23
CHAPTER

Chief Complaints and ICD Codes

Michael M. Wagner and William R. Hogan
RODS Laboratory, Center for Biomedical Informatics
University of Pittsburgh, Pittsburgh, Pennsylvania

Wendy W. Chapman
Center for Biomedical Informatics
University of Pittsburgh, Pittsburgh, Pennsylvania

Per H. Gesteland
Inpatient Medicine Division, Department of Pediatrics
University of Utah School of Medicine, Salt Lake City, Utah

1. INTRODUCTION

A *chief complaint* is a concise statement in English or other natural language of the symptoms that caused a patient to seek medical care. A triage nurse or registration clerk records a patient's chief complaint at the very beginning of the medical care process (Figure 23.1).

In contrast, an ICD code is a number (e.g., 558.9) that a clinician or professional coder uses to represent a medical diagnosis, syndrome, or symptom–usually for the purpose of billing. The ICD-coding system allows physicians and professional coders to express their diagnostic impression of a patient at different levels of diagnostic precision, ranging from very precise (e.g., ICD code 022.1 for inhalational anthrax) to syndrome (e.g., ICD code 079.99 for viral syndrome) to symptom (e.g., ICD code 780.6 for fever). The diagnosis may be a working diagnosis (a provisional diagnosis) or a definitive diagnosis, although ICD does not allow the clinician or coder to indicate this distinction. A clinician or professional coder may record an ICD code early in the process of medical care. Professional coders, not clinicians, invariably encode hospital discharge diagnoses, which are not available until after a patient is discharged from a hospital. The important points to remember about ICD coding are the heterogeneity in diagnostic precision, who does the encoding, and when the encoding is done.

Chief complaints and ICD codes are used ubiquitously in medical care in the United States in both the civilian and military healthcare systems. Medicare and other third party payers require these data for billing and claims. As a result, the healthcare industry has built significant electronic infrastructure to capture chief complaints and ICD codes.

Over the past six years, researchers have studied methods to obtain and analyze patient chief complaints and ICD codes for the purpose of early detection of outbreaks. The intensity of research on these data has been motivated in part by their availability. The objective of research is to test hypotheses that these data can be used either alone or in conjunction with other data to improve the timeliness of outbreak detection (Wagner et al., 2001). As a result of the research, many health departments are now routinely monitoring chief complaints and ICD codes.

For clarity of exposition, we discuss chief complaints and ICD codes separately in this chapter. However, we do not wish to reinforce a somewhat prevalent impression that they are competing alternatives. Both types of data contain information that is useful in biosurveillance and together they are complementary. In the future, we expect that biosurveillance systems will collect both types of data routinely. They will link these data to other data about a patient to support more accurate inference about a patient's true disease state. We explore the future roles of chief complaints and ICD codes and their synergies in the final section of this chapter.

2. CHIEF COMPLAINTS

The concept of a *chief complaint* is important in medicine. It is a statement of the reason that a patient seeks medical care. Medical and nursing schools teach future clinicians to begin their verbal presentations of patient cases with a statement of the chief complaint. They teach them to record the chief complaint using the patient's words and to avoid replacing the patient's words with their diagnostic interpretation. It is considered bad form to proffer a diagnostic impression in a chief complaint.[1]

---

1 The reasons for this practice are myriad. One reason is that other clinicians such as consultants and supervisory clinicians in academic medical centers read a patient's chart like detectives: They wish to form an independent diagnostic impression. They are interested in knowing 'just the facts.'
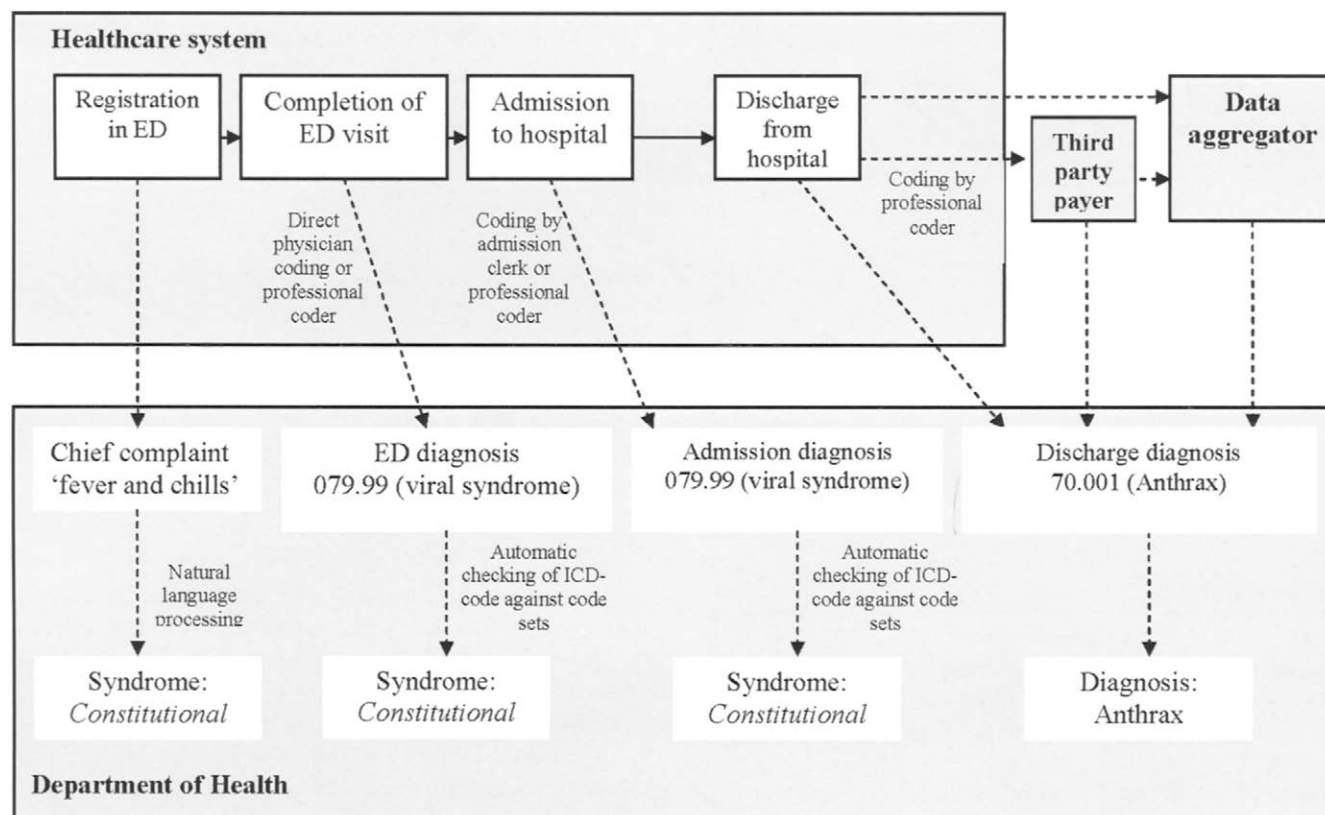
FIGURE 23.1 Points in the healthcare process at which chief complaints and ICD codes are recorded and transmitted to a health department or other biosurveillance organization. This figure illustrates a hypothetical patient with *anthrax* who seeks care at an emergency department (ED) and is subsequently admitted to a hospital. The patient's chief complaint is recorded at the time of registration and transmitted immediately to a health department via a HL7 message router. When the patient is discharged from the ED and admitted to the hospital, a professional coder reads the patient's ED chart and assigns ICD codes for billing purposes. The delay in transmission to a health department is indicated by a slanted arrow. Another ICD code may be assigned at the time of hospital admission. Finally, ICD codes are assigned by professional coders at the time of hospital discharge. These codes are transmitted to third party payers, who may submit them to data aggregators (e.g., commercial companies that analyze healthcare trends or health departments that assemble hospital discharge data sets for statistical purposes). In general, the diagnostic precision of the data available to a health department increases over time (moving from left to right in the figure). Note that there is variability from healthcare system to healthcare system. In some settings, the chief complaints are coded directly into ICD codes by physicians at the time of service (e.g., U.S. military).

As a result, the chief complaint usually states the key symptoms that a patient is experiencing.

During the process of medical care, a patient's chief complaint is recorded many times. Triage nurses and registration clerks create the first record at the time of initial registration for service at a clinic or emergency department (ED). Clinicians also record chief complaints in daily progress notes and discharge, transfer, and patient acceptance summary notes.

The research that we will discuss has shown that chief complaints contain information that may be very useful in biosurveillance. This result is not surprising. If a patient is ill with an infectious disease and presents to a physician, we would expect her chief complaint to reflect the nature of the illness.

2.1. Description of Chief Complaint Data Used in Biosurveillance

The recorded chief complaint of most interest to biosurveillance is the one recorded at the time a patient initially presents for medical care.[2] This chief complaint is often recorded directly into a registration computer by triage nurses or registration clerks and is highly available for biosurveillance purposes.

Table 23.1 is a sample of chief complaints from a registration computer in an ED. The chief complaints are more terse (four or five words) than those recorded in physician notes. They also contain misspellings, unconventional abbreviations, and unorthodox punctuation. Only two of these chief complaints contain diagnoses (*finger lac* and *uti*, which are abbreviations

---

2   Chief complaints are also recorded by call centers, as we discuss in Chapter 28. The methods and results discussed in this chapter apply equally to chief complaints obtained in these settings.

TABLE 23.1 Examples of Chief Complaints Recorded in an Emergency Department

Chief complaint	CoCo Syndrome
diff breathing	*Respiratory*
chest pain	*Other*
abd pain nausea vomiting	*Gastrointestinal*
Finger lac	*Other*
resp dist	*Respiratory*
Fever	*Constitutional*
nausea diarhea chest tightness sob	*Gastrointestinal, Respiratory*
chest pain vomiting	*Gastrointestinal*
r side pain	*Other*
rectal bleeding walkin	*Hemorrhagic*
chest pain	*Other*
Uti	*Other*
urinary problems	*Other*
abd pain	*Gastrointestinal*

Notes: These 14 examples come from a file used to train a Bayesian natural language processing (NLP) program called CoCo (described in Chapter 17). The second column shows the syndromes that a physician assigned to the chief complaints. For clarity, we adopt a typographical convention of italicizing syndromes.

for *finger laceration* and *urinary tract infection*, respectively). The rest describe the patient's symptoms. The second column of the table shows the syndromes that a human expert assigned to the patient for purposes of training a Bayesian natural language processor. We will discuss syndromes shortly.

2.1.1. Natural Language Processing of Free Text Chief Complaints

Before chief complaints can be analyzed by computerized biosurveillance systems, they must be converted from English (or other natural language) into computer-interpretable format. Biosurveillance systems typically use natural language processing (NLP) to convert chief complaints into computer-interpretable format. We are aware of one system that takes advantage of a routine translation of chief complaints into computer-interpretable form (Beitel et al., 2004).

There are two basic NLP methods for converting free-text chief complaints into computer-interpretable format—keyword parsing and probabilistic. We discussed these methods in Chapter 17 and will not repeat the discussion here.

The NLP component of a biosurveillance system analyzes a recorded chief complaint to classify a patient into a syndrome category. Some biosurveillance systems use NLP to identify syndromes directly and others use NLP to identify *symptoms* in the chief complaint and subsequently use Boolean (AND, OR, NOT) or probabilistic combinations of symptoms to assign a syndrome.

The subsequent (non-NLP) analysis performed by a biosurveillance system searches for clusters of syndromes in space,

time, and/or demographic strata of a population, as discussed in Part III.

2.1.2. Syndromes

The concept of a syndrome is important in medical care (and in epidemiology). A syndrome is a constellation of symptoms, possibly combined with risk factors and demographic characteristics of patients (e.g., age and gender). Familiar examples of syndromes are *SARS* (*severe acute respiratory syndrome*) and *AIDS* (*acquired immune deficiency syndrome*). A syndrome plays the same role as a diagnosis in medical care—it guides the physician in selection of treatments for patients.

In this chapter, we will be discussing syndromes such as *respiratory* that are far less diagnostically precise than SARS or AIDS. The syndromes used in automated analysis of chief complaints and ICD codes are diagnostically imprecise by intent. The developers of these syndromes recognize that chief complaints (and ICD-coded diagnoses obtained close to the time of admission) in general do not contain sufficient diagnostic information to classify a patient as having SARS or other more diagnostically precise syndrome. They create syndrome definitions that are sufficiently precise to be useful, but not so precise that few if any patients will be assigned to them automatically, based solely on information contained in a four- or five-word chief complaint (or ICD code assigned early during the process of medical care).

Table 23.2 is the set of syndromes used by the RODS system. The table shows each syndrome name (which is just a convenient handle to reference the syndrome definition) and its definition. The RODS system uses the syndrome definitions in two ways. First, it makes them available to epidemiologists and other users of the system to assist in interpreting time series and maps of chief complaint data that have been aggregated by syndrome. If the user sees an increase in a syndrome such as *respiratory*, his interpretation of the increase should be that it could be due to any disease that is consistent with the definition of the *respiratory* syndrome. Second, RODS provides the definitions to individuals who are developing training sets (discussed in Chapter 17) for the CoCo parser.[3]

Tables 23.3 and 23.4 are syndrome classification systems used by Maryland Department of Hygiene and Mental Health and the New York City Department of Health and Mental Hygiene, respectively.

A final subtle point about the definition of syndromes that applies to both chief complaint syndrome definitions and ICD-code sets described later: The field of artificial intelligence distinguishes between intensional and extensional definitions. Tables 23.2 and 23.4 are intensional definition of syndrome categories. They express the intent of the system designer.

---

3 In practice, we train the CoCo parser using a set of approximately 10,000 chief complaints that a human has manually classified into one of these eight categories. Note that there is nothing domain specific about a Bayesian classifier as it is a mathematical formalism. CoCo could just as easily be trained to recognize a set of injury-related chief complaints.

TABLE 23.2 CoCo Syndrome Definitions

CoCo Syndrome	Definition
Gastrointestinal	Includes pain or cramps anywhere in the abdomen, nausea vomiting, diarrhea and abdominal distension or swelling.
Constitutional	Is made up of nonlocalized, systemic problems including fever, chills, body aches, flu symptoms (viral syndrome), weakness, fatigue, anorexia, malaise, irritability, weight loss, lethargy, sweating (diaphoresis), light headedness, faintness and fussiness. Shaking (not chills) is not constitutional but is other. Includes all of the "vaguely unwell" terms: doesn't feel well, feels ill, feeling sick or sick feeling, feels bad all over, not feeling right, sick, in pain, poor vital signs. Shaking or shaky or trembling (not chills) are not constitutional but are other (8). However, tremor(s) is neurological (7). Note: cold usually means a URI (cold symptoms; 3), not chills. Weakness, especially localized, is often neurological (7), rather than constitutional.
Respiratory	Includes the nose (coryza) and throat (pharyngitis), as well as the lungs. Examples of respiratory include congestion, sore throat, tonsillitis, sinusitis, cold symptoms, bronchitis, cough, shortness of breath, asthma, chronic obstructive pulmonary disease (COPD), pneumonia, hoarseness, aspiration, throat swelling, pulmonary edema (by itself; if combined with congestive heart failure, it is 8). If both cold symptoms and flu symptoms are present, the syndrome is respiratory. *Note: "Sore throat trouble swallowing" is respiratory, not respiratory and botulinic. That is, the difficulty in swallowing is assumed to be an aspect of the sore throat.*
Rash	Includes any description of a rash, such as macular, papular, vesicular, petechial, purpuric or hives. Ulcerations are not normally considered a rash unless consistent with cutaneous anthrax (an ulcer with a black eschar). *Note: Itch or itchy by itself is not a rash.*
Hemorrhagic	Is bleeding from any site *except the central nervous system*, e.g., vomiting blood (hematemesis), nose bleed (epistaxis), hematuria, gastrointestinal bleeding (site unspecified), rectal bleeding and vaginal bleeding. Bleeding from a site for which we have a syndrome should be classified as hemorrhagic and as the relevant syndrome (e.g., <u>Hematochesia is gastrointestinal and hemorrhagic; hemoptysis is respiratory and hemorrhagic</u>). Bleeding from a site for which we have a syndrome should be classified as hemorrhagic only without reference to the relevant syndrome, *except* hematochesia... hemoptysis. *Note: "Spitting up blood" is assumed to be hemoptysis.*
Botulinic	Includes ocular abnormalities (diplopia, blurred vision, photophobia), difficulty speaking (dysphonia, dysarthria, slurred speech) and difficulty swallowing (dysphagia).
Neurological	Covers nonpsychiatric complaints which relate to brain function. Included are headache, head pain, migraine, facial pain or numbness, seizure, tremor, convulsion, loss of consciousness, syncope, fainting, ataxia, confusion, disorientation, altered mental status, vertigo, concussion, meningitis, stiff neck, tingling, numbness, cerebrovascular accident (CVA; cerebral bleed), tremor(s), vision loss or blindness (but changed or blurred vision or vision problem is botulinic). Dizziness is constitutional and neurological. *Note: headache can be constitutional is some contexts, for example, "headache cold sxs achey" or "headache flu sxs."*
Other	Is a pain or process in a system or area we are not monitoring. For example, flank pain most likely arises from the genitourinary system, which we are not modeling, and would be considered other. Chest pain with no mention of the source of the pain is considered other (e.g., chest pain [other] versus pleuritic chest pain [respiratory]). Earache or ear pain is other. Trauma is other. Hepatic encephalopathy (not neurological), dehydration (not constitutional), difficulty sleeping or inability to sleep (not constitutional), constipation (not constitutional), and choking (but aspiration is respiratory) are all other.

Note: A physician or other medical expert refers to these definitions when creating training examples for the CoCo Bayesian classifier (described in Chapter 17).

The extensional definition of each syndrome is the actual set of chief complaints that NLP parsers, trained or configured based on the intensional definitions, assign to the categories.

2.1.3. Symptoms

Some researchers have divided the one-step chief-complaint-to-syndrome assignment process into two steps. Rather than using NLP to assign a syndrome directly to a patient based on the chief complaint, they use NLP to find all of the symptoms embedded

TABLE 23.3 Syndrome Classification System Used by Maryland Department of Hygiene and Mental Health

Syndrome
Death
Gastrointestinal
Neurologic
Rash
Respiratory
Sepsis
Unspecified
Other

From Sniegoski, C. (2004). Automated syndromic classification of chief complaint records. *Johns Hopkins University APL Technical Digest* 25:68–75, with permission.

in the chief complaint. They define a syndrome as a Boolean or probabilistic combination of symptoms. The two-step process then is: (1) NLP extracts symptoms from the chief complaint, and (2) a non NLP process determines whether the symptoms satisfy a Boolean or probabilistic syndrome definition. As an example, consider the chief complaint "n/v/d." With a two-step process, an NLP system would extract three symptoms from the chief complaint: nausea, vomiting, and diarrhea. Any syndrome definition that required nausea, vomiting, or diarrhea would be satisfied by this chief complaint. Biosurveillance systems operated by Washington State, New York State, and those using the CoCo classifier translate free text directly to syndromes; ESSENCE, SyCo2, and MPLUS classify to symptoms first.

The two-step approach has several potential advantages. It is more natural for epidemiologists and physicians, who conceive of a syndrome as a combination of symptoms. In fact, *case definitions* (discussed in Chapter 3) are Boolean combinations of symptoms. Additionally, it is possible to create a new syndrome definition "on the fly" without retraining a Bayesian classifier or restructuring lists of keywords. To do this, one simply defines a Boolean or probabilistic combination

TABLE 23.4 Syndrome Classification System Used by New York City
Department of Health and Mental Hygiene

Syndrome	Includes	Excludes
Common cold	Nasal drip, congestion, stuffiness	Chest congestion, sore throat
Sepsis	Sepsis, cardiac arrest, unresponsive, unconscious, dead on arrival	
Respiratory	Cough, shortness of breath, difficulty breathing, croup, dyspnea, bronchitis, pneumonia, hypoxia, upper respiratory illness, chest congestion	Cold
Diarrhea	Diarrhea, enteritis, gastroenteritis, stomach virus	
Fever	Fever, chills, flu, viral syndrome, body ache and pain, malaise	Hay fever
Rash	Vesicles, chicken pox, folliculitis, herpes, shingles	Thrush, diaper and genital rash
Asthma	Asthma, wheezing, reactive airway, chronic obstructive airway disease	
Vomiting	Vomiting, food poisoning	

Heffernan, R., Mostashari, F., Das, D., et al. (2004b). Syndromic surveillance in
public health practice, New York City. *Emerg Infect Dis* 10:858–64.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&
dopt=Citation&list_uids=15200820 with permission.

of symptoms. This advantage is important because new syndromes emerge with regularity, so it is important to be able to create a new syndrome definition quickly. Human biology changes very slowly, so new symptoms do not occur and the NLP conversion from free-text to symptom will be relatively stable, except as the language patients and triage nurses use to record chief complaints slowly evolves (e.g., the first time a patient uttered "I think I have SARS").

A limitation of the two-step approach is that it has not been validated. A real concern is that users of such systems can define syndromes for which the system's sensitivity and specificity may be extremely poor. A user may create a syndromic definition that is rational from an epidemiological standpoint but is not well-suited to the input data being classified. Without deep knowledge of the underlying processing method and its assumptions, a user will be completely unaware of this phenomenon and may be falsely reassured by the absence of disease activity when the newly created syndrome is put into operational use. As an extreme example of this problem, consider that a user might define a syndrome as a Boolean conjunction (AND statement) of five symptoms. Since the average registration chief complaint comprises four words, it

is almost inconceivable that any patient would match such a syndrome definition.[4]

2.2. Availability and Time Latencies of Registration Chief Complaints

The chief complaints recorded at the time of registration are among the earliest data available electronically from a patient's interaction with the healthcare system. They are typically recorded before a doctor sees a patient. If the ED or a clinic is busy, many hours may pass before a registered patient is seen by a clinician.

The time latency between recording of a chief complaint and its availability to a biosurveillance organization can range from seconds to days, depending on whether the data collection system utilizes the HL7-messaging capability of a healthcare system for real-time communication or batch transfer of files (Figure 23.2). Hospitals are frequently capable of real-time transmission whereas office practices are not—unless they are associated with a larger organization (e.g., the Veterans Administration, U.S. military, or a large healthcare system).

Real-time transmission is possible when a healthcare system has a pre-existing Health Level 7 (HL7) messaging capability. Several publications describe the technical approach to HL7-based data collection and chief-complaint processing (Tsui et al., 2002, 2003, Gesteland et al., 2002, 2003, Olszewski, 2003a). Briefly, when a patient registers for care at an ED, a triage nurse or registration clerk enters the chief complaint into a registration system. This step is part of normal workflow in many U.S. hospitals (Travers et al., 2003). The registration system almost always transmits chief-complaints in the form of HL7 messages (Tsui et al., 2003) to an HL7-message router located in the healthcare system. To transmit these data to a biosurveillance organization, the healthcare system would configure the HL7-message router to de-identify these messages and transmit them via the Internet to a biosurveillance organization as they are received from the registration system. This configuration process is a native capability of commercial HL7-message routers and it is a routine task for an HL7 engineer or other information technology staff working in or for a healthcare system.

Batch transfer can either be automatic or manual. Automatic means that a computer program periodically queries the registration computer (or other system in which the chief complaint data are stored) for recent registrations, writes a file, and transmits the file to the biosurveillance organization via

---

4 Febrile syndromes provide a more realistic and common example of this problem. Many of the infectious diseases that represent threats to the public's health produce a febrile response in affected individuals early in the course of illness. From an epidemiological standpoint, monitoring febrile syndromes, such as Febrile Respiratory, will increase the chance that a positive patient actually has an infectious disease and will decrease the number of false positives. However, chief complaints—being terse–rarely describe both a syndromic symptom and fever. Of 610 patients who actually had febrile syndromes (Febrile Respiratory, Febrile GI, Febrile Neurological, Febrile Hemorrhagic, or Febrile Rash), only 5% of the chief complaints described both fever and the symptoms related to the organ (Chapman and Dowling, 2005).
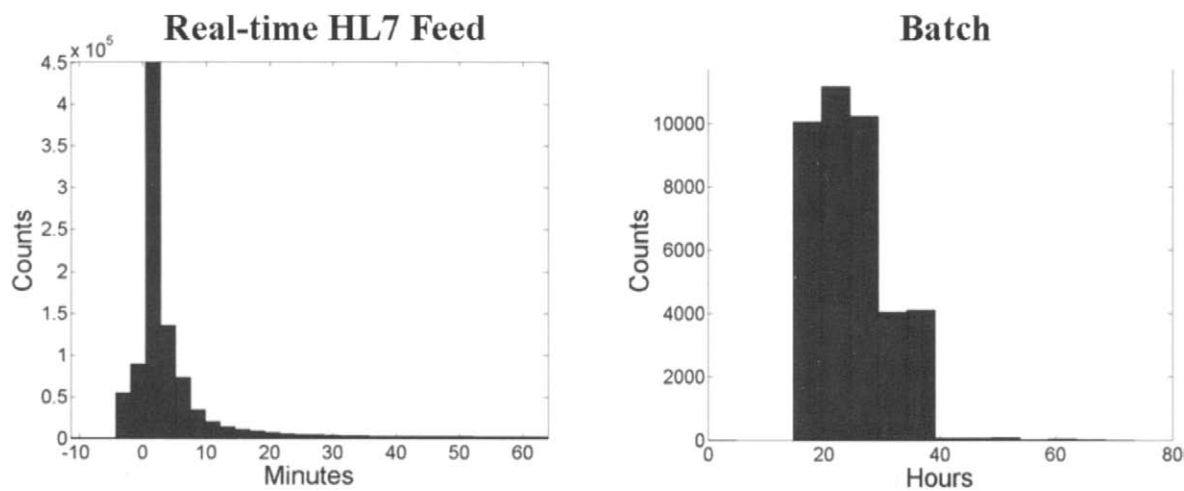
Real-time HL7 Feed

Batch

FIGURE 23.2 Comparison of time latencies of real-time and batch feeds. The negative time latencies associated with the real-time feed are due to slight clock differences between the biosurveillance system and the ED registration system.
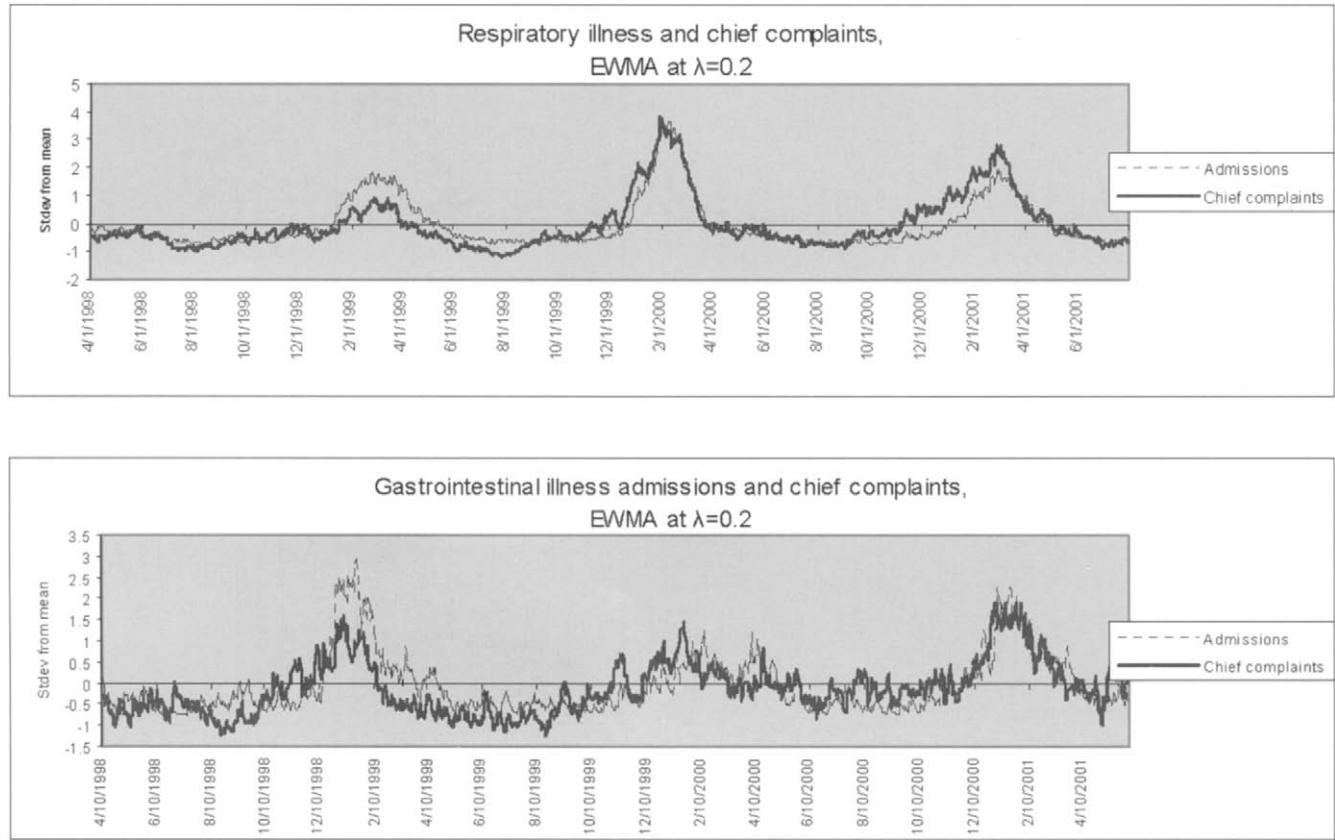
Respiratory illness and chief complaints, EWMA at λ=0.2

Gastrointestinal illness admissions and chief complaints, EWMA at λ=0.2

FIGURE 23.3 Detection of outbreaks in pediatric population from chief-complaint analysis, Utah 1998–2001.

the Internet. The transmission may use a secure file transfer protocol, non secure file transfer protocol, a web transfer protocol, or PHIN MS.[5] Manual means that someone working in the healthcare system must run a query and attached the results of the query to an email to the biosurveillance organization, or upload the file to a computer in the biosurveillance organization. In the past, manual data transfer often involved faxing of paper log files.

Tsui et al. (2005) studied the time latencies, data loss, and reliability associated with real-time HL7 feeds and batch feeds. Figure 23.3 compares the distribution of time delays between the time that a chief complaint was recorded during registration and receipt of that chief complaint by a biosurveillance system. The median time delay for a real-time feed was 0.033 hours and for batch was 23.95 hours.

The proportion of U.S. hospitals that are capable of sending a real-time HL7 feed appears to be approximately 84% based on our experience. Table 23.5 summarizes our experience with hospitals in the United States, suggesting that many hospitals have this capability.

2.3. Studies of Informational Value of Chief Complaints

Researchers have studied the ability of algorithms to detect syndromes and outbreaks of disease from chief complaints.

TABLE 23.5 Numbers of Hospitals Using Real-Time Versus Batch Connections to the RODS System, September 2005

Jurisdiction (project inception)	Healthcare Facilities		
	Real time	Batch	Location of Server
Pennsylvania (1999)	123	2	
Utah (2002)	26	0	
Ohio (2004)	50	14	Solaris/Linux/Oracle
Kentucky (2005)	0	5	(Running at the
Nevada (2005)	3	0	RODS Public Health
Atlantic City NJ (2004)	3	0	Data Center,
California (2005)	1	1	University of
Illinois (2005)	2	0	Pittsburgh)
Kentucky (2005)	0	5	
Michigan (2005)	2	0	
Los Angeles (2004)	1	4	Los Angeles County DOH
Houston TX (2004)	14	2	Houston DOH
Dallas Forth Worth Area TX (2004)	16	14	Tarrant County DOH
El Paso TX (2004)	2	0	El Paso
Totals	243	47	

Note: Many of these projects are statewide or citywide efforts that have an objective to connect every hospital to the health department. We provide the year of project inception to indicate the rate at which biosurveillance organizations are able to develop chief complaint data feeds from hospitals. DOH, Department of Health; RODS, Real-Time Outbreak and Disease Surveillance Laboratory, Center for Biomedical Informatics, University of Pittsburgh.

These studies contribute to our understanding of the informational value of chief complaints for the detection of cases and of outbreaks. The studies utilized experimental methods that we discussed in Chapter 21. In this section, we review these studies and discuss how they address the following three hypotheses of interest:

Hypothesis 1: A chief complaint can discriminate between whether a patient has *syndrome* or disease X or not (stated as the null hypothesis: the chief complaint cannot discriminate).

Hypothesis 2: When aggregated with the chief complaints of other patients in a region, chief complaints can discriminate between whether there is an *outbreak* of type Y or not.

Hypothesis 3: When aggregated with the chief complaints of other patients in a region, algorithmic monitoring of chief complaints can detect an *outbreak* of type Y earlier than current best practice (or some other reference method).

It is important to note that the experiments that we will discuss differ in many details. They differ in the hypothesis being tested; the syndrome or type of outbreak studied; the NLP method; the detection algorithm; and the reference standard used in the experiment. Thus, achieving a "meta-analytic" synthesis about the informational value and role of chief complaints in biosurveillance requires that we pay attention to these distinctions. The one thing these studies share in common, however, is that they are all studies of the informational value of chief complaints.

2.3.1. Detection of Cases

Table 23.6 summarizes the results of studies that are informative about Hypothesis 1: A chief complaint can discriminate between whether a patient has syndrome or disease X or not.

Methodologically, the studies measure the sensitivity and specificity with which different NLP methods (which we refer to as "classifiers") identify patients with a variety of syndromes using only the recorded chief complaints. The reference syndrome ("gold standard") for patients in these studies was developed by physician review of narrative medical records, such as ED reports, or automatically from ICD-9 primary discharge diagnoses. Most studies have evaluated detection of syndromes in adults, whereas a single study examined detection of syndromes in pediatric patients (Beitel et al., 2004).

Table 23.6 groups the experiments by syndrome because many experiments studied the same or similar syndromes. Each row in Table 23.6 reports the sensitivity and specificity of a classifier for a particular syndrome, and the likelihood ratio

---

5 Note that time latencies associated with automatic batch transfer cannot necessarily be decreased by decreasing the periodicity of the batch transfer. The query to the healthcare information system that generates the batch file may be resource intensive and a healthcare system may only be willing to schedule the query during non peak load periods (e.g., midnight).

TABLE 23.6 Performance of Bayesian and other classifiers in detecting syndromes

Classifier Being Tested	Reference Standard for Comparison	Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio (95% CI)	Negative Likelihood Ratio (95% CI)
Respiratory Syndrome					
Chief Complaint Bayesian Classifier *Respiratory* (CCBC)[a]	Utah Department of Health (UDOH) *Respiratory with fever*	0.52 (0.51–0.54)	0.89 (0.89–0.90)	5.0 (4.74–5.22)	0.54 (0.52–0.56)
CCBC[b]	Human review of ED reports	0.77 (0.59–0.88)	0.90 (0.88–0.92)	7.9 (5.8–10.8)	0.26 (0.13–0.49)
CCBC[a]	Utah ICD-9 list	0.60 (0.59–0.62)	0.94 (0.94–0.95)	10.45 (9.99–10.96)	0.25 (0.13–0.49)
Manual Assignment[c]	Human review of ED reports for *Pediatric respiratory illness*	0.47 (0.38–0.55)	0.99 (0.97–0.99)	56.73 (18.12–177.59)	0.53 (0.46–0.63)
CCBC[d]	ICD-9 list	0.63 (0.63–0.64)	0.94 (0.94–0.94)	11.14 (11.00–11.30)	0.39 (0.39–0.40)
CCBC[e]	Human review of ED reports	0.34 (0.30–0.38)	0.98 (0.97–0.99)	18.0 (11.24–28.82)	0.67 (0.63–0.71)
CCBC *Respiratory* and Keyword *Fever*[e]	Human review of ED reports for *Febrile respiratory*	0.02 (0.01–0.04)	0.99 (0.99–1.0)	20.83 (2.18–199.28)	0.99 (0.97–1.0)
Gastrointestinal (GI) Syndrome					
CCBC *GI*[a]	UDOH *Gastroenteritis without blood*	0.71 (0.69–0.74)	0.90 (0.90–0.90)	7.34 (6.98–7.72)	0.32 (0.29–0.35)
CCBC *GI*[f]	Human review of ED reports for *Acute infectious GI*	0.63 (0.35–0.85)	0.94 (0.92–0.96)	7.77 (4.77–12.65)	0.40 (0.20–0.80)
CCBC[a]	Utah ICD-9 list	0.74 (0.72–0.76)	0.92 (0.92–0.92)	9.5 (9.04–9.94)	0.28 (0.26–0.30)
CCBC[d]	ICD-9 list	0.69 (0.68–0.70)	0.96 (0.96–0.96)	15.70 (15.46–15.95)	0.32 (0.32–0.33)
CCBC[e]	Human review of ED reports	0.22 (0.16–0.29)	0.90 (0.88–0.91)	2.09 (1.51–2.91)	0.87 (0.80–0.95)
CCBC *GI* and Keyword *Fever*[e]	Human review of ED reports for *Febrile GI*	0.04 (0.02–0.08)	0.99 (0.99–1.0)	60.82 (7.65–483.45)	0.96 (0.93–0.99)
Neurologic/Encephalitic Syndrome					
CCBC Neurologic[a]	UDOH Meningitis/Encephalitis	0.47 (0.32–0.63)	0.93 (0.93–0.94)	4383.26[g] (1394.21–13780.56)	0.53 (0.39–0.72)
CCBC[a]	Utah ICD-9 list	0.72 (0.69–0.76)	0.95 (0.94–0.95)	13.5 (12.57–14.41)	0.29 (0.26–0.33)
CCBC[d]	ICD-9 list	0.68 (0.67–0.69)	0.93 (0.93–0.93)	9.25 (9.076–9.418)	0.35 (0.34–0.36)
CCBC[e]	Human review of ED reports	0.31 (0.27–0.35)	0.97 (0.95–0.98)	8.89 (6.25–12.65)	0.72 (0.68–0.76)
CCBC *Neurologic* and Keyword *Fever*[e]	Human review of ED reports for *Febrile Neurologic*	0.03 (0.01–0.07)	0.99 (0.99–1.0)	12.79 (2.89–56.59)	0.98 (0.95–1.0)
Constitutional Syndrome					
CCBC[d]	ICD-9 list	0.46 (0.45–0.47)	0.97 (0.97–0.98)	13.65 (13.30–14.0)	0.56 (0.55–0.57)
CCBC[e]	Human review of ED reports	0.27 (0.23–0.32)	0.95 (0.93–0.96)	5.12 (3.82–6.85)	0.77 (0.72–0.82)
Rash syndrome					
CCBC *Rash*[a]	UDOH *Febrile illness with rash*	0.50 (0.40–0.59)	0.99 (0.99–0.99)	55.6 (44.25–69.91)	0.51 (0.42–0.61)
CCBC[a]	Utah ICD-9	0.60 (0.52–0.67)	0.99 (0.99–0.99)	80.9 (67.43–97.07)	0.40 (0.33–0.49)
CCBC[d]	ICD-9 list	0.47 (0.45–0.49)	0.99 (0.99–0.99)	65.25 (61.79–68.90)	0.54 (0.52–0.56)
CCBC[e]	Human review of ED reports	0.31 (0.24–0.39)	0.99 (0.98–1.0)	34.01 (18.76–61.68)	0.70 (0.62–0.78)
CCBC *Rash* and Keyword *Fever*[e]	Human review of ED reports for *Febrile rash*	0.12 (0.05–0.27)	1.0 (0.99–1.0)	[h]	0.88 (0.78–1.0)
Hemorrhagic Syndrome					
CCBC[d]	ICD-9 list	0.75 (0.74–0.76)	0.99 (0.98–0.99)	49.01 (47.79–50.25)	0.25 (0.24–0.26)
CCBC[e]	Human review of ED reports	0.39 (0.34–0.44)	0.99 (0.98–0.99)	36.61 (20.96–63.93)	0.62 (0.57–0.68)
CCBC *Hemorrhagic* and Keyword *Fever*[e]	Human review of ED reports for *Febrile hemorrhagic*	0.0 (0–0.07)	1.0 (1.0–1.0)	[h]	1.0 (1.0–1.0)

TABLE 23.6 Performance of Bayesian and other classifiers in detecting syndromes—Cont'd

Classifier Being Tested	Reference Standard for Comparison	Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio (95% CI)	Negative Likelihood Ratio (95% CI)
Botulinic syndrome					
CCBC[a]	UDOH *Botulism-like*	0.17	0.998	104.45	0.83
		(0.05–0.45)	(0.998–0.999)	(28.57–381.86)	(0.64–1.07)
CCBC[a]	Utah ICD-9 list	0.22	0.999	166.94	0.78
		(0.13–0.36)	(0.998–0.999)	(89.07–312.90)	(0.67–0.91)
CCBC[d]	ICD-9 list	0.30	0.99	44.26	0.70
		(0.28–0.32)	(0.99–0.99)	(41.06–47.70)	(0.68–0.72)
CCBC[e]	Human review of ED reports	0.10	0.99	10.96	0.91
		(0.06–0.17)	(0.99–1.0)	(5.11–23.48)	(0.86–0.97)
Fever					
Keyword[g]	Human review of ED reports	0.61	1.0	h	0.39
		(0.51–0.69)	(0.96–1.0)		(0.31–0.49)

[a]From Wagner, M., Espino, J., Tsui, F.-C., et al. (2004). Syndrome and outbreak detection from chief complaints: the experience of the Real-Time Outbreak and Disease Surveillance Project. *MMWR Morb Mortal Wkly Rep* 53(Suppl):28–31, with permission.
[b]From Chapman, W. W., Espino, J. U., Dowling, J. N., et al. (2003). *Detection of Acute Lower Respiratory Syndrome from Chief Complaints and ICD-9 Codes.* Technical Report, CBMI Report Series 2003. Pittsburgh, PA: Center for Biomedical Informatics, University of Pittsburgh, with permission.
[c]From Beitel, A. J., Olson, K. L., Reis, B. Y., et al. (2004). Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Pediatr Emerg Care* 20:355–60, with permission.
[d]From Chapman, W. W., Dowling, J. N., and Wagner, M. M. (2005). Classification of emergency department chief complaints into seven syndromes: a retrospective analysis of 527,228 patients. *Ann Emerg Med* 46(5):445–455.
[e]From Chapman WW, unpublished results.
[f]From Ivanov, O., Wagner, M. M., Chapman, W. W., et al. (2002). Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. In: *Proceedings of American Medical Informatics Association Symposium*, 345–9, with permission.
[g]From Chapman, W. W., Dowling, J. N., Wagner, M. M. (2004). Fever detection from free-text clinical records for biosurveillance. *J Biomed Inform* 2004;120–7, with permission.
[g]Large positive likelihood ratio due to specificity of 0.9999.
[h]Not able to calculate (denominator is zero).
CCBC, Chief complaint Bayesian classifier CI, confidence interval.

positive and negative. The likelihood ratio positive is the purest measure of the informational content of a chief complaint for detecting a syndrome (i.e., its ability to discriminate between a person with the syndrome and one without the syndrome). In a Bayesian analysis, it is a number that indicates the degree to which a system should update its belief that a patient has the syndrome, given the chief complaint (see Chapter 13).

The gold standard used in these studies varied. The most valid standard used was classification based on review of patients' ED reports using random selection of patients. The earliest studies evaluating the ability of chief complaints to identify syndromes were able to use this gold standard because they studied common syndromes, such as *respiratory* (Chapman et al., 2003, Beitel et al., 2004) or *gastrointestinal* (Ivanov et al., 2002). When a syndrome is common, a pool of randomly selected patients will produce a sufficient sample of actual respiratory cases.

Later studies examined less common syndromes. To obtain a sufficient sample of patients with uncommon syndromes, researchers searched ICD-9 discharge diagnoses to find cases (Wagner et al., 2004, Chapman et al., 2005b, Mundorff et al., 2004). Using a patient's discharge diagnosis as the gold standard enabled these studies to acquire large numbers of patients—even for rare syndromes, such as *botulinic*. Chart review, however, probably provides more accurate gold standard classifications than ICD-9 codes (Chang et al., 2005).

A few recent studies have used chart review as the gold standard for evaluating a variety of syndromes, including syndromes of low prevalence (Chang et al., 2005, Chapman et al., 2005c). One study compared chief complaint classification during the 2002 Winter Olympic Games against gold-standard classification of potentially positive cases selected by Utah Department of Health employees who performed drop-in surveillance (Wagner et al., 2004).

Chapman et al. (2005c) used ICD-9 searching to find a set of patients with discharge diagnoses of concern in biosurveillance. Physicians then reviewed ED reports for each of the cases to finalize a reference syndrome assignment. Using ICD-9 codes to select patients made it possible to use chart review on a fairly small sample of patients while still acquiring a reasonably sized set of patients for seven different syndromes.

An important issue is whether the same classification accuracy observed in a study of chief complaints from hospital X will be observed for chief complaints from hospital Y. Levy et al. (2005) showed that classification accuracy of a keyword-based parser differed from hospital to hospital for *gastrointestinal* syndrome. Chapman et al. (2005b), however, showed that the classification accuracy of a Bayesian chief complaint classifier was no different when it was used on a set of chief complaints from a geographic region other than the one that it had been trained on.

There are a number of studies in the literature that we did not include in Table 23.6 because they measured the sensitivity and specificity of an NLP program's syndrome assignment *relative to a physician who is classifying a patient only from the chief complaint* (Chapman et al., 2005a, Olszewski, 2003a, Sniegoski, 2004). These studies report much higher sensitivities and specificities than those in Table 23.6. These studies represent formative studies of NLP algorithms. The accuracy of syndrome classification should always be measured relative to the actual syndrome of the patient as determined by a method at least as rigorous as medical record review or discharge diagnoses when accepting or rejecting Hypothesis 1 for a syndrome under study.

In summary, the experiments in Table 23.6, although somewhat heterogeneous methodologically, are similar enough to be considered meta-analytically. They made the same measurements (sensitivity and specificity), studied similar syndromes, used simple techniques for classifying chief complaints into syndromic categories, and used similar gold standards.

With respect to Hypothesis 1, these experiments demonstrate that:

1. Chief complaint data contain information about syndromic presentations of patients and various NLP techniques including a naïve Bayesian classifier and keyword methods can extract that information.

2. For syndromes that are at the level of diagnostic precision of *respiratory* or *gastrointestinal* it is possible to automatically classify ED patients (both pediatric and adult) from chief complaints with a sensitivity of approximately 0.60 and a specificity of approximately 0.95.

3. Sensitivity of classification is better for some syndromes than for others.

4. When syndromes are more diagnostically precise (e.g., *respiratory with fever*), the discrimination ability declines quickly.

5. The specificity of syndrome classification from chief complaints is less than 100%, meaning that daily aggregate counts will have false positives among them due to falsely classified patients.[6]

2.3.2. Detection of Outbreaks

This section describes studies that address Hypotheses 2 and 3, which we reproduce here for convenient reference:

Hypothesis 2: When aggregated with the chief complaints of other patients in a region, chief complaints can discriminate between whether there is an *outbreak* of type Y or not.

Hypothesis 3: When aggregated with the chief complaints of other patients in a region, algorithmic monitoring of chief complaints can detect an *outbreak* of type Y earlier than current best practice (or some other reference method).

As discussed in Chapter 20, studies of *outbreak* detection are more difficult to conduct than studies of syndrome (case) detection. These studies require chief complaint data collected from multiple healthcare facilities in a region affected by an outbreak. Research groups often expend significant time and effort building biosurveillance systems to collect the chief complaint data needed to conduct this type of research. Because outbreaks are rare events, many of the studies we will discuss are of common diseases such as *influenza* or of seasonal outbreaks (winter) that may be caused by multiple organisms.

Adequate sample size (of outbreaks) is difficult to achieve in studies of outbreak detection. Only one of the studies computed a confidence interval on its measurements of sensitivity or time of outbreak detection. The scientific importance of adequate sample size cannot be overstated. There are two possible approaches to increasing sample size: (1) conduct research in biosurveillance systems that span sufficiently large geographic regions that they are expected to encounter sufficient numbers of outbreaks, or (2) meta-analysis. To emphasize the importance of adequate sample size, we divide this section into studies of multiple outbreaks (labeled N>1), prospective studies, and studies of single outbreaks (N=1).

N>1 Studies. Ivanov and colleagues used the detection-algorithm method and correlation analysis to study detection of six seasonal outbreaks in children from CoCo classifications of patients into *respiratory* and *gastrointestinal* based on chief complaints (Ivanov et al., 2003). They studied the daily visits to ED of a pediatric hospital during annual winter outbreaks due to diseases such as rotavirus, gastroenteritis and influenza for the four-year period 1998–2001.

The researchers identified outbreaks for study using the following procedure: They created two ICD-9 code sets, corresponding to infectious diseases of children that are respiratory or gastrointestinal, and used them to create two reference time series from ICD9-coded hospital discharge diagnoses. Figure 23.3 (top) from the publication is a plot of the daily *respiratory* time series and the reference time series of respiratory disease created from hospital discharge diagnoses of children under age five. Figure 23.3 (bottom) is a similar comparison of *gastrointestinal* time series and the reference time series of infectious gastrointestinal conditions.

The detection algorithm (exponentially weighted moving average) identified three respiratory and three gastrointestinal outbreaks in the hospital discharge data (the reference standard

---

6 Even were the specificity to be 100%, the diagnostic precision of the syndrome categories studied is low. As a result, the syndromes used in research have many causes other than acute infectious diseases, which are the diseases of interest in biosurveillance. When monitoring aggregate daily counts, even a syndrome with 100% specificity will show baseline levels of counts in the absence of an outbreak due to the presence of these chronic and sporadic conditions.
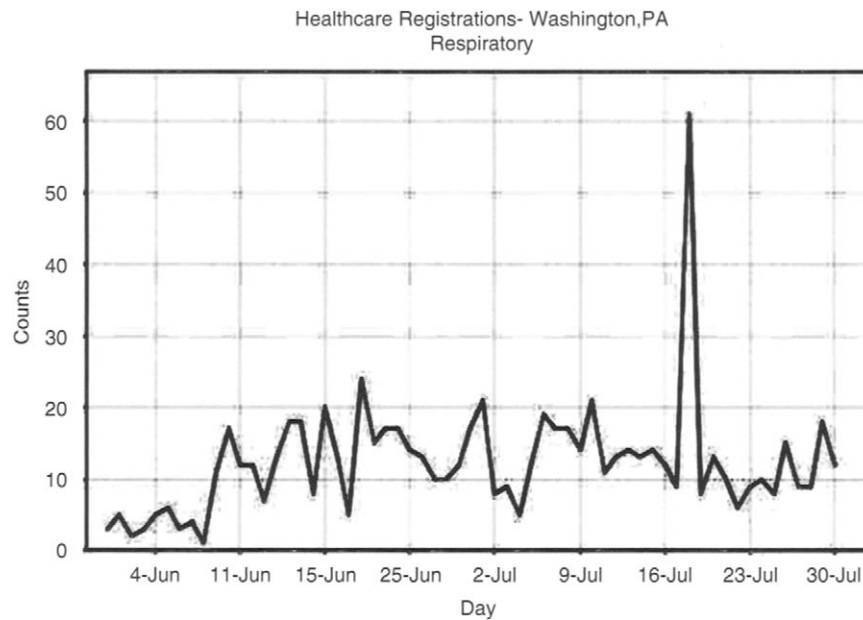
FIGURE 23.4 Daily counts of *respiratory* cases, Washington County, Pennsylvania, June–July 2003. The small increase in early June 2003 corresponds to new hospitals being added to the surveillance system.

for outbreaks). The detection from chief complaints preceded detection from automatic analysis of hospital discharge diagnoses by a mean of 10.3 days (95% confidence interval [CI], 15.15–35.5) for respiratory outbreaks and 29 days (95% CI, 4.23–53.7) for gastrointestinal outbreaks (Table 23.7). The researchers used the date of admission rather than the date of discharge in constructing the reference time series.

The correlation analysis of three respiratory outbreaks showed that on average the chief complaint time series was 7.4 days earlier (95% CI, 8.34–43.3), although the 95% CI included zero. For the three gastrointestinal outbreaks, the chief complaint time series was 17.6 days earlier (95% CI, 3.4–46.7).

Prospective Studies. Prospective studies are field evaluations of a biosurveillance system. In a prospective evaluation, the detection algorithms are operated at a fixed detection threshold for an extended period and the ability of the biosurveillance system to detect known outbreaks or to identify new outbreaks is measured.

Heffernan et al. (2004b) used the detection-algorithm method prospectively to study *respiratory and fever* syndrome monitoring in New York City (Heffernan et al., 2004b). They studied the New York City Department of Health and Mental Hygiene (DOHMH) syndromic system for the one-year period

November 2001-November 2002. Note they also report the DOHMH one-year experience monitoring *diarrhea* and *vomiting*, however, the paper by Balter, which we discuss next, included that year in a three-year analysis, so we do not discuss it here.

In New York City, EDs transmit chief complaints to the DOHMH on a daily basis as email attachments or via FTP. The researchers estimated that the DOHMH system received chief complaint data for approximately 75% of ED visits in New York City. The NLP program was a keyword-based system that assigned each patient to exactly one syndrome from the set: *common cold, sepsis/dead on arrival, respiratory, diarrhea, fever, rash, asthma, vomiting,* and *other* (Table 23.4). The NLP program was greedy, which means that the algorithm assigned a patient to the first syndrome from the list of syndromes whose definition was satisfied and did not attempt further assignment.

DOHMH used the detection-algorithm method to identify potential outbreaks from daily counts of *respiratory* and *fever*. They used a univariate detection algorithm on data aggregated for the entire city (*citywide*), and spatial scanning for data aggregated by patient home zip code and by hospital (separate analyses).

The citywide monitoring of *respiratory* found 22 above-threshold anomalies (called *signals*), of which the researchers

TABLE 23.7 Detection Algorithm Analysis of Timeliness of Detection from Chief Complaints

Syndrome	Gold Standard Outbreak	Sensitivity	Specificity	Timeliness (95% CI)
Respiratory	Seasonal outbreaks of pediatric respiratory illness (bronchiolitis, P&I)	100%	100%	10 days (−15–35)
Gastrointestinal	Seasonal outbreaks of pediatric gastrointestinal illness (rotavirus gastroenteritis)	100%	100%	29 days (4–53)

CI, confidence interval; P&I, pneumonia and influenza.

stated that 14 (64%) *occurred during periods of peak influenza activity.* The first citywide signal occurred in December 2001 and it was followed by additional signals in both *respiratory* and *fever* signals on the six successive days. The authors commented that these signals coincided with a sharp increase in positive influenza test results, but did not report a correlation analysis. They also commented that the reports of influenza-like illness (ILI) from the existing sentinel physician ILI system showed increases three weeks after the first signal. Three other *respiratory* signals occurred during periods of known increases in asthma activity. The remaining five signals occurred during periods of increasing visits for respiratory disease. Thus, there were no signals that could not be attributed to known disease activity.

The citywide monitoring of *fever* generated 22 signals, of which 21 (95%) occurred during periods of peak influenza activity.

The hospital monitoring of *respiratory* and *fever* produced 25 signals. The home zip code monitoring of these two syndromes produced 18 signals. Investigations of these 43 (25+18) signals found no associated increases in disease activity.

Balter and colleagues analyzed the DOHMH three-year experience (November 2001–August 2004) monitoring *diarrhea* and *vomiting* using the same biosurveillance system as described in the previous paragraphs (Balter et al., 2005). The authors estimate that by the end of the study period, the monitoring system received data for approximately 90% of ED visits in New York City.

During the three years, the DOHMH system signaled 236 times (98 citywide and 138 hospital or zip code) for *diarrhea* or *vomiting*. Of 98 citywide signals, 73 (75%) occurred during what the authors referred to as "seasonal" outbreaks likely due to norovirus (fall and winter) and rotavirus (spring). One citywide signal after the August 2003 blackout was *believed to have represented a true increase in diarrheal illness.* Their investigations of the 138 hospital or zip code signals found no increased disease activity.

During the same period, DOHMH investigated 49 GI outbreaks involving ten or more cases; none of which were detected by monitoring of *diarrhea* or *vomiting.* In 36 of these outbreaks, few or no patients went to EDs. In two outbreaks, the victims were visitors to New York City who returned to their homes before onset of symptoms. In three outbreaks, victims visited EDs not participating in the monitoring system. In three outbreaks, victims visited EDs over a "days or weeks" (the algorithms used by DOHMH were sensitive to rapid increases, not gradual increases in daily counts of syndromes). In two outbreaks, the victims presented to the ED as a group and their chief complaints were recorded by reference to the group (e.g., "school incident"). In two outbreaks, *a combination of the above causes explained the failure.*

N=1 Studies. Irvin and colleagues (Irvin et al., 2003) used the detection-algorithm method to retrospectively study the

ability of their *anthrax* syndrome to detect a single influenza outbreak. The paper is not explicit about the *anthrax* syndrome, but states, "The presence of any of the following symptoms were sufficient to categorize a patient into *anthrax*: cough, dyspnea, fever, lethargy, pleuritic chest pain, vomiting, generalized abdominal pain, or headache," suggesting that the researcher included symptoms with which pulmonary anthrax may present. They studied an atypical monitoring system based on numeric chief-complaint codes from a commercial ED charting system. This charting system, called E/Map (Lynx Medical Systems, Bellevue WA, *http://www.lynxmed.com*), offers clinicians charting templates for approximately 800 chief complaints. Each template has a numerical code. A clinician's selection of charting template reflects the patient's chief complaint. The detection algorithm used a fixed detection threshold set at two standard deviations from a recent two-month mean. The algorithm signaled when two of the previous three days exceeded the threshold. The reference standard was the Centers for Disease Control and Prevention (CDC) defined peak week of influenza activity. The system signaled one week prior to the CDC peak and signaled one false positive.

Yuan et al. (2004) used the detection-algorithm method to study the timeliness of detection of one influenza outbreak in southeastern Virginia. They manually assigned chief complaints to seven syndromes (*fever, respiratory distress, vomiting, diarrhea, rash, disorientation,* and *sepsis*). The detection algorithm was CUSUM, operated at three different moving averages (7-day window, window days 3–9, and 3-day window) and set at a threshold of 3 S.D. They reported that the CUSUM algorithm detected trends in *fever* and *respiratory* in one hospital that preceded the local sentinel influenza surveillance system by one week.

A key limitation of N=1 studies is that any correlation found may be spurious. Meta analysis could address this problem if differences among analytic and reporting methods used by studies were reduced so that studies of single outbreaks could be merged analytically. In 2003, the RODS Laboratory developed a case-study series that encourages the use of a standard method of studying single outbreaks that would enable the application of uniform analytic methods across outbreaks (or alerts) occurring in different regions (Rizzo et al., 2005). The objectives of the case report series are to: (1) ensure complete description of outbreak and analytic methods, and (2) collect the raw surveillance data and information about the outbreak in a way that future re-analyses are possible.

Each case study describes the effect of a single outbreak or other public health event, such as low air quality due to forest fires, on surveillance of data available for the event. At present, these case studies are available only to authorized public health users of the NRDM system because of legal agreements with organizations that provide surveillance data (employees of governmental public health organizations can

access case studies through the RODS interface by sending e-mail to nrdmaccounts@cbmi.pitt.edu).

Of the 15 case studies developed to date, eight are examples of outbreaks considered "detectable" from available surveillance data; six were not detectable. These case studies include outbreaks of influenza, salmonella, norovirus, rotavirus, shigella, and hepatitis A. One case study describes a false alarm investigation that resulted from a retailer recording error.

Figure 23.4 is taken from a case study of a large spike in CoCo *respiratory* cases in a single county outside Pittsburgh that resulted in an alert being sent automatically on Friday July 18, 2003 at 8 PM to an on-call epidemiologist. Normally, daily counts of respiratory cases numbered 10, but on that day they numbered 60 by 8 PM. The epidemiologist logged into the RODS web interface, reviewed the verbatim chief complaints of affected patients and discovered that the cases were related to carbon-monoxide exposure, which a phone call to an ED revealed to be related to a faulty furnace at a day-care center.

The case studies include three studies of the effect of influenza on emergency room visits for the CoCo *constitutional* and *respiratory* syndromes. Figure 23.5 illustrates the size of the influenza effect in 2003–2004 in Utah (middle spike) on *constitutional*, and *respiratory*, as well as sales of thermometers by pharmacies participating in the National Retail Data Monitor.

These case studies add to the previously described studies the following: Influenza has a strong early effect on free text chief complaints in the *constitutional* and *respiratory* categories. Air pollution and small carbon monoxide events may have marked effects on chief complaints in the *respiratory* category. The results for gastrointestinal outbreaks have been negative for relatively small, protracted outbreaks of Norovirus and Shigella.

Summary of Studies of Outbreak Detection from Chief Complaints. With respect to Hypotheses 2 and 3, the studies we reviewed demonstrate that:

1. Some large outbreaks causing respiratory, constitutional, or gastrointestinal symptoms can be detected from aggregate analysis of chief complaints. Small outbreaks of gastrointestinal illness generally cannot (Hypothesis 2).

2. Research to date is suggestive but not conclusive that influenza can be detected earlier by chief complaint monitoring than current best practice (Hypothesis 3).
3. The false-alarm rates associated with such monitoring can be low. In New York City for city-wide monitoring of *respiratory* and *fever*, there were few signals that did not correspond to disease activity. There were more signals that did not correspond to disease activity from monitoring of *diarrhea* and *vomiting*. Conversely, all of the signals from spatial monitoring of hospital or zip code were not correlated with known disease activity.
4. The methodological weaknesses in the studies included failure to describe or measure time latencies involved in data collection. Some studies did not report sampling completeness, the method by which chief complaints are parsed, or details of the syndrome categories.
5. In general, the number of published studies is small, perhaps due to the fact that chief complaint monitoring systems are still being constructed. We expect more studies to be published in the near future.

The answers to Hypotheses 2 and 3 for surveillance of chief complaints in isolation may not be as important long term as the question of whether chief complaints contain diagnostic information (Hypothesis 1). The reason is that chief complaints can be used with other surveillance data to detect outbreaks, either through linking at the level of the individual patient or as a second source of evidence. Nevertheless, because of their availability and earliness, and the threat of bioterrorism and large outbreaks, it is important to understand the ability to detect outbreaks solely from this type of data.

3. ICD CODES

The International Classification of Diseases, 9th Revision, Clinical Modification (ICD) is a standard vocabulary for diagnoses, symptoms, and syndromes (see Chapter 32). ICD has a code for each class of diagnoses, syndromes, and symptoms that it covers. For example, the ICD code 034.0 *Streptococcal sore throat* includes tonsillitis, pharyngitis, and laryngitis caused by any species of *Streptococcus* bacteria. There are more than
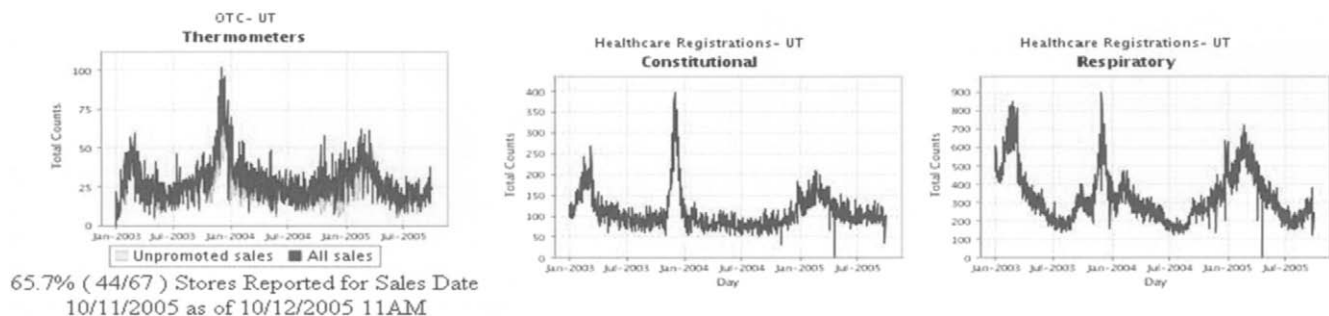

65.7% (44/67) Stores Reported for Sales Date
10/11/2005 as of 10/12/2005 11AM

FIGURE 23.5 Daily counts of *constitutional* and *respiratory* syndrome. January 2003–September 2005. The largest spikes correspond to the 2003–2004 influenza outbreak, which was more severe (involving more people) than the previous or following year's outbreak.

12,000 ICD codes. Internationally, some countries use the 10th revision of the International Classification of Diseases, or a modification of it.

Data encoded using ICD are widely available in the United States. Most visits to physicians or other healthcare providers and hospitalizations result in one or more ICD codes. The reason is that healthcare insurance corporations require providers of care to use ICD codes when submitting insurance claims to receive reimbursement for their services.

ICD codes range in diagnostic precision from the very imprecise level of symptoms to very precise diagnoses. There are precise codes for infectious diseases, specifying both the causative organism and the disease process (e.g., 481 *Pneumococcal pneumonia*). However, there are less precise codes that providers can use if the organism is unknown or not documented (e.g., 486 *Pneumonia, organism unspecified*). There are also ICD codes for syndromes (e.g., 079.99 is the code for *Viral syndrome*) and even for symptoms (e.g., 786.2 *Cough* and 780.6 *Fever*).

ICD codes may be assigned at different times during the course of care (Figure 23.1). As you go from left to right in Figure 23.1, who assigns the ICD code, how, and when vary. Physicians, when they do assign ICD codes to office or ED visits, usually do so during or within hours to days of the visit. They either enter ICD codes into a point-of-care system or record them on an encounter form (also sometimes known as a "superbill"). Professional coders usually assign the final, billing ICD codes for ED and office visits days later. They also assign ICD codes to hospital discharge diagnoses, typically days to weeks after the patient leaves the hospital. Professional coders often enter ICD codes into specialized billing software. ICD-coded data from organizations that collect large volumes of insurance claims data (we discuss these "data aggregators" in more detail below) are usually not available for months after visits or hospital stays.

The diagnostic precision of ICD codes generally increases with time, as you go from left to right in Figure 23.1. The reason that discharge diagnoses generally have higher diagnostic precision relative to visit diagnoses is that discharge diagnoses typically represent the outcome of a greater amount of diagnostic testing that leads to greater diagnostic certainty (i.e., providers are more likely to order—and have the results available from— laboratory tests, microbiology cultures, x-rays, and so on).

Health services researchers have established that the accuracy of ICD-coded data is highly variable and often only moderately high (O'Malley K et al., 2005, Hsia et al., 1988). They have identified several causes for inaccuracy (Peabody et al., 2004, O'Malley et al., 2005). One cause is that two different, highly trained, experienced coders may assign different codes to the same hospitalization (Fisher et al., 1992, Lloyd and Rissing, 1985, MacIntyre et al., 1997). One reason is that coders work from the patient chart, which is an imperfect

representation of the patient's true medical history and is subject to variable interpretation. Professional coders are typically not physicians or nurses, so their level of understanding of the medical process is imperfect. Finally, the rules for assigning codes are complex and change at least annually.[7]

The problem of correct assignment of ICD codes is compounded when clinicians encode the diagnoses (Yao et al., 1999). Clinicians rarely have formal training in the rules for assigning codes. They typically have little time to ensure that the codes they assign are accurate. They often view the assignment of ICD codes to patient encounters as a distraction from patient care. To address these problems, physicians often use preprinted encounter forms that have check boxes for an extremely small subset of commonly used codes. Although these forms typically include a blank space to write in additional ICD codes, clinicians are extremely busy so an open question is how often they use the space and how accurate are the ICD codes that are hand entered. Another question is whether busy clinicians, who do not use the data on the encounter form for subsequent patient care, completely code all diagnoses made during a patient visit. One study found that, during patient visits, physicians addressed an average 3.05 patient problems but documented only 1.97 on billing forms (they documented nearly as many problems in the paper record as they addressed) (Beasley et al., 2004).

ICD codes, because of their inaccuracy and the fact that their primary use and purpose is billing, are likely to be less than ideal when used for other purposes. One study found low accuracy of billing data about cardiac diseases relative to a clinical research database (Jollis et al., 1993). Another study found that one-third of patients who received an ICD code that indicated the presence of a notifiable disease did not truly have the notifiable disease (Campos-Outcalt, 1990). A third study found that data about prescriptions identified patients with tuberculosis more accurately than all 60 ICD codes for tuberculosis combined (Yokoe et al., 1999).

ICD codes might be less ideal for biosurveillance than other coding systems such as SNOMED-CT. The designers of ICD did not design it with biosurveillance requirements in mind. One study found that SNOMED-CT was superior to ICD for coding ED chief complaints (McClay and Campbell, 2002). SNOMED-CT had a term that was a precise match for 93% of chief complaints; ICD had a precise match for only 40% of chief complaints.

In summary, billing ICD codes from insurance claims and hospital discharge data sets are widely available, but at long time latencies (weeks to months). ICD codes at shorter time latencies ICD (within 24 hours of ED or office visit) are less available. Who assigns ICD codes and when and how influence the time latency, diagnostic precision, and accuracy of ICD-coded data. Thus, it is essential that studies describe the process that generated the ICD codes and measure time latency.

---

7 Hence the need for professionals to do the coding in the first place.

3.1. Categories of ICD Codes ("Code Sets")

Despite the potential for high diagnostic precision, biosurveillance researchers and developers group ICD codes into categories ("code sets") such as "respiratory illness." The set of all 60 ICD codes for tuberculosis mentioned above is an example of an ICD code set. It is not necessary to group ICD codes into codes sets, although it is almost always done. For example, we could monitor for the single ICD code for inhalational anthrax (022.1). Creators of code sets usually group codes of diseases and syndromes that share similar early presentations to form syndrome code sets. *Respiratory*, *gastrointestinal*, *neurological*, *rash* and *febrile illnesses* are representative of code sets in common use.

A key reason that developers create code sets is to improve the sensitivity of case detection, because patients with the same disease may be assigned different ICD codes. This variability may be due to variability in how coders assign codes or that the patients are at different stages in their diagnostic work-ups. For example, a patient with influenza who has not yet undergone definitive testing may be coded as 780.6 *Fever* (or any of a number of other ICD codes for symptoms of influenza), 079.99 *viral syndrome*, 465.9 *Acute upper respiratory infection NOS*, or 486 *Pneumonia, organism unspecified*.

The most difficult and "art-more-than-science" aspect of ICD-code monitoring is development of code sets. The next sections describe several code sets used in biosurveillance. Our purpose is to illustrate how a code set is developed. It is important to note that one code set could be superior (i.e., have better case detection and/or outbreak detection performance) to others for the detection of one disease (e.g., influenza), but inferior to other code sets for the detection of another disease (e.g., bronchiolitis due to respiratory syncytial virus). To date, only one study (Tsui et al., 2001) has compared the accuracy of two alternative code sets to determine their differential ability to detect the same set of cases or outbreaks. That study lacks generalizability because no other research groups have found the data they studied—ICD codes for chief complaints assigned by registration clerks—to be available at other institutions.[8] Therefore, which code sets are better than others for the detection of various outbreaks such as influenza or cryptosporidiosis remains unknown.

3.1.1. ESSENCE Code Sets

The Department of Defense (DoD) developed the first ICD code sets for use in biosurveillance as part of its Global Emerging Infections System (DoD-GEIS). The ESSENCE biosurveillance system[9] uses these code sets (DoD-GEIS, 2005) to aggregate ambulatory visits at DoD outpatient clinics into seven syndromes (Table 23.8). A number of researchers and system developers have used these code sets,

TABLE 23.8 The Seven ESSENCE ICD Code Sets

Code Set	Number of ICD Codes in Set
Coma/Shock/Death	10
Dermatologic–Hemorrhagic	9
Dermatologic–Infectious	18
Fever/Malaise/Sepsis	26
Gastrointestinal	118
Neurologic	72
Respiratory	175

Note: The code sets are available at the U.S. Department of Defense, Global Emerging Infections System website, http://www.geis.fhp.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE. asp#ICD9.

or slight modifications of them, in their work (Lazarus et al., 2002, Reis and Mandl, 2004, Lazarus et al., 2001, Lewis et al., 2002, Mocny et al., 2003, Magruder et al., 2005, Buckeridge et al., 2005, Henry et al., 2004).

3.1.2. CDC/DoD Categories

Representatives from the CDC and other stakeholders formed a working group to develop a suggested list of ICD code sets (CDC, 2003). These representatives reviewed all 12,000 ICD codes for possible inclusion in one of twelve ICD code sets (Table 23.9). They also looked at the frequency of code usage in both a DoD outpatient visit data set and a civilian ED data set. Based on the frequency of code usage and their knowledge of symptoms associated with diseases represented by ICD codes, they assigned ICD codes to syndrome code sets. They further specified three levels of membership in the categories, which they somewhat confusingly labeled *categories 1, 2,* and *3*:

> *Category 1:* ...codes that reflect general symptoms of the syndrome group and also ... the bioterrorism diseases of highest concern or those diseases highly approximating them.

TABLE 23.9 CDC/DoD ICD Code Sets

Code Set	Number of ICD Codes in Set
Botulism-like	41
Fever	87
Gastrointestinal, All	125
Gastrointestinal, Lower	108
Gastrointestinal, Upper	17
Hemorrhagic Illness	33
Lesion	71
Lymphadenitis	26
Neurological	95
Rash	44
Respiratory	171
Severe Illness or Death	14

Note: The code sets are available at the Centers for Disease Control and Prevention website, http://www.bt.cdc.gov/surveillance/syndromedef/index.asp. DoD, U.S. Department of Defense.

---

8 And even the health system that Tsui and colleagues studied no longer collects this data.

9 ESSENCE was first part of DoD-GEIS and now civilian health departments in a number of jurisdictions use it also.

Category 2: ...codes that might normally be placed in the syndrome group, but daily volume could overwhelm or otherwise detract from the signal generated from the Category 1 code set alone.

Category 3: ...specific diagnoses that fit into the syndrome category but occur infrequently or have very few counts. These codes may be excluded to simplify syndrome category code sets.

The end result was 12 ICD code sets, each with three categories of codes. The code sets are not mutually exclusive (some ICD codes appear in more than one code set). The report of the working group does not describe the procedure they followed to reach a consensus on—or resolve conflicts over—which ICD codes to include in a code set.

This effort made use of intensional definitions for each code set (Table 23.10). The extensional definition of the *Fever* ICD code set is given in Table 23.11.

As with the ESSENCE code sets, researchers and system developers have used the CDC/DoD code sets in their work (Yih et al., 2004, 2005).

3.1.3. Other Code Sets

No other code sets other than the DoD-GEIS and CDC/DoD code sets appear to be in use in a functioning biosurveillance system. At present, DoD-GEIS code sets are in use in the ESSENCE biosurveillance system (Lombardo et al., 2003) and

TABLE 23.10 Two Examples of Intensional Definitions of CDC/DoD ICD Code Sets

Syndrome	Definition	Category A Condition
Botulism-like	ACUTE condition that may represent exposure to botulinum toxin	Botulism
	ACUTE paralytic conditions consistent with botulism: cranial nerve VI (lateral rectus) palsy, ptosis, dilated pupils, decreased gag reflex, media rectus palsy	
	ACUTE descending motor paralysis (including muscles of respiration)	
	ACUTE symptoms consistent with botulism: diplopia, dry mouth, dysphagia, difficulty focusing to a near point	
Fever	ACUTE potentially febrile illness of origin not specified	Not applicable
	INCLUDES fever and septicemia not otherwise specified	
	INCLUDES unspecified viral illness even though unknown if fever is present	
	EXCLUDE entry in this syndrome category if more specific diagnostic code is present allowing same patient visit to be categorized as respiratory, neurological or gastrointestinal illness syndrome	

CDC, Centers for Disease Control and Prevention; DoD, U.S. Department of Defense.

the CDC/DoD code sets are in use in the National Bioterrorism Syndromic Surveillance Demonstration Program (Yih et al., 2004). The other code sets that researchers describe in the literature were created for research studies.

For example, Lazarus et al. (2002) used a slight modification of ESSENCE code sets and also created an *influenza-like illness (ILI)* code set. They based this code set on the CDC sentinel surveillance definition of *ILI*, which is fever (body temperature measured in the office >37.8°C) plus cough or sore throat without a known cause. Miller and colleagues created an *ILI* code set that includes 31 ICD codes (Miller et al., 2004). Most of the codes in the set are symptoms of influenza. Ritzwoller et al. (2005) created an *ILI* code set. This code set included codes for respiratory illness plus fever (780.6). Espino and Wagner (2001a) created a *respiratory* code set that includes 64 ICD codes. To create this code set, experts reviewed all ICD codes assigned as ED chief complaints over a three-year period. Tsui et al. (2001) created an *influenza* code set from this *respiratory* code set and viral illness ICD codes (Tsui et al., 2001). Ivanov created a *gastrointestinal* code set that includes 16 ICD codes (Ivanov et al., 2002). As with Espino and Wagner (2001a), experts reviewed all ICD codes assigned as ED chief complaints over a three-year period. Beitel et al. (2004) created *respiratory, lower respiratory*, and *upper respiratory* code sets. To create the *respiratory* code set, they merged ICD codes 460.00–519.xx (encompasses all "diseases of the respiratory system"), 786.xx (encompasses all "symptoms involving the respiratory system and other chest symptoms"), and the ESSENCE *respiratory* code set except ICD code 079.99 (*Viral syndrome*). They then divided *respiratory* into *lower respiratory* and *upper respiratory* based on the part of the respiratory system implied by the code. When a code did not imply any anatomy or implied anatomy that spanned the upper and lower parts of the respiratory system, they added it to the *upper respiratory* set.

3.1.4. Standardizing Code Sets

We note that some authors have referred to either the DoD-GEIS or the CDC/DoD code sets as "standard" ICD code sets (Heffernan et al., 2004a, Mandl et al., 2004). However it would be extremely premature to standardize on these code sets other than for research reporting purposes. Given the variability in diseases, ICD assignment by clinicians and coders, and diagnostic precision, it is unlikely that a one-size-fits-all ICD code set will be optimal for all settings and purposes. Moreover, it is trivial for computer systems to aggregate raw ICD-coded data using any desired code set in real time and as needed.[10]

We note that identifying an optimal code set is not trivial. A computer scientist would, in fact, call it an intractable problem, which means that it cannot be done. The reason: there are

---

10 A single microchip cannibalized from a discarded DVD player likely has sufficient processing power to keep up with all the ICD-to-code set translations required for biosurveillance by the entire world.

TABLE 23.11 Extensional Definition of the CDC/DoD *Fever* ICD Code Set

ICD9CM	ICD9DESCR	Consensus	ICD9CM	ICD9DESCR	Consensus
020.2	PLAGUE, SEPTICEMIC	1	054.5	HERPETIC SEPTICEMIA	3
020.8	OTHER TYPES OF PLAGUE	1	060.0	YELLOW FEVER, SYLVATIC	3
020.9	PLAGUE NOS	1	060.1	YELLOW FEVER, URBAN	3
021.8	TULAREMIA NEC	1	060.9	YELLOW FEVER, UNSPEC	3
021.9	TULAREMIA NOS	1	066.0	PHLEBOTOMUS FEVER	3
022.3	ANTHRAX, SEPTICEMIA	1	066.1	TICK-BORNE FEVER	3
022.8	ANTHRAX, OTHER SPECIFIED	1	066.2	VENEZUELAN EQUINE FEVER	3
022.9	ANTHRAX, UNSPECIFIED	1	066.3	MOSQUITO-BORNE FEVER NEC	3
038.3	ANAEROBES SEPTICEMIA	1	066.8	ARTHROPOD VIRUS NEC	3
038.40	GRAM-NEGATIVE ORGANISM UN	1	066.9	ARTHROPOD VIRUS NOS	3
038.49	SEPTICEMIA, OTHER GRAM-NEG	1	078.2	SWEATING FEVER	3
038.8	SEPTICEMIAS, OTHER SPECIF	1	080	LOUSE-BORNE TYPHUS	3
038.9	SEPTICEMIA, NOS	1	081.0	MURINE TYPHUS	3
079.89	VIRAL INFECTION OTHER S	1	081.1	BRILL'S DISEASE	3
079.99	VIRAL INFECTIONS UNSPECIFIED	1	081.2	SCRUB TYPHUS	3
780.31	FEBRILE CONVULSIONS	1	081.9	TYPHUS NOS	3
780.6	FEVER	1	082.8	TICK-BORNE RICKETTS NEC	3
790.7	BACTEREMIA	1	082.9	TICK-BORNE RICKETTS NOS	3
790.8	VIREMIA NOS	1	083.0	Q FEVER	3
002.0	TYPHOID FEVER	3	083.1	TRENCH FEVER	3
002.1	PARATYPHOID FEVER A	3	083.2	RICKETTSIALPOX	3
002.2	PARATYPHOID FEVER B	3	083.8	RICKETTSIOSES NEC	3
002.3	PARATYPHOID FEVER C	3	083.9	RICKETTSIOSIS	3
002.9	PARATYPHOID FEVER NOS	3	084.0	MALARIA, FALCIPARUM	3
003.1	SALMONELLA SEPTICEMIA	3	084.1	VIVAX MALARIA	3
023.0	BRUCELLA MELITENSIS	3	084.2	QUARTAN MALARIA	3
023.1	BRUCELLA ABORTUS	3	084.3	OVALE MALARIA	3
023.2	BRUCELLA SUIS	3	084.5	MIXED MALARIA	3
023.3	BRUCELLA CANIS	3	084.6	MALARIA UNSPECIFIED	3
023.8	BRUCELLOSIS NEC	3	086.2	CHAGA'S DISEASE WITHOUT	3
023.9	BRUCELLOSIS, UNSPECIFIED	3		MENTION OF ORG	
024	GLANDERS	3	086.3	TRYPANOSOMIASIS, GAMBIAN	3
025	MELOIDOSIS	3	086.4	TRYPANOSOMIASIS, RHODESIAN	3
027.0	LISTERIOSIS	3	086.5	TRYPANOSOMIASIS, AFRICAN	3
034.1	SCARLET FEVER	3	086.9	TRYPANOSOMIASIS, UNSPEC	3
038.0	SEPTICEMIA STAPHYLOCOCCAL	3	087.0	LOUSE-BORNE RELAPS FEVER	3
038.10	SEPTICEMIA STAPHYLOCOCCAL	3	087.1	TICK-BORNE RELAPS FEVER	3
038.11	SEPTICEMIA STAPHYLOC. AUR	3	087.9	RELAPSING FEVER NOS	3
038.19	SEPTICEMIA STAPHYLOCOCCAL	3	088.0	BARTONELLOSIS	3
038.2	PNEUMOCOCCAL SEPTICEMIA	3	088.81	LYME DISEASE	3
038.41	SEPTICEMIA, HEMOPHILUS INFLUENZAE	3	088.82	BABESIOSIS	3
038.42	SEPTICEMIA, E. COLI	3	088.89	OTHER ARTHROPOD-BORNE	3
038.43	PSEUDOMONAS	3	088.9	ARTHROPOD-BORNE DIS NOS	3
038.44	SEPTICEMIA SERRATIA	3	100.82	LEPTOSPIROSIS, OTHER	3

CDC, Centers for Disease Control and Prevention; DoD, U.S. Department of Defense.

approximately 12,000 ICD-9 codes and therefore $2^{12,000}$ possible ICD code sets. This number is far greater than the total number of atoms in the universe or the number of milliseconds since the beginning of time. Thus, it is not possible for a computer, let alone a human, to try every code set to find the best one. That is not to say that it is not possible to determine whether Code Set A is superior to Code Set B for some specific biosurveillance application, just that it is not possible to establish that Code Set A is superior to every other code set for all ICD-coded data.

3.2. Availability and Time Latencies of ICD-coded Diagnoses

The availability of ICD codes, by which we mean the proportion of patients being seen in a region for which ICD codes are available, in general is poorly understood. ICD coding of chief complaints is uncommon. ICD coding of ED diagnoses by

clinicians using clinical information systems is not universal. Nationwide, only 17% of physician practices and 31% of EDs have even adopted point-of-care systems (Burt and Hing, 2005). ICD codes from healthcare-insurance claims data are widely available, but the time latency is too long for many biosurveillance applications.

We discuss the empirical data about time latency from published reports shortly. But it is worthwhile to review the steps in the process of assigning ICD codes and transmitting them to a biosurveillance system (Figure 23.1), and how these steps may contribute to time latency. First, an individual clinician or coder must acquire enough information about a patient to assign an ICD code. This individual may be a triage nurse who obtains the patient's chief complaint within minutes of the patient's arrival to the ED, or a professional coder who studies a four-inch thick medical record weeks after the patient left

the hospital. Next, someone must enter an ICD code into a computer system.[11] For physician-assigned ICD codes, that may occur within minutes after she sees the patient (which in turn may be minutes to hours after the patient arrived at the office or ED) or even hours or days after the patient leaves the office or ED. Finally, the computer system must transmit the codes to a biosurveillance system or other computers that in turn transmit them to a biosurveillance system. Systems may either transmit ICD codes in real time (the exception) or in batch mode (the rule). Ideally, studies of ICD codes would measure and report the contribution of each step to overall time latency.

Studies to date—with one exception we discuss below—have not *measured* time latencies involved in ICD-code monitoring. Instead, authors provide general statements such as *most records were received by the next day*.

Lewis et al. (2000) report a one- to three-day time latency for diagnoses encoded with ICD by physicians working at DoD clinics. They state that this latency is from the time of the patient visit to the time data are available for viewing and analysis in the ESSENCE biosurveillance system. They note two factors that affect this time latency: time delays prior to physicians assigning codes and frequency of data transmission to ESSENCE from the DoD's Ambulatory Data System.

Note that the CDC's BioSense biosurveillance system receives the same ICD codes that Lewis and colleagues studied. Anecdotally, the time latency of ICD codes obtained by BioSense is not short enough to meet the needs of some state and local public health officials (U.S. Government Accountability Office, 2005). Such anecdotes highlight the importance of accurate *measurements* of each contribution to time latency and detailed descriptions of who assigns ICD codes and how and when.

Reports by the National Bioterrorism Syndromic Surveillance Demonstration Program suggest that the time latency of ICD codes assigned by physicians using a point-of-care system in an outpatient setting are "usually" less than 24 hours (Yih et al., 2004, Lazarus et al., 2002, Miller et al., 2004). This program involves multiple healthcare providers and health plans; thus, each healthcare system may have different latencies. Lazarus and colleagues report that at Harvard Vanguard Medical Associates in Boston, Massachusetts, ICD-coded diagnoses are available for "essentially all episodes by the end of the same day on which care is given" (Lazarus et al., 2002). Miller and colleagues report that at HealthPartners Medical Group in the Minneapolis-St. Paul region of Minnesota, ICD-coded diagnoses are available "within approximately 24 hours of a patient's initial visit" (Miller et al., 2004). Yih and colleagues report that, for the Demonstration Program as a whole, providers usually enter ICD-coded diagnoses into a point-of-care system on the same day as the patient visit, and that the systems put in place by the Program extract these data on a nightly basis (Yih et al., 2004).

Beitel and colleagues report that ED physicians at Boston Children's' Hospital code diagnoses with ICD within "hours" of the visit (Beitel et al., 2004). They do not describe how available these data are for biosurveillance nor whether they are transmitted in real time or via a daily batch-mode process. They also report that billing administrators assign "the final ICD-9 code to all charts, usually within 48 to 72 hours."

Suyama and colleagues report that billing ICD codes at the ED of University Hospital (which is part of the University of Cincinnati Medical Center) are available within 12 hours from the time the patient leaves the ED (Suyama et al., 2003). They do not state who assigns the ICD codes to each visit.

Begier and colleagues report that two community hospitals in the National Capital Region do not assign any ICD codes to patient visits within 24 hours of the patient leaving the ED (Begier et al., 2003). Whether ICD codes are available after 24 hours of the patient leaving the ED, they do not say.

Espino and colleagues conducted the only study that measured and reported time latencies for ICD codes (Espino and Wagner, 2001b). It is important to note that this study was of an atypical health system that expended effort to make physician-assigned ICD codes available to a biosurveillance system *in real time*, instead of using daily batch-mode extraction of ICD codes as other studies have described. Espino found that ICD codes for diagnoses (assigned by ED physicians) arrived at the RODS system on average 7.5 hours *after* ICD codes for chief complaints (assigned by ED registration clerks). The maximum time delay was 80.6 hours. Because the ICD codes for chief complaints and diagnoses were transmitted in real time, this measurement of time latency includes only a negligible transmission delay. Thus, nearly all of the 7.5 hours comprises patient waiting time, time for the physician to see the patient, and time for the physician to assign ICD codes to the visit using a point-of-care system.

In summary, a pervasive methodological limitation of reported studies is failure to measure and report the distribution of time latencies between patient presentation to the healthcare system and appearance of an ICD code in data systems. The single study that measured time latency was conducted in a best case—and non representative—situation. In the absence of definitive studies, our best assessment of time latencies and availability—based on the literature and our knowledge of biosurveillance systems and the healthcare system—is summarized in Table 23.12.

---

11   This someone may or may not be the person who originally elicited the information from the patient. For example, a triage nurse may elicit the chief complaint and write it on paper, and then the registration clerk, using the chief complaint on paper, enters an ICD code into a computer.

TABLE 23.12 Estimated Time Latencies and Availability of ICD Codes Originating from Different Points in Healthcare Process and Different Practice Settings

Point in Process of Care	Latency	Availability	Population Coverage
Registration (chief complaint)	Negligible if sent as HL7 ADT	Very low	Few EDs
Conclusion of ED or office visit (billing diagnosis)	VA: days DoD: days Civilian: days to months	VA: high DoD: high Civilian: unknown	Most EDs and offices
Admission (admission diagnosis)	Similar to ED diagnosis; or months if part of discharge data set	Unknown	All hospitalized patients
Discharge from hospital (discharge diagnosis)	Months	Very high	All hospitalized patients

DoD, U.S. Department of Defense; ED, emergency department; VA, Veterans Administration.

3.3. Studies of Informational Value of ICD Codes

There have been several studies that have measured the informational value of ICD codes for chief complaints and diagnoses. As with free-text chief complaints, these studies utilized experimental methods that we discussed in Chapter 21 and address the three hypotheses of interest (restated for ICD codes):

Hypothesis 1: An individual ICD code or a code set can discriminate between whether a patient has *syndrome* or disease X or not.

Hypothesis 2: ICD code sets can discriminate between whether there is an *outbreak* of type Y or not.

Hypothesis 3: Algorithmic monitoring of ICD code sets can detect an *outbreak* of type Y earlier than current best practice (or some other reference method).

3.3.1. Detection of Cases

Table 23.13 summarizes the results of studies of case detection using only ICD-coded data. The hypothesis underlying these studies is Hypothesis 1: *An individual ICD code or a code set can discriminate between whether a patient has syndrome or disease X or not.* We have grouped the studies in Table 23.13 by syndrome. Each row reports the sensitivity and specificity of an ICD code or code set for a particular syndrome or diagnosis. For convenient reference, we also computed the likelihood ratio positive and negative.

Unlike chief complaints, in which the syndromes studied are at coarse level of diagnostic precision (e.g., *respiratory*), research on ICD code sets has also studied code sets that are diagnostically precise. For this reason, we have grouped the published studies into those using diagnostically less precise and diagnostically more precise code sets.

Detection of Less Diagnostically Precise Syndromes. Four studies measured the case detection accuracy of less diagnostically precise ICD code sets. Only one study measured the case detection accuracy of a DoD-GEIS or CDC/DoD code set.

Espino and Wagner conducted the first study of case detection accuracy of an ICD code set for a diagnostically imprecise syndrome (Espino and Wagner, 2001). They studied the ability of ICD codes for ED chief complaints and ICD codes for ED diagnoses to detect patients with acute respiratory syndrome (defined as symptom duration of five days or less). In this ED, registration clerks assigned ICD codes to the chief complaints and ED physicians assigned ICD codes to the diagnoses at the time of the patient's visit. Two internists created the gold standard by reviewing the dictated ED visit note and assigning patients to *acute respiratory* if the duration of illness was five days or less and the patient had respiratory symptoms, abnormal pulmonary examination, or radiological evidence of pneumonia. The sensitivity and specificity of a *respiratory* ICD code set (that contained 64 ICD codes) for detecting patients with acute respiratory illness from chief complaints were 0.44 and 0.97, respectively. The sensitivity and specificity of the same *respiratory* code set for detecting patients with acute respiratory illness from diagnoses were 0.43 and 0.97, respectively. The accuracy of ED diagnoses was not significantly different from the accuracy of ED chief complaints.

Betancourt (2003) studied the case detection accuracy of three DoD-GEIS ICD code sets. He used manual review of medical records by two primary care physicians as the gold standard. He found that the sensitivity of the *respiratory*, *fever*, and *gastrointestinal* code sets was 0.65, 0.71, and 0.90, respectively. All three code sets had a specificity of 0.94.

Ivanov et al. (2002) studied the case detection accuracy of ICD codes for the detection of acute gastrointestinal illness (duration of illness two weeks or less) in EDs. ED physicians assigned ICD codes for diagnoses after seeing a patient. Review of the transcribed emergency-department report by two internists was the gold standard. As in the study by Espino and Wagner, the internists were instructed to label a patient as acute gastrointestinal if symptoms were of duration two weeks or less. They found that an ICD code set (that contained 16 ICD codes) had a sensitivity and specificity of 0.32 and 0.99, respectively, for the detection of *acute gastrointestinal* syndrome.

Beitel et al. (2004) studied the case detection accuracy of ICD codes for the detection of *respiratory, lower respiratory*, and *upper respiratory* syndromes in a pediatric ED. They do not state whether physicians or billing administrators assigned the ICD codes that they studied (but mention that both groups assign ICD codes to ED visits). They used single-physician review of the medical record as the gold standard. The sensitivity and specificity of their *respiratory* ICD code set was 0.70 and 0.98, respectively, for detection of *respiratory* syndrome. Similarly, the sensitivity and specificity of their *upper respiratory* ICD code set was 0.56 and 0.98 for *upper*

TABLE 23.13 Performance of ICD Codes and Code Sets in Detecting Syndromes and Diseases

ICD Code or Code Set Being Tested	Reference Standard for Comparison	Sensitivity	Specificity	Positive Likelihood Ratio	Negative Likelihood Ratio
Respiratory Syndrome					
Respiratory code set, ICD codes for ED chief complaints[a]	Two internist-review of dictated ED note	0.44	0.97	14.7	0.58
Respiratory code set, ICD codes for ED diagnoses[a]	Two internist-review of dictated ED note	0.43	0.97	14.3	0.59
Respiratory code set[b]	Two physician-review of medical record	0.65	0.94	10.8	0.37
Respiratory code set[c]	Single physician review of medical record	0.70	0.98	35.0	0.31
Combined chief complaint and respiratory code set[c]	Single physician review of medical record	0.72	0.97	24.0	0.29
Lower respiratory syndrome	Single physician review of medical record	0.87	0.99	87.0	0.13
Lower respiratory code set[c]					
Upper respiratory syndrome					
Lower respiratory code set[c]	Single physician review of medical record	0.56	0.98	28.0	0.45
Gastrointestinal syndrome					
Gastrointestinal code set[d]	Two internist-review of dictated ED note	0.32	0.99	32.0	0.69
Gastrointestinal code set[b]	Two physician-review of medical record	0.90	0.94	15.0	0.11
Fever					
Fever code set[b]	Two physician-review of medical record	0.71	0.94	11.8	0.31
Pneumococcal pneumonia					
038.00 Streptococcal septicemia[e]	Microbiology cultures	0.20	0.999	50.0	0.80
038.20 Pneumococcal septicemia[e]	Microbiology cultures	0.19	>0.999	190.0	0.81
481 Pneumococcal pneumonia (or lobar pneumonia, organism unspecified)[e]	Microbiology cultures	0.58	0.98	23.2	0.43
482.30 Pneumonia due to *Streptococcus*[e]	Microbiology cultures	0.11	0.99	11.0	0.90
486 Pneumonia, organism unspecified[e]	Microbiology cultures	0.14	0.48	0.3	1.78
518.81 Respiratory failure[e]	Microbiology cultures	0.15	0.92	1.8	0.93
Pneumococcal pneumonia code set (038.00, 038.20, 481, 482.30, 486, 518.81)[e]	Microbiology cultures	0.89	0.44	1.6	0.25
Pneumococcal pneumonia code set (038.00, 038.20, 481, 482.30)[e]	Microbiology cultures	0.81	0.96	20.3	0.20
Tuberculosis					
Tuberculosis code set (all ICD codes with tuberculosis in title)[f]	Review of medical record	0.57	0.75	2.3	0.57
HIV					
HIV/AIDS code set (042, 043, 044, 795.8)[g]	Review of medical record	0.98	0.89	8.9	0.02
AIDS					
HIV/AIDS code set (042, 043, 044, 795.8)[g]	Review of medical record	0.97	0.58	2.3	0.05

[a]From Espino, J., Wagner, M. (2001a). The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness. In: *Proceedings of American Medical Informatics Association Symposium*, 164–8, with permission.

[b]From Betancourt, J. A. (2003). An evaluation of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE). Doctoral dissertation, George Washington University, with permission.

[c]From Beitel, A. J., Olson, K. L., Reis, B. Y., et al. (2004). Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Pediatr Emerg Care* 20:355–60, with permission.

[d]From Ivanov, O., Wagner, M. M., Chapman, W. W., et al. (2002). Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. In: *Proceedings of American Medical Informatics Association Symposium*, 345–9, with permission.

[e]From Guevara, R. E., Butler, J. C., Marston, B. J., et al. (1999). Accuracy of ICD-9-CM codes in detecting community-acquired pneumococcal pneumonia for incidence and vaccine efficacy studies. *Am J Epidemiol* 149:282–9, with permission.

[f]From San Gabriel, P., Saiman, L., Kaye, K., et al. (2003). Completeness of pediatric TB reporting in New York City. *Public Health Rep* 118:144–53, with permission.

[g]From Rosenblum, L., Buehler, J. W., Morgan, M. W., et al. (1993). HIV infection in hospitalized patients and Medicaid enrollees: the accuracy of medical record coding. *Am J Public Health* 83:1457–9, with permission.

respiratory syndrome, respectively, and for their lower respiratory ICD code set was 0.87 and 0.99 for *lower respiratory* syndrome, respectively.

Notably, Beitel and colleagues studied the combined information present in both chief complaints and ICD codes. They assigned patients to *respiratory* if they had either a *respiratory* chief complaint or an ICD code in the *respiratory* code set. The sensitivity and specificity of the combined definition for *respiratory* were 0.72 and 0.97, respectively. These results vary only slightly from the sensitivity and specificity of the *respiratory* ICD code set alone.

Detection of More Diagnostically Precise Syndromes. Several studies have measured the case detection accuracy of ICD codes available only after a patient has left the hospital. We discuss three examples of such studies here.

Guevara et al. (1999) studied the sensitivity and specificity of ICD codes and code sets for the detection of the disease *pneumococcal pneumonia*. The ICD codes were part of data collected during a study of community-acquired pneumonia. They used the results of microbiology cultures as a gold standard. They found that ICD code 481 had the highest sensitivity—58.3%—for *pneumococcal pneumonia* (out of all six ICD codes studied) and a specificity of 97.5% (Table 23.13). Of the six ICD code sets they studied, an ICD code set with four ICD codes had the highest sensitivity and specificity—81.2% and 96.0%, respectively. One note of caution in interpreting these results is they limited the study to patients who met the inclusion criteria for community-acquired pneumonia. Thus the sensitivity and specificity results reported are specifically for discriminating among pneumococcal and other causes of community-acquired pneumonia in hospital patients.

San Gabriel et al. (2003) studied the accuracy of hospital discharge ICD codes for the detection *tuberculosis* in children. The gold standard against which they compared ICD codes was medical record review. An ICD code set that contained all ICD codes with "tuberculosis" in their definition had a sensitivity and specificity of 57% and 75%, respectively, for the detection of *tuberculosis* in children.

Rosenblum et al. (1993) studied the accuracy of hospital discharge ICD codes for the detection of human immunovirus (*HIV*) infection and acquired immunodeficiency syndrome (*AIDS*) (Rosenblum et al., 1993). They found that the sensitivity and specificity of an ICD code set (that contained four ICD codes for HIV infection and AIDS) for *HIV* was 98% and 89%,[12] respectively. Similarly, the sensitivity and specificity of the ICD code set for *AIDS* were 97% and 58%, respectively.[13]

Summary of Studies of Case Detection from ICD Codes. With respect to Hypothesis 1, the literature on accuracy of both disease-detection and syndrome-detection from ICD code analysis suggest that:

1. ICD-coded diagnoses and chief complaints contain information about syndromic presentations and diagnoses of patients and existing code sets can extract that information.
2. For syndromes that are at the level of diagnostic precision of *respiratory* or *gastrointestinal*, it is possible to automatically classify ED and office patients (both pediatric and adult) from ICD codes with a sensitivity of approximately 0.65 and a specificity of approximately 0.95. This level of accuracy is similar to that achievable with chief complaints.
3. For diagnoses that are at the level of diagnostic precision of disease (e.g., pneumonia) or disease and organism (e.g., pneumococcal pneumonia) it is possible to automatically classify patients from hospital discharge data.
4. Sensitivity of classification is better for some syndromes and diagnoses than for others.
5. The specificity of case detection from ICD codes and code sets is less than 100%, meaning that daily aggregate counts will have false positives among them due to falsely classified patients.

3.3.2. Detection of Outbreaks

This section describes studies that address Hypotheses 2 and 3, which we reproduce here for convenient reference:

Hypothesis 2: ICD code sets can discriminate between whether there is an *outbreak* of type Y or not.
Hypothesis 3: Algorithmic monitoring of ICD code sets can detect an *outbreak* of type Y earlier than current best practice (or some other reference method).

As with chief complaints, challenges in studying the outbreak detection performance of ICD code sets include collecting data from multiple healthcare facilities in the region affected by an outbreak and achieving adequate sample size of outbreaks for study. To emphasize the importance of sample size, we divide this section into studies of multiple outbreaks (N>1) and studies of single outbreaks (N=1). Unlike chief complaints, no published prospective studies exist.

N>1 Studies. Yih et al. (2005) used the detection algorithm method to study retrospectively 110 gastrointestinal disease outbreaks in Minnesota. They studied daily counts of the CDC/DoD *Gastrointestinal, All* code set. The detection algorithm they used was a space-time scan statistic algorithm

---

12   They computed a PPV of 97%. We analyzed the data reported in the paper to compute this specificity result.
13   They computed a PPV of 65%; We analyzed the data reported in the paper to compute this specificity result.

(Kleinman et al., 2005). The ICD code data they studied included 8% of the population. The outbreaks ranged in size from 2 to 720 cases with a median outbreak size of 7. Half of the outbreaks were foodborne and approximately half were caused by viruses. They defined a true alarm as a cluster of disease identified by the detection algorithm that was within 5 km of the outbreak, and that occurred any time from one week prior to the start of the outbreak to one week after the outbreak investigation began. The sensitivity of *Gastrointestinal, All* for the detection of small gastrointestinal outbreaks at a false alarm rate of 1 every 2 years was 1%. They also measured sensitivity at false alarm rates of one per year (sensitivity = 2%), 1 per 6 months (3%), 1 per 2 months (5%), 1 per month (8%), and 1 per 2 weeks (13%). They did not measure timeliness of detection. The low percentage of population covered may partly explain the low sensitivity.

N=1 Studies. Tsui et al. (2001) conducted the first study of outbreak detection accuracy and timeliness of ICD codes available soon after a patient visit. It is also the only study to compare the outbreak-detection performance of two ICD code sets for the same outbreak. They used correlation analysis and the detection-algorithm method to study the accuracy and timeliness of an *ILI* and a *respiratory* ICD code set for the detection of an influenza outbreak.[14] Pneumonia and influenza (P&I) deaths were used as the gold-standard determination of the outbreak. The correlation analysis showed a two-week lead time for both ICD code sets relative to P&I deaths.[15] The detection algorithm analysis used the Serfling algorithm. Both ICD code sets had a sensitivity of 100% (1/1). The false alarm rate for the *ILI* code set was one per year and for the *respiratory* code set was one per three months. Detection of the influenza outbreak occurred one week earlier from both code sets relative to detection from P&I deaths at these false alarm rates and measurements of sensitivity.

Miller et al. (2004) used correlation analysis and the detection algorithm method to study the ability of ICD data obtained from a large healthcare organization in Minnesota to detect an influenza outbreak. Using P&I deaths as a gold standard, Miller demonstrated a significant Pearson correlation of 0.41 of an *ILI* ICD code set with P&I deaths. When they lagged the *ILI* time series by one week, the Pearson correlation remained 0.41. For the detection-algorithm method, they used a CUSUM algorithm. CUSUM detected the outbreak from *ILI* (sensitivity of 1/1) one day earlier than the date the health department confirmed the first positive influenza isolate. They did not report the false alarm rate of the CUSUM algorithm for this detection sensitivity and timeliness.

Lazarus et al. (2002) used correlation analysis to study the ability of ICD codes to detect an influenza outbreak (Lazarus et al., 2002). They obtained the ICD codes from a point-of-care system used by a large multispecialty physician group practice. Clinicians assigned ICD-coded diagnoses at the time of either an in-person or phone consultation. They used hospital discharge data as a gold standard. They created two time series using the same *lower respiratory* ICD code set (based on the DoD-GEIS *respiratory* code set): one from their point-of-care data and one from the hospital discharge data. The maximum correlation of these time series was 0.92 when point-of-care ICD codes preceded hospital-discharge ICD codes by two weeks.

Lewis et al. (2002) used correlation analysis to study the ability of ICD codes to detect an outbreak of influenza. They used CDC sentinel physician data for influenza as a gold standard. Physicians assigned ICD codes at outpatient DoD facilities. The correlation of the DoD-GEIS *respiratory* code set with the CDC sentinel physician data was 0.7. They did not measure the correlation at time lags other than zero.

Suyama et al. (2003) used correlation analysis to study whether ICD codes contain a signal of notifiable diseases reported to the health department. They used health department data about notifiable diseases as a gold standard. Billing administrators assigned ICD codes to ED visits. They created time series from both data sets using the same ICD code sets: *gastrointestinal, pulmonary, central nervous system, skin, fever,* and *viral.* They found statistically significant correlations between *gastrointestinal, pulmonary,* and *central nervous system* code sets and notifiable diseases. Conversely, the *skin, fever,* and *viral* syndromes did not have a statistically significant correlation with notifiable diseases.

Summary of Studies of Outbreak Detection from ICD Codes. With respect to Hypotheses 2 and 3, the studies we reviewed demonstrate that:

1. Some large outbreaks of respiratory and ILI can be detected from aggregate analysis of ICD codes. Small outbreaks of gastrointestinal illness generally cannot. There were no studies of large outbreaks of gastrointestinal illness (Hypothesis 2).
2. Research to date is suggestive but not conclusive that outbreaks of influenza can be detected earlier than current surveillance methods (Hypothesis 3).
3. The methodological weaknesses of studies included:

   a. Not measuring all three characteristics of outbreak-detection performance (false alarm rate, sensitivity, and timeliness).

---

14   They used the same same *Respiratory* ICD code set as Espino and colleagues (Espino and Wagner, 2001).
15   They do not report the correlation measurement, but in one figure in the paper, the correlation of the *Respiratory* code set appears close to 1.0 at a time lag of two weeks.

b. Not reporting either the correlation or the time lag at which correlation was maximal.

c. Measuring correlation at too few time lags (sometimes at just one or two values).

d. Not reporting sampling completeness.

4. In general, the number of published studies is small, perhaps due to the fact that ICD monitoring systems are relatively recent. We expect more studies to be published in the near future.

4. USING CHIEF COMPLAINTS AND ICD-CODED DIAGNOSES IN BIOSURVEILLANCE

Although there is the occasional perception that a biosurveillance organization must choose between collecting chief complaints or ICD codes, this is a false issue. They each have strengths and limitations. They vary in availability by healthcare organization, time latencies, and the accuracy with which they can discriminate among different syndromes and diagnoses. Outbreaks vary in their rapidity, early syndromic presentation, and whether they cause people to seek hospital care versus outpatient care. It is likely that future research will show that chief complaints are superior to ICD codes for monitoring for some outbreaks and ICD codes are superior for other outbreaks.

At the present time, the strongest statement about their respective roles is that chief complaints—because of the low time latency—may prove to be more useful for detecting sudden events. ICD codes—because of their ability to support more diagnostically precise syndromes (or disease categories)—may prove to be more useful for detecting smaller outbreaks.

A biosurveillance organization should collect and analyze both. Algorithms for case detection can utilize both types of data (Chapter 13). Because chief complaints and ICD-codes are associated with a specific patient, it is possible to link them (and other data) to improve the sensitivity and specificity of case detection.

5. SUMMARY

Chief complaints and ICD-coded diagnoses are among the most highly available types of biosurveillance data that we discuss in this section of the book and they are being monitored by many biosurveillance organizations.

Chief complaints are routinely collected by the healthcare system. They are available early in the course of the clinical care process and they contain information about a patient's symptoms. Research to date has demonstrated that they can be readily obtained from emergency departments and hospitals, more often than not in real time as HL7 messages. They are less accessible in the outpatient setting. Chief complaints must be subject to NLP to extract information for biosurveillance. The two basic approaches are keyword matching and Bayesian text processing. The NLP assigns a chief complaint

either directly to a syndrome category, or extracts symptoms that are then further processed in a second step to determine whether the patient matches a Boolean case definition. Research on the accuracy of automatic syndrome assignment from chief complaints shows that automatic classification from chief complaint data into syndromes such as *respiratory* and *gastrointestinal* can be accomplished with moderately good sensitivity and specificity. This accuracy is not sufficient to support detection of small outbreaks. When researchers have attempted to automatically assign patients to more diagnostically precise syndromes such as *febrile respiratory*, sensitivity declines quickly—chief complaints simply do not contain sufficient information to support such automatic assignments. Research has also demonstrated that detection algorithms can use daily counts of syndromes derived in this manner to detect large respiratory or diarrheal outbreaks such as those due to influenza and rotavirus. It is likely that large outbreaks of diseases that initially present with other syndromes monitored for would also be detected, although studies of such outbreaks do not yet exist, so this point remains a matter of opinion.

ICD codes are also a type of data routinely collected by the healthcare system. They are far more heterogeneous in meaning than chief complaints as the ICD coding system contains codes for symptoms, syndromes, and diagnoses at different levels of diagnostic precision. Additionally, codes are collected at multiple points during the course of a patients illness and the accuracy and diagnostic precision of coding may vary across these points. ICD codes become available later in the course of the clinical care than chief complaints. From the perspective of biosurveillance, they contain indirect information about a patient's symptoms: the reasoning process that developers use when they include an ICD code in a "code set" is "would a patient with disease X have respiratory symptoms? If so, let's include the code for disease X in a code set for respiratory." Research to date has demonstrated that ICD codes can be readily obtained from the healthcare system. The military fortuitously uses ICD to encode all services (and doctors do the encoding). There are several examples of ICD code sets developed for biosurveillance. A surprisingly small amount of work has been done on developing very specific ICD codes sets to automate the surveillance for conditions such as pneumonia. Research on the accuracy of automatic syndrome assignment shows results similar to chief complaints. As with chief complaints, this accuracy is not sufficient to support detection of small outbreaks. Research on the accuracy of automatic disease assignment using hospital discharge diagnoses shows higher accuracy. Research has also demonstrated that detection algorithms can use daily counts of syndromes derived in this manner to detect large respiratory or diarrheal outbreaks such as those due to influenza and rotavirus. It is likely that large outbreaks of diseases that initially present with other syndromes monitored for would also be detected, although there are no natural occurring examples, so this point remains a matter of opinion.

The results suggest a few immediate applications such as influenza monitoring and early warning of cohort exposures. They also carry an important implication: users of systems should not be overly reassured about the absence of a spike of activity as these data when used alone are not likely to detect a small outbreak. Monitoring of chief complaints and ICD-coded diagnoses alone simply does not have high sensitivity for small outbreaks (unless they are tightly clustered in space and time and possibly demographic strata and the detection system is capable of exploring those dimensions).

It is unfortunate that some authors over-interpret these limited results to conclude that "syndromic surveillance" does not work. A more accurate summarization of the state-of-the-art might be that surveillance of diagnostically imprecise syndrome categories is not capable of detecting small outbreaks because of the background level of patients satisfying the broad syndrome definition. When the availability of additional clinical data (e.g., temperatures and laboratory test orders or results) allows monitoring of more diagnostically precise syndromes, we expect that smaller outbreaks will be detectable and the time of detection of larger outbreaks will improve. This type of syndromic surveillance has been practiced for quite some time (e.g., polio, AIDS), albeit manually due to the lack of automatic access to required surveillance data.

ADDITIONAL RESOURCES

- CDC site with annotated bibliography that will include future publications on this topic: *http://www.cdc.gov/epo/ dphsi/syndromic/*
- Syndromic.org "latest studies" site: *http://www.syndromic. org/index.php*

ACKNOWLEDGMENTS

REFERENCES

Balter, S., Weiss, D., Hanson, et al. (2005). Three years of emergency department gastrointestinal syndromic surveillance in New York City: what have we found? *MMWR Morb Mortal Wkly Rep* 54(Suppl):175–80. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi? cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16177711.*

Beasley, J. W., Hankey, T. H., Erickson, R., et al. (2004). How many problems do family physicians manage at each encounter? A WReN study. *Ann Fam Med* 2:405–10. *http://www.ncbi.nlm.nih.gov/entrez/ query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids= 15506571.*

Begier, E. M., Sockwell, D., Branch, L. M., et al. (2003). The National Capitol Region's Emergency Department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results, *Emerg Infect Dis* 9:39–6. *http://www.ncbi.nlm.nih.gov/entrez/ query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids= 12643841.*

Beitel, A. J., Olson, K. L., Reis, B. Y., et al. (2004). Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Pediatr Emerg Care* 20:355–60. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd= Retrieve&db=PubMed&dopt=Citation&list_uids=15179142.*

Betancourt, J. A. (2003). An Evaluation of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE). Doctoral dissertation, George Washington University.

Buckeridge, D. L., Switzer, P., Owens, D., et al. (2005). An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. *MMWR Morb Mortal Wkly Rep* 54(Suppl): 109–15. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd= Retrieve&db=PubMed&dopt=Citation&list_uids=16177701.*

Burt, C., Hing, E. (2005). *Use of Computerized Clinical Support Systems in Medical Settings: UnitedStates, 2001–03.* Atlanta, GA: Centers for Disease Control and Prevention.

Campos-Outcalt, D. E. (1990). Accuracy of ICD-9-CM codes in identifying reportable communicable diseases. *Qual Assur Util Rev* 5:86–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd= Retrieve&db=PubMed&dopt=Citation&list_uids=2136670.*

Centers for Disease Control and Prevention. (2003). Syndrome Definitions for Diseases Associated with Critical Bioterrorism-Associated Agents. *http://www.bt.cdc.gov/surveillance/syndromedef/ index.asp.*

Chang, H., Cochrane, D., Tserenpuntsag, B., et al. (2005). ICD9 as a surrogate for chart review in the validation of a chief complaint syndromic surveillance system. In: *Proceedings of Syndromic Surveillance Conference,* Seattle, WA.

Chapman, W. W., Christensen, L. M., Wagner, M. M., et al. (2005a). Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artif Intell Med* 33:31–40. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db= PubMed&dopt=Citation&list_uidsg=15617980.*

Chapman, W. W., Dowling, J. N., Wagner, M. M. (2005b). Classification of emergency department chief complaints into seven syndromes: a retrospective analysis of 527,228 patients. *Ann Emerg Med.* vol. 46, no. 5, pp. 445–55.

Chapman, W. W., Dowling, J. N., Wagner, M. W. (2005c). Generating a reliable reference standard set for syndromic case classification. *J Am Med Inform Assoc.* vo. 12, no. 6, pp. 618–29 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db= PubMed&dopt=Citation&list_uids=16049227.

Chapman, W. W., Espino, J. U., Dowling, et al. (2003). *Detection of Multiple Symptoms from Chief Complaints.* Technical Report, CBMI Report Series. Pittsburgh, PA: Center for Biomedical Informatics, University of Pittsburgh.

Espino, J., Wagner, M. (2001a). The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness. In: *Proceedings of American Medical Informatics Association Symposium*, 164–8.

Espino, J. U., Wagner, M. M. (2001b). Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. In: *Proceedings of American Medical Informatics Association Symposium*, 164–8. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11833477*.

Fisher, E. S., Whaley, F. S., Krushat, et al. (1992). The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health* 82:243–8. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1739155*.

Gesteland, P. H., Gardner, R. M., Tsui, F. C., et al. (2003). Automated Syndromic Surveillance for the 2002 Winter Olympics. *J Am Med Inform Assoc. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12925547*.

Gesteland, P. H., Wagner, M. M., Chapman, W. W., et al. (2002). Rapid deployment of an electronic disease surveillance system in the State of Utah for the 2002 Olympic Winter games. In: *Proceedings of American Medical Informatics Association Symposium*, 285–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12463832*.

Guevara, R. E., Butler, J. C., Marston, B. J., et al. (1999). Accuracy of ICD-9-CM codes in detecting community-acquired pneumococcal pneumonia for incidence and vaccine efficacy studies. *Am J Epidemiol* 149:282–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9927225*.

Heffernan, R., Mostashari, F., Das, D., et al. (2004a). Syndromic surveillance in public health practice, New York City. *Emerg Infect Dis* 10:858–64. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15200820*.

Heffernan, R., Mostashari, F., Das, D., et al. (2004b). Syndromic surveillance in public health practice, New York City. *Emerg Infect Dis* 10:858–64. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15200820*.

Henry, J. V., Magruder, S., Snyder, M. (2004). Comparison of office visit and nurse advice hotline data for syndromic surveillance—Baltimore-Washington, D.C., metropolitan area, 2002. *MMWR Morb Mortal Wkly Rep* 53(Suppl):112–6. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15714639*.

Hsia, D. C., Krushat, W. M., Fagan, A. B., et al. (1988). Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. *N Engl J Med* 318:352–5. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3123929*.

Irvin, C. B., Nouhan, P. P., Rice, K. (2003). Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. *Ann Emerg Med* 41:447–52. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12658241*.

Ivanov, O., Gesteland, P. H., Hogan, W., et al. (2003). Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. In: *Proceedings of American Medical Informatics Association Symposium*, 318–22. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14728186*.

Ivanov, O., Wagner, M. M., Chapman, W. W., et al. (2002). Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. In: *Proceedings of American Medical Informatics Association Symposium*, 345–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12463844*.

Jollis, J. G., Ancukiewicz, M., DeLong, E. R., et al. (1993). Discordance of databases designed for claims payment versus clinical Information systems. Implications for outcomes research. *Ann Intern Med* 119:844–50. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8018127*.

Lazarus, R., Kleinman, K., Dashevsky, I., et al. (2002). Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerg Infect Dis* 8:753–60. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12141958*.

Lazarus, R., Kleinman, K. P., Dashevsky, I., et al. (2001). Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. *BMC Public Health* 1:9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11722798*.

Levy, R., Cochrane, D., Chang, H., et al. (2005). Components of Inter-hospital variability in chief complaint classification of gastrointestinal syndrome. In: *Syndromic Surveillance Conference*, Seattle, WA.

Lewis, M. D., Pavlin, J. A., Mansfield, J. L., et al. (2002). Disease outbreak detection system using syndromic data in the greater Washington, DC area. *Am J Prev Med* 23:180–6. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12350450*.

Lloyd, S. S., Rissing, J. P. (1985). Physician and coding errors in patient records. *JAMA* 254:1330–6. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3927014*.

Lombardo, J., Burkom, H., Elbert, E., et al. (2003). A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). *J Urban Health* 80(Suppl 1):i32–42. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12791777*.

MacIntyre, C. R., Ackland, M. J., Chandraraj, E. J., et al. (1997). Accuracy of ICD-9-CM codes in hospital morbidity data, Victoria: implications for public health research. *Aust N Z J Public Health* 21:477–82. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9343891*.

Magruder, S. F., Henry, J., Snyde, M. (2005). Linked analysis for definition of nurse advice line syndrome groups, and comparison to encounters. *MMWR Morb Mortal Wkly Rep* 54(Suppl):93–7.

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16177699.

Mandl, K., Reis, B., Cassa, C. (2004). Measuring outbreak-detection performance by using controlled feature set simulations. *MMWR Morb Mortal Wkly Rep* 53(Suppl):130–6.

McClay, J. C., Campbell, J. (2002). Improved coding of the primary reason for visit to the emergency department using SNOMED. In: *Proceedings of American Medical Informatics Association Symposium*, 499–503. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=etrieve&db=PubMed&dopt=Citation&list_uids=12463874.*

Miller, B., Kassenborg, H., Dunsmuir, W., et al. (2004). Syndromic surveillance for influenzalike illness in ambulatory care network. *Emerg Infect Dis* 10:1806–11. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15504267.*

Mocny, M., Cochrane, D. G., Allegra, J. R., et al. (2003). A comparison of two methods for biosurveillance of respiratory disease in the emergency department: chief complaint vs ICD9 diagnosis code. *Acad Emerg Med* 10:513. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12720909.*

Mundorff, M. B., Gesteland, P., Haddad, M., et al. (2004). Syndromic surveillance using chief complaints from urgent-care facilities during the Salt Lake 2002 Olympic Winter Games. *MMWR Morb Mortal Wkly Rep* 53(Suppl):254.

Olszewski, R. (2003a). Bayesian classification of triage diagnoses for the early detection of epidemics. In: *Recent Advances in Artificial Intelligence: Proceedings of Sixteenth International FLAIRS Conference*, 412–16.

O'Malley K, J., Cook, K. F., Price, M. D., et al. (2005). Measuring diagnoses: ICD code accuracy. *Health Serv Res* 40:20–39. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16178999.*

Peabody, J. W., Luck, J., Jain, S., et al. (2004). Assessing the accuracy of administrative data in health Information systems. *Med Care* 42:1066–72. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15586833.*

Reis, B. Y., Mandl, K. D. (2004). Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. *Ann Emerg Med* 44:235–41. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15332065.*

Ritzwoller, D. P., Kleinman, K., Palen, T., et al. (2005). Comparison of syndromic surveillance and a sentinel provider system in detecting an influenza outbreak—Denver, Colorado, 2003. *MMWR Morb Mortal Wkly Rep* 54(Suppl):151–6. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids?16177707.*

Rizzo, S. L., Grigoryan, V. V., Wallstrom, G. L., et al. (2005). *Using Case Studies to Evaluate Syndromic Surveillance.* RODS Technical Report. Pittsburg, PA: Real-Time Outbreak and Disease Surveillance Laboratory, Center for Biomedical Informatics, University of Pittsburgh.

Rosenblum, L., Buehler, J. W., Morgan, M. W., et al. (1993). HIV infection in hospitalized patients and Medicaid enrollees: the accuracy of medical record coding. *Am J Public Health* 83:1457–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8214239.*

San Gabriel, P., Saiman, L., Kaye, K., et al. (2003). Completeness of pediatric TB reporting in New York City. *Public Health Rep* 118:144–53. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12690068.*

Sniegoski, C. (2004). Automated syndromic classification of chief complaint records. *Johns Hopkins University APL Technical Digest* 25:68–75.

Suyama, J., Sztajnkrycer, M., Lindsell, C., et al. (2003). Surveillance of infectious disease occurrences in the community: an analysis of symptom presentation in the emergency department. *Acad Emerg Med* 10:753–63. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12837650.*

Travers, D. A., Waller, A., Haas, S. W., et al. (2003). Emergency department data for bioterrorism surveillance: electronic data availability, timeliness, sources and standards. In: *Proceedings of American Medical Informatics Association Symposium*, 664–8. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14728256.*

Tsui, F., Espino, J., Wagner, M. (2005). *The Timeliness, Reliability, and Cost of Real-time and Batch Chief Complaint Data.* RODS Laboratory Technical Report. Pittsburg, PA: Real-Time Outbreak and Disease Surveillance Laboratory, Center for Biomedical Informatics, University of Pittsburgh.

Tsui, F., Wagner, M., Dato, V., et al. (2001). Value of ICD-9–coded chief complaints for detection of epidemics. In: *Proceedings of the Fall Symposium of the American Medical Informatics Association*, 711–5.

Tsui, F.-C., Espino, J. U., Dato, V. M., et al. (2003). Technical description of RODS: A real-time public health surveillance system. *J Am Med Inform Assoc* 10:399–408. *http://www.jamia.org/preprints.shtml.*

Tsui, F. C., Espino, J. U., Wagner, M. M., et al. (2002). Data, network, and application: technical description of the Utah RODS Winter Olympic Biosurveillance System. In: *Proceedings of American Medical Informatics Association Symposium*, 815–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12463938.*

U.S. Government Accounting Office. (2005). *Information Technology: Federal Agencies Face Challenges in Implementing Initiatives to Improve Public Health Infrastructure.* GAO-308. Washington, DC: Government Accounting Office. *http://www.gao.gov/new.items/d05308.pdf.*

U.S. Department of Defense, Global Emerging Infections System. (2005). ICD9 Code Sets Used in ESSENCE. *http://www.geis.fhp.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp#ICD9.*

Wagner, M., Espino, J., Tsui, F.-C., et al. (2004). Syndrome and outbreak detection from chief complaints: the experience of the Real-Time Outbreak and Disease Surveillance Project. *MMWR Morb Mortal Wkly Rep* 53(Suppl):28–31.

Wagner, M. M., Tsui, F. C., Espino, J. U., et al. (2001). The emerging science of very early detection of disease outbreaks. *J Public Health Manag Pract* 7:51–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11710168.*

Yao, P., Wiggs, B. R., Gregor, C., et al. (1999). Discordance between physicians and coders in assignment of diagnoses. *Int J Qual Health Care* 11:147–53. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10442845.*

Yih, K. W., Abrams, A., Danila, R., et al. (2005). Ambulatory-care diagnoses as potential indicators of outbreaks of gastrointestinal illness—Minnesota. *MMWR Morb Mortal Wkly Rep* 54(Suppl):157–62. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16177708.*

Yih, W. K., Caldwell, B., Harmon, R., et al. (2004). National Bioterrorism Syndromic Surveillance Demonstration Program. *MMWR Morb Mortal Wkly Rep* 53(Suppl):43–9. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15714626.*

Yokoe, D. S., Subramanyan, G. S., Nardell, E., et al. (1999). Supplementing tuberculosis surveillance with automated data from health maintenance organizations. *Emerg Infect Dis* 5:779–87. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10603211.*

Yuan, C. M., Love, S., Wilson, M. (2004). Syndromic surveillance at hospital emergency departments—southeastern Virginia. *MMWR Morb Mortal Wkly Rep* 53(Suppl):56–8. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15714630.*