or an event is censored, but a value of *one-1* when an event occurs. *TTE*, a variable for event; required for a visual predictive check (VPC); *OTTE*, a variable for counting event; this ensures that in a TTE model only one event occurs; both *TTE* and *OTTE*, values of *zero-0* when no event occurs, and values of *one-1* for events (censored and occurring).

When the simulation is initiated:
At time = 0, *DV*, *TTE*, an d *OTTE* are set to zero. Whenever a new ID is called [*NEWIND* not equal *(NE.)* to *2*], a random value (USUR) is picked from a uniform distribution and compared to the survival at every time point (SUR).

If USUR > SUR and time ≤ 100, an event occurs, and DV, TTE, and OTTE are set to 1. If USUR ≤ SUR and time = 100, the event will be censored, and DV is set to 0 but TTE and OTTE are set to 1.

The estimation step features various methods including Laplace, stochastic approximation expectation-maximization, and importance sampling methods; the details are in the work of Karlsson et al. [5].

```
;Sim_start
$EST METHOD=COND LAPLACE LIKE MAXEVAL=9999 PRINT=1 MSFO=msfb101 NSIG=3 SIGL=9
$COV PRINT=E
;$SIMULATION (5988566) (39978 UNIFORM) ONLYSIM NOPREDICTION SUB=100
;Sim_end
```

In the above code, commands *;Sim_start* and *;Sim_end* are again used to switch from estimation to simulation. When either comment is called, the PsN will feature this line in the code:

```
; Sim_start
;$EST METHOD=COND LAPLACE LIKE MAXEVAL=9999 PRINT=1 MSFO=msfb101 NSIG=3 SIGL=9
;$COV PRINT=E
$SIMULATION (5988566) (39978 UNIFORM) ONLYSIM NOPREDICTION SUB=100
;Sim_end
```

The table output file is used for visualization:

```
; Xpose
$TABLE ID TIME SUR EVID ONEHEADER NOPRINT FILE=sdtab101
```

Running the model:

```
execute run101.mod
```

Selection of the model:
During TTE analysis, attention should be paid to the objective function between nested models, the precision of the parameter estimates, and scientific plausibility. In addition, a model is selected based on a visual predictive check. To execute the VPC, the PsN command is:

```
vpc run101.mod -flip_comments -tte=TTE -sample=n stratify_on=SEX -rplot=1
```

Here:
   -*flip_comments* inverts the comments between the tags *;Sim_start* and *;Sim_end*.
   -*tte* = *TTE* is required in a TTE model used for simulation, and will have a value of *zero-0* if

an observation is not an event, but *non-zero* (for example, *1*) if an observation is an event (occurring and censored events). PsN will add this value to the Table file of the simulation as a column name in the TTE, and will format it unlike a regular VPC (to specifically engage the kalplan.plot functionality of Xpose). For more information on the VPC options for TTE modeling, refer to the VPC and NPC user guides [6].

-*samples = n*: The simulation will run with a sample size of *n*. Generally, to adequately evaluate the appropriateness of a developed model, the VPC should run with a sample size of 1,000.

-*stratify_on = SEX* (optional): This option allows the user to specify the VPC as groups divided by (for example) gender or the drug dose.

-*rplot = 1* (optional): Basic VPC plots are generated.

Plotting the VPC:

When plotting the VPC using Xpose, the following R code is used. The code is in the **Supplementary Data 3** (part 2 in R_code.docx).

```
# set up the directory working in R
setwd('C:/Users/quyen/OneDrive/NM TTE')
# Load xpose4 package for running VPC
require(xpose4)
# name the object
new.runno <- '101'
# create a xpose.data object
newdb <- xpose.data(new.runno)
# call option to create VPC of Kaplan-Meier plot
# option "by" allows specifying plot as group, for example, groups as SEX
kaplan.plot(VPC = T, object = newdb, by = c("SEX == 0", "SEX == 1"), main = NULL, main.sub = c("Female", "Male"),
ylab = "Survival Probability (%)", xlab = "Time (Day)", main.sub.cex = 1.2, real.col = "black", poly.fill = "green", poly.alpha
= 0.35, probs = c(0.05, 0.95))
```

Results:

Table 2 and **Fig. 5** present the estimated parameters and the VPC, respectively, after estimation and evaluation using the example dataset in the **Supplementary Data 1**.
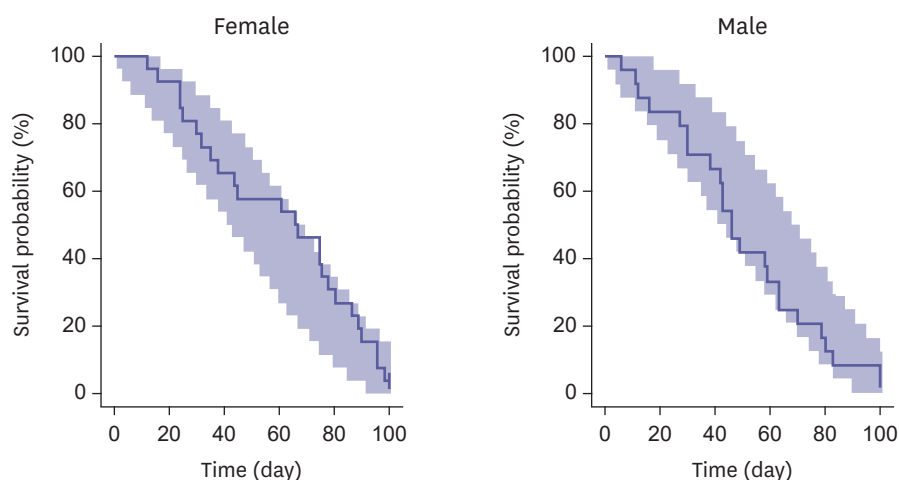

Figure 5. The visual predictive check of the time-to-event model. Solid line represents observed survival probability, and the shaded area represents 90% prediction interval of simulated data.

Table 2. The estimated parameters of the example TTE dataset

Parameters	Estimated value	Relative standard error
Rate factor (λ)	0.0052	32%
Shape factor (γ)	0.0279	18%
Random effect*	0 FIX	-

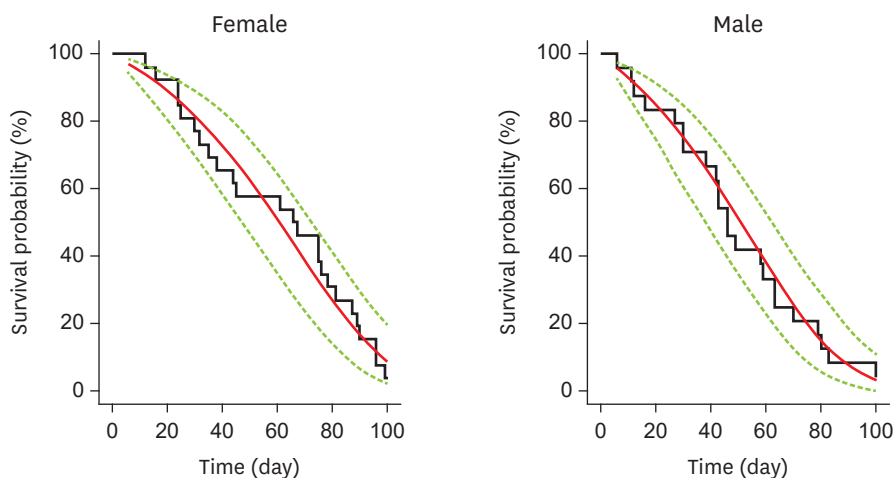*A random effect describing the log-normal distribution of the hazard.

Figure 6. Plots of survival probability according to gender. Black line indicates observed survival probability, red line indicates fitted curve, and green lines indicate 95% confidence interval of fitted curve.

ANALYZING TTE DATA IN R

TTE data can be analyzed by using package "flexsurv" [7]. This package provides users with fitting options of various distributions such as exponential, gompertz, weibull, etc. The example code was provided in the **Supplementary Data 3**, part 3 in R_code.docx.

The function "flexsurvreg" gave the results as same as in the NONMEM model with shape parameter = 0.0279 and rate parameter = 0.0052. The plots of survival probability were shown in the **Fig. 6**.

REFERENCES

1. Lim HS. Brief introduction to parametric time to event model. Transl Clin Pharmacol 2021;29:1-5.
   PUBMED | CROSSREF
2. Juul RV, Rasmussen S, Kreilgaard M, Christrup LL, Simonsson US, Lund TM. Repeated time-to-event analysis of consecutive analgesic events in postoperative pain. Anesthesiology 2015;123:1411-1419.
   PUBMED | CROSSREF
3. Kleinbaum DG, Klein M. Statistics for biology and health. New York (NY): Springer-Verlag; 2012.
4. Bauer RJ. NONMEM tutorial part I: description of commands and options, with simple examples of population analysis. CPT Pharmacometrics Syst Pharmacol 2019;8:525-537.
   PUBMED | CROSSREF
5. Karlsson KE, Plan EL, Karlsson MO. Performance of three estimation methods in repeated time-to-event modeling. AAPS J 2011;13:83-91.
   PUBMED | CROSSREF
6. VPC and NPC user guide [Internet]. https://nanopdf.com/download/vpc-and-npc-user-guide_pdf. Accessed May 4, 2022.
7. Jackson C, Metcalfe P, Amdahl J, Warkentin MT, Kunzmann K. Package 'flexsurv' [Internet]. https://cran.r-project.org/web/packages/flexsurv/flexsurv.pdf. Accessed May 4, 2022.

SUPPLEMENTARY MATERIALS

Supplementary Data 1

Click here to view

Supplementary Data 2

Click here to view

Supplementary Data 3

Click here to view

Supplementary Data 4

Click here to view