Revolutionizing COVID-19 research: an EHR-enabled Learning Health Network

Ann Marie Navar  
MD, PhD, Associate Professor of Medicine, Duke Clinical Research Institute, Duke University School of Medicine, Strategic Advisor, Cerner  
Ann.Navar@duke.edu

Eric Peterson  
MD, MPH, Distinguished Professor of Medicine, Duke Clinical Research Institute, Strategic Advisor, Cerner  
Eric.Peterson@duke.edu
Real-World Data in a pandemic

The critical value Real-World Data can provide:

- Understanding disease epidemiology
- Identifying prognostic factors
- Determining effects on other disease states
- Generating hypotheses re: new therapies

Clinical Characteristics of Covid-19 in New York City

TO THE EDITOR: The world is in the midst of the coronavirus disease 2019 (Covid-19) pandemic, and New York City has emerged as an epicenter. Here, we characterize the first 793 consecutive patients with Covid-19 who were admitted to two hospitals in New York City.

This retrospective case series includes adults 18 years of age or older with confirmed Covid-19 who were consecutively admitted between March 3 (the date of the first positive case) and March 21, 2020, at an 861-bed quaternary referral center and an affiliated 185-bed non-teaching community hospital in New York City. Both hospitals adopted an antimicrobial strategy with limited use of high-flow nasal cannula during this period. Cases were confirmed through reverse-transcriptase polymerase-chain-reaction assays performed on nasopharyngeal swab specimens. Data were manually abstracted from electronic health records with the use of a query-embedded protocol and structured abstraction tool macros provided in the InESHealth software in the InEsHealth registry, available with the full text of this letter at NEJM.org.

Among the 793 patients, the median age was 62.2 years, 72% were male, and 75.5% had obesity (BMI ≥ 30). The most common presenting symptoms were cough (75.8%), fever (53.7%), dyspnea (35.8%), myalgia (23.8%), diarrhea (3.7%), and nausea and vomiting (3.1%) (Table 1, in the Supplementary Appendix). Illness of the patients (69%) had lymphopenia, 2% had thrombocytopenia, and many had altered comorbidity function values and inflammatory markers. Between March 3 and April 18, mortality rates were higher among patients in both the New Eschler Health System and in Northern California on April 26, 2020. We compared these data with data from the same period of 917 incident cases among 407,500 population members of the Covid-19 period in March 4, 2020, adjusted for the period from 2017-2018 and 2019). We compared these data with data from the corresponding period in 2019 (January 1 through April 15, 2019). In addition, outcomes for both weeks from March 3, 2020, a total of 369 patients tested positive, of whom 145 were excluded (due to minimal data on race or ethnic group. Of the 140 Covid-19-positive patients included in our analysis, 60% were female, 35% were black non-Hispanic, and 26% were white non-Hispanic. Mean number of deaths was 95. Mean number of deaths was 95.

Hospitilization and Mortality among Black Patients and White Patients with Covid-19

Tomi G. Price-Haywood, M.D., M.P.H., Jeffrey Burton, Ph.D., Daniel Fort, Ph.D., and Leonardo Samaire, M.D.

ABSTRACT

TO THE EDITOR: During the Covid-19 pandemic, reports have suggested that the mortality rate among black patients with Covid-19 is increased compared to that among white patients. In this analysis, we compared hospitalization and mortality rates among white and black patients with Covid-19 in two non-teaching community hospitals in New York City and Northern California.

We found that hospitalization rates among white patients with Covid-19 were higher than those among black patients (8.9% vs. 6.2%, respectively). The difference in hospitalization rates was not statistically significant (OR, 1.46; 95% CI, 0.78 to 2.68, P = 0.21).

ORIGINIAL ARTICLE

Renin–Angiotensin–Aldosterone System Blockers and the Risk of Covid-19

Giovannetti Mancini, M.D., Federico Res, Ph.D., Monica Legnani, M.Sc., Giovann Apollone, M.D., and Giovanni Cenni, Ph.D.

ABSTRACT

A potential association between the use of angiotensin-converting enzyme inhibitors and the risk of coronavirus disease 2019 (Covid-19) has been well studied.

METHODS

We carried out a population-based case-control study in the Lombardy region in Italy. A total of 627 case patients in whom infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was confirmed between February 21 and March 11, 2020, were matched to 6270 beneficiaries of the Regional Health Service (4 to 1 ratio according to age, sex, and municipality of residence). Information about the use of selected drugs and parity, clinical profile was drawn from regional databases of health care use. Odds ratios and 95% confidence intervals for associations between drugs and infection, with adjustment for confounders, were estimated by means of logistic regression.

RESULTS

Among both case patients and controls, the mean (SD) age was 68.2 (18.1) years, and 57% were women. The use of ACE inhibitors and ARBs was more common among case patients than among controls, so was the use of other antihypertensive and non-angiotensin-converting enzyme inhibitors, and case patients had a worse clinical profile. Use of ARBs or ACE inhibitors did not show any association with Covid-19 among case patients overall (adjusted odds ratio [AOR], 0.97; 95% confidence interval [CI], 0.73 to 1.28 for ARBs and 0.96 [0.89, 1.03] for ACE inhibitors) in a meta-anal
Data quality and reliability

Data quality issues include:

• Accuracy
• Completeness
• Representativeness
• Comparability across systems

Data integrity issues include:

• Replication
• External validation
• Transparency
Open science is good for researchers, good for innovation, and good for patients and the public. “The question at the center of the open science discussion is not whether data should be shared, but how we can usher in responsible methods for doing so.”

Eric Peterson, MD
Data access procedures

Step 1: Academic centers invited to apply

Step 2: Application fields:
- Study questions
- Data elements needed
- Statistical expertise

Step 3: Cerner team reviews:
- Feasibility
- Warnings about duplication, not prohibited

Step 4: Institution obtains IRB approval

Step 5: Data use agreement between institution and Cerner
Data protection

**Use of data**
- Only approved researchers granted access
- Research proposals overseen by Cerner review team

**Safety and security**
- Data fully de-identified
  - Dates, extreme ages
- Data only analyzed on Amazon Web Services (AWS) cloud
- Local internal review board (IRB) approval required

**Health system and provider protection**
- Sites anonymized
- Provider identifiers removed
- User agreement prohibits site or provider identification in publications
COVID-19 cohort: approved organizations
Early use cases for COVID-19 data set

Studies that are feasible:

- Comparative effectiveness of treatments
- Racial disparities
- Complication rates

Studies that are not feasible:

- Trends in time
- Geographic disparities
- In-depth socioeconomic evaluations
- Studies requiring NLP/access to notes:
  - Imaging characteristics, etc.

- Risk factors and comorbidities 31%
- Social determinants of health and demographics 14%
- Care pathways 19%
- Intubation and ventilation 7%
- Lab testing 5%
- Medications 24%
Beyond COVID: Cerner Real-World Data

Data contributor extraction

De-identification per HIPAA standards and also applying safe harbor methodology

Data mart Creation

Data delivery HealtheDataLab
Data elements

- Demographic
- Immunization
- Medication
- Medication administration
- Encounter
- Condition
- Result
- Procedure
- Order list
Cerner Real-World Data metrics*

89M patients

450M outpatient encounters

18M inpatient encounters

48M emergency encounters

161M procedures

593M diagnoses

903M medications

6M microbiology results

11.5B lab results

28M immunizations

12M type II diabetes

16M hypertension

3M atrial fibrillation

3.5M asthma

2.1M COPD

*All data pulled from HealtheIntent™ and current as of January 2020.1 Calculated using distinct person IDs which leverage a multipoint match algorithm to account for and remove duplicates; 2Leveraging Cerner standard ontologies to standardize and account for results among disparate coding systems and using unique encounter IDs to prevent over-inflation of data.

© Cerner Corporation. All rights reserved.
Cerner Real-World Data metrics

Gender

- Female
- Male
- Transgender

Ethnicity

- Hispanic
- Non-Hispanic

Real-World Data race distribution

- Other Race
- Unknown
- American Indian or Alaska Native
- Native Hawaiian or Other Pacific Islander
- White
- Black/African American
- Asian

Patient observation period length

- Percentage of patients

Encounters per U.S. Region

- Patient observation period (years)

*All data current as of January 2020 and pulled from HealtheIntent™. ¹Leveraging Cerner standard ontologies to standardize and account for results among disparate coding systems; ²Observation period calculated by comparing the first to the last encounter determined by distinct person ID; ³Calculated using unique encounter IDs per US region
The Cerner Learning Health Network
Governance council

The Learning Health Network (LHN) Governance Council is tasked with ensuring that the use of data is ethical, confidential, and represents the interests of patients, providers and member organizations.

The Governance Council helps ensure:

✓ Confidentiality of patient, provider and site data
✓ Data source linkages are appropriate, guided by science, only for research purposes and protect patient confidentiality
✓ Appropriate and ethical use of research including registry, trial or implementation projects
✓ All conducted research considers the rights and perspectives of patients, providers and health care systems
Learning Health Network and Real-World Data uses

Data science
• Near real-time data access across multiple systems
• Improved data quality and more structured elements, harnessing natural language processing (NLP)
• Cloud-based analytic tools
• Data privacy protections and patient input

Registries
• Key elements enabled auto-capturing
• Broader patient enrollment
• Long-term safety surveillance

Facilitate pragmatic RCTs
• Single, harmonized data stream from various sites
• EHR-based patient identification, recruitment and PRO collection

Implementation science
• Alerts, decision support and provider feedback
• Tools across Cerner Learning Health Network tested for effectiveness
HealtheIntent enabled Learning Health Network

Creating real-world data (RWD) resources and empowering research collaboration

Create networks for:
- Rare disease
- Comparative research and best practices
- Clinical trials
- Cross-organizational quality improvement

Sources (ex: EHRs)  Tenant A  Sources (ex: EHRs)  Tenant B

ID program  Populations

Chronic condition
Chronic condition
Study population
Study population (consented)
Study population
Study population (consented)

Study network
CV Mobius: EHR-based longitudinal CV registry

Consented, prospective enrollees
-~250 sites, N=8500
-Eligible ASCVD patients, elevated LDL-C

Consent/enrollment and baseline data
eCRF by study coordinator
-Consent for long-term follow-up
-Contact info, provider info and preferences

Follow-up visits
-Labs, vitals, medications, events, and clinical information collected via eCRF

Passive (consented) data collection
-Mortality through national databases
-Linked claims/registry data

Baseline EHR data for all patients presenting with ASCVD
-40 Health systems
-N = Millions of CVD patients
-Medications, comorbidities, lipids

Annual EHR-based patient updates
-Demographics
-Risk factor control (lipids, BP)
-Medication use
-In-system event rates
Janssen pilot cardiovascular disease registry

Collaborators
- University of Missouri
- UT Dell Medical School/Dell Seton

Data
- De-identified EHR data on all patients with ASCVD
- Linked to NDI and CMS claims by Cerner with IRB waiver

Chart reviews
- Validate clinical phenotypes
- Assess longitudinal events
- Capture non-standard data
Phase 2: quality improvement

Expand to 20 health systems

Rapid validation of computable phenotypes

Track impact of QI interventions for PAD
Future directions

Development of custom, study-specific tools
- Interoperable frameworks
- Shared decision-making
- Patient-based tools

Linkage of new data sources

Tailored data exploration experiences

Advancements in NLP and data curation

Exploration, standardization and incorporation of genomic data types

New mechanisms for data cleaning and research-readiness
Q & A
Thank you.

Ann Marie Navar  
MD, PhD, Associate Professor of Medicine, Duke Clinical Research Institute, Duke University School of Medicine, Strategic Advisor, Cerner  
Ann.Navar@duke.edu

Eric Peterson  
MD, MPH, Distinguished Professor of Medicine, Duke Clinical Research Institute, Strategic Advisor, Cerner  
Eric.Peterson@duke.edu