Application of a Mixed Effects Model for Biosurveilliance of Regional Rail Systems

Robert J. Gallop
West Chester University

Abstract

Although United States government planners and others outside government had recognized the potential risk of attacks by terrorists, the events of September 11, 2001, vividly revealed the nation's vulnerabilities to terrorism. Similarly, the 2004 terrorist attacks in Madrid, illustrate vulnerabilities to terrorism extend beyond the United States. Those attacks were obvious destructive acts with a primary purpose of massive causalities. Let's consider a bioterrorist attack which is conducted subtly through the release of a Chemical/Biological agent. If such an attack occurs through release of a specific biological agent, an awareness of the potential threat of this agent in terms of the number of infections and deaths that could occur in a community is of paramount importance in preparing the public health community to respond to this attack. An increase in biosurveillance and novel approaches to biosurveillance are needed. This paper illustrates the use of mixed effects model for biosurveillance based on commuter size for regional rail lines first reported in Gallop et al. (2005). With mixed effects model we can estimate for any station on a given rail system the expected daily number of commuters and establish an acceptability criterion around this expected size. If the actual commuter size is significantly smaller than the estimate, then this could be an indicator of a possible attack. We illustrate this method through an example based on the 2001 daily totals for the Port Authority Transportation Company (PATCO) rail system, which serves residents of southern New Jersey and Philadelphia region in the United States. In addition, we discuss ways to put this application in a real time setting for continuous biosurveillance.

Key words: Bioterrorism, biosurveillance, mixed effects models, Best linear unbiased prediction (BLUP) estimates, effect size.

1. **Introduction:**

The primary goal of biosurveillance is to minimize the effect of a bioterrorist attack. While most communities have disaster response systems, a major bioterrorism attack will produce a catastrophe primarily affecting the community health care system (Flowers et al., 2002). Bioterrorism response requires precise distribution of available personnel and resources. This distribution of personnel and resources must be based on the expected severity of the attack for a given community. Poor distribution of personnel and resources could result in an increase in infections and mortality; therefore, reliable prediction of the expected severity is crucial.

A key location for a bioterrorist attack is a large transportation hub such as Philadelphia's 8th Street and Market Street station. This station is popular not only for the large number of businesses in the neighboring vicinity but also for the large number of tourist attractions. Comparable stations are easily identified for any major city in the United States such as Penn Station in New York and the Metro Center in Washington, D.C. There are various types of biological pathogens that potential enemies may employ, which could be released in a large crowd, released in a food or water supply, or released through ventilation systems of a targeted hub. One concern is the incubation time before infection is recognized. For example, if the agent was anthrax, infection is not immediately apparent. A person may be infected but not symptomatic. Upon presentation, infection may appear as cutaneous anthrax, inhalational anthrax, or gastrointestinal anthrax. Depending on how the anthrax infection appears, determines the severity of the infection as well as the urgency of immediate medical response. Death due to non-treated inhalational and gastrointestinal anthrax is much more rapid than cutaneous anthrax (Bartlett et al., 2002).

The majority of the people in the transportation hub are using one of the available commuter services such as the regional rail lines available at Philadelphia's 8th Street and Market Street station to go to and from work. During the work week, the same commuters will consistently use the same transportation service. If anthrax was released in the ventilation system, then upon presentation of the infection as inhalational anthrax, many commuters may be too sick to go to work; therefore, the number of commuters, which we'll refer to as passenger loads, on any specific rail line would be lower than expected. Many infected individuals will not attribute the infection to bioterrorism. They are more likely to attribute the symptoms to a cold or flu. While we do expect some variability from day to day, if the number of commuters was substantially lower than the expected value, indicating a large number of people potentially experiencing infection symptoms, would raise some level of suspicion.

As discussed by Evans et al. (2002), one approach to biosurveillance is the use of mixed effects models, treating census tracts as the repeated observations over time for a given cluster where the cluster is the unit of analysis. Using historical data, the model estimates month, day of the week, and holiday fixed effects, in addition to random effects for each census tract. These random effects will account for the within correlation for a given unit as

it is measured over time. Evans et al. (2002) described using the estimated random effects to get different estimates for each tract on each day, which may be compared to observed counts. In response to the need for increased biosurveillance, we consider monitoring the passenger loads for regional rail systems through this mixed effects model approach.   This mixed effects model approach will allows us to forecast the expected number of commuters on a regional rail system. This forecast will be adjusted for day of the week differences as well as stations on the rail system. If the forecast is not within an established criterion, then it provides a reason for concern, which will require further investigation of possible causes for the decreased passenger load.

   Gallop et al. (2005) illustrated the use of mixed effects models as a method of biosurveillance. The mixed effects model will be used to forecast the expected passenger load on a given regional rail system. The mixed effects model allows us to properly address various sources of variability. We illustrate this method on available data from the Port Authority Transportation Company (PATCO) rail system, which serves New Jersey residents who commute to Philadelphia and Philadelphians who commute to New Jersey. The PATCO line is illustrated in Figure 1.

Figure 1: PATCO Regional Rail Line
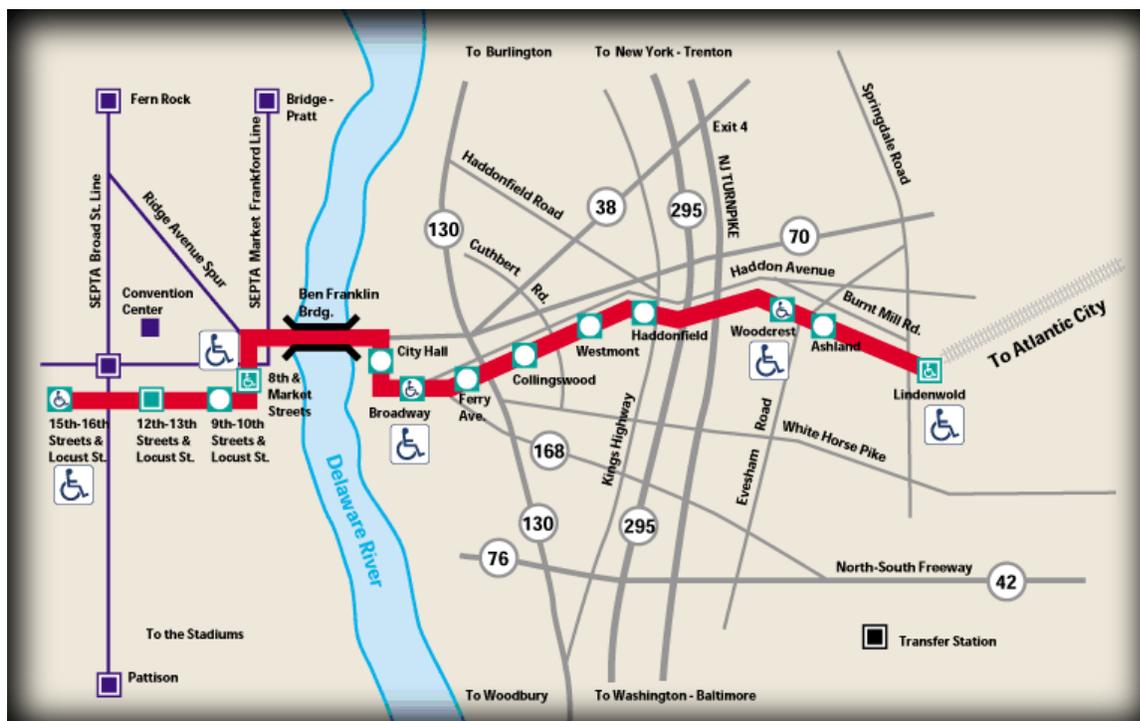
 DRPA: PATCO regional rail map *http://www.drpa.org/patco*.

We have data for each station in New Jersey and Philadelphia for 2001. The data consists of daily total counts of the number of passengers entering each station.   Mirroring the terminology of Evans et al. (2002), the PATCO data consists of 13 census tracts, one for each rail station on the PATCO line, consisting of 365 data points for each tract, corresponding to the daily total per station for the 2001 calendar year. We will estimate the expected number of passengers for any station and particular day through the mixed effects model.

   The remainder of the paper is organized as follows. In the next section, we review the mixed effects model. In Section 3, we discuss our Biosurveillance application to the PATCO regional rail system. Some concluding remarks are made in Section 4.

2. Mixed Effects Model Applied to the PATCO data

   With the PATCO data, we expect some similarity in the daily totals for a given station over time. To get a better understanding of the dynamics of these station tracts over time, we consider daily totals over time for each
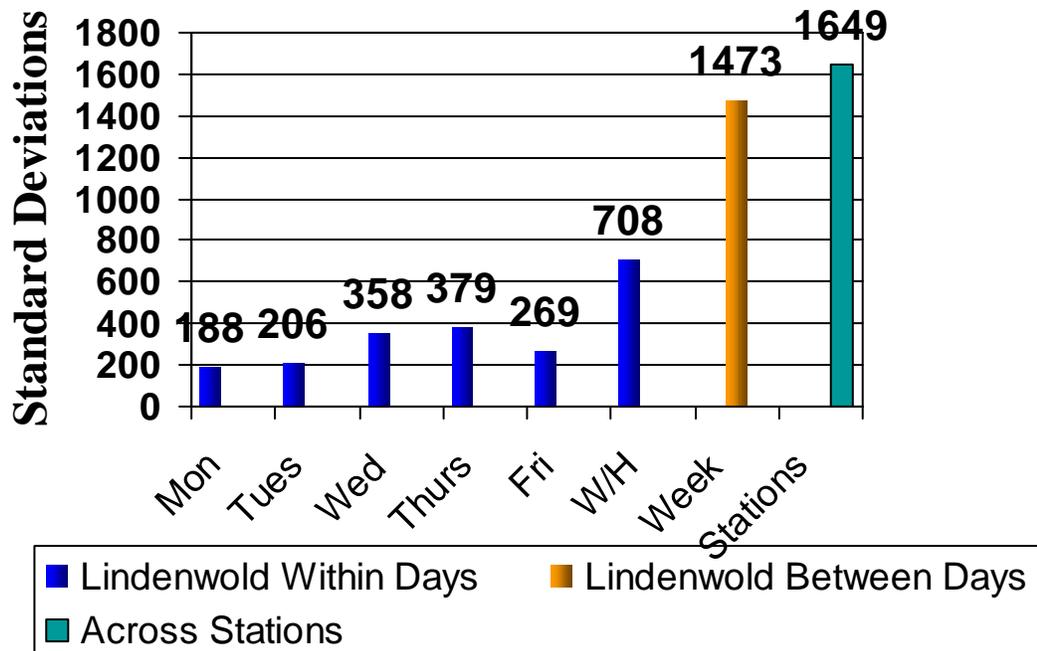
station. Figure 2 illustrates the daily totals for the Lindenwold, 8th Street and Market Street, and 16th Street and Locust Street stations for the first 60 days of 2001.

Figure 2. Daily Totals for the first 60 days of 2001.
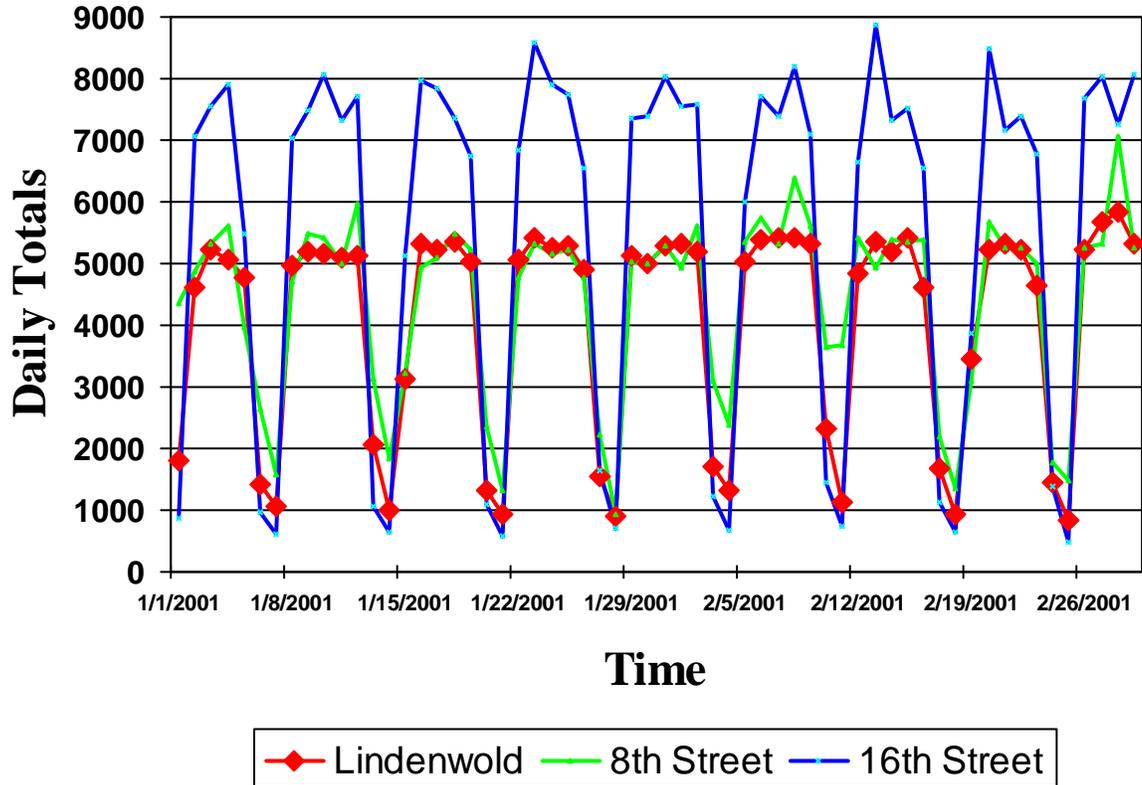
As illustrated in Figure 2, there is a rise and fall for the daily totals that appears cyclic with a weekly periodicity. This pattern holds over the entire year, as well as for the other stations. This cyclic behavior makes sense due to expected lower number of passengers during the weekends and the beginning of the work week (Monday) and the end of the work week (Friday). Usually one expects more passengers during the middle of the week.   Passengers on average use the same station primarily due to proximity to their residence or parking availability.   In addition, days in which passengers use the rail system are usually consistent over time. For example, some passengers may work from home on Mondays; therefore, they will consistently use the rail system only four of the available seven days per week, and in addition, they use the rail system on the same four days. Focusing on the Lindenwold station's tract in Figure 2, we see the Monday daily total from 1/1/2001 – 2/26/2001 range from 1806-5239. Tuesday daily totals from 1/2/2001 – 2/27/2001 range from 4613-5682. Wednesday daily totals from 1/3/2001 – 2/28/2001 range from 5157-5842.   The same period of time was investigated for the Philadelphia 8th Street and Market Street stations with daily totals ranging from 3083-5434 for Mondays, 4885-5748 for Tuesdays, and 5096-7067 for Wednesdays. These summary statistics illustrate that for a given station there is variability within the various Monday daily totals as well as for the other days; therefore, a within day classification per station source of variability. Similarly, within a station, there is variability across the days of the week. Finally there is variability across stations. This dissection of the variability in passenger load is allocated to three sources: Error variance (Variability across the replicates for a given day per station), variance across day of the week classification per station, and variability between stations.   These sources of variability are illustrated further in Figure 3 based on simple summary statistics from the available data.

Insert Figure 3 here.

Figure 3. Standard Deviations by Mon, Tues, Wed, Thurs, Fri, W/H, Week, Stations. Lindenwold Within Days: 188, 206, 358, 379, 269, 708. Lindenwold Between Days (Week): 1473. Across Stations: 1649.

Legend: ■ Lindenwold Within Days ■ Lindenwold Between Days ■ Across Stations

Per day of the week, standard deviation estimates are acquired for the 52 weeks of the PATCO data. Means per day of the week per station are derived. The Week column in Figure 3 is the standard deviation in these means for the Lindenwold station over the 6 classifications for the 7 days of the week. The last bar is the standard deviation for the mean estimates per station over the 365 data points. From these summary statistics, we see variability within a station for a given day of the week, which show heterogeneity per day, variability across the Day of the Week within a station, and variability across stations. Thus, we identified a nesting structure to the PATCO data. Now we need a statistical model which accommodates this structure.

Data that have a nested or hierarchical structure are common in a wide variety of disciplines, and similar methods for analyzing such data are found in these disciplines under different guises. The analyses fall under the heading of **random coefficient models (RCM), hierarchical linear models (HLM), and multilevel linear models (MLM)** (Littell et al., 1996;Goldstein, 1987; Bryk and Raudenbusch, 1996). Laird and Ware (1982) and Strenio, Weisberg, and Bryk (1983), working independently, proposed essentially identical approaches to the analysis of hierarchical data, both using the EM algorithm. Despite the prevalence of hierarchical data structures, classical analysis ignored such structures for many years, partly due to the underdevelopment of statistical models as well as statistical packages to model such data (Plewis, 1997; Singer, 1998; Suzuki and Sheu, 1998). With software such as PROC MIXED in SAS, HLM, MLWin, and SPSS Mixed Models, implementation of mixed effects models in a variety of fields has become more common. Recently, we've seen mixed effects models used in the transportation field (Deaton and Winebrake, 2000). The mixed effects model offers investigators a model to accommodate the hierarchy of the data. The main advantage of this approach is the variability can be partitioned at each level through the inclusion of random effects (Raudenbush, 1993). Ignoring the hierarchical structure by fitting a fixed effects model would result in an underestimation of the variance of the outcome, which results in underestimation of standard errors and "smaller" levels of statistical significance.

For the PATCO data, the hierarchy of the data consists of three levels:
- Reported observation for each day of the week and station combination
- Day of the week classification within station
- Station.

Days of the week are classified into six classifications: Monday – Friday, and Weekend/Holiday. Holidays will consist of major national and religious holidays. The three level mixed effects model is:

$$Y_{ijk} = \beta \text{ (grand average)} + \varepsilon_k + \varepsilon_{j(k)} + \varepsilon_{i(j(k))}$$

where

Y_{ijk} is the ith observation for the jth day classification of the kth station,

$\varepsilon_{i(j(k))}$ accounts for the random variance within a day classification,

$\varepsilon_{j(k)}$ accounts for the random day classification variance nested within station, and

(1)

ε_k accounts for the random station variance.

Note for Y_{ijk}, $i = 1, 2, ..., 52$ corresponding to the 52 weeks of the year, $j = 1, 2, ..., 6$ corresponding to the day of the week classifications, and $k = 1, 2, ..., 13$ corresponding to the 13 rail stations on the PATCO line. The assumptions for the random elements are:

$$\varepsilon_{i(j(k))} \quad NIID(0, \sigma_e^2)$$

$$\varepsilon_{j(k)} \quad NIID(0, \sigma_{DS}^2)$$

(2)

$$\varepsilon_k \quad NIID(0, \sigma_S^2).$$

The variance of Y_{ijk} is:

$$Var(Y_{ijk}) = \sigma_S^2 + \sigma_{DS}^2 + \sigma_e^2. \tag{3}$$

In this approach, each station's passenger load is characterized by a set of station and day classification random effects. These estimates are termed Best Linear Unbiased Prediction (BLUP) estimates. BLUP estimates are linear in the sense that they are linear functions of the data; unbiased in the sense that the average value of the estimates is equal to the average value of the quantity being estimated; best in the sense that they have the minimum mean squared error within the class of linear unbiased estimators; and prediction estimates to distinguish them from estimation of the random effects (Robinson, 1991). The station by day classification estimates will be used as the "expected values of passenger loads" for a given station on a specific day classification. BLUP Estimates are derived as follows:

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^{\mathbf{T}}\mathbf{V}^{-1}(\mathbf{Y} - \beta),$$

(4)

where $\mathbf{Y}$ is the vector of passenger loads. The superscript T indicates the transpose of $\mathbf{Z}$ and $\mathbf{V}^{-1}$ is the inverse of the estimate of $\mathbf{V}$. These estimates indicate deviations for each station from the grand average and day classification within station from the grand average.

To establish a threshold for concern, we focus on an effect size for the reduction of the grand average in equation (1). Effect sizes are unitless measures, which are usually applied to measure the magnitude of a treatment effect (Rosnow and Rosenthal, 1996). These effect sizes are standardized values centered with respect to a specified mean and standard deviation. The particular effect size used here was defined by Cohen (1988) and is commonly referred to as Cohen's d. The formula is:

$$d = \frac{M_1 - M_0}{s} \tag{5}$$

where M_1 is current passenger load, M_0 is the BLUP estimate of the passenger load. The standard deviation, s, is estimated by equation (3) evaluated at the estimates for each variance component. For the PATCO data we have: variability from station to station is 3,801,074; variability across days within a station is 12,987; and, the within station and day classification with station variability is 47,880. Each estimate is inserted in equation (3) to derive an estimate for the variability. The standard deviation estimate is acquired by taking the square-root of the variance estimate.

The main benefit of an effect size is the unitless property which allows for comparison to other studies regardless of outcome, design, or analysis. In our application, the effect size provides a unitless measure for the magnitude of change from the best linear unbiased predicted estimate.   We establish an acceptable reduction of less than 35% reduction in average estimated number of passengers for a station on a given day. If an observed passenger load reduces by more than 35%, then we have rise for concern.   This 35% reduction threshold is based on an examination of a 35% reduction in the grand average, which corresponds to an effect size of 0.5 as illustrated below:

$$d = \frac{M_1 - M_0}{s} = \frac{0.35 \times 2841.38}{\sqrt{3,801,074 + 12,987 + 47,880}} \approx 0.50$$

Cohen defined an effect size of 0.5 as a "medium" effect size, which corresponds to the expected number of passengers is at the 69th percentile of the observed number of passengers. While the effect size estimate per station and day within station based on the 35% threshold deviates around this 0.50 estimate for the grand average, using the grand average as a reference point for setting this acceptability threshold seems reasonable.

3.   Biosurveillance Application

We split the 2001 PATCO data in two halves. We use the first six months to estimate expected daily totals for each station on the PATCO line. The mixed effects model of equation (1) is used. Refer to the appendix to see the SAS PROC MIXED syntax. For simplicity, the mixed effects model is implemented on only the work week (6) (Monday through Friday). This reduces one level of complexity for the limited available data. Indices for (1) in this reduced sample are: Y_{ijk}, $i = 1, 2, ..., 26$ corresponding to the 26 weeks for the first half of the year, $j = 1, 2, ..., 5$ corresponding to the day of the week classifications for the 5 workdays, and $k = 1, 2, ..., 13$ corresponding to the 13 rail stations on the PATCO line.

Variance estimates for the three sources of variability were reported above. Corresponding standard deviations are 1949 for station-to-station, 113 for the across time within a station, and 218 for within time and station. The standard deviation estimates for station to station variability and the within station and day classification are consistent with the summary statistic estimates illustrated in Figure 3. The time within station source of variability is smaller here due to the exclusion of the weekend/holiday data points, which usually have a large deviation from the average due to much less commuters as compared to the work-week.

We produce BLUP estimates for the number of total passengers per station and day classification pair. This is implemented in SAS PROC MIXED through ESTIMATE statements as illustrated in the appendix. We use the second half of the data to determine if any daily totals fall below the threshold values (35% reduction in the BLUP estimates) and determine if there is some reason for the reduced passenger load.   Second half U.S. national and major religious holidays are identified and classified with a day classification of Weekend/Holiday. These days include July 4, Labor Day, Thanksgiving, Christmas Eve, Christmas, and New Year's Eve; therefore, these values are excluded from this example. This split sample approach was used by Stern and Lightfoot (1999) in their investigation of salmonella infection for the National Enteric Pathogen Surveillance Scheme data, although their statistical approach differed from the methods discussed here.

BLUP estimates are generated by equation (4) and passenger loads per station and day classification are produced, and illustrated in Table 1.
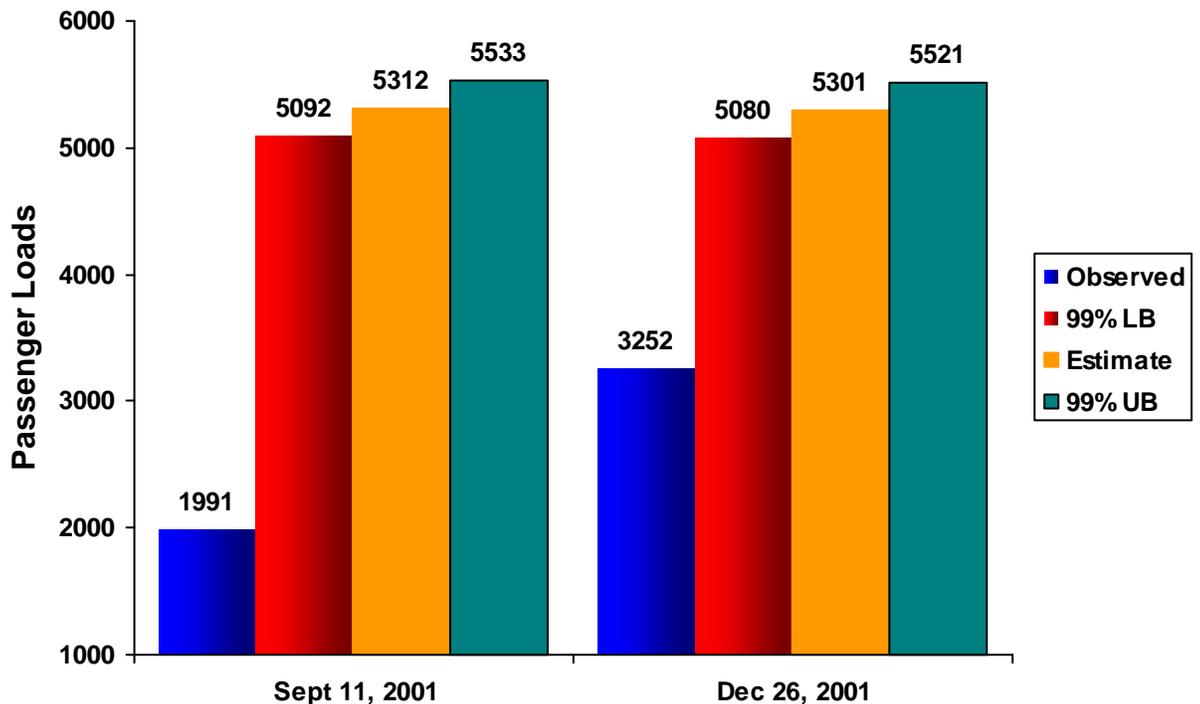
Table 1: Passenger Load Estimates

STATION	Monday	Tuesday	Wednesday	Thursday	Friday
NEW JERSEY STOPS					
Lindenwold	5140.4	5233.5	5221.7	5197.4	5092.7
Ashland	2049.7	2142.8	2131.0	2106.7	2002.0
Woodcrest	2975.4	3068.5	3056.7	3032.4	2927.8
Haddonfield	1699.4	1792.5	1780.7	1756.3	1651.8
Westmont	1700.0	1793.1	1781.4	1757.1	1652.4
Collingswood	1640.8	1733.9	1722.2	1697.8	1593.2

Ferry Ave.	2777.7	2870.8	2859.0	2834.7	2730.1
Broadway	1785.9	1879.0	1867.3	1842.9	1738.3
City Hall	770.0	863.1	851.4	827.0	722.4
PHILADELPHIA STOPS					
Phila 8th & Market	5219.4	5312.5	5300.7	5276.4	5171.8
Phila 10th & Locust	1494.8	1587.9	1576.1	1551.8	1447.2
Phila 13th & Locust	1721.5	1814.6	1802.8	1778.5	1673.9
Phila 16th & Locust	7478.2	7571.3	7559.5	7535.2	7430.6

Focusing our attention on Philadelphia's 8th Street and Market street station, there are two times when the observed passenger counts had more than a 35% reduction from the BLUP estimate for the given day classification. Figure 4 illustrates these two occurrences.

Figure 4: Philadelphia's 8th Street and Market Street station sizable reduction from expected passenger loads.

Both of these "small" passenger load estimates would raise some level of concern. With this data being historical data, we can find causes for the large declination of passengers for both dates. The September 11, 2001 occurrence corresponds to the terrorist attacks on New York and Washington, D.C. The Mayor of Philadelphia urged many workers to stay home; therefore, reduced number of passengers is expected. Similarly, December 26,

2001 is the day after Christmas; therefore, while this is still a "workday", many companies were closed and many workers did not return to work.

In both these scenarios, we were able to justify the large reduction. What if we can not? Are passengers starting to develop symptoms due to exposure to a biological pathogen? While the answer to this question is not known, the reduced passenger load could provide some evidence of peculiar behavior.

4. Conclusion

We have presented an application of mixed effects model for estimation of passenger loads on the PATCO regional rail line for any day or station. This mixed effects model has three benefits. First, it accommodates the hierarchical structure of the data, which has repeated observations per day classification nested in station. Second, it provides best estimates of the expected number of passengers per station for any day through the derivation of BLUP estimates. Third, there are a vast number of statistical packages to implement this model. We were able to establish a rule which would indicate a potential concern if the observed passenger load fell substantially below the expected count. A more than 35% decrease was our established rule, which resulted in two days where the passenger load fell below the criterion for the Philadelphia 8th Street and Market Street station.

Two alternate analytical methods are Cross Sectional Time Series (CSTS) and Generalized Estimating Equations (GEE). CSTS approach was used by Williamson and Hudson (1999) in their time-series analysis of public health surveillance reports. The CSTS approach usually focuses on time-series observations on each of several cross-sectional units, where the cross-sectional units are independent. The CSTS is a 2-level approach with time-series nested within station. For a given station and day of the week classification, we expect the number of passengers to be correlated. An advantage of the mixed effects approach is it allows for this third level of clustering and therefore, models this correlation directly. While for the CSTS model, accounting for this might be more challenging. Under certain specifications of the mixed effects model, the model can fit time-series type models (Verbeke and Molenberghs, 2000). So the mixed effect model can assess if the three-level model is a substantial improvement compared to the two-level CSTS model. This process was done by Gmel et al. (2001) in their assessment of Alcohol Consumption for European men. Their conclusions were "For pooled cross-sectional time series data, control of the potential impact of historical time is of the utmost importance. Hierarchical linear models constitutes a superior alternative to analyze such complex data". The second alternative is GEE. The GEE approach of Liang and Zeger (1986) is extremely popular for repeated measures designs, namely because of their consistent estimates of parameters in the model. The GEE approach is commonly called a population-averaged (PA) approach. In the PA approach, one is not concerned with random deviations per subject. For the PATCO data we're interested in the random deviations per day for a given station. We need this individual deviation which is sometimes refereed to as a subject-specific effect, which are used to produce the BLUP estimates. Hence, one limitation of the GEE approach is there are no BLUP estimates, and these BLUP estimates are needed to establish the threshold cutoff for each day of the week per station. These limitations of both the CSTS and GEE approach indicate that the mixed effects approach is as good as and possibly better than these alternate approaches.

A third approach would be to fit a fixed effects model, which ignores the within cluster per station, Fitting this model to the first quarter PATCO data, produces a variance estimate of 218.81. Compared to the solution of equation (3), 1965.18. In addition, the passenger load estimates from the Fixed effects model for the Philadelphia 8th Street and Marker Street station for Monday through Friday are, respectively, 5020.1, 5183.5, 5417.2, and 5246.9. These estimates do not match Table 1. Therefore, we have differences in variance estimates and the estimated average levels per station and day combination. We recognize the fixed effects model is not capturing all sources of variability properly; therefore, we have more confidence in the means estimated by the more complex HLM model.

Two limitations of this application focus on limitation of the BLUP estimates. Distribution theory associated with BLUP estimates is not nearly as well-understood as it is with conventional estimable functions (Littell et al., 1996). The variance of the BLUP estimates may experience shrinkage, since the observed data are shrunk towards the overall average since the prior means of the random effects is zero (Verbeke and Molenberghs, 2000). An additional limitation is the acceptability of the derived threshold rule. While the rule is based in a statistical framework, collaboration between statisticians, medical community, government, and regional rail managers is needed to derive a more justified rule.

The statistical community has made strides to become integral members of the war on Terrorism and is eager to help U.S. government respond to recent terrorist attacks and assist in planning for potential attacks (Banks, 2002). There are a vast number of diverse applications of statistical models for the war on terrorism. Recently, we've seen discrete event simulation models used to develop the public health infrastructure for bioterrorism response (Hupert et al., 2002). Space-time permutation scan statistics have been implemented for spatial disease

surveillance (Kulldorf et al., 2003). Time series analysis has been used to track over-the-counter medication sales as an early statistical detection of anthrax outbreaks (Goldenberg et al., 2002).

One statistical model is not going to provide us the answers to all questions and concerns. The model we discuss, as well as the papers discussed above focus on one area of biosurveillance. For the PATCO, the reduced passenger load may not be due to terrorist activities, as we saw in our illustration. For the most part, the reduced passenger load will not be due to terrorism. Our model serves as warning signs or one piece of the puzzle. Linking multiple warning signs could provide the necessary evidence of bioterrorist activities.

Consider a situation where one station on the PATCO on a random day fell below the established criterion. This on its own is only one piece of evidence. In addition, suppose there is an increase in the number of over the counter medication sales in the communities surrounding the train station, as well as an increase in patients seen by primary care physicians or patients entering emergency rooms. These three pieces of evidence, could provide necessary information to determine that a biological agent was used on the passengers at the specific train station. This would allow immediate notification to other passengers who have not become symptomatic or who have ignored the symptoms.

References:
Banks, D. (2002). *Statistical Issues in Counter-Terrorism*. Retrieved 10 October, 2002, from the American Statistical Association's Committee for Statisticians in Defense and National Security Web site: *http://www.statdns.org*.

Bartlett, J.G., Inglesby, T.V., and Borio, L. (2002). Management of Anthrax. Clinical Infectious Diseases, 35, 851-858.

Byrk, A., and Raudenbush, S. (1996). Hierarchical Linear Modeling: applications and data analysis methods. Newbury Park, CA: Sage Publishing

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd ed., Hillsdale, NJ: Lawrence Earlbaum Associates.

Deaton, M.L., and Winebrake, J.J. (2000). The Use of Mixed Effects ANCOVA to Characterize Vehicle Emission Profiles. Journal of Transportation and Statistics, 3, 49-64.

Evans, S., Kleinman, K., and Pagano, M. (2002). Bioterrorism and Biosurveillance. AMSTAT NEWS, 301, 41-43.

Flowers, L.K., Mothershead, J.L., and Blackewell, T.H. (2002). Bioterrorism preparedness II: the community and emergency medical services. Emergency Medicine Clinics in North America, 20, 457.

Gallop, R.J., Mode, C.J., Blank, K., Jurgens, S.M., and Schaben, C. (2005). Application of a Mixed Effects Model for Biosurveillance of a Regional Rail System, Journal of Data Science, *3, 353-370.*

Gmel, G., Rehm, J., and Frick, U. (2001). Methodological Approaches to Conducting Pooled Cross-Sectional Time-Series analysis: The Example of the Association between All-Cause Mortality and per capita Alcohol Consumption for Men in 15 European States. European Addiction Research, 7, 128-137.

Goldenberg, A., Shmueli, G., Caruana, R.A., and Fienberg, S.A. (2002). Early Statistical Detection to Anthrax Outbreaks by Tracking Over the Counter Medication Sales. Proceedings of the National Academy of Sciences of the United States of America, 99, 5237-5240.

Goldstein, H. (1987). Models in Educational and Social Research. New York, NY: Oxford University Press.

Hupert, N., Mushlin, A.L., and Callahan, M.A. (2002). Modeling the public health response to bioterrorism: Using discrete event simulation to design antibiotic distribution centers. Medical Decision Making, 22: S17-S25 Supplemental S.

Kulldorf, M., Mostashari, F., Heffernan, R., and Hartman, J. (2003). A Space-time permutation scan statistics for spatial disease surveillance. International Biometric Society Eastern North American Region Program and Abstracts for 2003 Spring Meeting, 32, 129.

Laird, N.M., and Ware, J.H. (1982). Random-effects models for longitudinal data. Biometrics, 38, 963-974.

Liang, K.Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, Biometrics, 73, 13-22.

Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). <u>SAS System for Mixed Models.</u> Cary, NC: SAS Institute Inc.

Plewis, I. (1997). <u>Statistics in Education,</u> London: Arnold Publishing.

Raudenbush, S.W. (1993). Hierarchical Linear Models and experimental design. In Lynne, E.K. (ed.), <u>Applied Analysis of Variance in Behavioral Science,</u> New York: M. Dekker.

Robinson, G.K. (1991). That BLUP is a good thing, <u>Statistical Science</u>, 6, 15-51.

Rosnow, R.L., and Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. <u>Psychological Methods</u>, 1, 331-340.

Singer, J.D. (1998). Using SAS Proc Mixed to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models, <u>Journal of Educational and Behavioral Statistics</u>, 24, 323-355.

Stern, L. and Lightfoot, D. (1999). Automated Outbreak Detection: A Quantitative Retrospective Analysis, <u>Epidemiology and Infection,</u> 122, 103-110.

Strenio, J.F., Weisberg, H.I., and Bryk, A.S. (1983). Empirical Bayes estimation of individual growth-curve parameters and their relationship to covariates, <u>Biometrics</u>, 39, 71-86.

Suzuki, S. and Sheu, C.F. (1998). Using PROC MIXED in Hierarchical Linear Models: Examples from two- and three-level school-effect analysis, and meta-analysis research. <u>1998 Midwestern SAS Users Group Conference Proceedings.</u> Cary, NC: SAS Institute Inc.

Verbeke, G., & Molenberghs, G. (2000). <u>Linear Mixed models for Longitudinal Data</u> New York: Spinger-Verlag.

Williamson, G.D, and Hudson, G.W. (1999) A Monitoring system for Detecting Aberrations in Public Health Surveillance Reports. <u>Statistics in Medicine</u>, 18, 3283-3298.