

**Measuring Service Accessibility in Telephone Nurse Triage Services**

Li Zeng  
Assistant Professor  
The University of Texas at Arlington  
Arlington, TX 76019

Ronnie Peterson  
Manager of Clinical Support  
University of Wisconsin Medical Foundation  
Middleton, WI 53563

**Abstract**

Telephone nurse triage (TNT) services is an important component of the health care delivery system. Thanks to the advancement of telephone/information technologies, abundant data on the daily operations of telephone triage centers are now readily available which can be used to assess their performance. However, the current way to use such data is very primitive, which mainly focuses on simple descriptive statistics of the data. This study provides an approach to mine information from count data from the TNT service, including the daily inbound and outbound call volumes. A class of overdispersed generalized linear models (OGLMs) is proposed to model the outbound-inbound relationship and Bayesian algorithms for model estimation and comparison are provided. The established model reveals the variation in the accessibility of the triage service and the effect of weekday/weekend, which will provide significant information for performance assessment and improvement in such services. In the case study, the proposed approach is applied to a dataset from a triage call center in Wisconsin to demonstrate its use in practice. A numerical study is also done to examine the properties of the proposed model and accuracy of parameter estimation.

**Keywords:** Telephone nurse triage services; Outbound calls; Service accessibility; Overdispersion; Bayesian inference.

**1. Introduction**

Telephone nurse triage (TNT) services have grown rapidly since the 1990s, as a means of demand management in healthcare<sup>1-3</sup>. Such services are designed to allow patients to speak directly with registered nurses about healthcare concerns via phone. The nurses, using computer-

accessed clinical protocols and judgment, assess the severity of the concerns and then provide recommendations/dispositions. Health-related information and provider referral may also be offered. The TNT services are an important component of the health care delivery system which promote cost-effective care as well as educate patients on self-care and the appropriate use of medical services<sup>4,5</sup>. It is estimated that today more than 100 million Americans have access to these services through a particular hospital or health maintenance organization (HMO), and they are getting more and more popular<sup>6,7</sup>.

A typical TNT service system can be described by the call routing path illustrated in Figure 1. 1. When a patient dials the triage center, the call will be first connected to a voice response unit (VRU) that queries on his/her needs, and then put into the *inbound* queue and served by a nurse. When no nurse is available to answer the call, the receptionists will record the contact number of the caller and place it into the *outbound* queue. Free nurses will serve these callers later by calling them back.

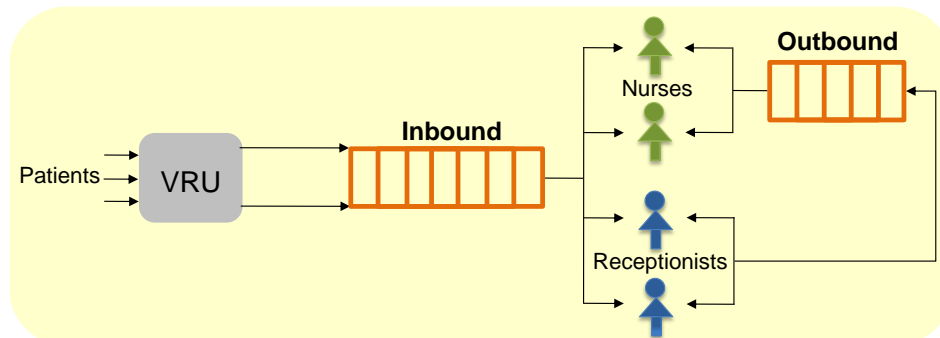


Figure 1. A typical telephone nurse triage service system

Thanks to the advancement of telephone and information technologies, together with the use of electronic medical records software, abundant measurements on the daily operations of telephone triage centers are now readily available, such as counts, duration and waiting time of calls. These data have been widely used in practice to assess the performance of the TNT service. Typically, simple descriptive statistics of the data, such as monthly/quarterly averages, are used as performance measures and compared to national or local standards to determine if the performance is satisfying. Obviously, such primitive analyses cannot fully extract information in

the data. Advanced data mining techniques need to be developed to mine the data in order to provide reliable support for performance assessment and improvement.

For this purpose, this paper provides an approach to analyze count data in the TNT service, including the number of inbound and outbound calls per unit period. In the current practice, the number of outbound calls is widely used as a measure of the service *accessibility*, which is one important dimension of the performance of TNT services. The intuition is that ideally all the calls should be answered immediately so that patients can receive timely advices on their problems, thus a larger number of outbound calls indicating worse service accessibility. However, the number of outbound calls alone may be misleading as it depends on the number of inbound calls: an increase in the outbound volume may be simply due to an increase in the volume of incoming calls and has nothing to do with the performance of the triage service. A real-world example is given in Figure 2 which shows daily inbound/outbound call volumes in a nurse triage center in Wisconsin. The data represent a common situation in practice where the demand for health services, represented by the inbound call volume, varies from day to day, and the outbound volume tends to fluctuate with it, as shown in Figure 2(a). Extreme cases often occur during pandemic periods, such as the H1N1 (swine flu) period in 2009, when outbound volumes increase dramatically because telephone triage becomes a better way to acquire health care than going to the hospital in person<sup>8</sup>. Figure 2(b) plots the inbound and corresponding outbound volumes, where the correlation between the two quantities is apparent.

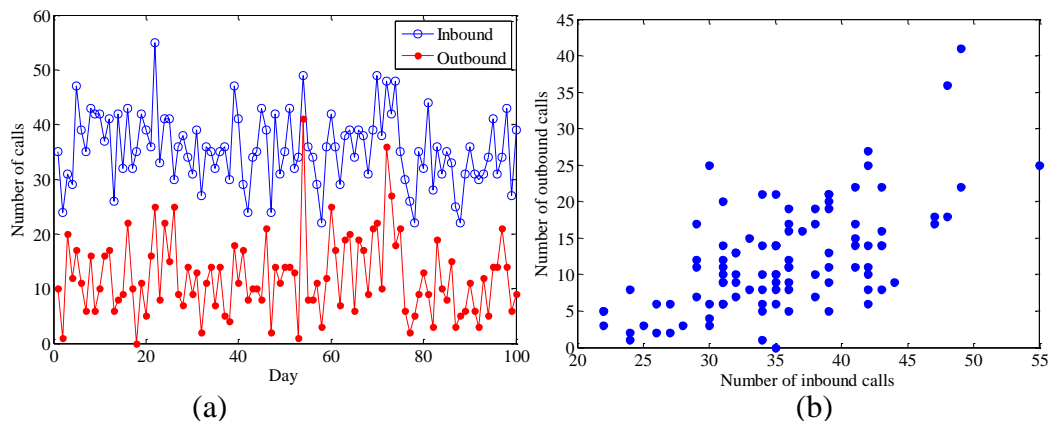


Figure 2. Daily volumes of inbound and outbound calls in a triage call center

For the above reason, the proposed approach focuses on modeling the relationship between the outbound and inbound call volumes. A class of overdispersed generalized linear model (OGLM) is developed, which characterizes the dependency of the outbound volume on the inbound volume and the effect of weekday/weekend on this dependency. Bayesian algorithms for model estimation and comparison are provided. In the case study, the proposed approach is applied to a dataset from a nurse triage center to demonstrate its use in practice. A numerical study is also done to examine the properties of the proposed models compared to conventional models and the performance of model estimation.

It deserves to mention that data mining/statistical analysis of operational data from call centers is an underdeveloped research area due to the lack of data collected at the level of individual calls rather than periodic aggregates<sup>9</sup>. Recently, Brown *et al.* make the first attempt on this topic using data from a bank call center<sup>10</sup>. Their study finds empirical models for each individual performance measures such as the number of incoming calls per hour and service time on each call. Our study differs from it in that we concern the relationship between two individual measures, the outbound and inbound call volumes, to provide better quantification for the service accessibility which is critical in health care services.

The proposed approach is very useful to practitioners in the TNT service industry: First, the proposed models reveal the variation in the call volumes, which provides a better understanding of the service system. Second, based on the established model, new measures of service accessibility can be defined for performance assessment, which are more reliable than the simple aggregates used in the current practice. In addition, the identified day-to-day variation and the weekday/weekend difference in service accessibility will also provide useful information for the scheduling of nurses to meet patient needs.

The remainder of the paper is organized as follows. Section 2 presents the conventional and proposed models for characterizing the outbound-inbound relationship. Section 3 describes the Bayesian algorithms for estimating the parameters of the proposed models and comparing the

models to determine the best model for a given dataset. Section 4 and 5 report the results of the case study and numerical study. Section 6 concludes this study and discusses future work.

## 2. Models for Outbound-Inbound Call Volumes

Let  $(y_i, x_i)$  be the observation on day  $i$  in a TNT center, where  $x_i$  is the number of inbound calls and  $y_i$  is the corresponding number of outbound calls. Assume the outbound volumes in different days are independent given the inbound volumes. This assumption is consistent with the common practice of telephone triage services where all arrived calls are required to be answered within the same day. Let the probability of making an outbound call be  $p$ , which represents the *accessibility rate* of the triage service system (more precisely, the inaccessibility rate as higher rate of outbound calls indicates lower rate of accessibility). The objective here is to build a statistical model to characterize the variation in  $p$ . In the follows, some conventional models for  $p$  are first introduced and then the proposed models are given.

### 2.1 Conventional models

The simplest model for the data is the Binomial model (BM) with a constant parameter  $p$

$$y_i \sim \text{Bin}(x_i, p) \quad (1)$$

This model rests on the assumption that the service accessibility maintains at a constant level from day to day. The mean of this model is

$$E(y_i) = x_i p$$

which is linearly dependent on the number of inbound calls. This is consistent with the relationship shown in Figure 2(b).

However, the constancy assumption of  $p$  may be too simple to describe the case in practice. More reasonably, the service accessibility depends on the inbound volume. The intuition is that when the call arrivals increase, the capacity of the service system to handle the calls tends to decrease, and thus the service becomes less accessible. The Generalized linear model (GLM)<sup>11</sup> can be used to describe the dependency of  $p$  on the inbound call volume

$$\begin{aligned} y_i &\sim \text{Bin}(x_i, p_i) \\ \eta(p_i) &= a + bx_i \end{aligned} \quad (2)$$

where  $p_i$  is the accessibility rate on day  $i$ ,  $\eta$  is the link function that connects  $p_i$  to a linear function of  $x_i$ , and  $a$  and  $b$  are the coefficients of the linear function. A popular choice of the link function is the logit function<sup>12</sup>

$$\eta(p_i) = \log \frac{p_i}{1-p_i} \Rightarrow p_i = \frac{e^{a+bx_i}}{1+e^{a+bx_i}}$$

The mean of the GLM is

$$E(y_i) = x_i \eta^{-1}(a + bx_i)$$

which is a complex nonlinear function of the inbound call volume.

## 2.2 The proposed models

The GLM may still not be adequate to characterize the variations in the call volume data as, in practical sense, the service accessibility may not depend on the inbound volume in a deterministic way; rather, the actual variations are likely to be larger than what is explained by the GLM due to the various variation sources existing in the service environment. In other words, there is possibly an *overdispersion* problem that often appears in model fitting in practice<sup>13-15</sup>. Moreover, the effect of weekday/weekend may also exist, that is, the service accessibility varies among weekdays and weekends. In particular, the weekday accessibility rate and weekend rate may be different as it is common in service industries that different schedules are followed on weekdays and weekends. Such variation should also be considered in the modeling.

To incorporate the weekday/weekend effect, we add another index,  $j, j=1, \dots, 7$ , corresponding to Monday, ..., Sunday, in the modeling, which denotes the day of the week for day  $i$ . The proposed overdispersed generalized linear model (OGLM), like the GLM, contains the model of  $y_{i(j)}$  and the link function of  $p_{i(j)}$ .  $y_{i(j)}$  is assumed to follow a Binomial distribution as in the GLM

$$y_{i(j)} \sim \text{Bin}(x_{i(j)}, p_{i(j)})$$

The link function may follow one of the following forms

Model 1:

$$\eta(p_{i(j)}) = a_j + b_j \cdot x_{i(j)} + \varepsilon_{i(j)}, \quad \varepsilon_{i(j)} \sim N(0, \sigma_j^2) \quad (3)$$

Model 2:

$$\begin{cases} \eta(p_{i(j)}) = a_I + b_I \cdot x_{i(j)} + \varepsilon_{i(j)}, & \varepsilon_{i(j)} \sim N(0, \sigma_I^2) \quad \text{for } j=1, \dots, 5 \\ \eta(p_{i(j)}) = a_{II} + b_{II} \cdot x_{i(j)} + \varepsilon_{i(j)}, & \varepsilon_{i(j)} \sim N(0, \sigma_{II}^2) \quad \text{for } j=6, 7 \end{cases} \quad (4)$$

Model 3:

$$\begin{cases} \eta(p_{i(j)}) = a_{Ij} + b_I \cdot x_{i(j)} + \varepsilon_{i(j)}, & \varepsilon_{i(j)} \sim N(0, \sigma_I^2) \quad \text{for } j=1, \dots, 5 \\ \eta(p_{i(j)}) = a_{IIj} + b_{II} \cdot x_{i(j)} + \varepsilon_{i(j)}, & \varepsilon_{i(j)} \sim N(0, \sigma_{II}^2) \quad \text{for } j=6, 7 \end{cases} \quad (5)$$

Model 4:

$$\begin{cases} \eta(p_{i(j)}) = a_I + b \cdot x_{i(j)} + \varepsilon_{i(j)}, & \varepsilon_{i(j)} \sim N(0, \sigma_I^2) \quad \text{for } j=1, \dots, 5 \\ \eta(p_{i(j)}) = a_{II} + b \cdot x_{i(j)} + \varepsilon_{i(j)}, & \varepsilon_{i(j)} \sim N(0, \sigma_{II}^2) \quad \text{for } j=6, 7 \end{cases} \quad (6)$$

In the above models, the random error  $\varepsilon$  in each model is used to characterize the overdispersion in the data. It is assumed to be normally distributed with mean 0 and variance  $\sigma^2$ . The effect of weekday/weekend is represented by different settings of the model parameters, including the intercept ( $a$ ), the slope ( $b$ ) and the variance of the random error ( $\sigma^2$ ).

Specifically, *Model 1* is the full model which assumes the service accessibility on each day of the week follows a different pattern and thus requires a separate model with parameters  $a_j$ ,  $b_j$ , and  $\sigma_j^2$ . Model 2-4 assume that similar patterns are followed among the weekdays ( $j=1, \dots, 5$ ) and the weekends ( $j=6, 7$ ), so separate models are considered for the two groups. The three models differ in the way to model the variation within the same group and between the two groups: *Model 2* assumes that the accessibility in the five weekdays follow the same model with parameters  $a_I$ ,  $b_I$ , and  $\sigma_I^2$ , while those on the two weekends follow another model with parameters  $a_{II}$ ,  $b_{II}$ , and  $\sigma_{II}^2$ ; *Model 3* is similar to Model 2, except that the intercepts on different days in each group are different; *Model 4* is also similar to Model 2, except that the weekday model and the weekend model share the same slope, i.e., the influence of inbound volumes is the same on weekdays and weekends. The interpretation of these models is given in Table 1. It is

worth mentioning that the OGLMs are hierarchical models with two levels (the binomial model of  $y$  and the regression model of  $\eta$ ).

Table 1. The proposed models of the link function

<b>Model</b>	<b>Interpretation</b>
1	a separate model for each day of the week (Monday model, Tuesday model, .....)
2	weekday model and weekend model
3	weekday models and weekend models (with different intercepts)
4	weekday model and weekend model (with same slope)

The models will be built in the following way: for Model 1, the model for each day of the week is built separately using related data; for Model 2 and 3, data on all the weekdays will be pooled together to build the weekday model, while those on the weekends will be pooled together to build the weekend model; for Model 4, all the data will be used to build the model. For a given dataset, we need to determine which one among the four OGLMs in (3)-(6) fits the data the best. Bayesian approaches will be used to build each model and find the best model, the details of which will be given in the next section.

### 3. Bayesian Algorithms for Model Estimation and Comparison

Consider a period of  $m$  days which yields a set of call volume observations  $\{(y_i, x_i), i=1, \dots, m\}$ . This section will present the Bayesian algorithms for model estimation and comparison based on these data. We adopt the Bayesian approach to build the proposed OGLMs for two reasons: first, it is well known that hierarchical models like the OGLM can be easily built under the Bayesian framework<sup>16-19</sup>. Second, Bayesian statistics provides convenient tools to compare models of different forms such as the four variants of the OGLM in (3)-(6). Details of the Bayesian approaches for model estimation and comparison are given in Section 3.1 and 3.2 respectively.

#### 3.1 Model estimation

Let  $\theta$  be the parameter of an OGLM. For the proposed models, the parameter is



$$\boldsymbol{\theta} = \begin{cases} [a_1, b_1, \sigma_1^2, \dots, a_7, b_7, \sigma_7^2] & \text{Model1} \\ [a_1, b_1, \sigma_1^2, a_{II}, b_{II}, \sigma_{II}^2] & \text{Model2} \\ [a_{II}, \dots, a_{IV}, b_1, \sigma_1^2, a_{III}, \dots, a_{IV}, b_{II}, \sigma_{II}^2] & \text{Model3} \\ [a, b, \sigma_1^2, \sigma_{II}^2] & \text{Model4} \end{cases}$$

Given a model, estimates of its parameters need to be found. In the Bayesian framework, this is to find the posterior distribution of the parameter

$$P(\boldsymbol{\theta} | (y_1, x_1), \dots, (y_m, x_m)) \propto \pi(\boldsymbol{\theta}) \cdot f((y_1, x_1), \dots, (y_m, x_m) | \boldsymbol{\theta}) \quad (7)$$

where  $P(\cdot)$  denotes probability,  $\pi(\cdot)$  is the prior of the parameter, and  $f(\cdot)$  is the sampling density of the data. The posterior distribution of the OGLMs is analytically intractable, so Markov chain Monte Carlo (MCMC) algorithms<sup>20</sup> need to be used to generate samples from the posterior (called *posterior sampling* later). Then the location statistics, e.g., mean, median or mode, of the samples will be taken as the point estimates of the parameter. The posterior distributions and the procedure for posterior sampling will be given in the following.

The posterior distribution in (7) cannot be sampled directly as the data  $\{(y_i, x_i), i=1, \dots, m\}$  depend on the accessibility rates  $\{p_1, \dots, p_m\}$ , rather than  $\boldsymbol{\theta}$ , directly, which means that the sampling density  $f((y_1, x_1), \dots, (y_m, x_m) | \boldsymbol{\theta})$  is not available. Here  $\{p_1, \dots, p_m\}$  are unknown nuisance parameters. Therefore, we need to sample from the joint posterior of  $\boldsymbol{\theta}$  and the nuisance parameters

$$P(\boldsymbol{\theta}, \{p_1, \dots, p_m\} | (y_1, x_1), \dots, (y_m, x_m))$$

to obtain estimates of  $\boldsymbol{\theta}$ . In general, sampling from such joint posteriors is not trivial due to the high dimension of the parameters. A popular MCMC method to conquer this problem is *Gibbs sampler*<sup>21,22</sup>. The basic idea of this method is to divide the parameters into subgroups and sample from the conditional posterior of each subgroup given all other parameters. As the subgroups have smaller dimensions and their conditional posteriors are simpler than the joint posterior, the sampling can be efficiently implemented using standard algorithms. Specifically, in our problem, the Gibbs sampling procedure includes two steps:

### **Gibbs sampling procedure**

**Step 1:** Sampling from the conditional posterior of  $\theta$  given  $\{p_1, \dots, p_m\}$ ,

$$P(\theta | \{p_1, \dots, p_m\}, (y_1, x_1), \dots, (y_m, x_m)) = P(\theta | \{p_1, \dots, p_m\})$$

**Step 2:** Sampling from the conditional posterior of  $\{p_i, i=1, \dots, m\}$  given  $\theta$ , i.e.,

$$P(\{p_1, \dots, p_m\} | \theta, (y_1, x_1), \dots, (y_m, x_m))$$

Note that the equality in Step 1 achieves because  $\theta$  does not depend on the data directly. To implement the Gibbs sampling procedure, some starting values of  $\{p_1, \dots, p_m\}$  need to be specified. Conditioning on those values, Step 1 will be performed to obtain one sample of  $\theta$ , and then given that sample, Step 2 will be performed to obtain one sample of  $\{p_1, \dots, p_m\}$ . Then Step 1 will be repeated using the current values of  $\{p_1, \dots, p_m\}$ , and so on. Through such an iterative process, a sequence of samples of  $\theta$  and  $\{p_1, \dots, p_m\}$  from their joint posterior will be obtained, and the samples of  $\theta$  will be used to calculate point estimates of the parameters. The key in this procedure is sampling from the two conditional posteriors, which will be described below.

In the Step 1 sampling, given  $\{p_1, \dots, p_m\}$ ,  $\{\eta(p_1), \dots, \eta(p_m)\}$  are known. Viewing the models of the link function in (3)-(6) as linear regression models with  $\eta(p)$  being the “response”,  $x$  being the “regressor”, and  $\theta$  being the unknown parameters, the sampling can be conducted following the well-established *Bayesian linear regression models*<sup>23,24</sup>. Let  $\{(y_i, x_i), i=1, \dots, n_j\}$  be all the data on day  $j, j=1, \dots, 7$ , of the week,  $\{(y_i, x_i), i=1, \dots, n_I\}$  be data on all the weekdays, and  $\{(y_i, x_i), i=1, \dots, n_{II}\}$  be data on the weekends. The sampling algorithms for the proposed models are displayed in Figure 3, where “IG” means Inverse-Gamma distribution. Note that: (1) Noninformative prior  $\pi(a, b, \sigma^2) \propto 1$  is used in all the sampling since prior knowledge on the parameters is typically not available<sup>25-27</sup>. (2) The algorithm under “Model 1 and 2” is for day  $j$  of Model 1 which uses data  $\{(y_i, x_i), i=1, \dots, n_j\}$ . To obtain the weekday model and weekend model under Model 2,  $\{(y_i, x_i), i=1, \dots, n_I\}$  or  $\{(y_i, x_i), i=1, \dots, n_{II}\}$  should be used instead.

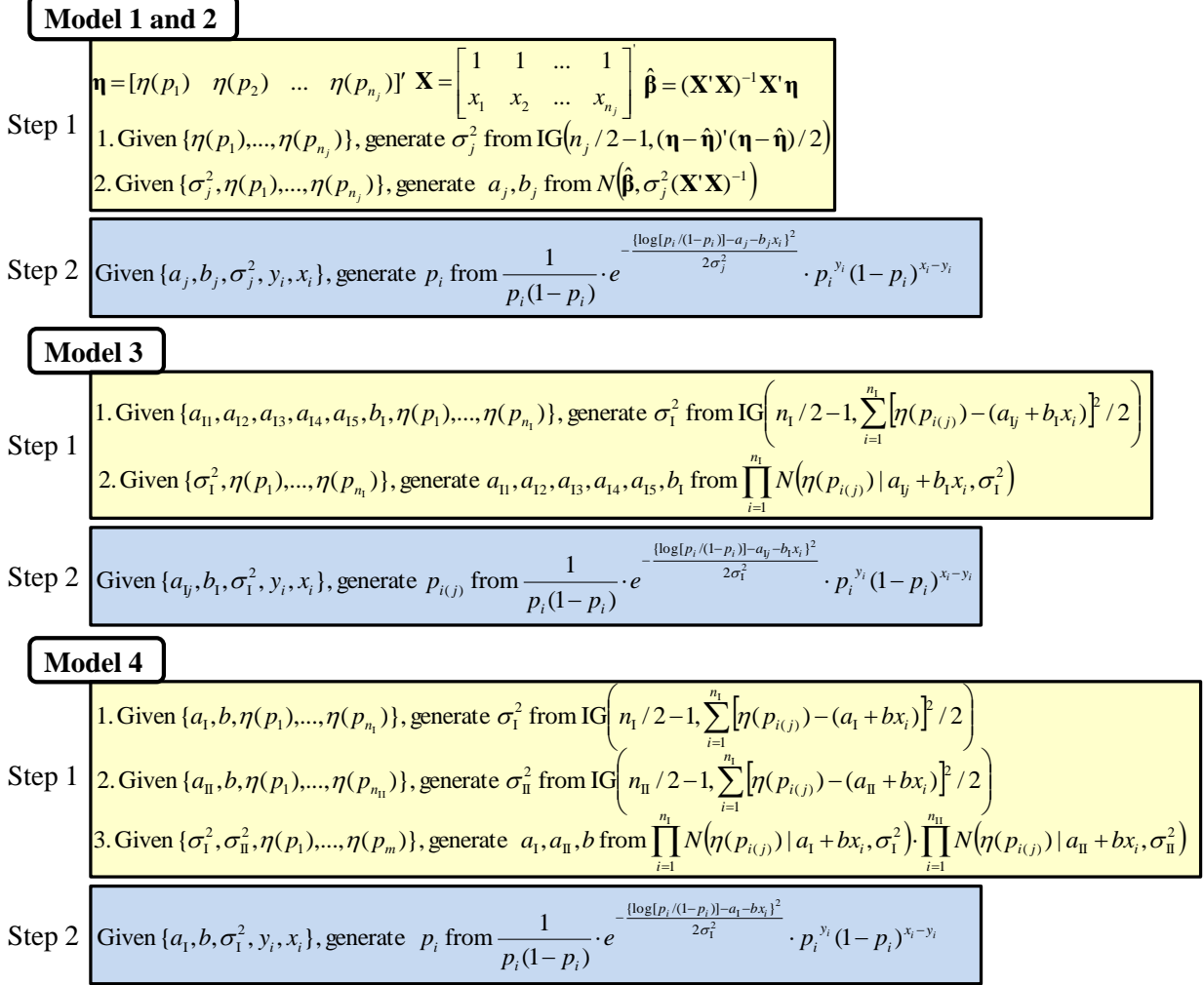


Figure 3. Algorithms of Gibbs sampling for estimating the proposed models

In the Step 2 sampling, given  $\boldsymbol{\theta}$ ,  $p_i$  only depends on  $(y_i, x_i)$ , so the posterior of  $\{p_i, i=1, \dots, m\}$  can be sampled separately. According to the structure of the proposed models,

$$P(p_i | \boldsymbol{\theta}, (y_i, x_i)) \propto \pi(p_i | \boldsymbol{\theta}) \cdot f(y_i | x_i, p_i)$$

where  $f(y_i | x_i, p_i)$  is the Binomial density, and  $\pi(p_i | \boldsymbol{\theta})$  is the prior of  $p_i$  given  $\boldsymbol{\theta}$ . Since  $p_i$  does not depend on  $\boldsymbol{\theta}$  directly,  $\pi(p_i | \boldsymbol{\theta})$  is not readily available. To find it, the conditional distribution of the link function  $\eta(p_i |)$  given  $\boldsymbol{\theta}$  is obtained first and then the conditional distribution of  $p_i$  can be induced through the change-of-variables transformation. The resulting conditional posteriors of  $p_i$  under the proposed models are given in Figure 3. Note that the posteriors are non-normalized.

The algorithms in Figure 3 require sampling from regular distributions including normal distributions and Inverse-Gamma distributions (e.g., Step 1 in Model 1 and 2), and from irregular distributions (e.g., Step 2 in all models). The former can be realized using standard random generators such as the built-in routines in Matlab, while the latter can be realized using MCMC algorithms. In this study, a novel MCMC algorithm, *slice sampler*<sup>28,29</sup> will be used for its convenience in use: it only needs two inputs to work: the posterior distribution to be sampled from and initial values. In addition, it does not require the posterior be normalized.

### 3.2 Model comparison

The proposed models can be compared using the deviance information criterion (DIC), which is a popular tool for model comparison, especially hierarchical models, in Bayesian statistics<sup>21,30</sup>. Let  $D$  represent the deviance of a model with parameter  $\boldsymbol{\theta}$

$$D((y_1, x_1), \dots, (y_m, x_m) | \boldsymbol{\theta}) = -2 \sum_{i=1}^m \log f((y_i, x_i) | \boldsymbol{\theta}) \quad (8)$$

The DIC is defined to be

$$\text{DIC} = 2\bar{D}((y_1, x_1), \dots, (y_m, x_m) | \boldsymbol{\theta}) - D((y_1, x_1), \dots, (y_m, x_m) | \hat{\boldsymbol{\theta}}) \quad (9)$$

where  $\bar{D}(\cdot | \boldsymbol{\theta})$  is the expected deviance, and  $D(\cdot | \hat{\boldsymbol{\theta}})$  is the deviance corresponding to a point estimate of  $\boldsymbol{\theta}$  such as its posterior mean. A smaller value of DIC indicates a better model.

The DIC can be estimated using the posterior samples of  $\boldsymbol{\theta}$  obtained through the Gibbs sampling procedures described in Section 3.1. Specifically, given a set of posterior samples of  $\boldsymbol{\theta}$ , the deviance of each sample will be calculated by (8), and their average will be used as an estimate of  $\bar{D}(\cdot | \boldsymbol{\theta})$ . Meanwhile, the mean of the samples will be found, and its corresponding deviance, i.e.,  $D(\cdot | \hat{\boldsymbol{\theta}})$ , will be calculated. The DIC statistic will then be obtained by (9). The method to calculate the deviance of a given posterior sample of  $\boldsymbol{\theta}$  will be described as follows.

For the proposed OGLMs, the sampling density can be obtained by

$$\begin{aligned} f((y_i, x_i) | \boldsymbol{\theta}) &= \int_0^1 f((y_i, x_i), p_i | \boldsymbol{\theta}) dp_i \\ &= \int_0^1 f((y_i, x_i) | p_i) \cdot f(p_i | \boldsymbol{\theta}) dp_i \end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \text{Bin}(y_i; x_i, p_i) \cdot f(p_i | \boldsymbol{\theta}) dp_i \\
&= E_{p_i}[\text{Bin}(y_i; x_i, p_i)]
\end{aligned} \tag{10}$$

where the expectation is with respect to  $p_i$  which follows the conditional distribution given  $\boldsymbol{\theta}$

$$p_i \sim f(p_i | \boldsymbol{\theta}) \tag{11}$$

Consequently, the deviance is

$$D((y_1, x_1), \dots, (y_m, x_m) | \boldsymbol{\theta}) = -2 \sum_{i=1}^m \log E_{p_i}[\text{Bin}(y_i; x_i, p_i)] \tag{12}$$

By the above formulas, the deviance can be estimated through the following steps: First, generate a sequence of samples of  $p_i$  by (11). This can be realized by generating a sequence of samples of  $\eta_i$  under the given OGLM and then obtaining the corresponding values of  $p_i$  through its relationship with  $\eta_i$  defined by the link function. Second, for each sample of  $p_i$ , the Binomial density  $\text{Bin}(y_i | x_i, p_i)$  will be calculated, and their average will be used as an estimate of the expectation in (10). Finally, the deviance in (12) will be obtained using these estimates.

## 4. Case Study

The proposed approach has been applied to a dataset from an after-hour triage center in Wisconsin. Part of the data have been shown in Figure 2. The whole dataset contains daily inbound and outbound call volume data in the center during 2005~2009. In this study, a preliminary analysis is first done to obtain some descriptive statistics and primitive estimates of the accessibility rate. Then the proposed models will be estimated using the algorithms described in Section 3.1 and compared using the procedure in Section 3.2.

### 4.1 Preliminary analysis

The first step in the preliminary analysis is to remove some potential outliers in the dataset, which are extreme values compared to the majority of observations in the same day of the week. The remaining observations are used in the following analysis. Table 2 lists some descriptive statistics of the data from each day of the week, including the number of observations and daily average number of inbound calls and outbound calls. One clear fact is that the call volume on

weekends is more than 3 times higher than weekdays. The reason is that the triage service is available for a longer period on weekends (starting from 5pm on weekdays and 8am on weekends). The call volume reaches the peak on Saturdays. Figure 4 shows the empirical distribution of the number of inbound calls on each day of the week. Since count data like the call volumes are usually modeled using a Poisson distribution, the fitted Poisson distribution ( $\lambda$ =sample mean) is also given in each plot of Figure 4 (the dashed curve). We can see that the empirical distributions have heavier tails than the Poisson distributions in all the cases, meaning that the real data contain more extreme values and Poisson distribution is not adequate to model the variation in the data.

Table 2. Descriptive statistics of data on each day of the week

	<b>Mon</b>	<b>Tue</b>	<b>Wed</b>	<b>Thu</b>	<b>Fri</b>	<b>Sat</b>	<b>Sun</b>
#obsns	222	234	238	230	235	242	239
avg(inbound)	39.5	38.0	36.3	35.2	39.3	136.7	111.4
avg(outbound)	17.0	17.4	15.5	14.4	18.5	54.9	37.5

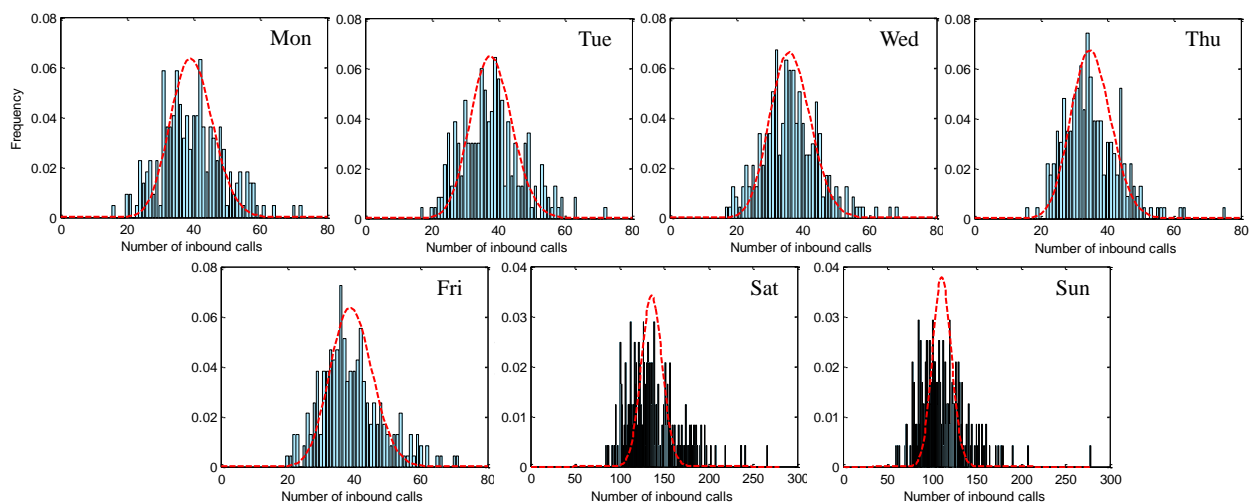


Figure 4. Empirical distribution of inbound call volume and fitted Poisson distribution

To examine the outbound-inbound relationship, Figure 5 (left panel) shows the plots of inbound call volumes and corresponding outbound call volumes. An apparent linear relationship can be seen on all the days with considerable noises. Another observation is that the data on the five weekdays show similar patterns, while those on the two weekends show similar patterns.

This suggests that it is reasonable to build a separate model for the weekdays and the weekends. To further explore the dependency of the accessibility rate  $p$  on the inbound call volume, we can check the maximum likelihood estimate (MLE) of  $p$  for a given inbound volume. Let  $x$  be the given inbound volume which occurred  $w$  times in the dataset and  $y_1, \dots, y_w$  be the corresponding outbound volumes. The MLE of  $p$  given  $x$  is

$$\hat{p}(x) = \sum_{t=1}^w y_t / wx$$

The MLEs for each day of the week are shown in the right panel of Figure 5. Clearly,  $p$  has a positive linear relationship with the number of inbound calls, and this relationship is subject to considerable noises. The noise effect can be handled by the random errors in the OGLMs. This suggests that the proposed models are reasonable to the data. Moreover, the slopes and intercepts of the accessibility rate on the weekdays/weekends look similar, indicating that Model 2 in (4) is likely to be an appropriate model for the data.

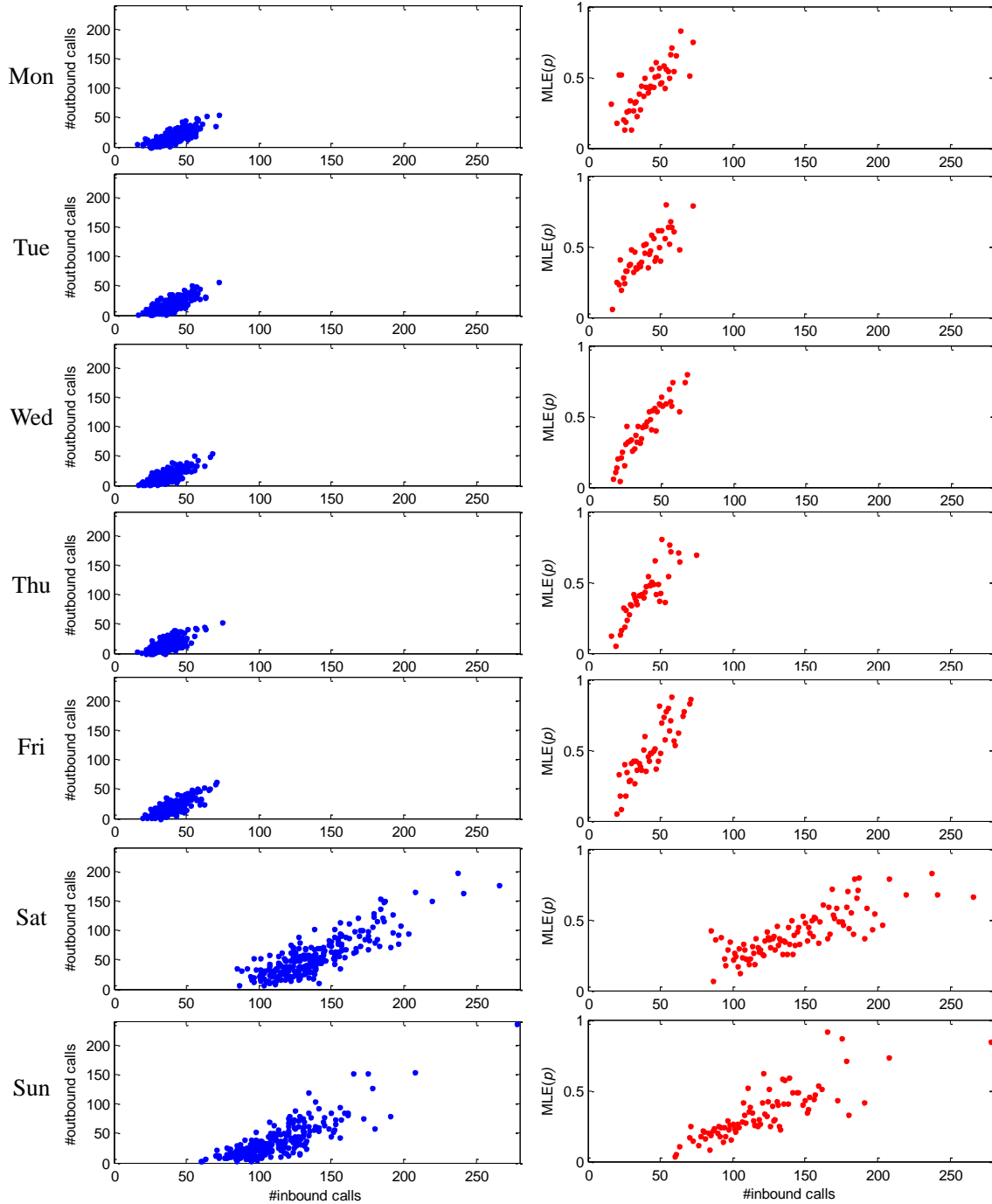


Figure 5. Data used in case study: call volumes (left) and MLE of  $p$  (right)

## 4.2 Results of model estimation

We fit each of the four OGLMs to the data using the algorithms in Figure 3. 5000 samples are generated in the posterior sampling. As an example, Figure 6 shows the first 500 samples in fitting the weekday model in Model 2 and Model 4 respectively. Fitting Model 2 requires



sampling from normal distribution and inverse-gamma distribution, which are conducted using random generators in Matlab, while fitting Model 4 requires sampling from irregular posteriors which are conducted using the *slicesample* function in Matlab. We can see that in all the plots, the effect of starting values disappear very quickly and then the samples become stationary. Samples from Model 2 converge faster since it is easier to sample from regular distributions and the dimension of its parameters is smaller than Model 4. To be conservative, the 500 starting samples will be discarded and the following analysis will be done on the remaining samples.

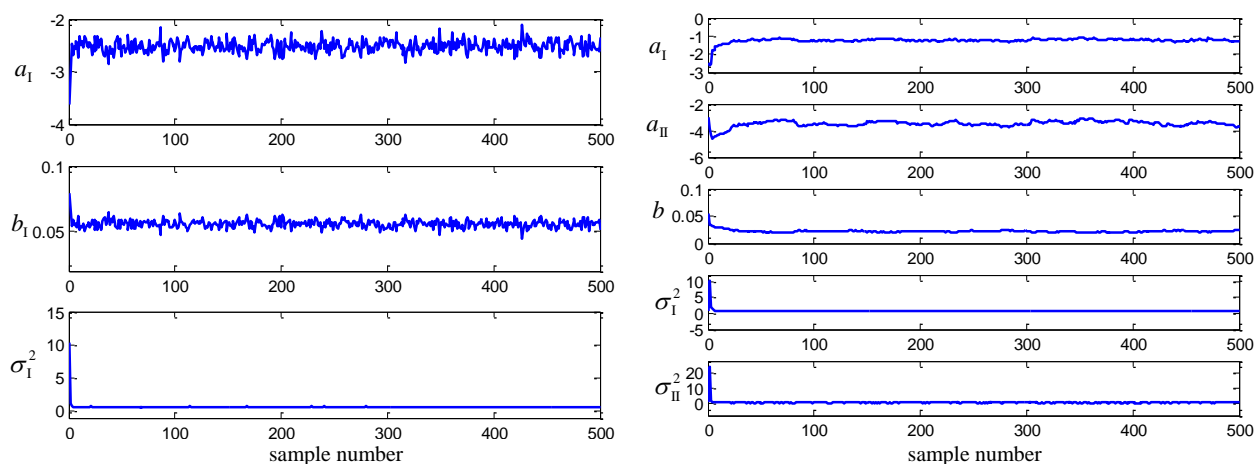


Figure 6. Samples from posterior sampling in Model 2 (left) and Model 4 (right)

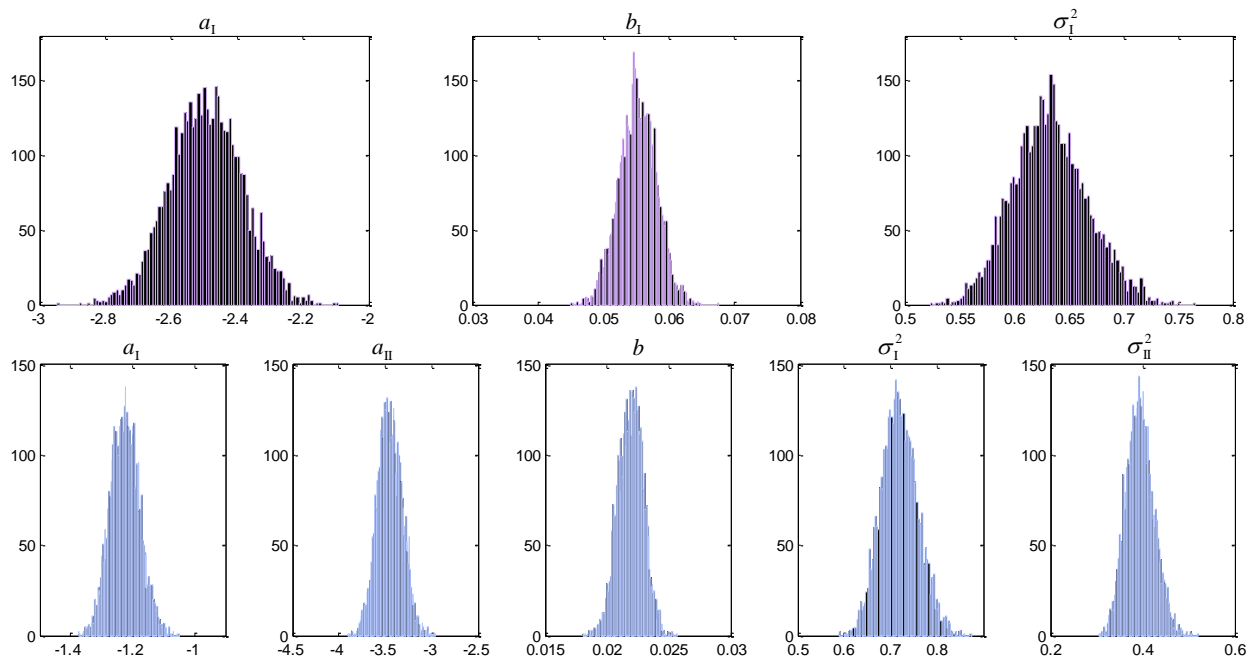


Figure 7. Histograms of posterior samples in Model 2 (upper) and Model 4 (lower)

Using the samples, we can obtain the empirical posterior distribution of each parameter by plotting the histograms of the samples. As an example, Figure 7 shows empirical posteriors of the weekday model in Model 2 and Model 4. The posteriors are mostly symmetric, meaning that the sample mean can be used as point estimates of the parameters. These estimates are summarized in Table 3. In the results of Model 1, the parameter estimates among the weekdays and the weekends are very similar, but the difference between the two groups is large. This indicates that it might be better to build one model for weekdays and another for weekends, which is consistent with the preliminary results in Figure 5; the results of Model 3 show that the difference in the intercepts is trivial among the weekdays and weekends, suggesting a common model for each group too; in Model 4, the estimate of the slope is between the two estimates ( $b_I$  and  $b_{II}$ ) in Model 2 since the weekday model and weekend model are assumed to share the same slope here. Correspondingly, the variances of random errors in the two models are larger than in Model 2, implying that the same-slope assumption may not be appropriate.

Table 3. Point estimates and DIC of the proposed OGLMs

Model	Point Estimates of Parameters	DIC
1	$a_1=-2.5460, b_1=0.0528, \sigma_1^2=0.4848$ (Mon)	11651.0
	$a_2=-2.0314, b_2=0.0456, \sigma_2^2=0.6574$ (Tue)	
	$a_3=-2.7571, b_3=0.0628, \sigma_3^2=0.5893$ (Wed)	
	$a_4=-2.5513, b_4=0.0567, \sigma_4^2=0.7531$ (Thu)	
	$a_5=-2.6811, b_5=0.0611, \sigma_5^2=0.6799$ (Fri)	
	$a_6=-2.8678, b_6=0.0170, \sigma_6^2=0.3256$ (Sat)	
	$a_7=-3.2702, b_7=0.0212, \sigma_7^2=0.4217$ (Sun)	
2	$a_I=-2.4941, b_I=0.0553, \sigma_I^2=0.6329$ (Weekday)	11634.8
	$a_{II}=-2.9631, b_{II}=0.0180, \sigma_{II}^2=0.3733$ (Weekend)	
3	$a_{11}=-2.6486, a_{12}=-2.4018, a_{13}=-2.4822, a_{14}=-2.4943$	11638.0
	$a_{15}=-2.4478, b_I=0.0553, \sigma_I^2=0.6306$ (Weekday)	
4	$a_{16}=-2.9874, a_{17}=-3.1101, b_{II}=0.0187, \sigma_{II}^2=0.3731$ (Weekend)	11787.0
	$a_I=-1.2259, a_{II}=-3.4548, b=0.0219, \sigma_I^2=0.7219, \sigma_{II}^2=0.3998$	

### 4.3 Results of model comparison

To find the best model for the data, DIC for each model is calculated. First, for each posterior sample of  $\theta$ ,  $m=100$  samples of  $p_i$  is generated by (11), and using these samples,  $f((y_i, x_i)|\theta)$  is calculated by (10). Then the deviance for the given  $\theta$  is obtained by (12) and the DIC under the considered model is obtained by (9). The resulting DIC values for the four OGLMs are listed in the last column of Table 3. Model 2 has the smallest DIC, while Model 4 has the largest DIC. Thus, Model 2 is chosen to be the best model. Note that Model 2 is also the simplest model.

Some interpretation of the chosen model is provided as follows: the accessibility of the call center follows the same pattern on the five weekdays and the two weekends; the accessibility depends on the number of incoming call volume, with higher incoming volume leading to lower accessibility; the influence of incoming volumes on the accessibility is higher on weekdays than weekends ( $b_I > b_{II}$ ), and the variation on weekdays is also larger than weekends ( $\sigma_I^2 > \sigma_{II}^2$ ). A possible explanation for this is that the call center opens after working hours (5pm~midnight) on weekdays and the whole day (8am~midnight) on weekends. As a result, there is more flexibility for patients to contact the call center on weekends and thus the triage service is more accessible and stable on weekends.

#### 4.4 Model validation

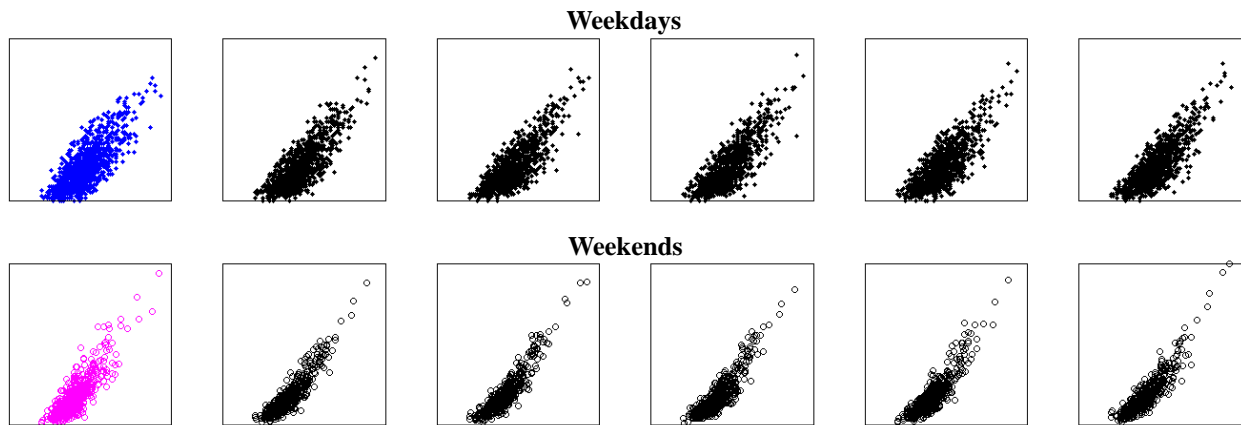


Figure 8. Real data (1<sup>st</sup> plot) and simulated data from the selected model

One way to check the adequacy of the selected model is to randomly generate datasets of the same size as that used in the case study from the selected model and compare the simulated data

to the real data graphically<sup>21</sup>. Such graphs are shown in Figure 8, where the first plot in each row is the real data and the following plots are random samples from the selected model given the same inbound volumes. We can see that the simulated datasets exhibit similar patterns as the real data, suggesting that the selected model fits the data well. The weekday model works relatively better than the weekend model, which is possibly due to the larger sample size of weekday data.

## 5. Numerical Study

We have also done some simulations to examine the properties of the OGLM. For simplicity, the effect of weekday/weekend is not considered here, that is, the OGLM is assumed to be

$$\begin{aligned} y_i &\sim \text{Bin}(x_i, p_i) \\ \eta(p_i) &= a + bx_i + \varepsilon_i \end{aligned}$$

where  $\varepsilon_i \sim N(0, \sigma^2)$ . Our study focuses on two aspects:

**Study 1.** *Graphical characteristics of the OGLM compared to the conventional models:* Data from the OGLM and conventional models (BM and GLM) are simulated and graphed to show the key features of each model and their differences. These graphs are useful explorative tools for practitioners to determine which model to use before conducting formal statistical analysis.

**Study 2.** *Performance of the Bayesian estimator for OGLM:* Errors of the Bayesian estimator given in Section 3.1 are evaluated under a set of sample sizes. The purpose is to show the accuracy of the proposed estimation approach and its sensitivity to sample size.

### 5.1 Results of Study 1

1000 data points from the three models are generated with the following parameter specification:  $p=0.8$  (BM),  $a=-4$ ,  $b=0.1$  (GLM), and  $a=-4$ ,  $b=0.1$ ,  $\sigma^2=1$  (OGLM). The inbound call volume is assumed to follow a Poisson distribution with  $\lambda=40$ . The simulated datasets are shown in the upper panel of Figure 9. The data from the three models appear to have similar patterns, especially the BM and the GLM, implying that it is not obvious which model is appropriate for the data by directly observing the data. A better method to explore the data is to check the MLE of  $p$  described in Section 4.1, which is shown in the lower panel of Figure 9. We

can see that the MLE of  $p$  in the BM fluctuates around a constant level, that in the GLM is a linear function of the inbound call volume with small noises, and that in the OGLM is similar to the GLM except with larger noises. This suggests that the plot of  $\text{MLE}(p)$  can be a good explorative tool to determine the appropriate model for the data.

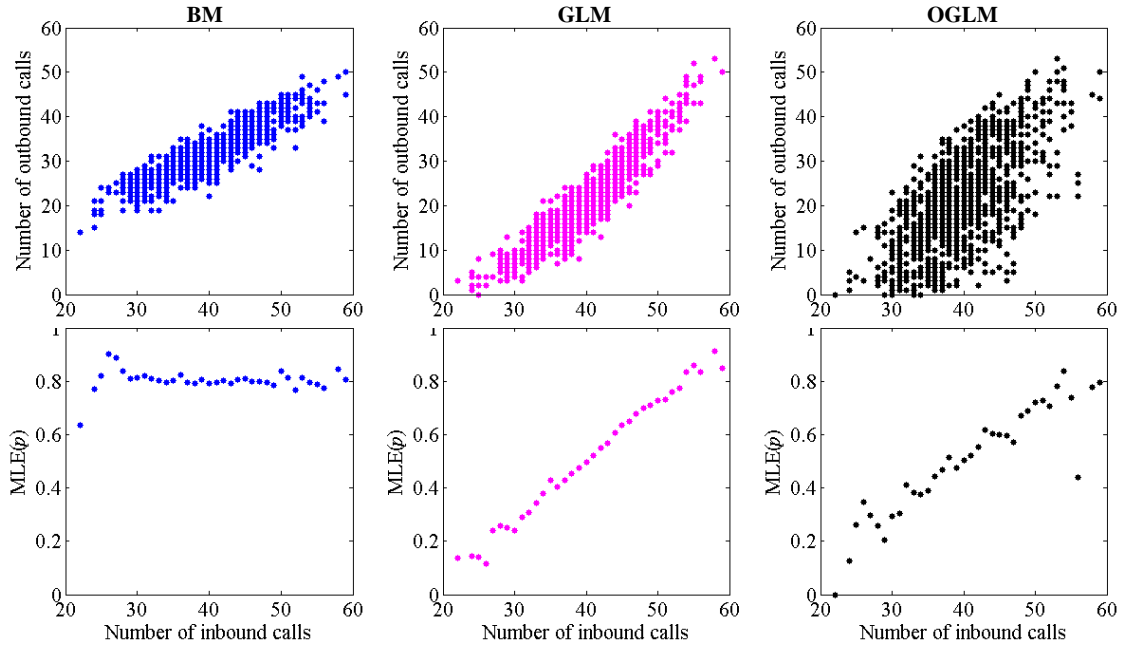


Figure 9. Simulated data from the models (upper) and MLE of  $p$  (lower)

## 5.2 Results of Study 2

Two measures are used to evaluate the errors of the Bayesian estimator for the OGLM, which are defined as follows

$$\text{relative bias: } \left| \frac{\text{avg}(\hat{\theta}) - \theta}{\theta} \right| \quad \text{relative standard deviation: } \left| \frac{\text{std}(\hat{\theta})}{\theta} \right|$$

where  $\theta = a, b, \sigma^2$  is a parameter of the OGLM, and  $\text{avg}(\hat{\theta})$  and  $\text{std}(\hat{\theta})$  are the average and standard deviation of the estimate of  $\theta$  respectively. Intuitively, the two errors depend on the sample size of data, so they are calculated under a set of specified sample sizes from 20 to 500. Given each sample size, 100 datasets are generated from the OGLM with the same parameter specification as in Study 1, and point estimates of the parameters are obtained for each dataset

using the Gibbs sampling procedure given in Section 3.1. Then the performance measures are calculated based on those estimates.

Figure 10 shows the results of the simulations. Both the bias and standard deviation measures have large values when the sample size is very small (e.g., 20), and decrease dramatically as sample size increases, which is consistent with intuition. When the sample size is adequate e.g., 400, the bias and standard deviation of the parameter estimates are very small, meaning that accurate estimation is achieved.

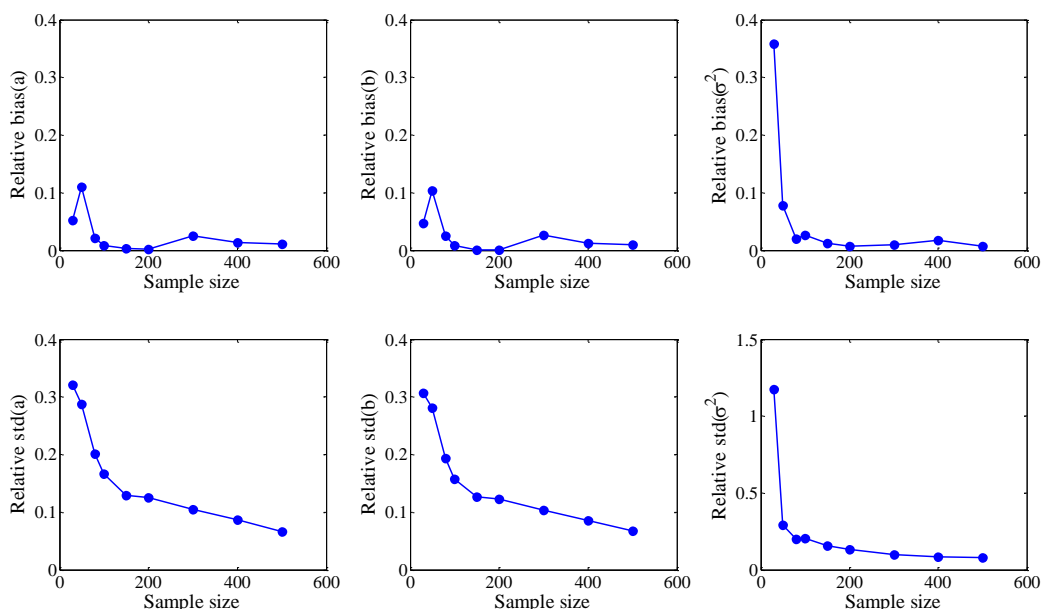


Figure 10. Errors of the Bayesian estimator for OGLM: relative bias (upper) and relative standard deviation (lower)

## 6. Conclusions and Discussions

This study develops an approach to model the relationship between the outbound and inbound call volumes to measure the accessibility of the telephone nurse triage service. The proposed models consider both the overdispersion in the data and weekday/weekend effect. Bayesian approaches are used to estimate parameters of the proposed models and select the best model. In the case study, the proposed models are applied to a dataset from a nurse triage call center. It is found that the accessibility of the center follows the same model among the weekdays and the weekends, and the accessibility on weekends is more stable and less dependent

on the incoming call volume than weekdays. The numerical study shows that the MLE of  $p$  can be used as an explorative tool in practice to determine the appropriate model for the given data, and errors of the estimators of the OGLM are acceptable when the sample size is adequate.

Empirical analysis of the operational data in the telephone nurse triage services is an emerging research topic and the current study is just a starting effort on this topic. There are many interesting unresolved problems that will be considered in our future studies. For example, methods for monitoring the service accessibility will be developed based on the constructed models. A straightforward idea for this is to build the OGLM for the data in each specified period and then monitor the parameter estimates of the models continuously. Since the parameters of the OGLM can be interpreted in an intuitive way, that is,  $a$  and  $b$  represent the base-line performance of the nurses, while  $\sigma^2$  represents the noises in the service system, monitoring of these parameters can not only detect changes in the system, but also provide guidance on root causes of the changes. The challenge here is to find the statistical properties of the parameter estimates and appropriate monitoring schemes. Moreover, besides the call volume data, there are many other types of measurements available in the TNT services such as duration and waiting time of calls. Efforts will also be made to mine the information in those data for a better characterization of service performances.

### References

- [1] Cariello, F.P. (2003) Computerized telephone nurse triage—an evaluation of service quality and cost. *Journal of Ambulatory Care Management*, **26**(2), 124-137.
- [2] Rutenberg, C.D. (2000) Telephone triage. *The American Journal of Nursing*, **100**(3), 77-81.
- [3] Ledlow, G.R., O’Hair, H.D., and Moore, S. (2003) Predictors of communication quality: the patient, provider, and nurse call center triad. *Health Communication*, **15**(4), 431-455.
- [4] O’Connell, J.M., Johnson, D.A., Stallmeyer, J., Cokingtin, D. (2001) A satisfaction and return-on-investment study of a nurse triage service. *The American Journal of Managed Care*, **7**(2), 159-169.
- [5] Richards, D.A., Meakins, J., Tawfik, J., Godfrey, L., Dutton, E., and Heywood, P. (2004a) Quality monitoring of nurse telephone triage: pilot study. *Journal of Advanced Nursing*, **47**(5), 551-560.

- [6] Hutcherson, C.M. (2001) Legal considerations for nurses practicing in a telehealth setting. *The Online Journal of Issues in Nursing*, **6**(3), 4.
- [7] Richards, D.A., Meakins, J., Godfrey, L., Tawfik, J., and Dutton, E. (2004b) Survey of the impact of nurse telephone triage on general practitioner activity. *British Journal of General Practice*, **54**, 207-210.
- [8] Oak ridge institute for science and education (ORISE) (2001) *Coordinating call centers for responding to pandemic influenza and other public health emergencies: a workbook for state and local planners*.
- [9] Gans N., Koole, G., and Mandelbaum, A. (2003) Telephone call centers: tutorial, review and research prospects. *Manufacturing and Service Operations Management*, **5**, 79-141.
- [10] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. and Zhao, L. (2005) Statistical analysis of a telephone call center: a queueing-science perspective. *Journal of the American Statistical Association*, **100**(469), 36-50.
- [11] Myers, R.H., Montgomery, D.C., Vining, G.G., and Robinson, T.J. (2010) *Generalized linear models with applications in engineering and the sciences*, John Wiley & Sons, Hoboken, New Jersey.
- [12] Chambers, E.A., and Cox D.R. (1967) Discrimination between alternative binary response models. *Biometrika*, **54**, 573-578.
- [13] Efron, B. (1986) Double exponential families and their use in generalized linear regression. *Journal of American Statistical Association*, **81**, 709-721.
- [14] Ganio, L.M., and Schafer, D.W. (1992) Diagnostics for overdispersion. *Journal of the American Statistical Association*, **87**, 795-804.
- [15] Gelfand, A.E., and Dalal, S.R. (1990) A note on overdispersed exponential families. *Biometrika*, **77**, 55-64.
- [16] Gelman, A., and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/hierarchical Models*. Cambridge University Press, New York, NY.
- [17] Daniels, M. J., and Gatsonis, C. (1999), "Hierarchical Generalized Linear Models in the Analysis of Variations in Health Care Utilization", *Journal of the American Statistical Association*, **94**(445): 29-42.
- [18] Daniels, M. J., and Gatsonis, C. (1997), "Hierarchical Polytomous Regression Models with Applications on Health Services Research", *Statistics in Medicine*, **16**:2311-2325.
- [19] Woodard, R., Sun, D., He, Z., and Sheriff, S.L. (1999) Estimating hunting success rates via Bayesian generalized linear models. *Journal of Agricultural, Biological, and Environmental Statistics*, **4**(4), 456-472.
- [20] Robert, C.P., and Casella, G (2004) *Monte Carlo statistical methods, 2nd edition*, Springer, New York, NY.
- [21] Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (2004) *Bayesian data analysis, 2nd edition*, Chapman & Hall/CRC.
- [22] Zeger, S.L., and Karim, M.R. (1991) Generalized linear models with random effects. *Journal of the American Statistical Association*, **86**(413), 79-86.



- [23] Mitchell, T.J., and Beauchamp, J.J. (1988) Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, **83**(404), 1023-1032.
- [24] West, M. (1984) Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society. Series B*, **46**(3), 431-439.
- [25] Kadane, J.B., and Wolfson, L.J. (1998) Experiences in elicitation. *The Statistician*, **47**, 3-19.
- [26] Bedrick E.J., Christensen R, Johnson W. (1996) A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, **91**,1450-1460.
- [27] Chen M, Ibrahim J.G. (2003) Conjugate priors for generalized linear models. *Statistica Sinica*, **13**, 461-476.
- [28] Neal, R.M. (2003) Slice sampling. *The Annals of Statistics*, **31**, 705-741.
- [29] Zeng, L., and Zhou, S. (2011) A Bayesian approach to risk-adjusted outcome monitoring in healthcare. *Statistics in Medicine*, **30**(29), 3431-3446.
- [30] Millar, R.B. (2009) Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics*, **65**, 962-969.