Health Outcomes in Congenital Cytomegalovirus, A Systematized and Unbiased Approach In the Electronic Medical Record Era (Poster #101)

Ryan H. Rochat MD, PhD, GEMS[1], Gail J. Demmler-Harrison MD[1]

Affiliations: 1. Baylor College of Medicine , Texas Children's Hospital, Department of Pediatrics, Houston, Texas, USA

Mailing Address: 1120 Bates Ave, Houston, Texas, 77030

Email Address: rochat@bcm.edu

Background:

There is limited data on the indirect and non-medical costs associated with congenital cytomegalovirus (cCMV). Attempts to predict the economic impact of disease often rely on secondary analyses of large private databases, and may not capture the full spectrum of a disease. Fortunately, the granularity of billing codes in the Electronic Medical Record (EMR) make it possible to track health outcomes. Unfortunately, with over 70,000 unique codes in the latest version of ICD-10, selecting the appropriate codes requires specific content knowledge and can lead to bias in categorization. The Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT)® provides physicians a tool to find specific ICD-10 on the basis of semantic terms. The semantic terms outlined by the SNOMED-CT® can be used to build disease state specific clusters of ICD-10 codes to study economic impact of this potentially devastating congenital infection.

Theory:

Clinical outcomes for a given condition (e.g. cCMV) can be measured in aggregate by retrieving ICD-10 diagnoses from the EMR. These, coupled with the metadata present in the EMR, make it possible to time specific clinical conditions (e.g. hearing loss). Since ICD-10 codes are hierarchically clustered on broad terms (**Figure 1**), these codes must be further interpreted in a specific clinical context. Numerous tools have been developed to help clinicians find the most appropriate ICD-10 for a specific clinical scenario. One such tool is the SNOMED-CT®, which has been developed to facilitate the electronic exchange of clinical health information by the International Health Terminology Standards Organization. This tool consists of hundreds of thousands semantic terms which are hierarchically related to specific ICD-10 codes (**Figure 2**).

Patient Selection:

There are many auditing tools/widgets available to query large EMRs for identifying subsets of patients who meet specific clinical criteria. Using SlicerDicer™, an auditing tool included with our release of Epic®, we identified all patients who had been seen in our cCMV clinic by Dr. Gail Demmler-Harrison for extraction of their clinical information (**Figure 3**).

Methods:

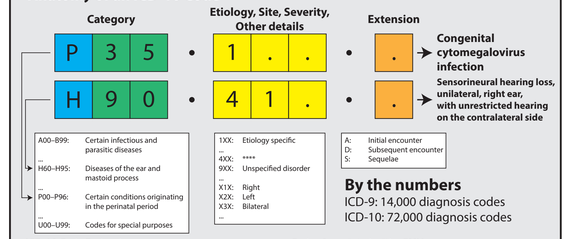
Using SAS V9.4 (Cary, N.C.) we developed a series of data parsing/processing scripts (Poster #132), we extracted the diagnosis codes for 190 patients seen in our Congenital Cytomegalovirus Clinic at Texas Children's Hospital in Houston, Texas (**Figure 3**). This data was consolidated into a relational database of clinical information which was used for statistical analyses of the clinical outcomes for these patients (**Figures 4-6**). A second program, which generalizes the SNOMED-CT® criteria, was used to categorize ICD-10 codes by semantic terms known to be associated with cCMV (e.g. "hearing problem") (**Figure 7**), and this data was again subject to further statistical analysis (**Figure 8,9**).

Results (Figures 3-9):

Figure 3: Schematic breakdown of patient's seen in our CMV clinic.
Figure 4: Distribution of cumulative number of ICD-10 diagnoses for patients with perinatal vs congenital CMV.
Figure 5: Cumulative distribution of ICD-10 diagnoses for a single patient with confirmed cCMV.
Figure 6: Graphical representation of cumulative distribution of ICD-10 codes for our cohort of cCMV patient.
Figure 7: Groups of ICD-10 codes made on the basis of semantic terms related to cCMV.
Figure 8: Whisker plot of the age at which a diagnosis within the Semantic Term groupings was made for patients with cCMV.
Figure 9: Overall distribution of ICD-10 codes for patients with cCMV grouped by conventional ICD-10 category "blocks" compared to systems identified by specific semantic terms.

Figure 1

Anatomy of an ICD-10 Code†



By the numbers
ICD-9: 14,000 diagnosis codes
ICD-10: 72,000 diagnosis codes

ICD-10 Code Redundancy

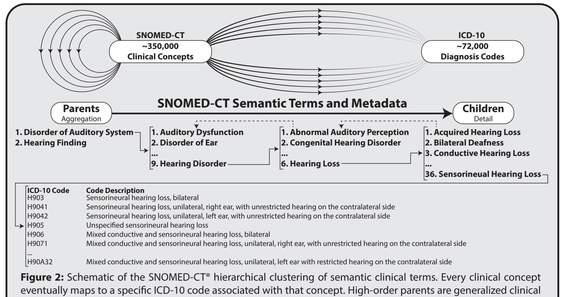
H90	Conductive and sensorineural hearing loss
H900	Conductive hearing loss, bilateral
H901	Conductive hearing loss, unilateral with unrestricted hearing on the contralateral side
H9011	Conductive hearing loss, unilateral, right ear, with unrestricted hearing on the contralateral side
H9012	Conductive hearing loss, unilateral, left ear, with unrestricted hearing on the contralateral side

Codes Associated with cCMV

H903	Sensorineural hearing loss, bilateral
G40009	Localization-related (focal) (partial) idiopathic epilepsy and epileptic syndromes with seizures of localized onset, not intractable, without status epilepticus
H540X33	Blindness right eye category 3, blindness left eye category 3
Z9621	Cochlear implant status
M958	Other specified acquired deformities of musculoskeletal system
G4730	Sleep apnea, unspecified

† https://www.cms.gov/Medicare/Coding/ICD10/

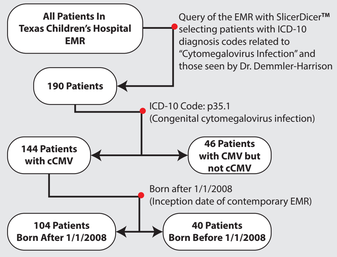
Figure 2



ICD-10 Code	Code Description
H903	Sensorineural hearing loss, bilateral
H9041	Sensorineural hearing loss, unilateral, right ear, with unrestricted hearing on the contralateral side
H9042	Sensorineural hearing loss, unilateral, left ear, with unrestricted hearing on the contralateral side
H905	Unspecified sensorineural hearing loss
H906	Mixed conductive and sensorineural hearing loss, bilateral
H9071	Mixed conductive and sensorineural hearing loss, unilateral, right ear, with unrestricted hearing on the contralateral side
...	
H90A32	Mixed conductive and sensorineural hearing loss, unilateral, left ear with restricted hearing on the contralateral side

Figure 2: Schematic of the SNOMED-CT® hierarchical clustering of semantic clinical terms. Every clinical concept eventually maps to a specific ICD-10 code associated with that concept. High-order parents are generalized clinical statements (e.g. "Hearing Finding"). The "Children" of these clinical terms are further refined to become more specific with each branch of the tree: "Hearing Finding" -> "Hearing Disorder" -> "Hearing Loss" -> "Sensorineural Hearing Loss" -> "Sensorineural hearing loss, unilateral, right ear, with unrestricted hearing on the contralateral side". More specific clinical concepts can share multiple "Parents" in this hierarchical structure. By starting with general semantic terms, it is possible to recursively search the SNOMED-CT® hierarchy and find groups of ICD-10 related to these semantic terms. This approach allows for an unbiased selection of ICD-10 codes from which to classify specific clinical syndromes (e.g "Developmental Disability").

Figure 3



Demographics

Race	n (%)
White	63 (60%)
African American	25 (25%)
Asian	5 (5%)
Not Specified	11 (10%)

Ethnicity	n (%)
Hispanic	22 (21%)
Non-hispanic	75 (72%)
Not Specified	7 (7%)

Sex	n (%)
Female	49 (47%)
Male	55 (53%)

This recursive search includes all ICD-10 codes related to CMV infection as identified from SNOMED-CT®

1. Congenital cytomegalovirus infection
2. Cytomegaloviral pneumonitis
3. Cytomegaloviral pancreatitis
4. Cytomegalovirus hepatitis
5. Other cytomegaloviral disease
6. Cytomegaloviral disease, unspecified
7. Cytomegaloviral mononucleosis without complications
8. Cytomegaloviral mononucleosis with polyneuropathy
9. Cytomegaloviral mononucleosis with meningitis
10. Cytomegaloviral mononucleosis with other complication

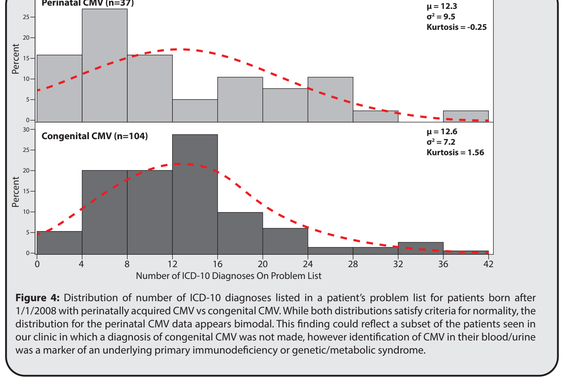


Figure 4: Distribution of number of ICD-10 diagnoses listed in a patient's problem list for patients born after 1/1/2008 with perinatally acquired CMV vs congenital CMV. While both distributions satisfy criteria for normality, the distribution for the perinatal CMV data appears bimodal. This finding could reflect a subset of the patients seen in our clinic in which a diagnosis of congenital CMV was not made, however identification of CMV in their blood/urine was a marker of an underlying primary immunodeficiency or genetic/metabolic syndrome.

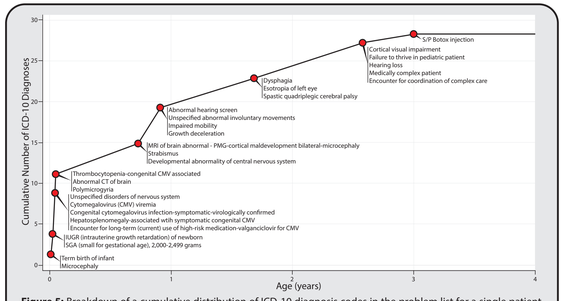


Figure 5: Breakdown of a cumulative distribution of ICD-10 diagnosis codes in the problem list for a single patient with cCMV (4 years old). The date in which the diagnosis is entered into the EMR is used to define the age at diagnosis, entry of a diagnosis into the EMR allows for "Date Noted" to be recorded and retrieved. For this patient, cCMV was noted within the first month of life, in addition to abnormalities on CNS imaging, hearing abnormalities were noted by age 1, cerebral palsy by age 2, and visual/hearing loss before age 3. This data was extracted from the most recent clinical summary for the patient, any inactive or resolved problems on the problem list are not captured by this cross-sectional extraction process.

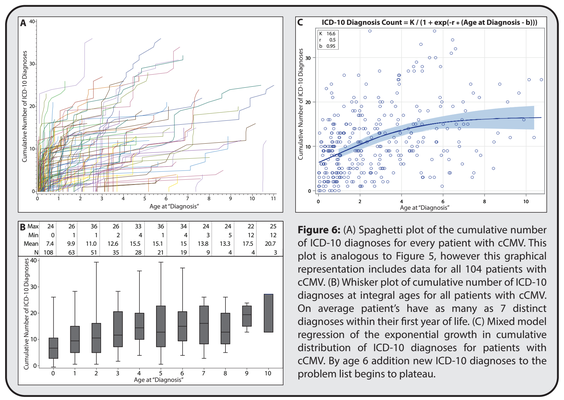
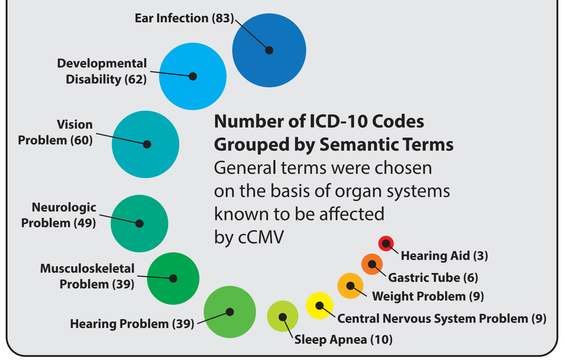


Figure 6: (A) Spaghetti plot of the cumulative number of ICD-10 diagnoses for every patient with cCMV. This plot is analogous to Figure 5, however this graphical representation includes data for all 104 patients with cCMV. (B) Whisker plot of cumulative number of ICD-10 diagnoses at integral ages for all patients with cCMV. On average patient's have as many as 7 distinct diagnoses within their first year of life. (C) Mixed model regression of the exponential growth in cumulative distribution of ICD-10 diagnoses for patients with cCMV. By age 6 addition new ICD-10 diagnoses to the problem list begins to plateau.

Figure 7



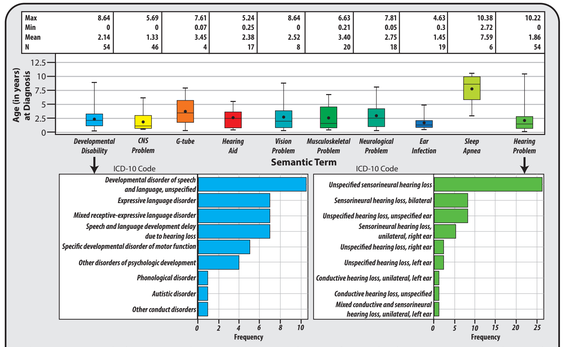


Figure 8: Age at which patients with congenital cytomegalovirus infection develop symptoms within the noted categories. Whisker plots with mean and standard deviation shown. Deficits within each system can be broken down into the specific ICD-10 code associated with each diagnosis.

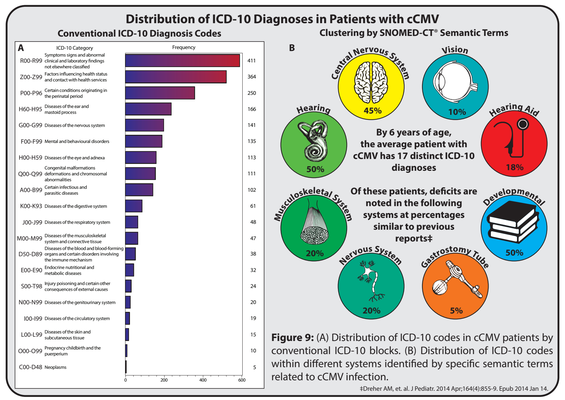


Figure 9: (A) Distribution of ICD-10 codes in cCMV patients by conventional ICD-10 blocks. (B) Distribution of ICD-10 codes within different systems identified by specific semantic terms related to cCMV infection.

‡Dreher AM, et. al. J Pediatr. 2014 Apr;164(4):855-9. Epub 2014 Jan 14.

Discussion:

One advantage of an EMR is the ease with which clinical information can be retrieved for research. Much of our previous work on cCMV was derived from paper charts, a time-intensive process that was limited by the data entry step. These data extraction and parsing scripts (Poster #132) make it possible to capitalize on the structure of the EMR and extract relevant clinical information expeditiously. For this project, we focused on health outcomes related to cCMV and so diagnosis codes were our priority. Fortunately, the increase in granularity between ICD-9 and ICD-10 codes allows for more accurate tracking of clinical outcomes, however this comes at the cost of needing additional analytics to aggregate similar diagnoses (**Figure 7**). Despite our unconventional approach to "chart review" and data analysis for this patient population, our findings recapitulate what has been seen in previous studies.

Limitations:

The biggest limitation of this study is the semi-cross-sectional approach to data collection. As diagnoses in the EMR can be "resolved", this approach has the potential to miss certain non-chronic diagnoses that physicians remove from the a patient's problem list. Furthermore, progression of a specific disease state may only be captured as the most recent diagnosis in the chart. Finally, if physicians are not prudent in how they enter medical diagnoses into the EMR (e.g. Congenital CMV diagnosed at age 10 years), accurately defining rates of disease progression can be obscured.

Future Directions:

The spectrum of disease of cCMV is broad and has been well studied in the past. The EMR gives us the potential to further study this disease in finer detail and identify rates of disease progression by mining the ICD-10 codes associated with these patients throughout time. This fine granular approach to disease progression should be invaluable when coupled with cost modeling rates to develop life-time adjusted cost analyses for cCMV. While this analysis was focused on the rate of disease progression for cCMV, the programs employed here can be generalized to other clinical questions.

Conclusion:

The spectrum of disease of cCMV is broad and has been well studied in the past. The EMR gives us the potential to further study this disease in finer detail and identify rates of disease progression by mining the ICD-10 codes associated with these patients throughout time. These results should prove invaluable for generating cost-models for the economic impact of cCMV. Integrating the principles behind clinical informatics into the daily work-flow of the physician will enable similar work to be completed expeditiously, especially with the widespread adoption of EMRs.

The ability to exhaustively analyze hundreds to thousands of clinical records, in a fraction of the time it took prior to widespread adoption of EMRs, opens the door to medical education at the level of the physician as a data entry specialists with specific content knowledge. When approached from this perspective, accurate documentation of clinical information in areas of the EMR should permit outcomes based research and best care practice guideline development in a timely manner. It is conceivable that in the future, iterative refinements or updates of these guidelines could be performed at the local level with a simple query of the EMR followed by a standardized analysis of the data.

Acknowledgments:

We would like to thank the Pediatric Infectious Diseases Program at Texas Children's Hospital a member of the Baylor College of Medicine Affiliated Hospitals for their support of Dr. Rochat during his fellowship training.

DIGITAL POSTER