Daily Mortality Surveillance Systems: Integration of Data Collection, Data Analysis and Reporting System Components

Tiago Chaves

Under supervision of
Mário Silva, INESC-ID/Instituto Superior Técnico, University of Lisbon
Baltazar Nunes, DEP/National Health Institute Dr. Ricardo Jorge

July 1, 2015

Abstract

The daily mortality surveillance system (VDM) monitors the observed all-cause mortality in Portugal. VDM, developed by the National Institute of Health Dr. Ricardo Jorge (INSA), monitors the effects that disease outbreaks, extreme weather conditions or other health-related events have on the Portuguese mortality, by detecting and estimating the excess deaths associated with such events. The workflow of the system includes data collection, analysis and report generation activities. The resulting statistics and reports are disseminated to multiple authorities, at the national and European level. This paper presents the development of a new version of VDM, with new added features that substantially optimize it and enable more effective data analysis. Additions include the calculation of the expected mortality baselines for each data stratification, the implementation of a nowcasting approach, to predict upcoming death notifications, and an automatic alert system for excess deaths. In the remodeled VDM system, public health practitioners now have a more user-friendly interface for generation and visualization of data for each stratification against the respective baseline, within a user defined time period. Report generation is now performed automatically, producing documents ready to be sent to the authorities. The new version of the VDM software was user-validated and is now in production at INSA.

Keywords: Mortality Surveillance, Surveillance System, Public Health Informatics

1 Introduction

Throughout history mortality data has been used to measure the burden of diseases and, ultimately, to assess the potential impacts of future epidemics. Being one of the oldest disease registries, mortality surveillance plays an important role in public health, consensually pointed out as one of its most important functions [1]. It provides data to support the planning of public health programmes and regional and national health policies. This enables the implementation and evaluation of public health actions, aiming to mitigate the impacts of future health-related events leading to deaths. Therefore, it is important to analyse these data in order to provide adequate and timely epidemiologic information to decision makers through surveillance systems.

To serve these purposes, the National Health Institute Dr. Ricardo Jorge (INSA) developed the Daily Mortality Surveillance System (VDM), which tracks all-cause mortality [2].

The system has been functional for years. However, recent information requests from public entities couldn’t be provided. For instance, the Regional Administrations of Health (ARS) were interested in obtaining accurate information about their areas of influence and the system was only assigning Eurostat NUTS II regions. To public health practitioners that use VDM on a daily basis, the workflow of the system was not very user-friendly, requiring many operator-supervised data exchanges from one software module to another to handle different tasks. This made the system patchy and not fully automated,
especially regarding the generation of reports that required multiple interventions. There was a web application (ICAROweb) implemented to ease the data exchange between users and the VDM database, but it was not harnessed to the fullest, given that it only served as an intermediary between the database records and the analysis/reporting tools.

With the wider availability of new software frameworks, it should be possible to group these multiple components into one, improving the system by also providing new and better features. Therefore, remodelling the VDM front-end so that it automatically integrates all the components would bring benefits both to the user and to the system efficiency point-of-view. The data analysis could be complemented with new additional features, such as baselines for all daily and weekly stratifications, the integration of a daily nowcasting algorithm to correct the number of deaths that had yet to be notified, and a new automatic alert system designed for all daily and weekly stratifications. Such features should result on an improvement of the system by adding robustness to the information, now more detailed, reducing the resources needed and providing timely and faster information.

The main goal of this project was to develop a new integrated version of all the VDM system components: data collection, data analysis and reporting. This included the following:

1. Data ingestion through the front-end application;
2. Data visualization of all stratifications, regions by ARS and user-defined age groups;
3. Generation of baselines for all stratifications and display of dynamic charts;
4. Integration of a nowcasting method;
5. Implementation of an automatic alert system;
6. Automatic generation of reports.

At the end of this project INSA had the new VDM software deployed in its intra-network, replacing the previous. All system functions are automatically operated through the application environment, speeding up the daily monitoring process carried out by INSA public health experts. I also revised and updated the VDM database scheme so that it can handle all the new features that were implemented.

Data ingestion into the database is now performed through the VDM front-end. In this new application of the VDM system (VDMweb) data analysis is carried out by an R software environment, which is an open-source software platform for statistical data analysis, implemented within the system. Scripts designed in this environment were created to implement the generation of baselines, nowcasting values and alerts. As INSA’s public health experts are used to the R functional language and the scripts are independent from the VDM environment, they can be developed in an R environment, updated and loaded to the VDM system by the experts with direct outcomes in the front-end data display. Regarding reports, the feature was successfully implemented and these are now automatically generated.

2 Mortality Surveillance

Surveillance systems consist on the ongoing systematic, continuous collection, analysis and interpretation of health-related data needed to plan, implement and evaluate the public health practice, by disseminating data to the ones responsible for making the decisions that affect the prevention and control of both diseases and injury [3]. By doing so, surveillance is an important tool to either provide information that the decision makers can rely on to decide what interventions need to be done or to get feedback information about the effects of those intervention strategies [4]. Besides that, it can also serve as a warning system to detect outbreaks or other hazard events, which can lead to immediate public health actions, and support multiple research hypotheses, by establishing possible correlations between some abnormal observed indicators [5]. Furthermore, it can also be used to prioritize population age groups or geographical areas to target interventions, if its data stratification is done at that level.

Regarding mortality in specific, since it is the final event of the most serious outcome of a disease, it is a basic indicator of health and the epidemiological value that its monitoring has is crucial when it comes to providing health information to support effective health planning and
action. Distribution of mortality through time is regarded as seasonal, mostly due to the effects of influenza epidemics (reaching peaks more often in the winter months) among the elderly and other vulnerable groups of the population. Monitoring it during the winter months provides timely information to decision makers on whether and when to increase the use of antiviral and vaccines [6]. Besides influenza, it can also be used to detect excess peaks of mortality related to environmental conditions, for instance heat waves, cold snaps and the release of biological agents, like several mortality systems across Portugal [7], France [8], UK [9] and the United States [10] have proved its value. The timeliness of the mortality monitoring is vital to guide public health measures, for instance focusing interventions on vulnerable groups. With the purpose of pooling mortality data of the excess number of deaths related to influenza and other health-related events at an European level, it was implemented in 2008, a three-year project denominated European Monitoring of Excess Mortality for Public Health Action (EuroMOMO) [149], in which Portugal participates with VDM.

3 VDM - Daily Mortality Surveillance System

The VDM system was initially designed in 2003 as a pilot, to evaluate the effects of the heatwave at the time in Portugal. It has been active since its permanent implementation in 2004 and with a new version implemented in 2007 [2].

The system receives information on a daily basis, with a national geographical coverage, and disseminates a daily bulletin to multiple entities, the State Secretary of the Ministry of Health, the General Directorate of Health (DGS), the National Authority of Civil Protection (ANPC), the ARSs and the Portuguese Institute of Sea and Atmosphere (IPMA). Besides, it also provides data to EuroMOMO.

VDM is used to detect and estimate the magnitude of deaths that are caused by different health-related events, such as influenza epidemics and extreme weather conditions. The system is therefore an important tool for public health action. VDM has three major system sub-systems:

1. Data collection, consisting of the gathering and posterior data ingestion in the VDM database of the mortality information collected in mortality records files (encoded in XML). These files contain the death records registered in the previous day and are sent by the Institute of Financial Management and Equipment of Justice (IGFEJ) on a daily basis.

2. Data analysis, performed through daily and weekly plots for different stratifications: age groups, regions and gender. These analysis are based on the information extracted from the front-end application of the VDM, and are carried out to determine excess mortality values.

3. Reporting, related to the production and dissemination of reports with monitoring outcomes.

The VDM system deployed in INSA’s intranet relied on a database within a Microsoft SQL Server 2008 R2 and it had a front-end application, ICAROweb, since early 2013. ICAROweb is an ASP.NET web application developed with .NET framework 2.0 and coded in C#. Several Microsoft Office tools were used to complete the daily workflow, involving charts visualization, analysis and reporting. The initial daily workflow of INSA professionals is represented in Figure 1.

3.1 Limitations

The VDM system had been functional and successfully used at INSA for years. However, from the user point-of-view, its workflow was neither efficient nor very user-friendly. For instance, integrating the mortality records file in the VDM database was not intuitive, since the public health practitioner had to save the file in a shared folder and wait for a data ingestion tool to be executed. Additionally, the VDM system was stratifying regions by NUTS only, despite new requests by the ARSs, and the generation of baselines was limited to the total weekly data. Within the stratification by region, recently there has been a recent introduction of alphanumeric codes for the geographical codes of region of death, which in the previous VDM version were not being assigned to any region. The system analysed data in the pre-defined 3 stratifi-
In VDM, there was not implemented an automatic parametrized stratification (e.g., region by age group by gender). Finally, regarding the daily generation of reports, it was not automatic and involved multiple software components.

4 New VDM System and Implementation

4.1 Conceptualization of the System

New Features

The approach for implementing a new VDM version was the reformulation of the ICAROweb front-end into an optimised and more robust application capable of handling data ingestion, analysis, visualization and reporting. This new front-end application was named VDMweb. Additionally, VDMweb now offers new analysis features through R, an open-source software environment, for statistical data analysis. This R environment can interact permanently with the web application. The objective was to centralize the system workflow on the VDMweb application (see Figure 2).

- Data ingestion of the daily mortality records file through VDMweb;
- Automatic generation and integration in database of daily and weekly baselines, calculated by R scripts, for all stratifications;
- Daily and weekly data stratification tables:
  - Pre-determined (0-14, 15-24, 25-44, 45-64, 65-74, 75-84 and ≥85 years) and a dynamic user-defined age group;
  - Regions (Norte, Centro, Lisboa, Alentejo, Algarve, Madeira and Açores) by ARS and NUTS II, Portugal and Portugal Mainland;
  - Gender (Male and Female);
  - User-customized grouping of strata (e.g., regions plus age groups) in the tables displayed;
  - User-parametrized stratification (e.g., region by gender by age group);
- Dynamic mortality variation plots within a user-defined time period for daily and weekly data and its baselines. Additionally, a yearly mortality comparison plot, up to 5 user-defined years, is also displayed. The plots by stratification display data of each stratification;
- Online display of numerical data of the last 15 days and comparison of these values to one of the following reference values selected by user:

Implementing the new VDM version involved redesigning the front-end, to provide users with new added features, and modifying the database to store and provide all the new information and analysis data results. The main new features implemented in VDM are listed below:
– Average of the last 4 years for the respective day;
– Mortality baseline;
– 95% upper confidence limit of the baseline;

• Automatic generation of daily bulletin with new plots including more stratifications:
  – Age groups;
  – Regions are now user-selected, whether ARS or NUTS II, and include Madeira and Açores;
  – Gender;

• User-defined time periods and years to display in the bulletin charts. Baselines are also selected by the user to be displayed in the respective charts or not;
• Automatic generation of mortality alerts, using R scripts, for each daily and weekly data stratification. Additionally, a comparison map between all strata of each stratification was also implemented;
• Integration of a nowcasting method, implemented in R. Plot display of the method results with observed, nowcasting and baseline values, and display of its alerts;

Figure 3: VDM system deployment architecture.

4.2 System Deployment and Information Architectures

Since the purpose of VDMweb was to replace ICAROweb, its system architecture, consisting of an application server hosting the front-end application and a database cluster with SQL Server 2008 R2, was unmodified (see Figure 3). However, the application server now hosts an R environment in which R scripts can be loaded for data analysis. Additionally, the application server also hosts the interactive tool for daily ingestion of records files into the database. The access to the database server is secured by SQL connections, consisting of trusted connections (Windows authentication) from the ASP.NET user session in the application server. The client connects to the application server using the TCP/IP network protocol.

To implement the new features, new modules were created and implemented in the VDMweb, resulting in the information architecture represented in Figure 4.

Figure 4: VDMweb information architecture.

4.3 VDMweb Modules

Just like in the previous VDM front-end, each web page of each module communicates with the SQL Server in order to get data and dynamically display it as the user requests. Some of the pages, depending on the module, also involve interactions with R scripts to analyse data. Each one of the modules is described next:

I. Home is where data ingestion is performed and a VDM description, as well as the web application, is presented. In this module, the public health practitioner uploads the daily mortality records file, whose records are ingested into the VDM database;
II. *Records* displays details about the successfully integrated mortality records files;

III. *Daily Mortality Surveillance* consists of multiple daily data display in tables. These are all downloadable as CSV files. The module is divided by stratification, with all the new features implemented for each, such as the regions by ARS, the new age groups or the new parametrized stratification.

IV. *Weekly Mortality Surveillance* groups multiple weekly data display, analogous to the previous module with one more page consisting in the display of data to be sent to EuroMOMO;

V. *Reporting* groups the data visualization and the reporting components. The latter consists on the automatic generation of the daily reports. Furthermore, the addition into the system of the baselines for every stratification enables their display in the respective stratum plot, overlaying the mortality time series. In each of these plots the user can also select the time period of data to be plotted.

VI. *Alerts* consists of the automatic alert system implemented for daily and weekly aggregations. It displays tables with the daily or weekly observed values and its alert status, represented by the colours red (alert) or green (no alert). A comparison map of alerts is also provided in this module, to compare the alerts of strata within each stratification, being useful to determine which regions have been more affected by a certain event. The definition of the alert status (1 in case of alert, 0 in the absence of it) is implemented in the R environment within the application server. An alert is identified when 2 consecutive counts (days or weeks) record observed mortality values above the 95% upper confidence limit of the mortality baseline, or when 1 count has an observed value above the 99% limit. Then, the alert status is only cleared when 2 consecutive counts register observed values below the 95% upper confidence limit.

VII. *Nowcasting* refers to the nowcasting method integrated in VDMweb, displaying its daily plot and alerts. The nowcasting method consists of an algorithm developed in the R environment covered in [11], and its script was provided by INSA with the purpose of integrating it into the new VDM version as a test module. It updates the mortality values of the past 15 days by correcting the observed number of deaths taking into account the delay in their notification, hence it is basically a prediction of the number of deaths of the last 15 days. By doing so, they can be used in the detection of alerts corresponding to excess deaths, for which an alert table is provided. This enables an earlier observation of the mortality variation, especially in the most recent days where the observed values time series tends to decay abruptly due to the low number of deaths that were registered and integrated in the VDM database.

VIII. *Data Management* consists of pages where the user can manage the generation of the baselines and update the table of the correspondences between geographical codes and ARS/NUTS regions. Baselines are generated using the R environment, based on the algorithm being used in the previous VDM version for the total weekly data, covered in [12]. The algorithm consists of cyclic regression models and data consisting of the observed mortality values by day or week, excluding values that were registered during mortality events (influenza, heat-waves, etc). For the implementation of the VDMweb, the algorithm was adapted and expanded in order to generate baseline values for all stratifications.

4.4 Development Tools and Base Software

VDMweb was developed using MS Visual Studio 2013 Community version with the .NET framework version 4.5, implemented in C#. To operate data analysis, besides its Microsoft SQL Server 2008 R2 to manage the VDM database, VDMweb benefits from the R programming language. The R setup comes with a tool "RScript.exe" that is used to execute R scripts without the need of opening an R terminal and can be used as a shell script. Using this tool en-
ables the automation of running R scripts from the front-end application.

4.5 Extended Database Model

Most of the tables that were added to the VDM database, with the implementation of VDMweb, had the purpose of storing the multiple baseline values, the alerts generated for each stratification, the nowcasting values and the defined periods of mortality events used in the generation of baselines. One new table named Regions, however, it’s of important notice since it’s the basis for the new method of stratifying data by region - ARS and NUTS. It is a robust and complete table that groups every correspondence between geographical codes and the ARS and NUTS (I, II and III) regions, among other information. This table also covers the recently introduced alphanumeric geographical codes, that were not being assigned to any region in the previous VDM version.

5 Data Visualization Showcase

VDMweb provides new and more powerful features than the previous, ICAROweb. Overall, the new VDM was praised by INSA professionals, due to its simpleness and automatic workflow. Data ingestion is more intuitive, data analysis have new useful features implemented and the automatic reporting dispenses a lot of resources previously used in the creation of the reports. Regarding data visualization, it is now performed a lot faster since it is done directly in VDMweb, with added features. Next, I provide an example of the VDMweb potential by showcasing the National influenza epidemic that recently hit Portugal between January and February of 2015.

The annual comparison chart of the number of registered deaths by day of death, in Figure 5, confirms that in January and February of 2015 there was an increase in mortality, when compared to the last 4 years.

This increase in the observed mortality is also confirmed by the weekly plot of the number of deaths displayed in VDMweb. This plot, represented in Figure 6, enables the observation of the mortality evolution through time by week, important to identify patterns or abnormal events.

The excess mortality related to this event is also identified using the new automatic alert system in VDMweb. By doing so, the periods of excess are easily identified. A table displaying absolute values for each stratification and its alert colour is displayed. To identify which age group, region or sex stratum was more affected, the comparison map of alerts for each stratification can be used. Figure 7 presents an example of one of the four comparison tables of the various ARSs. Observing it, it is clear that the most affected regions were Norte, Centro and LVT, although Alentejo only had two weeks less with alerts. This same exercise can be performed for
Figure 6: Plot of weekly number of deaths (light blue) in the weeks between 2013-06-29 and 2015-05-25 and its upper baselines: the baseline (dark blue), the 95% (orange) and 99% (red) upper confidence limits.

Figure 7: Screenshot of the table of alerts for weekly data by ARS regions between 2014-12-15 and 2015-03-15. Fields in red represent alerts.

the other stratifications, in the case of gender the alerts were equally assigned, as expected since influenza does not target a specific gender, and the most affected age groups were the elderly, the 75-84 and the ≥85. However, this identifies the excesses qualitatively and not quantitatively. To do so, the table of alerts for each data stratification is exported and the difference between the observed number of deaths and the baseline (present in each alert table) is calculated, in periods where the alert is equal to 1 (red).

An alternative approach, although less accurate, is to observe the plots for each stratification by age groups, regions and gender stratum. As these plots are displayed in VDMweb with their baselines, the deviations between the observed values’ series and the corresponding baselines enable users to assess the gravity of the excesses.

All the plots and alerts can also picture daily aggregated data, since they are also provided in the application. However, in the case of influenza epidemics the daily values present a lot more variation than the weekly ones and therefore a weekly data visualization is more readable.

In the above example, all the information is related to a past time period. However, as all of these data are constantly updated, both the plots and alerts are always related to current events that might be happening.

6 Conclusions and Future Work

The new VDM version was successfully designed, implemented and is now deployed at INSA. The implementation covered the integration of the data ingestion, analysis and reporting components, and was user-validated and tested. It is now in production at INSA, and is used in
a daily basis by its professionals.

The new analysis features implemented benefit from the VDMweb versatility, along with the storing and query processing of the SQL Server database and the statistical computing power of the R environment. The integration of these three environments is key to the system, with the integration of the R environment providing the most innovative feature now available in the new system. The possibility of manipulating and changing the R scripts outside of the application server opens a new door to researchers. Since VDMweb operation is now completely automated, the opened possibility of testing new algorithms and methods for the generation of baselines, nowcasting or alerts is rather useful. The most adequate method can be chosen by updating the R scripts. The new VDM application allows the direct observation of the resulting changes that it induces in each of the respective modules in the VDMweb. This gives to public health experts a more manageable automatic system, allowing it to be always updated with better and more accurate analysis algorithms. By doing so, this VDM version provides precise and valuable data relevant for an extended lifetime.

The automatic reporting reduces the time and computing resources used in the generation of the daily reports. Besides, with the introduction of the baselines, these reports were also complemented with new information. This was appreciated by INSA public health practitioners, since it dispenses the exhaustive routine of manually shipping data between software modules. Regarding the reports, as new information was added, there are also improvements to the information sent to decision makers. They can now rely on more information to support interventions or implement new health policies.

To sum up, the new VDM represents an improvement to the previous, with new added and user-customized features. Its workflow is more efficient and autonomous, resulting into faster and more detailed information to be used on the detection and estimation of excess deaths.

6.1 Future Work

One improvement to VDMweb would be the introduction of information related to the incidence rates of influenza like illness (ILI). As ILI are a major cause of mortality, it would be interesting to automatically overlay an incidence rate series in the mortality charts displayed, with 2 Y axis. Similarly to ILI, the ICARO alerts could also be used in this series overlay instead of ILI, providing that option to users. This allows the comparison between the distribution of mortality through time and the ILI incidence rates or ICARO-index, to determine if they are related or to refute their relationship. By doing so, these two events of mortality peaks would be rapidly ruled out by observing the plots. This would let researchers focus on other events.

Another improvement would be the development of a framework to test, and ultimately adopt, different algorithms and methods to generate outbreak alerts. In this case, the outbreaks consist of periods associated with excess deaths. Statistical algorithms such as the ones described by Farrington [13] and Stroup [14] can be tested and adopted if they present more timely results.

A third improvement would be the development of a framework to test, and ultimately adopt, different algorithms and methods to generate outbreak alerts. In this case, the outbreaks consist of periods associated with excess deaths. Statistical algorithms such as the ones described by Farrington [13] and Stroup [14] can be tested and adopted if they present more timely results.

A third improvement would be the development of a framework to test, and ultimately adopt, different algorithms and methods to generate outbreak alerts. In this case, the outbreaks consist of periods associated with excess deaths. Statistical algorithms such as the ones described by Farrington [13] and Stroup [14] can be tested and adopted if they present more timely results.

Lastly, as the parametrized queries are the result of parameters selected within each stratification, their baselines are not generated. The baselines are calculated only for the strata already determined in the system. It would be interesting if they could be calculated on-the-fly for the variation displayed in the parametrized query chart. This would provide more customization to analyse new different strata that are not covered by the pre-determined stratifications.

References


[6] euroMOMO, “Pooled analyses of all-cause mortality indicates low excess mortality in Europe in the winter of 2013/14, in particular amongst the elderly.” euroMOMO publication available on website.


