Using Big Data to Develop the Epidemiology of Orthopedic Trauma

Yi-Hui Lee1,2 and Nan-Ping Yang3,4*

1Department of Nursing, College of Medicine, National Taiwan University, Taipei, Taiwan, ROC
2Department of Nursing, School of Nursing, College of Medicine, Chang-Gang University, Taoyuan, Taiwan, ROC
3Department of Orthopedic Surgery, Keelung Hospital, Ministry of Health and Welfare, Keelung, Taiwan, ROC
4Institute of Public Health, Medical College, National Yang-Ming University, Taipei, Taiwan, ROC

Abstract

In many countries, a long-term registered electronic medical database has been established by public authorities, which could be used to investigate the epidemiology of orthopedic injuries or degenerative disorders. Three basic epidemiological study designs could be applied to a big-data for healthcare to study orthopedic medicine: cross-sectional study, retrospective case-control study, and retrospective cohort study. From the viewpoint of epidemiology, nationwide descriptive estimations of orthopedic injuries or disorders presented as the point prevalence (or incidence) with the 95% confidence interval and subsequent analysis of risk factors (or prognostic factors) evaluated by various statistic methods are valuable to policymakers when considering the redistribution of medical and social resources.

Keywords: Epidemiology; Big data; Orthopedics

Short Communication

In 2012, the US government unveiled the “Big Data” initiative, with $200 million committed to research across several agencies [1]. Integrating “big data” science into the practice of epidemiology has been recommended because the unquestionable reality of 21st century epidemiology is a tsunami of data spanning the spectrum of genomic, molecular, clinical, epidemiologic, environmental, and digital information [2]. In many countries, a long-term registered electronic medical database has been established by public authorities, which could be used to investigate the descriptive epidemiology of orthopedics, such as orthopedic injuries or degenerative disorders, in order to analyze the potential risk factors of these orthopedic issues and to predict their future consequences (which could be thought of as analytic epidemiology of orthopedics).

The number of students and disciplines requiring basic instruction in epidemiologic methods is growing. Many approaches to teaching epidemiology generally start with labeling key concepts and then move on to explain them [3]. It is important for an orthopedic researcher to learn some modern biostatistics to deal with nationwide big data. For example, three basic epidemiological study designs can be applied to an existing big-data for healthcare to study orthopedic medicine: cross-sectional study [4-9], retrospective case-control study [10-12], and retrospective cohort study [13-15]. Methods to estimate the prevalence or incidence of some orthopedic disorders or injuries, to evaluate their associated factors, to select other comparable groups and to use some suitable statistical models are shown in Table 1.

Based on our restricted experience, the descriptive epidemiology of orthopedic trauma could use big data such as a nationwide health insurance databank to calculate prevalences [4,6,7] or incidences [5,8-10] with 95% confidence intervals (CIs), and analytic epidemiology of orthopedics could be also performed. All the enrolled cases could be divided into several subgroups for comparison [11,12,14,15], but a control group may be randomly selected from the other normal population by a matched method [13]. In a retrospective data analysis, the odds ratio (OR) with the 95% CI is evaluated [11-13], usually by a binominal logistic regression (LR) model or an innovative multi-level logistic regression (MLLR) model [11,12]. In a retrospective cohort data analysis, the relative risk (RR) with the 95% CI is also evaluated [10,13-15], usually by a traditional survival analysis method, although an unconventional Poisson regression model [14,15] can be used in some specific situations. A long-term registered database could be used to observe trends in some orthopedic issues, and furthermore, the period effect could be evaluated [15].

Some criticism still focuses on the limited information contained in epidemiologic methods.

Table 1: Descriptive or analytic methods applied to orthopedics-related big data.

<table>
<thead>
<tr>
<th>Suitable statistic models</th>
<th>Cross-sectional study</th>
<th>Retrospective case-control study</th>
<th>Cohort study*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression (LR)†</td>
<td>Logistic Regression (LR)†</td>
<td>Multi-level LR (MLLR)</td>
<td>Cox Regression Poisson Regression</td>
</tr>
<tr>
<td>Logistic Regression (MLLR)</td>
<td>Logistic Regression (MLLR)</td>
<td>Multi-level Logistic Regression (MLLR)</td>
<td>Multi-level Logistic Regression (MLLR)</td>
</tr>
</tbody>
</table>

*: A retrospective cohort study could be designed based on an existing big-data. ICD: International Statistic Classification of Diseases
CI: Confidence Interval
OR: Odds Ratio; RR: Relative Risk
†: Conditional logistic regression (CLR) should be suitable if the matched controls were used.

Received December 20, 2014; Accepted January 21, 2015; Published January 23, 2015


Copyright: © 2015 Lee Y, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
in these registered medical big datasets and their accuracy. From the viewpoint of epidemiology, nationwide descriptive estimations of orthopedic injuries or disorders presented as point prevalence (or incidence) with the 95% CI and subsequent analysis of risk factors (or prognostic factors) evaluated by various statistical methods are valuable to policymakers when considering the redistribution of medical and social resources.

References