An Antibiogram Data Model and Implementation Schema

by

David B. Ross, MD, PhD

A Capstone Project

Presented to the Department of Medical Informatics and Clinical Epidemiology

and the Oregon Health & Sciences University School of Medicine

in partial fulfillment of

the requirements for the degree of

Master of Biomedical Informatics

June 2012

School of Medicine

Oregon Health and Sciences University

CERTIFICATE OF APPROVAL

_____

This is to certify that the Master's Capstone Project of

David B. Ross, MD, PhD

An Antibiogram Data Model and Implementation Schema

has been approved

_____

Capstone Advisor

Table of Contents

Disclaimer

The work presented here represents the opinions of the author and does not necessarily represent the views of the United States Department of Veterans Affairs or the United States Government.

Acknowledgements

This capstone would not have been possible without the patient guidance of my advisor, Dr. Judy Logan. I am greatly indebted to her for her willingness to share her expertise, her sense of humor, and her extraordinary commitment to teaching.

I am also very grateful for Diane Doctor's support and assistance during my studies towards this degree. Her calm response and practical solutions for a series of academic and administrative challenges were invaluable. More generally, I would like to thank Dr. William Hersh, the DMICE faculty and staff, and my fellow students for creating and maintaining an incredibly collegial environment for studying biomedical informatics.

This work has benefited greatly from discussions with several individuals within the Veterans Health Administration interested in potential applications of medical informatics to the problem of antimicrobial resistance. I would particularly like to thank Dr. Richard Pham, for suggesting antibiogram data modeling as the topic for this capstone and for many hours of stimulating discussions. I am also extremely grateful to Dr. Melinda Neuhauser for encouraging me to pursue this and for her unflagging support and excellent scientific counsel.

I am very appreciative to Dr. Daniel Maddix of the San Francisco Veterans Affairs Medical Center (VAMC) and Mr. Neil Puzon of the Washington DC VAMC for their time and patience in educating me about clinical microbiology and antibiogram construction procedures, as well as Dr. Virginia Kan and Dr. Robert Williamson of the DC VAMC for discussions about the structure of microbiology data in VistA. Thanks

also to Dr. Makoto Jones, of the Salt Lake City VAMC, Dr. Michael Matheny of Vanderbilt University, Dr. Cynthia Brandt of Yale University, and Dr. Woody Levin at the West Haven VAMC for helpful discussions on data standardization. Last, but not least, much thanks to Dr. Larry Mole, of the VHA Office of Public Health for sharing his unique expertise on VHA databases and for keeping me sane.

Finally, my thanks to my family for their love, support, and endless patience.

Abstract

Antimicrobial resistance (AR) is a serious clinical and public health problem, driven largely by inappropriate use of antimicrobials, particularly in the setting of empiric therapy of infections. Provision of local AR prevalence data to clinicians in the form of a cumulative antibiogram (CABGM) can provide decision support for antimicrobial prescribing, decreasing inappropriate use, improving treatment outcomes, and minimizing selection of AR organisms. However, CABGMs are produced by widely varying methods from different data sources with different vocabularies, resulting in inaccurate data presentations, lack of compliance with published standards, and inability to compare AR rates between different health care facilities for quality improvement.

To address this, I describe a data model constraining a CABGM to a published standard using appropriate vocabularies. This standard data model can be used to construct CABGMs for major use cases, including use as a decision support tool for clinicians choosing empiric antimicrobial therapy, as a surveillance tool for hospital epidemiologists and public health officials comparing AR prevalence rates between hospitals, and as a quality improvement tool for aggregating AR prevalence data from different institutions.

This data model may be used to construct a relational database model; however, clinical microbiology and patient data are frequently stored in electronic medical records using a hierarchical database model. Using the electronic medical record and clinical data warehouse employed by the Veterans Health Administration as an example, I outline an implementation scheme for migrating clinical and microbiological data from hierarchical databases to a relational database defined by the model. Finally, vocabularies

used to describe microbiologic concepts and store clinical microbiology data frequently differ between different health care facilities. I discuss approaches to mapping semantic equivalents used for microbiology data to a common set of standard terms.

Abbreviations

ABGM	Antibiogram
AR	Antimicrobial resistance
AST	Antimicrobial susceptibility test
BLA	β-lactamase
CABGM	Cumulative antibiogram
CLSI	Clinical and Laboratory Standards Institute
EMR	Electronic Medical Record
ESBL	Extended spectrum β-lactamase
FDA	Food and Drug Administration
HL7	Health Level 7
LOINC®	Logical Observation Identifiers Names and Codes
MIC	Minimum inhibitory concentration
MRSA	Methicillin-resistant *Staphylococcus aureus*
MUMPS	Massachusetts General Hospital Utility Multi-Programming System
PBP	Penicillin binding protein
PRSP	Penicillin-resistant *Streptococcus pneumoniae*
SNOMED-CT®	Systematized Nomenclature of Medicine – Clinical Terms®
VHA	Veterans Health Administration
VistA	Veterans Information System Technology Architecture

Introduction

Antimicrobial resistance

Antimicrobial therapy represents a signal triumph of 20th century Western medicine. In combination with mass immunization of susceptible populations against common infectious diseases, antimicrobials have dramatically lowered morbidity and mortality from bacterial pneumonia, tuberculosis, human immunodeficiency virus infection, and many other previously untreatable infections (Fauci and Morens 2012). Beyond these direct benefits, anti-infective therapy has revolutionized treatments for many non-infectious conditions that rely on control or cure of infection to achieve their therapeutic goals. Prosthetic joint replacement, cancer chemotherapy, and liver transplantation would not be possible without antimicrobials.

Paradoxically, the success of antimicrobial therapy has become self-defeating. Widespread use of antimicrobials has rendered many antimicrobials less effective now than when they were first introduced. For example, when first developed for clinical practice in the 1940's, penicillin could reliably cure almost all infections caused by *Staphylococcus aureus*. Today, fewer than 10% of infections due to this pathogen can be successfully treated with penicillin (Lowy 2003).

This decrease in antimicrobial efficacy across the population, largely due to a phenomenon termed *antimicrobial resistance* (AR)*, is a major public health problem in

* *Bacterial Resistance* and *Antibiotic Resistance* are sometimes used as synonyms for *Antimicrobial Resistance*. Because these terms are more restrictive (pathogens other than bacteria can develop resistance, and not all antimicrobials are antibiotics), this document uses *Antimicrobial Resistance*. In addition, although the examples used in this capstone and the scope

the United States. In patients with serious bacterial infections, AR-related treatment failure leads to worse clinical outcomes, increased lengths of hospitalization, and significantly higher utilization of health care resources (Fish and Ohlinger 2006).

Decreased drug efficacy at the individual patient level is common in many therapeutic areas. For example, ovarian cancer cells in patients receiving carboplatin are likely to become resistant to this agent with repeated exposure (Stordal, Pavlakis, and Davey 2007). As another example, selective serotonin re-uptake inhibitors can lose their effectiveness over time (Solomon et al. 2005). Treatment failure due to AR is unique, however, because it affects entire populations, not just individual patients, with development of AR in one patient leading to an increased risk of AR infection in other patients.

This phenomenon occurs via emergence of resistant organisms at the cellular level, selection for these organisms at the patient level, and transmission of these microbes at the population level. As discussed below, the selection step occurs to a greater or lesser degrees with different antimicrobials, making the choice of anti-infective therapy an important factor in accelerating or retarding the spread of AR.

Emergence of antimicrobial resistance

Antimicrobials inhibit physiologic processes essential to a microbe's survival, usually by binding to an intracellular molecular target involved in such processes (Gumbo 2011). For example, penicillin acts primarily by binding to penicillin-binding proteins (PBPs) involved in bacterial cell wall synthesis, disrupting their function and killing susceptible bacteria. Antimicrobials are designated as *broad* or *narrow spectrum* depend-

---

of the data model described are restricted to bacteria, the principles apply to other microorganisms such as fungi and viruses.

ing on the extent of their activity against one or more classes of microorganisms. Common bacterial classifications are based on morphology (cocci or bacilli), growth requirements (aerobic or anaerobic conditions), and staining characteristics (Gram stain-positive or negative).

Antimicrobial resistance is a lack of pharmacologic activity of a particular antimicrobial against one or more microbial species at concentrations usually achievable *in vivo*. If the molecular target is absent, a pathogen has *innate* AR to the antimicrobial. For example, *Pseudomonas aeruginosa* is never susceptible to penicillin because it lacks the appropriate PBPs (Alvarez-Ortega *et al*. 2011).

AR is more often *acquired* via mutation in the gene encoding an antimicrobial's target, or acquisition of genetic *resistance factors* encoding proteins that degrade antimicrobials or pump them out of the microbe (Tenover 2006). For example, penicillin resistance is often due to a resistance factor encoding the enzyme β-lactamase, which readily destroys penicillin. Microorganisms readily acquire such factors from other microbes via plasmid-mediated conjugation, frequently by cell-cell interactions in the gastrointestinal tract of an individual patient. Such transfers can occur across species barriers, allowing resistance factors carried by nonpathogenic organisms to be acquired by pathogenic microbes.

A resistance factor may protect an organism against multiple antimicrobials in the same class. Extended-spectrum β-lactamases (ESBLs) cause resistance not just to penicillin G, but also to semi-synthetic penicillins such as piperacillin, which was designed to be resistant to simpler β-lactamases (Jacoby and Munoz-Price 2005). In addition, mobile genetic elements can carry gene cassettes carrying different resistance factors, allowing

rapid transfer of resistance to multiple classes of drugs. Over a hundred such cassettes have been described, conveying resistance to β-lactams, cephalosporins, aminoglycosides, macrolides, and many other antimicrobials (Partridge *et al*. 2009).

This introductory discussion of principles of antimicrobial resistance and empiric antimicrobial therapy provides a basis for consideration of how appropriately presented data, such as antibiograms, might be useful in addressing the significant clinical and public health problems created by AR. In the next section, I will discuss the informatics implications of AR, providing a rationale for the design decisions underlying the data model in this capstone, and a possible implementation schema.

Background

General Principles

Antimicrobials eradicate susceptible organisms, selecting for growth of AR organisms, and increasing the probability of infection by or transmission of such organisms. Normal individuals are colonized by microorganisms of various species in various organs and tissues without infection being present. Although some of these organisms may be pathogenic, other, nonpathogenic microbial species competing for nutrients restrict their growth, a phenomenon known as *colonization resistance* (Vollaard and Clasener 1994). Their ability to cause infection is further limited by immune defenses in the local environment.

For example, although *S. aureus* is a virulent pathogen frequently found on the skin (particularly the nares), it only causes disease if a sufficiently large inoculum exists that can enter subcutaneous tissues, for example through a break in the skin. The same is true for AR organisms; usually (although not always), these pathogens are not more virulent than their more susceptible cousins. Thus, methicillin-resistant strains of *S. aureus* (MRSA), which are resistant to most antimicrobials active against susceptible *S. aureus*, do not cause infection unless they multiply and gain access in sufficient quantities to tissues where they can cause infection[*] (Gordon and Lowy 2008).

Colonization resistance thus represents a major defense mechanism against infection, including infection by AR organisms. It also represents a major defense against transmission of AR organisms from colonized to uncolonized individuals. Thus,

[*] Some MRSA strains are in fact more virulent than more susceptible strains of *S. aureus*, but infection still requires enough growth in the local environment to achieve a sufficient inoculum.

individuals who have intact colonization resistance are much less likely to acquire or develop infection from AR organisms.

Unfortunately, colonization resistance can be easily disrupted. Advanced age, malnutrition, serious illness, and prolonged hospitalization are associated with loss of colonization resistance, overgrowth of AR organisms in previously colonized individuals, and acquisition of AR organisms to uncolonized individuals. In turn, these individuals have a substantially increased risk of serious infections and death. They may also transmit AR organisms to other susceptible individuals.

Antimicrobials have particularly pernicious effects on colonization resistance (Rubin and Samore 2002). First, even an antimicrobial with a narrow spectrum of activity will kill off a significant amount of normal microbial flora, impairing colonization resistance. Second, AR organisms are, by definition, much less likely to be affected by antimicrobials. Thus, antimicrobial use selects for growth of AR organisms. Once established as a dominant species in the local environment of a colonized individual, an AR strain may prevent re-establishment of colonization resistance, even if the inciting antimicrobial has been discontinued. As an example, even a single dose of an antimicrobial can lead to overgrowth of the gut pathogen *Clostridium difficile*, causing life-threatening colitis. This condition is of particular concern because of the attributable mortality it causes, particularly in vulnerable patients with other serious co-morbid conditions (Bajaj *et al*. 2010), and the difficulty of successfully eradicating this pathogen once it has established itself through selection and overgrowth (Gerding, Muto, and Owens 2008).

Spread of antimicrobial resistance

One last step is required for spread of AR organisms – transmission. For this to occur efficiently, colonized and susceptible individuals must be brought together in large numbers, and a route of transmission provided. If one had to design a setting for such transmission, it would be difficult to improve on the modern hospital. The patient population has a heavy representation of seriously ill individuals (particularly in intensive care units), who are likely to have risk factors for loss of colonization resistance, and are extremely likely to undergo interventions that breach normal host defenses and promote infection, particularly by AR organisms, such as use of indwelling vascular catheters and endotracheal intubation. The population consists of both colonized and uncolonized patients who are both cared for by teams of health care workers (HCWs) that can serve as vectors for transmission of AR organisms. The situation is exacerbated by intermittent adherence to hand washing by HCWs (Harris *et al*. 2000).

This fertile soil is almost literally watered by large-scale use of antimicrobials. As discussed above, any antimicrobial use may affect colonization resistance, but the classes of antimicrobials used among hospitalized patients – parenterally administered agents with broad antimicrobial spectra – are particularly likely to disrupt colonization resistance and promote acquisition of AR organisms.

While the hospital has traditionally been the venue for heavy use of antimicrobials, it is worth noting that over the last quarter-century, the use of these agents has increased dramatically in the outpatient setting as well, usually in the setting of acute respiratory tract infections (McCaig and Hughes 1995). While this has had the advantage of avoiding admission for parenteral antimicrobial therapy, it has also led to an increasing preva-

lence of colonization of individuals by AR organisms. For example, prior receipt of oral antimicrobials is an extremely strong predictor of colonization by penicillin-resistant *Streptococcus pneumoniae* in children (Samore *et al*. 2001).

Empiric antimicrobial therapy and antimicrobial resistance

The epidemiologic mechanisms underlying emergence and spread of antimicrobial resistance has created an interesting clinical and public health problem. In most clinical situations where there are multiple drugs available to treat a particular condition, the choice depends on patient-specific factors: how effective the drug is likely to be in a particular patient, the risk of toxicity, and the cost. However, in the case of an infectious disease, the choice of agent may affect the risk to the population as a whole. This point is nicely illustrated by the correlation between the level of fluoroquinolone use in Canada and the prevalence of resistance to this class of agents in isolates of *Streptococcus pneumoniae*, a virulent pathogen that is the primary cause of community-acquired bacterial pneumonia (Adam *et al*. 2009). In addition, as discussed above, antimicrobial therapy may treat an infection but select for AR organisms, raising a treated patient's risk for future problems.

The problem is exacerbated by the nonspecific clinical presentation for many acute infections requiring anti-microbial therapy, leading to selection of antimicrobial therapy on empiric (as opposed to microbiologic) grounds. For example, although there are relatively few bacterial etiologies of community-acquired pneumonia in immunocompetent individuals, it is usually not possible to distinguish between them solely on the basis of clinical signs and symptoms. Sputum and blood cultures are relatively insensitive in this setting, and in any event do not yield results for days, if at all. Thus, initial antibacterial

therapy for this infection is usually chosen so as to cover all reasonable possibilities, even though only one of them is likely to be responsible (Mandell et al. 2007). Even if a specific organism is identified, it is extremely common for the broad initial treatment to be continued, even if narrower-spectrum antimicrobials targeting the identified pathogen are available.

Antimicrobials used for empiric therapy, because of their broader spectrum of activity, differ in their effects on colonization resistance (Rice *et al*. 2004). Agents with a relatively narrow antimicrobial spectrum have less effect than do broader spectrum agents (Rice 2012). Other factors, such as the antimicrobial concentrations achieved in various body fluids and an antimicrobial's half-life, may also contribute to a given antimicrobial's relative influence on colonization resistance.

At the same time, while there is a common perception that the broader spectrum antimicrobials used for empiric therapy or for treatment of infections due to AR organisms are "better" or "stronger" than narrower-spectrum antimicrobials for a specific infection, this is not correct. The overwhelming majority of antimicrobials marketed in the U.S. are approved on the basis of noninferiority trials, which only require showing that a new agent is not much worse than existing therapy (Fleming and Powers 2008). Even in cases where a new antimicrobial appears to have better activity *in vitro* against AR organisms, this does not imply superior clinical activity. For example, telithromycin, a ketolide antimicrobial, was widely marketed for empiric therapy of community-acquired respiratory tract infections due to resistant pathogens, despite the absence of clinical trial data to support this claim. Unfortunately, telithromycin turned out to cause se-

vere hepatotoxicity, thus increasing the risk to patients without any corresponding clinical benefit (Ross 2007).

Thus, empiric (or microbiologically defined) therapy with inappropriately broad spectrum antimicrobials is unlikely to yield a clinical benefit, but increases the risk of selection for, colonization by, or infection with AR organisms. Despite this, physicians frequently feel compelled to choose overly broad spectrum antimicrobials because of anxiety over the consequences of "missing" a resistant pathogen, pressure from patients, or lack of information about the characteristics of different antimicrobials and their appropriate use (DiNubile 1990). The last factor is of particular interest to medical informaticists because of the potential application of decision support systems to this gap.

Antimicrobial stewardship

Multiple integrated public health strategies have been devised to address AR, such as development of new drugs and increased attention to hand hygiene and other infection control measures. However, *antimicrobial stewardship* – organizing health care delivery in both the hospital and community to maximize appropriate use of antimicrobials – holds particular promise for combatting AR (Tamma and Cosgrove 2011). Antimicrobial stewardship refers to a discrete set of structures, processes, and outcomes used to optimize appropriate use of antimicrobials. When properly designed, implemented, and executed, a stewardship program can decrease emergence of ARs, improve treatment outcomes, avoid toxicities associated with antimicrobial use, and lower health care costs (Goff 2011). The Infectious Diseases Society of America and Society for Hospital Epidemiology of America have jointly issued evidence-based guidelines on design and implementation of antibiotic stewardship programs (Dellit *et al*. 2007).

Although evaluation of a stewardship program's effectiveness relies on multiple process and intermediate outcome measures, the prevalence of AR in a particular practice setting is a key metric. In addition, clinically relevant AR prevalence data is itself a key stewardship tool, providing decision support for clinicians choosing empiric antimicrobial therapy, allowing them to integrate the probability that a given patient has an infection due to an AR organism with other data, and avoid unnecessarily broad treatment.

The antibiogram – decision support for microbiologically defined therapy

An *antibiogram*[*] (ABGM) is the profile of susceptibility or resistance of an organism isolated from a patient (an *isolate*) to a panel of antimicrobial agents, usually those that are clinically relevant to treatment of infections caused by the particular species isolated. ABGMs are produced from data derived by antimicrobial susceptibility testing (AST). AST consists of testing the *in vitro* activity of an antimicrobial against a pure strain of an organism isolated from a patient, often using automated analyzers[†]. Depending on the technique used, the *in vitro* activity may be expressed as the lowest concentration of the antimicrobial that inhibits growth of the organism by 99.9% (the *minimum inhibitory concentration*, or *MIC*, which is usually expressed as μg/mL). The lower the MIC, the more susceptible the isolate is to the antimicrobial. Alternatively, the antimicrobial's activity against an isolate may be measured by the extent to which microbial growth is inhibited in the area surrounding a paper disc impregnated with an antimicro-

[*] The term *antibiogram* generally refers to susceptibility profiles for bacteria; profiles for other classes of microorganisms are designated by terms such as *phenotyping*. Because of its common usage in the literature, I will use antibiogram throughout this document, but the principles apply to any susceptibility profile.

[†] AST can also be performed against non-clinical isolates, such as those cultured from environmental surfaces. For simplicity, this discussion will focus on clinical isolates.

bial, usually expressed as millimeters; the larger the zone of inhibition, the more susceptible the isolate is to the antimicrobial (CLSI 2012).

The AST result is then compared to the activity of the antimicrobial against reference strains of known susceptibility, using reference values called *breakpoints* established during the development and clinical testing of the drug, and translated into an *interpretation*. Most antimicrobials have two interpretative breakpoints for a given microbial species, a *susceptibility breakpoint* and a *resistance breakpoint*. If the antimicrobial's activity against the isolate (expressed as either the MIC or zone diameter) is equal to or greater than the susceptibility breakpoint for the corresponding reference strain, than the isolate is *susceptible* to the antimicrobial. If the activity is less than the resistance breakpoint, the isolate is *resistant.* If the activity is between the breakpoints, then the isolate is *intermediate*. In some cases, no resistance breakpoint has been defined for the activity of an antimicrobial, usually because there is insufficient data. In such cases, an antimicrobial that does not reach the susceptibility breakpoint is defined as *nonsusceptible.*

AST methods are relatively standardized, with the overwhelming majority of microbiology laboratories in the U.S. following standards published by the Clinical Laboratory Standards Institute (CLSI), an ISO-certified standards organization. CLSI also publishes breakpoints for interpreting AST results by correlating *in vitro* antimicrobial activity with clinical outcomes[*]. CLSI regularly revises breakpoint determinations based on new clinical, epidemiologic, and microbiologic data, and publishes updated interpretations, usually annually. For example, if surveys of clinical isolates of *S. pneu-*

[*] The U.S. Food and Drug Administration independently determines breakpoints for antimicrobials and uses them for the approved labeling for antimicrobials. FDA breakpoints are usually, but not always, identical to those published by CLSI. For simplicity, I will only consider CLSI breakpoints.

moniae show an increase in the median MIC, CLSI may revise the susceptibility break-point accordingly.

It's important to note that not all antimicrobials are tested against all isolates. As discussed earlier, a particular microbial species may have innate AR to a particular antimicrobial, making AST against that drug pointless. AST generally uses selective testing based on features such as the isolate's morphology, growth characteristics, Gram staining characteristics, and the presumed location of the infection. For example, as discussed above, *E. coli* is never susceptible to oxacillin, so that AST would never be performed using this drug against *E. coli*. However, it is frequently active against *S. aureus*, making it part of the standard AST panel tested against this species.

In the case of bacteria, standard AST panels differ between different classes of organisms. For a given class, the AST panel includes antimicrobials that almost always active against that class, but which should be reserved for infections in which the pathogen is resistant to narrower spectrum antimicrobials. For example, isolates of *E. coli* are almost always susceptible to ertapenem, a broad-spectrum anti-infective in the carbapenem class, but if a particular isolate is known to be also susceptible to ampicillin – a narrower spectrum agent – the latter agent is preferable because it is much less likely to select for AR organisms or cause *C. difficile* colitis. Ampicillin would also have similar clinical activity to ertapenem (because of the use of NI trials for antimicrobial development, most FDA-approved antimicrobials have comparable efficacy for similar indications), and would cost much less. However, because of the inaccurate perception that broader spectrum antimicrobials are "better", many clinicians are likely to choose

ertapenem over ampicillin in such a situation, despite the lack of added benefit and the increased risks.

To address this, CLSI divides antimicrobials into *report groups* specific to particular classes of organisms (CLSI 2012). Group A consists of antimicrobials that may always be used for initial testing (*primary* antimicrobials), and for which the results should always be reported. Group B consists of antimicrobials that may be used for initial testing, but for which the results should be reported *selectively* to clinicians, in those instances in which the isolate is resistant to Group A agents. Group C consists of *supplemental* antimicrobials, which should be used for AST only in instances or institutions where there is a high prevalence of AR; the results for these should also be reported selectively. Group U consists of antimicrobials used for AST of isolates from urine cultures, on the basis of their pharmacokinetics; usually, only Group U AST is performed for urine cultures[*]. Not all classes of organisms have all groups.

A particular antimicrobial may be assigned to different reporting groups, depending on the class of organism. For example, ampicillin/sulbactam, which combines ampicillin with a β-lactamase inhibitor to allow activity against β-lactamase-producing organisms, is in Group B for *E. coli* for aerobic Gram-negative enteric bacteria (*Enterobacteriaceae*), but is in Group A for anaerobes that always express β-lactamase, such as *Bacteroides fragilis*.

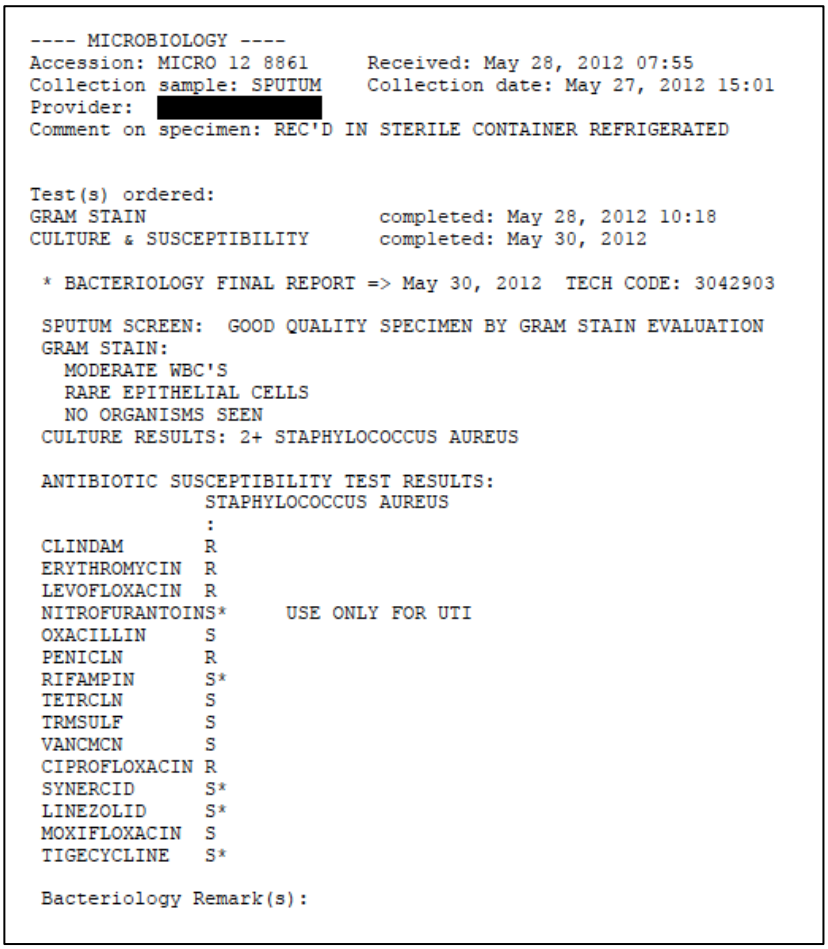
Furthermore, within any given group, the antimicrobials are arranged by class, with laboratories having the option of using one of several agents from a particular class.

[*] The CLSI classification also includes Groups O ('Other') and Inv ('Investigational'). These are rarely used in routine clinical microbiology practice, and will not be included in the scope of the data model. However, expansion of the model to include these would be straightforward.

For example, in the example just mentioned, that of AST against *B. fragilis*, a laboratory may use any one of four different β-lactam/β-lactamase inhibitor drugs for AST of *B. fragilis* or *E. coli*, not just ampicillin/sulbactam. This introduces an obvious source of variability in AST results between laboratories.

Once AST has been performed and the results interpreted, a single isolate ABGM can be constructed, as shown in Figure 1. These data are extremely useful to clinicians, guiding them to antimicrobials known to be active against the isolate infecting a patient, and allowing them to avoid antimicrobials that are not (Lambke 2012). When generating ABGMs for clinical use, microbiology laboratories generally employ *selective reporting*, displaying results for broader spectrum antimicrobials only if the isolate is resistant to more commonly used, narrower spectrum antimicrobials (Pakyz 2007). In addition, the ABGM is typically not displayed for isolates that are identical to previous isolates from the same patient. For example, a patient with *S. aureus* endocarditis may have persistent bacteremia with this organism and repeatedly positive blood cultures; the microbiology lab will typically only report the ABGM for the first *S. aureus* isolate.

Figure 1. An antibiogram for a single isolate from a single culture from a single patient. The asterisks indicate antimicrobials for which AST results are selectively reported. The hospital where this report was produced generally reports all AST results for *S. aureus* isolates because of the high prevalence of resistant strains.



```
---- MICROBIOLOGY ----
Accession: MICRO 12 8861     Received: May 28, 2012 07:55
Collection sample: SPUTUM    Collection date: May 27, 2012 15:01
Provider:  [redacted]
Comment on specimen: REC'D IN STERILE CONTAINER REFRIGERATED


Test(s) ordered:
GRAM STAIN                    completed: May 28, 2012 10:18
CULTURE & SUSCEPTIBILITY      completed: May 30, 2012

 * BACTERIOLOGY FINAL REPORT => May 30, 2012  TECH CODE: 3042903

 SPUTUM SCREEN:  GOOD QUALITY SPECIMEN BY GRAM STAIN EVALUATION
 GRAM STAIN:
   MODERATE WBC'S
   RARE EPITHELIAL CELLS
   NO ORGANISMS SEEN
 CULTURE RESULTS: 2+ STAPHYLOCOCCUS AUREUS

 ANTIBIOTIC SUSCEPTIBILITY TEST RESULTS:
               STAPHYLOCOCCUS AUREUS
               :
 CLINDAM       R
 ERYTHROMYCIN  R
 LEVOFLOXACIN  R
 NITROFURANTOINS*     USE ONLY FOR UTI
 OXACILLIN     S
 PENICLN       R
 RIFAMPIN      S*
 TETRCLN       S
 TRMSULF       S
 VANCMCN       S
 CIPROFLOXACIN R
 SYNERCID      S*
 LINEZOLID     S*
 MOXIFLOXACIN  S
 TIGECYCLINE   S*

 Bacteriology Remark(s):
```

The cumulative antibiogram – a broader decision support tool

Results from ABGMs for different isolates can be aggregated by species to create a *cumulative antibiogram* (CABGM), consisting of a tabular presentation of susceptibility rates for a defined set of pathogens to a defined set of antimicrobial agents over a defined period of time (For clarity, I will refer to any ABGM containing data on multiple isolates as a CABGM.) For a given species, the rates shown are derived from the standard

AST panel for that species. If generated for clinical use, CABGMs generally employ selective reporting and only report data to clinicians based on the first isolate of a particular species from a particular patient. This avoids introduction of bias caused by more frequent microbiologic testing of sicker patients, which would tend to inflate resistance rates. This first-isolate-only approach has been validated by comparison with results from CABGMs reporting all results (Shannon and French 2002). For epidemiologic purposes, CABGMs can be constructed that do include all isolates, including duplicates, in the numerator and denominator.

Figure 2. Sample CABGM summarizing susceptibility rates for selected pathogens, aggregating data for multiple isolates grown from multiple cultures obtained from multiple patients. This report follows CLSI guidelines for analysis and presentation of cumulative susceptibility data (CDC 2004). The CABGM is produced annually by manual methods and is distributed as a paper product; it is also available as a Web version.

		Percent Sensitive										
ORGANISM	**TOTAL #ISO***	**AMP**	**ZOS**	**ERTA**	**CIP**	**T/S**	**CZOL**	**CTRI**	**CPIM**	**GEN**	**TOB**	**AMI**
Enterobacter cloacae	30	**NA**	**NA**	87	97	77	**NA**	**NA**	100	100	100	100
***Escherichia coli*†**	75	56	93	100	73	71	58	88	89	92	87	100
***Klebsiella oxytoca*†**	18	0	88	93	100	100	36	94	94	100	100	100
***Klebsiella pneumoniae*†**	41	0	100	96	100	83	58	93	93	100	93	100
Proteus mirabilis	18	78	100	100	78	67	0	78	78	94	100	100
Pseudomonas aeruginosa	47	**NA**	93	**NA**	85	**NA**	**NA**	**NA**	96	87	96	96
Serratia marcescens	22	**NA**	**NA**	100	100	95	**NA**	**NA**	100	100	100	100
				VANC	**OXAC**	**T/S**	**ERYT**	**CLIN**	**TCN**			
Staphylococcus aureus	297			100	55	96	45	78	92			
MRSA	133			100	0	95	**NA**	69	92			
MSSA	164			100	100	97	**NA**	85	92			
***Staphylococcus* coag neg**	197			100	52	73	50	69	84			

NA = not available
AMP - ampicillin, ZOS - Zosyn, ERTA - ertapenem, CIP - ciprofloxacin, T/S - trimethoprim/sulfamethoxazole, CZOL - cefazolin, CTRI - ceftriaxone, CPIM - cefepime, ERTA - ertapenem, GEN - gentamicin, TOB - tobramycin, AMI - amikacin, VANC - vancomycin, OXAC - oxacillin, ERYT - erythromycin, CLIN - clindamycin, TCN - tetracycline
*Statistical validity of % susceptible is decreased if fewer than 30 isolates are tested.
14.8% (67/454) of all enterococcal isolates were vancomycin-resistant
†Extended-Spectrum Beta-Lactamase (ESBL) Positive - *E. coli - 9%; K. oxytoca - 6%; K. pneumoniae - 7%*

CABGMs serve multiple purposes for different users:

- Individual clinicians: Decision support for choosing empiric antimicrobial therapy. In combination with the other components of an antimicrobial stewardship program, this can be a powerful tool for choosing empiric therapy that has an antimicrobial spectrum broad enough to cover likely pathogens while minimizing the risk of selecting for AR organisms by using overly broad anti-infective treatment. For example, a clinician treating a patient suspected (but not confirmed) of having bloodstream infection due to *E. coli* could use the CABGM in Figure 2 to choose ceftriaxone for initial treatment, in place of a broader-spectrum agent such as ertapenem that carries a higher risk of selecting for resistant organisms or causing *C. difficile* colitis.
- Hospital epidemiologists, infectious disease pharmacists, microbiology laboratory directors, and public health departments: Surveillance of AR trends in individual medical facilities, communities, geographic region, countries, or even continents. Monitoring of such trends is essential in constructing and revising evidence-based recommendations on treatment of infectious diseases (Critchley and Karlowsky 2004).
- Quality managers and health care organization executives: Evaluation of the effectiveness of components of antimicrobial stewardship programs, such as policies on antimicrobial utilization and educational campaigns, as well as correlating AR rates with other outcome measures such as antimicrobial utilization, treatment outcomes, and costs.

Practical issues with antibiogram construction

While design and construction of a CABGM is seemingly straightforward, especially in a hospital using an electronic medical record, the reality is far different due to the fragmentation of information systems and databases in a typical health care organization and the lack of automation of many processes, and CABGM's are rarely the same from one hospital to the next.

Some of this variability reflects the simultaneous existence of multiple accepted methods for AST testing and reporting. For example, one laboratory may perform testing by MIC determination while another uses disk diffusion. Labs may use similar but nonidentical panels for AST on the same microbial species.

However, these issues are remediable, and are nowhere as serious as the interoperability problems created by the typical workflow for capturing AST data into an electronic medical record and constructing a CABGM:

1. A culture received in the microbiology laboratory is assigned a unique identifier (the accession number, which is entered manually into the EMR, along with patient data. The accession number is manually entered by a separate proces into the Laboratory Information System (LIS).
2. If the culture is positive, the LIS is electronically updated and the EMR is manually updated.
3. Organisms grown from the culture are speciated by a mixture of semi-automated and manual methods. Intermediate results are recorded manually on paper.
4. AST is performed on the isolate (or isolates) obtained from the culture by a mixture of automated and manual methods. If an automated AST analyzer is used,

the results are captured electronically in the LIS, but are entered manually into the EMR.

5. Cumulative antibiograms are produced by a series of manual steps: Semi-automated searching within the EMR for AST results by species, manual export of single isolate ABGM results into a spreadsheet, manual deletion of duplicate isolates from the same patient, manual deletion of AST results for broad spectrum antimicrobials to allow selective reporting, manual export of associated patient data into the spreadsheet, generation of frequency counts, susceptibility rates, and resistance rates, and production of a list of organisms and susceptibility rates to selected organisms.
6. Production of a CABGM as a paper product.

From an informatics perspective, this process has the following flaws:

- Significant variations in vocabulary exist between different laboratories (and even within the same laboratory over time).
- Data and database models are not standardized
- Analyses are not standardized and are potentially inaccurate
- Data displays are not standardized and are potentially misleading data
- There is a high potential for entry of invalid data

These issues substantially decrease the utility of CABGMs. Problems include the following:

- Construction of potentially inaccurate and/or ineffective CABGMs
- Difficulty tracking AR trends at the local, regional, and national levels

- Difficulty correlating AR data with data outside the LIS
- Unable to compare AR rates across different institutions
- Inability to meaningfully aggregate AR data from different institutions

These are not just theoretical problems. Surveys have demonstrated substantial variability in CABGMs across the U.S. (Zapantis *et al*. 2005; Lautenbach and Nachamkin 2006)., reflecting a lack of standardization. Even worse, it is not uncommon for CABGMs to report inaccurate results (even when compiled from accurate AST data), or present AR prevalence rates that are biased by oversampling of seriously ill patients.

Although CLSI has issued multiple revisions of a standard designated M39 for collecting, analyzing, and presenting AR data for CABGMs (CLSI 2009), the surveys described above demonstrate that they are rarely followed, even though CLSI is the recognized standard-setting organization for AST. Examination of CABGM survey results suggests that the variability is due instead to lack of an interoperable data model and vocabularies constraining collection, manipulation, and analysis of data used for construction of CABGMs. Laboratory information systems are often isolated from electronic medical records. Names for pathogens and antimicrobials are often inconsistent between labs.

This suggests that a first step in resolving the interoperability problems involved in CABGM is construction of a data model reflecting the M39 standard, based on common use cases. With this as a starting point, relational database models can be created, schema for importing data from EMRs can be devised and implemented, and applications for constructing CABGMs can be built.

Prior work

There is an extensive body of work on clinical decision support tools to guide antimicrobial prescribing, starting in the mid-1970's with MYCIN (Wraith *et al*. 1976). One of the first clinical decision support tools constructed for any therapeutic area, MYCIN was an application that provided recommendations to clinicians on appropriate antimicrobial therapy, based on a built-in knowledge base of several hundred rules, an inference engine, and data provided by the clinician. Interestingly, recommendations made by MYCIN were appropriate in 69% of cases, compared to a range of 42.5% to 62.5% for infectious disease physicians (Yu *et al*. 1979).

Since that time, there has been an enormous amount of effort expended in constructing point-of-care tools to improve the quality of empiric antimicrobial prescribing (Kilroy et al. 1984; Mullett *et al*. 2004; Sintchenko *et al*. 2005; Rubin *et al*. 2006; Thursky 2006; Thursky and Mahemoff 2007; Buising *et al*. 2008), with data to support their utility in both the inpatient and outpatient settings (Samore *et al*. 2005; Leibovici *et al*. 2007). There has also been theoretical work done to define an ontology of antimicrobial resistance (Schober *et al*. 2010).

However, these efforts have not addressed the issue of standardization of antibiograms. Of note, MYCIN was never used as a production system, primarily because of the need for manual entry of clinical and microbiologic data (an early incarnation of the interoperability problem described above).

There are currently no domain or refined message data models specific to CABGM construction. Existing or candidate standards relevant to CABGM creation include a draft Clinical Document Architecture standard for submission of hospital-

acquired infection reports to CDC's National Healthcare Safety Network (HL7 2009); HL7 2.7-compliant laboratory reporting standards, as well as HL7 2.5-compliant implementation guides for electronic laboratory and public health reporting; and an HL7 draft template for antibiotic result templates. Relevant vocabularies include LOINC, SNOMED-CT, and the Public Health Information Network (PHIN) Vocabulary.

The Veterans Health Administration (VHA) as an implementation model

The Veterans Health Administration (VHA) represents a useful system for implementing HL7-based standards to generate CABGMs from different health care facilities. VHA operates the largest civilian health-care organization in the U.S., comprising 152 medical centers and more than 800 outpatient clinics located across the country. VHA provides care to 5 million Veterans, with 70 million outpatient visits annually.

All VHA medical facilities use an electronic health record system, the Veterans Health Information Systems and Technology Architecture, or VistA (Brown *et al*. 2003). VistA was developed using MUMPS, and consists of over 100 software modules, providing functions such as computerized provider order entry, picture archiving and communication, and display of progress notes and laboratory results. VistA employs a client-server architecture, with installation and maintenance of data and applications managed by local facilities.

Although the overall architecture of VistA applications and data files is similar between facilities, there is considerable flexibility in how data can be entered and stored. This is especially true in the case of antimicrobial susceptibility data. For example, susceptibility results are entered as free text, with a recommendation that the entry consist

of either a numeric result (the MIC) or a character (e.g., ‘S’, ‘I’, or ‘R’), along with a free text interpretation. Individual facilities can choose their own nomenclature for adding newly introduced antimicrobials. To complicate matters, there are not necessarily any procedures in place for validation, error checking, or error handling.

These features create substantial variability in microbiology datasets throughout VA, complicating the task of aggregating AR data to generate CABGMs at the national, regional, and local levels. This is perhaps best illustrated by the dramatically different results obtained within the VHA system by three studies of the effects of interventions to control the incidence of infections due to methicillin-resistant *S. aureus* (Jain *et al*. 2011; Kennedy *et al*. 2010; Tracy *et al.* 2011). The first, a national level study aggregating unstandardized, unvalidated AR data, concluded that use of of active surveillance cultures to detect MRSA colonization and application of contact precautions to colonized patients decreased infections due to MRSA. The other two studies, reporting single-center experiences (with automatic standardization of AR data) did not find the same benefit. Clearly, the inability to standardize and validate underlying AST data severely hampers any effort to define the benefit of this or other infection control and antimicrobial stewardship practices.(Graves *et al*. 2011). However, this situation is not unique to VHA, and the breadth of experience within VHA with an EMR integrating clinical and microbiologic data makes it a good platform for understanding and modeling the real-life challenges of constructing accurate, standardized CABGMs.

In this capstone, I define a standardized data model for specific use cases, along with appropriate business rules for data manipulation. I then present representative examples of the current universe of data structures currently employed within VHA to organ-

ize AR data, and describe a schema for migrating VHA data from its current unstructured, non-standardized form to a relational database. I then discuss how this could be feasibly implemented, allowing design and construction of applications to produce CABGMs at different levels of specificity within VHA.

Methods

The objectives of this project are to:

1. Create a data model and associated data dictionary for generating CABGMs for specific use cases, based on existing standards and relevant vocabularies
2. Describe business rules that will allow constrain applications for generating CABGMs
3. Describe a schema processes for cleaning, transforming, and mapping existing AR data to a common data structure, using the VHA as a representative platform.

The scope of the data model includes clinically relevant bacterial organisms for which antimicrobial susceptibility testing guidelines have been issued by CLSI. Fungi, viruses, and parasites are excluded. Antibacterials for which breakpoints have been established by CLSI are included. Investigational agents and other classes of antimicrobials (e.g., antifungals and antivirals) are excluded.

Standards

Standards used in the data model include those published by CLSI, SNOMED-CT, and LOINC. The overwhelming majority of clinical microbiology laboratories in the U.S. follow standards published by the CLSI for antimicrobial susceptibility testing and reporting. CLSI document M100-S22 (CLSI 2012) is an appropriate standard; this is the latest in a series of annual updates to the basic technical standards. With regard to modeling data requirements for cumulative antibiograms, CLSI document M39-A3 (CLSI 2009) is the most recently published standard for designing, formatting, and reporting CABGMs.

LOINC version 2.38 provides a standard vocabulary for antimicrobial susceptibility tests and tests for specific resistance factors. LOINC provides a standardized, unambiguous set of identifiers for these concepts.

SNOMED-CT (31 Jan 2012 Release) provides a standard vocabulary for culture specimen type and collection site, Isolates, and reference bacterial strains used for determining breakpoints. SNOMED-CT provides a standardized, unambiguous set of identifiers for these concepts.

Users and use cases

This data model is designed for two sets of end-users: clinicians choosing empiric antibacterial therapy, and hospital epidemiologists (or other antimicrobial stewardship personnel) tracking AR prevalence rates. Three use cases are described (Boxes 1-3). The first illustrates typical use by a clinician. The second and third describe some of the specialized uses by hospital epidemiologists or other stewardship personnel.

Box 1. Use case: Choosing empiric antimicrobial therapy.

Choosing empiric antimicrobial therapy

Scenario: A physician is caring for a patient who has developed hospital-acquired pneumonia; no pathogen has been identified. The patient is at risk for infection by resistant Gram-negative bacteria, anaerobes, and *Staphylococcus aureus.* The patient has had previous episodes of *Clostridium difficile* colitis due to prolonged treatment with antimicrobials. The patient also has renal insufficiency. The physician would like to choose antimicrobials that have a high probability of active against likely pathogens, while avoiding renal toxicity to avoid further compromising the patient's kidney function. The physician has reasonable knowledge regarding the relative nephrotoxicity of different agents. However, the physician does not have specific training or expertise regarding the antimicrobial spectra of the possible choices, and is not aware that different anti-infectives are associated with different risks for *C. difficile* colitis

Actors:

- Physician
- System

Steps

a. User action: The physician launches a CABGM viewer application based on the data model, via Electronic Medical Record (EMR) application. Alternatively, the system could be configured to launch the viewer automatically if the physician starts to order an antimicrobial and there are no recent AST results.

b. System response: The application displays a table similar to that shown in Figure 2, showing susceptibility rates for common nosocomial pathogens. The table is formatted so as to highlight antimicrobials associated with a lower risk of *C. difficile* colitis, as well as antimicrobials with higher susceptibility rates against various pathogens.

i. The display appears within a few seconds after launching the application.

ii. Ideally, the view can be filtered to present results for a particular unit (e.g., the medical intensive care unit) or particular patient demographics (e.g., patients 65 years of age and older).

c. User action: Based on the data in the table and knowledge about the nephrotoxicity of the various agents, the physician orders empiric antimicrobial therapy that maximizes the probability of activity against possible hospital-acquired pneumonia etiologies and minimizes the risks described above.

Box 2. Use case: Validating AST data in the microbiology laboratory

Validating AST data in the microbiology laboratory

Scenario: The director of a hospital microbiology laboratory performs weekly audits of the quality and consistency of AST data manually transferred from the Laboratory Information System to the clinical database supporting the hospital's EMR.

Actors:

- Lab Director
- System

Steps:

d. User Action: The director launches an AST Consistency Checker based on the data model.

e. System Response: The Checker generates a line listing of all isolates with AST results suggesting internally inconsistent data (e.g., an interpretation of Resistant for an isolate with an MIC below the susceptibility breakpoint, or MRSA isolates reported as susceptible to cephalosporins), incomplete data (e.g., absence of results for the primary AST panel), or inappropriate testing (e.g., testing with secondary agents when the isolate is susceptible to primary agents).

f. User Action: Based on the report, the director:

i. Reviews the flagged isolates

ii. Arranges for correction of the data where appropriate

iii. Uses the results for quality improvement by identifying patterns of data errors or inconsistencies (e.g., identifying inappropriate AST testing requested by specific clinicians).

Box 3. Use case: Evaluating an antimicrobial stewardship intervention

Evaluating an antimicrobial stewardship intervention.

Scenario: The Chief of Clinical Pharmacy is evaluating whether a policy of requiring prior approval for particular antimicrobials has affected the prevalence of multi-drug resistant Gram-negative bacilli within the hospital over a 36-month period.

Actors:

- Chief of Clinical Pharmacy
- System

Steps:

g. User Action: The pharmacy chief launches an AR Trend Analyzer that is based on the data model.

h. System Response: The Analyzer displays graphs showing rates of susceptibility, resistance, and nonsusceptibility over the time period. The display includes data for all antimicrobials (primary, supplemental, and secondary).

i. Ideally, the view can be filtered to present results for a particular unit (e.g., the medical intensive care unit) or particular patient demographics (e.g., patients 65 years of age and older).

ii. Additional features could include simultaneous graphing of utilization of specific antimicrobials or classes of antimicrobials.

The CABGM Data Model

The CABGM Data Model is described here in several ways. First, entity-relationship diagrams illustrate the data model in three sections: those dealing with general health care delivery (Figure 3), those dealing with microbiologic specimen collection and processing (Figure 4), and those dealing with antimicrobial susceptibility testing (Figure 5).. Table 1 then shows the complete list of entities and relationships. Table 2 shows a data dictionary with entities, their attributes, and attribute data type. Table 3 shows allowable values for attributes with value constraints. Following these diagrams and tables are text explanations of each model, with a narrative description of relevant entities, their characteristics (including how they are uniquely identified), and their relationships. Entity-relationship diagrams were constructed with ER-Assistant 2.10 (Mosor, Inc.).

Figure 3. Clinical entities, attributes, and relationships.

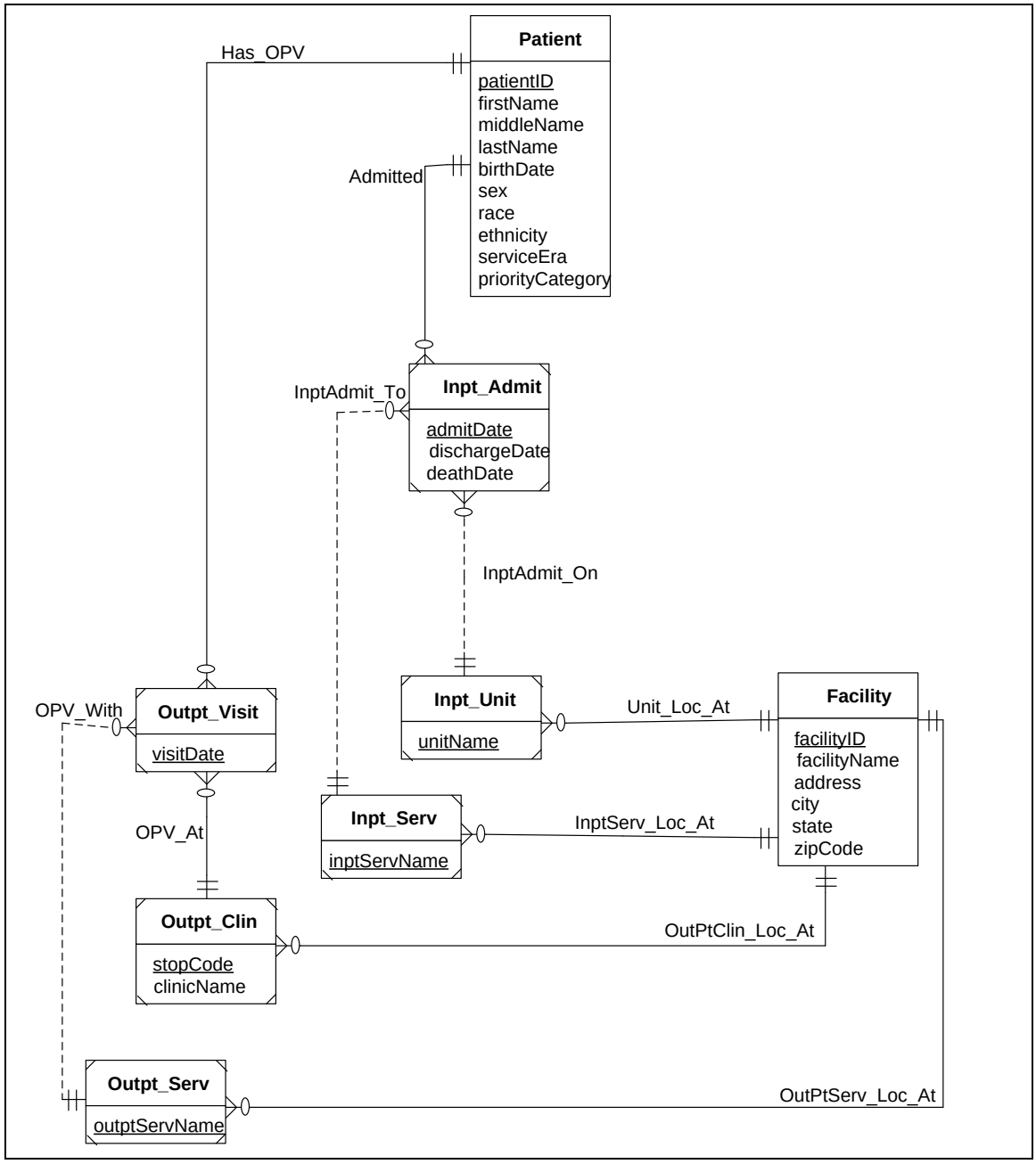


Figure 4. Microbiologic entities, attributes, and relationships.

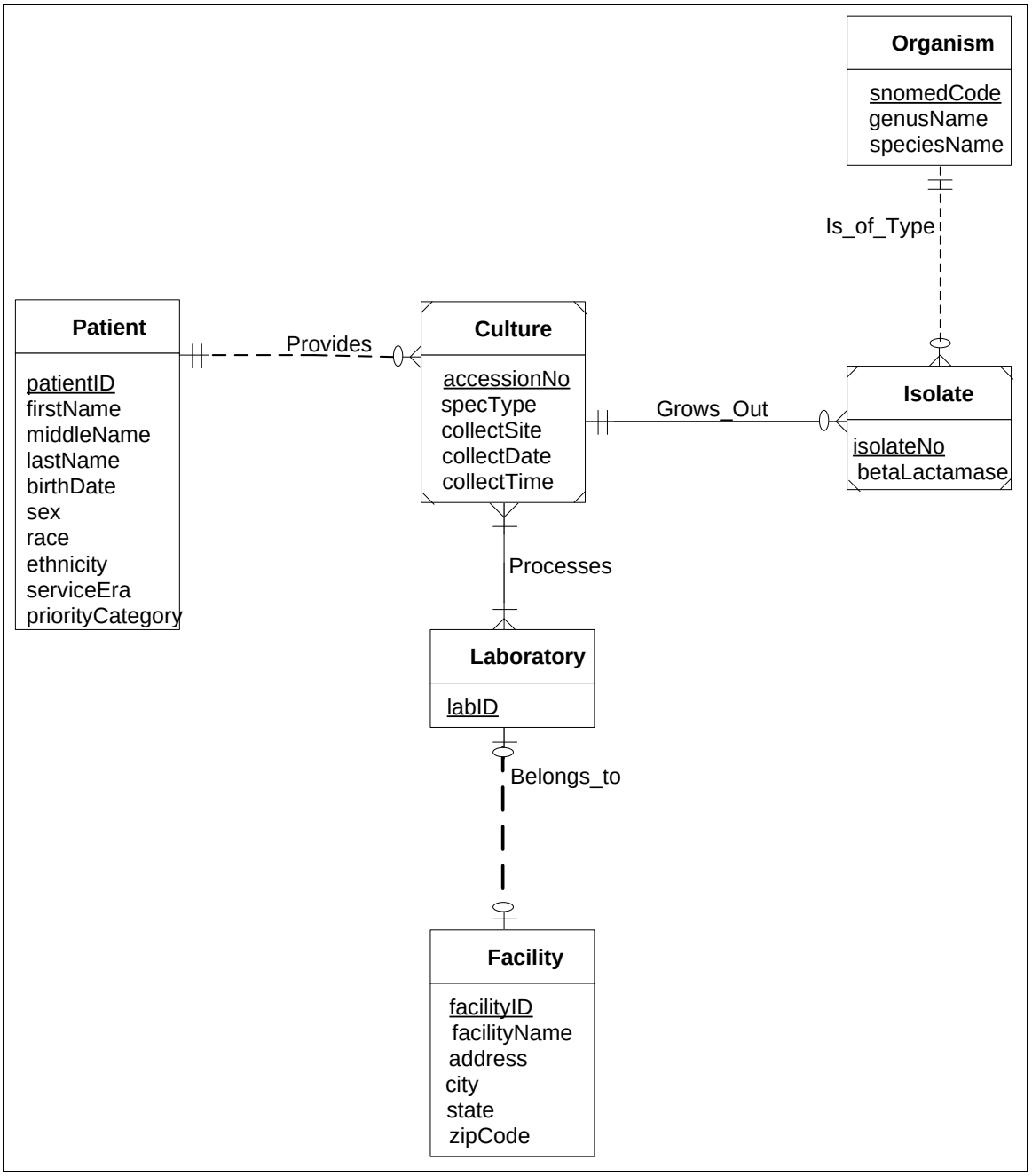


Figure 5. AST entities, attributes, and relationships.

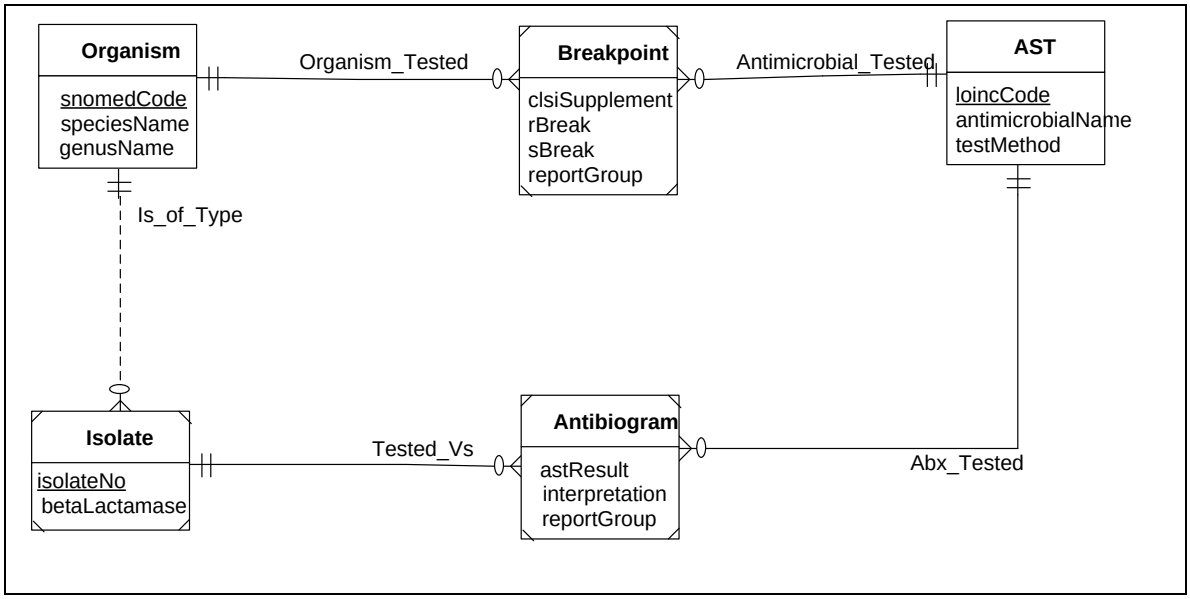


Table 1. The CABGM Data Model. Entities and Relationships.
Strong Entities are bolded; weak entities are in regular type. The primary key is underlined. Identifying relationships are bolded.

Entity	Description	Attributes	Relationships and cardinality
Patient	Individual patient	**patientID** firstName middleName lastName birthDate sex race ethnicity serviceEra priorityCategory	o ***Admitted*** to Inpt _Admit (1-1) o ***Has_OPV*** to Outpt_Visit (1-1) o *Provides* to Culture (1-1)

Facility	VHA Facility	**facilityID** facilityName address city state zipCode	o ***Unit_Loc_At*** to Inpt_Unit (1-1) o ***InptServ_Loc_At*** to Inpt_Serv (1-1) o ***OutptClin_Loc_At*** to Outpt_Clin (1-1) o ***OutptServ_Loc_At*** to Outpt_Serv (1-1) o *Belongs_To* to **Lab** (0-1)
Inpt_Unit	Inpatient unit at a **Facility**	unitName	o ***Unit_Loc_At*** from **Facility** (0-M) o *InptAdmit_On* to Inpt Admit (0-M)
Outpt_Clin	Outpatient clinic at a **Facility**	clinicName stopCode	o ***OutptClin_Loc_At*** from **Facility** (0-M) o *OPV_At* to Outpt_Visit (0-M)
Inpt_Serv	Inpatient Service at a Facility	inptServName	o ***InptServ_Loc_At*** from **Facility** (1-1 o ***InptAdmit_To*** to InptAdmit
Outpt_Serv	Outpatient Service at a Facility	outptServName	o ***OutptServ_Loc_At*** from **Facility** (0-M) o ***OPV_With*** to Outpt Visit (0-M)
Inpt_Admit	Inpatient Admission to an Inpt_Unit on an Inpt_Serv	admitDate dischargeDate deathDate	o ***Admitted*** from **Patient** (0-M) o *InptAdmit_On* from Inpt Unit (0-M) o *InptAdmit_To* from Inpt Service (0-M)
Outpt_Visit	Outpatient Visit to an Outpt_Clin run by an Outpt_Serv	visitDate	o ***Has_OPV*** from **Patient** (0-M) o ***OPV_At*** from Outpt_Clin (0-M) o *OPV_With* from Outpt_Serv (0-M)
Culture	Single culture from a single **Patient**	accessionNo specType collectSite	o ***Provides*** from **Patient** (0-M) o ***Grows_Out*** to Isolate (1-1) o *Processes* to **Laboratory**

		collectDate collectTime	(1-M)
Laboratory	A microbiology laboratory	**labID**	o *Processes* from Culture (1-M) o *Belongs_To* from **Facility** (0-1)
Isolate	A specific Organism isolated from a Culture	isolateNo betaLactamase	o ***Grows_Out*** from Culture (0-M) o *Is_of_Type* from **Organism** (0-M) o ***Tested_Vs*** to Antibiogram (1-1)
Antibiogram	AST results for a single Isolate vs. a single antimicrobial	astResult interpretation reportGroup	o ***Abx_Tested*** from **AST** (0-M) o ***Tested_Vs*** from Isolate (0-M)
AST	An AST for a specific anti-microbial using a specific method	loincCode antimicrobialName testMethod	o ***Abx_Tested*** to Antibiogram (1-1) o ***Antimicrobial_Tested*** to Breakpoint (1-1)
Breakpoint	Susceptibility and resistance break-points for a partic-ular antimicrobial tested against a reference strain	clsiSupplement testMethod sBreak rBreak reportGroup	o ***Antimicrobial_Tested*** from **AST (**0-M) o ***Organism_Tested*** from **Organism** (0-M)
Organism	A reference bacterial strain used for AST	snomedCode genusName speciesName	o ***Organism_Tested*** to Breakpoint (1-1) o *Is_of_Type* to Isolate (1-1)

Table 2. Data dictionary for the CABGM data model.
Strong entities are shown in bold; primary keys are underlined.

Entity	Attribute	Data type	Description	Null-able?
Antibiogram	interpretation	CHAR	Qualitative interpretation of astResult	Yes
Antibiogram	astResult	FLOAT	MIC, zone diameter, or E-test result	Yes

Antibiogram	reportGroup	CHAR	Reporting group for antibiograms	No
AST	antimicrobialName	VARCHAR2 (20)	Name of antimicrobial corresponding to loincCode	No
AST	loincCode	VARCHAR2 (7)	LOINC code corresponding to testMethod	No
AST	testMethod	VARCHAR2 (20)	Method used for susceptibility testing	No
Breakpoint	clsiSupplement	VARCHAR2 (10)	M100 supplement used as source for breakpoints	No
Breakpoint	reportGroup	CHAR	Reporting group for antibiograms	No
Breakpoint	rBreak	FLOAT	Resistance breakpoint value for testMethod	Yes
Breakpoint	sBreak	FLOAT	Susceptibility breakpoint value for testMethod	Yes
Culture	accessionNo	INTEGER	Accession number for the culture – unique to the facility	No
Culture	collectDate	DATE	Culture collection date	No
Culture	collectSite	INTEGER	SNOMED Code for the collection site	No
Culture	collectTime	TIME	Culture collection time	Yes
Culture	specType	INTEGER	SNOMED Code for the type of collection	No
Facility	address	VARCHAR2 (20)	Street address for facility	No
Facility	city	VARCHAR2 (20)	Facility city	No
Facility	**facilityID**	VARCHAR2 (10)	Facility identifier – usually a 3-digit number with an optional letter and number	No

Facility	facilityName	VARCHAR2 (20)	Official VHA facility name	No
Facility	state	VARCHAR2 (20)	Facility state	No
Facility	zipCode	VARCHAR2 (10)	Facility Zip code	No
Inpt_Admit	admitDate	DATE	Date of admission	No
Inpt_Admit	deathDate	DATE	Date of death	Yes
Inpt_Admit	dischargeDate	DATE	Date of discharge	Yes
Inpt_Serv	inptServName	VARCHAR2 (20)	Name of inpatient service admitting patient	No
Inpt_Unit	unitName	VARCHAR2 (20)	Physical location to which patient is admitted	No
Isolate	isolateNo	INTEGER	If multiple isolates from a culture	No
Isolate	betaLactamase	BOOLEAN	Presence or absence of a β-lactamase	Yes
Laboratory	**labID**	VARCHAR2 (10)	Identifies lab. If a VHA lab, set to facilityID for the facility where the lab is located	No
Organism	genusName	VARCHAR2 (20)	Genus name for organism	No
Organism	**snomedCode**	INTEGER	SNOMED code for organism	No
Organism	speciesName	VARCHAR2 (20)	Species name for organism	Yes
Outpt_Clin	clinicName	VARCHAR2 (20)	Name of outpatient clinic	No
Outpt_Clin	stopCode	INTEGER	Three-digit identification code for type of outpatient clinic	No
Outpt_Serv	OutptServName	VARCHAR2 (20)	Name of outpatient service	No

Outpt_Visit	visitDate	DATE	Date of outpatient visit	No
Patient	birthDate	DATE	Patient's birth date	No
Patient	ethnicity	VARCHAR2 (20)	Patient ethnicity (white, Hispanic, or >1)	Yes
Patient	firstName	VARCHAR2 (20)	Patient's first name	No
Patient	lastName	VARCHAR2 (20)	Patient's last name	No
Patient	middleName	VARCHAR2 (20)	Patient's middle name	Yes
Patient	**patientID**	VARCHAR2 (9)	Social Security Number	No
Patient	priorityCategory	INTEGER	Enrollment category	Yes
Patient	race	VARCHAR2 (20)	Patient race	Yes
Patient	serviceEra	VARCHAR2 (20)	Era of military service	Yes
Patient	sex	CHAR	Patient sex	Yes

Table 3. Attributes of the CABGM data model with constrained values

Entity	Attribute	Allowed Values
Antibiogram	astResult	S – Susceptible R – Resistant I – Intermediate N – Nonsusceptible U – Undefined
Antibiogram	reportGroup	A – Primary test, always report B – Primary test, selectively report C – Supplemental test, selectively report U – Test and report only for urine cultures
Breakpoint	reportGroup	A – Primary test, always report B – Primary test, selectively report C – Supplemental test, selectively report

		U – Test and report only for urine cultures
Isolate	betaLactamase	0 – Negative test for β-lactamase 1 – Positive test for β-lactamase
Patient	ethnicity	W – White H – Hispanic 2 – More than one ethnicity
Patient	priorityCategory	Enrollment Category (1-8)
Patient	race	W – White B – Black A – Asian P – Pacific Islander N – Native American 2 – More than one race
Patient	sex	M – Male F – Female I – Intersex T - Transgender

Clinical Entities and Attributes

This portion of the data model consists of concepts related to general health care activities that affect individual patients. Figure 3 shows clinical entities, their attributes, and their relationships.

A **Patient** is uniquely identified by his or her social security number, and has attributes describing the patient's first name, middle name, last name, birth date, sex, race, ethnicity, era of military service (e.g., Vietnam era), and enrollment priority category (an integer between 1 and 8, inclusive, representing the level of benefits for which the patient is eligible.

A **Facility** is identified by a unique character string, and has a name, street address, city, state, and Zip Code. Facilities may have Inpatient Units and/or Outpatient

Clinics. An Inpatient Unit is identified by the combination of the identifier for the Facility to which it belongs and the unit's name. An Outpatient Clinic is identified by a combination of the identifier for the Facility to which it belongs and a three-digit number called a *stop code*; for example, an Infectious Diseases Outpatient Clinic in VHA always has the stop code 320. An Outpatient Clinic also has a name.

A **Laboratory** may be located at a VHA Facility or be a non-VHA laboratory (e.g., a laboratory at an academic medical center or a commercial laboratory). If the former, its identifier is a string equivalent to the Facility identifier; if the latter, it's identifier is the name of the non-VHA laboratory.

Facilities may also have Inpatient Clinical Services and Outpatient Clinical Services. Both of these are identified by a combination of the facility identifier and the service's name, e.g., 'General Surgery'. Inpatient Clinical Services are associated with particular Inpatient Units at a Facility, while Outpatient Clinical Services are associated with particular Outpatient Clinics at a Facility.

Patients may be admitted to an Inpatient Service on an Inpatient Unit at a particular Facility. An Inpatient Admission is uniquely identified by the Facility ID, the Patient ID, and the admission date. The admission also has discharge date and date of death as attributes.

Patients may also have an Outpatient Visit to an Outpatient Clinic. A Visit is uniquely identified by the Patient ID, the stop code for the Clinic, the ID for the Facility to which the Clinic belongs, and the visit date.

Microbiologic Entities, Attributes, and Relationships

This part of the data model consists of concepts related to microbiologic specimen collection and processing. Figure 4 shows microbiologic entities, their attributes, and their relationships.

A Patient may provide Cultures. A Culture is uniquely identified by the combination of its accession number (which is unique within the facility) and the ID for the Facility at which it is collected. A Culture also has the attributes of specimen type, collection site, collection date, and collection time. Cultures are processed by a **Laboratory**, which has a unique string identifier.

An **Organism** is uniquely identified by a SNOMED code. It also has a genus name and a species name. An Organism that grows from a Culture is referred to as an Isolate. Cultures may grow out one or more Isolates; an Isolate is uniquely identified by the combination of its isolate number, the accession number of the Culture from which it grew, the SNOMED code of the type of organism, and the ID for the Facility at which it was collected.

AST Entities, Attributes, and Relationships

This part of the data model consists of concepts related to antimicrobial susceptibility testing; figure 5 shows AST entities, their attributes, and their relationships.

An **AST** is uniquely identified by a LOINC code, and has attributes of antimicrobial name (e.g., ampicillin) and test name (e.g., disk diffusion).

An Organism can be tested via an AST to produce a Breakpoint, which is uniquely identified by the combination of the SNOMED code for the Organism being tested, the LOINC code for the susceptibility method used, and the CLSI Supplement document used as the reference. It has attributes of a susceptibility breakpoint value, a resistance breakpoint value, and a report group.

An Isolate can be tested via an AST to yield an Antibiogram, which is uniquely identified by the combination of the Isolate's number, the accession number for the Culture from which it grew, the Facility ID for the Culture, and the LOINC code for the AST. An Antibiogram has a test result, interpretation, and report group.

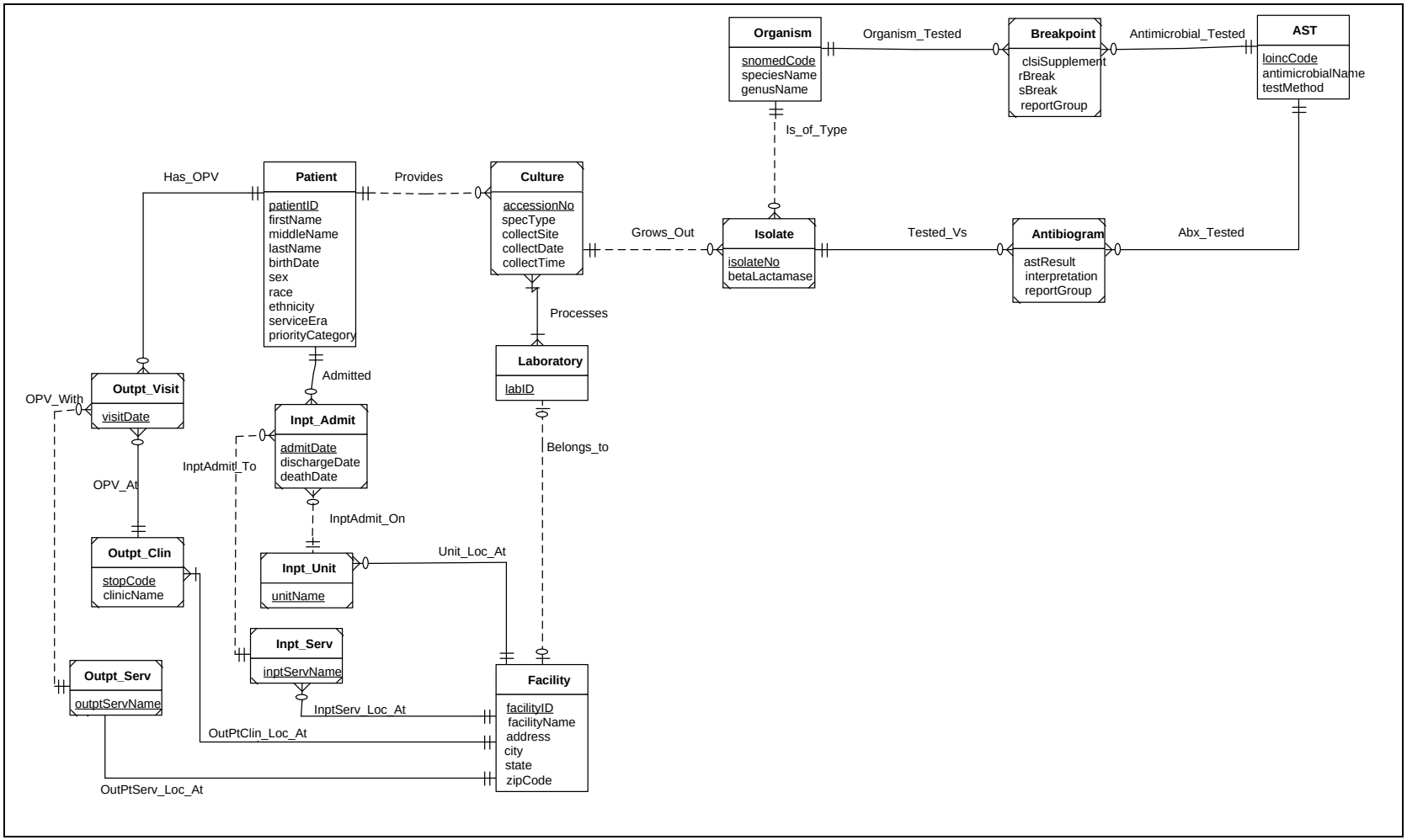
Isolates can be tested for zero to many resistance factors; the results are attributes of the isolate. For simplicity, only β-lactamase testing is included in this model, with the attribute having a value of TRUE or FALSE (i.e., the factor is present or absent).

The full model is shown in Figure 6.

Validation

Although the data model complies with formal completeness and consistency rules, and employs appropriate standards, this by no means guarantees that it accurately models the data needed to accurately produce CABGMs and satisfy the use cases described in above. From a business need validation perspective, the first step would involve review of the model by potential end-users to determine if the model completely captures the data elements (entities, attributes, and relationships) needed for the use cases described above. Once the model is revised based on this input, the next step would be to construct a database model that could be tested and validated.

Figure 6. Full model.



Implementation of a Cumulative Antibiogram

Business Rules

For purposes of generating a database model, the data model in Figure 6 must be combined with business rules that specify how results such as AST interpretations are derived, constrain attribute values to the values in the data dictionary, prevent logical inconsistencies, and incorporate the logic used for antimicrobial susceptibility testing. Business rules for the CABGM include the following:

1. The value of an Inpt_Admit dischargeDate must be equal to or greater than the value of the Inpt_Admit admitDate.
2. If the value for an Inpt_Admit deathDate is not NULL, it must be equal to the value for the Inpt_Admit dischargeDate.
3. Antibiogram interpretations have a restricted set of values, as follows:
    a. If an Antibiogram testResult is less than or equal to the sBreak value for the corresponding Breakpoint, the Antibiogram interpretation has a value of **S**.
    b. If the Antibiogram testResult is greater than the rBreak value for the corresponding Breakpoint, the Antibiogram interpretation has a value of **R**.
    c. **I**f an Antibiogram testResult is less than or equal to the rBreak value for the corresponding Breakpoint, and greater than the sBreak, the Antibiogram interpretation has a value of **I**.

d. If the Antibiogram testResult is greater than the sBreak value for the corresponding Breakpoint, but_the rBreak value is **NULL** the Antibiogram interpretation has a value of **N**.

e. If the Antibiogram testResult is less than the rBreak for the corresponding Breakpoint, but the sBreak value is **NULL**, the Antibiogram Interpretation is **U**.

f. If the rBreak value and the sBreak value are both NULL, the Antibiogram interpretation is **U**.

g. Some Antibiogram interpretations are overridden by the interpretation values for other Antibiograms for the same isolate, e.g., if the interpretation for *S. aureus* tested against oxacillin is R, interpretations for cephalosporins should also be R except for ceftaroline.

h. If Isolate.betaLactamase is **TRUE**, then the interpretation for any AST using penicillin or ampicillin (e.g., for ampicillin, loincCode is equal to ‘18864-9’, ‘29-9’, ‘6979-9’, or ‘28-1’) should always be **R.**

4. The interpretation for Antibiograms testing *Klebsiella pneumoniae* tested against ampicillin (i.e snomedCode = 56415008 and loincCode equal to ‘18864-9’, ‘29-9’, ‘6979-9’, or ‘28-1’) should always be **R**.

5. The interpretation for Antibiograms testing *S. aureus* against vancomycin (i.e., snomedCode equal to 3092008 and loincCode equal to ‘19000-9’, ‘525-6’, ‘7059-9’, or ‘524-9’) should always be **S**.

Display Rules for Antibiograms for a Single Isolate

1. For clinical use in choosing microbiologically defined antimicrobial therapy, selective reporting is used as follows:

    a. Only interpretations with a non-NULL value are displayed.

    b. Interpretations for Antibiograms with a reportGroup value of ‘A’ are always displayed.

    c. Interpretations for Antibiograms with a reportGroup value of ‘B’ are displayed only if all of the interpretations for Group A Antibiograms are equal to **R**.

    d. Interpretations for Antibiograms with a reportGroup value of ‘C’ are displayed only if all of the interpretations for Group A Antibiograms are equal to **R**.

    e. Group U Interpretations are only displayed for Antibiograms for urine cultures (i.e., cultures with a specType equal to 411852016 or a related SNOMED-CT Code).

2. Full displays should be available to microbiology laboratory directors, infectious disease physicians, infectious disease pharmacists, and infection control staff.

Rules for constructing Cumulative Antibiograms

1. CABGMs consist of susceptibility rates for specific antimicrobials for specific organisms of interest (i.e., a genus/species mapping to a unique SNOMED code).

2. The susceptibility rate is calculated by dividing the number of Isolates with an interpretation of ‘S’ (i.e., susceptible) to a particular antimicrobial by the number of Isolates tested against that antimicrobial, regardless of test method used.
3. For Use Case 1 (i.e., for data presentation for clinicians choosing empiric antimicrobial therapy for clinicians):
    a. Only susceptibility rates are displayed.
    b. Only the first Isolate with a particular snomedCode for each patient in a reporting period is included in the analysis for a reporting period.
    c. Antibiogram interpretations for Group B and Group C Antimicrobials are only included in the numerator and denominator if the values for the Antibiogram Interpretations for Group A Antimicrobials for an Isolate all have values of **R;** in other words, only antibiograms that would be displayed as individual antibiogram results to clinicians are used in the analysis.
    d. Rates for Group U interpretations are reported separately.
4. For Use Cases 2-4:
    a. Rates for resistance, intermediate susceptibility, and nonsusceptibility may be displayed.
    b. Only Antibiogram interpretation values for the first Isolate of a given snomedCode for each patient in a reporting period is included in the numerator and denominator.

With these rules in place, a relational database model can be constructed in a straightforward fashion from the data model shown in Figure 6. Each entity becomes a

table, with attributes representing individual columns, and foreign keys used to create relationships between tables. Queries can then be constructed to determine susceptibility rates for particular species of interest. Selective reporting can be implemented using views, with cumulative antibiograms restricted by producing logical tables that do not contain data on Group B or C interpretations except for those isolates resistant to all Group A antimicrobials.

Populating a relational antibiogram database

Construction of a cumulative antibiogram is only possible if the underlying relational database is populated. However, the data stored in electronic medical records tends to be structured using a hierarchical, rather than a relational, database model. Such structures provide a more natural model for the realities of clinical care because they allow for efficient storage and processing of repeated information. For example, a individual patient may have multiple outpatient visits, or many laboratory test results, or many hospital admissions. Since patient care generally involves focusing on one individual at a time, a hierarchical data structure provides an excellent basis for modeling patient-centered medical records and prescribing.

However, this model is not always particularly well-suited or efficient for performing queries because of the redundancy inherent in this structure and the time needed to traverse the hierarchy. Thus, from a practical viewpoint, for the antibiogram model described above to be useful, it is necessary to migrate hierarchically structured clinical data into a relational database.

VHA's electronic medical record system, the Veterans Information System Technology Architecture (VistA), is a good platform for addressing this issue. The VistA clinical database is composed of several hundred hierarchically structure files managed by Fileman, a database management system (DBMS) written in the Massachusetts General Hospital Utility Multi-Programming System (MUMPS) (Andrews and Beauchamp 1989). Fileman files are analogous to relational database tables, and are divided into fields that are analogous to columns.

However, unlike a relational database, the fields of Fileman files are not necessarily individual values. Instead, they are frequently pointers to a *subfile* (i.e., another table). In essence, Fileman files are denormalized tables in which some fields point to a subtable with multiple values.

So, for example, the Fileman County File (File 5.1) has the structure below. Field 1 contains a pointer to the State File (File #5) rather than an actual value, in essence yielding a file containing a subfile, which may contain or point to subfiles of its own.

Field #	Name	Loc	Type	Details
.01	Name	0;1	Free Text	
1	State	0;2	Point to State File (#5)	State (#5)
2	SEER County Code	0;3	Free Text	
3	Abbreviation	0;4	Free Text	
4	VA County Code	0;5	Free Text	
5	Catchment Code	0;6	Free Text	

Despite the differences between the VistA data model and the structure of relational databases, migrating data from the former to the latter is not particularly difficult. The VistA data model is well-documented, and mapping the VistA data model to the current data model is straightforward. The Vista Cross Reference Documentation Web site (Anon.) provides an excellent resource for such mapping.

Table 4 shows a cross-walk between VistA data files and the current model. With this cross-walk in place, a database management system such as Intersystems Caché can be used to construct a relational database from the hierarchically structured data in VistA.

Table 4. Cross-walk between VistA and the CABGM data model

Entity	Attribute	Data type	VistA File/ Field Number	Data type
Antibiogram	interpretation	CHAR	63.05/5-200	Free Text
Antibiogram	astResult	FLOAT	63.05/5-200	Free Text
Antibiogram	reportGroup	CHAR	62.06/7	CHAR
AST	antimicrobialName	VARCHAR2 (20)	63.05/5-200	Free Text
AST	loincCode	VARCHAR2 (7)	95.3/.01	INTEGER
AST	testMethod	VARCHAR2 (20)	63.061/.01	Free Text
Breakpoint	clsiSupplement	VARCHAR2 (10)	N/A	N/A
Breakpoint	reportGroup	CHAR	62.06/7	CHAR
Breakpoint	rBreak	FLOAT	N/A	N/A
Breakpoint	sBreak	FLOAT	N/A	N/A
Culture	accessionNo	INTEGER	63.05/.06	Free Text

Culture	collectDate	DATE	63.05/.01	DATE
Culture	collectSite	INTEGER	63/2	Free Text
Culture	collectTime	TIME	63.05/.01	DATE
Culture	specType	INTEGER	61/2	Free Text
Facility	address	VARCHAR2 (20)	4/1.01	Free Text
Facility	city	VARCHAR2 (20)	4/1.03	Free Text
Facility	**facilityID**	VARCHAR2 (10)	4/99	Free Text
Facility	facilityName	VARCHAR2 (20)	4/100	Free Text
Facility	state	VARCHAR2 (2)	4/.02	Free Text
Facility	zipCode	VARCHAR2 (10)	4/1.04	Free Txt
Inpt_Admit	admitDate	DATE	45/2	DATE
Inpt_Admit	deathDate	DATE	63/12	DATE
Inpt_Admit	dischargeDate	DATE	45/70	DATE
Inpt_Serv	inptServName	VARCHAR2 (20)	42/.03	Free Text
Inpt_Unit	unitName	VARCHAR2 (20)	42/.02	Free Text
Isolate	genusName	VARCHAR2 (20)	63.05/12	Free Text
Isolate	isolateNo	INTEGER	63.3/.001	INTEGER
Isolate	snomedCODE	INTEGER	61.2/2	Free Text
Isolate	speciesName	VARCHAR2(20)	63.05/12	Free Text
Isolate	betaLactamase	BOOLEAN	63.061/.01	Free Text
Laboratory	labID	INTEGER	63.5/.112	Pointer to Institution File (#4)

Organism	genusName	VARCHAR2 (20)	63.05/12	Free Text
Organism	snomedCode	INTEGER	61.2/2	Free Text
Organism	speciesName	VARCHAR2 (20)	63.05/12	Free Text
Outpt_Clin	clinicName	VARCHAR2 (20)	40.7/.01	Free Text
Outpt_Clin	stopCode	INTEGER	40.7/1	INTEGER
Outpt_Serv	outptServName	VARCHAR2 (20)	42.4/1	Free Text
Outpt_Visit	visitDate	DATE	9000010/.01	DATE
Patient	birthDate	DATE	2/.03	DATE
Patient	ethnicity	VARCHAR2 (20)	2/.06	Free Text
Patient	firstName	VARCHAR2 (20)	2/.01	Free Text
Patient	lastName	VARCHAR2 (20)	2/.01	Free Text
Patient	middleName	VARCHAR2 (20)	2/.01	Free Text
Patient	patientID	VARCHAR2 (9)	2/.09	Free Text
Patient	priorityCategory	INTEGER	8.1/3	Integer
Patient	race	VARCHAR2 (20)	2/.06	Free Text
Patient	serviceEra	VARCHAR2 (20)	21/20	Free Text
Patient	sex	CHAR	2/.02	CHAR

Although there is some variability in the exact storage locations for microbiologic data between different VHA facilities, the number of alternatives is relatively small, making it extremely feasible to pull data from a clinical database for a facility and migrate it to a relational database. Such migration could be done a regular basis; for example, the VHA's national clinical case registries for patients with HIV or chronic hepatitis

C sweep the data from individual facilities on a nightly basis, allowing the population of a VHA-wide relational database dealing with these diseases, and providing the basis for sophisticated queries and analyses (Backus *et al*. 2009; Backus *et al*. 2010).

However, the cross-walk reveals some disconcerting discrepancies in data types between the VistA data model and the CLSI-based data model described here, indicated by shading. These primarily involve string variables in VistA being mapped to integer or floating-point values in the data model. To make matters worse, much of the VistA database – particularly for AST results – has been populated by manual entry, as described in the introduction, with constraints on values being the exception rather than rule. Thus, migration of VistA data to a relational database for purposes of generating CABGMs involves not just type conversion, but also semantic and syntactic mapping of free text to standard vocabularies. These problems are discussed below.

Discussion

The data model described here provides a basis for standardizing construction and analysis of CABGMs, not only facilitating CABGM generation at the individual facility level, but also allowing aggregation of data and comparison of AR prevalence rates across multiple institutions. The migration of data from a hierarchical clinical data repository to a relational database is relatively straightforward.

However, as described earlier, microbiology laboratory data is frequently manually entered into VistA, creating a source for variability in such data. This problem is amplified substantially by the use of strings as a standard data type in MUMPS. While this provides enormous flexibility, it also introduces complex type conversion problems, as illustrated by the type discrepancies shown in Table 4.

To make matters worse, VHA medical facilities frequently use different terminologies for microbiology data and different configurations of data files. For example, some facilities will designate ampicillin, a widely used antimicrobial, by the string "AMPICILLIN," while others will use the string "AMP," and others will use a numeric value of 2.006. Some facilities will use File 62.06 (Antimicrobial Susceptibility) for storing AST results, while others will use File 63.05 (Laboratory Data). This is illustrated in Table 5, which is a flat file showing the genus and species name and which antimicrobials were used for testing for a single isolate from a single VHA facility. It shows substantial variation in data formats for antimicrobials, even though the data come from a single center. The antimicrobial column contains a mixture of generic and trade names, as well as codes such as 2.006, which represents AST with ampicillin.

Table 5. Representative microbiologic VistA data	
Organism	**Antimicrobial**
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	No DD(63.3,2.0003
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	No DD(63.3,2.0004
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	No DD(63.3,2.0006
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	No DD(63.3,2.003
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	CIPROFLOXACIN
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	SYNERCID
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	LEVOFLOXACIN
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	LINEZOLID
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	MOXIFLOXACIN
METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	TIGECYCLINE

This lack of standardization is not unique to the VHA, and creates immediate barriers to interoperability for reporting or analyzing AST data. While a particular medical center may be able to create CABGMs for its own use, variations in data types, formats, and storage locations represent obstacles to aggregating or comparing AST data from multiple centers.

Manual mapping is one possible approach to this problem. This method catalogs all existing values for a particular concept and maps them all to a standard vocabulary term. Thus, “AMPICILLIN”, “AMP”, and 2.006 would all be mapped to the LOINC code 18664-9. This method has been highly successful in creating the HIV and hepatitis C clinical case registries within VHA described above. It is feasible when there are a relatively small number of variations on the terminology for concept, and would be easily applicable to mapping antimicrobial names.

More complex issues may arise in connection with differences in bacterial nomenclature, where there may be a large number of variant terms for many different

classes of organisms. This situation arises from periodic revisions in bacterial taxonomy. For example, the Gram-negative bacillus *Stenotrophomonas maltophilia* was originally named *Pseudomonas maltophilia* when originally described in 1958, and multiple organisms thought to be separate species with distinct names (e.g., *Bacterium booker, P. melanogena,* and *P. alcaligenes*) were classified as being identical to this species. Analysis of ribosomal RNA genes led to reclassification and renaming of this organism as *Xanthomonas maltophilia* in 1983, with subsequent reclassification and renaming as *S. maltophilia* in 1993 (Denton and Kerr 1998).

This is just one example of how a single concept in bacterial taxonomy can be represented by multiple synonyms. Additional variants can arise because of the use of historical synonyms, variant (or incorrect) spellings, and use of abbreviations.

Although these differences can be also be resolved via manual mapping, natural language processing (NLP) represents an alternative approach, which may also be useful for mapping free text entries for bacterial tests to a common term. NLP has been used within VHA to retrieve data on organisms and susceptibilities, such as MRSA *(Jones et al.* 2012) with high levels of precision and recall. The availability of open-source clinical systems such as cTAKES makes NLP a promising complement to manual mapping (Garla *et al*. 2011) for standardization of microbiologic data.

Summary and Conclusions

The data model described here provides a basis for both a relational database model and construction of applications to construct CABGMs for a variety of purposes, including internal validation of microbiology data. Constraining the model to the published CLSI standard creates a mechanism for assuring standardization and accuracy of CABGMs. In addition, as discussed above, migration of the data from a hierarchical to relational database is relatively straightforward.

Data standardization of microbiologic information is a higher and more difficult hurdle to clear. However, the techniques such as manual mapping and natural language processing have been used successfully to create registries within VHA for concepts and terminologies that are at least as complicated as those related to antimicrobial susceptibility testing, and should, in theory, be applicable to this area as well.

Although the VHA VistA system has been used as a representative platform for antibiogram modeling and generation, the issues in VHA are not unique, and the results described here are likely applicable to resistance surveillance systems elsewhere in the U.S. This model may thus be useful for initiatives such as the National Healthcare Safety Network operated by the U.S. Centers for Disease Control and Prevention, an Internet-based surveillance system, as well as national monitoring of antimicrobial resistance by the U.S. FDA and other public health agencies.

Literature cited

Adam, Heather J, Daryl J Hoban, Alfred S Gin, and George G Zhanel. 2009. "Association Between Fluoroquinolone Usage and a Dramatic Rise in Ciprofloxacin-resistant Streptococcus Pneumoniae in Canada, 1997-2006." *International Journal of Antimicrobial Agents* 34 (1) (July): 82–85.

Alvarez-Ortega, Carolina, Irith Wiegand, Jorge Olivares, Robert E W Hancock, and José Luis Martínez. 2011. "The Intrinsic Resistome of Pseudomonas Aeruginosa to β-lactams." *Virulence* 2 (2) (April): 144–146.

Andrews, Robert D., and Charles Beauchamp. 1989. "A Clinical Database Management System for Improved Integration of the Veterans Affairs Hospital Information System." *Journal of Medical Systems* 13 (6): 309–320.

Backus, Lisa I, Derek B Boothroyd, Barbara R Phillips, Pamela S Belperio, James P Halloran, Ronald O Valdiserri, and Larry A Mole. 2010. "National Quality Forum Performance Measures for HIV/AIDS Care: The Department of Veterans Affairs' Experience." *Archives of Internal Medicine* 170 (14) (July 26): 1239–1246.

Backus, Lisa I, Sergey Gavrilov, Timothy P Loomis, James P Halloran, Barbara R Phillips, Pamela S Belperio, and Larry A Mole. 2009. "Clinical Case Registries: Simultaneous Local and National Disease Registries for Population Quality Management." *Journal of the American Medical Informatics Association: JAMIA* 16 (6) (December): 775–783.

Bajaj, Jasmohan S, Ashwin N Ananthakrishnan, Muhammad Hafeezullah, Yelena Zadvornova, Alexis Dye, Emily L McGinley, Kia Saeian, Douglas Heuman, Arun J Sanyal, and Raymond G Hoffmann. 2010. “Clostridium Difficile Is Associated with Poor Outcomes in Patients with Cirrhosis: A National and Tertiary Center Perspective.” *The American Journal of Gastroenterology* 105 (1) (January): 106–113.

Brown, Steven H, Michael J Lincoln, Peter J Groen, and Robert M Kolodner. 2003. “VistA--U.S. Department of Veterans Affairs National-scale HIS.” *International Journal of Medical Informatics* 69 (2-3) (March): 135–156.

Buising, Kirsty L, Karin A Thursky, James F Black, Lachlan MacGregor, Alan C Street, Marcus P Kennedy, and Graham V Brown. 2008. “Improving Antibiotic Prescribing for Adults with Community Acquired Pneumonia: Does a Computerised Decision Support System Achieve More Than Academic Detailing alone?--A Time Series Analysis.” *BMC Medical Informatics and Decision Making* 8: 35.

CDC. 2004. “Antibiogram-method.pdf.” Available at http://www.cdc.gov/abcs/reports-findings/downloads/antibiogram-method.pdf, accessed on 9 June 2012.

CLSI. 2009. “Analysis and Presentation of Cumulative Antimicrobial Susceptibility Data; Approved Guideline - Third Edition”. Clinical and Laboratory Standards Institute.

———. 2012. “Performance Standards for Antimicrobial Susceptibility Testing; Twenty-second Informational Supplement. CLSI Document M100-A22”. Clinical and Laboratory Standards Institute.

Critchley, IA, and JA Karlowsky. 2004. “Optimal Use of Antibiotic Resistance Surveillance Systems.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 10 (6) (June): 502–511.

Dellit, Timothy H, Robert C Owens, John E McGowan Jr, Dale N Gerding, Robert A Weinstein, John P Burke, W Charles Huskins, et al. 2007. “Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America Guidelines for Developing an Institutional Program to Enhance Antimicrobial Stewardship.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 44 (2) (January 15): 159–177.

Denton, Miles, and Kevin G. Kerr. 1998. “Microbiological and Clinical Aspects of Infection Associated with Stenotrophomonas Maltophilia.” *Clinical Microbiology Reviews* 11 (1) (January): 57–80.

DiNubile, MJ. 1990. “Antibiotics: The Antipyretics of Choice?” *The American Journal of Medicine* 89 (6) (December): 787–788.

Fauci, Anthony S, and David M Morens. 2012. “The Perpetual Challenge of Infectious Diseases.” *The New England Journal of Medicine* 366 (5) (February 2): 454–461.

Fish, Douglas N, and Martin J Ohlinger. 2006. “Antimicrobial Resistance: Factors and Outcomes.” *Critical Care Clinics* 22 (2) (April): 291–311, vii.

Fleming, Thomas R., and John H. Powers. 2008. “Issues in Noninferiority Trials: The Evidence in Community-Acquired Pneumonia.” *Clinical Infectious Diseases : an Official*

Publication of the Infectious Diseases Society of America 47 (Suppl 3) (December 1): S108–S120.

Garla, Vijay, Vincent Lo Re 3rd, Zachariah Dorey-Stein, Farah Kidwai, Matthew Scotch, Julie Womack, Amy Justice, and Cynthia Brandt. 2011. "The Yale cTAKES Extensions for Document Classification: Architecture and Application." *Journal of the American Medical Informatics Association: JAMIA* 18 (5) (October): 614–620.

Gerding, Dale N, Carlene A Muto, and Robert C Owens Jr. 2008. "Treatment of Clostridium Difficile Infection." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 46 Suppl 1 (January 15): S32–42.

Goff, Debra A. 2011. "Antimicrobial Stewardship: Bridging the Gap Between Quality Care and Cost." *Current Opinion in Infectious Diseases* 24 Suppl 1 (February): S11–20.

Gordon, Rachel J, and Franklin D Lowy. 2008. "Pathogenesis of Methicillin-resistant Staphylococcus Aureus Infection." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 46 Suppl 5 (June 1): S350–359.

Graves, Nicholas, Adrian G Barnett, Kate Halton, Christopher Crnich, Ben Cooper, Jan Beyersmann, Martin Wolkewitz, Matthew Samore, and Stephan Harbarth. 2011. "The Importance of Good Data, Analysis, and Interpretation for Showing the Economics of Reducing Healthcare-associated Infection." *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America* 32 (9) (September): 927–928; author reply 928–930.

Gumbo, T. 2011. “Chapter 48. General Principles of Antimicrobial Therapy.” In *Goodman & Gilman’s The Pharmacological Basis of Therapeutics.*, ed. Brunton LL, Chabner BA, Knollmann BC,. Vol. 12th ed. New York: McGraw-Hill.

Harris, AD, MH Samore, R Nafziger, K DiRosario, MC Roghmann, and Y Carmeli. 2000. “A Survey on Handwashing Practices and Opinions of Healthcare Workers.” *The Journal of Hospital Infection* 45 (4) (August): 318–321.

Jacoby, George A, and Luisa Silvia Munoz-Price. 2005. “The New Beta-lactamases.” *The New England Journal of Medicine* 352 (4) (January 27): 380–391.

Jain, Rajiv, Stephen M Kralovic, Martin E Evans, Meredith Ambrose, Loretta A Simbartl, D Scott Obrosky, Marta L Render, et al. 2011. “Veterans Affairs Initiative to Prevent Methicillin-resistant Staphylococcus Aureus Infections.” *The New England Journal of Medicine* 364 (15) (April 14): 1419–1430.

Jones, Makoto M, Scott L Duvall, Joshua Spuhl, Matthew H Samore, Christopher Nielson, and Michael Rubin. 2012. “Identification of Methicillin-resistant Staphylococcus Aureus Within the Nation’s Veterans Affairs Medical Centers Using Natural Language Processing.” *BMC Medical Informatics and Decision Making* 12 (1) (April 25): 34. doi:10.1186/1472-6947-12-34.

Kennedy, Leigh A, Janet A Gill, Maureen E Schultz, Monica Irmler, and Fred M Gordin. 2010. “Inside-out: The Changing Epidemiology of Methicillin-resistant Staphylococcus Aureus.” *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America* 31 (9) (September): 983–985.

Lambke, Michael. 2012. “Local Antibiograms Can Reduce Inappropriate Antibiotic Prescribing.” *American Family Physician* 85 (10) (May 15): 948–950.

Lautenbach, Ebbing, and Irving Nachamkin. 2006. “Analysis and Presentation of Cumulative Antimicrobial Susceptibility Data (antibiograms): Substantial Variability Across Medical Centers in the United States.” *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America* 27 (4) (April): 409–412.

Leibovici, Leonard, Mical Paul, Anders D Nielsen, Evelina Tacconelli, and Steen Andreassen. 2007. “The TREAT Project: Decision Support and Prediction Using Causal Probabilistic Networks.” *International Journal of Antimicrobial Agents* 30 Suppl 1 (November): S93–102.

Lowy, Franklin D. 2003. “Antimicrobial Resistance: The Example of Staphylococcus Aureus.” *The Journal of Clinical Investigation* 111 (9) (May): 1265–1273.

Mandell, Lionel A, Richard G Wunderink, Antonio Anzueto, John G Bartlett, G Douglas Campbell, Nathan C Dean, Scott F Dowell, et al. 2007. “Infectious Diseases Society of America/American Thoracic Society Consensus Guidelines on the Management of Community-acquired Pneumonia in Adults.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 44 Suppl 2 (March 1): S27–72.

McCaig, LF, and JM Hughes. 1995. “Trends in Antimicrobial Drug Prescribing Among Office-based Physicians in the United States.” *JAMA: The Journal of the American Medical Association* 273 (3) (January 18): 214–219.

Mullett, Charles J, John G Thomas, Connie L Smith, Arif R Sarwari, and Rashida A Khakoo. 2004. “Computerized Antimicrobial Decision Support: An Offline Evaluation of a Database-driven Empiric Antimicrobial Guidance Program in Hospitalized Patients with a Bloodstream Infection.” *International Journal of Medical Informatics* 73 (5) (June 15): 455–460.

OSEHRA. 2012. “OSEHRA VistA Code Documentation.” Available at http://code.osehra.org/dox_beta/, accessed on 9 June 2012

Pakyz, Amy L. 2007. “The Utility of Hospital Antibiograms as Tools for Guiding Empiric Therapy and Tracking Resistance. Insights from the Society of Infectious Diseases Pharmacists.” *Pharmacotherapy* 27 (9) (September): 1306–1312.

Partridge, Sally R, Guy Tsafnat, Enrico Coiera, and Jonathan R Iredell. 2009. “Gene Cassettes and Cassette Arrays in Mobile Resistance Integrons.” *FEMS Microbiology Reviews* 33 (4) (July): 757–784.

Rice, Louis B. 2012. “Mechanisms of Resistance and Clinical Relevance of Resistance to B-lactams, Glycopeptides, and Fluoroquinolones.” *Mayo Clinic Proceedings. Mayo Clinic* 87 (2) (February): 198–208.

Rice, Louis B, Rebecca Hutton-Thomas, Viera Lakticova, Marion S Helfand, and Curtis J Donskey. 2004. “Beta-lactam Antibiotics and Gastrointestinal Colonization with Vancomycin-resistant Enterococci.” *The Journal of Infectious Diseases* 189 (6) (March 15): 1113–1118.

Ross, David B. 2007. “The FDA and the Case of Ketek.” *The New England Journal of Medicine* 356 (16) (April 19): 1601–1604.

Rubin, Michael A, Kim Bateman, Sharon Donnelly, Gregory J Stoddard, Kurt Stevenson, Reed M Gardner, and Matthew H Samore. 2006. “Use of a Personal Digital Assistant for Managing Antibiotic Prescribing for Outpatient Respiratory Tract Infections in Rural Communities.” *Journal of the American Medical Informatics Association: JAMIA* 13 (6) (December): 627–634.

Rubin, Michael A., and Matthew H. Samore. 2002. “Antimicrobial Use and Resistance.” *Current Infectious Disease Reports* 4 (6) (December): 491–497.

Samore, MH, MK Magill, SC Alder, E Severina, L Morrison-De Boer, JL Lyon, K Carroll, et al. 2001. “High Rates of Multiple Antibiotic Resistance in Streptococcus Pneumoniae from Healthy Children Living in Isolated Rural Communities: Association with Cephalosporin Use and Intrafamilial Transmission.” *Pediatrics* 108 (4) (October): 856–865.

Schober, Daniel, Martin Boeker, Jessica Bullenkamp, Csaba Huszka, Kristof Depraetere, Douglas Teodoro, Nadia Nadah, Remy Choquet, Christel Daniel, and Stefan Schulz. 2010. “The DebugIT Core Ontology: Semantic Integration of Antibiotics Resistance Patterns.” *Studies in Health Technology and Informatics* 160 (Pt 2): 1060–1064.

Shannon, KP, and GL French. 2002. “Validation of the NCCLS Proposal to Use Results Only from the First Isolate of a Species Per Patient in the Calculation of Susceptibility Frequencies.” *The Journal of Antimicrobial Chemotherapy* 50 (6) (December): 965–969.

Solomon, David A, Andrew C Leon, Timothy I Mueller, William Coryell, Jedediah J Teres, Michael A Posternak, Lewis L Judd, Jean Endicott, and Martin B Keller. 2005. "Tachyphylaxis in Unipolar Major Depressive Disorder." *The Journal of Clinical Psychiatry* 66 (3) (March): 283–290.

Stordal, Britta, Nick Pavlakis, and Ross Davey. 2007. "A Systematic Review of Platinum and Taxane Resistance from Bench to Clinic: An Inverse Relationship." *Cancer Treatment Reviews* 33 (8) (December): 688–703.

Tamma, Pranita D, and Sara E Cosgrove. 2011. "Antimicrobial Stewardship." *Infectious Disease Clinics of North America* 25 (1) (March): 245–260.

Tenover, Fred C. 2006. "Mechanisms of Antimicrobial Resistance in Bacteria." *The American Journal of Medicine* 119 (6 Suppl 1) (June): S3–10; discussion S62–70.

Tracy, LaRee A, Jon P Furuno, Anthony D Harris, Mary Singer, Patricia Langenberg, and Mary-Claire Roghmann. 2011. "Staphylococcus Aureus Infections in US Veterans, Maryland, USA, 1999-2008." *Emerging Infectious Diseases* 17 (3) (March): 441–448.

Vollaard, EJ, and HA Clasener. 1994. "Colonization Resistance." *Antimicrobial Agents and Chemotherapy* 38 (3) (March): 409–414.

Wraith, SM, JS Aikins, BG Buchanan, WJ Clancey, R Davis, LM Fagan, JF Hannigan, *et al*. 1976. "Computerized Consultation System for Selection of Antimicrobial Therapy." *American Journal of Hospital Pharmacy* 33 (12) (December): 1304–1308.

Yu, VL, LM Fagan, SM Wraith, WJ Clancey, AC Scott, J Hannigan, RL Blum, BG Buchanan, and SN Cohen. 1979. “Antimicrobial Selection by a Computer. A Blinded Evaluation by Infectious Diseases Experts.” *JAMA: The Journal of the American Medical Association* 242 (12) (September 21): 1279–1282.

Zapantis, Antonia, Melinda K Lacy, Rebecca T Horvat, Dennis Grauer, Brian J Barnes, Brian O’Neal, and Rick Couldry. 2005. “Nationwide Antibiogram Analysis Using NCCLS M39-A Guidelines.” *Journal of Clinical Microbiology* 43 (6) (June): 2629–2634.