A Novel Bayesian Approach to Assessing the Risk of QT Prolongation

Suraj P. Anand,[1,*] and Sujit K. Ghosh[2]

[1]Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA
[2]Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA

Institute of Statistics Mimeo Series #2620

Abstract

The ICH E14 guidelines recommend performing a 'thorough QT/QTc study' to support the safety profile of a drug. The standard way of analyzing a 'thorough QT/QTc study' to assess a drug for its potential for QT prolongation is to construct a 90% two-sided (or a 95% one-sided) confidence interval (CI), for the difference in baseline corrected mean QTc (heart-rate corrected version of QT) between drug and placebo at each time-point, and to conclude non-inferiority if the upper limit for each CI is less than 10 ms. The intent of the ICH E14 guidelines is to establish that the mean effect of the drug is less than 5 ms and the standard approach may not be well suited to achieve this goal. In this paper, we propose a novel Bayesian approach to address this problem directly keeping in line with the intent of the ICH E14 guidelines. We assess the performance of our proposed approach using simulated data, discuss its advantages over the standard approach, and illustrate the method by applying it to a real data set obtained from a GlaxoSmithKline (GSK) conducted thorough QT study.

Keywords: Bayesian inference; ICH E14; MC simulations; QT/QTc; thorough QT/QTc Study.

[*]Correspondence: Suraj Anand, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA; E-mail: spanand@ncsu.edu

1 Introduction

An undesirable property associated with some non-antiarrhythmic drugs is their ability to delay cardiac repolarization, more generally known as QT prolongation. The QT interval is a segment of the surface electrocardiogram (ECG) and represents the duration of ventricular depolarization and subsequent repolarization, and is measured from the beginning of the QRS complex to the end of the T wave. QT values are correlated with heart rate and hence a corrected version (QTc) is used for data analysis (see International Conference of Harmonization (ICH) E14 guidelines, available at `www.fda.gov/cder/guidance/`). The ICH E14 guidelines recommend conducting a 'thorough QT/QTc study' to determine whether the drug has a threshold pharmacological effect on cardiac repolarization, as detected by QTc prolongation. The study is typically carried out in healthy volunteers (as opposed to individuals at increased risk of arrhythmias) and is used to determine whether or not the effect of a drug on the QTc interval in target patient populations should be studied intensively during later stages of drug development.

The standard approach to investigating a drug for its potential for QTc prolongation is to construct a 90% two-sided confidence interval (CI), for the difference in mean QTc between drug and placebo at each time-point, and to conclude non-inferiority if the upper limits for all these CIs are less than a pre-specified constant (Patterson, Jones and Zariffa, 2005). A review of statistical design and analysis in thorough QT studies can be found in Patterson et al., 2005. More recently, Journal of Biopharmaceutical Statistics issued a special edition with a wide range of articles on statistical issues in design and analysis of thorough QT studies (Volume 18, Issue 3, 2008). Under the standard normality assumptions, the standard approach corresponds to both intersection-union test and the likelihood ratio test but is conservative and biased (Patterson, Jones and Zariffa (2005), Eaton et al. (2006)). However, this approach is conceptually simple to understand and easy to implement, and being conservative it goes well with the regulatory agencies. The industry needs better approaches that are less stringent while safeguarding the interests of the consumer and regulatory agencies.

According to the ICH E14 guidelines,

> "a negative 'thorough QT/QTc study' is one in which the upper bound of the 95% one-sided confidence interval for the largest time-matched mean effect of the drug on the QTc interval excludes 10 ms."

Hence, an alternative to the standard approach is to base the non-inferiority inference on the largest time-matched mean difference in population mean QTc between drug and placebo. Constructing an interval estimate for the maximum difference however is a non-trivial task. Recent attempts to address this problem in the context of thorough QT studies include articles by Eaton et al. (2006) and Boos et al. (2007), and both articles provide ingenious frequentist methods for the construction of an approximate large-sample based interval estimate for the maximum difference. Recently, Anand and Ghosh (2009) provided a Bayesian method for constructing a finite-sample based credible interval for the maximum difference parameter, based on a class of conjugate priors.

The major issue with QTc assessment is the correct interpretation and implementation of the ICH E14 guidelines, that appear to have a disconnect between the intent and the definition of a negative thorough QT study. While the definition of a negative thorough QT study is based on comparision of an upper bound of the 95% one-sided confidence interval for the largest time-matched mean effect of the drug with the threshold of 10 ms, the actual intent appears to be ruling out a drug with a mean effect of over 5 ms. The guidelines state that,

> "It is difficult to determine whether there is an effect on the mean QT/QTc interval that is so small as to be of no consequence. However, drugs that prolong the mean QT/QTc interval by around 5 ms or less do not appear to cause TdP."

and

> "This definition is chosen to provide reasonable assurance that the mean effect of the study drug on the QT/QTc interval is not greater than around 5 ms. When the largest time-matched difference exceeds the threshold, the study is termed 'positive'. A positive study influences the evaluations carried out during later stages of drug development, but does not imply that the drug is pro-arrhythmic."

If we denote by μ_{1k} and μ_{2k} the mean baseline corrected QTc values at time-point $k = 1, \ldots, p$, for drug and placebo respectively, then the actual object of interest indicated by the above excerpts from the ICH E14 guidelines appears to be an assessment of the chance of the event $\bigcap_{k=1}^{p} \{\mu_{1k} - \mu_{2k} \leq 5\}$, which is equivalent to the event $\{\max_{1 \leq k \leq p} (\mu_{1k} - \mu_{2k}) \leq 5\}$. If we denote the true largest time-matched mean difference by $\theta = \max_{1 \leq k \leq p} \{\mu_{1k} - \mu_{2k}\}$, then the actual event of interest is to know whether $\theta \leq 5$. To declare a thorough QT study negative one would want this event to have a high probability if θ is considered random.

Under the frequentist paradigm, θ is considered as a fixed parameter so it does not make sense to consider evaluating the probability of such an event. This is the reason why the current ICH E14 definition of a *negative* thorough QT study focusses on constructing an intervel estimate for θ and comparing the upper limit to a regulatory threshold (currently set at 10 ms). This is an indirect way of addressing the real problem at hand and may not meet the intent of the guideline. One might acheive a 95% one-sided upper confidence limit (UCL) for θ below 10 ms even with a mean effect larger than 5 ms by rigorously controlling the variability of QTc measurements. In a short simulation study, we repeatedly computed the 95% one-sided UCLs for θ using the standard method for a parallel design (described in more details in Section 3), by generating data sets with a between subject variability of 7 ms and an overall sample size of 200 subjects, for different values of θ. Figure 1 presents a plot of the median value (over 1000 simulated data sets) of the 95% one-sided UCLs for θ versus the true value of θ. It is easy to see that median of the 95% one-sided UCLs is less than the regulatory threshold of 10 ms even for relatively large values of θ (e.g., $\theta = 7, 8$) as compared to the regulatory threshold of $\theta = 5$.

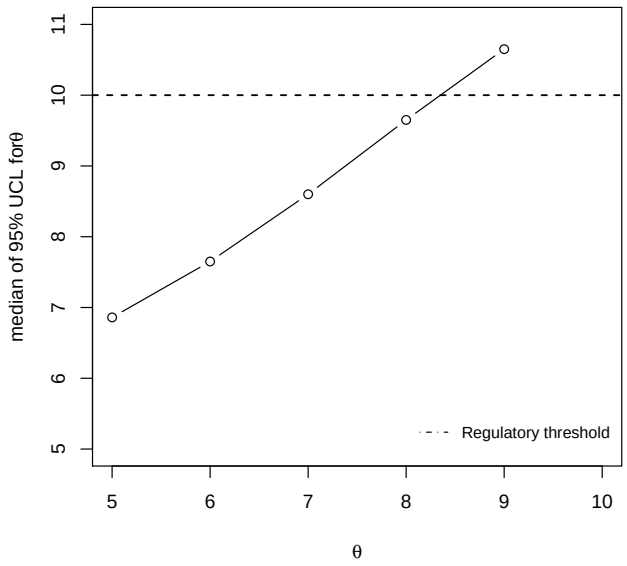

Figure 1: The median values of the 95% one-sided UCLs versus the true values of θ based on 1000 simulated data sets.

Additionally, Anand, Murray and Koch (2008) provide sample size calculations for thorough QT studies for both parallel and crossover studies based on multivariate methods similar to the standard approach. In their simulated examples, they found out that with a parallel study, the sample size required to show non-inferiority to 10 ms when the true maximum (θ) is 5 ms may be quite large. However, for a crossover study, even with a moderate sample size of 50, a power of over 90% could be obtained to establish non-inferiority when the true maximum was 5 ms, with within subject variability as small as 7 ms for the *hill* and the *steady state* mean structures. Further details about these mean structures are given in Section 3. Such a value for the within subject variability is not unusual and can be obtained by employing rigorous control over study conditions to ensure precise measurement of QTc effects. This defeats the intent of the ICH E14 guidelines that try to ensure with reasonable confidence that the mean effect of the study drug on the QT/QTc interval is not greater than 5 ms.

A more direct way to address this problem could be formulated under a Bayesian paradigm. Under the Bayesian paradigm, the true largest time-matched mean difference (θ) can be considered as a random variable and once we obtain the posterior density of θ under suitable specification of priors, it is a relatively straightforward procedure to compute the probability of the actual event of interest $\mathrm{P}(\theta \leq 5|data)$. Although a Bayesian counterpart of the current frequentist definition of a *negative* thorough QT study could also be easily formulated under a Bayesian paradigm by comparing the $95th$ percentile of the posterior distribution of θ with the regulatory threshold of 10 ms, we argue that such an approach does not meet the actual intent of the ICH E14 guidelines. Recently, Anand and Ghosh (2009) proposed a Bayesian

approach to solving this problem by constructing a posterior interval estimate for θ based on independent samples generated from its posterior density. Furthermore, the entire posterior cumulative distribution function of θ can be obtained to make inferences about the true maximum time-matched mean difference (θ).

In this paper, we propose a *direct* Bayesian approach to assess the risk of QT prolongation that is completely in line with intent of the ICH E14 guidelines. We provide a description of the Bayesian framework for analyzing QTc prolongation in Section 2. In Section 3, we use simulated data to assess the validity of this approach in a variety of cases, and in Section 4 we apply the proposed method to a real data set on QTc, obtained from a thorough QT study conducted at GlaxoSmithKline (GSK). We present the main conclusions in Section 5 and finally, in Section 6, we present some discussions of the future research in this area.

2 A Bayesian Framework for QTc Data

We develop a general statistical model for the parallel study design for our data. Suppose n_1 and n_2 denote the number of subjects receiving drug and control (placebo) respectively, with QTc (baseline corrected) calculated at p time points. We denote the vector of p measurements taken on the subjects on drug by $\boldsymbol{x}_i$, where $i = 1, \ldots n_1$, and those on control by $\boldsymbol{y}_j$, where $j = 1, \ldots n_2$, respectively. The measurements are assumed to be distributed identically and independently *(iid)*, arising from two independent p-variate normal distributions (denoted by $\mathcal{N}_p$) having a common covariance structure, and can be represented as follows:

$$\begin{aligned} \boldsymbol{x}_i &\overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \quad i = 1, \ldots, n_1, \text{ and} \\ \boldsymbol{y}_j &\overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}), \quad j = 1, \ldots, n_2, \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the population means for the drug and the control groups, respectively, and the common covariance matrix $\boldsymbol{\Sigma}$ is assumed to be unknown and unstructured. Let $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ be the vector of differences in mean QTc between drug and control. Then the parameter of interest which is the maximum difference between the two groups is denoted by

$$\theta = \max_{1 \le k \le p} \delta_k \equiv \max_{1 \le k \le p} (\mu_{1k} - \mu_{2k}). \tag{2}$$

The goal of a QT study is to obtain statistical inference about θ. Notice that θ is a non-differentiable function of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and hence we cannot use the standard delta-method (Casella and Berger, 2002, p.245) based on the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{x}}$ and $\hat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{y}}$, where $\bar{\boldsymbol{x}} = \sum_i \boldsymbol{x}_i / n_1$ and $\bar{\boldsymbol{y}} = \sum_j \boldsymbol{y}_j / n_2$ denote the sample means. It is well known (e.g., Anderson, 1984, p.167) that as $n \to \infty$,

$$\sqrt{n} \boldsymbol{S}^{-1/2} (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{d} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I}_p),$$

where $n = n_1 n_2 / (n_1 + n_2)$, $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2$ is the MLE of $\boldsymbol{\delta}$ and $\boldsymbol{S} = \{(\sum_{i=1}^{n_1} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^t + \sum_{j=1}^{n_2} (\boldsymbol{y}_j - \bar{\boldsymbol{y}})(\boldsymbol{y}_j - \bar{\boldsymbol{y}})^t)\}/(n_1 + n_2)$ is the MLE of $\boldsymbol{\Sigma}$. The standard method consists of computing the maximum of 95% one-sided upper confidence limits for δ_k's based on the above

asymptotic distribution of $\hat{\boldsymbol{\delta}}$. Marcheselli (2000) provides a generalization of the regular delta method to obtain the asymptotic distribution of a function $g(\boldsymbol{\delta})$ (e.g., θ in eq. (2)), that is not necessarily a differentiable function of $\boldsymbol{\delta}$ but satisfies some weak regularity conditions, which can be shown to have a non-Gaussian limiting distribution. But the sample sizes in thorough QT studies may not be large enough to justify the use of such asymptotic results. Cheng et al. (2008) provide arguements for asymptotic normality of the maximum difference random variable under a compound symmetric covariance structure. They also propose a small sample correction to obtain mean and variance of such an estimator in small sample settings. But they use normality for the test which may not be a valid assumption in small samples.

Alternatively, within a Bayesian framework, in order to obtain the posterior distribution of θ we first derive a class of conjugate priors for $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$, and then derive the posterior distribution of θ from the closed form posterior distribution of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$. It is well known that the following class of prior distributions is conjugate for the two-sample problem (Box and Tiao, 1973, chap.8),

$$\begin{aligned} \boldsymbol{\mu}_1 \mid \boldsymbol{\Sigma} &\sim \mathcal{N}_p(\boldsymbol{\mu}_{10}, \frac{1}{n_{10}}\boldsymbol{\Sigma}), \\ \boldsymbol{\mu}_2 \mid \boldsymbol{\Sigma} &\sim \mathcal{N}_p(\boldsymbol{\mu}_{20}, \frac{1}{n_{20}}\boldsymbol{\Sigma}) \text{ and} \\ \boldsymbol{\Sigma}^{-1} &\sim \mathcal{W}_p(a_0, \boldsymbol{B}_0), \end{aligned} \tag{3}$$

where $\mathcal{W}_p$ denotes a p-dimensional Wishart distribution with degrees of freedom a_0 and scale matrix $\boldsymbol{B}_0$ (Anderson, 1984, p.245). In the absence of prior information we can use the reference prior (Bernardo, 1979) which turns out to be the Jeffreys prior given by

$$\pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}. \tag{4}$$

Notice that the Jeffreys prior (4) can be obtained as a limiting case of the conjugate prior (Geisser and Cornfield, 1963) described in (3) by setting $\boldsymbol{\mu}_{k0} = 0$ and letting $a_0 \to 0$ and $n_{k0} \to 0$ for $k = 1, 2$.

Thus, it is enough to obtain the posterior density of θ under the conjugate prior (3) and then take suitable limits to obtain the posterior distribution under the Jeffreys prior (4). By conjugacy it follows that the joint posterior distribution of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ is given by

$$\begin{aligned} \boldsymbol{\mu}_1 \mid \boldsymbol{\Sigma}, \boldsymbol{X}, \boldsymbol{Y} &\sim \mathcal{N}_p\left(\frac{n_{10}\boldsymbol{\mu}_{10} + n_1\bar{\boldsymbol{x}}}{n_{10}+n_1}, \frac{1}{n_{10}+n_1}\boldsymbol{\Sigma}\right), \\ \boldsymbol{\mu}_2 \mid \boldsymbol{\Sigma}, \boldsymbol{X}, \boldsymbol{Y} &\sim \mathcal{N}_p\left(\frac{n_{20}\boldsymbol{\mu}_{20} + n_2\bar{\boldsymbol{y}}}{n_{20}+n_2}, \frac{1}{n_{20}+n_2}\boldsymbol{\Sigma}\right) \text{ and} \\ \boldsymbol{\Sigma}^{-1} \mid \boldsymbol{X}, \boldsymbol{Y} &\sim \mathcal{W}_p(a_0 + n_1 + n_2, \boldsymbol{B}), \end{aligned} \tag{5}$$

where $\boldsymbol{B}^{-1} = {\boldsymbol{B}_0}^{-1} + (n_1+n_2)\boldsymbol{S} + \frac{n_{10}n_1}{n_{10}+n_1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_{10})(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_{10})^t + \frac{n_{20}n_2}{n_{20}+n_2}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}_{20})(\bar{\boldsymbol{y}} - \boldsymbol{\mu}_{20})^t$ and $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote the list of observed $\boldsymbol{x}_i$'s and $\boldsymbol{y}_j$'s, respectively. It follows from (5) that the marginal posterior distribution of $\boldsymbol{\delta}$ is a p-variate t distribution (Anderson, 1984, p.283) and we present the result as the following Lemma:

Lemma 2.1. *If* $\boldsymbol{\delta}|\boldsymbol{A} \sim \mathcal{N}_p(\boldsymbol{\delta}_0, \boldsymbol{A}^{-1})$, *and* $\boldsymbol{A} \sim \mathcal{W}_p(m, \boldsymbol{A_0}^{-1})$, *where* $m > p-1$ *and* $\boldsymbol{A_0}$ *is a* $p \times p$ *positive definite symmetric matrix, then* $\boldsymbol{\delta} \sim t_p(m-p+1, \boldsymbol{\delta}_0, \frac{1}{m-p+1}\boldsymbol{A_0})$, *where* $t_p(m, \boldsymbol{\delta}, \boldsymbol{A})$ *denotes a* p*-variate* t *distribution with degrees of freedom* m, *location* $\boldsymbol{\delta}$ *and scale matrix* $\boldsymbol{A}$.

Lemma 2.1 shows that multivariate t can also be represented as a matrix mixture of multivariate normal variates, which extends the well known result that multivariate t can be represented as a scaler mixture of multivariate normal variates. The proof of Lemma 2.1 is provided in the Appendix. Although Lemma 2.1 provides a closed form posterior distribution for $\boldsymbol{\delta}$, it is not straightforward to obtain the posterior distribution of θ analytically. However, the actual object of interest $p_{neg} = P(\theta \leq 5|data)$, where p_{neg} is the probability of declaring a thorough QT study negative, does not necessarily require sampling from the posterior distribution of θ. Notice that

$$P(\theta \leq t|\boldsymbol{X}, \boldsymbol{Y}) = P(\delta_1 \leq t, \delta_2 \leq t, \ldots, \delta_p \leq t|\boldsymbol{X}, \boldsymbol{Y}), t \in \mathbb{R},$$

which is the posterior cummulative distribution function (cdf) of $\boldsymbol{\delta}$ at $t\mathbf{1}_p$, which by Lemma 2.1 has a p-variate t distribution given by

$$\boldsymbol{\delta}|\boldsymbol{X}, \boldsymbol{Y} \sim t_p(\nu, \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\Sigma}}), \tag{6}$$

where the degrees of freedom $\nu = n_1 + n_2 + a_0 - p + 1$, location parameter $\tilde{\boldsymbol{\delta}} = \frac{n_{10}\boldsymbol{\mu}_{10} + n_1\bar{\boldsymbol{x}}}{n_{10}+n_1} - \frac{n_{20}\boldsymbol{\mu}_{20} + n_2\bar{\boldsymbol{y}}}{n_{20}+n_2}$, and dispersion matrix $\tilde{\boldsymbol{\Sigma}} = \frac{1}{\nu}(\frac{1}{n_{10}+n_1} + \frac{1}{n_{20}+n_2})\boldsymbol{B}$. So, p_{neg} can be calculated directly using the cdf of a p-variate t distribution with the parameters mention above. Arellano-Valle and Genton (2008) derived the exact probability density function of the maximum of arbitrary absolutely continuous dependent random variables and they provide an example based on the multivariate t distribution which can be used to derive the exact posterior distribution of θ. However, numerical algorithms are still needed to obtain the cdf of such a distribution. So, we use the computational methodology proposed by Genz and Bretz (2002) to compute p_{neg} numerically, using the cdf of a p-variate t distribution as explained above. This has been implemented using the `pmvt` function available in the `R` package `mvtnorm`. The underlying idea employed by the function is that the cdf of multivariate t can be expressed as an one dimensional integral involving a multivariate normal probability. Computations are then made using randomized quasi-Monte Carlo methods as described in Genz and Bretz (2002). It should be noted that the order of numerical errors for this method are generally less than 10^{-5} for most applications.

Alternatively, we can also use a simple Monte Carlo (MC) sampling scheme to obtain iid samples from the (marginal) posterior distribution of θ given $\boldsymbol{X}$ and $\boldsymbol{Y}$ as follows.

For $l = 1, 2, \ldots$ generate

step 1: $\boldsymbol{\delta}^{(l)}|\boldsymbol{X}, \boldsymbol{Y} \sim t_p(\nu, \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\Sigma}})$,

step 2: compute $\theta^{(l)} = \max_{1 \leq k \leq p} \boldsymbol{\delta}^{(l)}$.

It follows that $\theta^{(l)}$'s generated by the above two steps are *iid* samples from the posterior distribution of θ given $\boldsymbol{X}$ and $\boldsymbol{Y}$. Thus, we can compute any desired posterior summary as an estimate

of θ using the sequence of *iid* Monte Carlo samples $\{\theta^{(l)} : l = 1, 2, \ldots\}$. For example, we could compute a 90% highest posterior density (HPD) region, or a 95% quantile for the posterior density of θ if we were interested in an interval estimate for θ. The main quantity of interest, p_{neg} can be easily estimated easily based on a finite MC sample $\{\theta^{(l)} : l = 1, \ldots, N\}$ by calculating the proportion of times $\theta^{(l)} \leq 5$, where the MC sample size N is chosen so that the overall MC error is less than some pre-assigned value $\epsilon > 0$. More specifically, $\hat{p}_{neg} = \sum_{l=1}^{N} I[\theta^{(l)} \leq 5]/N$, where $I[\cdot]$ denotes the indicator function, provides an arbitrarily close estimate of p_{neg} if we choose N large enough. For example, choosing $N > \frac{1}{\epsilon^2}$ ensures that $\hat{p}_{neg}$ is within $\pm\epsilon$ of p_{neg} with 95% probability.

3 Simulation Studies

We conducted a Monte Carlo simulation study to assess the performance of our proposed approach. For the purpose of simulations, we used a parallel study setup with $p = 10$, and true values of $\boldsymbol{\mu}_1 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ and $\theta = 2, 5, 7$. Depending on the value of θ we assessed the performance of our approach for different structures of the time-matched mean difference vector $\boldsymbol{\delta}$. More specifically, we performed the simulations for three mean structures: (i) *constant* mean, (ii) *hill* mean and (iii) *steady state* mean structure. In the *constant* mean structure, the baseline corrected QTc is expected to remain constant across the time-points. In the *hill* mean structure, the baseline corrected QTc is expected to initially increase and then decrease after reaching the maximum, whereas, in the *steady state* mean structure, it is expected to increase initially and then stabilize after reaching the maximum. Although the *hill* mean and the *steady state* mean structures appear to be the most plausible ones (Anand et. al. (2008)) we also provide results for the *constant* mean structure as a worst case scenario. Our proposed method does not require $\boldsymbol{\Sigma}$ to have a specific covariance structure but for simplicity we performed the simulations under an exchangeable covariance structure for the true $\boldsymbol{\Sigma} = \sigma^2[\rho\boldsymbol{I} + (1-\rho)\mathbf{1}\mathbf{1}^t]$. However, we do not use the exchangeable form while obtaining the posterior distribution of θ but rather use a Wishart prior for $\boldsymbol{\Sigma}^{-1}$.

We used a class of normal conjugate priors for $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ with the hyperparameters $\boldsymbol{\mu}_{k0} = 0$, and tuning parameters $n_{k0} = 0.001$ for $k = 1, 2$, and a Wishart prior for $\boldsymbol{\Sigma}^{-1}$ with the hyperparameters $a_0 = p + 2$ and $\boldsymbol{B}_0 = \boldsymbol{I}$ (see eq. (3)) to obtain the corresponding closed form posterior distribution of $\boldsymbol{\delta}$ (as in (6)). We conducted the Monte Carlo simulation study for different combinations of the operating characteristic drivers, namely, θ, n_1, n_2, σ, and ρ, under the mean structures as described above. We vary $n = n_1 + n_2 \in \{100, 150, 200\}, n_1 = n_2 = n/2$, $\theta \in \{2, 5, 7\}$, $\rho \in \{0.5, 0.8\}$ and $\sigma \in \{7, 12\}$. These operating characteristics provide a $3\times3\times2\times2$ experimental design and we computed p_{neg} for all possible 36 combinations of the operating characteristics. We used the `pmvt` function available in the `R` package `mvtnorm` to obtain a numerical estimate for $p_{neg} = P(\theta \leq 5|data)$ as described in Section 2. We repeated this procedure $B = 1000$ times to obtain an MC estimate of p_{neg} as $\bar{p}_{neg} = (\sum_{b=1}^{B} \hat{p}_{neg}^{(b)})/B$, where $\hat{p}_{neg}^{(b)}$ denotes the numerical estimate of $p_{neg}^{(b)}$ based on the bth simulated dataset. We present the MC average value, $\bar{p}_{neg}$ in Table 1 for various operating characteristics.

Table 1: MC average values of estimated $p_{neg} = P[\theta \leq 5 \mid \boldsymbol{X}, \boldsymbol{Y}]$ for different values of operating characteristics.

				Total Sample Size (n)		
θ	$\boldsymbol{\delta}$	σ	ρ	$n = 100$	$n = 150$	$n = 200$
2	$(0, 0.5, 1, 1.5, 2, 2, 1.5, 1, .5, 0)$	7	0.5	0.84	0.92	0.96
	(Hill)		0.8	0.89	0.94	0.97
		12	0.5	0.53	0.62	0.72
			0.8	0.67	0.76	0.81
	$(0, 0, 0.25, 0.5, 1, 1.5, 1.75, 2, 2, 2)$	7	0.5	0.80	0.90	0.95
	(Steady state)		0.8	0.88	0.93	0.97
		12	0.5	0.51	0.61	0.68
			0.8	0.65	0.73	0.79
	$(2, 2, 2, 2, 2, 2, 2, 2, 2, 2)$	7	0.5	0.70	0.83	0.90
	(Constant)		0.8	0.84	0.89	0.94
		12	0.5	0.41	0.50	0.56
			0.8	0.58	0.66	0.73
5	$(1, 2, 3, 4, 5, 5, 4, 3, 2, 1)$	7	0.5	0.25	0.28	0.28
	(Hill)		0.8	0.40	0.39	0.40
		12	0.5	0.21	0.23	0.23
			0.8	0.34	0.35	0.37
	$(5, 5, 5, 5, 5, 5, 5, 5, 5, 5)$	7	0.5	0.09	0.09	0.09
	(Constant)		0.8	0.23	0.24	0.24
		12	0.5	0.08	0.09	0.09
			0.8	0.23	0.23	0.22
7	$(2, 3, 4, 5, 6, 7, 6, 5, 4, 3)$	7	0.5	0.05	0.04	0.03
	(Hill)		0.8	0.11	0.08	0.06
		12	0.5	0.08	0.07	0.06
			0.8	0.18	0.16	0.15

Probabilities for the Steady state mean structure for $\theta = 5$ were similar to those for the Hill mean structure and hence omitted.

Results are based on $B = 1000$ simulations with the largest MC standard error not exceeding 0.01, across all operating characteristics.

It is clear from Table 1 that p_{neg} increases with ρ. It is interesting to note that while p_{neg} increases with the overall sample size (n) for $\theta \leq 5$, it appears to decrease with n for $\theta > 5$ as desired. It can also be seen that while p_{neg} decreases with σ for $\theta \leq 5$, it appears to increase with σ for $\theta > 5$ as expected. Our choice of values for the operating characteristics were chosen so as to accommodate the plausible values as expected in a realistic parallel study design. In a carefully designed thorough QT study, with rigorous control in the variability of QTc measurements, the σ values are expected to be less than 10 ms, whereas, the correlation values (ρ) between QTc measurements taken on the same individual are widely expected to be in the range of (0.5, 0.9) (personal communication from a GSK scientist). The probabilities for the *hill* and the *steady state* mean structures are comparable and are much larger than those for the *constant* mean structure.

The results reported in Table 1 suggest that the Bayesian method is fairly strict for large values of θ. The p_{neg} values increase with the sample size for $\theta \leq 5$ but the increase is gradual and seems to level off at its asymptotic value for large sample size. Given the strict nature of the Bayesian method, a p_{neg} of 0.5 or above appears to provide reasonable assurance that the thorough QT study is successfully negative. One should note that p_{neg} can also be used to construct a test statistic within a frequentist paradigm. If we use a cutoff value for p_{neg}, say p_{crit}, then the empirical probability of incorrectly declaring a thorough QT study negative when true value of $\theta > 5$, i.e., $P[p_{neg} > p_{crit}|\theta > 5]$ is akin to type I error rate for the null hypothesis $H_0 : \theta > 5$, whereas, $P[p_{neg} > p_{crit}|\theta < 5]$ is similar to power in the frequentist sense. In practice, for any given configuration of the operating characteristics such a cutoff may be choosen based on simulations such that it maintains a linear combination of the type I and type II error rates to a given level.

Though our Bayesian method cannot be compared directly to the standard method, we also conducted a small case simulation study to evaluate these methods for their ability to correctly declare a thorough QT study negative. The empirical probability of declaring a thorough QT study negative for the Bayesian method $P[p_{neg} > p_{crit}|\theta]$ was estimated by $\sum_{b=1}^{B} I[\hat{p}_{neg}^{(b)} > p_{crit}]/B$, when we set the true value of θ to 2 and 7. We chose to use $p_{crit} = 0.5$ for the purpose of our illustration. Simulations for the standard method were also conducted to match a few of the operating characteristic settings that we used for our proposed Bayesian method. The empirical probability of declaring a thorough QT study negative for the standard method was estimated by calculating the proportion of times the largest two-sided upper 90% confidence limit (obtained from the large sample distribution of $\hat{\boldsymbol{\delta}}$) was less than the regulatory threshold of 10 ms. It should be noted that a two-sided upper 90% confidence interval is equivalent to a 95% one-sided upper confidence interval under the standard normality assumptions. Table 2 summarizes these results.

These results reveal the somewhat unexpected patterns of the empirical probabilities based on standard method. First, the results reported in Table 2, clearly indicate that for the Bayesian method the empirical probability of declaring a thorough QT study negative improves with an increase in sample size for $\theta \leq 5$ but it decreases for $\theta > 5$. On the other hand, the standard method yields a larger empirical probability (of declaring a thorough QT study negative) for larger sample sizes regardless of the true mangitude of θ. In other words, even with moderate increase in sample sizes, the empirical probability of declaring a thorough QT study negative

Table 2: Empirical probability of declaring a thorough QT study negative using Bayesian and Standard methods

					Empirical Probability	
θ	$\boldsymbol{\delta}$	σ	ρ	n	Bayesian	Standard
2	$(2,2,2,2,2,2,2,2,2,2)$	7	0.8	100	0.90	1.00
				150	0.97	1.00
				200	0.99	1.00
7	$(2,3,4,5,6,7,6,5,4,3)$	7	0.8	100	0.06	0.67
				150	0.03	0.82
				200	0.01	0.92

Results are based on $B = 1000$ simulations with the MC margin of error not exceeding 0.03.

increases rapidly even for a relatively large value of θ (e.g., $\theta = 7$). Thus, by using the standard method, there is a high probability of declaring a thorough QT study negative with a large sample size when in fact the true value of $\theta > 5$, thus posing a higher risk of incorrectly declaring a thorough QT study negative. On the contrary, by using the proposed Bayesian method, the chances of incorrectly declaring a thorough QT study negative are very small and they decrease with larger sample sizes. Under the above set of operating characteristics, our results in Tables 1 and 2 indicate that $p_{crit} = 0.5$ appears to be a reasonable choice as a cutoff value for p_{neg}. In other words, to test the null hypothesis $H_0 : \theta > 5$, we may use the test that rejects H_0 when $p_{neg} > p_{crit} = 0.5$. Additional simulation studies (not reported here) suggest that such a test will have reasonably good performance in a variety of cases. The `R` code to carry out such a test is available from the first author upon request.

4 Application to a Thorough QT Study

The data used in this paper were obtained from a GSK conducted thorough QT study with placebo, a positive control *(Moxifloxacin)*, and multiple drug concentration arms for the study drug. QTc *(Fridericia)* measurements were taken during the 24 hour baseline period (mean of triplicates) and after dosing at multiple time-points ($p = 9$ in eq. (1)). Although the actual study design was crossover, only data from a single period of each sequence were used and only data on placebo and the drug arm with the supratherapeutic dose (called drug henceforth) were included to mimic a parallel study design, with sample sizes $n_1 = 23$ and $n_2 = 20$ subjects in the placebo and drug arms respectivily, for the purpose of our illustration. The end-point of interest was the change from baseline QTc values and the parameter of interest was the maximum time-matched mean difference between drug and placebo for baseline corrected QTc.

The primary aim of this research was to obtain an estimate for $p_{neg} = P(\theta \leq 5|data)$, i.e., the probability of declaring the thorough QT study negative. As in the case of analyzing simulated data, we used a class of normal conjugate priors for $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ with the hyperparameters $\boldsymbol{\mu}_{k0} = 0$, and tuning coefficients $n_{k0} = 0.001$ for $k = 1, 2$, and a Wishart prior for $\boldsymbol{\Sigma}^{-1}$

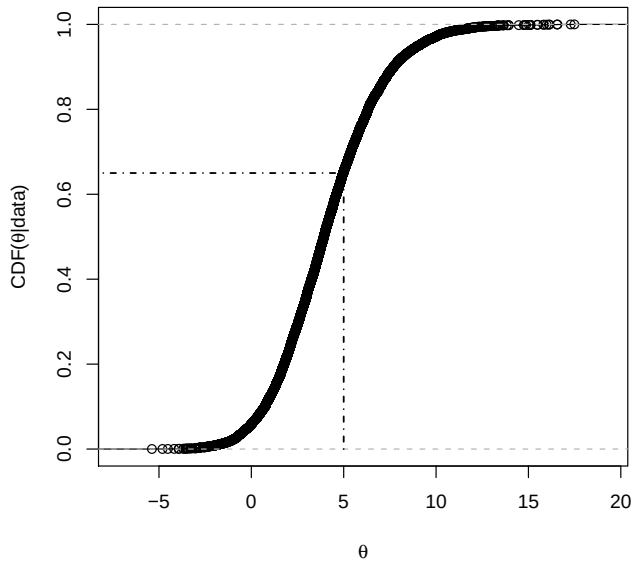

Figure 2: Posterior cdf of θ given $\boldsymbol{X}$ and $\boldsymbol{Y}$ based on the GSK data.

with the hyperparameters $a_0 = p + 2$ and $\boldsymbol{B}_0 = \boldsymbol{I}$ to obtain the corresponding closed form posterior distribution. To obtain a numerical estimate for p_{neg} we used the closed form posterior distribution of $\boldsymbol{\delta}$ and implemented it by using the `pmvt` function available in the `R` package `mvtnorm`. The resulting probability estimate was $\hat{p}_{neg} = 0.65$. In addition, we also present the numerical estimate of the posterior cdf of θ given $\boldsymbol{X}$ and $\boldsymbol{Y}$ in Figure 2, based on generating $N = 10,000$ MC samples from the posterior distribution of θ using the algorithm provided at the end of Section 2. By using $p_{crit} = 0.5$ as the cutoff value for p_{neg} we can declare this thorough QT study negative with reasonable confidence. For the sake of completeness, we also computed 90% two-sided confidence intervals for the data using the standard method. The standard method based on a mixed effect model with drug, time-point and drug by time-point interaction as fixed effects, and subject nested in drug as a random effect (confidence intervals calculated at each time-point), yielded the largest 90% two-sided upper confidence limit to be 8.13, which is smaller than the regulatory threshold of 10 ms indicating the same outcome. Figure 3 displays the individual 90% two-sided upper confidence intervals for the GSK data. It should be noted that the actual study was crossover and hence the confidence intervals for the overall data were much below the regulatory threshold of 10 ms owing to a smaller variability estimate from the full crossover design.

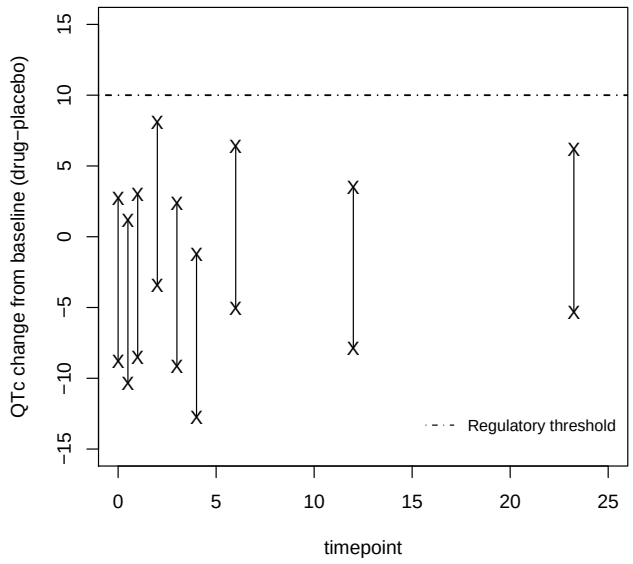

Figure 3: 90% two-sided confidence intervals for the time-matched mean differences (δ_k's) using the standard approach for the GSK data.

5 Conclusions

The Bayesian methodology that we have proposed in this paper is a novel approach, aimed at resolving the important problem arising in the assessment of the risk of QT prolongation in thorough QT studies. The standard method and the other recently proposed methods attempt to address this problem in an indirect way that may not meet the intent of the ICH E14 guidelines. This is a crucial issue especially for the case when the true value of θ is larger than 5 ms but the maximum of the 90% two-sided upper confidence limits turns out to be below 10 ms (e.g., by rigorously controlling the variability of QTc measurements). Our approach attempts to *directly* calculate the probability that the mean effect is no larger than 5 ms, thereby, providing a *direct measure of evidence* of whether the drug prolongs mean QTc beyond the tolerable threshold of 5 ms. Our approach is easy to implement and can be used efficiently to make inferences about a thorough QT study. The proposed statistical inference does not rely on any restrictive assumptions (e.g., $\mathbf{\Sigma}$ is allowed to be unstructured), it is not based on any large sample theory or approximations, and it can be used for any sample size (e.g., moderate values of $n = 50$ to large values of $n = 200$) and potentially for a large number of time points (e.g., $p = 20$). Finally, a simple `R` code (available from the first author) can be used to obtain $\hat{p}_{neg}$ and the entire posterior cdf of θ.

6 Discussion and Future Research

A Bayesian approach provides a promising alternative to the standard approach for analyzing a 'thorough QT/QTc study.' While the current ICH E14 definition of a negative thorough QT study mandates comparing the upper bound of the 95% one-sided confidence interval for θ with the threshold of 10 ms, the actual intent appears to be ruling out a drug with a mean effect of over 5 ms. Under the frequentist paradigm, as θ is considered to be a fixed but unknown parameter it does not provide a logical foundation to evaluate the probability of such an event (i.e., $P(\theta \leq 5)$). The standard frequentist interval estimate-based approach is an indirect way of addressing the real problem at hand and may not meet the intent of the regulatory guidelines. The upper 95% confidence limit for θ being less than 10 ms does not necessarily provide a reasonable guarantee that the mean effect of drug is less than 5 ms. In contrast, under the Bayesian paradigm, θ is considered to be a random variable and once we obtain the posterior density of θ under a relatively non-informative specification of priors, it is relatively straightforward to compute the probability of the actual event of interest, i.e., $\mathrm{P}(\theta \leq 5|data)$. In addition, we can obtain the entire posterior cdf of θ by using numerically efficient algorithms. It is to be noted that our proposed Bayesian methodology does not require the use of so-called Markov Chain Monte Carlo (MCMC) methods which are known to be computationally intensive.

As illustrated by extensive simulation studies (in Section 3), our proposed methodology appears to work fairly well under many plausible settings of the operating characteristics that are typically observed in a realistic study. For example, for small values of θ (e.g., $\theta = 2$) the probability of declaring the study negative is reasonably high as desired for almost all configurations, whereas, for large values for θ $(e.g., \theta = 5, 7)$, p_{neg} is quite small for most configurations under various mean structures. So, one could choose a cutoff value for p_{neg}, to be, say, 0.5 or some other high value, to rule out a thorough QT study as negative. This cutoff value could be decided upfront based on expert opinion and/or cost-benefit ratio and other statistical considerations. In the absence of additional information/considerations we recommend a value of 0.5 as an operational cutoff value, i.e., to declare a thorough QT study to be negative if $p_{neg} > 0.5$. More thorough studies to determine optimal sample sizes that achieve a given power using this cutoff value, are being currently investigated by these authors and the results will be reported elsewhere.

A cautious reader may question our selection of priors for this approach. We selected conjugate but relatively non-informative priors for simplicity, with extremely small values of n_{10} and n_{20} so as to draw very less information from the priors. We found the reported results to be rather insensitive to moderate variations of the priors in our simulation studies. Alternatively, the prior parameters can be updated to elicit available information based on historical data and scientific knowledge that may be available from the past and the concurrent investigations, to yield more realistic and informative priors. Though we did not formally explore this aspect in this project, it would be interesting to see how the posterior density of θ changes with changes to the prior parameters elicited from historical data (see Anand and Ghosh, 2008).

In our additional simulations (not reported here) we found out that the hyperparameter a_0

appears to have a substantial impact on p_{neg} and larger values of a_0 result into larger p_{neg}. To be on the conservative side we chose the lowest allowable value of $a_0 = p + 2$ to derive our results which in turn also leads to a relatively non-informative prior for $\boldsymbol{\Sigma}$. Another possibility is to estimate a_0 using the marginal likelihood obtained by integrating out all the parameters. However, such an empirical Bayes approach results into an increasing marginal likelihood of a_0. Yet another alternative would be to use information criteria (e.g., DIC) to choose an appropriate value of a_0 (see Anand and Ghosh, 2008).

Finally, the model in (1) assumes a common $\boldsymbol{\Sigma}$ for both arms (i.e., for both drug and placebo groups). Within our Bayesian framework, we can relax this assumption and allow different $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ for drug and placebo groups, respectively. The inference will then proceed with the priors, $\boldsymbol{\Sigma}_1^{-1}$, $\boldsymbol{\Sigma}_2^{-1} \overset{iid}{\sim} \mathcal{W}_p(a_0, \boldsymbol{B}_0)$ with the same choices for a_0 and B_0 that we used for our simulated and real data sets. The MC sampling algorithm described in Section 2 can also be extended to obtain MC samples from the posterior distribution of θ. Last but not the least, one may explore the possibility of relaxing the normality assumption in (1). This appears to be a nontrivial extension, especially if we choose to develop a non-asymptotic method to obtain a 95% upper interval estimate of θ (e.g., see Anand and Ghosh, 2008 for other flexible distributions). A fully non-parametric Bayes method based on normal mixtures of Dirichlet processes, may also be used to obtain $p_{neg} = P(\theta \leq 5|data)$, but such an approach may require computationally intensive methods.

References

Anand, S. P. and Ghosh, S. K. (2008). A Flexible Class of Models for Data Arising from a 'thorough QT/QTc study', In *JSM Proceedings*, Biopharmaceutical Section. (`www.stat.ncsu.edu/library/papers/ISMS2610.pdf`)

Anand, S. P. and Ghosh, S. K. (2009). A Bayesian approach for investigating the risk of QT prolongation, *Journal of Statistical Theory and Practice*, (To appear in **3(2)**).

Anand S. P., Murray S. C., Koch G. G. (2008). A multivariate approach to sample size calculations for Thorough QT studies, *Institute of Statistics Mimeo Series*, # **2608**. (`www.stat.ncsu.edu/library/papers/ISMS2608.pdf`)

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd ed.*, John Wiley and Sons, New York.

Arellano-Valle, R. B. and Genton M. G. (2008). On the exact distribution of the maximum of absolutely continuous dependent random variables, *Statistics & Probability Letters*, **78(1)**, 27-35.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference, *Journal of the Royal Statistical Society B*, **41**, 113-147.

Boos, D., Hoffman, D., Kringle, R., Zhang, J. (2007). New confidence bounds for QT studies, *Statistics in Medicine*, **26**, 3801-3817.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison Wesley, New York.

Casella, G. and Berger, R. L. (2002). *Statistical inference*, Duxbury Press, New York.

Cheng, B., Chow, S., Burt, D., Cosmatos, D. (2008). Statistical Assessment of QT/QTc Prolongation Based on Maximum of Correlated Normal Random Variables, *Journal of Biopharmaceutical Statistics*, **18(3)**, 494-501.

Eaton, M. L., Muirhead, R. J., Mancuso, J. Y., Lolluri, S. (2006). A confidence interval for the maximal mean QT interval change due drug effect, *Drug Information Journal*, **40**, 267-271.

Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters, *Journal of the Royal Statistical Society, Series B*, **25**, 368-376.

Genz, A. and Bretz, F. (2002), Comparison of Methods for the Computation of Multivariate Probabilities, *Journal of Computational and Graphical Statistics*, **11**, 950-971.

International Conference on Harmonisation (2005). Guidance for Industry: E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrythmic Potential for Non-Antiarrythmic Drugs. Available at: `www.fda.gov/cber/gdlns/iche14qtc.htm`

Marcheselli, M. (2000). A Generalized Delta Method with Applications to Intrinsic Diversity Profiles, *Journal of Applied Probability*, **37(2)**, 504-510.

Patterson, S., Agin, M., Anziano, R., et al. (2005). Investigating Drug-Induced QT and QTc Prolongation in the Clinic: A Review of Statistical Design and Analysis Considerations: Report from the Pharmaceutical Research and Manufacturers of America QT Statistics Expert Team, *Drug Information Journal*, **9**, 243-266.

Patterson, S., Jones, B., Zariffa, N. (2005). Modeling and interpreting QTc prolongation in clinical pharmacology studies, *Drug Information Journal*, **39**, 437-445.

Appendix

Lemma 2.1. Let $f(\mathbf{u}|\mathbf{v})$ denote the conditional density of a random variable $\mathbf{u}$ given a random variable $\mathbf{v}$. Notice that,

$$\begin{aligned} f(\boldsymbol{\delta}|\boldsymbol{A}) &= (2\pi)^{-p/2}|\boldsymbol{A}|^{-1/2}exp\{-\frac{1}{2}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t\boldsymbol{A}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\} \text{ and} \\ f(\boldsymbol{A}) &= 2^{-mp/2}(\boldsymbol{\Gamma}_p(m/2))^{-1}|\boldsymbol{A_0}|^{m/2}|\boldsymbol{A}|^{(m-p-1)/2}exp\{-\frac{1}{2}tr\boldsymbol{A}\boldsymbol{A_0}\}. \end{aligned}$$

Thus, the marginal density of $\boldsymbol{\delta}$ is given by

$$\begin{aligned} f(\boldsymbol{\delta}) &= \int_{\boldsymbol{A}} f(\boldsymbol{\delta} \mid \boldsymbol{A})f(\boldsymbol{A})d\boldsymbol{A} \\ &= \boldsymbol{C}_{p,m}|\boldsymbol{A_0}|^{m/2}\int_{\boldsymbol{A}}|\boldsymbol{A}|^{(m-p)/2}exp(-1/2)\{tr\boldsymbol{A}[\boldsymbol{A_0}+(\boldsymbol{\delta}-\boldsymbol{\delta}_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t]\}d\boldsymbol{A} \end{aligned}$$

(where $\boldsymbol{C}_{p,m} = (2\pi)^{-p/2}2^{-mp/2}(\boldsymbol{\Gamma}_p(m/2))^{-1}$)

$$\begin{aligned} &= \boldsymbol{C}_{p,m}|\boldsymbol{A_0}|^{m/2}\frac{2^{(m+1)p/2}}{(\boldsymbol{\Gamma}_p[(m+1)/2])^{-1}}|\boldsymbol{A_0}+(\boldsymbol{\delta}-\boldsymbol{\delta}_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t|^{-(m+1)/2} \\ &= \frac{\boldsymbol{\Gamma}_p[(m+1)/2]}{\boldsymbol{\Gamma}_p(m/2)}\left(\frac{1}{\pi}\right)^{p/2}|\boldsymbol{A_0}|^{m/2}|\boldsymbol{A_0}|^{-(m+1)/2}\left[1+(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t(\boldsymbol{A}_0)^{-1}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\right]^{-(m+1)/2} \end{aligned}$$

(since $|\boldsymbol{A_0}+(\boldsymbol{\delta}-\boldsymbol{\delta}_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t| = |\boldsymbol{A}_0||1+(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t(\boldsymbol{A}_0)^{-1}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)|$)

$$\begin{aligned} &= \frac{\boldsymbol{\Gamma}_p[(m+1)/2]}{\boldsymbol{\Gamma}_p(m/2)}\left(\frac{1}{\pi}\right)^{p/2}|\boldsymbol{A_0}|^{-1/2} \\ &\quad\left[1+\frac{1}{(m+p-1)}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t\left(\frac{1}{(m+p-1)}\boldsymbol{A}_0\right)^{-1}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\right]^{-((m-p+1)+p)/2} \\ &= \frac{\boldsymbol{\Gamma}_p[(m+1)/2]}{\boldsymbol{\Gamma}_p[(m-p+1)/2]}\left(\frac{1}{(m-p+1)\pi}\right)^{p/2}|\frac{1}{m+p-1}\boldsymbol{A_0}|^{-1/2} \\ &\quad\left[1+\frac{1}{(m+p-1)}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)^t\left(\frac{1}{(m+p-1)}\boldsymbol{A}_0\right)^{-1}(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\right]^{-((m-p+1)+p)/2}. \end{aligned}$$

Hence, it follows that
$\boldsymbol{\delta} \sim t_p(m-p+1, \boldsymbol{\delta}_0, \frac{1}{m-p+1}\boldsymbol{A_0})$. □