A. Ugon et al. (Eds.)

doi:10.3233/978-1-61499-852-5-376

Process Mining in Primary Care: A Literature Review

Richard WILLIAMS[a,b,1], Eric ROJAS[c], Niels PEEK[a,b] and Owen A JOHNSON[d]
[a] *MRC Health eResearch Centre, Division of Informatics, Imaging and Data Science, University of Manchester, Manchester, UK*
[b] *NIHR Greater Manchester Patient Safety Translational Research Centre, University of Manchester, Manchester, UK*
[c] *Computer Science Department, School of Engineering, Pontificia Universidad Católica de Chile, Chile*
[d] *School of Computing, University of Leeds, Leeds, UK*

Abstract. Process mining is the discipline of discovering processes from event logs, checking the conformance of real world events to idealized processes, and ultimately finding ways to improve those processes. It was originally applied to business processes and has recently been applied to healthcare. It can reveal insights into clinical care pathways and inform the redesign of healthcare services. We reviewed the literature on process mining, to investigate the extent to which process mining has been applied to primary care, and to identify specific challenges that may arise in this setting. We identified 143 relevant papers, of which only a small minority (n=7) focused on primary care settings. Reported challenges included data quality (consistency and completeness of routinely collected data); selection of appropriate algorithms and tools; presentation of results; and utilization of results in real-world applications.

Keywords. Process mining; workflow; primary care; care pathways.

1. Introduction

Healthcare systems worldwide are trying to reduce costs by moving as much care as possible out of the hospital environment into other settings such as primary care. Primary care covers many aspects of healthcare, in a generalist manner, with a particular focus on the management of chronic conditions. In countries such as the UK where primary care plays a key role in the delivery of healthcare, patients are enrolled with a local, community-based general practitioner (GP) who is the first point of contact for non-emergency healthcare needs. Typically GPs act as the "gatekeeper" for referral to specialist care services and have responsibility for managing the lifelong care of the patient. Primary care is less structured than in-hospital care: patients are not physically present for the duration of their care pathway; care frequently transfers between healthcare settings and providers; and patients have a greater responsibility for the self-management and treatment of their conditions.

[1] Corresponding Author, Richard Williams, G.303a, Jean McFarlane Building, University of Manchester, M13 9PL; E-mail: richard.williams2@manchester.ac.uk.

Process mining, also called "process discovery" or "workflow mining", is the discipline of discovering processes from event logs, checking the conformance of real world events to idealized processes, and ultimately finding ways to improve those processes. It was originally applied to business processes over 20 years ago [1], and more recently has been applied to the healthcare domain. There are examples in secondary care [2, 3], tertiary care [4] and dentistry [5], but there would appear to be little published work in community or primary care.

Three recent literature reviews [6–8] related to process mining in healthcare have variously reported on: the volume of research over time; the algorithms and techniques used; the tools and software used; the geographical distribution of datasets and the medical domains studied. None of the reviews have specifically focused on the healthcare setting in which process mining was applied.

Therefore the objectives of this paper are: (1) to review the scientific literature on process mining in healthcare as it relates to community-based and primary care, (2) to summarize the data sources, geographical location and medical domains that were reported, and (3) identify challenges that may appear when applying process mining in primary care.

2. Method

We aimed to review articles written in English that are related to the application of process mining within healthcare and describe the use of a real world data source i.e. not simulated data or methods presented without data. A previous literature review [6] executed on 8th February 2016 used similar criteria, with the exception that they included papers with methods but no data. We therefore restricted our search to articles published since February 2016 until the present and included all the articles in the previous review that had a real world data source. Other literature reviews were considered out of scope on the grounds that they didn't contain a data source.

Following [6], the databases searched were PubMed, dblp and Google Scholar. The Google Scholar searches were performed in an incognito mode to remove the effects of previous browser history. The searches were performed by the lead author on 5th October 2017 using the query: *("process mining" OR "workflow mining") AND healthcare*. Due to the domain specific nature of PubMed the "healthcare" part of the query was omitted for this database. We were careful to reproduce the same search strategies as [6] and this review also included the list of papers from the main process mining research community website at http://www.processmining.org/ so we also looked for papers here.

In total 579 papers were found after removing 31 duplicates. One was the previous literature review, 380 were excluded based on the title and 95 based on the abstract, leaving 103 to read in full. At this stage 73 of the 74 papers from the previous literature review were added after one duplicate had been removed. From the 176 papers read in full, 33 were excluded. For papers excluded at any stage, the most commonly used exclusion criteria were: not about healthcare (n=382); not process mining (n=67); and no data – just methods, discussion, literature review or simulation (n=24). This review is based on the 143 papers read, after duplicates and exclusions removed (see Figure 1).

The primary focus for data extraction was the healthcare setting of the datasets used for the process mining. The precise classification was deliberately left open ended to allow for unexpected domains, but where the dataset contained one or more of the primary, secondary and tertiary care settings, we aimed to classify as either: "hospital" to include any inpatient or outpatient care from a secondary or tertiary care unit but without any primary or community based care data; or "primary" to include any dataset that contains, or may contain, some primary care data. We also recorded the country and medical domain of the datasets used.

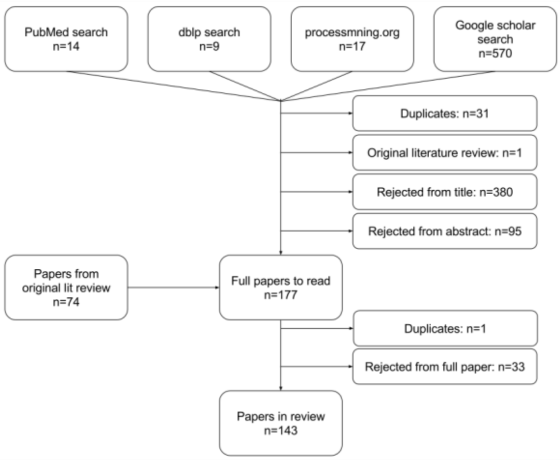

Figure 1. The full workflow of papers screened for this review.

3. Results

Datasets from hospitals were used in 91% (n=130) of papers, far more than the number of papers that used, or may have used, primary care data (n=7). Additional domains that were found during data extraction were dentistry (n=4), public health (n=1) and nursing homes (n=1). Occasionally, for example with insurance datasets, it was unclear whether the dataset contained primary care data. In these instances the medical domain was used to help classify so that, for example, surgery data would be assumed to occur in hospital, but chronic conditions such as type 2 diabetes were assumed likely to contain at least some primary care data.

Of the 7 papers with a dataset which may contain primary care data: 4 used datasets from insurance providers (2 using ICD codes [9, 10], 1 using Belgian insurance codes [11] and 1 unspecified [12]), 1 had limited information about their dataset [13], 1 mentioned preliminary results but didn't actually present them [14], and 1 had primary and secondary data for type 2 diabetes but limited results [15].

Europe (n=68) contributed the largest number of papers, though at 48% of papers it was less dominant than at the time of the previous literature review when it accounted for 73% of papers. North America (n=31) and Asia (n=22) have increased their share, while work has also appeared in South America (n=8) which was absent previously. The Netherlands (n=35) remain the country with the most papers, but second and third are now USA (n=25) and Australia (n=8); previously it was Germany and Belgium.

Oncology (n=33) is the most prevalent area, then cardiology (n=13), emergency care (n=11), stroke (n=10), surgery (n=8), diabetes (n=6), asthma (n=6) and others (n=61).

Many study challenges were identified by the authors of the reviewed papers. These include: data quality and how to assess or correct for consistency and completeness of routinely collected data; which of several competing process mining tools to use; which of an increasing number of algorithms to consider; how to validate the results; how to give insight into the discovered processes either by improved visualizations or comprehendible models; and how to utilize the results in a clinical setting. No specific primary care challenges were mentioned in the reviewed papers, however all of these are likely to be present in primary care.

The full list of the 143 reviewed papers and the data extracted is omitted due to the restrictions on paper length. However this information is available at https://zenodo.org/badge/latestdoi/110376986.

4. Discussion

Our review demonstrates there is little research in the area of process mining within primary care. Of the limited research, none is done exclusively in primary care.

The relevance, remit and extent of primary care varies from country to country. However, primary care plays a key role in most of the countries that we identified with at least 4 papers on process mining; only China, USA and Germany don't require registration with a GP, or use primary care as the gatekeeper of healthcare services [16].

Acute care pathways within secondary care, where the patient is physically located within the hospital, have tight and well defined boundaries – you can monitor, interact with, and record info on the patient for the entire duration of the pathway. This is also true to some extent in outpatient settings for disease specific processes, such as cancer, when managed within specialist tertiary centres. In such cases there is a tight boundary in that all aspects of treatment are within, and recorded within, the centre. It is perhaps therefore unsurprising that these domains are popular with process mining, especially to researchers interested in method development looking for easy data sets.

Fragmentation of data may be an issue in some countries, however large primary care databases have been used for research globally with examples in USA [16], UK [17] and the Netherlands [17]. Although the boundaries of primary care are less well defined, there are still opportunities to use these data to look at processes that are exclusive to primary care. This could include various stages in chronic disease management such as monitoring, diagnosis and treatment. Medication management, and safe prescribing are other areas with potential – especially within the UK where the combination of large primary care datasets and universal electronic prescribing is particularly attractive.

The strengths of this paper are that: we have based the search on a previously published peer review, giving it increased validity; and we have explored a clinical field that is not currently well reported within the process mining community. The limitations are that: the literature search and data extraction were performed by a single author (RW) which may have introduced bias but as our intention was simply to investigate healthcare setting rather than to systematically extract more complex concepts we believe this to be sufficient; and we explicitly rejected other literature reviews that may have contained papers not found by our search, however given the breadth of our search we believe it unlikely that many papers have been missed and our results and conclusion would remain.

5. Conclusion

The lack of published papers to date suggests there are challenges to be overcome when applying process mining to primary care, so future work should look to identify and resolve these problems. There is a wealth of primary care data available for research and a big, as yet unrealized, opportunity to analyze this data with process mining.

6. Acknowledgements

Funded by the National Institute for Health Research Greater Manchester Patient Safety Translational Research Centre (NIHR GM PSTRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. NP and OJ were supported through Connected Health Cities, a Northern Health Science Alliance led programme funded by the Department of Health and delivered by a consortium of academic and NHS organizations across the north of England.

References

[1] Aldin L, de Cesare S., "A literature review on business process modelling: new frontiers of reusability," *Enterp. Inf. Syst.*, vol. 5, no. 3, pp. 359–383, Aug. 2011.
[2] Huang H, Jin T, Wang J., "Extracting Clinical-event-packages from Billing Data for Clinical Pathway Mining," Springer, Cham, 2017, pp. 19–31.
[3] Rismanchian F, Lee YH., "Process Mining–Based Method of Designing and Optimizing the Layouts of Emergency Departments in Hospitals," *Heal. Environ. Res. Des. J.*, vol. 10, no. 4, pp. 105–120, Jul. 2017.
[4] Wang T, Tian X, Yu M, Qi X, Yang L., "Stage division and pattern discovery of complex patient care processes," *J. Syst. Sci. Complex.*, vol. 30, no. 5, pp. 1136–1159, Oct. 2017.
[5] Mans R, Reijers H, van Genuchten M, Wismeijer D., "Mining processes in dentistry," in *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*, 2012, p. 379.
[6] Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D., "Process mining in healthcare: A literature review," *J. Biomed. Inform.*, vol. 61, pp. 224–236, 2016.
[7] Erdogan T, Tarhan A., "Process mining for healthcare process analytics," in *Proceedings - 26th International Workshop on Software Measurement, IWSM 2016*, 2017, pp. 125–130.
[8] Ghasemi M, Amyot D., "Process Mining in Healthcare — A Systematised Literature Review," *Int. J. Electron. Healthc.*, vol. 9, no. 110, 2016.
[9] Hilton RP, Serban N, Zheng RY., "Uncovering Longitudinal Healthcare Utilization from Patient-Level Medical Claims Data," 2016.
[10] Tóth K, Kósa I, Vathy-Fogarassy A., "Frequent Treatment Sequence Mining from Medical Databases," *Stud. Health Technol. Inform.*, vol. 236, pp. 211–218, 2017.
[11] Lismont J, Janssens AS, Odnoletkova I, vanden Broucke S, Caron F, Vanthienen J., "A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways," *Comput. Biol. Med.*, vol. 77, pp. 125–134, Oct. 2016.
[12] Lakshmanan GT, Rozsnyai S, Wang F., "Investigating Clinical Care Pathways Correlated with Outcomes," Springer, Berlin, Heidelberg, 2013, pp. 323–338.
[13] Fernandez-Llatas C, Martinez-Millana A, Martinez-Romero A, Benedi JM, Traver V., "Diabetes care related process modelling using Process Mining techniques. Lessons learned in the application of Interactive Pattern Recognition: Coping with the Spaghetti Effect," in *Conf Proc IEEE Eng Med Biol Soc.*, 2015, vol. 2015–Novem, pp. 2127–2130.
[14] Meneu T, Traver V, Guillen S, Valdivieso B, Benedi J, Fernandez-Llatas C., "Heart Cycle: Facilitating the deployment of advanced care processes," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 6996–6999, 2013.
[15] Dagliati A, Sacchi L, Cerra C, et al., "Temporal data mining and process mining techniques to identify cardiovascular risk-associated clinical pathways in Type 2 diabetes patients," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2014, pp. 240–243.
[16] Meiman J, Freund JE., "Large data sets in primary care research.," *Ann. Fam. Med.*, vol. 10, no. 5, pp. 473–4, Sep. 2012.
[17] Afonso A, Schmiedl S, Becker C, et al., "A methodological comparison of two European primary care databases and replication in a US claims database: inhaled long-acting beta-2-agonists and the risk of acute myocardial infarction," *Eur J Clin Pharmacol*, vol. 72, no. 9, pp. 1105–1116, Sep. 2016.