bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Automated collection of pathogen-specific diagnostic data for real-time syndromic epidemiological studies

Lindsay Meyers[1*], Christine C. Ginocchio[1, 2, 3], Aimie N. Faucett[1], Frederick S. Nolte[4], Per H. Gesteland[5], Amy Leber[6], Diane Janowiak[7], Virginia Donovan[8], Jennifer Dien Bard[9,10], Silvia Spitzer[11], Kathleen A. Stellrecht[12], Hossein Salimnia[13], Rangaraj Selvarangan[14], Stefan Juretschko[15], Judy A. Daly[16], Jeremy C. Wallentine[17], Kristy Lindsey[18], Franklin Moore[18], Sharon L. Reed[19], Maria Aguero-Rosenfeld[20], Paul D. Fey[21], Gregory A. Storch[22], Steve J. Melnick[23], Christine C. Robinson[24], Jennifer F. Meredith[25], Camille V. Cook[1], Robert K. Nelson[1], Jay D. Jones[1], Samuel V. Scarpino[26], Benjamin M. Althouse[27, 28], Kirk M. Ririe[2], Bradley A. Malin[29], Mark A. Poritz[30]

[1] BioFire Diagnostics, Salt Lake City, UT 84103
[2] bioMérieux USA, Durham, NC 27712
[3] Hofstra Northwell School of Medicine, Hempstead, NY 11549
[4] Medical University of South Carolina, Charleston, SC 29425
[5] University of Utah, Salt Lake City, UT 84112
[6] Nationwide Children's Hospital, Columbus, OH 43205
[7] South Bend Medical Foundation, South Bend, IN 46601
[8] Winthrop University Hospital, Mineola, NY 11501
[9] Children's Hospital of Los Angeles, Los Angeles, CA 90027
[10] Keck School of Medicine, University of Southern California, Los Angeles, CA 90039
[11] Stony Brook University Medical Center, Stony Brook, NY 11794
[12] Albany Medical Center, Albany, NY 12208
[13] Detroit Medical Center, Detroit, MI 48201
[14] Children's Mercy Hospital, Kansas City, MO 64108
[15] Northwell Health, Lake Success, NY 11042
[16] Primary Children's Hospital, Salt Lake City, UT 84132
[17] Intermountain Medical Center, Murray, UT 84107
[18] University of Massachusetts Medical School-Baystate, Springfield, MA 01199
[19] UC San Diego Health, San Diego, CA 92093
[20] New York University Langone Health, New York, NY 10016
[21] University of Nebraska Medical Center, Omaha, NE 68198
[22] Washington University Medical Center, St. Louis, MO 63110
[23] Nicklaus Children's Hospital, Miami, FL 33155
[24] Children's Hospital Colorado, Aurora, CO 80045
[25] Greenville Health System, Greenville, SC 29605
[26] University of Vermont, Burlington, VT 05405
[27] University of Washington, Seattle, WA 98105
[28] New Mexico State University, Las Cruces, NM 87131
[29] Vanderbilt University, Nashville, TN 37232
[30] BioFire Defense, Salt Lake City, UT 84107

* Corresponding author. E-mail: lindsay.meyers@biofiredx.com

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Health-care and public health professionals rely on accurate, real-time monitoring of infectious diseases for outbreak preparedness and response. Early detection of outbreaks is improved by systems that are pathogen-specific. We describe a system, FilmArray® Trend, for rapid disease reporting that is syndrome-based but pathogen-specific. Results from a multiplex molecular diagnostic test are sent directly to a cloud database. www.syndromictrends.com presents these data in near real-time. Trend preserves patient privacy by removing or obfuscating patient identifiers. We summarize the respiratory pathogen results, for 20 organisms from 344,000 patient samples acquired as standard of care testing over the last four years from 20 clinical laboratories in the United States. The majority of pathogens show influenza-like seasonality, rhinovirus has fall and spring peaks and adenovirus and bacterial pathogens show constant detection over the year. Interestingly, the rate of pathogen co-detections, on average 7.7%, matches predictions based on the relative abundance of organisms present.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The availability of real-time surveillance data that can monitor the spread of infectious diseases benefits public health [1-3]. At present, tracking of respiratory or foodborne outbreaks relies on a variety of methods ranging from automated real-time electronic reporting to manual web entry of test results. Systems such as the Centers for Disease Control and Prevention's (CDC) FluView [4, 5], National Respiratory and Enteric Virus Surveillance Systems (NREVSS, [6]), National Electronic Disease Surveillance System (NEDSS [7, 8]), Global Emerging Infections Surveillance (GEIS) [9], and others, although web-based still require manual entry of data from laboratories, resulting in data that are often incomplete or not current.

Newer efforts at syndrome-based surveillance [10-12] include BioSense (extraction of symptomatic data from electronic health records [13]), Google Flu (tracking of Internet search queries [14] but recently discontinued [15]), and Flu Near You (voluntary reporting [16]). Additionally, numerous next generation, syndromic surveillance systems; e.g., pharmacy sales records [17, 18], Twitter conversations [19, 20], and Wikipedia hits [21, 22], have come online in the past five years. However, these systems cannot report the specific pathogen causing an increase in a particular set of symptoms. Finally, there are more localized efforts such as GermWatch in Utah [23, 24] and the Electronic Clinical Laboratory Reporting System (ECLRS) in New York [25, 26] that draw from hospital (HIS) and laboratory (LIS) information systems. This disparity in technologies and data collection methods results in incomplete surveillance.

Comprehensive and uniform diagnostic test data will aid in the identification of potential outbreaks. A combination of broad respiratory pathogen testing and an internal electronic surveillance system enabled the rapid dissemination of data across the largest health care system in New York, the North Shore-LIJ Health System (now Northwell Health) during the influenza A H1N1-2009 pandemic in the New York City area. Pathogen-specific molecular testing permitted rapid a) notification to state epidemiologists; b) tracking of the virus so that health care resources could be managed effectively; and c) evaluation of influenza diagnostics [27, 28]. Today, with the threat of emerging pathogens such as Middle East Respiratory Syndrome

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

coronavirus (MERS CoV), avian influenza, enterovirus D68, and Ebola virus, real-time surveillance programs are critical [1, 29, 30].

It is not always possible to accurately diagnose the causative agents of most infectious diseases from symptoms alone due to overlapping clinical presentation. Thus, to achieve maximal utility, infectious disease surveillance systems should move beyond syndrome-based reporting and be pathogen-specific and comprehensive, reporting on as many of the common pathogens for a particular syndrome as possible. Sensitive and specific automated molecular diagnostic systems that detect up to four different pathogens in a single sample have been available from in vitro diagnostic (IVD) manufacturers for some time [31-33]. However adoption of IVD platforms with broad multiplexing capability has become widespread only in the last few years. Commercially available systems that can detect most of the known etiological agents for respiratory, gastrointestinal and other multi-pathogen syndromes [34-36] include the BioFire (Salt Lake City, UT) FilmArray® System [37]; the GenMark (Carlsbad, CA) eSensor XT-8® [38] and ePlex® [39]; and the Luminex (Austin, TX) xTAG® [40], nxTag® [41] and Verigene® systems [42].

Multi-analyte diagnostic tests provide the raw data needed for real-time pathogen-specific surveillance but there remain a number of obstacles to sharing these results (reviewed in [43]). The obstacles largely center on information privacy and network security. A real-time surveillance system using diagnostic test results requires safeguards for protected health information (PHI). Medical records and devices have become attractive targets for cyber attackers in recent years [44], which has made hospitals and clinics reluctant to connect their Local Area Networks (LANs) to the Internet. Releasing patient test results requires the removal of PHI or authorization from the patient. Studies have shown that de-identification of patient data is not as simple as removing all specific identifiers because, in the age of big data, under the right circumstances it is possible to re-associate patients and their data using publicly available information [45-48].

We describe here the implementation of a real-time pathogen-specific surveillance system that overcomes the PHI concerns noted above. FilmArray Trend de-identifies, aggregates and exports

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

test results from FilmArray Instruments in use in US clinical laboratories. Although data from all commercially available FilmArray Panels (Methods and [49]) is exported to the Trend database, we focus here on the Respiratory Panel (RP), which can detect 17 viral (adenovirus [ADV], coronaviruses [CoV OC43, 229E, NL63, HKU-1], human metapneumovirus [hMPV], human rhinovirus/enterovirus [HRV/EV]; influenza A [Flu A, subtyping H1N1, 2009 H1N1, H3N2], influenza B [Flu B]; parainfluenza viruses [PIV1-4], and respiratory syncytial virus [RSV]), and three bacterial (*Bordetella pertussis, Chlamydia pneumoniae, and Mycoplasma pneumoniae*) pathogens [37, 50-56].

With more than 344,000 patient results for the FilmArray RP test alone, the Trend database has many of the properties associated with "big data" as it applies to infectious disease [57]. After describing how the dataset can be cleaned of non-patient tests, we make some observations on the seasonality of the different respiratory pathogens and we apply the ecological concept of "species diversity" [58] to observe a correlation between the abundance of each pathogen and the rate at which co-detections (more than one positive result per test) occur.

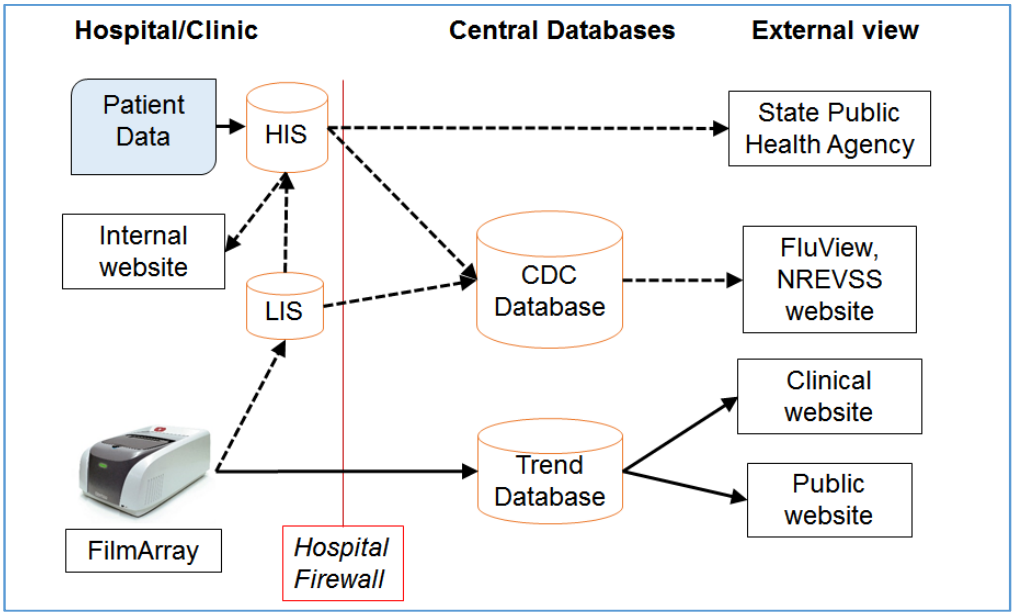
bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Sending FilmArray data directly to the cloud

The FilmArray Trend public website is an outgrowth of the Trend clinical website, which was developed to provide clinical laboratories using the FilmArray System with up-to-date information on the respiratory, gastrointestinal and meningitis/encephalitis pathogens circulating at their institutions. The most general and efficient way to export both the clinical and public Trend information is to follow a "Bottom-Out" approach to data export (Figure 1). In this scheme, the FilmArray Instrument sends data via the Internet directly to a single cloud database where it can be viewed by health care providers at the originating institution. This data export pathway contrasts with a "Top-Out" approach (Figure 1) in which diagnostic test results are pushed from the instrument up to the LIS, then to the HIS and, finally, a subset of this information is forwarded to cloud-based databases.

Figure 1. Schema for export of IVD test results to an external database



Bottom-Out and Top-Out approaches for data export are indicated by solid and dashed lines respectively. Some institutions have developed their own systems for aggregating and displaying infectious disease data (indicated by "Internal website"). Other abbreviations are defined in the text.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Initial testing of the Trend export mechanism was performed in collaboration with the Clinical Laboratories of the Medical University of South Carolina. This trial allowed us to develop and test auto-export functions and de-identification protocols for the Trend software. The de-identification requirement of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 [59], specifically the Safe Harbor provision, requires the removal of 18 enumerated variables that could directly or indirectly identify an individual. In accord with this requirement, the first stage study did not export test identifiers, or free form text fields and only returned the year of the test. The initial dataset provided low-resolution information but was a useful platform to evaluate the proposed system. Further development to enable export of higher resolution data required the design of routines that would adhere to an alternative HIPAA de-identification strategy, namely, the Expert Determination approach, which requires a risk assessment demonstrating that the chance of re-identifying an individual is sufficiently small [60, 61]. The Expert Determination process identified and made recommendations for fields that could facilitate disclosure of PHI. A summary of the Expert Determination results detailing the risk of Trend data in regard to replicability, availability and distinguishability is shown in Methods, Table 2.

Because the Expert Determination study established that no PHI will be disclosed by the Trend data export algorithm, Data Use Agreements (DUA), rather than Business Associates Agreements (BAA, see Methods for the difference between a DUA and a BAA) were executed with each of the collaborating institutions (listed in Table 3). The DUAs define for the clinical laboratory how BioFire will manage and make use of the Trend data. The Trend client software residing on the FilmArray computer queries the FilmArray test result local database and exports the results to an Amazon Web Services (AWS) database (see Methods). The Trend client software performs de-identification on the FilmArray computer prior to export as detailed in Table 2. Health care providers are granted access to their institutions' Trend data by the Laboratory Director. Since web access to view the data is restricted to the local site, de-identification of geographic indicators is not required. However, in implementation of the public Trend website, which presents the national syndromic test surveillance (Figure 1), we have further aggregated the data with respect to geographic origin and obfuscated the date of the test

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(see Methods). Since only de-identified data are exported from the clinical institutions, no PHI is sent to or stored on the cloud server.

All sites submitted the Trend project for review by their local Institutional Review Board (IRB); all but one of the 20 IRBs deemed the project "exempt" because of the absence of PHI export. Thus the security requirements for the database and the controls necessary for storage and transport of de-identified data are significantly reduced.

Following the protocol established by Expert Determination review, the Trend software delays the export of data until the number of tests queued for export exceeds a minimum threshold for each type of FilmArray panel. In practice, this results in an average time to export of less than two hours from each site that has multiple instruments. 98.5% of the tests are exported within 24 hours.

Characteristics of the FilmArray sites used in the Trend pilot study

The 20 sites contributing to the Trend pilot project (Methods, Table 3) have the same average number of instruments, six (range one to 22) as for all US FilmArray customers. The Trend pilot sites have been using the FilmArray RP test for an average of 3.8 years (range one to six) prior to July 2017. The size of the institutions participating ranges from 300 to 6400 beds with the majority being large hospitals and health care networks with an average of 1,100 beds. Six (30%) sites are pediatric hospitals and one is a reference laboratory. Fifteen (75%) of the sites have uploaded archived FilmArray RP test results to the Trend database, with eight (40%) reporting results dating back to 2012. Unless stated otherwise, the data presented here covers the period from July, 2013 to July, 2017.

The algorithm used to diagnose the cause of respiratory disease varies by site. More than half of the Trend sites do not enforce an institutional respiratory testing protocol and, even within sites that have a required protocol, some discretionary use of FilmArray RP is allowed. Without detailed records from each institution's HIS, it is not possible to determine whether the FilmArray RP was used as a front line test or as a reflex test (typically following a negative result for influenza and/or RSV).

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cleaning non-patient test results from the FilmArray Trend database

To determine the prevalence of respiratory pathogens, we needed to expunge the Trend database of test results that are not derived from clinical patient samples. Non-patient results come from a variety of sources including verification testing, routine Quality Control (QC) and proficiency testing (PT) (see Methods, Cleaning the Trend data). Despite this complexity, the majority of non-patient test results can be identified and distinguished from the patient-derived data because of the high number of positive organism calls in a single test and because of the temporal aspects of verification and control testing (see Methods). QC tests are estimated to account for half of all FilmArray RP results in which more than 3 organisms are detected. A combination of filters (Methods) removes 3.52% of the total tests from the Trend data set. We estimate that 3% of the remaining dual and 6% of the remaining triple detection tests respectively may come from QC testing. In addition to the exclusion of tests temporally associated with validation events, all results with four or more positives were removed from further analysis (0.9% of the filtered total total). This includes the 0.16% of test results with exactly four organisms (Table 1, “Tests after event removal” column) because the minority are derived from patient testing.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: Distribution of number of positive results per test for Trend RP tests

# Positives per RP Test	All	Tests after event removal	Final Trend Dataset
0	48.16	48.47	48.91
1	42.65	43.04	43.43
2	6.64	6.65	6.71
3	0.96	0.94	0.95
4	0.19	0.16	0.00
5	0.07	0.06	0.00
6	0.06	0.04	0.00
7	0.54	0.26	0.00
8	0.01	0.01	0.00
9	0.06	0.03	0.00
10	0.03	0.02	0.00
11	0.03	0.02	0.00
12	0.50	0.26	0.00
13	0.00	0.00	0.00
14	0.01	0.00	0.00
15	0.01	0.01	0.00
16	0.03	0.02	0.00
17	0.03	0.01	0.00
18	0.02	0.00	0.00
19	0.00	0.00	0.00

Percent of FilmArray RP tests in the Trend dataset that are positive for the indicated number of organisms for all data in the Trend dataset (All), after removing known control tests (Tests after event removal), and what is analyzed here (Final Trend Dataset). Data are scaled to sum to 100%. The panel tests for 20 organisms but a positive for Flu A H1-2009 is reported as negative for Flu A H1 so the maximum possible number of positive results is 19.

Table 1-Figure Supplement 1: QC runs are performed when a new lot of FilmArray RP pouches is introduced.

Detection of respiratory pathogens in Trend samples from 2013 to 2017

The detection counts and pathogen detection rates derived from the Trend data set for each organism in the FilmArray RP are shown in Figure 2. Other views of these data, including

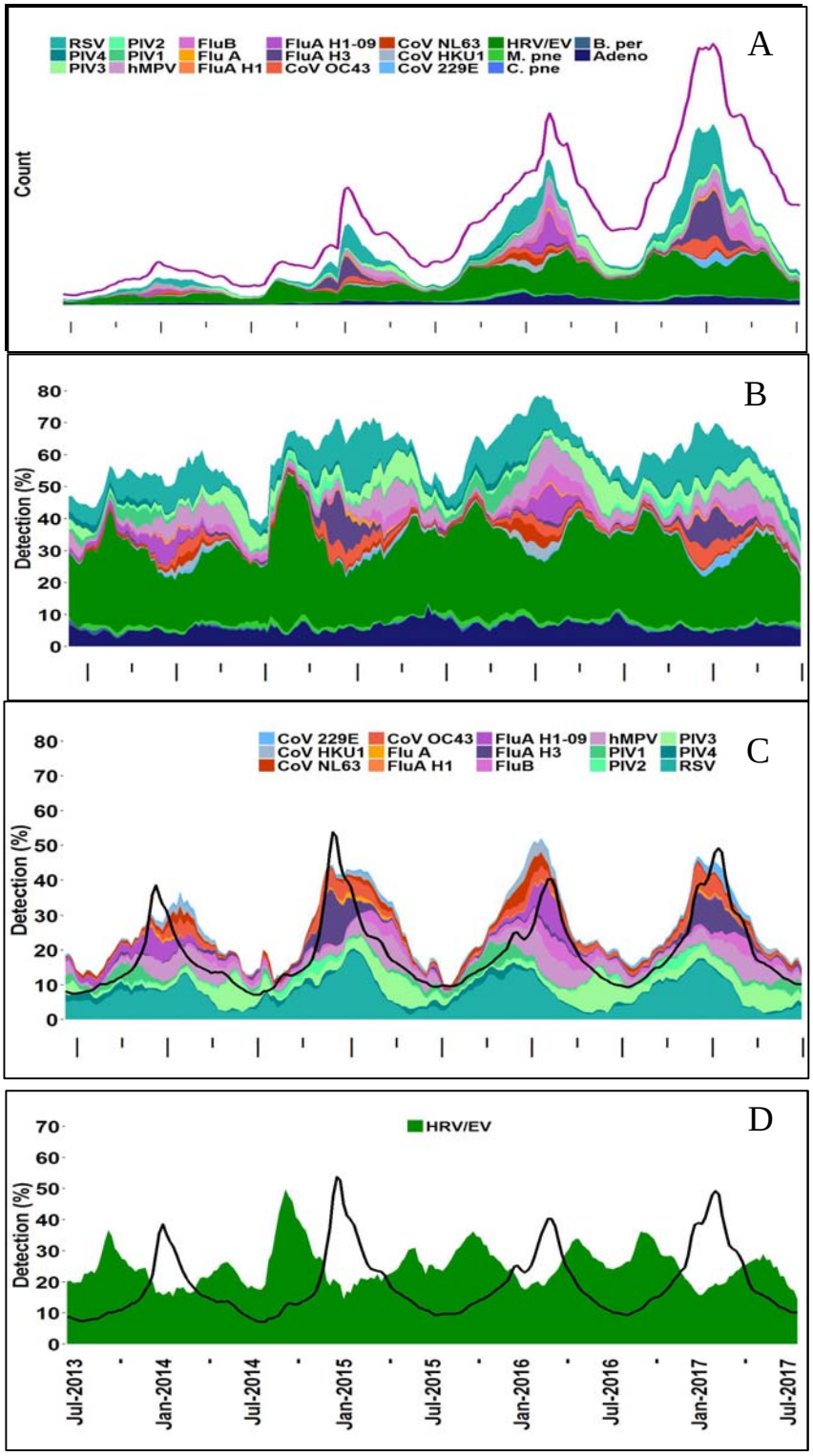
bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

percent detection of individual organisms or combinations of organisms, are available on the FilmArray Trend website: www.syndromictrends.com. The FilmArray RP Test Utilization Rate (RP TUR, see Methods) and the individual organism detection counts increased over this period because the Trend clinical sites increased their utilization of the FilmArray RP tests (Figure 2A). Seasonal fluctuations can also be seen within this growth pattern, with use increasing up to four-fold each winter when compared to the previous summer. HRV/EV, the most common pathogen detected group, is identified in approximately 30% of all samples tested each year (Figure 2-figure supplement 1). Other pathogens detected in five to 10% of the samples include: RSV, the PIVs, ADV, Flu A and hMPV. *M. pneumoniae, C. pneumoniae, B. pertussis* and Flu B are detected in less than 2% of all samples. The average percentage of each organism is relatively constant over the four years of data in the Trend database (Figure 2-figure supplement 2).

The pathogens' seasonal variability measured by percent detection can be classified into at least three groups. **Group 1**: The majority of organisms follow the classical "respiratory" season (October through March) and increase by more than 10-fold above their baseline detection rate (Figure 2C). These include the CoVs, Flu A, Flu B, hMPV, the PIVs, and RSV (PIV3 is a slight exception to this rule in that it peaks in the summer months, and has a winter peak that is only detected regionally (data not shown)). Within this group, all but five viruses demonstrate significant biennial fluctuations; Flu B, hMPV, OC43, and PIV3 and RSV experience relatively consistent annual peaks. **Group 2**: HRV/EV is in a class by itself in that it is detected in a high percentage of tests over time, (minimum of 10% in winter) and experiences moderate peaks of two to three fold outside the respiratory season baseline, in the early fall and spring (Figure 2D). **Group 3**: The bacteria and adenovirus are present at a relatively constant rate (Figure 2-Figure supplement 1). The CDC FluView reported rate of Influenza-Like-Illness (ILI) tracks moderately well with the Group 1 organisms (cross correlation of 0.85) and not with HRV/EV or with ADV and the bacteria.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2 Detection of RP organisms over time across all sites



255

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2 legend:

Detection of FilmArray RP pathogens in the Trend dataset displayed as a stacked area graphs. All panels have the same time period (July 2013 to June 2017). A) Count of each organism. The TUR metric (red line, units are FilmArray RP tests performed), and count of FilmArray RP tests that are negative (white area between pathogen count and TUR) are indicated. The y-axis values are not indicated as this is considered proprietary information. B) Pathogen detection rates for all organisms. C) Pathogen detection rates for the subset of organisms that show seasonality (see Results and the legend for the list of organisms). D) HRV/EV detection rates. CDC measured ILI is indicated (black line) in panels C and D. Organisms follow the same color scheme in all panels; the order of organisms in the legend (down then across) matches the stacked area graph top to bottom.

Figure 2-Figure Supplement 1: Detection of adenovirus and three bacteria

Figure 2-Figure Supplement 2: Detection of FilmArray RP organisms by type

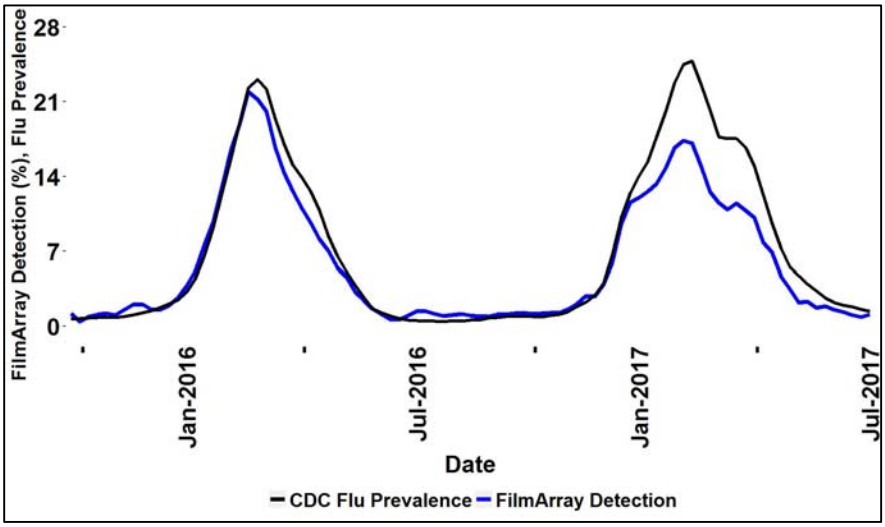
Figure 2-Figure Supplement 3: Organism detection by type and year

Comparison of Trend to CDC measures of influenza

The CDC FluView network [4, 5] gathers information about influenza prevalence from a large number of public health and clinical laboratories in the US. FluView is considered the gold standard for these measures. We compared the Trend detection rates for Flu A (all subtypes) plus Flu B to the FluView Influenza (A and B) from September 2015 to July 2017 (Figure 3). The analysis was restricted to this time period because of a change in the CDC's reporting of Flu prevalence in the fall of 2015. A cross-correlation of 0.974 was observed between the Trend Flu A/B percent detection and FluView reported influenza prevalence. Notably, the onset, peak and duration of the influenza season coincide between the two measures.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3: Trend Influenza Detection Rate compared to CDC Influenza activity



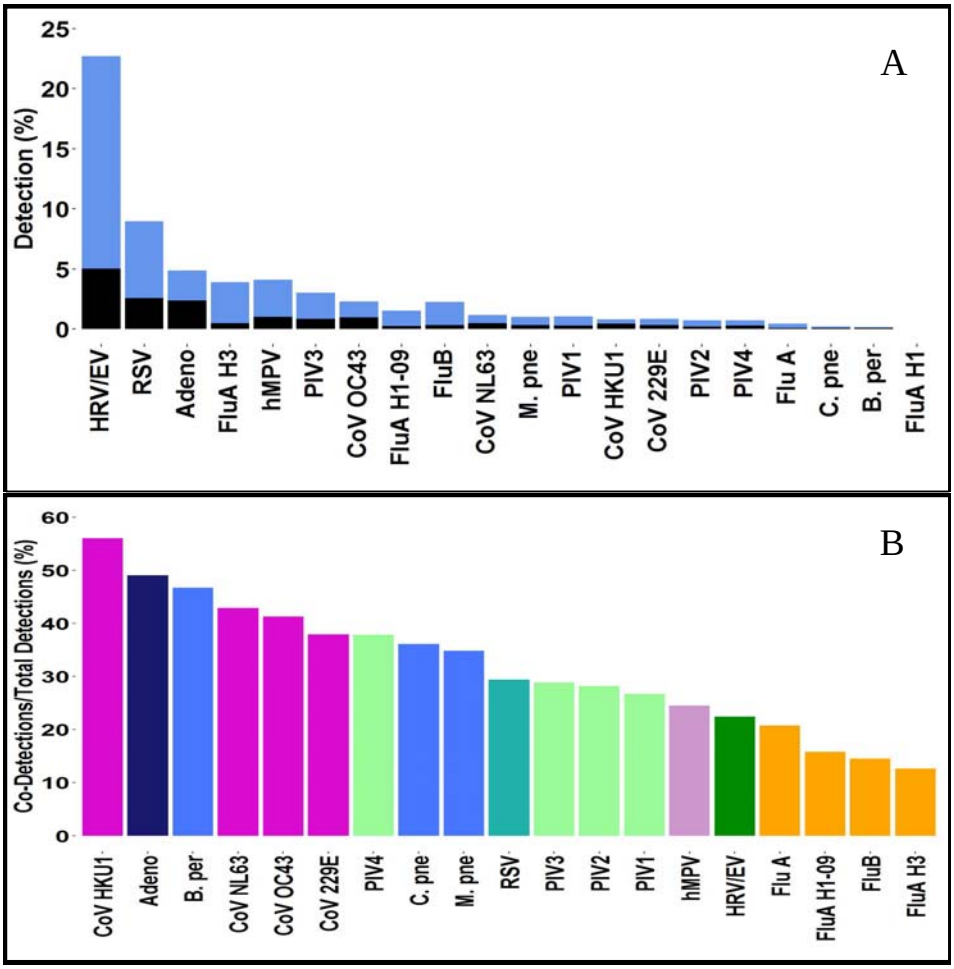
Percent of combined FilmArray Flu A (all subtypes) and Flu B detections (blue line) and CDC reported Influenza prevalence (black lines). CDC data are aggregated only from regions with participating Trend sites.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Respiratory Panel Co-Detections

We found that 26,347 FilmArray RP tests in the Trend dataset had 2 or 3 co-detections (7.7% of total FilmArray RP tests) and that the co-detection rate of each organism varies widely (10-55%). Although an additional pathogen was detected in half of the ADV and CoV positive samples, co-detections were observed in only 10-15% of the samples positive for either Flu A or Flu B (Figure 4).

Figure 4: Detection rates for all organisms compared to co-detections



(A) Percent total positive detections for each organism in the RP Trend data set is presented in stacked bars, showing the rate of detection of a single organism (A, blue) and those involved in a co-detection (A, black). Data are calculated for each site during the period of July 2013 to March 2017, when available, and then aggregated. **(B)** Percentage of each organism involved in a co-detection is shown. Bars are colored by pathogen family (CoV, purple; Bacteria, blue; PIV, green; Flu A, yellow).

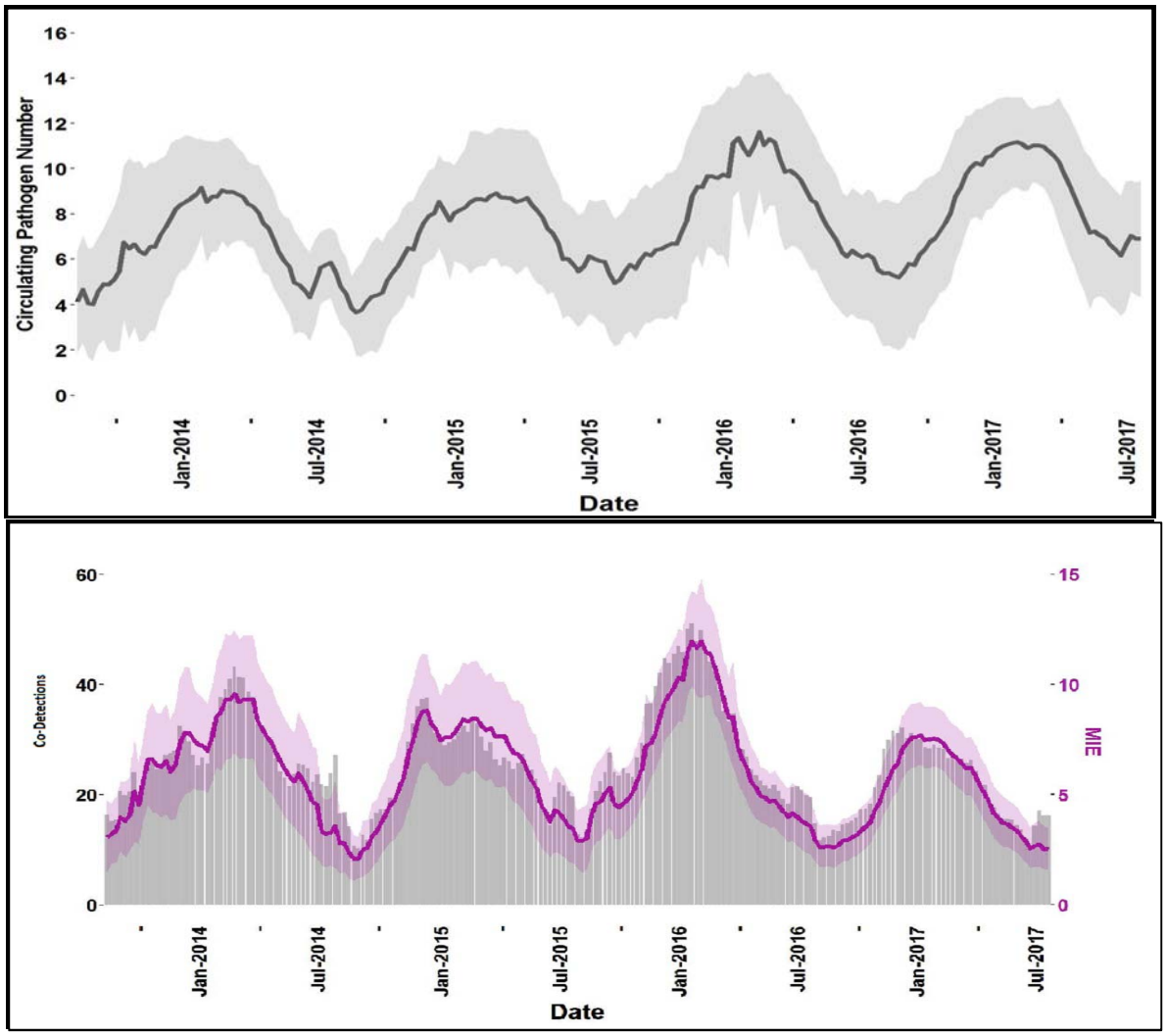
bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Trend data have high temporal, spatial, and organism-specific resolution. These three properties allow for a novel evaluation of co-detections. The observed rates of co-detections should be influenced by the number of circulating pathogens detected by the FilmArray RP test at a particular site. Figure 5A shows the average number of unique organisms detected at each site in a given week (see Methods: Calculation of co-detection rates). This number fluctuates from a summer low of four to a winter high of 11 pathogens. Figure 5B (grey bars) shows that the total rate of organism co-detections in the Trend dataset fluctuates annually with peak rates occurring in the winter months. The average rates have been as high as 12% in the winter of 2016 and as low as 2% in the summer of 2014.

From the Trend data, a Measure of Interspecific Encounter (MIE) can be calculated as the probability of a co-detection, weighted by the prevalence of each circulating pathogen at a site. Although the value of the MIE metric is higher than the actual co-detection rate it correlates well (Figure 5B, purple line compared to the grey bars has a cross correlation of 0.9488 at a lag of 0). The magnitude adjustment between MIE and the observed co-detections is calculated by the slope of the linear regression of the two metrics (Figure 5 – figure Supplement 1) and has a value of 4.05 ($R^2 = 0.9003$).

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5: Seasonal variation in pathogen diversity and co-detections



A) Average circulating pathogen number (black line) and one standard deviation computed across all Trend sites (grey area). **B)** Rate of co-detections in the RP Trend dataset (grey bars, left axis) the MIE index (purple line, right axis) and MIE confidence intervals (shaded purple area).

Figure 5-Figure supplement 1: Linear regression of MIE and observed co-detections

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

This article describes "FilmArray Trend", a new system for real-time reporting of widespread pathogen-specific syndromic data. This system already has many of the important properties of big data. We consider Trend in terms of the "V"s that are often used to describe big data: volume (amount), velocity (speed of acquisition), veracity (accuracy), variety (diversity of information) and value (utility) [57, 62].

Trend Volume

The Trend RP dataset is growing at an average rate of >200,000 pathogen test results per month. Connecting the first 20 clinical sites has provided insight into the principal concerns that will be raised by the legal, IT and administrative departments of the healthcare providers that house FilmArray instruments. It should be possible, therefore, to expand the Trend installed base by 10 to 20-fold over the next few years. Similarly, the existence of Trend should enable other IVD manufacturers to build their own Trend-like systems with greater acceptance on the part of their customers, thereby allowing a more global and comprehensive surveillance perspective.

Trend Velocity

The data in Figure 2 are similar to previous demonstrations of the seasonality associated with different respiratory viruses [63-66]. What is novel is that this data is generated automatically, on site, and in close to real-time compared to other surveillance systems. Greater than 98% of the test results are exported to the Trend database within 24 hours of being generated. As part of the de-identification protocol, sequential FilmArray RP tests of the same type are put into the same time bin. This has the effect that test results are exported faster during periods of peak use such as during the peak of the respiratory season or during an outbreak. Trend should be instrumental at a local level to determine the start of respiratory season; many hospitals make significant changes to their operations based on this event however, at present, data collection to track the respiratory season is often slow and manual, or semi-automated at best.

The key to implementing Trend clinical sites was to demonstrate that FilmArray test results can be exported without the risk of breaching PHI confidentiality either directly or through some combination of the data that was exported. Trend successfully used the Expert Determination process as prescribed by the HIPAA guidelines (see Methods), which greatly simplified the data

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sharing agreement between BioFire Diagnostics and the clinical site and allowed health care providers to use Trend without risk of inadvertently disclosing PHI.

The software architecture underlying the Trend system is both simple and secure: 1) no changes to the institutional firewall or LAN are needed; 2) the Trend database cannot reach back and query the FilmArray computer due to the institutional firewall, which is set to outbound data only; 3) Trend software can only submit data to the cloud database and cannot query the database. Yet, despite this security, authorized users of the Trend database can mine the de-identified data to look for novel patterns in respiratory pathogen epidemiology.

Trend Veracity

The goal of an epidemiological surveillance network is to infer which infectious diseases are circulating in the general population based on testing a sample of patients [67]. Different surveillance systems have different biases in their data, biases that perturb the ability to predict true population prevalence.

While the removal of all PHI has great benefits in terms of implementation, it also has several shortcomings that complicate interpretation of the data. First, Trend cannot account for the variability in the diagnostic testing algorithms applied to the selection of samples to be tested by the FilmArray Instruments. During respiratory season, health care providers may prescreen patients with other diagnostic tests including rapid antigen or molecular assays for influenza and RSV, and/or commercial and laboratory developed molecular tests for a mix of other respiratory pathogens. Depending upon the sensitivity of these upstream tests, 50-90% of influenza and RSV for the subset of the patients screened would be excluded from the Trend dataset if the front line test is positive. This testing protocol may skew the actual prevalence of not only influenza and RSV but all other individual respiratory pathogens and co-infections detected by the FilmArray. In some institutions, testing is reserved for hospitalized patients and others at risk for developing complications of respiratory tract infections including the very young, very old and immunocompromised patients, so Trend data may represent a less healthy patient population and not necessarily general community prevalence. Conversely there are sites that perform a significant number of tests for the outpatient setting. This may create variability among the

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

clinical sites' percent positivity and introduces a challenge to comparing pathogen intensity between sites.

The uncertainties surrounding the testing algorithm and the precise patient population tested should not interfere with determining the onset, peak and duration of the pathogen season at each institution. These limitations on the data are likely to be common among almost all current surveillance systems for similar reasons. Given these concerns, the agreement between the percent positivity of Flu A/B as determined by Trend and the percent positivity reported by CDC FluView Influenza is striking (Figure 3), supporting the validity and utility of the Trend data.

A second source of concern in the Trend data set is a consequence of removal of sample identification such that we cannot directly determine whether the sample was from a patient or was a non-clinical sample (verification test, QC or PT) and should be removed from further epidemiological analysis. We estimate that non-patient testing makes up approximately 1.8% of the total FilmArray RP tests. Automated detection algorithms remove 3.5% of the total RP tests, including approximately half of the non-clinical samples. With the exception of the four positive tests, the clinical samples removed by filtering should be a random sampling of all patient tests. The remaining 1% fraction of non-patient tests have essentially no impact on the Trend evaluation of pathogen prevalence but they do make it more difficult to perform high resolution analysis of pathogen co-detections. This is especially true for co-detections of low prevalence organisms where QC positives are likely to be more common than real positives. Future updates to the FilmArray software will simplify the process by which the instrument operator can tag tests of non-patient samples, thereby largely eliminating the need to filter such test results from the Trend database before analysis.

Trend Variety:

The total positivity rate of the FilmArray RP test varies from a low of 38% in the summer months to a high of 75% in December and January, with a yearly average of approximately 60%. Figure 5A shows that the average number of different circulating pathogens at a single institution can vary from eight up to 11 during the winter months. Even during the peak periods of ILI, many respiratory infections are due to other viruses (Figure 2C) that can present clinically in a

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

similar fashion [68, 69]. Therefore, the presumption of an influenza infection based on reported influenza percent positivity, without diagnostic testing for the virus, can lead to the inappropriate use of anti-viral agents [70]. Conversely, without comprehensive testing, a negative influenza or RSV test can lead to prescription of an unnecessary antibiotic. FilmArray Trend data can be a valuable aid for antimicrobial stewardship programs because it provides real-time information regarding the causes of respiratory infections and highlights the prevalence of viral infections.

As previously observed [66], the viruses that share the winter seasonality of influenza demonstrate annual or biennial behavior. It is possible that the viruses that share an influenza-like seasonality but do not show a two year cycle (RSV and hMPV) are actually alternating strains but the FilmArray RP Test does not detect this difference (e.g. the FilmArray RP does not differentiate between RSV A and RSV B). Adenovirus and the bacteria show constant occurrence through the year; HRV is in a unique class with peaks in the fall and spring.

Detection of multiple respiratory viruses in the same patient has been reported before. In the Trend dataset the rate of dual and triple co-detections was approximately 7.7 %, with HRV/EV as the organism most commonly observed in a co-detection. Some viruses, such as ADV and CoVs are detected in a mixed infection more than 50% of the time (Figure 4). In principle a FilmArray RP positive result may represent detection of residual pathogen nucleic acid from a previous infection that has resolved. However, several studies suggest that coinfections are associated with more severe disease ([71-73], see also discussion in [74]). In such cases, information about multiple detections can provide infection control practitioners with data that can assist in bed management and in the assessment of risk for nosocomial infections in a patient population that has been segregated by the occurrence of a common pathogen. Such information can prevent the introduction of a new pathogen associated with cohorting patients during busy respiratory seasons [75-77].

The question of whether different respiratory pathogens interfere with, or facilitate, growth in a human host is of some interest and not well understood. With the right data it can be studied at the population [78], individual [79], and cellular level [74]. Because the Trend data still includes

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

some non-patient tests, we have chosen not to analyze every possible dual or triple infection individually. Rather we have taken a global approach and compared the overall rate of observed co-detections with MIE, which is a measure of the diversity of viruses circulating in a specific region and time period. MIE is similar, but not identical, to PIE (Probability of Interspecific Encounter [80]), also referred to as the Gini-Simpson index (1-D, where D is the Simpson's index [81, 82]), which is used in ecology as a measure of the species diversity of a region. Similarly, the circulating pathogen number of Figure 5A is identical to the Species Richness measure of ecology. We calculate MIE using frequencies (P_i) of pathogen positivity per FilmArray test and note that the sum of all pathogen frequencies can add up to more than 100% because of co-detections or be less than 100% because of the presence of negative tests. In this regard, MIE differs from PIE because it is not a probability measure.

Figure 5B shows that the observed rate of co-detections is a constant fraction of MIE (approximately one quarter as indicated by the linear regression of Figure 5-Figure supplement 1). This observation suggests that, in the aggregate, respiratory pathogens are appearing in co-infections at a rate that can be predicted by their observed abundance. Data however may be biased by the patient population tested and the type of respiratory disease. The data also does not rule out that there are particular respiratory pathogens that occur more or less often in mixed infections than predicted by their individual percent positivity rates. [74, 83]. As we improve our ability to remove non-patient test results from the Trend dataset we will be able to characterize specific virus co-detection rates and their significance [65, 66, 78, 79, 84-86].

Trend Value

As with weather forecasting, there is both a theoretical and a practical interest in predicting the next few weeks or months of the respiratory season [87-90]. FilmArray Trend contributes to infectious disease forecasting efforts because the data is timely and comprehensive. As the number of sites participating in Trend increases it will be possible to localize the reported infections to smaller geographical regions. At a high enough density of Trend sites, patterns of movement of respiratory pathogens across the US will become visible in a way that has not been easily observed before now.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The Trend RP data is summarized as a stacked graph that shows the percentage contribution of each pathogen to what is currently being detected by FilmArray RP testing (Figure 2B, www.syndromictrends.com). This analysis does not take into account changes in the rate of testing over a given season, information that should provide additional data regarding disease intensity and severity. In contrast the simple metric, Test Utilization Rate (TUR), describes the non-normalized rate of FilmArray test usage and serves as a surrogate for the level of syndromic disease that health care providers observe (Figure 2A).

TUR suffers from two defects. First, it is closely linked to the sales of the FilmArray test and thus is proprietary data that BioFire does not share (Google took a similar position in regard to releasing the search queries used by Google Flu Trends [14]). Second, TUR is driven by both the demand for testing and the growth in FilmArray product adoption and increasing acceptance and usage by health care providers. A useful step beyond TUR would be a normalized Test Utilization Rate that can adjust for the underlying growth of testing unrelated to the intensity and duration of the respiratory disease season. An increase in a normalized TUR metric may indicate the prevalence of circulating respiratory viruses and the intensity of respiratory disease overall. Likewise, an increase in the normalized metric, concomitant with an increase in negative tests, may indicate the occurrence of an outbreak caused by an emerging pathogen.

Public health agencies (PHAs), which include local and state health departments and the CDC, are specifically exempt under an HIPAA provision that allows clinical laboratories to disclose PHI to the PHA for specified public health purposes [91]. The exemption includes follow up studies on reportable infectious diseases. Real-time pathogen specific syndromic surveillance systems such as Trend will allow state PHAs to more rapidly identify, acquire, and test residual samples from potential outbreaks. Conversely, perceived “outbreaks” may actually be coincidental multi-organism seasonal surges, and rapid analysis by Trend-like systems could prevent timely and costly outbreak investigation.

Given the movement in healthcare technology towards greater vertical integration of a hospital’s data, the Bottom-Out approach exemplified by Trend will face more competition from Top-Out

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

approaches (Figure 1, see for example GermWatch in Utah, [24]), because these systems can capture patient information (e.g. age, gender, and patient address) that is critical for more detailed epidemiological analysis. However, combining PHI with the diagnostic test result in the Top-Out approach makes these systems more complex and difficult to implement and may limit participation by health care institutions. Ironically, Bottom-Out data export systems have a role to play in the development of Top-Out systems because Bottom-Out export provides a rapid and efficient means to quality check the data flowing from Top-Out systems. FilmArray Trend data could also be combined with data derived from other automated diagnostic platforms [92, 93]. This work might best be accomplished by a third party that is viewed as independent and impartial. For example, in the case of data originating in the US, a federal institution or a private foundation could host a database to which IVD manufacturers would contribute their different syndromic test results. The benefits of a more complete view of circulating pathogens should outweigh the complexities of combining data from different platforms.

Outlook

FilmArray Trend is a novel surveillance tool for simultaneously monitoring multiple syndromic diseases that has demonstrated promise in expanding our knowledge of the epidemiology of infectious diseases. Indeed, the close correlation of seasonal respiratory viruses tracked by Trend with reported CDC ILI highlights the major contributory role of multiple respiratory pathogens beyond influenza to ILI. The national and global expansion of Trend will provide a comprehensive tool to study the impact of co-infections, understand the role of previously underappreciated pathogens, and clarify true disease epidemiology. Finally, systems like Trend will be essential for the rapid identification of disease anomalies indicating potential emergent outbreaks, thereby providing an independent tool for public health surveillance.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

FilmArray System and data output

The FilmArray System performs nucleic acid purification, reverse transcription, nested multiplex PCR amplification and DNA melt curve analysis for up to 30 targets in 63 minutes ([37], www.biofiredx.com). The FilmArray disposable pouch is prepared by injecting hydration solution and, separately, the patient sample, i.e., for the FilmArray RP test, a nasopharyngeal swab in viral transport media. The pouch is loaded into the FilmArray Instrument and the operator scans the barcode (containing panel type, lot and serial numbers) and then scans or manually enters information into the "Sample" field. Additional free form text can be entered into a "Tag" field. Software on the computer connected to the FilmArray Instrument analyses the amplification products and indicates each pathogen target detected. The results are presented as a PDF report and can be exported to the laboratory's LIS. Information about the patient (gender, age, date of birth etc.) is not recorded by the instrument unless specifically entered.

As of 2016, more than 4,000 FilmArray instruments have been placed in clinical use worldwide. Approximately 80% of these systems are in the US. The FilmArray RP test is United States Food and Drug Administration (US FDA) designated as Clinical Laboratory Improvement Act (CLIA) moderate status (can be used in moderate to high complexity laboratories).

HIPAA and the use of de-identified patient data for epidemiological tracking

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH, a part of the American Recovery and Reinvestment Act of 2009) apply to the use of IVD test results. The US Department of Health and Human Services (DHHS) has issued regulations that set standards for the use of patient data (see links at [94]). Under HIPAA, laboratories that perform the FilmArray test on patient samples are "Covered Entities". If laboratories send PHI to a third party (such as the IVD manufacturer) then that party is acting as a business associate (BA) of the Covered Entity and a Business Associate Agreement (BAA) would be required.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The HIPAA Privacy Rule [95] dictates that Covered Entities and, by extension, their BAs, may only disclose PHI with the patient's written authorization for the purposes of treatment, payment or normal health care operations, and a small number of additional exemptions [96]. Trend software resolves HIPAA concerns by removing all PHI before the test results leave the Covered Entity. This makes it highly unlikely that BioFire, or a malicious intruder into the Trend database, could associate the FilmArray test results with a patient. Because PHI is not exported, BioFire is not a BA and a Data Use Agreement (DUA) between the laboratory and BioFire addresses the export of FilmArray data to the Trend database. Demonstrating that Trend database does not contain PHI has been critical to recruiting institutions to this project.

Expert determination of de-identification

The HIPAA Privacy Rule allows for release of patient data without prior authorization as long as it has been properly de-identified [97]. There are two acceptable routes to de-identification, 1) the Safe Harbor approach wherein the data are stripped of an enumerated list of 18 identifiers classified as PHI and there is no indication (i.e., actual knowledge, that the remaining information would lead back to the individual) and 2) the Expert Determination approach wherein a person with experience in relevant statistical and scientific principles evaluates the PHI in conjunction with other reasonably available records, establishes a protocol for de-identification and certifies that the protocol allows only a small risk that PHI may be disclosed to an anticipated recipient.

The goal of the Trend project is to provide a near real-time view of the changes in pathogen prevalence; therefore, it is important to be able to retrieve the date when a FilmArray test is performed. However, retrieving the date conflicts with the Safe Harbor approach because the date of a test is PHI (the year of the test is not PHI but working with just the year defeats the purpose of tracking prevalence through the season). For this reason we followed the Expert Determination approach to manage data export.

The study took into consideration data that are available on participating clinical laboratory FilmArray Instruments, BioFire's own customer database, the proposed Trend database and publicly available data sources. We analyzed how combinations of this information could be

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

used by an adversary to identify an individual in the dataset thereby disclosing PHI [98]. The results of this study (summarized in Table 2) provided recommendations for development and site enrollment criteria for Trend, and for BioFire operating procedures.

Table 2: Qualitative analysis for the Expert Determination

Field	Replicability	Availability	Action
Sample	**High** (continually associated with a patient)	**High** (name exits in public records)	Deleted, the Trend database is not configured for the capture of this field.
Test Time and Date	**Medium** (only associated with patient once)	**Medium** (may be self-disclosed in public)	Binned into date range when minimum threshold of individuals is met
FilmArray Pouch Serial Number	**Medium** (only associated with patient once)	**Low** (only available to institution)	Truncated (last digit removed)
Test Location	**High** (likely continually associated with a patient)	**Medium** (may be self-disclosed in public)	Limited to sites drawing from > 20,000 pop.
Instrument Name	**Low** (not associated with patient)	**Low** (only available to institution)	No Action

In accord with the recommended actions, information in the fields that may be used to distinguish a patient is obfuscated (through truncation or binning) to ensure that a combination of these fields cannot be used to identify a specific patient [97]. For example, the time and date of the test are dynamically binned so that a minimum number of tests of one panel type are included in each bin prior to export. This ensures that a sufficient quantity of test results are uploaded to the database from one site at one time so that there is very low risk that patient identity can be inferred from knowledge of the start time of the test.

If an adversary were to infiltrate the safeguards of the database, and wished to know specific patient test results from a specific location on a given day, no unique records would exist. The combination of deleting the sample identification (ID) field, binning the test date range, and truncating the FilmArray pouch serial number ensures that the remaining information is never unique, which indicates that there is a low risk of misuse of data.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FilmArray Trend Client Software and Database

The Trend client software resides on the computer associated with the FilmArray Instrument(s). The computer makes a secure, HTTPS, connection to services hosted by Amazon Web Services. Authenticated data submissions are stored in a database hosted by Amazon Relational Database Service. Both services have been configured to be HIPAA Security Rule compliant [99]. Trend client software requires that the computer has Internet access to make secure outbound web requests. The HTTPS protocol is industry standard technology used for secure banking and web applications. The Website software authenticates the Trend data before saving the exported de-identified data to the Trend Database.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sites used in the Pilot Trend project:

The clinical laboratory sites participating in the pilot Trend project are shown in Table 3. The HHS regions are as defined by the CDC [5].

Table 3: Clinical laboratories participating in the initial Trend study

HHS region	Name	City, State	Hospital, Facility Type	Bed Count
1	University of Massachusetts Medical School-Baystate	Springfield, MA	University Affiliated	1,000
2	Albany Medical Center	Albany, NY	Medical College Affiliated	800
2	Northwell Health	Lake Success, NY	Network	6,400
2	NYU Langone Health	New York, NY	Network	1,500
2	Stony Brook University Hospital	Stony Brook, NY	University Affiliated	600
2	Winthrop University Hospital	Mineola, NY	University Affiliated	600
4	Greenville Health System	Greenville, SC	Network	1,400
4	Nicklaus Children's Hospital	Miami, FL	Children's	300
4	Medical University of South Carolina	Charleston, SC	University Affiliated	700
5	South Bend Medical Foundation	South Bend, IN	Commercial lab serving regional hospitals	NA
5	Detroit Medical Center	Detroit, MI	Network	2,000
5	Nationwide Children's Hospital	Columbus, OH	Children's	600
7	Children's Mercy	Kansas City, MO	Children's	400
7	Washington University Medical Center	St. Louis, MO	University Network	1,700
7	Nebraska Medical Center	Omaha, NE	University Network	600
8	Children's Hospital Colorado	Denver, CO	Children's Hospital	500
8	Intermountain Medical Center	Salt Lake City, UT	University Affiliated	500
8	Primary Children's Hospital	Salt Lake City, UT	Children's	300
9	Children's Hospital of Los Angeles	Los Angeles, CA	Children's, University Affiliated	500
9	UC San Diego Health	San Diego, CA	University Affiliated	700

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cleaning the Trend data

FilmArray Trend should display only valid test results acquired as part of the normal clinical testing of patients. Results from tests where an instrument, software or process control failure occurred have been removed from the data set. In addition, there are situations in which FilmArray customers test non-patient samples: FilmArray clinical customers can purchase Quality Control material from several different commercial sources and can make their own QC material by using cultured organism (from, for example, Zeptometrix, Buffalo, NY) either individually or in combination, or by using patient samples that previously tested positive for one or more organisms on the panel. Although BioFire provides recommendations for performing initial test validations [100] that involve combining pathogens in pools of 1-4 per test as well as negative controls, clinical laboratories may use other combinations of analytes. The signature of several of the commercial mixes can be identified simply by observing the set of organisms present. For example MMQC (Maine Molecular Quality Controls, Saco, ME) makes synthetic RNA mixes (M211 v1.1 and M212 v1.1) that explain 98% and 97% of the positive signals for seven and 12 analyte test results, respectively (Table 1, All, lines 7 and 12).

Additional QC testing is associated with specific events: a site adopts a new panel, qualifies an instrument returned from repair or, more frequently, qualifies a new lot of FilmArray RP pouches. Since we cannot be sure which tests are performed on control material, we identify events that are associated with test results containing more than three positive calls and flag all tests associated with these events for removal. Table 1-Figure Supplement 1 presents an example in the case of the QC tests performed to qualify a new lot of FilmArray RP pouches. In the first six days after a new lot of FilmArray pouches is received, tests with more than three positive organism calls make up 50, 48, 18, 10, 7, and 3% of the results. Thus, for a new lot of pouches, we remove the first three days of testing to minimize the contribution from control runs, thereby deleting 1.4% of all test results from Trend.

Clinical laboratories also perform periodic proficiency testing using test samples provided by either CAP (College of American Pathologists, Northfield IL) or API (American Proficiency Institute, Traverse City, MI). This testing occurs at defined time periods during the year and

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

BioFire is notified. Thus, we have knowledge of the organism mixes being tested as well as the number of such tests. Because only 0.05% of the results in the Trend data set are predicted to be part of CAP or API proficiency testing we have not removed them from the present study.

Because all patient identifiers are stripped before data are exported, there is no automated means to detect repeat tests on the same patient. However, the level of this testing is less than 1% of the total due to test cost and the fact that repeat testing should not be used for test-of-cure. For these reasons, the occurrence of patient duplicate testing would be rare and would not impact the overall Trend dataset. At present we cannot easily remove test results that are part of research protocols. Users at the Trend Pilot sites are instructed to add the Tag: "No Trend" to such tests (at present 0.03% of the Trend dataset) but due to the manual nature of the process we do not know how well this request is observed in practice.

Test Utilization Rate (TUR) and Pathogen Detection Rate

The FilmArray RP TUR metric is defined as the non-normalized number of RP patient test results generated each week on average across the Trend sites (computed as a centered three week moving average). To calculate the pathogen detection rate (as displayed in Figure 2A and on the Trend website) we compute the rate for each organism at each institution as a centered three week moving average. In order to adjust for the capacity differences between sites, a national aggregate is calculated as the unweighted average of individual site rates Only data from sites contributing more than 30 tests per week is included to avoid noise from small numbers of tests. Because the calculation of pathogen detection rate includes results from patients with multiple detections the detection rate for all organisms can, in theory, add up to greater than one. In practice this does not occur.

Comparison to the CDC Influenza observed rate of detection

The CDC FluView rate of Flu A and Flu B detections as well as the reported incidence of ILI are taken from the CDC website [101]. Only the CDC data from the HHS regions that contained Trend Pilot sites were used for calculating the rate of influenza detections (Table 3).

Calculation of co-detection rates and related measures

Pathogen co-detections are defined as FilmArray tests in which two or three organisms are detected. We also calculated two other measures that relate to co-detections: the Circulating

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Pathogen Number and the Measure of Interspecific Encounter (MIE). Both of these time series measures are calculated for each site and week, a centered five-week moving average is computed and then an unweighted average of all sites is used to create a national aggregate. The five week moving average is used to reduce noise due to small numbers of samples within a week at some sites.

More specifically the Circulating Pathogen Number is simply the count of the unique organisms detected at a site during a one-week period. MIE is calculated from the frequencies of each organism at a site (number of positive test results for an organism divided by the number of FilmArray tests performed at that site). To reduce noise, we only include site data if more than 10 FilmArray tests were performed in that week. If $P_1 \ldots P_N$ are the percentage detection of the N different organisms circulating at a single site over a single week then MIE is defined as:

$$MIE = \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} P_i P_j$$

Conceptually, MIE is an attempt to estimate the likelihood that a patient infected with one organism may be infected with another unique organism circulating in the population at a given period in time, resulting in a co-infection.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Additional information

Acknowledgements

We acknowledge funding from NIH grant 5U01AI074419 (L.M., K.M.R. and M.A.P) and NIH/NIAID contract HHSN272201600002C (MAP). We thank Chris Thurston and Spencer Rose (BioFire Defense) for building the final Trend public website; Andrew Wallin (BioFire Defense) for reviewing the MIE data analysis; Anna Hoffee (BioFire Diagnostics) for assistance with the figures; Mark Pallansch (CDC), Kirsten St. George (New York State Department of Health) and Allyn Nakashima (Utah Department of Health) for useful discussions; Anne Blaschke and colleagues at BioFire Diagnostics and BioFire Defense for reviewing the manuscript.

Author contributions:

L.M., K.M.R. and M.A.P. developed the Trend concept. P.H.G. and C.C.G provided early input on the scope of the project. L.M. designed the DUAs, recruited the sites and provided overall coordination of the project. C.V.C. coordinated the clinical site software installations, L.M. and R.N. designed and supervised the coding of the data export software, database and web interface for the data. L.M. and A.N.F. performed the data analysis and prepared the figures. B.A.M. reviewed the expert determination studies. S.V.S., B.M.A. and J.D.J. reviewed the epidemiological analysis. F.S.N. tested the first release of the data export software. F.S.N., P. H. G., A.L., D.J., V.D, J.D.B., S.S., K.A.S., H.S., R.S., S.J., J.A.D., J.C.W., K.L., F.M., S.L.R, M.A-R., P.D.F., G.A.S., S.J.M., C.C.R., J.F.M. implemented the Trend export software at their respective institutions. C.C.G. and M.A.P. provided periodic review of the project. M.A.P., L. M. and C.C.G. wrote the manuscript. All authors reviewed the final manuscript.

Author ORCIDs

Lindsay Meyers	0000-0003-2994-274X
Christine C. Ginocchio	0000-0002-8200-0324
Aimie N. Faucett	0000-0003-4590-4276
Frederick S. Nolte	0000-0002-4943-9220
Per H. Gesteland	0000-0002-9869-8122
Diane Janowiak	0000-0002-9208-0647
Virginia Donovan	0000-0002-9296-6931
Jennifer Dien Bard	0000-0003-0524-9473
Silvia Spitzer	0000-0002-2642-4771
Kathleen A. Stellrecht	0000-0002-0160-6034
Hossein Salimnia	0000-0002-8069-5185
Rangaraj Selvarangan	0000-0003-3275-6657

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stefan Juretschko	0000-0002-4658-6276
Judy A. Daly	0000-0003-4801-7023
Kristy Lindsey	0000-0002-8172-5964
Franklin Moore	0000-0002-3019-0437
Sharon L. Reed	0000-0001-7615-9366
Paul D. Fey	0000-0003-0939-6884
Camille V. Cook	0000-0001-6071-4898
Jay D. Jones	0000-0001-9091-5316
Samuel V. Scarpino	0000-0001-5716-2770
Benjamin M. Althouse	0000-0002-5464-654X
Kirk M. Ririe	0000-0003-0620-7419
Bradley A. Malin	0000-0003-3040-5175
Mark A. Poritz	0000-0002-5060-0560

Competing interests

L.M., A.N.F., R.K.N., C.V.C, J.D.J., K.M.R., C.C.G. and M.A.P. are present or former employees of bioMérieux, Inc. or its subsidiaries. bioMérieux markets the FilmArray System and Trend. F.S.N., P.H.G., D.J., V.D., A.L., J.D.B., S.S., K.A.S., H.S., R.S., S.J., J.A.D., J.C.W., K.L., F.M., S.L.R., M.A-R., P.D.F., G.A.S., S.J.M., S.V.S., B.M.A., are research contractors of BioFire Diagnostics for the development of the FilmArray Trend system. CCR and JFM are members of the Trend Working Group. B.A.M. is a paid consultant of BioFire Diagnostics.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Marinović, A.B., et al., *Quantifying Reporting Timeliness to Improve Outbreak Control.* Emerg Infect Dis, 2015. **21**(2)
2. Yang, W., A. Karspeck, and J. Shaman, *Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics.* PLoS Comput Biol, 2014. **10**(4)
3. van den Wijngaard, C.C., et al., *Evaluation of syndromic surveillance in the Netherlands: its added value and recommendations for implementation.* Euro Surveill, 2011. **16**(9)
4. CDC. *Flu Activity & Surveillance*. [cited April 30, 2017]; Available from: https://www.cdc.gov/flu/weekly/fluactivitysurv.htm.
5. CDC. *Overview of Influenza Surveillance in the United States*. [cited April 30, 2017]; Available from: http://www.cdc.gov/flu/pdf/weekly/overview.pdf.
6. CDC. *National Respiratory and Enteric Virus Surveillance System (NREVSS)*. [cited April 30, 2017]; Available from: http://www.cdc.gov/surveillance/nrevss/.
7. National Electronic Disease Surveillance System Working, G., *National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine.* J Public Health Manag Pract, 2001. **7**(6)
8. Lamb, E., et al., *Update on progress in electronic reporting of laboratory results to public health agencies - United States, 2014.* MMWR Morb Mortal Wkly Rep, 2015. **64**(12)
9. United States Department of Defense. *Global Emerging Infections Surveillance and Response System*. [cited April 30, 2017]; Available from: https://www.health.mil/Military-Health-Topics/Health-Readiness/Armed-Forces-Health-Surveillance-Branch/Global-Emerging-Infections-Surveillance-and-Response.
10. Nelson, N.P., et al., *Event-based internet biosurveillance: relation to epidemiological observation.* Emerg Themes Epidemiol, 2012. **9**(1)
11. Cheng, C.K., et al., *A profile of the online dissemination of national influenza surveillance data.* BMC Public Health, 2009. **9**
12. Hulth, A. and G. Rydevik, *GET WELL: an automated surveillance system for gaining new epidemiological knowledge.* BMC Public Health, 2011. **11**
13. CDC. *BioSense*. [cited September 20, 2017]; Available from: https://www.cdc.gov/nssp/biosense/index.html.
14. Ginsberg, J., et al., *Detecting influenza epidemics using search engine query data.* Nature, 2009. **457**(7232)
15. GoogleFlu. 2015, *The Next Chapter for Flu Trends*. [cited April 30, 2017]; Available from: http://googleresearch.blogspot.com/2015/08/the-next-chapter-for-flu-trends.html.
16. FluNearYou. *Flu Near You | HealthMap*. [cited April 30, 2017]; Available from: https://flunearyou.org/.
17. Das, D., et al., *Monitoring over-the-counter medication sales for early detection of disease outbreaks--New York City.* MMWR Suppl, 2005. **54**
18. Patwardhan, A. and R. Bilkovski, *Comparison: Flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator.* PLoS One, 2012. **7**(8)
19. Broniatowski, D.A., M.J. Paul, and M. Dredze, *National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic.* PLoS One, 2013. **8**(12)

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. Signorini, A., A.M. Segre, and P.M. Polgreen, *The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic.* PLoS One, 2011. **6**(5)

21. Generous, N., et al., *Global disease monitoring and forecasting with Wikipedia.* PLoS Comput Biol, 2014. **10**(11)

22. McIver, D.J. and J.S. Brownstein, *Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time.* PLoS Comput Biol, 2014. **10**(4)

23. IHC. *GermWatch - Intermountain Healthcare - Salt Lake City, Utah*. [cited April 30, 2017]; Available from: http://intermountainhealthcare.org/health-resources/germwatch/Pages/home.aspx.

24. Gesteland, P.H., et al., *Informing the front line about common respiratory viral epidemics.* AMIA Annu Symp Proc, 2007

25. NYSDOH. *Electronic Clinical Laboratory Reporting System*. [cited April 30, 2017]; Available from: http://www.health.ny.gov/professionals/reportable_diseases/eclrs/.

26. Shah, S. and C. Messina, *The time-associated impact of the Newborn Influenza Protection Act on infant influenza rates in New York State.* J Perinat Med, 2014. **42**(6)

27. Crawford, J.M., et al., *Laboratory surge response to pandemic (H1N1) 2009 outbreak, New York City metropolitan area, USA.* Emerg Infect Dis, 2010. **16**(1)

28. Ginocchio, C.C., et al., *Evaluation of multiple test methods for the detection of the novel 2009 influenza A (H1N1) during the New York City outbreak.* J Clin Virol, 2009. **45**(3)

29. Al-Tawfiq, J.A., et al., *Surveillance for emerging respiratory viruses.* Lancet Infect Dis, 2014. **14**(10)

30. Gautret, P., et al., *Emerging viral respiratory tract infections-environmental risk factors and transmission.* Lancet Infect Dis, 2014. **14**(11)

31. Hazelton, B., et al., *Detection of influenza A and B with the Alere i Influenza A & B: a novel isothermal nucleic acid amplification assay.* Influenza Other Respir Viruses, 2015. **9**(3)

32. Salez, N., et al., *Prospective and retrospective evaluation of the Cepheid Xpert(R) Flu/RSV XC assay for rapid detection of influenza A, influenza B, and respiratory syncytial virus.* Diagn Microbiol Infect Dis, 2015. **81**(4)

33. Nolte, F.S., L. Gauld, and S.B. Barrett, *Direct Comparison of Alere i and cobas Liat Influenza A and B Tests for Rapid Detection of Influenza Virus Infection.* J Clin Microbiol, 2016. **54**(11)

34. Buchan, B.W. and N.A. Ledeboer, *Emerging technologies for the clinical microbiology laboratory.* Clin Microbiol Rev, 2014. **27**(4)

35. Zhang, H., S. Morrison, and Y.W. Tang, *Multiplex polymerase chain reaction tests for detection of pathogens associated with gastroenteritis.* Clin Lab Med, 2015. **35**(2)

36. Salez, N., et al., *Evaluation of Four Commercial Multiplex Molecular Tests for the Diagnosis of Acute Respiratory Infections.* PLoS One, 2015. **10**(6)

37. Poritz, M.A., et al., *FilmArray, an automated nested multiplex PCR system for multi-pathogen detection: development and application to respiratory tract infection.* PLoS One, 2011. **6**(10)

38. Pierce, V.M. and R.L. Hodinka, *Comparison of the GenMark Diagnostics eSensor respiratory viral panel to real-time PCR for detection of respiratory viruses in children.* J Clin Microbiol, 2012. **50**(11)

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

39. GenMarkDx. 2016, *The ePlex® Respiratory Pathogen Panel (RP)**. [cited April 30, 2017]; Available from: https://genmarkdx.com/wp-content/uploads/2016/09/GNMK-IMC-1046-E-ePlex-RP-Data-Sheet-US.pdf.

40. Gadsby, N.J., et al., *Comparison of the Luminex Respiratory Virus Panel fast assay with in-house real-time PCR for respiratory viral infection diagnosis.* J Clin Microbiol, 2010. **48**(6)

41. Lee, C.K., et al., *Comparison of Luminex NxTAG Respiratory Pathogen Panel and xTAG Respiratory Viral Panel FAST Version 2 for the Detection of Respiratory Viruses.* Ann Lab Med, 2017. **37**(3)

42. Butt, S.A., et al., *Comparison of three commercial RT-PCR systems for the detection of respiratory viruses.* J Clin Virol, 2014. **61**(3)

43. van Panhuis, W.G., et al., *A systematic review of barriers to data sharing in public health.* BMC Public Health, 2014. **14**

44. Williams, P.A. and A.J. Woodward, *Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem.* Med Devices (Auckl), 2015. **8**

45. Benitez, K. and B. Malin, *Evaluating re-identification risks with respect to the HIPAA privacy rule.* J Am Med Inform Assoc, 2010. **17**(2)

46. El Emam, K., et al., *A systematic review of re-identification attacks on health data.* PLoS One, 2011. **6**(12)

47. El Emam, K., S. Rodgers, and B. Malin, *Anonymising and sharing individual patient data.* BMJ, 2015. **350**

48. Sarpatwari, A., et al., *Ensuring patient privacy in data sharing for postapproval research.* N Engl J Med, 2014. **371**(17)

49. BioFire Diagnostics. *FilmArray® Panels*. [cited April 30, 2017]; Available from: http://www.biofiredx.com/products/the-filmarray-panels/.

50. Hayden, R.T., et al., *Comparison of two broadly multiplexed PCR systems for viral detection in clinical respiratory tract specimens from immunocompromised children.* J Clin Virol, 2012. **53**(4)

51. Loeffelholz, M.J., et al., *Comparison of the FilmArray Respiratory Panel and Prodesse real-time PCR assays for detection of respiratory pathogens.* J Clin Microbiol, 2011. **49**(12)

52. Pierce, V.M., et al., *Comparison of the Idaho Technology FilmArray system to real-time PCR for detection of respiratory pathogens in children.* J Clin Microbiol, 2012. **50**(2)

53. Rand, K.H., H. Rampersaud, and H.J. Houck, *Comparison of two multiplex methods for detection of respiratory viruses: FilmArray RP and xTAG RVP.* J Clin Microbiol, 2011. **49**(7)

54. Xu, M., et al., *Implementation of filmarray respiratory viral panel in a core laboratory improves testing turnaround time and patient care.* Am J Clin Pathol, 2013. **139**(1)

55. Couturier, M.R., et al., *Evaluation of the FilmArray(R) Respiratory Panel for clinical use in a large children's hospital.* J Clin Lab Anal, 2013. **27**(2)

56. Popowitch, E.B., S.S. O'Neill, and M.B. Miller, *Comparison of the Biofire FilmArray RP, Genmark eSensor RVP, Luminex xTAG RVPv1, and Luminex xTAG RVP fast multiplex assays for detection of respiratory viruses.* J Clin Microbiol, 2013. **51**(5)

57. Bansal, S., et al., *Big Data for Infectious Disease Surveillance and Modeling.* Journal of Infectious Diseases, 2016. **214**(suppl 4)

58. Magurran, A.E., *Measuring biological diversity*. 2004, Malden, Ma: Blackwell Pub. viii.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

59. U.S. Dept. of Health and Human Services, *Standards for privacy of individually identifiable health information, Final Rule.* . 2002.

60. Malin, B., D. Karp, and R.H. Scheuermann, *Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research.* J Investig Med, 2010. **58**(1)

61. Office for Civil Rights, *Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, U.S. Department of Health and Human Services., Editor. 2012, U.S. Department of Health and Human Services.

62. IBM. *Infographic: The Four V's of Big Data*. [cited June 8, 2017]; Available from: http://www.ibmbigdatahub.com/infographic/four-vs-big-data.

63. Stockmann, C., et al., *Seasonality of acute otitis media and the role of respiratory viral activity in children.* Pediatr Infect Dis J, 2013. **32**(4)

64. Radin, J.M., et al., *Epidemiology of pathogen-specific respiratory infections among three US populations.* PLoS One, 2014. **9**(12)

65. Brittain-Long, R., et al., *Seasonal variations of 15 respiratory agents illustrated by the application of a multiplex polymerase chain reaction assay.* Scand J Infect Dis, 2012. **44**(1)

66. Olofsson, S., et al., *PCR for detection of respiratory viruses: seasonal variations of virus infections.* Expert Rev Anti Infect Ther, 2011. **9**(8)

67. Reed, C., et al., *Estimating influenza disease burden from population-based surveillance data in the United States.* PLoS One, 2015. **10**(3)

68. Ruuskanen, O., et al., *Viral pneumonia.* Lancet, 2011. **377**(9773)

69. Pavia, A.T., *Viral infections of the lower respiratory tract: old viruses, new viruses, and the role of diagnosis.* Clin Infect Dis, 2011. **52 Suppl 4**

70. Brendish, N.J., et al., *Routine molecular point-of-care testing for respiratory viruses in adults presenting to hospital with acute respiratory illness (ResPOC): a pragmatic, open-label, randomised controlled trial.* Lancet Respir Med, 2017. **5**(5)

71. Goka, E.A., et al., *Single, dual and multiple respiratory virus infections and risk of hospitalization and mortality.* Epidemiol Infect, 2015. **143**(1)

72. Martin, E.T., et al., *Epidemiology of multiple respiratory viruses in childcare attendees.* J Infect Dis, 2013. **207**(6)

73. Rhedin, S., et al., *Clinical utility of PCR for common viruses in acute respiratory illness.* Pediatrics, 2014. **133**(3)

74. Pinky, L. and H.M. Dobrovolny, *Coinfections of the Respiratory Tract: Viral Competition for Resources.* PLoS One, 2016. **11**(5)

75. Thorburn, K., et al., *RSV outbreak in a paediatric intensive care unit.* J Hosp Infect, 2004. **57**(3)

76. Gooskens, J., et al., *Morbidity and mortality associated with nosocomial transmission of oseltamivir-resistant influenza A(H1N1) virus.* JAMA, 2009. **301**(10)

77. Te Wierik, M.J., et al., *An outbreak of severe respiratory tract infection caused by human metapneumovirus in a residential care facility for elderly in Utrecht, the Netherlands, January to March 2010.* Euro Surveill, 2012. **17**(13)

78. Casalegno, J.S., et al., *Rhinoviruses delayed the circulation of the pandemic influenza A (H1N1) 2009 virus in France.* Clin Microbiol Infect, 2010. **16**(4)

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

79. Karppinen, S., et al., *Interference between respiratory syncytial virus and rhinovirus in respiratory tract infections in children.* Clin Microbiol Infect, 2016. **22**(2)

80. Hurlbert, S.H., *The Nonconcept of Species Diversity: A Critique and Alternative Parameters.* Ecology, 1971. **52**(4)

81. Jost, L., *Entropy and diversity.* Oikos, 2006. **113**(2)

82. Krebs, C.J., *Ecological methodology*. 2nd ed. 1999, Menlo Park, Calif.: Benjamin/Cummings. xii.

83. Essa, S., et al., *Mixed viral infections circulating in hospitalized patients with respiratory tract infections in kuwait.* Adv Virol, 2015. **2015**

84. Bhattacharyya, S., et al., *Cross-immunity between strains explains the dynamical pattern of paramyxoviruses.* Proc Natl Acad Sci U S A, 2015. **112**(43)

85. Wen, X., et al., *Structural basis for antibody cross-neutralization of respiratory syncytial virus and human metapneumovirus.* Nat Microbiol, 2017. **2**

86. Achten, N.B., et al., *Interference Between Respiratory Syncytial Virus and Human Rhinovirus Infection in Infancy.* J Infect Dis, 2017. **215**(7)

87. Nsoesie, E.O., et al., *A systematic review of studies on forecasting the dynamics of influenza outbreaks.* Influenza Other Respir Viruses, 2014. **8**(3)

88. Chretien, J.P., et al., *Influenza forecasting in human populations: a scoping review.* PLoS One, 2014. **9**(4)

89. Biggerstaff, M., et al., *Results from the centers for disease control and prevention's predict the 2013-2014 Influenza Season Challenge.* BMC Infect Dis, 2016. **16**

90. Moran, K.R., et al., *Epidemic Forecasting is Messier Than Weather Forecasting: The Role of Human Behavior and Internet Data Streams in Epidemic Forecast.* Journal of Infectious Diseases, 2016. **214**(suppl 4)

91. HHS. 2003, *Disclosures For Public Health Activities*. [cited April 30, 2017]; Available from: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/special/publichealth/publichealth.pdf.

92. Shekalaghe, S., et al., *Clinical performance of an automated reader in interpreting malaria rapid diagnostic tests in Tanzania.* Malar J, 2013. **12**

93. InSTEDD. *Connectivity Solutions for TB Diagnostics*. [cited April 30, 2017]; Available from: http://instedd.org/project/connectivity-solutions-for-tb-diagnostics/.

94. HHS. *Health Information Privacy*. [cited; Available from: http://www.hhs.gov/ocr/privacy/index.html.

95. HHS. *Summary of the HIPAA Privacy Rule*. [cited; Available from: https://www.hhs.gov/sites/default/files/privacysummary.pdf.

96. HHS. *Special Topics in Health Information Privacy*. [cited; Available from: http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/index.html.

97. Office for Civil Rights. 2012, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. [cited April 30, 2017]; Available from: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf.

98. Federal Committee on Statistical Methodology, *Statistical and Science Policy, Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology*. 2005, Office of Information and Regulatory Affairs, Office of Management and Budget,.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

99. Amazon. 2017, *Architecting for HIPAA Security and Compliance on Amazon Web Services*. [cited April 30, 2017]; Available from: https://d0.awsstatic.com/whitepapers/compliance/AWS_HIPAA_Compliance_Whitepaper.pdf.

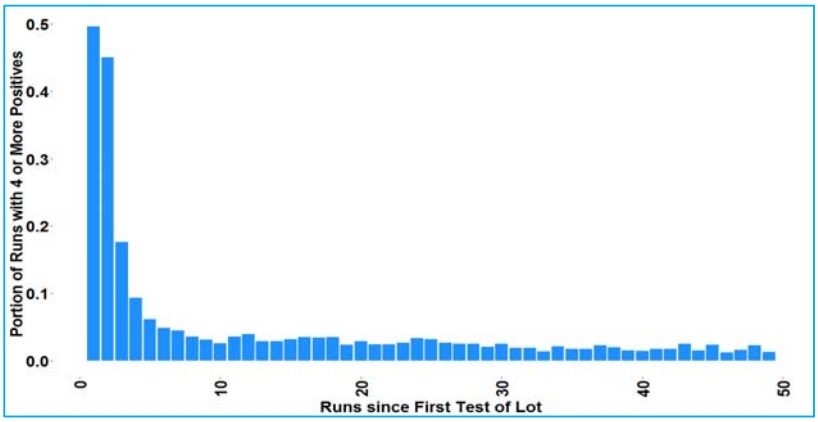
100. BioFire Diagnostics. *Guidelines for Laboratory Verification of Performance of the FilmArray® Respiratory Panel (RP)*. [cited April 30, 2017]; Available from: http://www.biofiredx.com/wp-content/uploads/2016/03/TN-FLM1-PRT-0060-07-Guidelines-for-Laboratory-Verification-of-Performance-of-the-FilmArray-Respiratory-Panel.pdf.

101. CDC. *Nation and Regional Level Outpatient Illness and Viral Surveillance*. [cited June 8, 2017]; Available from: https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

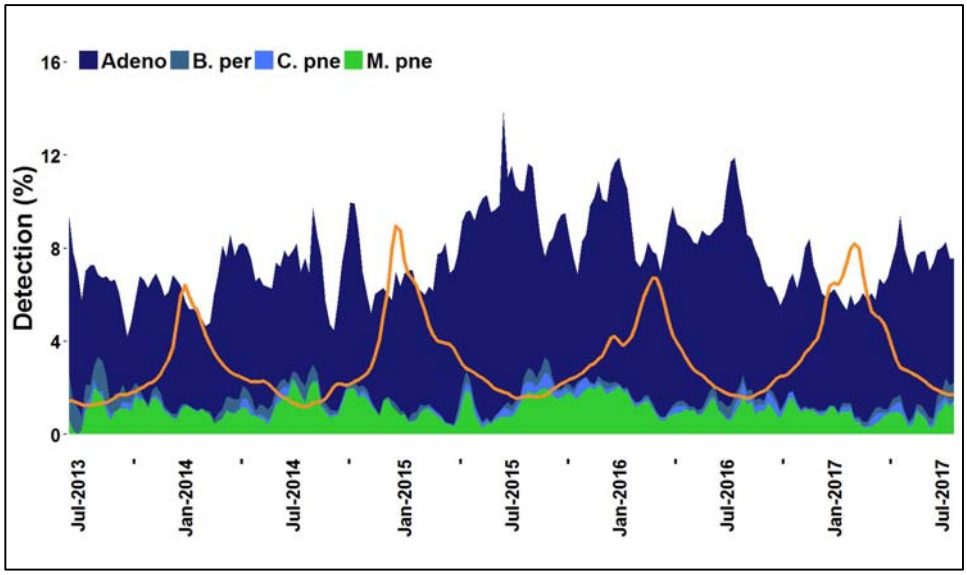
Supplementary Figures

Table 1-Figure Supplement 1: QC runs are performed when a new lot of FilmArray RP pouches is introduced



Histogram of count of FilmArray RP tests with more than three positive results after a new lot of FilmArray RP pouches is introduced at a clinical site. The X-axis is the sequential number of RP tests associated with the new lot that have been performed at that site.

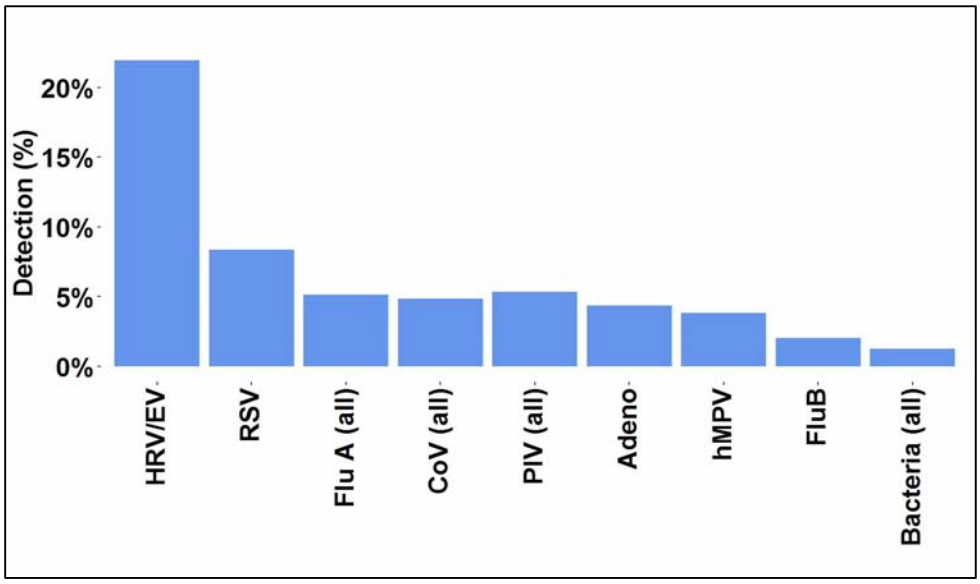
Figure 2-Figure Supplement 1: Detection of adenovirus and three bacteria



Detection of adenovirus and the three bacterial pathogens (*B. pertussis*, *C. pneumoniae*, and *M. pneumoniae*) in the Trend dataset displayed as a stacked area graph. CDC measured ILI is indicated (orange line).

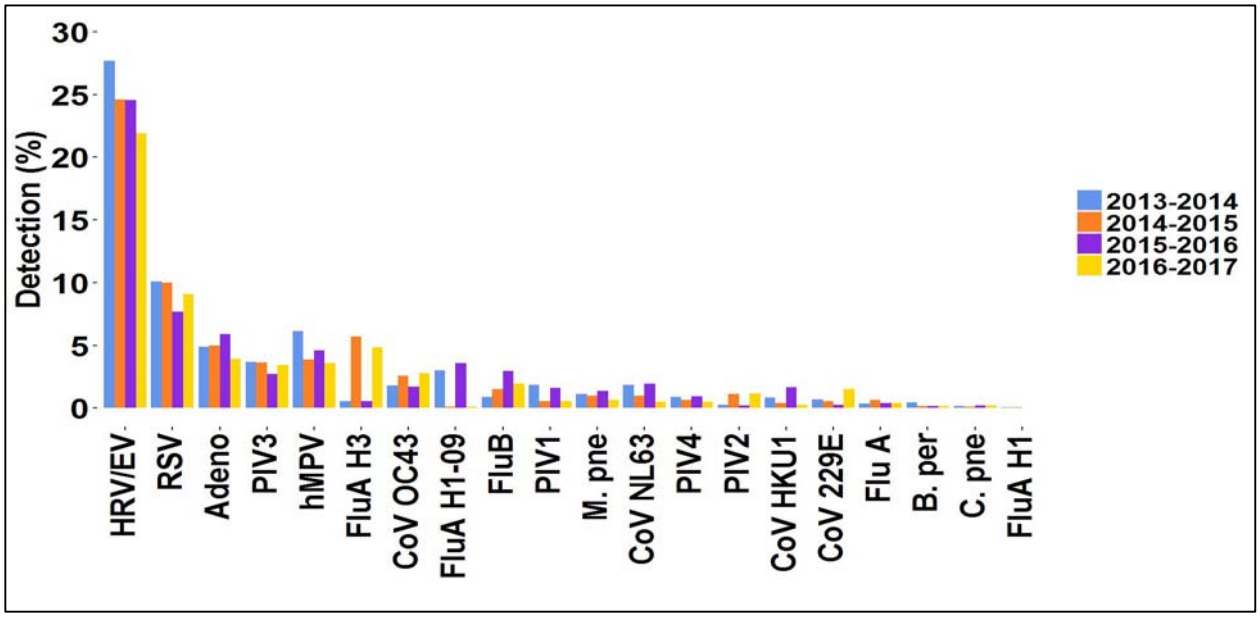
bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2-Figure Supplement 2: Detection of FilmArray RP organisms by type



Percent detection rates aggregated from all participating Trend sites. The percent positive detection per test for each organism group is ordered by abundance.

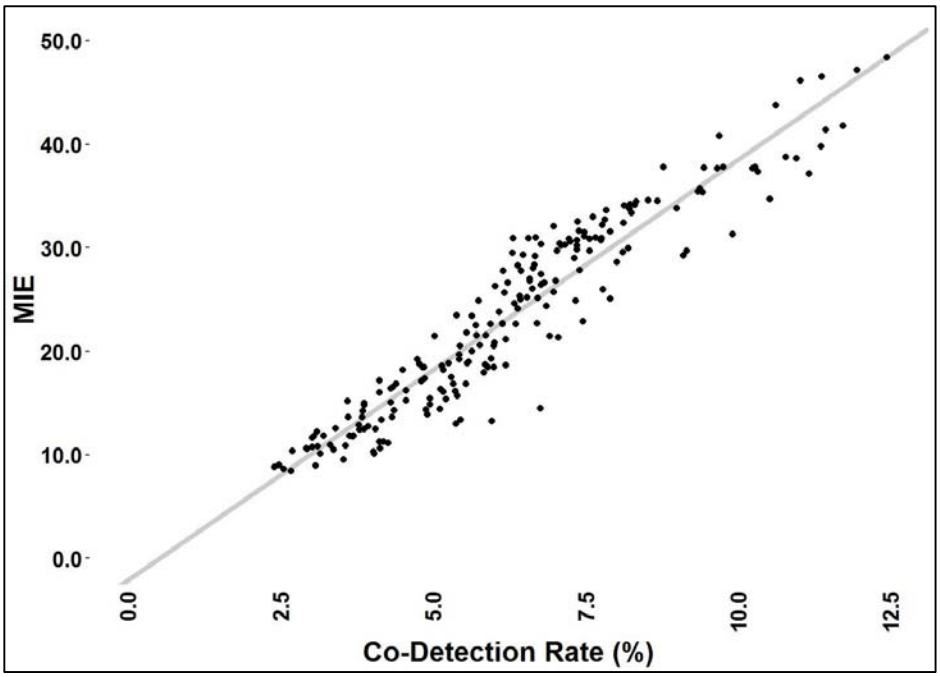
Figure 2-Figure Supplement 3: Organism detection by type and year



Percent detection rates from all participating Trend sites for each FilmArray RP organism are broken out by respiratory year. Organisms are ordered by total abundance across the four years.

bioRxiv preprint doi: https://doi.org/10.1101/157156; this version posted September 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5-Figure Supplement 1: Linear regression of MIE and observed co-detections



The time series data in figure 5B shown as a scatter plot. The equation of the linear regression is: MIE = 4.05153* Co-Detection rate -0.0206 with an R^2 value of 0.9003