

1. Introduction

Disease trajectory refers to the probable pathway of a patient with a given condition may have, taking into account the sequence of diagnoses over time, including the development of symptoms and possible complications [1]. This can include factors such as the severity of symptoms, the rate of progression, and the likelihood of recovery or recurrence. Understanding a disease trajectory is crucial for both diagnosis and treatment, as it provides insight into how a disease progresses, the symptoms it presents, and potential complications that may arise.

In recent years, research on disease trajectory has increasingly relied on data from Electronic Health Records (EHRs) and other sources to better understand the natural history of diseases and identify potential targets for intervention [2]. A real-world healthcare dataset is usually difficult to access due to its sensitivity [3]. But the Indonesian Health Insurance (BPJS Kesehatan) has provided real-world data samples of participants of BPJS Kesehatan from 2015-2018. The dataset covers membership data and the electronic health records service of participants acquired from BPJS Kesehatan data warehouse [4]. One of the data sample sub-set data is Advanced Referral Health Facilities Service Secondary Diagnoses.

Secondary diagnoses refer to additional diagnoses that a patient may have in addition to their primary diagnosis. These additional diagnoses include complications that arose during their hospital stay and comorbidities [5]. Secondary diagnoses can have an impact on the patient's treatment, recovery, and prognosis. They also can affect the cost of the treatment and the length of the hospital stay. According to the CMS ICD-10-CM Official Guidelines for Coding and Reporting [6], secondary diagnoses were judged by clinicians when the diagnosis process happens and are coded as 'additional codes' for the primary diagnosis with no sequencing or time factor applied to it.

Disease trajectory and secondary diagnoses can greatly impact patient outcomes, treatment, and management. By analyzing EHR data from patients with a specific patient group, we can identify patterns that may indicate an increased risk of developing certain secondary diagnoses. By identifying these patterns, we can take proactive measures to prevent or manage these secondary diagnoses. In this paper, we analyze the BPJS Kesehatan data sample for disease trajectory analysis using Process Mining [7], which can be utilized to make disease trajectory based on patients' diagnoses [8], creating a Data-Driven model and in turn able to support an Evidence-Based business process. The literature review revealed that recent studies have shown the potential of process mining in the analysis of disease trajectory [2, 9, 10]. Other studies have also shown that the BPJS Kesehatan data sample is viable to be used in disease trajectory analysis [11]. While some studies, like the ones that have been mention previously, have explored the use of process mining in healthcare, most have focused on primary diagnoses data. Since a patient have the possibility of having multiple secondary diagnoses on one primary diagnoses, the EHR will show multiple diagnoses on the same timestamp. And with no established technique to analyze these secondary diagnoses, there is a gap in the literature regarding the use of secondary diagnoses data in disease trajectory analysis.

In this study, we analyzed the BPJS Kesehatan data samples for the feasibility of disease trajectory analysis with insight of secondary diagnoses, using process mining as our primary approach. This study followed the established process mining project methodology (PM²) [12] and focused on specific diagnoses for our analysis. By analyzing event logs from EHRs, process mining can uncover hidden patterns that connect primary diagnoses with secondary diagnoses. By combining this information with data on disease trajectory and secondary diagnoses, we can gain a more comprehensive understanding of how patients' health changes over time and identify potential risk factors for developing additional medical conditions.