

P-MEDDS User Manual

Predictive Science Inc.

December 25, 2014

Contents

1	Introduction	1
2	Examples	2
3	Influenza Modeling	2
3.1	ILI Methods	2
3.2	ILI Data	2
3.3	ILI Results	4
4	SARS Modeling	4
4.1	SARS Methods	4
4.2	SARS Data	4
4.3	SARS Results	4

1 Introduction

P-MEDDS provides a robust set of tools for modeling the spread of infectious diseases. Using a deterministic compartmental S-I-R model and a robust Markov-Chain-Monte-Carlo (MCMC) fitting procedure P-MEDDS can quickly characterize an incidence profile in real time providing estimates for the:

- individual level epidemic severity (as described by the proportion p_C of infections that result in clinical cases)
- and the epidemic transmissibility (measured by the basic reproduction number R_0)

Publicly available weekly influenza-like-illness (ILI) data is included in the package (from the CDC, Google Flu Trends (GFT), and the World Health Organization (WHO)) along with weekly averaged specific humidity and school opening/closing schedule. The last two sets of data are needed for three of the five different models for the time dependence of the reproduction number $R(t)$ that the user can choose from.

Upon completion of the MCMC fitting procedure the P-MEDDS package analyses the results and produces an extensive set of publication quality plots (in PDF and PNG formats) and tables (in csv format). The complete history of the run and the fitting procedure is saved as an ‘.RData’ file which the user can later load and further analyze.

P-MEDDS also provides a computationally efficient version of the Wallinga-Teunis (W-T) likelihood-based procedure for estimating the daily effective reproduction number using an observed epidemic curve [insert ref.]. This algorithm can be applied to the 2003 SARS epidemic data which is included in the P-MEDDS package. As in the case of influenza modeling, P-MEDDS will produce a set of publication quality figures and tables when the estimation procedure is completed.

2 Examples

P-MEDDS includes an examples directory with three sample R drivers that demonstrate how to use the package:

- `example.driver.R`: This script can be used to model any of the ILI data in the P-MEDDS data base: Military, CDC or GFT. The script explains every parameter that the package requires and shows the default values for each one.
- `example.interactive.R`: This is an interactive version of the `example.driver.R`: it will prompt the user to select each of the parameters for a P-MEDDS run. It will explain the options and use their defaults if the user provides an incorrect value.
- `example.wt.R`: This script demonstrates how the P-MEDDS package can be used to model the 2003 SARS data using the Wallinga-Teunis procedure.

For more information on each of these R scripts see `README.example.md` file in the `examples` directory. Inside the `examples` directory there is a sub-directory `tests.output` with sample output files for the different models supported by P-MEDDS. After reading this manual we suggest that the User familiarize him/her-self with the package by using these scripts and modifying them as needed.

3 Influenza Modeling

put more text here

3.1 ILI Methods

3.2 ILI Data

We obtained data from the Armed Forces Health Surveillance Center (AFHSC) consisting of outpatient visits to permanent military treatment facilities (MTFs) by active duty military personnel for a range of ICD (international classification of diseases)-9 codes associated with respiratory-related illnesses between January 1, 2009 and April 30, 2011. For each record, the data contained: a unique study identifier for the individual; ICD-9 codes associated with that visit; and the zip code (5 digits) of the clinic location. We used the zip code of the reporting clinic as a proxy with which to define a military installation: we do not explicitly represent military installations or bases, rather, we assume that the case reports from the same zip code are from the same military installation. Each record (an anonymized Study ID) was assigned as either “ILI-large” ($n = 1,336,471$) or “ILI-small” ($n = 27,582$) using a set of classifications based on ICD-9 codes [ref here]. The definition of ILI-large was broader and included non-specific diagnosis such as ‘viral infection’ and ‘acute nasopharyngitis’. The definition of ILI-small was more constrained and included: ‘influenza w/other respiratory manifestations’ ($n = 25,293$), ‘influenza with manifestation not elsewhere classified (NEC)’ ($n = 1,006$), ‘infectious upper respiratory, multiple sites, acute NEC’ ($n = 897$), and ‘influenza with pneumonia’ ($n = 404$). We further trimmed the data temporally to cover the period from April 1, 2009 through June, 1, 2010 which is the 2009 H1N1 pandemic period. We ranked the military installations by size according to the total number of ILI-small cases they reported.

In the P-MEDDS data base we provide the data for the top-50 largest profiles, 47 of which, were located within the USA. Of the remaining three, one was located in Landstuhl, Germany, and two were located in Japan (Misawa and Yokosuka).

For each of these 50 military installations we provide our estimate of the total population (N_{total}), the “denominator data”. Our method for estimating these sizes relied on the use of the total number of visits to a clinic for all causes as a proxy for the total number of active duty

personnel at that location [ref]. The coefficient of proportionality Ω was estimated by using a subset of the installations for which reasonably reliable estimates for the total population have been published.

We obtained civilian data through a variety of means. County- and State-level data were generally acquired directly from the appropriate public health services department.

The CDC Influenza-like Illness Surveillance Network (ILINet) Human and Health Services (HHS) region and national data were downloaded from the CDC hosted web application Flu-View <http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html> for all available dates (Week 40, October 10th, 1997 through Week 45, November 9th, 2014).

The CDC Patient ILINet consists of more than 2,900 outpatient healthcare providers in all 50 states, Puerto Rico, the District of Columbia and the U.S. Virgin Islands reporting more than 30 million patient visits each year. Each week, approximately 2,000 outpatient healthcare providers around the country report data to CDC on the total number of patients seen and the number of those patients with influenza-like illness (ILI) by age group (0-4 years, 5-24 years, 25-49 years, 50-64 years, and ≥ 65 years). For this system, ILI is defined as fever (temperature of 100F or greater) and a cough and/or a sore throat without a KNOWN cause other than influenza. Sites with electronic health records use an equivalent definition as determined by public health authorities. The CDC Influenza surveillance data collection is based on a reporting week that starts on Sunday and ends on Saturday of each week. With the exception of the 2009 influenza pandemic, each flu season starts on Week 27 and ends on Week 26 of the following year. For 2009, the flu season started Week 13, 2009 (3-30-2009) and ended Week 26, 2010 (6-28-2010). For more information on ILINet see: <http://www.cdc.gov/flu/weekly/overview.htm>.

Because P-MEDDS requires an absolute number of incidences per week, the CDC ILINet data must be converted from percent ILI cases per patient to ILI cases. We estimate the absolute number of weekly ILI cases by dividing the weighted number of ILI cases in the CDC data ("X.WEIGHTED.ILI" column) by 100 and multiplying it by the total weekly number of patients ("TOTAL.PATIENTS" column).

Google Flu Trends data (at the national, state and city level) was obtained directly from the GFT web site: <http://www.google.org/flutrends/us/data.txt>. GFT attempts to quantify ILI cases through a proprietary model using key terms from search queries (originally discussed in Ginsberg et al., 2009; see also Cook, 2011 and Copeland, 2013). GFT ILI estimates were calculated from Week 39, September 28th, 2003 to Week 45, November 9th, 2014. The GFT data was aggregated to the same ten HHS regions as for the CDC data and the flu seasons were also defined using the same dates. For more information on GFT, see: <http://www.google.org/flutrends/about/how.html>.

Denominator civilian data was obtained from the US 2010 census data center: <http://www.census.gov/2010census/data/>. The estimated inter-censal state and national population sizes between 2000 and 2010 were obtained from <http://www.census.gov/popest/data/intercensal/state/state2010.html> and <http://www.census.gov/popest/data/intercensal/state/tables/ST-EST00INT-01.csv>.

The projected state and national population sizes from 2010-2013 were obtained from <http://www.census.gov/popest/data/datasets.html> and http://www.census.gov/popest/data/national/totals/2013/files/NST_EST2013_ALLDATA.csv.

Each region population size was calculated by summing the population sizes of every state in the region.

3.3 ILI Results

4 SARS Modeling

4.1 SARS Methods

4.2 SARS Data

4.3 SARS Results