

On Bayesian Analysis of Univariate Longitudinal Clinical Data (ULCD)

Ganiyu A. Dawodu, Ph.D.*; Olaniyi M. Olayiwola, Ph.D.; and Osebi E. Asiribo, Ph.D.

Statistics Department, Federal University of Agriculture, Abeokuta (FUNAAB)

E-mail: agad1963@yahoo.com*
dawoduga@funaab.edu.ng

ABSTRACT

ULCD are very common in public health delivery systems; their measurements are particularly entrenched in the process of taking vital statistics of the patients and play a crucial role in the management of terminal ailments (e.g. hypertension and diabetics). This measurement could be the sole determinant that captures the severity of the ailment, during the period that the patient is around the clinic for a medical examination. Patients, for the sake of confidentiality, rarely go for their medical re-examinations, more so, if they feel that their medical state is stable and “seem” safe. Hence a patient may completely jettison attendance at the clinic or handles it with levity and this will usually result into irregular attendance of the clinic. It is of paramount importance to note that a Bayesian methodology is the most appropriate for the analysis of the resultant irregular data. This work purports to provide such a methodology that can be adopted in the analysis of the said irregular or extremely few-entries, univariate data, such that the inference obtained will be as reliable as that from the analysis of regular and adequate data. A numerical illustration is shown to further exhibit the functionality of the proposed methodology.

(Keywords: Bayesian analysis; ULCD; Markov Chain Monte Carlo, MCMC simulations; probability density distribution, PDF, rjags)

INTRODUCTION

Clinical data usually originates from clinical research that can be categorized as; observational studies, experiments, surveys, and their hybrids (i.e., a combination of a pair or the three categories) (Suresh and Chandrashekar, 2012; Mathew and Murthy, 1998). All research has to be scientifically planned and properly laid-out with good randomization processes to ameliorate or usurp the presence of any pertinent

extraneous sources of errors that can possibly confound the essence of the research (Box et al, 2005; Cochran and Cox, 1957).

Univariate Longitudinal Clinical Data (ULCD) are born out of the observational study part of the research (Raffa, 2012; Wang et al., 2016; Barbieri et al., 2017). They are usually included in the vital statistics that are continuously taken whenever the experimental units (i.e., the concerned patients) come around for either a “fresh” medical consultation or subsequent re-examinations. This approach to data generation is particularly useful whenever the ailment under study has been socially stigmatized (Okeefe and Rubin, 2014).

The main problem of all categories of research is inability to quickly obtain the required, primarily sourced, sample size within the study period (Daniels and Hogan, 2008; Brown and Prescott, 2006). A reliable inference can only be drawn from a research in which, at least, an optimum sample size was utilized (Cochran, 1977). However, if the study statistics are robust to lack of normality or Bayesian, then the scourge of too few samples can be averted (Luo, 2014). Now, not too many reliably strong study statistics are so robust as to lack of normality and this virtually leaves us with no alternative than to always desire Bayesian study statistics whenever our study ailment is, at least socially stigmatized. Socially stigmatized ailments are usually lethally incurable, terminal, and contagious. Most people will not want sufferers of such ailments (i.e. the study units) around them, most especially, when they do not have any “protective gadgets” in place.

Bayesian study statistics are very useful even when there is only one sample datum as “prior” (Spiegelhalter et al., 2004; Luo, 2014). Before the application of Bayesian concepts in clinical research, many researchers have had to jettison

very expensive and useful research because of lack of adequate samples or data. They have also had to abandon research whenever some sizeable amount of the study units abscond or die (Norleans, 2001). This work is geared towards presenting a Bayesian approach to utilize very few samples from pertinent study units to synthesize the appropriate PDF with which the ailment could be properly studied with respect to the study units from which the data was obtained. The required data (i.e. ULCD) should be nonempty (i.e. must contain, at least one datum) and be taken with respect to particular members from the study units.

MATERIALS AND METHODS

Basic Theory

This methodology will ultimately be coded and run in the rjags software package. The rjags package provides an interface from R (Crawley, 2007) to the JAGS (Just Another Gibbs Sampler) library for Bayesian data analysis. JAGS uses Markov Chain Monte Carlo (MCMC) to generate a sequence of dependent samples from the posterior distribution of the parameters. JAGS is written in C++ and is portable to all major operating systems. JAGS is designed to work closely with the R language (facilitated by both rjags and R2jags packages) and environment for statistical computation and graphics (<http://www.r-project.org>).

Users of rjags should also endeavor to install the coda package for R to analyze the output. JAGS is a clone of BUGS (Bayesian Analysis Using Gibbs Sampling). Lunn et al (2009) contains a short history of the BUGS project. The rjags package does not include a copy of the JAGS library and hence it is required that its users must install this library separately. The rjags package does not use the command line interface, like JAGS, but provides equivalent functionality using R functions. Analysis using the rjags package proceeds in the following steps:

1. Define the model using the BUGS language in a separate file.
2. Read in the model file using the "jags.model" function. This creates an object of class "jags".
3. Update the model using the update method for "jags" objects. This constitutes a 'burn-in' period.

4. Extract samples from the model object using the "coda.samples" function. This creates an object of class "mcmc.list" which can be used to summarize the posterior distribution. The coda package also provides convergence diagnostics to check that the output is valid for analysis (Plummer et al., 2006).

RESULTS AND DISCUSSION

In the management of some terminal ailments (e.g., hypertension and diabetes) in patients, some crucial measurements are usually made. The said measurements are so important that they have to be done at regular interval, at least, until perhaps, a stability of the monitored "indicator" is noticed (i.e., when the quantified "trait" that is reflected in the measurements converges to a value that seems to be "safe").

Usually, at the beginning of the management of the ailment, the patient is put on admission in a ward, so that the said trait can be effectively monitored until, at least, when the stable period is reached and the patient may be discharged with an advise to "imbibe" the "habit" that he/she was "subjected to" during the "stability seeking" period (i.e., period of his/her admission). Now suppose that a hypothetical out-Patient of hypertension of a Medical Centre, only revisited for re-examinations for just twenty times after he was discharged and each time his systolic, "sphygmograph" reading was recorded along with his vital statistics. Assuming that his twenty (20) systolic readings are:

98,160,136,128,130,114,123,134,128,107,123,125,129,132,154,115,126,132,136,130,

then to ease task of following the four-steps algorithm described in the Methodology Section, the following R codes was written:

```
> require(MatrixModels)
> library(rjags)
> systolic <-
c(98,160,136,128,130,114,123,134,128,107,
123,125,129,132,154,115,126,132,136,130)
> length(systolic)
> sum(systolic)
```

```

> ( mn <- mean( xx <- rgamma( 10000, 2560.01,
20.01 ) ) )

> ( md <- median( xx ) )

> curve( dgamma( x, 2560.01, 20.01 ), from=120,
to=135, lwd=2 )

> abline( v=mn, col="red", lwd=3 )

> abline( v=md, col="blue" )

> a.dat<-
list(systolic=c(98,160,136,128,130,114,123,134,1
28,107,123,125,129,132,154,115,126,132,136,13
0,NA), l=21)

> cat("model { sigma2 <- 169; tau <-1/sigma2;
mu ~ dnorm(0,0.00001);
for (i in 1:21) { y[i] ~ dnorm(mu,tau)}
p.above <- step(mu-125); p.below <- step(125-
mu); y.new ~ dnorm(mu,tau) }",
file="m1.jag")

```

```

> a.ini <- list( list( mu=120 ), list( mu=125 ), list(
mu=130 ) )

> m <- jags.model( file = "m1.jag", data = a.dat,
n.chains = 3,
inits = a.ini, n.adapt = 2000 )

> res <- coda.samples( m, var = "mu" , n.iter =
10000, thin = 10 )

> class(res)

> str(res)

> summary( res )

> par( mfrow=c(1,2) )

> plot( res )

```

The output that was generated, when the codes were run on R (version 3.4.1) that has its allied packages (e.g. rjags, coda), was properly linked to JAGS 4.3.0 is as contained in Figure 1.

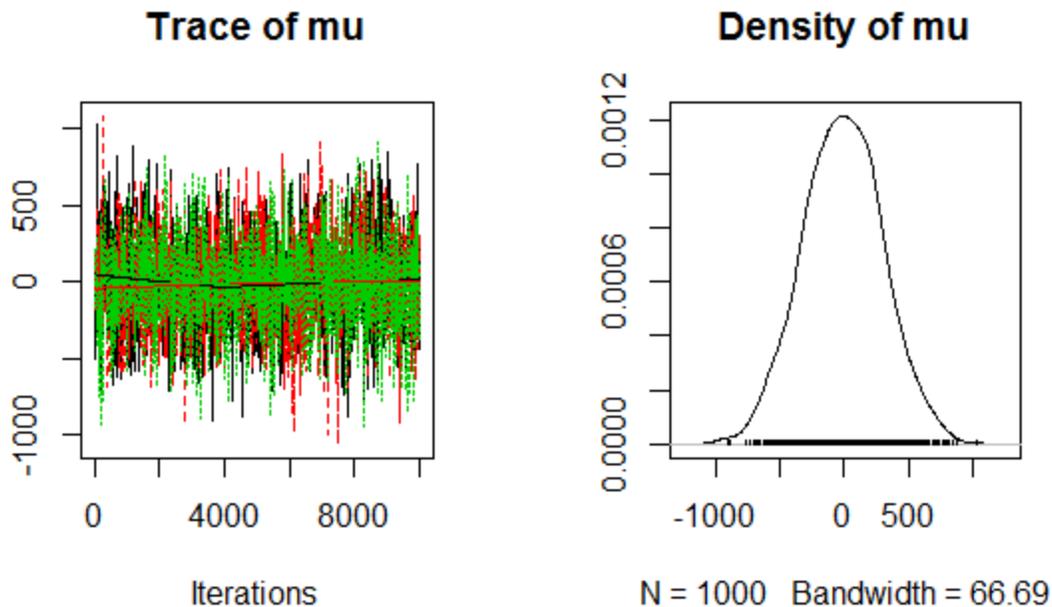


Figure 1: The Trace and Density Plots on the Population Parameter (μ) on the Systolic Data.

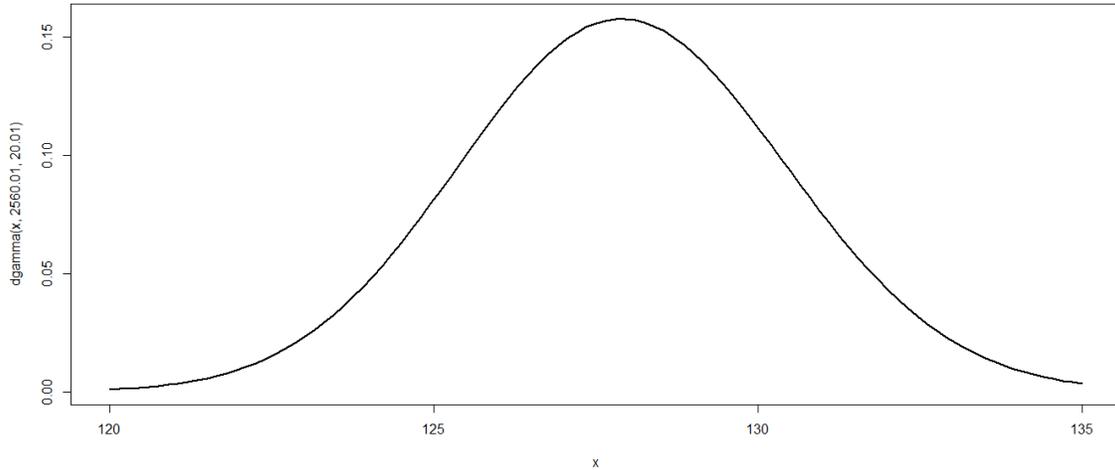


Figure 2: Sketch of the Posterior is “Bell-shaped” thus Stipulating that the Guess of Normality is Valid.

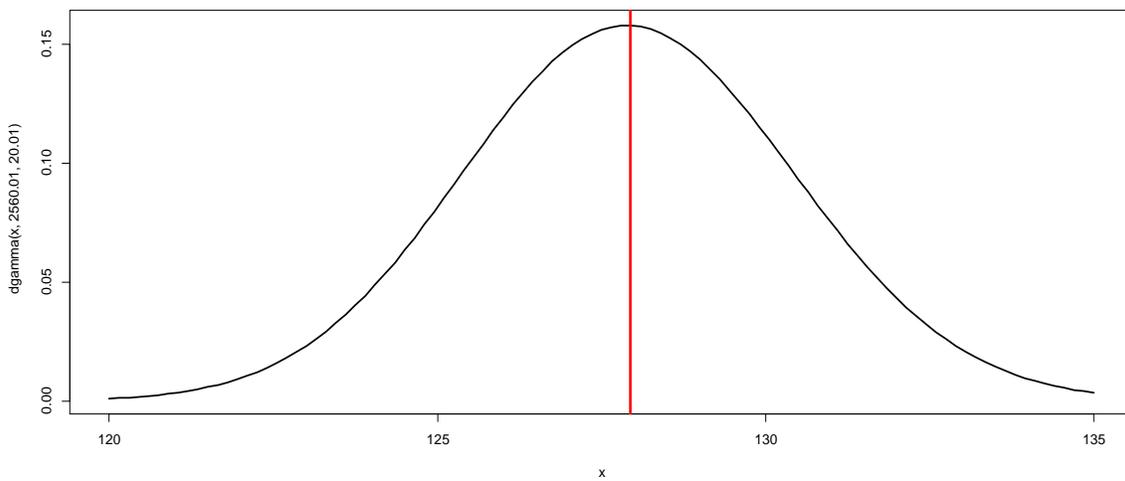


Figure 3: The Red Line is the Mean Line as well as the “Line of Symmetry”.

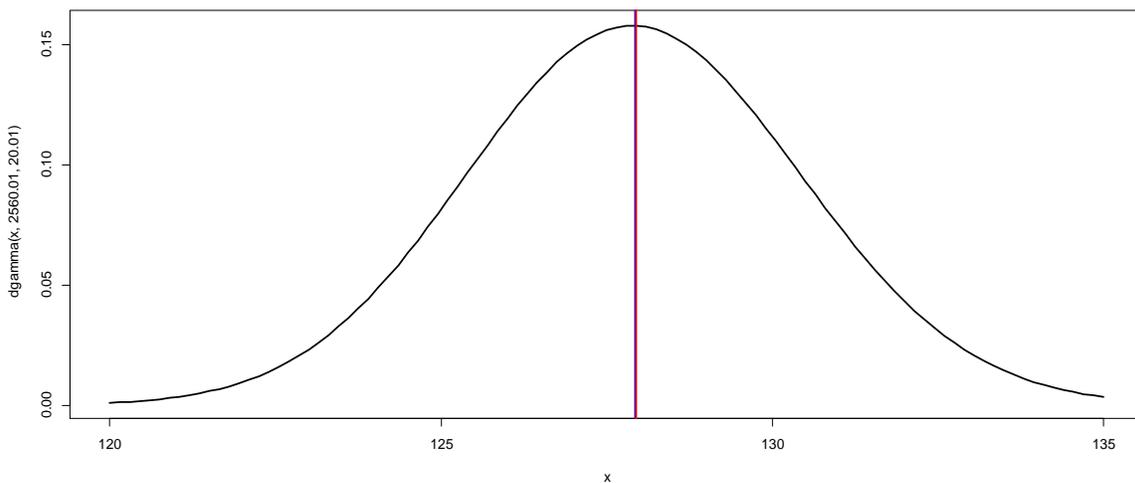


Figure 4: The Blue Line (i.e. the “Median line”) is Almost Superrimposing the Red Line (i.e. the “Mean line”).

CONCLUSION

The systolic data is here presented as an example of ULCD; other examples can be from diabetics (e.g. the quantity of sugar in urine test) origin, HIV origin, and the Jaundice babies' syndrome (e.g. the bilirubin monitoring situation), and any other situations in which a univariate variable, is used to quantify the severity of an ailment, has to be monitored. With respect to the R codes, its analysis is as contained below:

1. The first couple of lines merely prepare R for the running of rjags codes.
2. The third through fifth lines were used to feed the data.
3. The sixth and seventh were used to obtain the sample size and sum of the data entries.
4. The eighth through tenth were used to obtain measures of central tendencies (i.e. mn=mean and md=median)
5. A guess that the prior is normal suggests that the posterior should be gamma, as stipulated in the eleventh and twelfth lines (diagrammatically, shown in Figure 2). The mean (Figure 3) and median (Figure 4) approximates to one another thus suggesting that the guess of normality of the prior is appropriate.
6. The thirteenth through sixteenth lines resupply the data with an indication that it should be used for simulation later. The 21st position contains "NA" (i.e., the location for simulated data entries).
7. Lines 17 through 22 supply the model to JAGS 4.3.0 (through R).
8. Lines 23 and 24 supply the initial values of the population parameter to JAGS 4.3.0
9. Lines 25 through 27 inform JAGS 4.3.0 to execute all the simulation commands
10. Lines 28 and 29 inform JAGS 4.3.0 to keep the trace and density plots of the result in the memory location named "res".
11. Lines 30 and 31 are for routine checks to know whether the output of JAGS 4.3.0 was correctly received by R.

12. Line 32 informs R to summarize and present the result (as seen in Figure 1).

Finally, the result contains the forecasted "would be" medical history (i.e., the trace and density plots of μ) of the patient with respect to the ailment if the patient maintains his/her habits and uses his/her drugs regularly.

REFERENCES

1. Barbieri, A., J. Peyhardi, T. Conroy, S. Gourguo, C. Lavergne, and C. Moller. 2017. "Item Response Models for the Longitudinal Analysis of Health-Related Quality of Life in Cancer Clinical Trials". *BMC Medical Research Methodology*. 17:148. DOI 10.1186/s12874-017-0410-9.
2. Box, G.E.P., W.G. Hunter, and J.S. Hunter. 2005. *Statistics for Experimenters: An Introduction to Design Data Analysis & Model Building (2nd edition)*. John Wiley & Sons, Inc.: New York, NY.
3. Brown, H and R. Prescott. 2006. *Applied Mixed Models in Medicine. Second Edition*. John Wiley and Sons Ltd.: New York, NY. ISBN 0-470-02356-2.
4. Cochran, W.G. 1977. *Sampling Techniques (Third Edition)*. John Wiley and Sons Inc.: New York, NY. ISBN 0-471-16230-X.
5. Cochran, W.G. and G.M. Cox. 1957. *Experimental Designs (Second Edition)*. John Wiley & Sons, Inc.: London, UK.
6. Crawley, M.J. 2007. *The R Book*. John Wiley and Sons Ltd.: West Sussex: UK.
7. Daniels, M.J. and J.W. Hogan. 2008. "Missing Data in Longitudinal Studies Strategies for Bayesian Modelling and Sensitivity Analysis". Chapman and Hall/CRC: Boca Raton, FL.
8. Lunn, D., D. Spiegelhalter, A. Thomas, and N. Best. 2009. "The BUGS Project: Evolution, Critique and Future Directions". *Statistics in Medicine*. 28:3049-67.
9. Luo, S. 2014. "A Bayesian Approach to Joint Analysis of Multivariate Longitudinal Data and Parametric Accelerated Failure Time". *Stat. Med.* 33(4):580-594. DOI: 10.1002/sim 5956.
10. Mathew, A. and N.S. Murthy. 1998. "Clinical Research Methods". *The National Medical Journal of India*. 2(6).

11. Norleans, M.X. 2001. *Statistical Methods for Clinical Trials*. Marcel-Dekker Inc.: New York, NY. ISBN 0-8247-0467-3.
12. Okeefe, C. and D.B. Rubin. 2014. "Individual Privacy versus Public Good: Protecting Confidentiality in Health Research". *Statistics in Medicine*. Wiley Online Library. DOI: 10.1002/sim.6543.
13. Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. "CODA: Convergence Diagnosis and Output Analysis for MCMC". *R News*. 6:7-11.
14. Raffa, J.D. 2012. "Multivariate Longitudinal Data Analysis with Mixed Effects Hidden Markov Models". An unpublished PhD Thesis presented to the University of Waterloo: Ontario, Canada.
15. Spiegelhalter, D.J., K.R. Abrams, and J.P. Myles. 2004. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley and Sons, Ltd.: West Sussex, UK. ISBN 0-471-49975-7.
16. Suresh, K.P. and S. Chandrashekar. 2012. "Sample Size Estimation and Power Analysis for Clinical Research Studies". *Journal of Human Reproductive Sciences*. 5(1).
17. Wang, T., J. Malone, H. Fu, C. Heilmann, Y. Qu, and W.J. Huster. 2016. "Crossover Design and its Application in Late-phase Diabetes Studies". *Journal of Diabetes*. 8:610-618.

ABOUT THE AUTHORS

Dr. Ganiyu A. Dawodu, is a Senior Lecturer in Statistics at the Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta (FUNAAB). He is a Computational Statistician and is a member of six academic bodies including: Nigerian Statistical Association (NSA), Nigeria Statistical Society (NSS), International Biometric Society (IBS, Region Ni), and Royal Statistical Society (RSS). He holds a Doctor of Philosophy (Ph.D.) in Statistics from FUNAAB. His research interests are in statistical modelling and computing.

Dr. Olaniyi M. Olayiwola, is a Senior Lecturer of Statistics at the Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta (FUNAAB). He is a member of Nigeria Statistical Society (NSS), International Biometric Society (IBS, Region Ni), Royal Statistical Society (RSS). He holds a Doctor of Philosophy (Ph.D.) in Statistics from the University

of Ibadan, Oyo State, Nigeria. His research interests are in sampling techniques and statistical inference.

Prof. Osebi E. Asiribo, is a Professor in the Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta (FUNAAB). He is a member of five academic bodies which includes; Nigerian Statistical Association (NSA), Nigeria Statistical Society (NSS), International Biometric Society (IBS, Region Ni). He holds a Doctor of Philosophy (Ph.D.) in Statistics from the University of Wisconsin, Madison, WI, USA. His research interests are in statistical modelling and design of experiments.

SUGGESTED CITATION

Dawodu, G.A., O.M. Olayiwola, and O.E. Asiribo. 2018. "On Bayesian Analysis of Univariate Longitudinal Clinical Data (ULCD)". *Pacific Journal of Science and Technology*. 19(1):105-110.

 [Pacific Journal of Science and Technology](http://www.akamaiuniversity.us/PJST.htm)