Most Common Statistical Methodologies in Recent Clinical Studies of Community-Acquired Pneumonia

Stephen Furmanek
Division of Infectious Diseases / University of Louisville, Louisville, KY, USA

Connor L. English
University of Louisville

Thomas Chandler
University of Louisville

Timothy L. Wiemken PhD
University of Louisville

Follow this and additional works at: http://ir.library.louisville.edu/jri

Part of the Community Health and Preventive Medicine Commons, Epidemiology Commons, Health Information Technology Commons, Influenza Humans Commons, Influenza Virus Vaccines Commons, International Public Health Commons, and the Translational Medical Research Commons

Recommended Citation
Furmanek, Stephen; English, Connor L.; Chandler, Thomas; and Wiemken, Timothy L. PhD (2017) "Most Common Statistical Methodologies in Recent Clinical Studies of Community-Acquired Pneumonia," The University of Louisville Journal of Respiratory Infections: Vol. 1 : Iss. 4 , Article 9.
DOI: 10.18297/jri/vol1/iss4/9
Available at: http://ir.library.louisville.edu/jri/vol1/iss4/9

This Review Article is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in The University of Louisville Journal of Respiratory Infections by an authorized editor of ThinkIR: The University of Louisville's Institutional Repository. For more information, please contact thinkir@louisville.edu.
Most Common Statistical Methodologies in Recent Clinical Studies of Community-Acquired Pneumonia

Cover Page Footnote
Correspondence To: Stephen Furmanek Work Address: 501 East Broadway, Suite 120B Louisville, KY 40202
Work Email: stephen.furmanek@louisville.edu
Introduction

Community-acquired pneumonia (CAP) is the leading cause of death from infectious disease in the United States. [1] Guidelines for the management of hospitalized patients with CAP have been structured around the results of clinical research activities. Although there is a wealth of data in this area, there are still important knowledge gaps for the management of hospitalized patients with CAP that will need to be resolved with further clinical research. Training new individuals is imperative to produce a new generation of clinical investigators with the expertise necessary to fill gaps in knowledge. Within the University of Louisville Division of Infectious Diseases, we have implemented novel clinical research training programs with this goal in mind.

In these courses, several challenges were discovered when teaching biostatistics. First, although a comprehensive understanding of all statistical methodologies used in a particular field is ideal, in practice this may be unnecessary. By attempting to teach with a goal of comprehensive understanding, practical ideas are lost in a sea of formulas. Furthermore, clinical investigators are often intimidated by their unfamiliarity with statistics. In an attempt to resolve some of these challenges, we consider that it is important to limit statistical teaching to the most pertinent tests and topics in clinical research. Although this knowledge does not replace the importance of enlisting the expertise of a biostatistician for all stages of a research project, it does facilitate a broad understanding of the basic concepts of biostatistics and assists in critical analysis of study results. This knowledge leads to higher quality studies and a better understanding of how results of other studies can be incorporated into clinical practice.

The primary objective of this study is to define the most common statistical methodologies in recent clinical studies of CAP to inform teaching approaches in the field. Secondary objectives...
were: 1) define the most common study designs, 2) define statistical tests to be included in a curriculum to educate clinical investigators interested in clinical research of CAP, 3) generate a glossary defining the most common statistical tests.

Methods

This was a literature review including recent publications in the field of CAP. Articles were eligible for inclusion in this study if they: 1) were clinical research with an emphasis on incidence, epidemiology, or patient outcomes, 2) were searchable via PubMed or Google Scholar, 3) were published within the timeframe of January 1st 2012 to August 1st 2017, 4) contained Medical Subject Headings (MeSH) keywords of “pneumonia” and one of the following: “epidemiologic studies”, “health services research”, or “comparative effectiveness research” or search keywords of “community-acquired pneumonia” and one of the following: “cohort study”, “observational study”, “prospective study”, “retrospective study”, “clinical trial”, “controlled trial”, or “clinical study”. A Delphi panel of CAP clinical investigators decided by vote on which 30 articles to include in this review based on their clinical relevance and applicability to clinical practice.

Only statistics mentioned in the body of the article were included in our evaluation; information in supplementary appendices were not evaluated.

Statistical methodologies were divided into five categories: 1) descriptive statistics, 2) inferential statistics and procedures, 3) graphics and figures, 4) study design, and 5) statistical software.

Descriptive statistics were performed to summarize the tests and procedures performed, represented by frequency and percent. The most frequent statistical methodologies were reported. R software version 3.3.2 was used for all analysis.

A test was considered for curriculum if it was present in more than 30% of the articles reviewed. Based on this criterion, we identified statistical tests to be included in a curriculum to educate individuals interested in clinical research of CAP. Tests were grouped into sessions based on their application in clinical research.

Using this curriculum, we developed a glossary for the most common statistical tests to be included in a curriculum to educate clinical investigators interested in clinical research of CAP. Definitions were written so that they are meaningful to non-statisticians and did not include mathematical formulae.

Results

Thirty studies were evaluated based on decisions by the Delphi panel [2-31]. Study characteristics are shown in Table 1. Every study that was evaluated included a descriptive analysis of patient characteristics. Every study included frequencies for various descriptive analyses, notated by n, and percent. Medians were the most frequently reported measures of central tendency (n=22 [73%]), and interquartile ranges (IQRs; n=19 [63%]) were the most frequently reported measure of variability.

P-values were reported in 27 (90%) studies. In 19 (63%) studies, cutoff values for statistical significance were explicitly stated. Confidence intervals were reported in 21 (70%) studies. Bivariate comparisons were performed most commonly with Wilcoxon rank sum tests (n=16 [53%]) and Chi-squared tests (n=20 [67%]). Multivariable analyses were most commonly logistic regression (n=17 [60%]) and Cox Proportional Hazards regression (n=10 [33%]).

Table 1. Study characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Design</td>
<td></td>
</tr>
<tr>
<td>Retrospective</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Prospective</td>
<td>22 (73)</td>
</tr>
<tr>
<td>Cohort</td>
<td>19 (63)</td>
</tr>
<tr>
<td>Descriptive Statistics</td>
<td></td>
</tr>
</tbody>
</table>
| n          | 30 (100)
| %          | 30 (100)
| Mean       | 16 (53)
| Standard Deviation | 13 (43)
| Median     | 22 (73)
| Interquartile Range  | 19 (63)
| Minimum    | 1 (3)
| Maximum    | 3 (3)
| Analyses or procedures |       |
| Type I error rate or significance level (alpha) | 14 (47)
| p values   | 27 (90)
| Confidence Intervals | 21 (70)
| T-test (or Z-test) | 13 (43)
| Shapiro-Wilk’s test | 1 (3)
| Wilcoxon Rank Sum test | 16 (53)
| Chi-squared test | 20 (67)
| Fisher’s Exact test | 17 (57)
| McNemar’s test | 1 (3)
| One-Way Analysis of Variance | 4 (13)
| Linear Regression | 1 (3)
| Linear Mixed Model | 1 (3)
| Logistic Regression | 18 (60)
| Hosmer-Lemeshow test | 5 (17)
| Poisson Regression | 1 (3)
| Log-Rank test | 3 (10)
| Cox Regression | 10 (33)
| Positive/Negative Predictive Value | 2 (7)
| Area under the curve | 3 (10)
| Sensitivity analysis | 8 (27)
| Multiple imputation adjustment | 2 (7)
| Attributable Fractions | 1 (3)
| Non-Inferiority testing | 4 (13)
| Graphics or Figures |       |
| Study Flowchart or Diagram | 17 (57)
| Pie Charts | 2 (7)
| Bar Charts | 8 (27)
| Line Charts | 2 (7)
| Kaplan Meier Curves | 6 (20)
| ROC Curve | 6 (20)

The most common figure was study flow diagram (n=17 [57%]). The most common graphic was a bar chart (n=8 [27%]). The majority of studies were cohort studies (n=19, [63%]). The majority of cohort studies were prospective (n=13, [68%]). The most frequent software used was SPSS (n=9, [30%]). Five studies (17%) mentioned multiple software used, and eight studies (27%) did not specify which software was used for analysis.

Six statistical tests occurred in more than 30% of the articles reviewed. These tests were: Chi-squared tests, Fisher’s exact tests, t-tests, Wilcoxon rank sum tests, logistic regression, and Cox Proportional Hazards regression. Curriculum session are depicted in Table 2.

A glossary defining the most common statistical tests is depicted in Table 3.
The most common descriptive statistics for continuous data were medians and IQRs. The most common statistical tests can be split into two types: comparing study sample characteristics, and comparing study sample outcomes. Concerning study sample characteristics, the most common tests for categorical data were chi-squared tests, or Fisher’s exact tests. The most common test for comparing continuous data between two study groups was the Wilcoxon rank sum test.

Tests comparing outcomes were led primarily by study design. In cohort studies, logistic regression was the most common test for outcome analysis. In randomized controlled trials, analyses were much more varied, with the most common test for outcome analysis being Cox Proportional Hazards Regression analysis.

Sessions of curriculum were chosen based on the application of tests in clinical research. Our data suggests a minimum requirement of four sessions for a basic core statistical curriculum. The t-test and Wilcoxon rank sum test are both used to compare continuous data between two study groups, so these were combined in the curriculum. Likewise, the chi-squared test and Fisher’s exact test both compare categorical data between study groups, so these were also combined. While the curriculum will contain more detail on each analytical approach, the glossary provied in Table 3 gives a fact sheet for clinical investigators on the kinds of tests they should expect to see within clinical research in the field of CAP.

An important limitation is that we concentrated only on clinical research, which is a single topic within the large field of CAP. In order to develop a curriculum for generalized clinical research, a more comprehensive review of clinical studies should be performed. Furthermore, we did not evaluate whether or not each statistical methodology or study design was appropriate for each hypothesis tested. It is possible that other methodologies may have been better suited to answer each of the questions posed by the investigators.

Finally, as articles were chosen by vote from a Delphi panel, it is possible that there may be bias in our results, or that our results may not be replicated by other researchers.

In conclusion, we identified the most common clinical research tests performed in studies of hospitalized patients with CAP. Junior investigators should become very familiar with these tests early in their research careers.

### References


