This is a repository copy of *Process Mining and Analytics for Care Pathways: Proposal for a Tutorial at ICHI2017*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/125723/

Version: Accepted Version

Process Mining and Analytics for Care Pathways

Proposal for a Tutorial at ICHI2017

Owen A. Johnson

Leeds MRC Biomedical Informatics Research Center, University of Leeds, Leeds, UK

o.a.johnson@leeds.ac.uk

Abstract— **This tutorial demonstrates the application of a range of process analytics techniques to the study and improvement of clinical care pathways. The past decade has seen increasing interest in care pathway design, documentation and dissemination but formal methods for describing, monitoring and assessing pathways have yet to be established. Outside of healthcare other industries have well established techniques for business processes and there is much scope for translating these to fit the unique nature of healthcare. In particular data analytics, data mining, and machine learning have converged on a set of technologies called process mining which has the potential to lead to a step-change in using e-health record data to mine and manage care pathways. The tutorial presents an iterative method developed with the UK NHS and the Connected Health Cities programme that combines process mining with other process analytics methods including process modeling, process simulation and business process improvement. The session is highly interactive and based on a series of hands-on exercises around a worked example supplemented by case studies of completed work. Links to further study are provided and the aim is encourage further research and build a global community of practice.**

Keywords— **care pathways; data analytics, process mining, business process improvement, e-health records.**

I - INTRODUCTION

Process mining is an emerging field within process analytics which involves the data mining of event logs from business information systems. The approach uses machine learning to discover and analyse process models based on real activities as they are recorded in computer systems through routine use. The resulting models can be shown graphically following notations such as Business Process Modelling Notation (BPMN). Process mining can be used for analysis such as detecting deviations, bottlenecks, early warning of adverse outcomes and sub-optimal behaviour patterns. Process mining research methods have been successfully applied to the analysis of healthcare processes with the aim of improving quality of care, patient safety, and optimization of resources.

In the UK we found process mining techniques proved useful in the analysis of highly complex and variable care pathways such as cancer when used as part of an iterative method that incorporates other process analysis techniques. In a variant of the L* method we have combined process mining with "Plain Old Business Analysis" techniques for investigating business systems and processes, modern lean and Six Sigma process improvement methods and process simulation (using a locally developed software tool). One challenge we, and other e-health record researchers, have faced is poor data quality in the large and comprehensive e-health record data now available. Business analysis methods, constant engagement with domain experts and stakeholders and deep-dives into the root cause of data anomalies has proved essential. Our methods are evolving through refinement with a series of major case studies set in the UK NHS in the North of England as part of the £20m Connecting Health Cities programme.

II – THE TUTORIAL

The tutorial is aimed at data scientists, informatics and clinical researchers working with care pathways and linked e-health record data. The aim is to introduce delegates to a set of skills that we think can be combined to form a coherent informatics methodology for exploring, understanding, explaining and improving care pathways. The objectives are i) to introduce process analytics as a powerful informatics methodology; ii) to demonstrate practical applications; iii) to teach the basic steps and signpost tools and other resources; iv) to create an informal network of informatics researchers interested in care pathway improvement. The session will be highly interactive with demonstrations, three short hands-on exercises (laptops required) and a structured discussion. We are keen to create an international network of researchers interested in care pathway data mining and build sufficient support to develop this new field within health informatics.

III - BIOGRAPHY

Owen Johnson won the UK Research Council 2014 award for setting up the ResearchOne e-health records data service, one of the largest in the world. He is one of the founders of the £20m Leeds Institute for Data Analytics, a major new UK centre for big data research. He has taught process analytics and health informatics at Leeds for 15 years. His research interests are in developing computer science tools and process analytics methods for care pathway improvement. Owen is a professional software engineer and has worked in roles ranging from developer to IT manager; he was vice chair of £60m Bradford Housing Trust and is a founder and director of X-Lab, a successful e-health software company.

Part 2 – Full Tutorial Proposal

Description: *A description of the main topics to be covered during the tutorial.*

The Tutorial will address the challenge in healthcare of care pathway development and cover four related sets of applicable methods which collectively can be described as process analytics – process modelling, process mining, process simulation and process improvement. It will join these methods together in a straightforward and usable approach, give delegates practical experience of each and provide further examples and resources for further study.

Care pathway literature and emerging repositories of good practice (for example the UK NHS NICE Pathways https://pathways.nice.org.uk) will be reviewed along with their strengths and weaknesses and the principles and aims of care pathway improvement efforts. The opportunities for data-driven and health informatics informed approaches will be introduced at the start of the tutorial and this theme will be developed throughout the tutorial with the aim of identifying the contribution that delegates can make to better health care by leveraging their existing informatics knowledge and skills.

Process modelling has been established in engineering fields since at least 1911 when Frank Gilbreth published what became the ASME notation for modelling the stages in manufacturing processes. In software engineering the Unified Modelling Language (UML) became the de facto world standard for systems modelling in the early 2000s and in business the Business Process Modelling and Notation (BPMN) has achieved similar status with both standards now managed by the Object Management Group and expected to converge. The principles of process modelling, including agile methods are similarly well established. The tutorial will highlight the importance of standards, make connections to business analysis (systems) and quantitative methods (research) approaches and take delegates quickly through an agile approach using manual document reviews, observation and sticky notes.

Process Improvement has an even longer pedigree and arguably traces its origins to the first large organisations through to FW Taylor, Henry Ford and Deming. Apparently "modern" approaches from the USA, e.g. Michael Porter's Value Chains, and the Japanese influenced Lean and Six Sigma borrow extensively from Gilbreth and Deming and a rich tradition. The tutorial links approaches to process improvement to the current interest in care pathways and rephrases the language of business into the language of healthcare opening up this rich source of methods, tools and body of theory. Two specific examples, Value Chains and Six Sigma DMAIC method are used to illustrate the connections. Delegates are encouraged to use a combination of process mining and process modelling to evaluate current "as is" care pathways and then to used process simulation to explore candidate improvements as "could be" scenarios.

Process simulation is a computer based approach to modelling and "what if" evaluation drawing on visualisation-through-animation, Monte-Carlo simulation, discrete event simulation and agent based techniques. Good simulations allow health care interventions or pathway improvements to be carefully explored in the lab (in vitro) before being introduced for real (in vivo). The tutorial makes use of NETIMIS, a software product developed by our team specifically for care pathways. Delegates will follow a quick tutorial to create their first pathway simulation using the online NETIMIS tool and then be able to review a range of pre-built care pathway examples from recent projects in the UK and elsewhere. A simple evidenced-based form to underpin the simulation will be used in the tutorial and delegates will learn how to create an "as is" pathway, calibrate that to real life practice and then use the calibrated model to explore candidate "could be" scenarios. The example of novel diagnostic devices at various points in the pathway from home, to ambulance, to point of care is used to illustrate how multiple "could be" scenarios can be quickly evaluated in terms of cost, time, patient impact and clinical outcome. Delegates will also create a simulated event log for the next exercise.

Process mining is a relatively recent set of approaches largely based on machine learning algorithms for time-series analysis of event log data and popularised by Wil van der Aalst with the ProM toolkit analysis. There is a small but growing body of literature applying process mining techniques to healthcare and specifically linking to e-health records research and "big data" analytics method. The basic approach is to detect patterns that occur frequently in e-health records that indicate that many patients follow broadly similar pathways of care. This is an area of research interest to our group and we are developing novel algorithms to identify disease trajectories, spot repeating patterns, address data quality issues (e.g. through imputation), spot poor and divergent practice and identify early warning signs that could be used to develop precision interventions. Delegates will follow a quick tutorial using a process mining tool using a de-identified health care dataset and the simulated event log they will created with NETIMIS. The range of features in process mining software will be overviewed and a number of MOOCs for further study will be identified.

Process analysis combines all of the four methods covered in the tutorial and much more that delegates may well be familiar with. Our aim is therefore to overview the methods in sufficient detail and with sufficient enthusiasm to encourage delegates to further independent study and to explore the contribution that they can make to better care pathways. We will overview the L* process mining method and reflect on our own evolving iterative methods and supplement these with examples of work already done.

Case Studies (optional): *If relevant, a description of real-world systems, experiences, and/or case studies to be discussed.*
As described above.

Audience: *What background will an attendee need to follow the tutorial (e.g., novice, expert, etc)?*
The workshop is aimed at health data scientists, informatics and clinical researchers interested in working with care pathways and linked e-health record data. Industry practitioners may also benefit but the emphasis is on emerging practical and scientific techniques rather than an established technology. No strong technical skills are required but this does need an analytical mindset, some grounding in healthcare and familiarity with using software tools as an end user. We are hoping that delegates will bring laptops and have Internet access for the two hands-on computing exercises.

Aims and objectives: *State the learning objectives and describe them as needed.*
The aim is to introduce delegates to a set of skills that can be combined to form a coherent informatics methodology for exploring, understanding, explaining and improving care pathways.
The objectives are
> i) to introduce process analytics as a powerful informatics methodology;
> ii) to demonstrate practical applications;
> iii) to teach the basic steps and signpost tools and other resources;
> iv) to create an informal network of informatics researchers interested in care pathway improvement.

Outline Timetable
Duration (120 minutes)
Introductions (20m) – to the workshop, the team, delegates and to the field of process analytics and challenges for healthcare. Readiness checks for exercises. Brief demonstration of Connected Health Cities pathways and review of the method.
Modelling a care pathway in Gout (20m) – table-top exercise using Sticky Notes and extracting pathways details from the text of guidelines from NHS Choices http://www.nhs.uk and National Institute for Health and Care Excellence (NICE) http://pathways.nice.org.uk
Creating a pathway simulation model (30m) – online exercise using the outputs from the table-top exercise and the NETIMIS simulation tool www.netimis.co.uk. Exploring pathway flow and "What If" scenarios. Review and examples.
Process Mining a Care Pathway (30 min) – a worked exercise with sample data and the Disco process mining tool www.fluxicon/disco. Process variants and impact on clinical outcomes. Review and examples.
Structured Discussion (20 min) – Examples, method, challenges, opportunities, next steps. Review of outcomes.

Other Information *Any other information (e.g., previous tutorials given) that will assist the PC Co-Chairs and Conference Leadership to adequately review the tutorial proposal.*

This tutorial was initially developed for the UK Farr Institute Informatics for Health 2017 Conference in April 2017 (www.farrinstitute.org/events-courses/event/informatics-for-health-2017) and will be delivered twice at that conference (once as part of the training day for the PhD Symposium and then repeated as part of the main conference). It is being delivered in Leeds at the Leeds Institute for Data Analytics on 6th April to 25 PhD, academic and industry delegates. Some aspects of the tutorial are drawn from existing MSc Health Informatics courses at Leeds, specifically from the Process Modelling Benefits and Change module (http://medhealth.leeds.ac.uk/info/291/cpd/1154/process_modelling_benefits_and_change_ychi5030m). The tutorial will be delivered by Owen Johnson from the University of Leeds. Owen designed the Leeds MSc in Health Informatics and has taught over 200 health informatics student in the last ten years. He has also taught Process Analysis methods in Computing, Business and Medicine.

Short CV

Owen Johnson won the UK Research Council 2014 award for setting up the ResearchOne e-health records data service, one of the largest in the world. He is one of the founders of the £20m Leeds Institute for Data Analytics, a major new UK centre for big data research. He has taught process analytics and health informatics at Leeds for 15 years. His research interests are in developing computer science tools and process analytics methods for care pathway improvement. Owen is a professional software engineer and has worked in roles ranging from developer to IT manager; he was vice chair of £60m Bradford Housing Trust and is a founder and director of X-Lab, a successful e-health software company.

Publications Relevant to the Tutorial

Baker, K., Dunwoodie, E.H., Jones, R.G., Newsham, A., **Johnson, O**., Price, C.P., Wolstenholme, J., Leal, J., McGinley, P., Twelves, C. and Hall, G., 2017. Process Mining Routinely Collected Electronic Health Records to Define Real-Life Clinical Pathways during Chemotherapy. *International Journal of Medical Informatics*.

Johnson OA, Hall PS, Hulme C, NETIMIS: Dynamic Simulation of Health Economics Outcomes Using Big Data, *PharmacoEconomics*, February 2016, Volume 34, Issue 2, pp 107-114 DOI: 10.1007/s40273-016-0384-1.

Johnson OA, Fraser HS, Wyatt JC, Walley JD. Electronic Health Records in the UK and USA. The Lancet. 2014;384:954.

Crocker T, **Johnson OA**, King S. The suitability of care pathways for integrating processes and information systems in healthcare. Transform. Gov. People, Process Policy. 2009. p. 289–301.

King SF, **Johnson OA**. VBP: An approach to modelling process variety and best practice. Inf. Softw. Technol. 2006;48:1104–14.