

Power, Potential, and Pitfalls of Surveillance using Clinical Ancillary Services Data

Beth T. Poitras^{1,2}, Rebecca S. Payne^{3,2}, Nicholas D. Seliga^{4,2}

¹Defense Health Agency, Falls Church, Virginia, United States, ²Navy and Marine Corps Public Health Center, Portsmouth, Virginia, United States, ³Battelle Memorial Institute, Hampton, Virginia, United States, ⁴Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, United States

Objective

Discuss the power of utilizing DOD clinical ancillary services data for infectious disease surveillance, the steps used to mitigate pitfalls which may occur during the surveillance process, and the potential of adapting this data for surveillance of emerging infectious diseases.

Introduction

Military service members and their families work and live around the world where both endemic and emerging infectious diseases are common. Timely infectious disease surveillance helps to inform medical and policy decisions which ensure mission readiness and beneficiary health. The EpiData Center (EDC) at the Navy and Marine Corps Public Health Center has performed public health surveillance, including routine infectious disease monitoring among service members, their families, and others eligible for military medical benefits for the Department of the Navy (DON) and Department of Defense (DOD) since 2005. The EDC stores and maintains 15 databases totaling over 20 terabytes of health and administrative data. These include administrative data from outpatient encounters and inpatient admissions, Health Level-7 (HL7) formatted ancillary services data, and medical event reports. These data provide the potential for robust surveillance methodologies to monitor diseases of interest and identify trends and outbreaks. The primary intent and design of these data sources is not for disease surveillance, but rather for administrative and billing purposes. However, due to the availability of this data, it is routinely used by academic organizations, private industry, health systems, and government organizations to conduct health surveillance and research. Ancillary services data in particular can be very powerful for near-real time infectious disease surveillance in the DOD as the aggregated data is available within 1 to 2 days after processing. The EDC has demonstrated the value of using laboratory data for surveillance through outbreak detection and longitudinal health trends for specific diseases among select populations. The fact that this data is not designed for surveillance does present several pitfalls in regards to analysis, from issues ranging from free text interpretation to changing testing practices. These pitfalls can be mitigated through standardized processes and detailed quality assurance testing. The EDC has harnessed the power of available administrative health data to improve health outcomes and influence policy among military beneficiaries.

Methods

The EDC has established and validated methods for using and interpreting ancillary services data. Key steps involved in the process for infectious disease surveillance include:

- Reviewing diagnostic criteria;
- Defining relevant search terms and test types;-
- Consulting clinicians for technical input when needed;
- Developing algorithms using retrospective data;
- Developing quality checks;
- Automating the process to reduce daily workload;
- Documenting processes and methods.
- Variables essential to interpretation within ancillary services records are not standardized across the DOD.

Several pitfalls can occur during the surveillance process due to complexities related to free text, layout of the full results, and differences between laboratory practices. Typically, these pitfalls can be grouped into one of the following categories:

- **Data irregularities** that include unexpected abbreviations and numerous misspellings; this may result in misclassification or missed cases.



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- **Data changes** resulting from shifts in testing practices due to new or discontinued laboratory tests, or differing data entry methods.
- **Classification challenges** for diseases that require sequential testing or clinical compatibility information, which limits the ability to positively identify cases. However, records can be identified as ‘suspect cases’ (i.e., syphilis, Lyme disease, varicella, yellow fever and others).
- **Technical issues**, at the medical facility, server, or EDC level, often causes lapses in data, which results in a delay in case reporting.

Despite these pitfalls, their impact can be mitigated by routinely reviewing algorithms, employing data analytic techniques that account for likely misspellings and abbreviations, and incorporating data quality checks that flag unexpected or unclassifiable results. Outside of automated processes, human interaction is important; EDC analysts must remain astute and vigilant to investigate unusual or unexpected occurrences, shifts in the volume of cases or data.

Results

Due to the pitfalls outlined, the EDC has developed powerful and robust methods to circumvent the issues of using administrative health data for near real-time clinical ancillary services based disease surveillance. The methods developed to address the pitfalls of working with administrative health data have been used in the daily active surveillance of over fifty reportable infectious diseases, weekly surveillance of influenza, and monthly surveillance of malaria and tuberculosis. In addition to using these methods for routine surveillance, the EDC adapts this methodology for new reports for specific concerns. Further, the EDC continues to develop and adapt these methodologies to quickly address emerging infectious threats and the pitfalls associated with the data. Pharmacy transactions and administrative data from outpatient encounters and inpatient discharges supplement and enhance laboratory-based surveillance, particularly when only a diagnosis or presumptive treatment occurs (such as with influenza). While this method provides timely information, built in quality assurance checks and routine reviews of algorithms must occur to address changes in testing practices, the use of new tests, variation in laboratory technician entry of results, and to ensure data integrity.

Conclusions

The EDCs comprehensive surveillance provides the DON and DOD leadership and preventive medicine community with the ability to monitor and respond to ongoing and emerging infectious disease threats. While the primary purpose of administrative health data is not for health surveillance, the EDC has recognized the rich source of health information which may be extracted from this data. Processes have been developed to mitigate the pitfalls that may occur when administrative data is adapted for health surveillance. This data provides a real-time snapshot of the health of military beneficiaries and provides awareness of possible outbreaks, health trends, and geographic hotspots. Beyond routine surveillance this data has the potential to be used to rapidly create new methodologies to detect emerging infections which can be combined with other data sources, such as pharmacy transactions and medical encounters, to provide a more robust picture of cases by accounting for variance in clinical practice. This data often guides military health policy and procedures and is essential for a medically ready force.

