

### ORIGINAL RESEARCH

## The Community-Acquired Pneumonia Organization (CAPO) Cloud-Based Research Platform (the CAPO-Cloud): Facilitating Data Sharing in Clinical Research

William A Mattingly,<sup>†</sup> Timothy L Wiemken, Stephen P Furmanek, Paula Peyrani, Robert R Kelley, Julio A Ramirez

### Abstract

**Background:** Pneumonia is a costly and deadly respiratory disease that afflicts millions every year. Advances in pneumonia care require significant research investment and collaboration among pneumonia investigators. Despite the importance of data sharing for clinical research it remains difficult to share datasets with old and new investigators. We present CAPOCloud, a web-based pneumonia research platform intended to facilitate data sharing and make data more accessible to new investigators.

**Methods:** We establish the first two use cases for CAPOCloud to be the automatic subsetting and constraining of the CAPO database and the automatic summarization of the database in aggregate. We use the REDCap data capture software and the R programming language to facilitate these use cases.

**Results:** CAPOCloud allows CAPO investigators to access the CAPO clinical database and explore subsets of the data including demographics, comorbidities, and geographic regions. It also allows them to summarize these subsets or the entire CAPO database in aggregate while preserving privacy restrictions.

**Discussion:** CAPOCloud demonstrates the viability of a research platform combining data capture, data quality, hypothesis generation, data exploration, and data sharing in one interface. Future use cases for the software include automated univariate hypothesis testing, automated bivariate hypothesis testing, and principal component analysis.

DOI: 10.18297/jri/vol1/iss3/9

Received Date: April 13, 2017

Accepted Date: May 8, 2017

Website: <https://ir.library.louisville.edu/jri>

#### Affiliations:

Department of Medicine, Division of Infectious Diseases, University of Louisville, Louisville, KY: (WAM, TLW, SPF, PP, RRR, JAR)

© 2017, The Author(s).



## 1 Introduction

Of all infectious diseases, community-acquired pneumonia (CAP) continues to be one of the most common causes of morbidity for adults in the world.<sup>1</sup> Improving patient outcomes is facilitated by the generation of new knowledge through clinical research and quality improvement activities. One of the largest CAP studies is The Community Acquired Pneumonia Organization (CAPO), originally developed in 1999 to facilitate collaborative international research on CAP<sup>2</sup>. CAPO currently includes investigators from over 30 countries and 130 healthcare facilities, and the CAPO database houses data for nearly 20,000 hospitalized patients with CAP.

The CAPO database uses the Research Electronic Data Capture software or REDCap. REDCap is a clinical data management software that is accessible online through a secure portal. CAPO members access the database with a unique username and password in order to enter data. The database and CAPO web portal are maintained by the CAPO support group and any pneumonia investigator may contact the group for the creation of a free account.<sup>3</sup>

The primary benefit of this type of shared collaborative database is the large number of case records it makes available to researchers. Since 2003 the NIH has maintained an official policy regarding data sharing which highlights the importance of collaboration to research.<sup>4</sup> This policy emphasizes the importance of data sharing to the translation of research into practice and requires that NIH funded studies include a data sharing plan that makes available the timely release and sharing of data. The increasing availability of the internet has provided a greater capability to share data among researchers, but successful dissemina-

<sup>†</sup>Correspondence To: William A. Mattingly, PhD  
Instructor  
University of Louisville Division of Infectious Diseases  
501 E. Broadway Suite 120; Louisville, KY 40202  
Office Phone: 502-852-1555  
Email: [bill.mattingly@louisville.edu](mailto:bill.mattingly@louisville.edu)

tion is still challenging for many reasons including differences in data formats, institutional intellectual property restrictions, and cost for de-identifying data.<sup>5,6</sup> Since its creation, CAPO investigators have published many studies based on data contributed by CAPO members<sup>7-35</sup> and each of these studies required careful analysis of data, as well as interpretation of results.

### 1.1 The Problem

Despite the utility that the CAPO system makes available to investigators, there are several challenges that need to be overcome. One of the primary problems is that investigators need to contact the CAPO support group to determine if a study population or study variable of interest is available in the CAPO database. Once the population and variables are identified, the data must be analyzed appropriately. Currently, analyses are conducted through two means: 1) The CAPO support group at the University of Louisville provides a raw dataset to investigators for analysis, or 2) the CAPO support group analyzes data and reports back to the investigator. Both of these approaches are cumbersome and time-consuming, typically necessitating several meetings to explain the data, methods, and results.

These issues are not unique to the CAPO database. Presently the only way to determine if a large pneumonia database might be appropriate for secondary analysis is to contact the original investigators and request the data. Even if the original investigators are willing to collaborate, there may be added costs of data de-identification and intellectual property agreements between institutions. These will add weeks and months to the time needed to form the first hypothesis for a secondary analysis.

### 1.2 Objective

The objective of this project was to develop and implement a cloud-based, interactive analytic and data visualization research platform with the following aims: 1) Promote new research 2) facilitate education of new researchers 3) enable the exploration of research ideas not envisioned by the initial protocol 4) make possible the testing of new or alternative hypothesis and 5) facilitate exploratory data analysis of the original study database. We believe that using data analytics and visualization techniques can improve secondary data analysis not only for the field of pneumonia but also for clinical research in general.

## 2 Methods

**Table 1** summarizes many of the terms used in the design and development of software for assisting with data analysis.

### 2.1 Use Cases

When designing software, use cases are used to describe the purpose that software affords to its users. Use cases are described from the perspective of a typical user of the software. Once use cases are established, software features are created that will afford and support all of the necessary use cases. In the case of CAPO-cloud, we focus on two primary use cases: 1) A CAPO Investigator wants to constrain, or subset, the CAPO database and

**Table 1** Glossary

Key Terms	Description
aggregate	data combined from several measurements
API	application programmer interface - flexible software tools for modifying program functionality
checkbox	An on-screen graphical element showing a check when clicked
open source	a software licensing model that allows for no cost redistribution of software and editing of the program
proprietary	a software licensing model that does not allow editing of the program
server	a computer program that provides data to other computers on a network
subset	to reduce the size of a dataset by excluding records
use case	a desired activity a user wishes to perform

2) A CAPO investigator wants to summarize or describe the CAPO database in aggregate.

Every feature of the app will be made to address a primary use case. Addressing the first use case of subsetting the database, requires a feature that allows selecting features of the database such as demographics. These selections effectively exclude portions of the CAPO database that do not meet the inclusion criteria of a research question and is usually done in software by means of a checkbox. After subsetting the database, an additional feature is needed to provide summarization of the subset. This feature can be added by performing calculations and presenting results in a tabular or graphical format.

To satisfy these use cases, we integrate three open source software projects: 1) Research Electronic Data Capture (REDCap)<sup>36</sup>, 2) The R Programming Language<sup>37</sup>, and 3) R-Studio/Shiny<sup>38</sup>.

### 2.2 REDCap

REDCap is an electronic data capture software that is developed and maintained by Vanderbilt University. It is web-based software that can be obtained for eligible organizations at no cost. The CAPO REDCap database is housed on a secure web server and comprises multiple data collection forms. REDCap has application programmer interface (API) functionality, allowing REDCap to communicate with a variety of other software and programming languages.

### 2.3 R

R is a statistical programming language based around S<sup>39</sup>. R is open source and free, allowing contributors to write software packages and contribute to software repositories such as CRAN<sup>40</sup>, GitHub<sup>41</sup>, and BIOConductor<sup>42</sup>.

### 2.4 Shiny

Shiny is an R software package created by RStudio<sup>43</sup> that allows programmers to build interactive applications that allow for the visualization and sharing of data. As with web applications, a

Shiny application is a web page that accepts inputs from users via on-screen controls, puts the user inputs into R running on a server or local computer, and then outputs results back to the web page, updating or generating the information displayed. Shiny applications are usually composed of two key units of code: a script for the functions run by the server and a script for the user interface.

### 3 Results

#### 3.1 Integration

Although REDCap does not use R, the REDCap API allows it to work in conjunction with many modern programming languages including R, Python<sup>43</sup>, PHP<sup>44</sup>, and Perl. Through the API, a software developer can create a REDCap "plug-in" or a feature that is added directly to the REDCap software to be used through its native interface. The REDCap plug-in runs on the same web server as the REDCap instance, and connects to the Shiny server which can be running on the same hardware or on a remote system, depending on performance and memory requirements. **Figure 1** depicts the CAPO-Cloud plug-in for REDCap.

#### 3.2 CAPO-Cloud

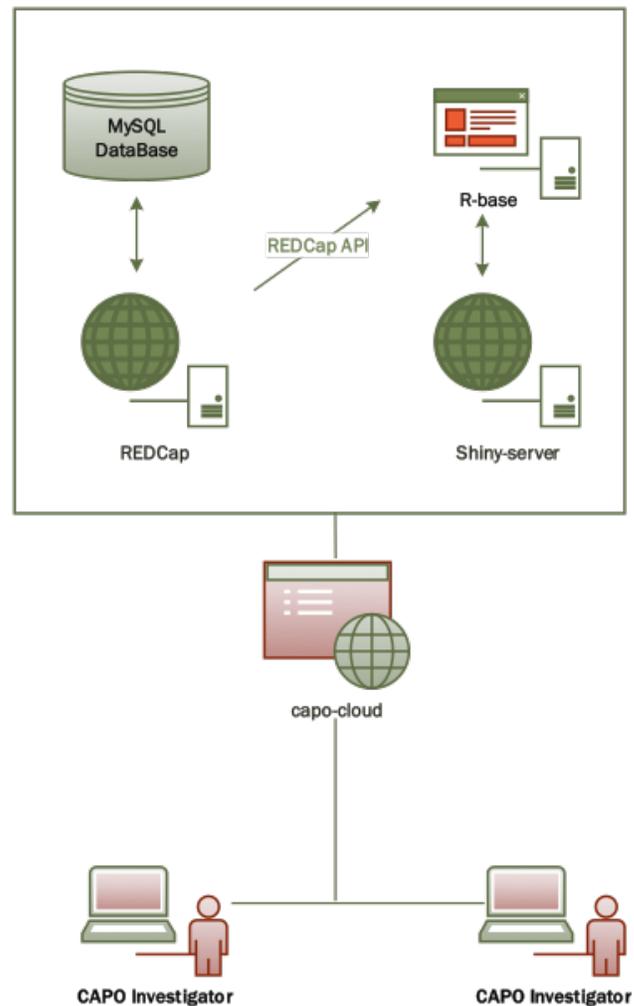
After logging in to the CAPO REDCap instance, investigators can navigate to the CAPO-Cloud plug-in by clicking the CAPO-Cloud link on the sidebar. This will take them to the CAPO-Cloud About page, which offers a brief description of the CAPO study and a patient characteristics table for the whole dataset (see **Figure 2**).

The next tab, *Subset Data*, provides the subset selection feature for the CAPO dataset. Currently, investigators can subset based on date of admission, sex of patient, age of patient, patient history of comorbidities, and region. The subsets made on this page will carry over to all subsequent pages, allowing investigators to get a unique snapshot of the database from a perspective that meets their research needs. **Figure 3** shows the total number of male patients aged 85 and over admitted between 2000-04-17 and 2016-11-03 with a history of COPD in the United State and Canada.

The third tab, *Descriptive Statistics*, shown in **Figure 4** allows investigators to look at summary statistics in more detail. Statistics are accompanied by tables and graphs such as histograms and bar charts. This tab is separated into multiple sections, allowing investigators to look at patient history and demographics, physical exam values, laboratory values, severity of disease markers, and outcomes such as length of stay or in-hospital mortality.

#### 3.3 Security and Privacy

The CAPO database is not completely de-identified, but the summary statistics provided by CAPO-cloud effectively mask protected health information. Access is only available upon requesting a CAPO REDCap account, and all access is tracked using REDCap's audit trail.



**Fig. 1** CAPO-Cloud architecture. Using the REDCap API to connect to R, CAPO-Cloud makes use of statistical functions using R-base (core R programming language) and applies them to data stored in REDCap. These functions and results are then displayed to the user, via the shiny-server web interface.

### 4 Discussion

In this study, we developed and implemented a new cloud-based interactive analytic and data visualization research platform, CAPO-Cloud, to facilitate hypothesis generation and secondary data analysis of the CAPO international database. The platform allows investigators to start secondary data analysis in an interactive way that is less time consuming and more accessible than what was previously available. It supports real-time subsetting and summarizing of the CAPO clinical database.

Increased focus on research has improved the availability of software which supports exploratory analysis. Proprietary software like Microsoft Excel and Tableau can load a data file exported from REDCap or other data collection systems and display graphs and tables similar to what is seen in CAPO-Cloud. Companies like IBM and Google have also invested heavily in providing open

**(ID) Community-Acquired Pneumonia Organization International Cohort Study**

CAPO-Cloud About Subset Data Descriptive Statistics

The Community-Acquired Pneumonia Organization (CAPO) was created in 1999, with the goal of facilitating international research in the field of community-acquired pneumonia (CAP). Led by Dr. Julio Ramirez, Chief of Infectious Diseases at the University of Louisville, CAPO includes investigators from over 30 countries and 130 hospitals; it strives to advance the state-of-the-art in CAP therapy world-wide.

The primary objective of the core study is to evaluate the quality of care delivered to hospitalized patients with CAP. The core study protocol and case report form were developed by the original steering committee and have been through several updates over the years.

In addition to the original study, several secondary analyses have been performed using CAPO data on various topics including, but not limited to:

Variable	Overall
Total N	8210
Age at Admission (mean (sd))	64.48 (19.23)
Male sex (%)	4914 (59.9)
Pregnant (%)	24 ( 2.7)
Patients directly admitted to hospital ICU (%)	863 (10.5)
Patients needing ventilatory support on first day (%)	510 (13.9)
Patients needing vasopressors on first day (%)	232 ( 6.6)
History of neoplastic disease (%)	806 ( 9.8)
History of congestive heart failure (%)	1364 (16.6)
History of cerebrovascular disease (%)	1094 (13.3)
History of renal disease (%)	819 (10.0)
History of liver disease (%)	436 ( 5.3)
History of renal failure (%)	747 ( 9.1)
History of diabetes (%)	1439 (18.3)
History of cirrhosis (%)	66 ( 1.8)
History of asplenia (%)	18 ( 0.5)

**Fig. 2** The *About* page for the CAPO-Cloud plug-in. The standard REDCap command options are seen on the left side, with choices for determining project setup. Under the project title, the interface for CAPO-Cloud is shown, and lists a patient characteristics table for the complete dataset.

CAPO-Cloud About Subset Data Descriptive Statistics

**Subset by:**

- Date
- Sex
- Age
- Comorbidity or Risk Factor
- Region

**Include only admission dates in this range:**

2000-04-27 to 2016-11-03

**Include only:**

Males

**Include only patients in this age range:**

18 to 104

**Include only patients with:**

- Neoplastic disease
- Congestive heart failure
- Cerebrovascular disease
- Renal disease
- Liver disease
- Diabetes
- Cirrhosis
- Asplenia
- Chronic obstructive pulmonary disease
- HIV
- Nursing home resident
- Altered mental status on admission

**Include only patients from:**

- USA/Canada
- Europe
- Latin America
- Asia/Africa

Subset data based on selections    Reset selections and start over

**Total number of observations: 358**

**Fig. 3** The *Subset Data* page

CAPO-Cloud About Subset Data Descriptive Statistics

Demographics and Patient History    Physical Exam Results    Lab Values    Disease Severity and Outcomes

Variable	Overall
Total N	358
Height in cm (mean (sd))	175.30 (7.59)
Weight in kg (mean (sd))	77.41 (21.70)
Heart rate in beats/minute (mean (sd))	101.87 (22.17)
Respiratory rate in breaths/minute (mean (sd))	23.91 (6.05)
Systolic blood pressure in mmHg (mean (sd))	122.53 (27.12)
Diastolic blood pressure in mmHg (mean (sd))	64.66 (12.31)
Temperature in degrees Celsius (mean (sd))	36.27 (2.26)
O2 saturation in percent (mean (sd))	90.64 (8.43)

Select a variable: DBP (mmHg)

**Data are not normally distributed**

Frequency

DBP (mmHg)

**Fig. 4** The *Descriptive Statistics* page

and available means to visualize data with their products Many Eyes<sup>45</sup> and Fusion Tables<sup>46</sup>. For investigators who are skilled with these tools, they are often sufficient to provide a general overview of a clinical dataset.

Excel provides a convenient way to interact with and edit data, and although it is proprietary software, it is not prohibitively expensive, and available to most students and employees at a reduced price or for free. However, Excel has limits when it comes to the size of a dataset that can be managed, and to performing any specialized statistics requires familiarity with Excel's formula language.

Tableau supports a point and click interface, allowing investigators to drag variables names to create charts and other data visualizations based on their selections. To get meaningful visualizations, a Tableau user must know what type of data is associated with each variable name. Most investigators, and especially investigators new to clinical research and statistics, could not discern this from the raw data alone, and so many datasets will have an accompanying data specification or data dictionary describing this information. This dictionary provides the data type, units, validity ranges, and any other relevant information about a variable. Even having access to this dictionary does not make initial data exploration steps approachable for a new investigator, as the amount of options available to new users when first opening the dataset in Tableau can be overwhelming.

None of this software achieves an integration with the same data collection and quality framework on which it was collected, and this is a key advantage of the CAPO-Cloud. Having the analysis and visualization tools ensures that investigators will only see capabilities that apply to the CAPO pneumonia dataset. This is crucial in making investigation approachable and reproducible.

#### 4.1 Strengths

All software used for this project is available at no cost to eligible organizations, and anyone with access to REDCap can create a similar platform tailored to their studies. Any investigator interested in pneumonia data can become a member of CAPO and be granted access to the CAPO clinical database and CAPO-Cloud.

The platform provides visualization that aids clinicians and researchers without a strong statistical background in the generation of new ideas for research. This allows researchers who were not part of the original study to initiate secondary data analyses of the clinical database on their own, allowing the CAPO support unit to focus on sophisticated, high-level analysis developed from investigators' hypotheses.

#### 4.2 Limitations

Currently, this platform can only be applied to one database (the CAPO database). Applying to other REDCap projects would require some modification from a programmer experienced in R and Shiny. REDCap is not open source and is only available upon request from the REDCap consortium. REDCap can only be licensed for use that is considered not-for-profit. This includes academic,

government and non-profit organizations, but not industry or private enterprise.

#### 4.3 Future research

CAPO-Cloud supports use cases for the exploratory browsing and descriptive statistics of the CAPO database for CAPO investigators. In **Table 2** we list several use cases and the features to support them that we anticipate to be useful to CAPO investigators in the future.

**Table 2**

Use case	Feature(s) to be added
Investigators wish to: perform hypothesis testing between two variables.	Inferential Statistics (bivariate) Tab: Chi-square tests, two population t-tests, simple linear regression
perform multivariate hypothesis testing, allowing for adjustment of other variables.	Inferential Statistics (multiple regression) Tab: Multiple regression, logistic regression, parametric survival analysis
identify outliers and perform basic data quality monitoring.	Data Quality Improvements: Identification of outliers, tools to address data quality fixes
get more meaningful interpretations for statistical tests performed.	Educational component: Explain tests, explain what results mean, provide meaningful context for statistical
perform more complicated analyses or select variables for analysis based on data-driven methods.	Advanced Inferential tab: Principal component analysis, genetic algorithms for variable selection, machine learning processes for model selection
use other pneumonia clinical databases in conjunction with CAPO international database.	Expand application to be usable for other pneumonia databases
use other clinical databases for other diseases.	Expand application to be usable for any clinical database

Many of these functions are in use by the CAPO coordinating center as standalone R programs, and only need adjustment for presentation on a web page through Shiny.

In order to understand the impact that CAPO-Cloud has on investigators, it will be important to track CAPO-Cloud usage and follow up regularly with investigators using it, to determine which features are the most useful. Features that facilitate studies leading to publication will have the highest priority.

#### 4.4 Conclusions

In conclusion, we have presented our project CAPO-Cloud, a REDCap plug-in that aims to create a basic structure for a cloud-based research program facilitating collaborative research and data sharing at the international level. We believe that tools like these will greatly enhance the ability of non-statisticians to perform preliminary investigations into important new clinical research questions.

**Author contributions:** WAM, SPF, and TLW wrote the first draft of the manuscript. All authors critically reviewed the manuscript for important intellectual content. All authors agree with the manuscript results and conclusions.

**Funding source:** This study project was unfunded.

**Competing interests:** All authors declare no competing interests.

## References

- 1 World Health Organization, "The top 10 causes of death," 2015 [cited 2017]. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs310/en/>
- 2 CAPO, "Community-Acquired Pneumonia Organization," 2017 [cited 2017]. [Online]. Available: <http://caposite.com/>
- 3 CAPO, "CAPO REDCap," 2017 [cited 2017]. [Online]. Available: <http://caposite.com/>
- 4 National Institutes of Health *et al.*, "NIH data sharing policy," Retrieved from, 2003.
- 5 J. M. Drazen, S. Morrissey, D. Malina, M. B. Hamel, and E. W. Campion, "The importance-and the complexities-of data sharing," 2016.
- 6 C. J. Savage and A. J. Vickers, "Empirical study of data sharing by authors publishing in plos journals," *PloS one*, vol. 4, no. 9, p. e7078, 2009.
- 7 C. M. Luna, I. Palma, M. S. Niederman, E. Membriani, V. Giovini, T. L. Wiemken, P. Peyrani, and J. Ramirez, "The impact of age and comorbidities on the mortality of patients of different age groups admitted with community-acquired pneumonia," *Annals of the American Thoracic Society*, vol. 13, no. 9, pp. 1519–1526, 2016.
- 8 G. Levy, M. Perez, B. Rodríguez, A. H. Voth, J. Perez, M. Gnoni, R. Kelley, T. Wiemken, and J. Ramirez, "Adherence with national guidelines in hospitalized patients with community-acquired pneumonia: results from the capo study in venezuela," *Archivos de Bronconeumología (English Edition)*, vol. 51, no. 4, pp. 163–168, 2015.
- 9 M. J. Lanspa, P. Peyrani, T. Wiemken, E. L. Wilson, J. A. Ramirez, and N. C. Dean, "Characteristics associated with clinician diagnosis of aspiration pneumonia: a descriptive study of afflicted patients and their outcomes," *Journal of hospital medicine*, vol. 10, no. 2, pp. 90–96, 2015.
- 10 R. Cavallazzi, T. Wiemken, F. W. Arnold, C. M. Luna, J. Bordon, R. Kelley, C. Feldman, J. D. Chalmers, A. Torres, and J. Ramirez, "Outcomes in patients with community-acquired pneumonia admitted to the intensive care unit," *Respiratory medicine*, vol. 109, no. 6, pp. 743–750, 2015.
- 11 M. Carugati, F. Franzetti, T. Wiemken, R. Kelly, P. Peyrani, F. Blasi, J. Ramirez, and S. Aliberti, "De-escalation therapy among bacteraemic patients with community-acquired pneumonia," *Clinical Microbiology and Infection*, vol. 21, no. 10, pp. 936–e11, 2015.
- 12 J. Bordon, R. Fernandez-Botran, T. Wiemken, P. Peyrani, S. Uriarte, F. Arnold, L. Rodriguez-Hernandez, M. Rane, R. Kelley, L. Binford *et al.*, "Bacteremic pneumococcal pneumonia: clinical outcomes and preliminary results of inflammatory response," *Infection*, vol. 43, no. 6, pp. 729–738, 2015.
- 13 T. L. Wiemken, R. M. Carrico, S. L. Klein, C. B. Jonsson, P. Peyrani, R. R. Kelley, S. Aliberti, F. Blasi, R. Fernandez-Gonzalez, G. Lopardo *et al.*, "The effectiveness of the polysaccharide pneumococcal vaccine for the prevention of hospitalizations due to streptococcus pneumoniae community-acquired pneumonia in the elderly differs between the sexes: results from the community-acquired pneumonia organization (capo) international cohort study," *Vaccine*, vol. 32, no. 19, pp. 2198–2203, 2014.
- 14 A. Liapikou, E. Polverino, C. Cilloniz, P. Peyrani, J. Ramirez, R. Menendez, A. Torres, C.-A. P. O. C. Investigators *et al.*, "A worldwide perspective of nursing home-acquired pneumonia compared to community-acquired pneumonia," *Respiratory care*, pp. respcare–02788, 2013.
- 15 S. Haubitz, F. Hitz, L. Graedel, M. Batschwaroff, T. L. Wiemken, P. Peyrani, J. A. Ramirez, C. A. Fux, B. Mueller, and P. Schuetz, "Ruling out legionella in community-acquired pneumonia," *The American journal of medicine*, vol. 127, no. 10, pp. 1010–e11, 2014.
- 16 R. Cavallazzi, T. Wiemken, D. Christensen, P. Peyrani, F. Blasi, G. Levy, S. Aliberti, R. Kelley, J. Ramirez, C.-A. P. O. C. Investigators *et al.*, "Predicting mycobacterium tuberculosis in patients with community-acquired pneumonia," *European Respiratory Journal*, vol. 43, no. 1, pp. 178–184, 2014.
- 17 A. Griffin, P. Peyrani, T. Wiemken, J. Ramirez, and F. Arnold, "Empiric therapy directed against mrsa in patients admitted to the intensive care unit does not improve outcomes in community-acquired pneumonia," *Infection*, vol. 41, no. 2, pp. 517–523, 2013.
- 18 F. W. Arnold, T. L. Wiemken, P. Peyrani, J. A. Ramirez, G. N. Brock *et al.*, "Mortality differences among hospitalized patients with community-acquired pneumonia in three world regions: Results from the community-acquired pneumonia organization (capo) international cohort study," *Respiratory medicine*, vol. 107, no. 7, pp. 1101–1111, 2013.
- 19 F. W. Arnold, T. L. Wiemken, P. Peyrani, M. Mirsaeidi, and J. A. Ramirez, "Outcomes in females hospitalised with community-acquired pneumonia are worse than in males," *European Respiratory Journal*, vol. 41, no. 5, pp. 1135–1140, 2013.
- 20 R. Riquelme, P. Jimenez, A. Videla, H. Lopez, J. Chalmers, A. Singanayagam, M. Riquelme, P. Peyrani, T. Wiemken, G. Arbo *et al.*, "Predicting mortality in hospitalized patients with 2009 h1n1 influenza pneumonia," *The International Journal of Tuberculosis and Lung Disease*, vol. 15, no. 4, pp. 542–546, 2011.
- 21 G. N. Brock, C. Barnes, J. A. Ramirez, and J. Myers, "How to handle mortality when investigating length of hospital stay and time to clinical stability," *BMC medical research methodology*, vol. 11, no. 1, p. 144, 2011.
- 22 M. Malinis, J. Myers, J. Bordon, P. Peyrani, R. Kapoor, R. Nakamatzu, G. Lopardo, A. Torres, C. Feldman, M. Allen *et al.*, "Clinical outcomes of hiv-infected patients hospitalized with bacterial community-acquired pneumonia," *International Journal of Infectious Diseases*, vol. 14, no. 1, pp. e22–e27, 2010.
- 23 A. Griffin, P. Peyrani, T. Wiemken, and F. Arnold, "Macrolides versus quinolones in legionella pneumonia: results from the

- community-acquired pneumonia organization international study,” *The International Journal of Tuberculosis and Lung Disease*, vol. 14, no. 4, pp. 495–499, 2010.
- 24 F. W. Arnold, G. N. Brock, P. Peyrani, E. L. Rodríguez, A. A. Díaz, P. Rossi, and J. A. Ramirez, “Predictive accuracy of the pneumonia severity index vs crb-65 for time to clinical stability: results from the community-acquired pneumonia organization (capo) international cohort study,” *Respiratory medicine*, vol. 104, no. 11, pp. 1736–1743, 2010.
  - 25 S. Aliberti, F. Blasi, A. Zanaboni, P. Peyrani, P. Tarsia, S. Gaito, and J. Ramirez, “Duration of antibiotic therapy in hospitalized patients with community-acquired pneumonia,” *European Respiratory Journal*, vol. 36, no. 1, pp. 128–134, 2010.
  - 26 F. W. Arnold, A. S. LaJoie, G. N. Brock, P. Peyrani, J. Rello, R. Menéndez, G. Lopardo, A. Torres, P. Rossi, and J. A. Ramirez, “Improving outcomes in elderly patients with community-acquired pneumonia by adhering to national guidelines: Community-acquired pneumonia organization international cohort study results,” *Archives of Internal Medicine*, vol. 169, no. 16, pp. 1515–1524, 2009.
  - 27 S. Aliberti, J. Myers, P. Peyrani, F. Blasi, R. Menendez, P. Rossi, R. Cosentini, G. Lopardo, L. de Vedia, and J. Ramirez, “The role of neutropenia on outcomes of cancer patients with community-acquired pneumonia,” *European Respiratory Journal*, vol. 33, no. 1, pp. 142–147, 2009.
  - 28 S. Aliberti, G. Brock, P. Peyrani, F. Blasi, J. Ramirez *et al.*, “The pneumonia severity index and the crb-65 in cancer patients with community-acquired pneumonia,” *The International Journal of Tuberculosis and Lung Disease*, vol. 13, no. 12, pp. 1550–1556, 2009.
  - 29 J. Bordón, P. Peyrani, G. N. Brock, F. Blasi, J. Rello, T. File, and J. Ramirez, “The presence of pneumococcal bacteremia does not influence clinical outcomes in patients with community-acquired pneumonia: results from the community-acquired pneumonia organization (capo) international cohort study,” *CHEST Journal*, vol. 133, no. 3, pp. 618–624, 2008.
  - 30 J. Ramirez, “Fostering international multicenter collaborative research: the capo project [research methods],” *The International Journal of Tuberculosis and Lung Disease*, vol. 11, no. 10, pp. 1062–1065, 2007.
  - 31 F. W. Arnold, J. T. Summersgill, A. S. LaJoie, P. Peyrani, T. J. Marrie, P. Rossi, F. Blasi, P. Fernandez, T. M. File Jr, J. Rello *et al.*, “A worldwide perspective of atypical pathogens in community-acquired pneumonia,” *American journal of respiratory and critical care medicine*, vol. 175, no. 10, pp. 1086–1093, 2007.
  - 32 P. Peyrani, D. Christensen, A. LaJoie, R. Nakamatsu, F. Arnold, P. Schulz, I. Toala, A. Ramirez, S. Lobo, B. Johnson *et al.*, “Antibiotic therapy of hospitalized patients with community-acquired pneumonia: an international perspective from the capo cohort study,” *The Journal of the Kentucky Medical Association*, vol. 104, no. 11, pp. 513–517, 2006.
  - 33 F. Arnold, A. LaJoie, T. Marrie, P. Rossi, F. Blasi, C. Luna, P. Fernandez, J. Porras, K. Weiss, C. Feldman *et al.*, “The pneumonia severity index predicts time to clinical stability in patients with community-acquired pneumonia,” *The International Journal of Tuberculosis and Lung Disease*, vol. 10, no. 7, pp. 739–743, 2006.
  - 34 J. A. Ramirez, C.-A. P. O. C. Investigators *et al.*, “Worldwide perspective of the quality of care provided to hospitalized patients with community-acquired pneumonia: results from the capo international cohort study,” in *Seminars in respiratory and critical care medicine*, vol. 26, no. 06. Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA., 2005, pp. 543–552.
  - 35 D. Christensen, C. Feldman, P. Rossi, T. Marrie, F. Blasi, C. Luna, P. Fernandez, J. Porras, J. Martinez, K. Weiss *et al.*, “Hiv infection does not influence clinical outcomes in hospitalized patients with bacterial community-acquired pneumonia: results from the capo international cohort study,” *Clinical infectious diseases*, vol. 41, no. 4, pp. 554–556, 2005.
  - 36 P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, “Research electronic data capture (redcap)—a metadata-driven methodology and workflow process for providing translational research informatics support,” *Journal of biomedical informatics*, vol. 42, no. 2, pp. 377–381, 2009.
  - 37 R Core Team, *R language definition*, R Foundation for Statistical Computing, Vienna, Austria, 2000.
  - 38 RStudio Team, *RStudio: Integrated Development Environment for R*, RStudio, Inc., Boston, MA, 2015. [Online]. Available: <http://www.rstudio.com/>
  - 39 R. A. Becker, J. M. Chambers, and A. R. Wilks, “The new s language,” *Pacific Grove, Ca.: Wadsworth & Brooks*, 1988, 1988.
  - 40 The R Foundation, “The Comprehensive R Archive Network,” 2017 [cited 2017]. [Online]. Available: <https://cran.r-project.org/>
  - 41 GitHub, Inc., “GitHub,” 2017 [cited 2017]. [Online]. Available: <https://github.com/>
  - 42 R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Detting, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry *et al.*, “Bioconductor: open software development for computational biology and bioinformatics,” *Genome biology*, vol. 5, no. 10, p. R80, 2004.
  - 43 Python Software Foundation, “Welcome to Python.org,” 2017 [cited 2017]. [Online]. Available: <https://www.python.org/>
  - 44 The PHP Group, “PHP: Hypertext Processor,” 2017. [Online]. Available: <https://secure.php.net/>
  - 45 F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon, “Manyeyes: a site for visualization at internet scale,” *IEEE transactions on visualization and computer graphics*, vol. 13, no. 6, 2007.
  - 46 H. Gonzalez, A. Halevy, C. S. Jensen, A. Langen, J. Madhavan, R. Shapley, and W. Shen, “Google fusion tables: data management, integration and collaboration in the cloud,” in *Proceedings of the 1st ACM symposium on Cloud computing*. ACM, 2010, pp. 175–180.