

Comparison of Air Passenger Travel Volume Data Sources for Biosurveillance

Diana Y. Wong*, Teresa Quitugua and Julie Waters

Department of Homeland Security (DHS), Office of Health Affairs (OHA), National Biosurveillance Integration Center (NBIC), Washington DC, DC, USA

Objective

To evaluate different government and commercial air travel route and volume data sources for utility in determining likely points of arrival and subsequent spread of communicable diseases originating from outbreaks outside the United States.

Introduction

The National Biosurveillance Integration Center (NBIC) has the responsibility to integrate, analyze, and share the nation's biosurveillance information provided from capabilities distributed across public and private sectors. The integration of information enables early warning and shared situational awareness of biological events to inform critical decisions directing response and recovery efforts.

Understanding travel trends and volumes is essential to managing public health and emerging infectious diseases as travelers move across today's increasingly globalized world. Travel routes, seasonal trends, and general passenger flow volume help determine where communicable diseases are more likely to arrive and spread within the US.

This type of data is useful in informing policy decisions, especially with respect to resource allocation and border screening procedures. Accuracy in the underlying data is important to build better predictive models

Methods

We conducted extensive research into currently available government and commercial air passenger data sources to evaluate data quality, utility, and accessibility for the purposes of biosurveillance. The goal was to identify the most complete and accurate data source of air passenger counts and flight schedules. The scope was limited to commercial passenger air flights.

The commercial data sources were limited to two industry leaders: the International Air Transport Association (IATA) and the Official Airline Guide (OAG). Evaluation of their products included interviews with data users in government, interviews with company representatives and some of their analytics staff, system demonstrations and examination of the results of sample data queries. The search for government sources of data included interviews of personnel within DHS Customs and Border Protection (CBP), the Federal Aviation Administration (FAA), the Centers for Disease Control and Prevention (CDC), and Department of Defense (DoD).

Identical queries were made of each data source. Examination of the differences between the volume values then led to the exploration of how each data source defined its designated variables and derived its calculations.

The purpose and use of air travel data to support decision-making was strongly considered during evaluation. Some examples of these concerns include resource allocation of staff, communication or messaging of health advisories, or planning for border screening processes and training staff. This plays a role in determining what kind of questions to ask of the data and whether a data source is flexible in altering its parameters to allow that query.

Results

The following are preliminary findings:

Commercial data sources differ in multiple ways:

- Not all commercial sources provide passenger volume data.
- OAG volume data is based on booked seats as gathered from Global Distribution Systems (GDS) and, with data delay, from Computer Reservation Systems (CRS). IATA volumes are based on the Business Settlement Process (BSP) surrounding ticket purchasing.

• Comparison of the results of OAG and IATA system query results is difficult as both draw from different data systems and use proprietary algorithms to calculate results.

• Most single commercial sources appear to require additional estimations to fill data gaps.

Government data sources:

• Government sources are usually limited to flight schedules or volumes of passengers crossing US airspace or borders, unless their mission space is in intelligence.

• DHS Customs and Border Protection (CBP) US entry data systems are considered the gold standard as they contain complete arriving passenger data.

• The results of government data systems queries differ from that of commercial sources due to differences in parameter definition.

Conclusions

The commercial data often used by academia and private performers to build disease spread models can differ significantly from government data not only due to the methodology to derive the passenger volumes (i.e. tracking passenger arrivals as opposed to ticket sales estimations based on flight routes), but also due to differing parameter definitions.

Keywords

air travel; border screening; commercial and government data; translocation

*Diana Y. Wong

E-mail: Diana.Wong@hq.dhs.gov

