

ESSENCE, the Electronic Surveillance System for the Early Notification of Community-Based Epidemics

Howard S. Burkom^{1*}, Wayne A. Loschen¹, Richard A. Wojcik¹, Rekha S. Holtry¹,
Monika A. Punjabi¹, Martina M. Siwek¹, Sheri H. Lewis¹

¹Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, United States of America

*Corresponding Author:

Howard.Burkom@jhuapl.edu

1 **ABSTRACT**

2 The Electronic Surveillance System for the Early Notification of Community-Based Epidemics
3 (ESSENCE) is a secure web-based tool that enables health care practitioners to monitor health
4 indicators of public health importance for detection and tracking of disease outbreaks,
5 consequences of severe weather, and other events of concern. The ESSENCE concept began in an
6 internally funded project at the Johns Hopkins University Applied Physics Laboratory (JHU/APL),
7 advanced with funding from the State of Maryland, and broadened in 1999 as a collaboration with
8 the Walter Reed Army Institute for Research. Versions of the system have been further developed
9 by JHU/APL in multiple military and civilian programs for timely detection and tracking of health
10 threats. Features of ESSENCE include spatial and temporal statistical alerting, custom querying,
11 user-defined alert notifications, geographical mapping, remote data capture, and event
12 communications. These features allow ESSENCE users to gather and organize the resulting wealth
13 of information into a coherent view of population health status and communicate findings among
14 users. The resulting broad utility, applicability and adaptability of this system led to adoption of
15 ESSENCE by the Centers for Disease Control and Prevention (CDC), numerous state and local
16 health departments, and the Department of Defense (DOD) both nationally and globally. With
17 emerging high-consequence communicable diseases and other health conditions, the continued
18 user-requirements-driven enhancements of ESSENCE demonstrate an adaptable disease
19 surveillance capability focused on the everyday needs of public health. The challenge of a live
20 system for widely distributed users with multiple different data sources and high throughput
21 requirements has driven an novel, evolving architecture design.

22
23 Key Words: biosurveillance; outbreak detection; population health

24 **INTRODUCTION AND BACKGROUND**

25 **Automated Routine Health Monitoring**

26 Until recent decades, public health disease surveillance relied upon laboratory confirmation and
27 passive participation. Often, the lack of automated detection and reporting resulted in time delays
28 that impeded prompt mitigation activities. Public health institutions thus began using enhanced
29 surveillance techniques with the potential for more timely epidemic detection and tracking. These
30 techniques were incorporated in electronic and increasingly internet-based biosurveillance systems
31 for everyday use by health monitors.

32 **Origin and Early Development**

33 Intensive efforts to establish biosurveillance systems occurred at multiple institutions and
34 government agencies in the late 1990s. The ESSENCE system originated as a collaboration of
35 two such projects. A project of Dr. Michael Lewis, then a medical resident under Dr. Julie Pavlin
36 at the Walter Reed Army Institute for Research, applied US military clinic visit data for outbreak
37 detection in a project called “ESSENCE”, Electronic Surveillance System for the Early
38 Notification of Community-based Epidemics (1). Concurrently, JHU/APL was combining civilian
39 data from hospital emergency departments, physician office visits, over-the-counter (OTC) sales,
40 and school absenteeism records, first with internal funding, then for the State of Maryland(2). The
41 two groups joined forces in anticipation of possible bioterrorist activity at the turn of the century
42 as January 1, 2000 approached, and this collaboration along with similar efforts at other
43 institutions led to further funded development in the multi-center Bio-Event Advanced Leading

44 Indicator Recognition Technology (BIOALIRT) program of 2001-3 for the Defense Advanced
45 Research Project Agency (DARPA) (3) . In leading one of four BIOALIRT research teams, the
46 JHU/APL and WRAIR groups further matured the ESSENCE system concept.

47
48 Following the terror attacks of September 11, 2001, ESSENCE was expanded in Maryland,
49 increasingly operationalized in state and local civilian health-monitoring agencies, and
50 implemented in all global US military treatment facilities. Since then, fueled by a succession
51 of initiatives driven by both bioterrorism and natural public health concerns, versions of
52 ESSENCE have been implemented in the Department of Defense (DoD), the Veterans
53 Administration (VA), the Centers for Disease Control and Prevention (CDC), and state and
54 regional public health agencies across the United States and, through collaborative DoD
55 efforts, internationally. Following widespread use of ESSENCE and enhancements at
56 JHU/APL to meet users' evolving needs, the CDC National Syndromic Surveillance
57 Program adopted ESSENCE in 2014 as the standard analytic surveillance and
58 visualization engine on the BioSense platform for state and local public health monitors
59 (4, 5). User benefits of ESSENCE have long surpassed the “Early Notification” part of
60 its acronym to include aspects of situational awareness such as tracking and
61 characterization of known health events, assessing the burden of all-hazards threats such
62 as severe weather, environmental hazards, and substance abuse, and rumor control to
63 enable improved public health response.

64

65 **ESSENCE History Highlights**

66 1997: Early research in hospital information management led to a request by Maryland's Secretary
67 of Health to work with the Maryland Department of Health and Mental Hygiene to develop an
68 automated system to detect bioterrorist attacks.

69
70 1999: The precursor to the current ESSENCE system, the Maryland Disease Surveillance and
71 Reporting System (MDSRS), was implemented to monitor for bioterrorism.

72 Subsequent collaboration with the Department of Defense Global Emerging Infection System
73 led to DARPA's multi-center BIOALIRT program enabling improvements in algorithmic
74 detection and biosurveillance system architectures.

75
76 2001: ESSENCE moved from a dynamic web framework to the static pre-built visualizations to
77 exploit commercial GIS technologies. Limitations on user capabilities forced a rapid return to
78 the dynamic web framework with a database backend for ESSENCE II.

79 Soon after the terrorist attacks of September 11, ESSENCE II was transformed from a limited-
80 use research project to a live operational civilian surveillance system in Maryland and for all
81 global US military treatment facilities.

82
83 2003: In response to the need to monitor the health of deployed US troops in the Iraq War,
84 ESSENCE III was developed and deployed in March 2003. This version featured novel data
85 sources, customizable syndrome groupings, and the first incorporation of external detection
86 algorithms. The ability to implement external analytic methods prompted and validated the
87 addition of an Application Program Interface (API) to the analytics architecture.

88

89 2003: Global emergence of the Severe Acute Respiratory Syndrome (SARS) spurred the need to
90 sharpen and narrow case definitions in place of general syndromes. This threat solidified the
91 decision to include free-text chief complaint queries along with diagnosis codes in the
92 aggregations of monitored counts and rates. Tracking of SARS highlighted the power of
93 automated systems for situational awareness, a broader and far more frequently useful
94 objective than novel event detection. Subsequent experience began the expansion of syndromic
95 surveillance from monitoring broad infectious syndrome groups to ad hoc categories of greater
96 clinical specificity and a broader variety of population health threats.

97

98 2004: Expansion: the DoD Joint Services Installation Pilot Project (JSIPP) and the Department of
99 Homeland Security (DHS) BioWatch program proliferated ESSENCE instances across the US.
100 The influx of new users and widespread user sentiment that shareable and explainable results
101 are no less challenging and important than event detection drove enhancements in visualization
102 flexibility and exportability.

103

104 2004: The states of Maryland and Virginia and the District of Columbia agreed to share data across
105 state lines, and the “Aggregated National Capital Region” (ANCR) version of ESSENCE was
106 created. The Enhanced Surveillance Operating Group (ESOG) of ANCR users was formed to
107 help guide ESSENCE development. The close tie between users and developers became a
108 foundation for future ESSENCE enhancements.

109

110 2006: Expanded usage of ESSENCE including patient data from hundreds of hospitals at the
111 Florida Department of Health led to additions to analytics including “time of arrival”
112 monitoring in which anomalous groups of emergency department (ED) arrivals at hourly
113 intervals were detected and reported with adjustments for patient volume, time of day, and
114 subsyndrome type.

115
116 2007: The high-profile mass gathering events of the Super Bowl and the 2009 Presidential
117 Inauguration impelled testing of information-sharing, distinct from data-sharing, strategies
118 including an InfoShare tool to enable timely sharing by ANCR users with national level
119 authorities. This effort foreshadowed multiple events or threats in which local restrictions
120 prevented ESSENCE user sites from sharing explicit data, but derived reports, aggregates,
121 data-free query language, or just descriptions could be legally shared. Implementation of
122 ESSENCE features to facilitate such sharing has continued since then.

123
124 2009: Collaboration of the ESSENCE team with DoD-GEIS clarified the need to transition the
125 technology and lessons learned from ESSENCE to an open source version that would be easy
126 to operate, cost-free, and deployable around the world. The Suite for Automated Global
127 Electronic bioSurveillance (SAGES) toolkit was developed, including OpenESSENCE and
128 ESSENCE Desktop editions. While based on the features and functionality of ESSENCE, tools
129 within SAGES were developed specifically for use in low- and middle-income countries.

130 Working with the Veterans Affairs and DoD, the ESSENCE team investigated the benefits and
131 obstacles of including elements of electronic medical records beyond the demographics and

132 chief complaints. Working with multi-terabyte database sites prepared ESSENCE developers
133 to deal with large datasets in every jurisdiction.

134
135 2013: The Boston Marathon bombings led to ESSENCE application to monitor non-infectious
136 health categories including Anxiety, Depression, Suicidal Tendencies and Hearing Loss.
137 These additions helped track the health of the Boston communities in the aftermath of the
138 attacks.

139
140 2014: CDC announced the transition of the National Syndromic Surveillance Program (NSSP) to
141 ESSENCE for syndromic systems analytics, to be hosted on the cloud-based BioSense
142 Platform.

143
144 2016: Non-syndromic alerting algorithms were added to ESSENCE for rare, nonmedical, or
145 uncommonly frequent terms in hospital free-text chief complaints.

146
147 2017: The opioid overdose crisis led to new, sustained partnerships between ESSENCE users and
148 other public health divisions, such as injury prevention, behavioral health, and drug abuse.
149 Analytic methods were explored to combine multiple data types for improved situational
150 awareness. The MyESSENCE feature was added to allow users to customize dashboards for
151 concise daily monitoring.

152
153 2018: The NSSP Syndrome Definition Committee users began sharing combined free-text and
154 diagnosis code-based definitions for inclusion in local ESSENCE sites. Advanced natural

155 language processing methods such as word embeddings were developed and shared among
156 CDC and local users

157
158 2019: Nationwide monitoring of outbreaks of lung injuries associated with the use of vaping
159 products increased demand for and development of system features to assess current data
160 quality, including availability and informative content of the records underlying ESSENCE
161 data displays.

162
163 2020: Collaboration among NSSP ESSENCE user groups intensified with evolution and sharing
164 of queries related to COVID-19 pandemic surveillance.

165 **SYSTEM DESCRIPTION AND PRINCIPAL FEATURES**

166 **System Overview**

167 ESSENCE is an enhanced health surveillance system using advanced mathematical/visual
168 analytics which can detect anomalies in both traditional and non-traditional public health data,
169 with the goal of enabling public health to find and monitor outbreaks of health events and make
170 decisions. Users of ESSENCE access a secure web-based tool to conduct disease surveillance for
171 the purpose of timely detection, situational awareness, and descriptive epidemiologic analysis of
172 baseline disease patterns and outbreaks. For effective public health response, public health
173 authorities must have the ability to identify the infected population so further spread can be
174 contained. Leveraging the near real-time availability of an increasing number of data sources,
175 ESSENCE analytical and alerting capabilities provide an opportunity for public health users to

176 capture the early stages of an outbreak and track its progress. ESSENCE enables integration of
177 electronic data from both clinical and nonclinical sources to enhance situational awareness. For
178 most users, the primary clinical data source is hospital emergency department chief complaint
179 records. Additionally, based on availability, public health agencies incorporate other data types
180 such as over-the-counter medication sales, poison control call center data, prescription drug data,
181 reportable disease data, vital statistics mortality data, and school absentee data. Once raw data
182 reach ESSENCE, the analysis, visualization, and communication features of ESSENCE allow the
183 end user to gather and organize the resulting wealth of information into a coherent view of
184 population health status and to communicate findings among other users and stakeholders.

185
186 As a result of ongoing user feedback daily and event-based disease surveillance needs, ESSENCE
187 features include spatial and temporal statistical alerting, custom querying, user-defined alert
188 notifications, geographical mapping, remote data capture, and event communications.

189 **Main Functions**

190 ESSENCE provides three main functions:

- 191 1) *Data ingestion/preprocessing.* Traditional and non-traditional data sources are
192 electronically received by ESSENCE and many are mapped to syndrome groupings.
193 During the ingestion process, data are cleansed (e.g. duplicate records removed, invalid
194 characters removed, etc.), updated (e.g. existing records information is updated with
195 new information) and categorized (e.g. syndrome and subsyndrome categorization
196 assigned to records).

197 2) *Alerting*. Multiple temporal and spatial alerting algorithms are applied to each data set
198 to develop a list of alerts or flags for further investigation by public health officials. In
199 addition to algorithms developed by JHU/APL and Walter Reed Army Institute of
200 Research (WRAIR), ESSENCE can incorporate algorithms required by the jurisdiction
201 where the system is deployed.

202 3) *Analysis and Visualization*, ESSENCE data and alerts can be analyzed and visualized
203 in multiple ways in the system, both spatially and temporally.

204 Section 4 describes these functions in three subsections: Data Management, Alerting Algorithms,
205 and Visualization.

206 **TECHNICAL DETAILS**

207 **Data Management**

208 **Architecture**

209 The software architecture employed for ESSENCE is a three-tier web application with a
210 presentation layer as a user frontend, a business layer for application of algorithms, and a backend
211 for databases. This architecture runs on modular server configurations, with the number of servers
212 contingent upon the data volume, number of active users, and frequency of required analysis
213 operations. The most common configurations comprise three servers for smaller instances and
214 five servers for larger ones. For systems with larger numbers of data sources and/or data volumes
215 reaching billions of records, the architecture can support additional servers to spread the functional
216 load of the processing. The backend databases are Microsoft SQL Server relational database
217 management systems (RDMS). Database functions include an ingestion database layer that

218 facilitates extract-transform-load (ETL) operations and performs deduplication and other data
219 cleaning operations. Depending on the user site, the ETL operations are performed by
220 Rhapsody,(6) Mirth,(7) or locally developed scripts to populate the database, and then ESSENCE's
221 Groovy-based data flow management system (8) controls data flow and business logic to transfer
222 data from ingestion to detection to web databases.

223 A detection database layer holds data and manages cube tables for fast algorithm access and
224 execution. Java-coded algorithms access the data and cubes for efficient signal detection on the
225 detection database to separate algorithm processing from user query management. The web
226 database layer expedites rapid formation and display of interactive screens for visualization and
227 communication. Web applications encoded in Java and JavaScript utilize this database via a
228 Tomcat web application server.(9) For display purposes, mapping and other geographic
229 information system (GIS) operations employ the open source tool GeoServer.(10) Users can
230 access the web application through standard web application displays or via a web service API
231 layer for direct access to ESSENCE data and functionality.

232 **Data Types**

233 The types of data analyzed in ESSENCE are the prerogative and responsibility of the jurisdiction,
234 though JHU/APL provides capability for basic types. The ESSENCE system is data agnostic—the
235 only requirement for a monitored data type is the inclusion of a data field. Data time resolution is
236 also unrestricted. Data frequencies in ESSENCE have ranged from seconds to years, though daily
237 data have been most common. All but a few users monitor hospital emergency department (ED)
238 data. Most ESSENCE user jurisdictions face the burden of acquiring their data sources, gaining
239 approval for their routine intended use, and extracting features to monitor. That level of effort

240 varies greatly by data source. Various users also monitor or have monitored records of the data
241 sources listed below.

- 242 • over-the-counter remedy sales
- 243 • physician office visits
- 244 • laboratory test orders and results
- 245 • school absenteeism and health office/nurse records
- 246 • reportable disease cases
- 247 • poison center call records
- 248 • ASPR DMAT data for disaster response (11)
- 249 • cardiovascular and other chronic diseases from inpatient encounters
- 250 • livestock and companion animal health encounters
- 251 • human vitals measurements
- 252 • death records
- 253 • emergency medical services (EMS) and 911 calls, and
- 254 • climate and air quality data

255 Users have also incorporated or adapted data from other surveillance systems such as the National
256 Poison Data System (NPDS).(12) While not all of these sources have proven useful for routine
257 surveillance, users have employed the system to investigate their utility and to seek best data usage
258 practices.

259 In separate studies or projects, individual user jurisdictions or their research partners have also
260 used ESSENCE to analyze records of radiology impressions, genomic sequencing data, zoo animal
261 health, environmental sensor outputs, sales of specific products such as thermometers, orange
262 juice, and tissues, social media posts and searches, and even fantasy sports data. Multiple content

263 formats used in ESSENCE have included Health Level-7 (HL7) formats for hospital data, National
264 Emergency Medical Services Information System (NEMESIS) formats for EMS data, and others.
265 File formats have included delimited/tabbed/fixed-width American Standard Code for Information
266 Interchange (ASCII) text, extensible markup language (XML) and JavaScript Object Notation
267 (JSON).

268

269 **Data Security**

270 All automated data transfers occur over a secure virtual private networks (VPNs), e.g. via secure
271 file transport protocol (SFTP); or over VPN tunnels, e.g. HL7; or by web transfer from a secure
272 website, e.g. NPDS and weather. The use of web application access control allows limited,
273 hierarchical access rights for every user by data source, by fields within a data source (e.g.
274 geography or jurisdiction, syndrome), and by website function (e.g. data details, time series,
275 alerts). These access privileges are customizable in the sense that one user site may control access
276 by health district, others by county or by neighborhood. Some sites build aggregated datasets that
277 meet the monitoring needs of users with limited access rights. Weather and other data sources with
278 no privacy concerns may be made available to all users.

279 **Data Preprocessing and Quality Management**

280 Multiple, data-dependent preprocessing steps include deduplication procedures, formation of
281 syndrome fields, calculation of distances, and deriving additional fields and flags based on
282 jurisdictional business rules and logic.

283

284 Procedures for managing data quality issues such as deduplication, temporary dropouts of data
285 feeds, data field value validation, and management of free-text or pick-list entries are incorporated
286 in ESSENCE standard business rules. By these rules, ESSENCE does not update records in place
287 but employs a delete-and-replace approach that has proven faster. For data with the functionality
288 enabled, a history system integrates all prior instances of each record to produce a single master
289 record with the relevant fields for each encounter. An extensive set of reference tables and
290 business logic allows conversion of field entries such as patient age, race, ethnicity, and vitals
291 measurement such as temperature to categorical values from standard sources such as the Public
292 Health Information Network Vocabulary Access and Distribution System (PHIN VADS). A
293 “region” data field is used for general spatial aggregation of patient records and is most often
294 employed to combine count data from collections of zip codes to approximate county-level counts
295 when the county field is unavailable. Beyond these features and conventions, ESSENCE includes
296 a substantial website section with guidance and analysis tools dedicated to helping users manage
297 the quality of their data.

298 **Alerting Methods using Statistics and Artificial Intelligence**

299 **Machine Learning and Natural Language Processing for Trackable Data** 300 **Features**

301 Data sources used or considered for health surveillance include medical encounter billing
302 records, emergency service calls, nurse hotline calls, prescription and over-the-counter remedy
303 sales, absenteeism records, and more recently, social media data such as tweets and web
304 searches. Each data source has its own challenges for user jurisdictions to obtain sustained
305 electronic access from data providers and any requisite government approval. When a data stream

306 of any of these sources is acquired for routine monitoring, an immediate question is how to use the
307 streaming data to track health outcomes of concern. An often applied procedure is to track counts
308 of subcategories of the data expected to correspond to these outcomes. These subcategories are
309 commonly called *syndromes*, generalizing the medical definition of this term denoting disease-
310 related collections of signs and symptoms. Thus, in the surveillance context a *syndrome* may refer
311 to grouped hospital visits associated with a fixed collection of symptoms, laboratory tests ordered
312 for certain conditions, web searches containing sets of terms, billing records covering any of a
313 class of remedies, or other subgroups depending on the data source. Syndrome formation is a
314 critical step that may use only a fraction of all streaming data and may produce few or many groups
315 to track. The number and composition of syndromes depends on the richness of the data, the
316 number of outcomes of interest, and the resources of the monitoring institution for investigation
317 and response.

318
319 Syndromes and subsyndromes used in ESSENCE vary depending on the clinical grouping systems
320 available and the needs of the user site. Early versions of ESSENCE formed syndrome groups
321 using diagnosis codes, which have disadvantages of late assignment and emphasis on billing
322 practice in many medical systems. Examples were Respiratory and Gastrointestinal with
323 subsyndromes such as Asthma. Categorization soon switched to the use of free-text chief
324 complaint (CC) or reason-for-visit data fields. For this categorization, the JHU/APL ESSENCE
325 team developed the Chief Complaint Processor (CCP), a versatile, stand-alone program for
326 weighted keyword-based classification by free-text fields. The CCP is highly configurable, with
327 tables including sets of syndromes and subsyndromes with classification rules allowing complex
328 logic, positive and negative weighting of component terms, abbreviation and spelling rules, and a

329 list of unmodifiable terms. For example, CCP puts a record with a CC of "NAUSEA" or
330 "VOMITING" in the Gastrointestinal (GI) category. The CCP creates a ChiefComplaintsParsed
331 field for use of classification rules after treatment of abbreviations, some misspellings, and other
332 cleanup.(13) These classifications have enabled additional natural language processing and
333 machine learning initiatives by both ESSENCE developers and users, and findings from these
334 initiatives are shared among users with each emerging health threat.(14-16)
335 As done with diagnosis code-based processing, syndrome groups are tabulated, plotted and
336 monitored each day with statistical alerting algorithms for early potential outbreak indications.

337 **Application Principles for Alerting Algorithms**

338 Individual alerting algorithms implemented in ESSENCE are listed and described in the Supportint
339 Materials file "S1 Technical_Details_of_ESSENCE_Alerting_Algorithms.docx".

340 The following principles were derived with users to guide method selection and to clarify
341 interpretation of results:

342

343 General considerations:

- 344 • These methods are not intended to positively identify outbreaks without supporting
345 evidence. Their purpose is to direct the attention of a limited monitoring staff with
346 increasingly complex data streams to data features that merit further investigation. They
347 have also been useful for corroboration of clinical suspicions, rumor control, tracking of
348 known or suspected outbreaks, monitoring of special events and health effects of severe
349 weather, and other locally important aspects of situational awareness. Successful users

350 value these methods more for the latter purposes and do not base public health responses
351 solely on algorithm alerts.

352

353 • These algorithms are one-sided tests that monitor only for unusually high counts, not low
354 ones. Low counts could result from a critical outbreak situation that prevents data reporting,
355 but there are many more common reasons for low counts (such as unscheduled closings or
356 system problems), so the algorithms do not test for abnormally low counts.

357

358 • In addition to data- and disease-specific considerations below, algorithm selection was also
359 driven by system considerations. Users need to monitor many types of data rapidly.
360 External covariates such as climate data or clinic schedules may not be available for prompt
361 analysis. Many methods in the literature, armed with substantial retrospective data of a
362 certain type, depend on analysis of substantial history. Day-to-day users, often with only a
363 small fraction of time available for monitoring, will not wait several minutes for each
364 query. In the absence of data history and data-specific analysis time for each stream,
365 ESSENCE methods have been adapted from the literature and engineered to system
366 requirements.

367

368 • If the time series monitored by algorithms represent many combinations of clinical
369 groupings, age groups, and geographic regions, excessive alerting may occur simply
370 because of the number of tests applied. The Summary Alert method was implemented to
371 limit such excessive alerting. This method is based on control of the false discovery rate,
372 i.e. the expected ratio of false alerts to the total alert count, and its statistical implementation

373 in ESSENCE is detailed in the Summary Alert section below. Aside from analytic methods
374 to control alerting, default alert lists should be limited to results from those time series of
375 concern to the user, either by system design or by active specification by the user. For
376 example, one method of reducing the default alert list is to restrict algorithms to all-age
377 time series groupings. Depending on the scope of the user's responsibility, the alert list
378 may also be restricted according to both epidemiological interest and the resources
379 available for investigation. For example, a monitor of a national-level system with
380 algorithms applied to many facilities may be interested only in alerts with at least 5-10
381 cases. In circumstances of heightened concern, these restrictions can be relaxed, or the user
382 can use ESSENCE advanced querying methods to apply algorithms to age groups and/or
383 subsyndromes.

384

385 **Visualization**

386 Both standard and user-customizable visualizations are available in ESSENCE. Highly
387 configurable and interactive modes of data stratification and filtering, graphical and tabular
388 customization, user preference management, and sharing features allow users to query data and
389 view geographic representations, time series and data details pages, and reports. The following
390 sections summarize key features. Screenshots of selected novel visualizations are shown below.

391 **Standard Visualizations**

392 Following are the most commonly used ESSENCE visualizations:

- 393 • The Time Series View provides a graphical display of the temporal behavior of the data
394 with the ability to stratify by specific parameters, view aggregated counts, and infuse data
395 quality factors to improve understanding of data features.
- 396 • The Data Details Page provides line listings of individual records and pie/bar chart
397 representations of query results.
- 398 • The Map View allows the user to view both data and alerts on a geographical display of
399 the specified region.
- 400 • The Alert List provides a view of signals generated by the alerting algorithms. Each row
401 of the list includes a link to the corresponding Time Series View or Data Details Page.

402

403 To facilitate selection of desired data by space, time, subpopulation, and clinical specificity, the
404 Query Portal provides the user with tools for detailed design and management of simple or
405 complex queries for routine or ad hoc monitoring. An online Query Wizard facilitates these
406 processes as in Fig 1. For the desired analysis, the user chooses the data source, date range, time
407 resolution (daily, weekly, monthly, quarterly, or yearly), temporal detection algorithm, and
408 selections from a variety of component filters depending on the data source. For each data source,
409 the user can define custom filters through configuration files that customize visualizations.
410 Options include Free-Text, Reference List, Number Range, and Dates filters. Examples of query
411 field options for filter creation are age group, geography system, syndrome or medical category of
412 interest.

413 Details of the attributes available for selection, stratification, and filtering are:

- 414 • **Geography System:** Data can be viewed in many geography systems such as Region, Zip
415 Code, Hospital, Region of the Hospital, Military Treatment Facility (MTF), School, or

416 Store. Each geography system defines a system for geographically filtering your data.

417 Regions are a generic term that defines the default geographic way to view data. Regions

418 normally map to a set of zip codes that closely resemble a county or health district.

419 • **Medical Grouping System:** When viewing the data via the Query Portal, the user has the

420 ability to choose between how the various data are presented. Data may be viewed by many

421 grouping systems, including ESSENCE Syndrome, International Classification of Diseases

422 (ICD) Code, Chief Complaint SubSyndrome, Chief Complaint, OTC Category for records

423 of sales of over-the-counter remedies, or Call Center Guideline, depending on which data

424 source is queried.

425 • **Syndrome Grouping:** The syndrome groupings used in the ESSENCE system vary

426 depending on the medical grouping systems used and the needs of the user site. These

427 groupings are used to filter data into medically similar sets. Examples of a syndrome

428 grouping are: Respiratory, Rash, Cough/Cold, Sinus, Asthma, Chest Pain, etc. Each

429 medical grouping system will have a set of syndromes or the ability to perform free text

430 queries. A Syndrome Definitions user interface provides a stepwise mechanism for viewing

431 rules that define a syndrome or subsyndrome. In addition to syndrome categories the user

432 may also query based on a Chief Complaint Discharge Diagnosis (CCDD) field, a

433 concatenation parsed chief complaint and discharge diagnosis. Filtering by CCDD

434 category uses Structured Query Language (SQL) “Where” clauses to select records

435 meeting user criteria. In general, this filtering takes the form of simple keyword matching

436 with inclusion of wildcard matching and negation terms.

437 • **Detector:** The ESSENCE tools refer to alerting algorithms as *detectors*. For temporal

438 alerting methods, users may choose any of the methods described in Section 4.2.1.

439 **Fig 1: The ESSENCE Query Wizard**

440 For queries involving more complex data selection and free-text logic than the query wizard
441 provides, ESSENCE provides an Advanced Query Tool, shown in Fig 2.

442

443 **Fig 2: The Advanced Query Tool in ESSENCE**

444

445 Once a user has created a query, an action button is available to indicate the use and disposition of
446 the query. Options include formation of time series and tables. Complex queries can be saved for
447 reuse and for application to other appropriate data sources.

448

449 **User-Customizable Visualizations**

450 In addition to the standard visualizations, ESSENCE offers additional views and analysis modes.
451 Of these, the most commonly used are myESSENCE and myAlerts.

452

453 myESSENCE: The myESSENCE feature on the main ESSENCE menu facilitates creation of
454 multiple customized dashboards of graphs, charts, tables, maps, and alerts. Users may create
455 separate dashboards by using a widget interface to select, drag, and drop widgets. They may share
456 each dashboard with other users by providing a copy to the dashboard or by providing a read-only
457 copy over which they maintain control. Possible choices for the widgets are time series graphs,
458 maps, and listings of data details. Parameters for each view, such as start and stop dates, are
459 modifiable. These views may be arranged in 1-, 2-, or 3-column format. Fig 3 exemplifies the
460 dashboard creation process.

461 The creating user gives each dashboard a text description and a note that is modifiable by creator
462 or shared user. Once created, each dashboard appears as a separate tab on the myESSENCE
463 webpage.

464 By default, each dashboard applies to data from the geographic regions selected for each graph
465 when added to the dashboard. The creator or sharing users may change the region, and ESSENCE
466 will change all views on the dashboard to reflect data from the new region. Users may revert to
467 the default view by choosing “Original Query”.

468

469 **Fig 3: Formation of Customized Dashboard Using MyESSENCE Feature**

470

471 myAlerts: The myAlerts feature enables users to set up automatic notifications of query results
472 reflecting specific subpopulations, algorithms, and conditions. Two types of these notifications
473 are available. Notifications triggered by receipt in ESSENCE of data records that satisfy a query
474 definition are *Records of Interest* alerts. Notifications triggered by minimum count requirements
475 or by algorithms detecting sets of records returned by a query are *Detector* alerts. Algorithms
476 specified may be any of the methods described in Section 4. The user may also specify a
477 minimum number of records requirement in addition to an alerting algorithm. Fig 4 shows the
478 web page used to create this customized alerting feature. Alerts that require algorithm results that
479 cross a designated statistical threshold may be restricted to require threshold crossings on m
480 consecutive days, or on m of the last n consecutive days. The user may set up automatic email
481 prompts for each customized alert and may choose automated sharing of selected alerts with
482 other designated ESSENCE users.

483 Developers of ESSENCE have also created single-purpose views and tools for specific
484 jurisdiction needs, including statistics tables and side-by-side graphs, as broadly applicable as
485 possible to all data sources.

486

487 **Fig 4: Creation of MyAlert for Customized Anomaly Detection**

488 **USER APPLICATIONS AND SUCCESSES**

489 This section describes ESSENCE applications in health departments for analysis of a wide range
490 of public health threats in monitored populations.

491 **Infectious disease applications**

492 Early applications of ESSENCE focused on detection of infectious disease outbreaks, with much
493 attention on influenza-like illness (ILI) because ILI is a prodrome for multiple naturally occurring
494 diseases and for many potentially weaponized for bioterrorism (35, 36). Civilian (37), military
495 (36), and VA (38) ESSENCE users monitored for outbreaks of seasonal and nonseasonal outbreaks
496 of febrile respiratory infections, GI infections caused by contaminated food or water, and rarer
497 infections.

498 Surveillance systems were particularly helpful to track the pandemic of novel H1N1 influenza in
499 2009 (38). The pandemic was important as an interregional use of surveillance systems to track a
500 common threat. For example, the National Capital Region (NCR) Disease Surveillance Network,
501 comprising ESSENCE users at health departments in the District of Columbia and parts of the
502 states of Maryland and Virginia, shares population-level disease incidence information to promote
503 inter-jurisdictional surveillance. In 2009, this network allowed NCR public health practitioners to

504 track the course of the pandemic from the spring through the fall, comparing the overall and age-
505 specific burden of illness to national and neighboring state trends. A broader collaboration of both
506 ESSENCE and non-ESSENCE system users collaborated on a standardized definition of ILI to
507 enable uniform local tracking of the pandemic across the US.(39)

508 Users have applied ESSENCE to form and share queries for indications of other infectious diseases
509 using text from chief complaints, discharge diagnosis, and triage notes when available. Infectious
510 threats tracked in published examples include general waterborne diseases (40), tuberculosis (16),
511 rabies (41), and Middle East Respiratory Syndrome (14). The recent years' concerns over
512 mosquito-borne diseases has also generated new queries by multiple users and occasionally
513 uncovered important cases.(15)

514
515 The health department of Maricopa County, Arizona presented an example of the benefit of
516 surveillance systems in June, 2018. The department had added an ESSENCE query for signs of
517 Rocky Mountain spotted fever, which is not endemic to that county, because of concerns that cases
518 transferred from endemic areas might be missed. A child's patient record was signaled by the
519 query, and the department contacted the hospital. This contact led to reversal of a medication
520 decision that might have been fatal to the child.(42)

521 Such monitoring activity has repeatedly uncovered unreported cases of diseases for which
522 reporting is mandatory. (43) These findings illustrate the importance of redundancy with systems
523 such as ESSENCE to avoid missing important cases, even when traditional reporting mandates
524 exist.

525 At submission of this manuscript, intense collaboration of ESSENCE users is focused on tracking
526 of the COVID-19 pandemic. General queries on COVID-like illness and specific ones involving
527 pneumonia and specific symptom sets are being refined and shared.(44)

528 **Applications for tracking burden of severe weather, natural disaster** 529 **events**

530 Health departments have used ESSENCE for preparedness, health burden assessment, and
531 response to severe storms and other natural disasters. The state of Oregon conducted a successful
532 program to mitigate the effects of wildfires.(45) This program featured customized ESSENCE
533 queries with other coordinated efforts among state and local health departments and preparedness
534 teams. Among several states that using ESSENCE to monitor the effects of hurricanes, the
535 Tennessee health department devised queries to determine the volume and clustering of patients
536 in local hospitals because of storms in other states.(46) Effective monitoring of some events
537 requires combination of multiple data sources. Following a prolonged storm-related power outage,
538 the health department of Seattle-King County, Washington combined data from ESSENCE
539 emergency department data with ambulance call and public utility data to monitor for cases and
540 clusters of carbon monoxide (CO) poisoning and food poisoning.(47) More recently, the Florida
541 health department monitored for CO poisoning after Hurricane Irma in 2017.(48) Institutions
542 using ESSENCE are increasingly incorporating environmental and other data sources in their
543 systems for richer situational awareness of disaster-related health threats. (49, 50)

544

545 **Applications for mass gathering surveillance**

546 Scheduled mass gathering events such as political conventions and major athletic competitions
547 concern population health monitors because a) such events are bioterrorism opportunities to affect
548 many victims and gain media attention, b) infections through contaminated food or water could
549 spread rapidly through the expanded population, c) those visiting for several days could import
550 infections or take them back to their own cities, and d) a surge of patients could overwhelm local
551 care provider resources. Adequate preparedness and response require coordination across
552 jurisdiction boundaries, but privacy laws often restrict patient-level data-sharing. ESSENCE
553 syndrome definitions and queries have been customized for many such events. In 2007, Marion
554 County, Indiana and Cook County, Illinois were home counties for the competing teams in Super
555 Bowl XLI, and the game was hosted in Miami-Dade, Florida. Health departments of these
556 geographically distant counties were ESSENCE users, and customization of their systems for the
557 days surrounding the event helped coordinate surveillance despite only two weeks' notice after
558 teams were determined.(51)

559 A partnership of the Florida Department of Health and the U.S. HHS Office of the Assistant
560 Secretary for Preparedness and Response (ASPR) to improve response by Disaster Medical
561 Assistance Teams (DMAT) produced a new ESSENCE module that was deployed for health
562 monitoring of the 2012 Republican National Convention in Tampa.(50) State and county health
563 departments used ESSENCE for coordinated monitoring of crowds at the US Olympic Trials in
564 July 2016.(52) In January 2017, the Washington DC Department of Health used ESSENCE
565 queries along with other data sources for health surveillance at the 58th US presidential
566 inauguration.(53)

567 For monitoring events outside the US, JHU/APL used the SAGES system in 2014 to monitor the
568 8th Micronesian Games held in Pohnpei Federated States of Micronesia and the 3rd International
569 Conference on Small Island Developing States in Apia, Samoa.(54)

570 **Applications for chronic disease and mental health surveillance**

571 Usage of ESSENCE to monitor risk factors and incidence of chronic disease and mental health
572 disorders has proliferated since a DoD ESSENCE study utilizing clinic and prescription data to
573 monitor behavioral health in 2004. (55) The Boulder County, Colorado Health Department
574 recently implemented and tested multiple queries to monitor mental health.(56)

575 Addressing ESSENCE utility for chronic diseases in general, the Cook County Illinois health
576 department applied machine learning methods to assess the utility of ESSENCE emergency
577 department (ED) data for monitoring cardiovascular disease (CVD), acute myocardial infarction
578 (AMI), acute coronary syndrome, stable angina, stroke, diabetes, hypertension, asthma, and
579 chronic obstructive pulmonary disease. From correlational validation testing based on eight full
580 years of chief complaint text and electronic medical record data, they concluded that ESSENCE
581 data are suitable for monitoring all of these conditions except stable angina and hypertension “at
582 local, state, or national levels”.(57) The Nebraska state health department has used ESSENCE to
583 monitor for CVD for several years, and the Florida state department similarly monitors AMI
584 incidence.(58)

585 **Applications for injury and substance abuse surveillance**

586 An unexpected but arguably the most helpful benefit of ESSENCE to health department users has
587 been to facilitate communication and collaboration among agency divisions. An important

588 example in the context of the ongoing opioid overdose crisis has been the strengthening of
589 connections between syndromic surveillance specialists and groups specializing in injury
590 prevention, behavioral health, and drug abuse.

591 Multiple health departments have applied ESSENCE to gain awareness of locations and
592 subpopulations at risk for injuries from falls.(59) The Boston Public Health Commission in
593 Massachusetts used it to monitor for hearing loss, acute depression, and explosion-related injuries
594 in the aftermath of the 2013 Boston Marathon bombing and subsequent manhunt. The St. Louis,
595 Missouri health department established ESSENCE queries for injuries indicative of bomb-making
596 activities.(60)

597 Monitoring for substance abuse has been common among ESSENCE users. The Tri-County
598 Health Department in Colorado uses its system to seek adverse effects of marijuana use.(61) The
599 Florida state department queries for ED visits resulting from synthetic marijuana(62) and for novel
600 street drugs such as Flakka as they become known public health problems.(63) Most recently, the
601 opioid crisis has stimulated intense collaboration including shared syndrome definitions and
602 analytic case-finding tools among geographically scattered institutions using ESSENCE.(64-66)

603 **CONCLUSIONS**

604 **Lessons Learned**

605 Several principles have driven the success of ESSENCE since its origins in the late 1990's.

606 Versatility: Users have valued the configurability and adaptability of ESSENCE. Default
607 categorization of complex data into syndromic groupings has always been valuable to users who
608 are inexperienced or who do not have the time to formulate or validate their own categories for
609 monitoring. Conversely, health departments with more analysis capacity have long demanded
610 surveillance systems that let them create their own categories to track, and ESSENCE
611 customization with query-building features using both diagnosis codes and free-text has grown
612 along with the sophistication and broadening needs of health department users. Pre-computed,
613 canned analysis products are not found in ESSENCE. However, versatility presents challenges to
614 database design and to the selection and adaptation of statistical analysis tools. Surveillance data
615 evolve with institutional information systems and formats, coding practices, and epidemiological
616 concerns. Users typically cannot wait several minutes for data retrieval and time-consuming model
617 runs. Alerting algorithms applied prospectively to detect disparate events in a wide variety of data
618 types cannot match the detection performance of models developed retrospectively using historical
619 datasets labelled with target events for a particular syndrome. Algorithm baselines in ESSENCE
620 do not reach back for years, not only for storage and computational reasons, but also because for
621 many users' desired data types, stable data or any data are available only within the past year.
622 Hence, ESSENCE alerting algorithms, adapted from published applications of models and control
623 charts in healthcare settings, (23, 67, 68) use rolling baselines of weeks rather than years.

624

625 Facilitating communication: Multiple ESSENCE users have remarked that one of the system's
626 main benefits has been to facilitate communication with other divisions within a health department,
627 with external local and federal agencies, and with care facilities that provide data and can benefit
628 from the broader geographic perspective that a surveillance system enables. Hence, substantial
629 ESSENCE development has occurred in response to user requests for custom analysis
630 comparisons, visualizations, and report formats, allowing overburdened users to concentrate on
631 the task of routine health monitoring. In situations where data-sharing is precluded by county or
632 state regulations, ESSENCE communication tools have enabled information-sharing.

633

634 Multiple analysis modes: The applicability of individual analysis modes such as univariate and
635 multivariate time series monitoring, spatiotemporal cluster detection, and single case identification
636 all depend on the nature and quality of available data. For example, the spatial scan statistic
637 implemented in ESSENCE can avoid issues of jurisdictional boundaries, but only if data location
638 fields are present and reliably used in the data. In many data sources, the limitation of location
639 fields to zip codes or postal codes restricts the geographic precision of clusters of interest. Health
640 monitors generally need multiple ways to analyze population health data. The clearest example is
641 that ESSENCE users in multiple health departments have discovered unreported cases of
642 reportable disease that traditional sentinel surveillance is expected to communicate to public
643 health. The various analysis modes of ESSENCE provide affordable and sometimes beneficial
644 analytic redundancy.

645 **Innovation**

646 The development of ESSENCE has produced novel features in areas of complex, disparate data
647 management, analytical methods, and the enhancement of user reporting and collaboration,
648 interrelated efforts to empower health monitors. The data management advances include
649 architecture and data transfer capabilities to meet needs of institutions with varying resources.
650 Analytics advances have required data quality examination methods and alerting algorithms
651 appropriate for diverse data time series that meet rapid response needs and do not require more
652 than a few months of data history. User experience enhancements have included customizable
653 visualization and reporting features that provide unique time- and resource-saving capabilities.

654 **Impact**

655 Recent ESSENCE projects have produced a variety of user capability enhancements. For sharing
656 of information within and across jurisdictions, online features allow users to share what they are
657 doing within ESSENCE with peers and to see what others are querying and find interesting. Text
658 analysis and visualizations facilitate creation of ad hoc local free-text queries. These features
659 provide correlation, trend, and association analytics to help the user determine what terms/phrases
660 queries should or should not include.

661 Back-end tools and checks with visualizations allow the user to monitor closely the local
662 ESSENCE system for data issues and irregularities. These administrative capabilities help
663 managers and users maintain day-to-day system availability and improve visibility of issues that
664 may develop over time.

665 Recently added visualizations and cohort clustering analytic tools for longitudinal assessment
666 allow users to determine categories of patients who use healthcare systems that provide data to

667 ESSENCE. These tools can show patient-level usage trends to inform allocation of healthcare
668 resources in a community.

669 The ongoing adaptation of ESSENCE to meet the needs of the understaffed public health practice
670 community has provided a means to share methods and information, though data are often not
671 shareable. The common analytic platform has enabled an evolving user ecosystem of multiple
672 working groups, and the US CDC currently hosts the NSSP Community of Practice, with
673 subgroups including Syndrome Definition, Data Quality, and Technical committees.(69) The
674 Syndrome Definitions committee promotes the analysis of common queries among geographically
675 scattered user sites and US CDC,(70) thus improving communication between local and national
676 health monitors. A primary example is the opioid overdose crisis, a noninfectious threat. In
677 addition to analytics and visualizations to support the activities related to this crisis, developers
678 and users have worked together to acquire additional data sources such as emergency medical
679 incidents, poison center calls, and death records to determine the benefit of fusing information
680 varying in specificity and timeliness into a common surveillance picture to better inform awareness
681 and interventions.

682

683 **Future Directions: Challenges to ESSENCE and the Surveillance**

684 **Community**

685 Challenges to public health surveillance span multiple disciplines ranging from epidemiology and
686 applied microbiology to statistics and machine learning and to database and network
687 technology.(71) In ESSENCE, data processing/storage and analytics challenges go hand-in-hand.
688 A major data challenge is to integrate increasingly diverse and granular data sources while

689 preserving the essential features of ad hoc user syndrome/case definition and custom real-time
690 analysis and visualization. Public health staff need to extend the power of their systems to all-
691 hazards public health threats and to One Health issues including zoonotic diseases across species,
692 antimicrobial stewardship, and food safety. Users of ESSENCE have long imported non-clinical
693 and non-syndromic data sources such as pharmacy sales and school absentee rates along with
694 traditional medical encounter records.

695 In expansion to address all-hazards threats, data complexity, categorization, and linkage challenges
696 will multiply as genomic, environmental (including remote sensing) and social media data sources
697 are added.

698 Significant advances in disease surveillance will also require meeting key analytics challenges.
699 Beyond the monitoring of individual sources and syndromes with data dashboards, combining
700 disparate data types requires statistical and machine learning tools for appropriate weighting and
701 corroboration of evidence from disparate data types. Analytic fusion efforts employing Bayesian
702 networks and other machine learning tools have been applied in both military and civilian
703 ESSENCE systems (72-75). More efficient and explainable fusion methods will be needed to
704 enable operationalizable forecasting and prediction for greater decision-making power and
705 effective planning and response. Efficient analytic methods will also be needed to determine
706 optimal feature extraction and resultant surveillance value of social media and other nontraditional
707 sources.

708 Lastly, best practices for biosurveillance systems face both human and electronic communication
709 challenges, including interoperability with other electronic systems such as the above-mentioned
710 ASPR DMAT, NPDS, and NEMESIS systems. Despite the preeminence of ESSENCE, as indicated
711 by its adoption as the analytic engine of the CDC Biosense platform and the widespread application

712 discussed in this paper, the limitations described above and the data gaps exposed by the COVID-
713 19 pandemic must be addressed as new threats emerge and data sources proliferate. These
714 challenges must be addressed as ESSENCE matures to expand public health surveillance
715 capabilities and utility beyond what is imaginable today and propel public health toward new
716 frontiers of understanding and action.

717 **ACKNOWLEDGEMENTS**

718 We would like to acknowledge invaluable assistance from our colleagues and partners. The
719 continued ESSENCE research and development would not have been possible without the support
720 of the CDC, DoD, VA, numerous State and Local health departments, and JHU/APL over the last
721 20 years. We especially acknowledge numerous champions and advisors among ESSENCE users
722 who have dedicated time and effort to the betterment of public health surveillance.

723

724 **Technical Details of ESSENCE Alerting Algorithms**

725 **Univariate temporal alerting algorithms**

726 The default temporal algorithm in ESSENCE is an automated selection between data modeling
727 and control-chart-based algorithms, resorting to a simplistic Poisson distribution-based method if
728 only a few days of recent data are available. The modeling method is adaptive multiple regression,
729 while the control chart-like approach is an adaptive exponentially weighted moving average
730 (EWMA) algorithm. The primary regression and EWMA methods are first discussed separately.
731 Each description below gives a method category, purposes of the method, a brief technical
732 description, key benefits, limitations, and literature sources.

733

734 **Algorithm: Linear Regression**

735 Categorization: Adaptive Multiple Regression Model

736 Purposes: This model is an adaptive regression model applied to remove the systematic behavior
737 often seen in time series of daily, syndromic, clinical visit counts and in other surveillance
738 data. The reason for removing these common effects is to avoid bias in identifying unusual
739 behavior. For example, there is a customary jump in visits on Mondays because many clinics
740 resume normal hours, and this expected jump should not automatically increase the possibility
741 of an alarm(17, 18). Similarly, alarms should be possible on weekends even though visit counts
742 drop off from weekday levels.

743 Technical Details: This adaptive, multiple, least-squares regression algorithm contains terms to
744 account for linear trends, day-of-week effects, and holidays. Multipliers for these terms are
745 calculated using 4 weeks of recent counts as a training period. This training period is separated
746 from the date of the test data by a 2-day buffer intended to keep early outbreak effects from
747 contaminating the training. Extreme data values in the training period are reduced to reasonable
748 values in order to avoid exaggerated predictions. This outlier correction for model inference
749 avoids loss of sensitivity in the weeks after either data problems or true outbreaks. The
750 regression multipliers are recomputed each day for calculation of a predicted count based on
751 the expected data trends. The algorithm then subtracts this prediction from the observed visit
752 count, scales the excess by the standard error of regression, and applies a statistical hypothesis
753 test to determine whether to signal an alert. The test is a Student's t distribution at significance
754 levels of 1% for red alerts and 5% for yellow alerts, with the number of degrees of freedom
755 determined by the number of regression covariates and the baseline length(19). Covariates and
756 training intervals were chosen to obtain maximum sensitivity for detection of injected signals at
757 manageable background alert rates.

758 Benefits: The main benefit is avoiding alerting bias resulting from expected data trends. The length
759 for the training baseline is critical. Based on performance comparisons among multiple
760 baseline lengths, it was chosen to be short and recent enough to capture seasonal time series
761 behavior but long enough to smooth out daily fluctuations. Separate multipliers are updated so
762 that a data source with regular but unusual patterns such as high weekend counts will be
763 modeled correctly. While a better fit may often be obtained with a more complex model for a
764 given data stream with a certain syndromic filter for a certain subregion and analysis of

765 sufficient data history, the current regression approach is relatively robust across time series
766 employed in ESSENCE.

767 Limitations: If this algorithm is applied to a data series without the baseline weekly and seasonal
768 behavior, the model will not explain the data well, and the detection sensitivity and specificity
769 will be decreased. The automated switch in the default method is applied for this reason. There
770 is no claim of optimal modeling for a given time series. This general-use implementation does
771 not assume the availability of a denominator variable such as the total visit count that can be
772 used to adjust the counts to emulate series of rates rather than counts. This adjustment has been
773 implemented in particular versions of ESSENCE and in past versions of Biosense(20, 21) and
774 could be added as an option.

775 **Algorithm: Adaptive Exponentially Weighted Moving Average (EWMA)**

776 Categorization: Adaptive Control Chart

777 Purposes: This algorithm is appropriate for daily counts that do not have the characteristic features
778 modeled in the regression algorithm. It is more applicable for Emergency Department data
779 from certain hospital groups and for time series with small counts (daily average below 10)
780 because of the limited case definition or chosen geographic region.(22)

781 Technical Details: This algorithm compares a weighted average of the most recent visit counts to
782 a baseline expectation. For the weighted average to be tested, an exponential weighting gives
783 the most influence to the most recent observations. Two weightings are applied: the first gives
784 negligible weight to observations over 3 days old and is designed to detect sudden events where
785 most outbreak cases affect data within a few days. The second weighting distributes influence
786 further over the past week for sensitivity to more gradual outbreaks. These weightings emulate
787 a dual strategy published for the hospital setting.(23) The monitored weighted averages are the
788 S_k given by:

$$789 \quad S_k = \omega S_{k-1} + (1 - \omega) X_k,$$

790 for a constant smoothing coefficient ω with $0 < \omega < 1$ and X_k as the successive data counts,
791 with $X_0 = 0$ and $S_0 =$ half the alerting threshold for prompt sensitivity. (Occasionally a useful
792 starting value for X_0 is known, but restarts may occur for many reasons, so the conservative
793 initialization to 0 is used.) For separate monitoring of sudden and gradual events, smoothing
794 coefficients $\omega = 0.9$ and 0.4 are used. For both weighted averages, the 4-week baseline mean

795 is subtracted, with a 2-day buffer period to separate the baseline from the counts being tested.
796 The rationale for the baseline length was the same as described above for the regression method
797 above. The test statistic is then $(S_k - \mu_k) / \sigma_k$, where μ_k and σ_k are baseline mean and standard
798 deviation. As in the regression method, the hypothesis applied to determine alerting is a
799 Student's t distribution at significance levels of 1% for red alerts and 5% for yellow alerts. The
800 number of degrees of freedom assumed for this distribution is the baseline length + 1. This
801 EWMA implementation is designed for any series that does not fit the characteristic trends, so
802 a couple of safeguards are included. A “zero-filtration” algorithm is implemented for rapid
803 adjustment to and recovery from data dropouts and catch-ups. When counts are sparse but not
804 uniformly zero, a Poisson-based adjustment is added to the standard deviation scale factor to
805 avoid excessive alerts.(19)

806 **Benefits:** This method gives sensitivity to both sudden and gradual outbreaks and has demonstrated
807 prompt alerting capability. It is less susceptible than the Early Aberration Reporting System
808 (EARS) methods C1, C2, and C3 to trends and to day-of-week effects. The added recovery
809 features handle the most common problems in the data acquisition chain. Alerting is indirectly
810 adjusted for the data distribution via the standardized residual test statistic, which provides a
811 safeguard against excessive alerting when counts are small.

812 **Limitations:** This algorithm applied to pure daily counts does not control for expected trends or
813 cyclic effects as in the regression method.

814 **Algorithm: Poisson/Regression/EWMA (default)**

815 **Categorization:** Automated switch between data model and control chart

816 **Purposes:** Many researchers and developers have applied complex statistical models to
817 surveillance data for prediction and detection. However, the predictive capability of a model
818 varies according to the specific data stream and how it is filtered and aggregated. This
819 capability may also be affected by data behavior changes that result from seasonal variations,
820 population shifts, and changes in the informatics. To account for such day-to-day changes,
821 ESSENCE automatically monitors its predictive capability of its regression model each day.
822 When this test fails, indicating that the model is not helpful for explaining the data, the system
823 switches to the EWMA adaptation described above. The result is that the regression model is
824 usually applied for the common respiratory and gastrointestinal syndrome classifications

825 applied to county-level data, but EWMA is more commonly applied to rare syndrome data.
826 For situations where less than a week of recent baseline data exists, a simple Poisson detector
827 is applied. Such situations include new start-ups and more common restarts after long (several-
828 week) intervals of missing data.

829
830 Technical Details: Details for the separate regression and EWMA methods are given in the
831 preceding pages. The adjusted R^2 coefficient for the regression is tested each day. This
832 coefficient does not give the quality of regression but is employed here specifically as a
833 measure of daily predictive capability using an empirically derived threshold criterion. When
834 the data pass this test, the model is assumed to have explanatory value, and the regression
835 algorithm is applied. When the data fail this test, the EWMA algorithm is used. The Poisson
836 distribution test is applied when less than a week (3-6 days) of recent data is available. A
837 Poisson distribution is assumed with mean and variance equal to the mean of the recent counts.
838 An alert is issued if the current count exceeds this mean and if probability that the current count
839 was drawn from this distribution is less than 1% (red alert) or 5% (yellow alert). Practical
840 safeguards for the composite method are as described in the regression and EWMA sections
841 above.(19)

842 Benefits: This algorithm is the default because it is designed to avoid mismatching the method to
843 the data. The regression model accounts for the expected data trends when they are seen in the
844 baseline. When they are absent because of the case definition used to filter the data, because
845 of the size of the monitored region, or because of data problems, alerting is based on the
846 EWMA algorithm.

847 Limitations: The goodness-of-fit test occasionally misclassifies the data. The test is set to err
848 toward the more conservative EWMA to avoid misfitting the data model.

849 ***Algorithm: C1, C2, and C3***

850 Categorization: Adaptive Control Chart

851 Purposes: To purpose is to detect general data aberrations. Algorithms C1, C2, and C3 of the Early
852 Aberration Reporting System (EARS) developed at the Centers for Disease Control and
853 Prevention are used in many U.S. states and in numerous foreign countries.(24) They are

854 included in the ESSENCE suite because of their wide application. While they lack many of the
855 features described above, their simplicity has both benefits and limitations.

856 Technical Details: The C1 algorithm subtracts the daily count from the mean of a moving baseline
857 ending the previous day. In effect, it then divides this difference by the standard deviation of
858 counts in that baseline. If the result exceeds 3, indicating an increase above the mean of more
859 than 3 standard deviations, an alert is issued. The C2 algorithm does the same calculation but
860 imposes a 2-day buffer between the test day and the baseline. The C3 algorithm is a more
861 sensitive version of C2 that adds the values from the 2 previous days if they do not exceed the
862 threshold. All three algorithms use the same criterion of an increase of at least 3 baseline
863 standard deviations above the sliding baseline mean. An important implementation detail is
864 that ESSENCE does not use the standard 7-day baseline because substantial experience has
865 shown that for many time series, such a short baseline gives an unstable statistic that can lead
866 to a loss of confidence in the results. The implemented baseline is 28 days as in the EWMA
867 and regression methods.(21) There are no other changes to the standard EARS methods,
868 including retention of the flat 3-standard-deviation threshold regardless of the data stream.

869 Benefits: The methods are easy to understand and widely known.(25-27)

870 Limitations: Like the EWMA, the methods take no account of systematic data behavior such as
871 day-of-week effects or seasonal trends. C3 is the only one of these methods with sensitivity to
872 gradual outbreak effects, but it is known to produce high alarm rates. For all three methods,
873 threshold data values for alerting may fluctuate noticeably from day to day.

874 **Summary Alerts--adjustment for multiple testing**

875 Categorization: False Discovery Rate processing of multiple alerts

876 Purpose: The *parallel monitoring problem* is the monitoring of multiple separate time series
877 representing different physical locations, such as counties or treatment facilities, possibly
878 stratified by other covariates such as syndrome type or age group. The purpose of the Summary
879 Alert Algorithm is to maintain sensitivity while limiting the number of alerts that arise from
880 testing the numerous resulting time series.

881 Multiple testing can lead to uncontrolled alert rates as the number of data streams increases.
882 For example, suppose that a hypothesis test is conducted on a time series of daily diagnoses of
883 influenza-like illness. In a one-sided test, this test results in a statistic whose value in some

884 distribution yields a probability p that the current count is as large as observed. For a desired
885 Type I error probability of α , the probability is then $(1 - \alpha)$ that an alert will not occur in the
886 distribution assumed for background data. Thus, for the parallel monitoring problem of interest
887 here, if such tests are applied to N independent data streams, the probability that no background
888 alerts occur is $(1 - \alpha)^N$, which decreases quickly for practical error rates α . For a single-test
889 error rate of $\alpha = 0.05$, for example, the probability of at least one background alert exceeds 0.5
890 if more than 13 independent tests are applied.

891 Technical Details: For N tests, where N is the number of combinations of region, syndrome, age
892 group, and any other covariates affecting the number of tests, let $P_{(1)}, \dots, P_{(N)}$ be the p -values
893 sorted in ascending order, an ordering that puts the smallest and most significant p -value first.
894 The Summary Alert method applies the Simes-Seeger-Eklund criterion to reject the combined
895 null hypothesis of no anomaly for any series.(28) The null hypothesis is rejected if for some
896 j^* , $j^* = 1, \dots, N$, $P_{(j^*)} < j^* \alpha / N$. To interpret this condition, note that for the most significant
897 p -value, an alert requires that $P_{(1)} < \alpha/N$, the strict Bonferroni bound. If $\alpha=0.01$ and $N=50$,
898 then the condition becomes $P_{(1)} < 0.0002$. For the least significant p -value, the condition is
899 simply $P_{(N)} < \alpha$, highly unlikely for the weakest result.

900 If this condition is satisfied for any j^* , then test results are considered alerts for all $j < j^*$.(29)
901 The Summary Alert is implemented at two levels, FDR and FDR-Major. For the FDR level
902 applied to N time series, the implementation is as above. For a more liberal option appropriate
903 for certain syndromes or scenarios, FDR-Major applies the condition to two sets of $N/2$ time
904 series.

905
906 Benefits: In defining the false discovery rate as the expected ratio of false alerts to the total alert
907 count, Benjamini and Hochberg showed that the Simes-Seeger-Eklund criterion gives an
908 overall error rate of α if the N time series tested are statistically independent.(30) Overall, this
909 criterion avoids the excess alerting resulting from using the nominal threshold α for all data
910 streams and also avoids the loss of sensitivity from using only the Bonferroni bound α/N .

911 Limitations: If one of the p -values crosses the adjusted threshold, it is not obvious for
912 epidemiological or other reasons which tests to consider anomalous. Most users have followed
913 the natural procedure described by Simes to consider all p -values less than $P_{(j^*)}$ as individual
914 alerts. Another limitation is that in general the time series are not statistically independent.

915 For situations where dependence is known, Hommel recommended the condition $P_{(j)} < j \cdot$
916 $i / C \cdot N$, where $C = \sum 1/j$. In ESSENCE applications where many groups of time series may
917 be requested and dependence can change, the above condition with $C=1$ is applied.
918

919 **Spatial cluster determination**

920 Categorization: Spatial Scan Statistics

921 Purpose: A problem with sophisticated temporal detectors is choosing the appropriate size and
922 location of the collection region for time series counts. If this region is too small or mislocated,
923 cases may be missed and the baseline data may not have enough structure, but if the region is
924 too large, the scale and variability of the large-scale time series may reduce sensitivity by
925 masking clusters of interest. We apply spatiotemporal scan statistics in an attempt to promptly
926 localize public health problems. For ESSENCE, JHU/APL built and implemented a Java
927 version of the algorithm implemented in the SaTScan software of Martin Kulldorff originally
928 developed for spatial surveillance of cancer and subsequently used and enhanced for many
929 types of hotspot detection.(31)

930 Technical Details: The null hypothesis is that the set of data subregions (often zip codes) in the
931 recent time interval tested forms a random sample from an expected spatial distribution of
932 cases. The expected distribution is not uniform over subregions but reflects a “customary”
933 spatial case spread that reflects urban/suburban case ratios or other factors. ESSENCE
934 implementation calculates the expected spatial distribution using recent case counts from a
935 sliding baseline interval. In effect, the code is similar to a common application of SaTScan,
936 the space-time permutation scan statistic, restricted to test cases from only the most recent time
937 interval and assuming circular clusters.(32)

938 As in SaTScan, the method calculates a test statistic for each candidate cluster. The test statistic
939 in the ESSENCE implementation is Kulldorff’s Poisson log likelihood ratio. The set of
940 candidate clusters is generated by scanning over a set of cluster center locations, often taken
941 as centroids of all zip codes in the dataset, and considering all circles within a maximum radius
942 of each center, where the number of circles is limited by the number of data subregions within
943 each radius. The maximum test statistic over these candidates is then tested for significance.

944 Statistical significance inference does not depend on a theoretical distribution but on repeated
945 trials on simulated datasets randomly drawn using the baseline distribution. For each such
946 trial, the algorithm uses the same scanning procedure to derive a trial maximum.

947 For assessing the significance of the maximum test statistic over all observed clusters, the
948 ESSENCE code uses the Gumbel distribution method.(33) The code collects 99 trial maxima,
949 fits a Gumbel distribution to these values, and uses the fitted distribution to assign a p-value to
950 the test statistics of clusters found in the original data. The observed cluster with the maximum
951 test statistic is considered significant if its p-value is below a predetermined threshold, often
952 set to 0.01. This threshold criterion can yield multiple significant clusters in a given run if
953 more than one candidate cluster yields a test statistic whose p-value is below the threshold.
954 For each significant case cluster, the system shows the location, extent, and degree of
955 significance using the GIS software.

956 Benefits: The ESSENCE Java implementation inherits features that have popularized SaTScan.
957 Potential clusters of interest are localized without bias regarding the center or extent of the
958 cluster as well as the spatial resolution of the data allows. As noted in Kulldorff, Heffernan,
959 et al., the empirical significance testing with many repeated trials takes “into account the
960 multiple testing stemming from the many potential cluster locations and sizes evaluated.”(32)

961 Limitations: The most important limitation, applicable also to SaTScan and to all other spatial or
962 space-time cluster detection methods, is that the usefulness of the method strongly depends on
963 the reliability of the expected spatial distribution. The use of census-based distributions,
964 insurance eligibility lists, regression models, and other means have been used to derive the
965 expected distribution. The method implemented in ESSENCE infers this distribution from
966 recent data separated from the test date(s) by a 2-day buffer.

967 Evaluation of statistically significant clusters for epidemiological significance is a nontrivial
968 task which may be exacerbated if the number of significant clusters is misleading or excessive
969 because the expected distribution is unrepresentative or because investigation resources are
970 insufficient.

971 The use of this popular approach for prospective use has been criticized despite numerous
972 applications and published real-life successes,(34) and the ESSENCE implementation lacks
973 the prospective adjustment in SaTScan attempting to manage cluster rates for multiple
974 successive runs. The ESSENCE implementation also does not support elliptical cluster shapes,

975 simultaneous clustering of multiple data sources, or test statistics other than the Poisson log
976 likelihood ratio. The user with a sufficiently detailed dataset and an application that requires
977 these extended SaTScan features should be aware of these limitations.
978

979 **Time-of-arrival aberration detection**

980 Categorization: Multiple Automated Hypothesis tests

981 Purpose: This algorithmic approach was implemented to find and display unusual clusters of
982 syndromically related emergency department visits by patients arriving for care within a short
983 time interval.

984 Technical Details: Patient visit counts are tabulated by cells, with one cell for each hospital/time-
985 interval/sub-syndrome combination. See Figure 1.

986

987 **Figure 1: Schematic of time-of-arrival filtering and anomaly detection process**

988

- 989 • For the visit counts in each cell, a Poisson or negative binomial test is chosen using the
990 last 60 days of visit counts for that cell. The Poisson distribution is used unless the count
991 variance exceeds the mean by a factor of 1.1 or greater, and then the time series is
992 considered overdispersed. This situation occurs for relatively few cells, generally
993 corresponding to the more common (sub)syndromes for the largest hospitals at the busiest
994 times when most alerts would be generated. For this situation, a negative binomial
995 distribution is assumed. Parameter settings and distribution types were derived from
996 testing on several years of patient visit records from a variety of large and small hospitals.
- 997 • Once the distribution is chosen, parameters for each cell are calculated from the 60-day
998 baseline. For each cell, an alert is then flagged if the current count exceeds the upper
999 limit threshold for the chosen distribution based on a preselected p-value.
- 1000 • Based on empirical results using 12 years of data from 134 hospital EDs from a large
1001 state with labeled events, a threshold p-value of $p^* = 10^{-4}$ (0.0001) was chosen.
- 1002 • Time intervals for the cells are 30 min., 60 min. beginning on the hour, and 60 min.
1003 beginning on the half hour, again a result of empirical testing.

- 1004 • Practical overrides are implemented based on observed cell counts. At least three
1005 observed cases are required for an alert. This minimum may be increased for more
1006 common syndromes. Mandatory alerts may also be implemented for certain
1007 subsyndrome/count combinations, such as subsyndromes for severe illness, regardless of
1008 the hypothesis test.

1009
1010 **Benefits:** In validation testing to monitor visit clusters for 51 subsyndromes for 134 hospitals at
1011 the time intervals above with the chosen p-value threshold, alert rates were consistently
1012 manageable and found all known clusters from a small historical collection of events except
1013 for two groups of 3-4 visits at very busy times. The alert burden was still manageable at the
1014 county level when anomalous clusters for all hospitals within each county were combined. The
1015 simplicity of this approach allows multiple daily runs and adaptation to new improvised
1016 subsyndromes with rapid system response without impact on routine processing.

1017
1018 **Limitations:** The hypothesis tests include no direct modeling of seasonality or other systematic
1019 data behavior. They were implemented to enable county-level processing, and validation was
1020 conducted on a 12-year historical dataset from one state. Expanding the computational load to
1021 include much larger sets of hospitals or syndrome groups with limited investigation capability
1022 may require recalibration (p-value threshold, minimum alert counts) or an alternate approach
1023 to retain sensitivity with manageable alerting.

1024

1025 **Identification of term-based free-text anomalies**

1026 **Categorization:** Hypothesis test for excessive occurrence of selected chief complaint terms

1027 **Purpose:** This algorithm was implemented to point out unusual distributions of chief complaint
1028 terms of interest without dependence on syndrome definitions.

1029 **Technical Details:** For each term in any chief complaint in the most recent time interval, the
1030 number of chief complaints containing that term in this interval is tested relative to the number
1031 of chief complaints containing same term in a much larger sliding baseline interval, assumed
1032 to be representative of customary data.

- 1033 • The algorithm forms 2x2 contingency tables whose entries are:

1034 A=number of recent chief complaints with the term,
1035 B=number of recent chief complaints without the term,
1036 C=number of baseline chief complaints with the term,
1037 D=number of baseline chief complaints without the term

1038

- 1039 • For many such tables with small cell counts, Fisher’s Exact Test is applied to alert when
1040 the probability that count of the term of interest is $\geq A$, given the contingency table’s
1041 marginal totals, is below a critical p-value.
- 1042 • If the smaller of B and C is larger than 1000, then a chi-square test is applied to the same
1043 contingency table with negligible loss of accuracy. The critical threshold is then applied
1044 to half the resultant (two-sided) p-value to determine anomalous terms.
- 1045 • Results are shown only for candidate terms that have not been previously eliminated
1046 because they are stopwords with no informational content (such as “the”, “if”, “all”) or
1047 because users have previously added them to a list of terms to be ignored (such as
1048 “patient”, “complaint”, “test”).
- 1049 • Conventions adopted from empirical test results and also to avoid impacting ESSENCE
1050 processing are: The test interval is 24 hours, the baseline interval is 30 days, a buffer of 7
1051 days is implemented between test interval and baseline, and the threshold p-value is set at
1052 $p^* = 10^{-5}$ (0.00001).

1053

1054 **Benefits:** With the above p-value threshold and settings, this method detects as few as three
1055 instances of unusual terms (place names, rare signs/symptoms) and unusual concentrations of
1056 interesting common terms while averaging from 0-4 anomalous terms per day over small and
1057 large hospitals. Inspecting chief complaints containing each of a small number of terms each
1058 day and disqualifying some terms from further consideration is a manageable task that can
1059 uncover clusters of visits that could be missed by syndromic methods. In testing with historic
1060 data, chief complaints containing anomalous terms found with the strict p-value threshold
1061 adopted have included small clusters of visits resulting from documented events of food
1062 poisoning and heat-related illness. This analysis has also found new abbreviations used by
1063 hospitals in their chief complaints. These new abbreviations can then be added to the Chief
1064 Complaint Processor to improve syndrome and subsyndrome categorization.

1065 Limitations: While the simplicity of this method avoids impact on daily ESSENCE processing or
1066 user investigation of large collections of distributed data streams, the method has no means to
1067 interpret anomalous terms, identify phrases with multiple terms, or distinguish topics of
1068 concern in chief complaints that may not share specific terms. The only preprocessing of free-
1069 text terms is the application of ESSENCE Chief Complaint Processor spell checking and
1070 abbreviation/acronym expansion. Interpretation of the anomalies requires a human-in-the-
1071 loop, both to evaluate anomalous terms for investigation and to rule-out specific terms from
1072 future anomalies. The method is subject to effects of changing terminology (street drug names,
1073 triage vocabulary and abbreviations), and the user should be aware of current perceived health
1074 threats and corresponding emergency/urgent care language.

1075

1076 **REFERENCES**

- 1077 1. Lewis MD, Pavlin JA, Mansfield JL, O'Brien S, Boomsma LG, Elbert Y, et al. Disease
1078 outbreak detection system using syndromic data in the greater Washington DC area. *Am J*
1079 *Prev Med.* 2002;23(3):180-6.
- 1080 2. Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, et al. A systems
1081 overview of the Electronic Surveillance System for the Early Notification of Community-
1082 Based Epidemics (ESSENCE II). *J Urban Health.* 2003;80(2 Suppl 1):i32-42.
- 1083 3. Siegrist D, Pavlin J. Bio-ALIRT biosurveillance detection algorithm evaluation. *MMWR*
1084 *Morb Mortal Wkly Rep.* 2004;53 Suppl:152-8.
- 1085 4. National Syndromic Surveillance Program. BioSense Platform Quick Start Guide to Using
1086 ESSENCE (October 2016) 2016 [updated 10/14/2016.
1087 [https://www.cdc.gov/nssp/biosense/docs/biosense-platform-quick-start-guide-for-](https://www.cdc.gov/nssp/biosense/docs/biosense-platform-quick-start-guide-for-essence.pdf)
1088 [essence.pdf](https://www.cdc.gov/nssp/biosense/docs/biosense-platform-quick-start-guide-for-essence.pdf)].

- 1089 5. Kumar RS, White JR. Impact of the NSSP's transition to ESSENCE on chief complaint field-
1090 based syndromes. *Online J Public Health Inform.* 2018;10(1).
- 1091 6. IBM Knowledge Center. Overview of Rational Rhapsody Poughkeepsie, NY USA: IBM
1092 Knowledge Center; 2020 [Available from:
1093 [https://www.ibm.com/support/knowledgecenter/SSB2MU_8.3.0/com.ibm.rhp.overview.doc/t](https://www.ibm.com/support/knowledgecenter/SSB2MU_8.3.0/com.ibm.rhp.overview.doc/topics/rhp_c_po_rr_product_overview.html)
1094 [opics/rhp_c_po_rr_product_overview.html](https://www.ibm.com/support/knowledgecenter/SSB2MU_8.3.0/com.ibm.rhp.overview.doc/topics/rhp_c_po_rr_product_overview.html).]
- 1095 7. HealthCare Integrations. Interface Monitor 2019 [updated September 1, 2019. Available from:
1096 [https://interfacemonitor.com/?gclid=EAIaIQobChMI8Dl8s_T5gIVBmyGCh0EbQdGEAAy](https://interfacemonitor.com/?gclid=EAIaIQobChMI8Dl8s_T5gIVBmyGCh0EbQdGEAAyASAAEgKmUfD_BwE.)
1097 [ASAAEgKmUfD_BwE.](https://interfacemonitor.com/?gclid=EAIaIQobChMI8Dl8s_T5gIVBmyGCh0EbQdGEAAyASAAEgKmUfD_BwE.)]
- 1098 8. Apache Software Foundation. The Apache Groovy programming language Wakefield, MA:
1099 Apache Software Foundation; 2019 [Available from: <https://groovy-lang.org/>].
- 1100 9. Apache Software Foundation. Apache Tomcat 2019 [
- 1101 10. Foundation OSG. GeoServer. 2.16.2 ed. Beaverton, Oregon USA: Open Source Geospatial
1102 Foundation; 2020.
- 1103 11. Office of the Assistant Secretary for Preparedness and Response. Disaster Medical
1104 Assistance Teams Washington D.C., USA: U.S. Department of Health and Human Services;
1105 2019 [updated September 9, 2017. Available from:
1106 <https://www.phe.gov/Preparedness/responders/ndms/ndms-teams/Pages/dmat.aspx>.
- 1107 12. American Association of Poison Control Centers. AAPCC National Poison Data System
1108 Arlington, Virginia USA: American Association of Poison Control Centers; 2019 [December
1109 18, 2019]. Available from: <https://www.npds.us>.
- 1110 13. Sniegowski CA. Automated syndromic classification of chief complaint records. *Johns*
1111 *Hopkins Apl Technical Digest.* 2004;25(1):68-75.

- 1112 14. Munroe JG, Straver R, Rubino H, Pritchard S, Atrubin D, Hamilton JJ. MERS PUI
1113 Surveillance and Restrospective Identification in ESSENCE-FL, 2013-2015. Online J Public
1114 Health Inform. 2017;9(1).
- 1115 15. Atrubin D. Syndromic Surveillance Identifies Unreported Cases of Zika Virus Disease,
1116 2016–2017 2019 [updated January 15, 2019. Available from:
1117 [https://www.cdc.gov/nssp/documents/success-stories/NSSP-Success-Story-florida-zika-](https://www.cdc.gov/nssp/documents/success-stories/NSSP-Success-Story-florida-zika-508.pdf)
1118 [508.pdf.](https://www.cdc.gov/nssp/documents/success-stories/NSSP-Success-Story-florida-zika-508.pdf)]
- 1119 16. Wiese M, Atrubin D, McDougale W, Lewis J. Effectiveness of Using a Chief Complaint and
1120 Discharge Diagnosis Query in ESSENCE-FL to Identify Possible Tuberculosis Patients and
1121 Contacts in Hillsborough County, Florida. Online J Public Health Inform. 62014.
- 1122 17. Xing J, Burkom H, Tokars J. Method selection and adaptation for distributed monitoring of
1123 infectious diseases for syndromic surveillance. J Biomed Inform. 2011;44(6):1093-101. doi:
1124 [10.16/j.jbi.2011.08.012](https://doi.org/10.16/j.jbi.2011.08.012). Epub Aug 24.
- 1125 18. Brillman JC, Burr T, Forslund D, Joyce E, Picard R, Umland E. Modeling emergency
1126 department visit patterns for infectious disease complaints: results and application to disease
1127 surveillance. BMC Med Inform Decis Mak. 5. England2005. p. 4.
- 1128 19. Burkom HS, Elbert Y, Magruder SF, Najmi AH, Peter W, Thompson MW. Developments in
1129 the Roles, Features, and Evaluation of Alerting Algorithms for Disease Outbreak Monitoring.
1130 Johns Hopkins APL Technical Digest. 2008;27(4):313.
- 1131 20. Zhou H, Burkom H, Winston C, Dey A, Ajani U. Practical comparison of aberration
1132 detection algorithms for biosurveillance systems. J Biomed Inform. 2015.

- 1133 21. Tokars JI, Burkom H, Xing J, English R, Bloom S, Cox K, et al. Enhancing time-series
1134 detection algorithms for automated biosurveillance. *Emerg Infect Dis.* 2009;15(4):533-9. doi:
1135 10.3201/eid1504.080616.
- 1136 22. Ryan TP. *Statistical methods for quality improvement*: John Wiley & Sons; 2011.
- 1137 23. Morton AP, Whitby M, McLaws ML, Dobson A, McElwain S, Looke D, et al. The
1138 application of statistical process control charts to the detection and monitoring of hospital-
1139 acquired infections. *J Qual Clin Pract.* 2001;21(4):112-7.
- 1140 24. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and
1141 response Early Aberration Reporting System (EARS). *J Urban Health.* 2003;80(2 Suppl
1142 1):i89-96.
- 1143 25. Zhu Y, Wang W, Atrubin D, Wu Y. Initial evaluation of the early aberration reporting
1144 system--Florida. *MMWR Morb Mortal Wkly Rep.* 54 Suppl. United States2005. p. 123-30.
- 1145 26. Lawson BM, Fitzhugh E, S. H, Hutwagner L. From Implementation to Automation --- A
1146 Step-by-Step Approach to Developing Syndromic Surveillance Systems from a Public Health
1147 Perspective 2020 [Available from:
1148 <https://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a60.htm>.
- 1149 27. Zhang H, Li Z, Lai S, Clements ACA, Wang L, Yin W, et al. Evaluation of the Performance
1150 of a Dengue Outbreak Detection Tool for China. *PLoS One.* 92014.
- 1151 28. Simes RJ. An Improved Bonferroni Procedure for Multiple Tests of Significance.
1152 *Biometrika.* 1986;73(3):751-4.
- 1153 29. Hommel G. A stagewise rejective multiple test procedure based on a modified. 2019.

- 1154 30. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful
1155 Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*
1156 (Methodological). 1995;57(1):289-300.
- 1157 31. Kulldorff M. A spatial scan statistic. *Communications in Statistics-Theory and Methods*.
1158 1997;26(6):1481-96.
- 1159 32. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation
1160 scan statistic for disease outbreak detection. *Plos Medicine*. 2005;2(3):216-24.
- 1161 33. Abrams AM, Kleinman K, Kulldorff M. Gumbel based p-value approximations for spatial
1162 scan statistics. *International Journal of Health Geographics*. 2010;9.
- 1163 34. Correa TR, Assuncao RM, Costa MA. A critical look at prospective surveillance using a
1164 scan statistic. *Stat Med*. 2015;34(7):1081-93.
- 1165 35. Franz DR, Jahrling PB, Friedlander AM, McClain DJ, Hoover DL, Bryne WR, et al. Clinical
1166 recognition and management of patients exposed to biological warfare agents. *Jama*.
1167 1997;278(5):399-411.
- 1168 36. Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based
1169 syndromic surveillance for influenzalike illness by International Classification of Diseases,
1170 Ninth Revision. *Emerg Infect Dis*. 2007;13(2):207-16.
- 1171 37. Hopkins RS, Kite-Powell A, Goodin K, Hamilton JJ. The Ratio of Emergency Department
1172 Visits for ILI to Seroprevalence of 2009 Pandemic Influenza A (H1N1) Virus Infection,
1173 Florida, 2009. *PLoS Curr*. 2014;6.
- 1174 38. Schirmer P, Lucero C, Oda G, Lopez J, Holodniy M. Effective detection of the 2009 H1N1
1175 influenza pandemic in U.S. Veterans Affairs medical centers using a national electronic
1176 biosurveillance system. *PLoS One*. 2010;5(3):e9533.

- 1177 39. Burkom HS BA, Paladini M, Revere D, McIntyre A. Analytic Evaluation of a
1178 Standardization Effort for the Distribute Emergency Department Surveillance Project.
1179 Proceedings of the Joint Statistical Meetings. 2011:1450-62.
- 1180 40. Kuspis K, Jagger M, Powell M, Hillwig R. Recognizing Recreational Water Exposure and
1181 Habituating HAB Surveillance in ESSENCE. 10. 2018.
- 1182 41. Bemis K, Frias M, Patel MT, Christiansen D. Using an Emergency Department Syndromic
1183 Surveillance System to Evaluate Reporting of Potential Rabies Exposures, Illinois, 2013-
1184 2015. Public Health Rep. 2017;132(1 Suppl):59S-64S.
- 1185 42. Collins J. Arizona Monitors Transfer of Patients with RMSF from Tribal Lands to Facilities
1186 in Maricopa County 2019 [
- 1187 43. O'Connell EK, Zhang G, Leguen F, Llau A, Rico E. Innovative uses for syndromic
1188 surveillance. Emerg Infect Dis. 2010;16(4):669-71.
- 1189 44. NCoP. Nssp Knowledge Repository: Syndrome Definition Subcommittee Calls 2020
1190 [Available from: <https://www.cste.org/page/nsspcop>.
- 1191 45. Powell M, Boyd L. Tracking Health Effects of Wildfires: The Oregon ESSENCE Wildfire
1192 Pilot Project. 9. 2017.
- 1193 46. Wiedeman C, Shaffner J, Squires K, Leegon J, Murphree R, Petersen PE. Monitoring Out-
1194 of-State Patients during a 2017 Hurricane Response using ESSENCE. 10. 2018.
- 1195 47. Baer A, Elbert Y, Burkom HS, Holtry R, Lombardo JS, Duchin JS. Usefulness of syndromic
1196 data sources for investigating morbidity resulting from a severe weather event. Disaster Med
1197 Public Health Prep. 2011;5(1):37-45. doi: 10.1001/dmp.2010.32. Epub Sep 28.

- 1198 48. Falise AM, Griffin I, Fernandez D, Rodriguez X, Moore E, Barrera A, et al. Carbon
1199 Monoxide Poisoning in Miami-Dade County Following Hurricane Irma in 2017. *Disaster*
1200 *Med Public Health Prep.* 2019;13(1):94-6.
- 1201 49. Kyle Ryff M, Melissa Powell M, Michelle Barber M, and Laurel Boyd M, all of whom spent
1202 significant time in the development of OPHD's Rhapsody engine and informatics
1203 infrastructure. This publication was supported by Cooperat, Laing R. Integrating Poison
1204 Center Data into Oregon ESSENCE using a Low-Cost Solution. 9. 2017.
- 1205 50. Eggers C, Passman D, Chern A, Spector D, Kite-Powell A, Davis T, et al. Operational
1206 Experience: Integration of ASPR Data into ESSENCE-FL during the RNC. 5. 2013.
- 1207 51. Lombardo JS, Sniegowski CA, Loschen WA, Westercamp M, Wade M, Dearth S, et al. Public
1208 Health Surveillance for Mass Gatherings. *Johns Hopkins Apl Technical Digest.*
1209 2008;27(4):347-55.
- 1210 52. Jagger MA, Jaramillo S, Boyd L, Johnson B, Reed KR, Powell M. Mass Gathering
1211 Surveillance: New ESSENCE Report and Collaboration Win Gold in OR. *Online J Public*
1212 *Health Inform.* 2017;9(1).
- 1213 53. Garrett-Cherry TA, Hennenfent AK, McGee S, Davies-Cole J. Enhanced One Health
1214 Surveillance during the 58th Presidential Inauguration-District of Columbia, January 2017.
1215 *Disaster Med Public Health Prep.* 2019:1-7.
- 1216 54. Hoy D, Mercier A, White P, Saketa ST, Roth A, Souares Y, et al. Using SAGES
1217 OpenESSENCE for Mass Gathering Events. *Online J Public Health Inform.* 2015;7(1).
- 1218 55. Pavlin JA, Murdock P, Elbert E, Milliken C, Hakre S, Mansfield J, et al. Conducting
1219 population behavioral health surveillance by using automated diagnostic and pharmacy data
1220 systems. *MMWR Suppl.* 2004;53:166-72.

- 1221 56. Sussman L, Brown T. Using ESSENCE to Meet Local Needs for Mental Health Data: Query
1222 & Results. *Online J Public Health Inform.* 2019;11(1).
- 1223 57. Demian Christiansen KB, and Victoria Persky, Patel MT. Assessment of the use of ED Chief
1224 Complaint Data for monitoring Chronic Diseases. 10. 2018.
- 1225 58. Culpepper AB, Atrubin D, Hamilton JJ, Cui D. Day of Week Analysis of Myocardial
1226 Infarctions Using ESSENCE-FL Emergency Department Data. 8. 2016.
- 1227 59. Hensley J, Gonzalez S, Stover D, Safranek T, Qu M. Evaluation of ESSENCE Syndromic
1228 Definitions for ED Visits Related to Falls in Icy Weather. *Online J Public Health Inform.*
1229 112019.
- 1230 60. Torgerson A. Using ESSENCE to Detect Bomb-Making Activity: What's Appropriate? 10.
1231 2018.
- 1232 61. Marx GE, Chen Y, Askenazi M, Albanese BA. Syndromic Surveillance of Emergency
1233 Department Visits for Acute Adverse Effects of Marijuana, Tri-County Health Department,
1234 Colorado, 2016-2017. *Public Health Rep.* 2019;134(2):132-40.
- 1235 62. Wiese M, Clark CR. Analysis of ED and UCC Visits Related to Synthetic Marijuana in
1236 ESSENCE-FL, 2010-2015. *Online J Public Health Inform.* 82016.
- 1237 63. Atrubin D, Bowden S, Hamilton JJ. Using Syndromic Surveillance to Rapidly Describe the
1238 Early Epidemiology of Flakka Use in Florida, June 2014 – August 2015. 8. 2016.
- 1239 64. Hudson T-ML, Klekamp BG, Matthews SD. Local Public Health Surveillance of Heroin-
1240 Related Morbidity and Mortality, Orange County, Florida, 2010-2014: Public Health
1241 Reports. 2017.
- 1242 65. Jacquemin BJ, Hamby T, Tsai S. Using probabilistic matching to improve opioid drug
1243 overdose surveillance, New Jersey. *Online J Public Health Inform.* 2019;11(1).

- 1244 66. Hou W, Brutsch E, Dunn AC, Burnett CL, Dimond MP, Nakashima AK. Using Syndromic
1245 Data for Opioid Overdose Surveillance in Utah. *Online J Public Health Inform.* 2018;10(1).
- 1246 67. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for
1247 the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal*
1248 *Statistical Society Series a-Statistics in Society.* 2012;175:49-82.
- 1249 68. Benneyan J, Lloyd R, Plsek P. Statistical process control as a tool for research and
1250 healthcare improvement. *Qual Saf Health Care.* 2003;12(6):458-64.
- 1251 69. Nssp Community of Practice 2020 [Available from: <https://www.cste.org/page/nsspcop>.
- 1252 70. Surveillance UCDoH1a. National Syndromic Surveillance Program Update - December
1253 2019 | CDC 2019 [updated December 12, 2019. Available from:
1254 <https://www.cdc.gov/nssp/news.html>.
- 1255 71. Burkom H. Evolution of Public Health Surveillance: Status and Recommendations.
1256 *American Journal of Public Health.* 2017.
- 1257 72. Alemi F, Atherton MJ, Pattie DC, Torii M. Continuously rethinking the definition of
1258 influenza for surveillance systems: a Dependent Bayesian Expert System. *Med Decis*
1259 *Making.* 2013;33(6):860-8.
- 1260 73. Burkom HS, Ramac-Thomas L, Babin S, Holtry R, Mnatsakanyan Z, Yund C. An integrated
1261 approach for fusion of environmental and human health data for disease surveillance. *Stat*
1262 *Med.* 2011;30(5):470-9. doi: 10.1002/sim.3976. Epub 2011 Feb 3.
- 1263 74. Burkom H, Elbert Y, Ramac-Thomas L, Cuellar C. Validation of Analytic Methods for
1264 Combining Evidence Sources in Biosurveillance. *Online Journal of Public Health*
1265 *Informatics.* 2014;6(1).

- 1266 75. Lake IR, Colon-Gonzalez FJ, Barker GC, Morbey RA, Smith GE, Elliot AJ. Machine
1267 learning to refine decision making within a syndromic surveillance service. BMC Public
1268 Health. 2019;19(1):559.
1269

Data Source ER Data by Patient Location Start Date 12Jul2006 End Date 10Oct2006 Detector Regression/EWMA 1.2

Message Area

Query

[[SYNDROME="Fever"] AND [CHIEFCOMPLAINTS="chil*"]]

Example [AGEGROUP = "00-04"] OR ([SUBSYNDROME = "ACUTE BLOOD ABNORMALITIES"] AND [ZIPCODE = "21043"]) More...

Query Builder

AND OR () Add Expression Undo Last Change

Variable [T]q

- ZIPCODE-RACEASIANPERCENTAGE
- ZIPCODE-RACEBLACKORAFRICANAMERICANPERCENTAGE
- ZIPCODE-RACEHAWAIIANORPACIFICISLANDERPERCENTAGE
- ZIPCODE-RACEOTHERPERCENTAGE
- ZIPCODE-RACEWHITEPERCENTAGE
- ZIPCODEFREETEXT
-
- PRIVATE SAVED EXPRESSIONS
- PUBLIC SAVED EXPRESSIONS**
- ADMINISTRATOR SAVED EXPRESSIONS

Operator

Values

fever chills [SYNDROME="Fever"] AND [CHIEFCOMPLAINTS="chil*"]

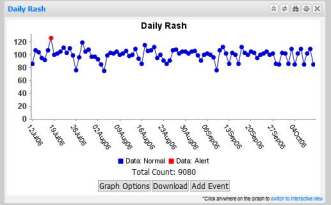
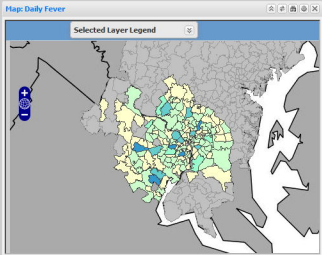
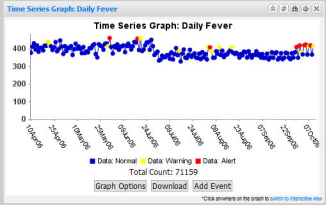
Validate Query

Save Private Expression

Save Public Expression

Clear Query

Execute



Data Details: Daily Fever and Blood (There are no additional records)

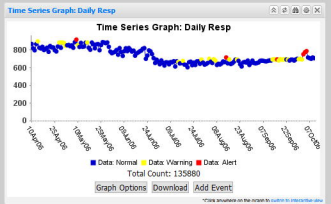
Date	Region	Category_flat
10Oct2006	Other_Region	:Fever;SI_Death;
09Oct2006	Prince Georges, MD	:Resp;
08Oct2006	Prince William, VA	:Hem_II;Resp;
07Oct2006	Other_Region	:Fever;SI_Death;
06Oct2006	Prince Georges, MD	:Resp;
05Oct2006	Prince William, VA	:Hem_II;Resp;
05Oct2006	Washington, DC	:Fever;Hem_II;
05Oct2006	Washington, DC	:Fever;Hem_II;
04Oct2006	Other_Region	:Fever;SI_Death;
03Oct2006	Prince Georges, MD	:Resp;
02Oct2006	Other_Region	:GI;Hem_II;
02Oct2006	Other_Region	:Hem_II;Resp;
01Oct2006	Prince William, VA	:Hem_II;Resp;
30Sep2006	Other_Region	:Fever;
29Sep2006	Montgomery, MD	:GI;Hem_II;
29Sep2006	Montgomery, MD	:Hem_II;Resp;
29Sep2006	Other_Region	:Fever;Hem_II;
28Sep2006	Other_Region	:Hem_II;Resp;

Rich Text Widget

Text to explain the dashboard.

My Alerts: Detection

Alerts	Alert Definition	Stratifications	Date	Data Source	Level
Daily Fever	Use Original	05Oct06	ER by Patient	0.018	
Daily Fever	Use Original	08Oct06	ER by Patient	0.020	
Daily AQT Fever	Use Original	05Oct06	ER by Patient	0.018	
Daily AQT Fever	Use Original	08Oct06	ER by Patient	0.020	
Fars C2 on Daily	Syndrom: Resp	05Oct06	ER by Patient	4.807	





Edit myAlert

Name of myAlert:

Query:

Enabled:



myAlert being created for: Records of Interest Detection




Stratifications: 

Detector: 

Threshold:

Minimum Count: 


 alerts in the past  days


consecutive alerts

Query Wizard

Datasource: ER Data by Patient Location
Time Resolution: Daily
Detector: Regression/EWMA 1.2
As Percent Query: No Percentage Query
Start Date: 12Jul06
End Date: 10Oct06

Available Query Fields

- ER Data by Patient Location
 - MyFilters
- Geography System
 - Region
 - Zipcode Free Text
- Facility
 - Facility Location
 - Zipcode
- Medical Grouping System
 - ChiefComplaintSubSyndromes
 - Syndrome
 - ChiefComplaints**
 - Age Group
 - CDC ILI Reporting Age Group
 - NCHS Age Group
 - Ten Year Age Group
 - Distribute Age Group

ChiefComplaints

Select Help Query Builder

^fever^,and,^chills^

Also apply the search string to:

Discharge Disposition Text
Discharge Diagnosis
CC and DD
Syndrome Free Text
SubSyndrome Free Text
Chief Complaint Orig Free Text
Clinical Impression

Select Help

Selected Query Fields

Geography System

Region

Medical Grouping System

ESSENCESyndromes

Sorts/Filters/Processing (Beta)

Apply R Script

Table Builder

Time Series

Data Details

Text Analysis

Graph Builder

Overview

Adv Qry

Explain Qry

Reset

MyFilter:

Create

