



Published in final edited form as:

Math Comput Model. 2008 ; 48(5-6): 929–939. doi:10.1016/j.mcm.2007.11.016.

Using Influenza-Like Illness Data to Reconstruct an Influenza Outbreak

Philip Cooley, Laxminarayana Ganapathi, George Ghneim, Scott Holmberg, and William Wheaton

RTI International

Abstract

The objective of this study was to reconstruct the type A influenza epidemic that occurred in the Research Triangle Park (RTP) region of North Carolina during the 2003–04 flu season. We describe an agent-based influenza transmission model that uses Influenza-like Illness (ILI) data gathered from state agencies to estimate model parameters. The design of the model is similar to models represented in the literature that have been used to predict the impact of pandemic avian influenza in Southeast Asia and in the continental United States and to assess containment strategies. The focus of this model aims to reconstruct a historical epidemic that left traces of its impact in the form of an ILI epidemic curve. In this context, the work assumes aspects of a curve fitting exercise.

Keywords

Validation; Infectious Disease Models; Influenza; Influenza-like Illness Surveillance

1. INTRODUCTION/APPROACH

The goal of the model we developed was to increase understanding of the underlying mechanisms of a particular phenomenon i.e., a previously occurring influenza epidemic. In this paper, we demonstrate how we reconstructed the type A influenza epidemic that occurred in the Research Triangle Park (RTP) region of North Carolina during the 2003–04 flu season. The RTP region is made up of six counties (see Table 1), and its population during that period was slightly more than one million persons. The model attempts to estimate the severity of a historic epidemic by reconstructing the epidemic curve. The reconstruction of the epidemic uses an agent-based model (ABM) of influenza transmission that generates epidemics in response to user supplied assumptions about the transmissibility traits of the flu pathogen and the number of social network specific contacts. The only available legacy information recorded about the epidemic is the number of cases of influenza like illness (ILI) processed by local area hospital emergency rooms during the epidemic period.

In the following sections, we describe the ILI data used and the structure of our model; we then define the model's parameters. Next, we describe the power sweep of the exogenous model

© 2007 Elsevier Ltd. All rights reserved.

Send all correspondence to: Phil Cooley, RTI International, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709, Voice: (919) 541- 6509, Fax: (919) 541- 6178, E-mail: pcc@rti.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

parameters to identify a candidate epidemic that resembles (by some criteria) the ILI epidemic. Finally, we provide a discussion of our results.

This study is important for several reasons. First, ILI data is a source of data that has not been (to our knowledge) used in the context defined here. Our assumption is that ILI epidemic curves coexist with corresponding past influenza epidemics; consequently, the influenza epidemic curve and the ILI epidemic curve will have the same footprint and shape. The ILI data uses a standardized case definition, and even without knowing the proportion of ILI cases that are ultimately diagnosed as influenza cases, these data could provide additional information on the extent and transmissibility of the epidemic. Furthermore, ILI data is being collected nationally and is growing in coverage and quality. Therefore, the first purpose of our study was to demonstrate how to use this burgeoning data to reconstruct past influenza epidemics. Second, our reconstruction provides a better understanding of the extent of historical epidemics, which will help other modelers predict future epidemics. Accurately reconstructing past epidemics provides an additional element for validating predictive models, through retrodiction or other methods.

1.1 Other Published Agent Based Models of Influenza Transmission

Developing strategies for mitigating the severity of a new influenza pandemic is a global public health priority. The Models of Infectious Disease Agent Study (MIDAS) is a research partnership between the National Institutes of Health (NIH) and the scientific community to develop computational models for policymakers, public health workers, and researchers to help all parties make better-informed decisions about emerging infectious diseases—both man-made and naturally occurring. MIDAS researchers have developed models that can assist the public health community understand how best to respond during outbreaks and epidemics. Influenza prevention and containment strategies include antivirals and vaccines, as well as nonpharmaceutical measures, which include case isolation, household quarantine, school or workplace closure, and travel restrictions. As part of the MIDAS initiative, Ferguson et al. developed large-scale mathematical models to explore the complex landscape of intervention strategies. Specifically, they developed a large-scale epidemic simulation to examine intervention options using Southeast Asia (attempting to contain the epidemic at its source) [1], Great Britain, and the United States as examples [2].

Longini et al. also examined the possibility of containing a flu epidemic at its source by modeling a rural region of Thailand [3].

Germann et al. [4] used a large-scale stochastic simulation model to investigate the spread of a pandemic strain of influenza virus through the U.S. population of 281 million individuals for R_0 (the basic reproductive number) from 1.6 to 2.4. They modeled the impact a variety of levels and combinations of influenza antiviral agents, vaccines, and modified social mobility (including school closure and travel restrictions) had on the timing and magnitude of this pandemic.

Eubank et al. [5] developed a discrete event simulation approach that uses random sample individual-specific activity structures. These structures provided the time, locations, and types of activities for all individuals in Chicago, Illinois. Contacts between individuals in the population were computed from the times and places of this activity set. An influenza model was used to simulate disease propagation through the population based on this contact structure. This model's main assumption regarding transmission is that it occurs at a fixed rate, so the probability of becoming infected is a function of duration of contact rather than type of contact. The model also collected and displayed characteristics of the disease, such as the attack rates by various demographics. Six scenarios were simulated and analyzed. These scenarios differ in the percent of cases diagnosed (and treated) and the percentage of persons complying with

social distancing directives. While the severity of the pandemic drastically changes as these two factors change, many similarities between the different simulations exist.

There are also unique and important models outside of MIDAS. For example, a study by Glass et al. [6] addressed the role of social distancing on the spread of influenza in the US. The powerful influence of school aged children was highlighted, with school closure and keeping teenagers at home reported to reduce attack rates by 90%. This model is an agent-based model with many characteristics similar to the MIDAS models. For example, the study emphasized the characteristics of social networks and their role in disease spread. One important difference in model structure is that this model focuses on a synthetic community substantially smaller than those studied by the MIDAS models. However, there are consistent results reported between this model and the MIDAS models.

Another example is provided by Haber et al. [7]. This model is a stochastic simulation model that is derived from a model described in Longini et al. [8]. The simulated region is described as a small urban US community that is infected by a H2N2 virus with properties that characterize the pandemic of 1957–58. For each simulated day, a susceptible person makes contacts with other persons within mixing groups that may lead to infection. The probability that a person becomes infected depends on the following:

- The number of different persons with whom the index person has contact with in each mixing group,
- The total duration, in minutes, of all the contacts with each person pair, and
- The per-minute rates of infection transmission if the contacted person is infectious.
- The number and duration of contacts may be different on weekdays and weekend days.

These traits differ from the MIDAS models and the model described here.

2. DESCRIPTION OF INFLUENZA-LIKE ILLNESS (ILI) DATA

The 2003–04 U.S. influenza season began earlier than previous seasons and was moderately severe; influenza A (H1), A (H3N2), and B viruses co-circulated, and the predominant strain was influenza A (H3N2) [9]. Fig. 1 displays the epidemic based on Influenza-Like Illness (ILI) sentinel data reported at participating hospital emergency rooms in the entire state of North Carolina as well as in the RTP region.

The ILI data has the following limitations.

- It describes a set of symptoms that embed influenza along with a number of other reportable conditions that include colds, pneumonia, and miscellaneous respiratory conditions.
- The sentinel hospitals that reported the ILI data in 2003 included only a small portion of North Carolina's emergency room facilities (under 10%).
- There is no estimate of the proportion of persons with flu that use the emergency room as their primary form of care.

The Flu-Isolates curve in Fig. 1 is defined as influenza virus isolates identified during the last 3 weeks by the State Laboratory of Public Health. Fig. 1 also shows ILI curves reported by the General Communicable Disease Control Branch of the North Carolina State Laboratory of Public Health, which is part of the U.S. Influenza Sentinel Physicians Surveillance Network. One of the network's functions is to monitor the status of statewide influenza activity. Sentinel physicians, university health centers, and public health agencies report ILI to the Centers for

Disease Control and Prevention (CDC) each week and collect representative samples for virus strain identification. Note that the data are tabulated by week. The ILI case definition criteria are any combination of fever (100°F or higher, oral or equivalent) and cough or sore throat. While information provides important epidemiologic information to the State Health Department for monitoring influenza activity in North Carolina, and supports CDC influenza surveillance activities throughout the United States, it does not provide a measure of the size of the epidemic in question.

Two ILI measures are shown in Fig. 1. The first represents all cases reported in North Carolina, and the second identifies all cases reported for patients living in the RTP region.

Fig. 1 suggests that the epidemic began in mid-October, peaked in mid-December, and extended into early February. The model we developed for the RTP region generated a flu epidemic with these characteristics.

2.1 Emergency Room Information

Table 2 summarizes the daily emergency room ILI visit data in the six-county RTP region. Fig. 1 shows these data as RTP Cases. In Table 2, the column Weekly Raw Counts identifies the actual visit data observed for all RTP residents as recorded by participating hospitals. The data in Table 2 represent ILI cases that are true influenza cases plus ILIs that are not part of the influenza epidemic (e.g., colds). We estimated non-flu cases by computing the average daily number of ILI cases observed prior to the rapid growth of cases, which began in mid-October 2003. We assumed that cases are the non-flu component of the ILI data. The Adjusted column represents an estimate of the actual epidemic with the non-flu illness cases removed from all reported cases.

The adjusted data suggest that the epidemic started the week of October 12 and ran through January 31st for a total of 105 days. Also, if we assume that the shapes of the ILI epidemic and flu epidemic are the same, we can determine a key feature of the epidemic, which is represented by the measures of skewness and kurtosis of the ILI distribution in Table 2 [10]. Assuming that the emergency room visit curve and the influenza epidemic curve for the RTP region have the same shape, we constructed an epidemic curve for the RTP region with skewness and kurtosis measures close to 1.284 and 3.661, respectively. This is a key assumption behind our method.

We assert that just as the ILI epidemic of 2003–04 began appearing in the RTP emergency rooms around October 12th (and was therefore circulating in the population up to two weeks earlier), the 2003–04 flu epidemic began to circulate at the same time. We also assert that the shape of the ILI epidemic and the flu epidemic are the same, and both can be represented by a curve with a skewness measure of 1.284 and a kurtosis measure of 3.661. However, a major problem was how to determine the scale of the influenza epidemic from the ILI data because while we could estimate them, we do not know what percentage of ILI cases are actually influenza cases.

3. MODEL DESCRIPTION

3.1 General Structure

The Susceptible-Infected-Recovered (SIR) model is frequently used to simulate the natural spread and course of an epidemic, like influenza, in a small community. The SIR model considers people born into a disease state labeled *susceptible*. On contact with people who are infectious, susceptible people (S) move into the infectious state (I) category. After a number of days, the people then recover and move to the recovered state (R). Because they are now immune to the pathogen, they remain in state R for the rest of the time period of interest.

We developed an agent-based version of the SIR model with 1,037,533 individuals (the 2000 population estimate for the six RTP counties) as circulating agents. Previous studies suggest that in modeling complex phenomena, the extra complexity captured in agent-based models (ABMs) sometimes leads to different conclusions from those reached by a differential equation model built to the same specifications, although in many cases the two models have very similar behavior. In our model, we used an ABM approach to describe the dynamics of disease spread because the period of performance overlaps with both school holidays and weekends. In both cases, attendance at school is interrupted, and because school closure is considered by many to be a major factor in disease spread [2], and [4] this interruption was explicitly factored into the reconstruction of the epidemic. Because of the simulation nature of the ABM, school closure processes are easily incorporated into the ABM approach.

Further analysis suggests that, in the context of diffusion, disaggregating a population into agents adds additional insights in dealing with a locally structured network. In the case of locally dense networks, the main mode of behavior is usually well captured by a calibrated differential equation model [11].

The specific design of our model also incorporates features similar to those defined by Ferguson et al. [2] and Germann et al. [4]. Agent characteristics that the model tracks are age, sex, occupation, household location, household membership, school assignment (if student or teacher), work location assignment (if employed adult), work status, and disease status. Agents dwell in households that are distributed in a manner consistent with the U.S. 2000 Census data at the block group level similar to those defined in [4]. School-aged children are assigned to a school according to a modified gravity model that also accounts for school enrollment information in ways similar to those defined in Ferguson et al. [2]. Working adults are assigned to workplace locations according to their occupational status. During the working day, these adults mix with other people assigned to that workplace. Specific data on the health care component of the population were purchased from an occupational survey and used to assign health care workers to known health care places of employment (hospital, clinic, office, etc.).

3.2 Social Network Structure

Complex social network patterns are described in Eubank et al. [5]. Our approach divided the complete set of contacts involved in disease transmission into six categories. Each agent interacts in some or all of five of six social network categories. We defined the categories as follows:

- **Schools:** We used a modified gravity model to assign school-aged children to school locations, with the location of each school, its enrollment, and the household location of each child known. Children interact at school only with children who attend the same school. Children interact with other children in their classroom more closely than their non-classroom peers.
- **Workplaces:** Using the modified gravity model, we assigned adults with appropriate occupation codes to work locations. The location of some places of work (e.g., hospitals and clinics), the number of persons employed at the site, and the household location of each adult are known. Adults with certain occupations interact at a workplace only with other adults who work at the same site. In general, most workers interact with persons employed at the same location; workers were assigned to a subgroup of peers that they interact with more closely. There are certain occupations (clerks, etc.) that interact with persons *outside* of their occupation code while at work. This is also represented in the model.
- **Public transportation:** A portion of the working population, the school population, and the nonworking population use public transportation to travel to workplaces, to

schools, and to shopping centers. Persons using public transportation interact only with other persons using public transportation. Currently, only school bus passengers were represented in the model.

- Family or households: We generated a synthetic population of RTP residents who occupy RTP households for each simulation experiment. This population represents families and/or roommates that occupy houses and apartment complexes. Families and roommates interact only with other family members or persons that are part of the same household.
- Neighborhoods: Interactions with neighbors occur based on relative household locations.
- Communities: Interactions with the community were also based on relative locations. For example, persons will tend to frequent the nearest mall and interact with other persons visiting that mall. This element was intended to represent shopping behaviors of adults and students and the “hanging out” behaviors of students.

3.3 Disease Natural History Data

We drew model parameters from the most recent and best available published literature sources. Model parameters used to describe the strain of flu were based on recent data that characterize the H3N2 strain of flu as defined in Longini et al. [8]. These parameters are the basis of the parameters used by Longini et al. [3], and Germann et al. [4].

4. THE MODEL EXPERIMENTS

As we discussed earlier, we used our model to reconstruct a prior epidemic. The reconstruction process consisted of the following steps:

- Develop a stochastic ABM that limns a SIR epidemic transmission process.
- Obtain the data that describes the RTP scenario from the disease transmission scenario point of view and link it to the ABM. This combination of model and data provides a mechanism for generating epidemics.
- Use the ILI data to assess “target” legacy epidemic characteristics.
- Run a power sweep of exogenous model parameters, including the number of seeds, that alter transmission effectiveness and number of contacts. Note that by including the number of initial infected persons seeding the epidemic, each unique combination of exogenous model parameters determines both the shape and scale of the generated epidemic.
- Replicate the epidemic 100 times with these parameters fixed but with different random number sequences. Use the “average” epidemic and calculate the skewness and kurtosis of this epidemic.
- Compare the skewness and kurtosis measures for the generated epidemics with the target values produced from the ILI data.
- Compile a distribution of “well-fitting” epidemics.

Figure 2 represents the process graphically. We used a SIR flu model generator in conjunction with a power sweep strategy to generate 9600 flu epidemics of the RTP region. We then applied a 5% and 1% ILI target criteria that resulted in a small number of epidemics that limn the traits of the 2003–2004 ILI epidemic in the RTP region. We reviewed the distribution of the resulting infection prevalence estimates that are an attribute of the epidemics for consistency and

stability, and the epidemic that best fit the ILI criteria was defined as the reconstructed epidemic.

4.1 Model Exogenous Parameters

This section defines the exogenous model parameters that we varied as part of the reconstruction process. The values of these model parameters determine the assumptions regarding the transmissibility of the 2003–2004 flu pathogen, the number of contacts made by infected persons, the number of infected persons at the start of the epidemic, and the seasonal characteristics of the epidemic. Note that these parameters influence both the shape and the scale (attack rate) of the epidemic. We defined the parameters and their feasible ranges as follows:

- The Transmission Multiplier (TM): this is a positive number (1.0–2.0) that scales all contact transmission probabilities. This parameter, multiplied by the transmission probability that is risk group specific, determines the transmissibility assumption.
- Social Network Multiplier (SNM): this is a positive number (0.5–1.0) that scales the size of each social network that an infected person makes daily contact with. This parameter, multiplied by the default size of the social network, determines the number of contacts. Note that this multiplier has no influence on number of household contacts.
- Seasonal Peak Day (SP): this is the length in days (40–100) before seasonal factors begin to reduce transmission. Seasonal characteristics are not understood. Many disease specialists feel that seasonal features are caused by changes in people’s behavior (e.g., greater outdoor activities during warmer weather periods).
- Seasonal Damper Multiplier (SDM): this is a damping factor (0.01–1.00) that causes the transmission multiplier to decline after the epidemic has progressed beyond the Seasonal Peak day.
- Model seeds (SEED): this is the number of persons (1–10) that are infected at the start of the epidemic. Each model run was started by seeding the epidemic with a number of infected persons (aged 15 to 60). The seeds were drawn randomly and with age restrictions. Depending on random actions, a number of the replicates did not result in an epidemic. These cases were excluded from the calculation of the “average” epidemic summary for each set of model parameters.

Note that seasonal effects are problematic for a number of reasons. First, there is no obvious biological reason that reduces pathogen infectiousness. However, many epidemics (such as the one that is the subject of this manuscript) start and stop within a single winter season for no apparent reason. There could be a reduction in mixing caused by a holiday or by milder weather that encourages people to move outside, but there is no readily apparent explanation.

4.2 Important Model Endogenous Parameters

There are a number of endogenous factors affecting disease transmission incorporated into the model that are fixed for each generated epidemic. These factors include:

- sickness—50% of children stay home from school and 50% adults stay home from work;
- death—persons who die from flu cease to interact with the survivors;
- day of the week—children stay home from school along with most working adults, who do not work on the weekends. This is offset by 50% more frequent community and neighborhood contacts on weekends; and

- vaccination rates—rates by state and by age group were obtained from http://wonder.cdc.gov/wonder/PrevGuid/m0049614/m0049614.asp#Table_1.

These data were used to assign a vaccinated status to susceptible persons in the simulation. Vaccinated persons were assumed to be unable to transmit disease.

5 RESULTS

Using the measures of skewness and kurtosis derived from the ILI emergency room visit information as target values, we performed a power sweep of the key exogenous model parameters to generate disease incidence patterns from 9,600 exogenous model parameter combinations (note: many of these combinations produce duplicate epidemic curves or produce epidemics that are far from satisfying the ILI conditions of 105 days, etc.). Our premise is that the best fitting and most feasible epidemics will be consistent with the shape of the curve defined by ILI reporting patterns, which could then be used to assemble the 2003–2004 RTP epidemic curve.

We used the ILI data to assess the length of the flu epidemic window (105 days) and the occurrence of the peak flu period during the window (56 days after the start of the epidemic). Each epidemic generated by the model recorded the influence of the exogenous parameters on the scale and shape of the epidemic curve as measured by the skewness and kurtosis. The power sweep exercise identified the set of model parameter values that generate a flu epidemic with the same kurtosis (and skewness) characteristics that were observed in the adjusted ILI visit data.

In summary, the SIR-based epidemic generator generates an epidemic for each unique combination of exogenous parameters. We performed a power sweep by varying the exogenous parameters and recording the measures of skewness and kurtosis that we are trying match and that represent the most feasible flu epidemics. By this process, we identified a distribution of feasible epidemics that also exhibit the shape properties we measured in the ILI data.

Table 3 presents a summary of the power sweep process. We performed the power sweep to determine the values of the exogenous parameters that produced an influenza epidemic with a target kurtosis value of 3.661 and a skewness value of 1.284 within an epidemic period of 105 days. This table portrays the prevalence of infections distribution for epidemics that are simultaneously within a 5% of the target values for skewness and kurtosis; i.e., with skewness that ranges from 1.220 to 1.348 and with kurtosis limits between 3.480 and 3.884. There were 117 unique epidemics generated that were within the 5% goodness-of-fit range. The range of attack rates among the 117 varied from 4.2% (42,358) to 12% (121,138).

Table 3 indicates that of the 117 (< 2% of the 9,600 total) epidemics that exhibit characteristics that fall within the 5% criteria, more than 50% of these exhibit prevalence estimates between 50,000 and 75,000 (60 of 117) and more than 80% fall into the range 50,000 to 95,000 (94 of 117). This suggests that epidemics with the desired ILI target properties that are produced by the SIR epidemic generation model are likely distributed as a log normal distribution (i.e., an error distribution that is constrained from below) with a mean prevalence of around 80,000 (8%).

The 5 “best” epidemic curves that best fit the target criteria are shown in figure 3.

The average number of infections recorded by the epidemic depicted in Fig. 3 is 70,696, which corresponds to an attack rate of 7%. The patterns of figure 3 also suggest a significant weekend effect that reduces infection due to school closures and partial workplace closures, even though a 50% increase in community contacts is assumed to occur when schools/workplaces are

closed. Because the epidemic began around October 12, the school break for Christmas fell approximately two months later, around mid December or Day 56.

The timing of the epidemic could also lead to the conjecture that the ILI visit results shown in Figure 1 could be affected by the annual school winter holiday. In this scenario, in the midst of the epidemic, schools close for the Christmas holidays. We attempted to represent this scenario in a second set of runs, in which children of school age stayed home from school starting in mid-December and remained home for 17 days. During that period, we assumed the populace engaged in weekend social network behaviors, which are characterized by more frequent mall and retail store visits and other community interactions but no classroom interactions. Using these assumptions, we generated a set of simulation runs that used the same shape criteria as above. Specifically, at day 56, the model assumes an extended weekend of 17 days occurs, which would correspond to a 17 day school closure.

This is readily accommodated by an ABM, and we modified the model to account for school closures in mid December (day 56) for 17 days to early January (day 73). We then performed a power-sweep of the exogenous model parameters to re-estimate the best fitting (with respect to the target skewness and kurtosis values) modified model. Table 4 summarizes the power-sweep results. Consistent with table 4, this table portrays the infection prevalence distribution for epidemics that are simultaneously within a 5% goodness-of-fit for skewness and kurtosis. There are 123 epidemics that satisfy the 5% goodness-of-fit criteria. The infection prevalence ranges from a low of 38,416 (3.8%) to 139,166 (13.9%) which is much larger than the pattern shown in Table 4. This is also demonstrated at the higher value for the standard deviation of the prevalence mean.

In summary, the epidemics generated under the school closure assumption produced an epidemic curve that was not demonstrably better than the well fitting epidemic curves that assumed schools were open. We provide the following interpretation:

- School holidays are unlike weekends, because many within school contacts continue throughout the holiday period,
- The model overstates the weekend effect observed in Fig. 2.
- An inaccurate epidemic period was estimated from the ILI data and biased the results.

Also, all of the epidemics that fit the criteria estimated seasonal effects of various degrees. One explanation is that the number of persons immune to the circulating pathogen was underestimated, either because of an vaccinations within the area were underreported, because persons acquired partial immunity from prior exposure to similar (past) influenza strains. This is a significant problem for reconstructing prior epidemics: and in that sense assuming no prior exposure (as one would with a new strain of flu) is an easier problem to model. We feel that the seasonal effects in part compensate for this underestimate of regional immunity.

Figure 4 presents both “best fitting” epidemics (with respect to the ILI criteria) with and without school closures. The School’s Closed Epidemic assumes school closure beginning in mid December and lasting for 17 days; the School’s Open Epidemic assumes no school closure through out the epidemic period. As figure 4 indicates, there is little difference between the two curves shown. The School’s Open epidemic has a slightly lower overall attack rate but conforms to the same overall weekday and weekend patterns.

7. DISCUSSION

We used local ILI data to reconstruct the shape of the local RTP 2003–04 flu epidemic. We used our epidemic model generator to conduct a power sweep of key model parameters. As

part of this process we generated many epidemics and recorded the impact of the values of the exogenous parameters on scale and shape measures of the epidemic curves. Our analysis suggested that the attack rate bounds were between 4.1% and 12.1%, but a likely best attack rate was approximately 8%, which is based on the mean attack rate of epidemics that closely fit the target ILI criteria.

We also solicited the opinions of NC flu experts. They bounded the scale of the attack rate of the epidemic between 10% and 15%, which was higher than we estimated. Finally, we used our assessment of school closure to develop an alternative epidemic curve estimate for the flu epidemic in the RTP region for the 2003–04 flu season. This estimate differed little from the original epidemic generated, but the standard deviation of the attack rate derived again from best fitting epidemics was substantially higher under the assumption of school closure.

The central issue we raise is: How can the accuracy of the developed model be assessed in the absence of accurate data that describes the flu epidemic for the 2003–04 flu season? This is a fundamental problem of the modeling process; there is usually sufficient data to build a model but insufficient data to accurately assess model performance. The solution we developed was to identify ancillary data that proxy evidence of the character of the phenomena being modeled and use that information in a curve fitting context to reconstruct the epidemic. It is also important to examine other influences on the phenomena that affect the interpretation of the results.

We believe that the epidemic curve of Fig. 3 is a crude representation of the actual epidemic that occurred in the six-county RTP region. However, while we believe that this figure is representative of the “actual” epidemic that occurred, we also acknowledge that the model was developed using many assumptions that are difficult, if not impossible, to corroborate with available information. A significant assumption we used in our approach is that the trends of ILIs, as reported to the emergency room in a single large RTP hospital, would mimic influenza trends in the RTP region. Thus, although ILI data significantly underestimate the scale of the true epidemic, the assumption behind the model is that the trends in those reports mirrored the trends (shape) of the true influenza epidemic within the RTP region, and that by using the SIR generator and treating the number of seeds needed to initiate the epidemic, we can also estimate epidemic attack rate.

A secondary issue raised by our model is whether a weekend effect is significant. Our model produced mixed results. On the one hand, it indicates that on the weekend, the reduction of mixing significantly reduces influenza transmission. Other ABMs that we are aware of assume business as usual, and the weekend dipping phenomena as recorded by our model is not present, see [1], [2], [3], [4], [5], and [6]. The Habler et al. model is an exception [7]. On the other hand, our reconstruction of an epidemic in the presence of a school closure suggests that weekend effects and school closure effects may not operate in the same fundamental manner. This is an important issue to understand because the weekend effect, if it exists, produces a flatter epidemic curve with a lower peak and thus has important preparedness implications.

We ultimately wonder “What does happen on the weekends? Is there a weekend effect? Do students have just as many contacts with their peers on the weekend as they do when they attend school?” We are hoping that the 8 NPI studies currently being developed under CDC auspices will shed some light on this important issue. Until those studies prove the contrary, we believe the evidence suggests a weekend effect and we continue to represent it in our results.

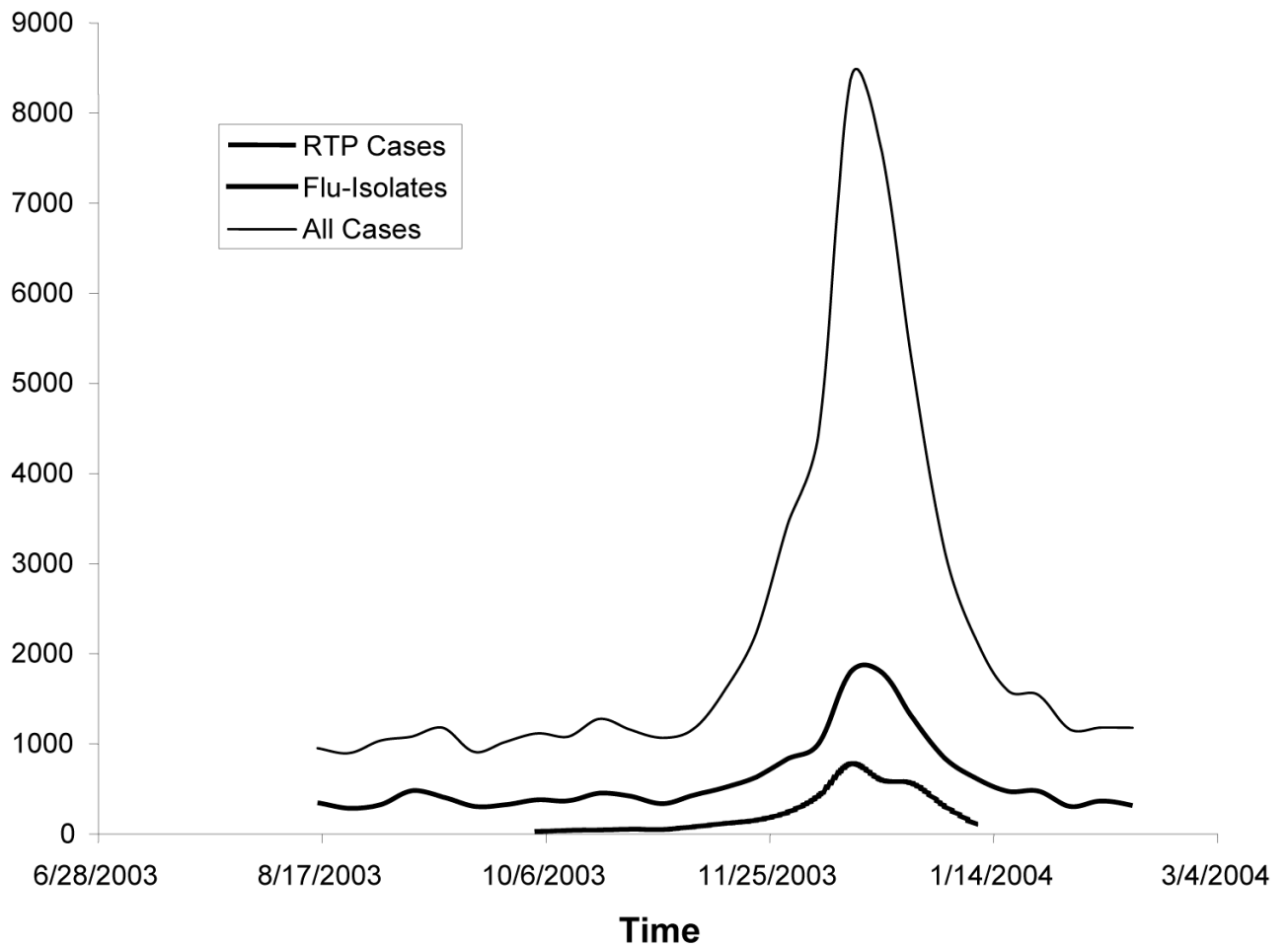
Acknowledgments

We thank the National Institute of General Medical Sciences MIDAS Program for research funding. We thank members of the MIDAS consortium and the North Carolina state public health officials: Jeffrey Engel, State

Epidemiologist, Megan Davies, CDC Career Epidemiological Field Officer, and Kristina Simeonsson, Medical Epidemiologist, in charge of influenza surveillance for the state of North Carolina, for useful discussions. We thank the MIDAS informatics group for computational resources.

9. REFERENCES

1. Ferguson N, Cummings DAT, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 2005 September 08;437:209–214. [PubMed: 16079797]
2. Ferguson N, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. *Nature* 2006 July 27;442:448–452. [PubMed: 16642006]
3. Longini I Jr, Nizam A, Xu S, et al. Containing pandemic influenza at the source. *Science* 2005 August 12;309:1083–1087. [PubMed: 16079251]
4. Germann TC, Kadau K, Longini IM Jr, Macken CA. Mitigation strategies for pandemic influenza in the United States. 2006 April 11; PNAS.
5. Eubank S, Guclu H, Anil Kumar VS, et al. Modelling disease outbreaks in realistic urban social networks. *Nature* 2004 May 13;429:180–184. [PubMed: 15141212]
6. Glass RJ, Glass LM, Beyeler WE, Min HJ. Targeted Social Distancing Design for Pandemic Influenza. *Emerging Infectious Diseases* 2006 November;12(11):1671–1681. [PubMed: 17283616]
7. Habler MJ, Shay DK, Davis XM, Patel R, Weintraub E, Orenstein E, Thompson WW. Effectiveness of interventions to reduce contact rates during a simulated infectious pandemic. *Emerging Infectious Diseases* 2007 April;13(4):581–589. [PubMed: 17553273]
8. Longini IM, Halloran ME, Nizam A, Yang Y. Containing pandemic influenza with antiviral agents. *Am J Epidemiol* 2004;159:623–633. [PubMed: 15033640]
9. Centers for Disease Control and Prevention. Update: Influenza activity--United States and worldwide, 2003–04 season, and composition of the 2004–05 influenza vaccine. *MMWR*. 2004 53(25):547–552. Available at: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5325a1.htm>. [PubMed: 15229411]
10. Engineering Statistics Handbook. [Accessed 12/12/2005]. Available at: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>
11. Rahmandad, H.; Sternman, J. MIT Sloan School of Management. Cambridge, MA 02142: System Dynamics Group; 2005 May [Accessed 12/12/2005]. Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models. published online Available at: http://www.mit.edu/~hazhir/papers/AgentBasedvsDE_July15.pdf
12. Laine, T. Methodology for comparing agent-based models of land-use decisions. Bloomington, IN: Indiana University, Computer Science Department and the Cognitive Science Program; 2004 [Accessed 12/12/2005]. Published online Available at: <http://simon.lrdc.pitt.edu/~iccm/proceedings/DCabstracts/Laine.Abs.pdf>

**Fig. 1.**

NC Sentinel Surveillance Data: January 15, 2004

Note: Flu-Isolates are defined as “Influenza Virus Isolates Identified during the last three weeks by the North Carolina State Laboratory of Public Health.”

Source: NC Influenza Surveillance Report, January 15, 2004; see

<http://www.charmeck.org/Departments/Health+Department/Top+News/News+Archive/2004/Home.htm> “RTP Cases” and “All Cases” are data reported by the General Communicable Disease Control Branch of the North Carolina State Laboratory of Public Health.

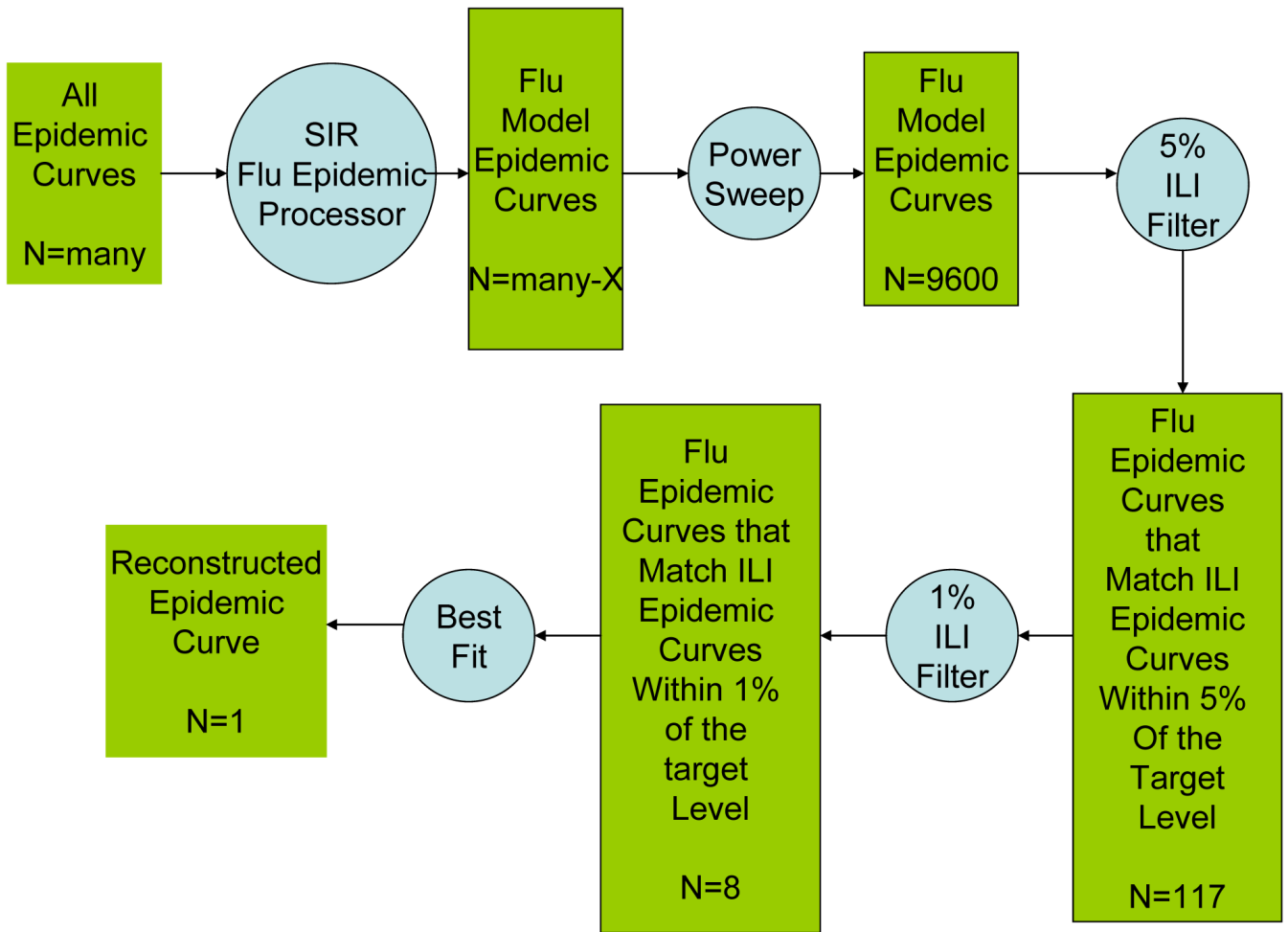


Fig. 2.
The Epidemic Curve Reconstruction Process

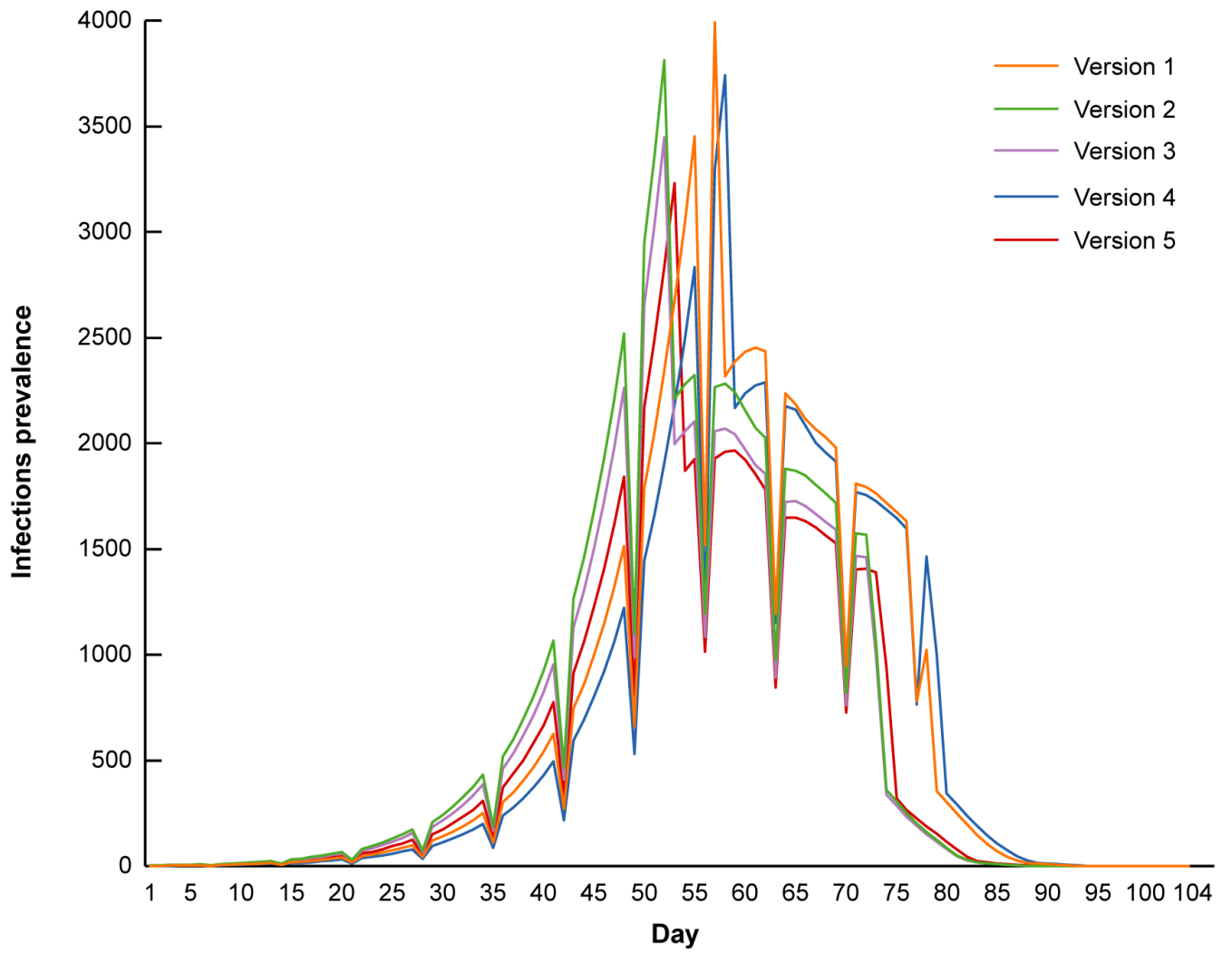


Fig. 3.
Number of Infections Recorded.

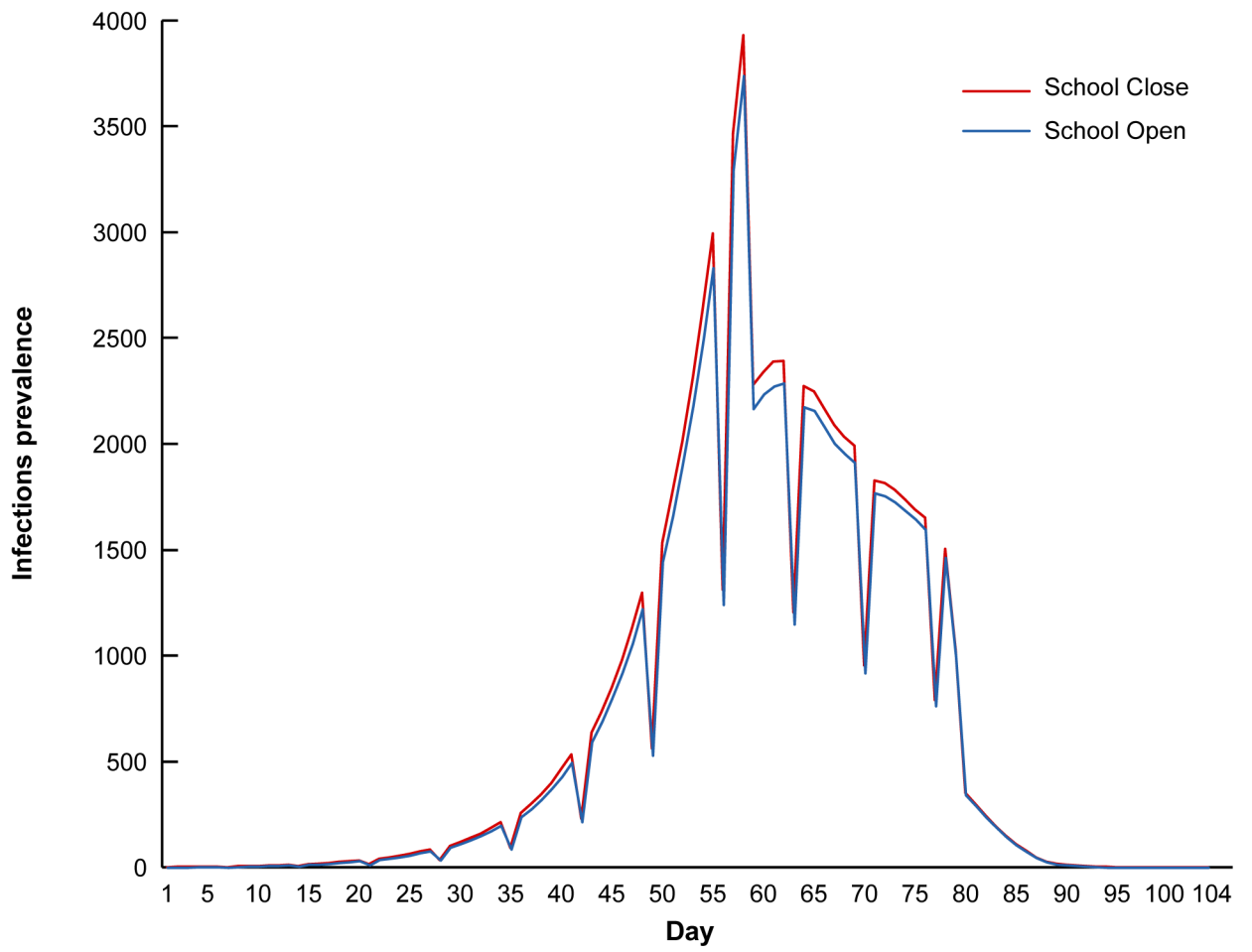


Figure 4.
Best Fitting Epidemics With and Without School Closure

Table 1

Counties in the Research Triangle Park (RTP) Metro Region in 2000.

Name	Population (2000)	Households	Area (mi ²)
Chatham	49,329	19,741	709
Durham	223,314	89,015	298
Franklin	47,260	17,853	495
Johnston	121,965	46,595	796
Orange	118,217	45,863	401
Wake	627,846	242,040	857
Total	1,187,931	461,107	3556

Table 2

Emergency Room Influenza-Like-Illness (ILI) Statistics

Measure	Weekly Raw Counts
Daily average over period before 10/25/2003	52.5
Average cases per day	111.4
Standard Deviation	120.90.1
Skewness	1.284
Kurtosis	3.661

* This table summarizes the RTP case curve displayed in Fig. 1.

Table 3

Distribution of Estimated Influence Prevalence.

Range Infections	Frequency	Cumulative Frequency	Percent
40 to 45	2	2	0.0171
45 to 50	5	7	0.0427
50 to 55	6	13	0.0513
55 to 60	10	23	0.0855
60 to 65	17	40	0.1453
65 to 70	10	50	0.0855
70 to 75	19	69	0.1624
75 to 80	6	75	0.0513
80 to 85	8	83	0.0684
85 to 90	10	93	0.0855
90 to 95	10	103	0.0855
95 to 100	4	107	0.0342
100 to 105	6	113	0.0513
110 to 115	3	116	0.0256
115 to 120	1	117	0.0085

Mean = 79,381; Standard Deviation = 181,016; Minimum = 42,358; Maximum = 121,138

Table 4

Distribution of Estimated Influence Prevalence - Schools Closed

Range Infections	Frequency	Cumulative Frequency	Percent
35 to 40	12	12	0.0081
40 to 45	2	3	0.0163
45 to 50	3	6	0.0244
50 to 55	6	12	0.0488
55 to 60	12	24	0.0976
60 to 65	9	33	0.0732
65 to 70	15	48	0.1220
70 to 75	7	55	0.0569
75 to 80	13	68	0.1057
80 to 85	8	76	0.0650
85 to 90	10	86	0.0813
90 to 95	7	93	0.0569
95 to 100	5	98	0.0407
100 to 105	7	105	0.0569
105 to 110	6	111	0.0488
110 to 115	6	119	0.0650
115 to 120	1	120	0.0081
120 to 125	1	121	0.0081
125 to 130	1	122	0.0081
130 to 135	1	123	0.0081

Mean = 84,302; Standard Deviation = 228,859; Minimum = 38,416; Maximum = 139,166