BMJ Open

Quantification and visualization methods of data-driven chronic care delivery pathways: protocol for a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033573
Article Type:	Protocol
Date Submitted by the Author:	11-Aug-2019
Complete List of Authors:	Siqueira do Prado, Luiza; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research Allemann, Samuel; University of Basel, Pharmaceutical Sciences Viprey, Marie; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research; Hospices Civils de Lyon, Pôle Santé Publique Schott, Anne-Marie; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research; Hospices Civils de Lyon, Pôle Santé Publique Dediu, Dan; Université Lumière Lyon 2, Laboratoire Dynamique du Langage UMR 5596 Dima, Alexandra; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research
Keywords:	clinical decision support systems, medical informatics application, data visualization, clinical pathway, delivery of health care, integrated, electronic healthcare databases

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title

Quantification and visualization methods of data-driven chronic care delivery pathways: protocol for a systematic review

Corresponding author

Luiza Siqueira do Prado

luiza.siqueira-do-prado@univ-lyon1.fr

EA 7425 HESPER
Health Services and Performance Research
Université Claude Bernard Lyon 1
Domaine Rockefeller- 2eme étage (aile CD)
8 avenue Rockefeller
69373 Lyon 8
+33 (0) 7 83 45 23 66

Authors

Luiza Siqueira do Prado (luiza.siqueira-do-prado@univ-lyon1.fr)[1],

Samuel Allemann (s.allemann@unibas.ch)[1,2],

Marie Viprey (marie.viprey@chu-lyon.fr)[1,3],

Anne-Marie Schott (anne-marie.schott-pethelaz@chu-lyon.fr)[1,3],

Dan Dediu (dan.dediu@univ-lyon2.fr)[4],

Alexandra Dima (alexandra.dima@univ-lyon1.fr)[1]

Affiliations

[1] EA 7425 Health Services and Performance Research HeSPeR – Université Claude Bernard Lyon 1, Lyon, France

[2] Pharmaceutical Care Research Group, Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland

[3] Pôle de Santé Publique, Hospices Civils de Lyon, Lyon, France

[4] Laboratoire Dynamique du Langage UMR 5596, Université Lumière Lyon 2, Lyon, France

Abstract

Introduction Chronic conditions require long periods of care and often involve repeated interactions with multiple healthcare providers. Faced with increasing illness burden and costs, healthcare systems are currently working towards integrated care to streamline these interactions and improve efficiency. To support this, one promising resource is the information on routine care delivery stored in various electronic healthcare databases (EHD). In chronic conditions, care delivery pathways (CDPs) can be constructed by linking multiple data sources and extracting time-stamped healthcare utilization events and other medical data related to individual or groups of patients over specific time periods; CDPs may provide insights into current practice and ways of improving it. Several methods

have been proposed in recent years to quantify and visualize CDPs. We present the protocol for a systematic review aiming to describe the content and development of CDP methods, to derive common recommendations for CDP construction.

Methods and analysis This protocol followed the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P). A literature search will be performed in PubMed (MEDLINE), Scopus, IEEE, CINAHL and EMBASE, without date restrictions, to review published papers reporting data-driven chronic CDPs quantification and visualization methods. We will describe them using several characteristics relevant for EHD use in long-term care, grouped into three domains: 1) clinical (what health-related events it includes and for what clinical aims?), 2) data science (how is the method developed and what data infrastructure it relies on?), and 3) behavioral (what behaviors and interactions does it promote in users and through what methods?). Data extraction will be performed via deductive content analysis using previously defined characteristics and accompanied by an inductive analysis to identify and code additional relevant features. Results will be presented in descriptive format and used to compare current CDPs and generate recommendations for future CDP development initiatives.

Ethics and dissemination Ethical approval is not required for this review. Results will be disseminated in peer-reviewed journals and conference presentations.

PROSPERO registration pending approval

Keywords clinical decision support systems; medical informatics application; data visualization; clinical pathway; delivery of health care, integrated; electronic healthcare databases.

Word count 2612

Strengths and Limitations

- While most reviews of health technology tools focus on clinical objectives and technical characteristics, we will also consider behaviours of and interactions between users to describe the selected methods.
- We will perform both deductive and inductive content analysis to fully describe the methods.
- We will focus on methods described in peer-reviewed papers and exclude conference proceedings and other types of reports, to obtain detailed validated descriptions; this may limit our access to more recent studies due to the fast-paced development in the field.
- Lack of completeness in methods descriptions may limit our ability to assess all characteristics, such as the stages of development, the involvement of stakeholders or experts prior to data acquisition and analysis.
- As this is a relatively new field of health technology, there are no guidelines for reporting and no consensus on quality criteria for the studies we will evaluate; our work will also contribute to the development of such recommendations.

Introduction

Effective delivery of integrated care is a priority for healthcare systems worldwide and has been the focus of considerable efforts in recent years, particularly in response to the increasing demands of chronic care[1,2]. Long-term conditions may require lifetime care, which may consist of multiple interactions with a variety of healthcare providers at variable time intervals[3,4]. When service delivery is fragmented, the overall effectiveness of these interactions in terms of long-term quality of life and health-related outcomes is reduced, and risk of harm is increased[5,6]. Centralizing patient information

produced by different providers in electronic healthcare databases (EHD) has the potential to help implementing new ways of service delivery to improve outcomes[7]. Several attempts have been made to link multiple data sources to generate comprehensive descriptions of patients' healthcare journeys in long-term conditions. These descriptions are produced by constructing longitudinal trajectories from various time-stamped healthcare utilization events and related medical data[8–13]. However, generating informative trajectories from disparate and often incompatible data sources proves challenging[14]. As various initiatives have been developed independently, with distinct methodologies and objectives, it is essential to examine systematically the proposed solutions in order to derive principles of action to stimulate convergence of methods.

In the context of chronic conditions, the way patient trajectories are established may be subject to multiple influences and analyzing routine care data can provide insights on how they have been drawn over time and their potential sources of variation[15,16]. In the literature, trajectories within healthcare systems have been described using many terms, which makes it challenging to build consensus on terminology and practical meaning[17,18]. We will use the term data-driven 'care delivery pathway' (CDP) to group several terms we will find in the selected studies to designate retrospective trajectories obtained from EHD. To describe the methods proposed for quantifying and visualizing chronic CDP, we will assess how they addressed three domains:

1) The selection of relevant clinical and health-related events.

This domain will examine how the methods define health status and evaluate disease progression or stabilization, and how they show transitions between health status and acute manifestations[13]. Usually, the trajectory timeline begins at diagnosis and involves more than one provider[13,16,19]. Treatment decisions are generally based on health status (indicated by biomarkers, clinical examination, self-declared levels of quality of life, etc.), care units and settings, treatment availability (medication, procedures, etc.), and patient-provider preferences[15].

2) The technological development itself and considering issues related to data quality and exchange.

This domain aims to describe how the method is built, which data sources and analyses are used, and the necessary infrastructure surrounding its implementation. Digitalization of health-related data is a global trend[20,21] and highly detailed data are being collected daily in diverse settings and healthcare services. Such methods may apply a range of techniques from basic algorithms to advanced statistical and machine learning models[22], which can provide useful insights into care delivery processes. Technological developments in this field also need to meet strict criteria of data security, accuracy of models and predictions, openness of development and validation processes, among others[19,23].

3) Considering behaviours of actors and interactions between them with the aim of effectively improving care delivery.

Integrated care depends on multiple actions and decisions made collaboratively by patients, healthcare providers, administrative staff and other actors concerning patients' course of treatment[24]. To inform these decisions, technological solutions must have access to clinical exams and provide key actors with relevant information, such as the patients' past interactions with other providers, the medical procedures performed, the medications prescribed[25]. To have a positive impact on improving care delivery, visualizations and quantitative indicators of the patient's prior care need to be adapted to the user's needs at specific points in the trajectory, like after acute events or hospitalizations. This domain will examine what behaviors and interactions the tools promote

(who are its target individuals, what actions need to be performed, in what context, when, and by whom)[26], and what strategies are proposed to encourage this performance.

Aims and objectives

We propose a systematic review of the methods to quantify and visualize data-driven chronic CDP. Given the complexity of their context of use, more than only reviewing technical methods, we aim to investigate how these tools have considered the three domains described above: how they have considered relevant clinical aspects, how they have addressed key technical challenges, and what behaviours and interactions they promote or facilitate between different providers in the context of chronic conditions and how. We will mix deductive and inductive content analysis[27] to appraise the selected studies: deductive when relying on pre-defined frameworks such as the categories previously described by Moreno-Conde et al. (2015)[28] (to describe the technical characteristics of the proposed solutions) and on the TACTA[26] (Target, Action, Context, Time, Actor, to describe the behavioural domain) and inductive when additional relevant characteristics need to be described.

For this end, we propose the following research questions:

Primary research question

1. What methods have been proposed to quantify and/or visualize data-driven CPDs of people living with chronic conditions?

Secondary research questions

2. What are the clinical aims of the method and what type of clinical information does it use?
3. How was the method developed and what data infrastructure does it relies on?
4. Which behaviours and interactions do they aim to promote among users and how?

Methods

The Cochrane Handbook[29] and the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P)[30] were used to write this protocol and the systematic review will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[31]. The review will be performed by one primary reviewer (LSP) and three secondary reviewers (AD, MV and SA) and will follow 6 steps: literature search, records screening and pre-selection (title and abstract), full-text screening and final selection, extraction of data, quality assessment, analysis and synthesis of data.

The studies expected to be analyzed in this work will likely be descriptive and not follow standard methodology (i.e., experimental or observational, method validation), yet considering the manuscripts as a qualitative corpus allows for coding the narratives according to the conceptual structure we propose[7,28]. Content analysis has been used in many studies in health sciences[27] and an inductive content analysis applied in a systematic review of clinical information modeling processes[28] has developed descriptive categories in a context similar to the one we propose here. As we consider them relevant to the studies we will review, they will be included in our coding framework, as detailed below.

We will consider CDPs to be a series of time-stamped events describing the sequence of care of users with a diagnosed chronic condition (conditions requiring medical attention for a period longer than 12 months). These events can be the diagnosis itself, routine, non-scheduled or emergency consultations with a general practitioner and/or specialist, therapeutic education sessions and other

health-related interventions. These can result in prescriptions of medications, medical procedures and tests, which may also appear in the trajectory.

Searches

A literature search will be performed in the following electronic databases: PubMed (MEDLINE), Scopus, IEEE, CINAHL and EMBASE. The search will be adapted to each database, and the resulting search strategies are provided as supplementary material. The terms searched will be related to three main categories, connected by the AND operator: “data-driven” (MeSH terms like “Electronic health record”, “data mining”, etc.), “clinical pathways” (MeSH terms like “clinical pathway”, “disease management”, etc.), and “chronic conditions” (MeSH term “chronic diseases”). Searches will be performed with MeSH terms or with keywords in Title/Abstract in PubMed; MeSH terms will be adapted for the databases that do not permit their usage or use different indexed terms. Bibliographies and citation tracking of relevant literature will be hand searched to identify additional relevant studies. A first selection will be performed using abstracts and titles, followed by full-text examination of entries selected.

Types of publications/studies and eligibility criteria

We will consider peer-reviewed publications (1) reporting methods for visualization or quantification of data-driven chronic CDP (including protocols and reports of study results), (2) using data from people living with chronic conditions retrieved from EHD and (3) published in English. No restrictions on publication date, study design, population characteristics, type of healthcare facility and level of care will be applied.

Data-driven CDP analyzed here will need to be composed of at least two time-stamped events recorded in EHD from people with the diagnosis of a chronic condition, with no duration restrictions (e.g., CDP may cover periods from days or few months to several years).

We will exclude studies that do not mention population or data characteristics or do not state they analyze data from people living with chronic conditions, papers with full-text not written in English, conference abstracts, systematic or narrative reviews, meta-analyses and grey literature.

Screening

We will use Covidence, an online systematic review management software, for records screening. After duplicates removal, titles and abstracts in the remaining records will be screened independently by two reviewers for full text appraisal. If reviewer discordance arises, consensus will be reached through discussion and arbitration with one of the secondary reviewers not involved in the selection of the record. Studies selected in the first step will go through full text screening using the same process to establish eligibility. Inter-rater reliability (Cohen’s Kappa) between primary and secondary reviewers will be computed after title and abstract and full text screening and reported.

Data management

We will report the number of included and excluded articles as well as the number of full-text papers obtained and assessed. Reasons for exclusion of screened full-text studies will also be stated in the final review. The data will be managed using Covidence and Microsoft Excel spreadsheets.

Data extraction

Data from included studies will be extracted using a customized electronic data extraction form. Information on study characteristics (authors, title, type of study, year and country of study,

objective and research questions); population characteristics (number of patients, age, gender, condition) will be extracted directly from the included studies.

Deductive-Inductive content analysis

We will perform a deductive content analysis following existing theories, as described below, and inductive analysis for observed relevant characteristics not yet covered by existing literature. If more than one selected record describe development, validation and/or implementation of the same method, we will extract basic paper characteristics, as described above, but the content analysis will be performed per method.

1) For the clinical domain, we will extract information on clinical or cost outcomes the studies might target (if reported and which ones) and on how the outcomes were considered relevant (e.g., involving experts, final users or other stakeholders).

2) For method development and data processing, we will analyze and compare to what has been proposed by Moreno-Conde et al. (2015)[28]. The categories detailed in the study are described below.

- Scope definition leading to selection of the domain and selecting relevant experts
- Analysis of the information covered in the specific domain
- Design of the tool
- Definition of implementable tool specifications
- Validation
- Publishing and maintenance
- Governance

Other information extracted from studies regarding this domain will be healthcare utilization characteristics (type of event, e.g., consultation, test, procedure) and data characteristics (sources of data, data preparation, data analysis).

3) To describe behaviour and interactions the method might promote or facilitate, we will apply the TACTA[26] (Target, Action, Context, Time, Actor) framework. Other information extracted from studies will be output characteristics like intended final users, purpose and use scenarios. We will also code the presence of strategies planned or performed to achieve these behavioural change objectives, such as training, organizational changes, evaluation of the performance of the method in routine care, if implemented, and other initiatives studies might present.

The primary reviewer and one secondary reviewer will pilot data extraction independently for a subset of 10% of selected records to compare and discuss data extraction process. If necessary, we will repeat the pilot extraction process (outlined above) until agreement is reached. Disagreements will be solved with the help of a third reviewer and piloting may consist of several interactions between reviewers to compare and reach consensus regarding relevant information to be extracted from full-text analysis. After this first step, a codebook will be developed, and data extraction of the remaining records will be performed by the primary reviewer.

Quality and bias assessment

As most quality assessment tools are developed for commonly-used study designs and there is no consensus regarding tools for generic use, we propose to evaluate quality from a different perspective. We will evaluate if main stakeholders (patients and/or family, healthcare professionals, administrative personnel) were involved at any stage of the development of the method. Research shows the importance of involving patients, the public and other stakeholders in health-related research to obtain experiential knowledge, setting research priorities and focus on practical

questions[32–35]. Also, it has been shown that trials funded by for-profit organizations can positively bias interpretation of trial results[36], and research in data usage can be funded by companies interested in selling their own methods. To assess potential bias, we will evaluate declared conflicts of interest and sources of funding. Quality assessment will be discussed in the review, but no study will be excluded from the analysis based on quality criteria.

Data analysis and synthesis

The technical methods will be synthesized using the content analysis described above and the studies will be categorized and described using the 3 domains, depending on study type and reporting. We will present it in tables along with study identification. We will compare the different characteristics within the 3 domains to identify common, infrequent, or missing features of these tools, and extract recommendations for future initiatives.

Patient and public involvement

A representative of a patients' association was involved in reading and approving of this protocol. This protocol of a systematic review is part of a larger project that will be developed closely with patients and healthcare providers.

Ethics and dissemination

Ethical approval is not required. Results will be disseminated in peer-reviewed journals and/or conference presentations. Data used in this review will be made available through supplementary materials and open trusted repositories.

Funding

DD was supported by an IDEXLYON (16-IDEX-0005) Fellowship grant (2018-2021), LSP was supported by a PhD funding within the same grant, AD by a Marie Curie Individual Fellowship from the European Commission (MCRA-IF n°706028) during the preparation of this review protocol, and SA by the Swiss Science Foundation (P2BSP3_ 178648).

Conflicts of interest

None declared.

Author statement

LSP, AD and SA designed the protocol and planned data extraction and quality assessment. LSP put together the search strategy and SA helped adapt it to the different databases. LSP and AD conceived the content analysis stages and conceptual framework. LSP wrote the first version of the manuscript, AD extensively reviewed it, SA, DD, AMS and MV revised it critically for important intellectual content. All authors have approved the publication of this protocol and contributed to the final manuscript.

Acknowledgements

The authors thank IDEXLYON with the University of Lyon and the Swiss Science Foundation for financially supporting this research.

References

1. Bodenheimer, T., Wagner, E. H. & Grumbach, K. Improving Primary Care for Patients With Chronic Illness: The Chronic Care Model, Part 2. *JAMA* **288**, 1909–1914 (2002).

2. Ouwens, M., Wollersheim, H., Hermens, R., Hulscher, M. & Grol, R. Integrated care programmes for chronically ill patients: a review of systematic reviews. *Int J Qual Health Care* **17**, 141–146 (2005).
3. Bodenheimer, T., Wagner, E. H. & Grumbach, K. Improving Primary Care for Patients With Chronic Illness. *JAMA* **288**, 1775–1779 (2002).
4. Pham, H. H., Schrag, D., O'Malley, A. S., Wu, B. & Bach, P. B. Care Patterns in Medicare and Their Implications for Pay for Performance. *http://dx.doi.org.docelec.univ-lyon1.fr/10.1056/NEJMsa063979* (2009). doi:10.1056/NEJMsa063979
5. Wagner, E. H. Meeting the needs of chronically ill people: Socioeconomic factors, disabilities, and comorbid conditions are obstacles. *BMJ* **323**, 945–946 (2001).
6. Campbell, H., Hotchkiss, R., Bradshaw, N. & Porteous, M. Integrated care pathways. *BMJ* **316**, 133–137 (1998).
7. Gooch, P. & Roudsari, A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *J Am Med Inform Assoc* **18**, 738–748 (2011).
8. Huang, Z. *et al.* Discovery of clinical pathway patterns from event logs using probabilistic topic models. *Journal of Biomedical Informatics* **47**, 39–57 (2014).
9. Perer, A., Wang, F. & Hu, J. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics* **56**, 369–378 (2015).
10. Gotz, D., Wang, F. & Perer, A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *Journal of Biomedical Informatics* **48**, 148–159 (2014).
11. Huang, Z., Lu, X., Duan, H. & Fan, W. Summarizing clinical pathways from event logs. *Journal of Biomedical Informatics* **46**, 111–127 (2013).
12. Yang, W. & Su, Q. Process mining for clinical pathway: Literature review and future directions. in *2014 11th International Conference on Service Systems and Service Management (ICSSSM)* 1–5 (2014). doi:10.1109/ICSSSM.2014.6943412
13. Jensen, A. B. *et al.* Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature Communications* **5**, 4022 (2014).
14. Denaxas, S. C. & Morley, K. I. Big biomedical data and cardiovascular disease research: opportunities and challenges. *Eur Heart J Qual Care Clin Outcomes* **1**, 9–16 (2015).
15. Vanasse, A., Courteau, M. & Ethier, J.-F. The '6W' multidimensional model of care trajectories for patients with chronic ambulatory care sensitive conditions and hospital readmissions. *Public Health* **157**, 53–61 (2018).
16. Zhang, Y., Padman, R. & Patel, N. Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. *J Biomed Inform* **58**, 186–197 (2015).
17. Bleser, L. D. *et al.* Defining pathways. *Journal of Nursing Management* **14**, 553–563 (2006).
18. Kinsman, L., Rotter, T., James, E., Snow, P. & Willis, J. What is a clinical pathway? Development of a definition to inform the debate. *BMC Medicine* **8**, 31 (2010).
19. Bettencourt-Silva, J. H. *et al.* Building Data-Driven Pathways From Routinely Collected Hospital Data: A Case Study on Prostate Cancer. *JMIR Med Inform* **3**, e26 (2015).
20. Adler-Milstein, J. *et al.* Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. *Health Affairs* **34**, 2174–2180 (2015).
21. Ben-Assuli, O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. *Health Policy* **119**, 287–297 (2015).
22. Zhang, Y. & Padman, R. Innovations in chronic care delivery using data-driven clinical pathways. *Am J Manag Care* **21**, e661-668 (2015).
23. Liaw, S. T. *et al.* Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int J Med Inform* **82**, 10–24 (2013).
24. Peikes, D., Chen, A., Schore, J. & Brown, R. Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures Among Medicare Beneficiaries: 15 Randomized Trials. *JAMA* **301**, 603–618 (2009).

25. Stille, C. J., Jerant, A., Bell, D., Meltzer, D. & Elmore, J. G. Coordinating care across diseases, settings, and clinicians: a key role for the generalist in practice. *Ann. Intern. Med.* **142**, 700–708 (2005).
26. Francis, J. J. & Presseau, J. Health Care Practitioner Behaviour. in *Cambridge Handbook of Psychology, Health and Medicine* (eds. Llewellyn, C. et al.) 324–324 (Cambridge University Press, 2019).
27. Elo, S. & Kyngäs, H. The qualitative content analysis process. *Journal of Advanced Nursing* **62**, 107–115 (2008).
28. Moreno-Conde, A. *et al.* Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. *J Am Med Inform Assoc* **22**, 925–934 (2015).
29. Cochrane Handbook for Systematic Reviews of Interventions. Available at: http://handbook-5-1.cochrane.org/. (Accessed: 27th June 2019)
30. Moher, D. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* **4**, 1 (2015).
31. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **6**, e1000097 (2009).
32. Staley, K. 'Is it worth doing?' Measuring the impact of patient and public involvement in research. *Research Involvement and Engagement* **1**, 6 (2015).
33. Hoffman, A., Montgomery, R., Aubry, W. & Tunis, S. R. How Best To Engage Patients, Doctors, And Other Stakeholders In Designing Comparative Effectiveness Studies. *Health Affairs* **29**, 1834–1841 (2010).
34. Selby, J. V., Beal, A. C. & Frank, L. The Patient-Centered Outcomes Research Institute (PCORI) National Priorities for Research and Initial Research Agenda. *JAMA* **307**, 1583–1584 (2012).
35. Ocloo, J. & Matthews, R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf* **25**, 626–632 (2016).
36. Als-Nielsen, B., Chen, W., Gluud, C. & Kjaergard, L. L. Association of Funding and Conclusions in Randomized Drug Trials: A Reflection of Treatment Effect or Adverse Events? *JAMA* **290**, 921–928 (2003).

COMPLETE SEARCH STRATEGY

Category	Medical Subject headings (MeSH)	CINAHL Search	EMBASE	Keywords
Data-driven	Electronic health record; data mining; machine learning; clinical decision support systems; analysis, cluster; medical informatics application	MH Electronic Health Records OR MH Data Mining OR MH Nursing Informatics OR MH Machine Learning OR MH Decision Support Systems, Clinical OR MH Cluster Analysis OR MH Medical Informatics OR MH Computer Graphics OR MH Algorithms OR TI Data-driven OR TI visualisation OR TI computer graphics OR TI process mining OR TI data mining OR TI visualization OR TI supervised learning OR TI unsupervised learning OR TI practice based OR TI modelling OR TI mapping OR TI cluster* OR TI data analys* OR AB Data-driven OR AB visualisation OR AB computer graphics OR AB process mining OR AB data mining OR AB visualization OR AB supervised learning	(Electronic health record or data mining or machine learning or clinical decision support systems or analysis, cluster or medical informatics application).sh. or (Data-driven or visualisation or computer graphics or process mining or data mining or visualization or supervised learning or unsupervised learning or practice based or modelling or mapping or cluster* or data analys*).ti. or (Data-driven or visualisation or computer graphics or process mining or data mining or visualization or supervised learning or unsupervised learning or practice based or modelling or mapping or cluster* or data analys*).ab.	Data-driven OR visualisation OR computer graphics OR process mining OR data mining OR visualization OR supervised learning OR unsupervised learning OR practice based OR modelling OR mapping OR cluster* OR data analys*

		OR AB unsupervised learning OR AB practice based OR AB modelling OR AB mapping OR AB cluster* OR AB data analys*		
Clinical pathways	Clinical pathways; delivery of health care, integrated; clinical practice pattern; disease management; care management, patient	MH Critical Path OR MH Health Care Delivery, Integrated OR MH Practice Patterns OR MH Disease Management OR MH Patient Care Plans OR TI Clinical course OR TI integrated care OR TI care map OR TI care pathway OR TI care plan OR TI treatment plan OR TI patient journey OR TI patient flow OR TI clinical redesign OR TI integrated care OR AB Clinical course OR AB integrated care OR AB care map OR AB care pathway OR AB care plan OR AB treatment plan OR AB patient journey OR AB patient flow OR AB clinical redesign OR AB integrated care	(Clinical pathways or delivery of health care, integrated or clinical practice pattern or disease management or care management, patient).sh. or (Clinical course or integrated care or care map or care pathway or care plan or treatment plan or patient journey or patient flow or clinical redesign or integrated care).ti. or (Clinical course or integrated care or care map or care pathway or care plan or treatment plan or patient journey or patient flow or clinical redesign or integrated care).ab.	Clinical course OR integrated care OR care map OR care pathway OR care plan OR treatment plan OR patient journey OR patient flow OR clinical redesign OR integrated care

Chronic conditions	Chronic diseases; chronic illness	MH Chronic Disease OR TI Integrated chronic care OR AB Integrated chronic care		Integrated chronic care

EQUATIONS

MEDLINE

((data-driven[Title/Abstract] OR health information[Title/Abstract] OR data analys*[Title/Abstract] OR computer graphics[MeSH Terms] OR visualization[Title/Abstract] OR machine learning[MeSH Terms] OR data mining[MeSH Terms] OR clinical decision support systems[MeSH Terms] OR medical informatics application[MeSH Terms] OR algorithm[MeSH Terms] OR supervised learning[Title/Abstract] OR unsupervised learning[Title/Abstract] OR analysis, cluster[MeSH Terms] OR practice-based[Title/Abstract] OR electronic health record[MeSH Terms] OR clinical decision support systems[Title/Abstract] OR process mining[Title/Abstract] OR data mining [Title/Abstract] OR machine learning [Title/Abstract] OR medical informatics application[Title/Abstract] OR cluster*[Title/Abstract] OR modeling[Title/Abstract] OR mapping[Title/Abstract])

AND

(chronic diseases[MeSH Terms] OR chronic illness[MeSH Terms] OR integrated chronic care[Title/Abstract])

AND

(delivery of health care, integrated[MeSH Terms] OR clinical practice pattern[MeSH Terms] OR clinical pathway[MeSH Terms] OR critical pathway[MeSH Terms] OR clinical course[Title/Abstract] OR integrated care[Title/Abstract] OR care map[Title/Abstract] OR care pathway[Title/Abstract] OR care plan[Title/Abstract] OR treatment plan[Title/Abstract] OR disease management[MeSH Terms] OR disease management[Title/Abstract] OR care management, patient[MeSH Terms] OR patient journey[Title/Abstract] OR patient flow[Title/Abstract] OR clinical redesign[Title/Abstract] OR integrated care[Title/Abstract]))

SCOPUS

(TITLE-ABS-KEY(“data-driven”) OR TITLE-ABS-KEY("health information") OR TITLE-ABS-KEY("data analys*") OR INDEXTERMS("computer graphics") OR TITLE-ABS-KEY(“visuali*ation”) OR INDEXTERMS("machine learning") OR INDEXTERMS("data mining") OR INDEXTERMS("clinical decision support systems") OR INDEXTERMS("medical informatics application") OR INDEXTERMS("algorithm") OR TITLE-ABS-KEY("supervised learning") OR TITLE-ABS-KEY("unsupervised learning") OR INDEXTERMS("cluster analysis") OR INDEXTERMS(“practice-based”) OR INDEXTERMS("electronic health record") OR TITLE-ABS-KEY("clinical decision support systems") OR TITLE-ABS-KEY("process mining") OR TITLE-ABS-KEY("data mining") OR TITLE-ABS-KEY("machine learning") OR TITLE-ABS-KEY("medical informatics application") OR TITLE-ABS-KEY(cluster*) OR TITLE-ABS-KEY("modelling") OR TITLE-ABS-KEY("mapping"))

AND (INDEXTERMS("chronic diseases") OR INDEXTERMS("chronic illness") OR TITLE-ABS-KEY("integrated chronic care"))

AND (INDEXTERMS("integrated delivery of health care") OR TITLE-ABS-KEY("clinical practice pattern") OR INDEXTERMS("clinical pathway") OR INDEXTERMS("critical pathway") OR TITLE-ABS-KEY("clinical course") OR TITLE-ABS-KEY("integrated care") OR TITLE-ABS-KEY("care map") OR TITLE-ABS-KEY("care pathway") OR TITLE-ABS-KEY("care plan") OR TITLE-ABS-KEY("treatment plan") OR INDEXTERMS ("disease management") OR TITLE-ABS-KEY("disease management") OR INDEXTERMS("patient care management") OR TITLE-ABS-KEY("patient journey") OR TITLE-ABS-KEY("patient flow") OR TITLE-ABS-KEY("clinical redesign") OR TITLE-ABS-KEY("integrated care"))

IEEE

((((((((((((((((((((((((((("Index Terms":electronic health records) OR "Index Terms":data mining) OR "Index Terms":machine learning) OR "Index Terms":clinical decision support systems) OR "Index Terms":cluster analysis) OR "Index Terms":medical informatics) OR "Index Terms":computer graphics) OR "Index Terms":algorithm) OR "IEEE Terms":medical information systems) OR "IEEE Terms":electronic medical records) OR "Author Keywords":healthcare practices) OR "Publication Title":data-driven) OR "Abstract":data-driven) OR "Publication Title":machine learning) OR "Abstract":machine learning) OR "Publication Title":cluster analys*) OR "Abstract":cluster analys*) OR "Publication Title":data mining) OR "Abstract":data mining) OR "Author Keywords":electronic health record) OR "IEEE Terms":Guidelines) OR "IEEE Terms":Data mining) OR "IEEE Terms":Algorithm design and analysis) OR data mining) OR data-driven) OR electronic health record) OR algorithm) OR visualization) OR clustering) OR algorithm)

AND ((((((((((("Index Terms":clinical pathway) OR "Publication Title":clinical pathway) OR "Abstract":clinical pathway) OR"Author Keywords ":clinical pathway) OR "Author Keywords":healthcare practices) OR"Author Keywords ":Pathway) OR clinical path*) OR care pattern) OR care plan) OR care map) OR critical path*)

AND (((((("Index Terms":chronic disease) OR "Publication Title":chronic disease) OR "Abstract":chronic disease) OR "IEEE Terms":Diseases) OR chronic*)

Reporting checklist for protocol of a systematic review.

Based on the PRISMA-P guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the PRISMA-P reporting guidelines, and cite them as:

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1.

		Reporting Item	Page Number
Title			
Identification	#1a	Identify the report as a protocol of a systematic review	1
Update	#1b	If the protocol is for an update of a previous systematic review, identify as such	n/a

Registration			
	#2	If registered, provide the name of the registry (such as PROSPERO) and registration number	1
Authors			
Contact	#3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contribution	#3b	Describe contributions of protocol authors and identify the guarantor of the review	7
Amendments			
	#4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support			
Sources	#5a	Indicate sources of financial or other support for the review	7
Sponsor	#5b	Provide name for the review funder and / or sponsor	7
Role of sponsor or funder	#5c	Describe roles of funder(s), sponsor(s), and / or institution(s), if any, in developing the protocol	n/a
Introduction			
Rationale	#6	Describe the rationale for the review in the context of what is	2,3

		already known	
Objectives	#7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	4
Methods			
Eligibility criteria	#8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	4,5
Information sources	#9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	4
Search strategy	#10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	4
Study records - data management	#11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	5
Study records - selection process	#11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	5
Study records - data collection	#11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any	5,6

process		processes for obtaining and confirming data from investigators	
Data items	#12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	5,6
Outcomes and prioritization	#13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	n/a
Risk of bias in individual studies	#14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	6
Data synthesis	#15a	Describe criteria under which study data will be quantitatively synthesised	6
Data synthesis	#15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I2, Kendall's τ)	n/a
Data synthesis	#15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
Data synthesis	#15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	#16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within	n/a

		studies)	
Confidence in cumulative evidence	#17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

BMJ Open

Quantification and visualization methods of data-driven chronic care delivery pathways: protocol for a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033573.R1
Article Type:	Protocol
Date Submitted by the Author:	22-Jan-2020
Complete List of Authors:	Siqueira do Prado, Luiza; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research Allemann, Samuel; University of Basel, Pharmaceutical Sciences Viprey, Marie; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research; Hospices Civils de Lyon, Pôle Santé Publique Schott, Anne-Marie; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research; Hospices Civils de Lyon, Pôle Santé Publique Dediu, Dan; Université Lumière Lyon 2, Laboratoire Dynamique du Langage UMR 5596 Dima, Alexandra; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research
Primary Subject Heading:	Health informatics
Secondary Subject Heading:	Health services research, Public health
Keywords:	clinical decision support systems, medical informatics application, data visualization, clinical pathway, delivery of health care, integrated, electronic healthcare databases

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title

Quantification and visualization methods of data-driven chronic care delivery pathways: protocol for a systematic review

Corresponding author

Luiza Siqueira do Prado

luiza.siqueira-do-prado@univ-lyon1.fr

EA 7425 HESPER
Health Services and Performance Research
Université Claude Bernard Lyon 1
Domaine Rockefeller- 2eme étage (aile CD)
8 avenue Rockefeller
69373 Lyon 8
+33 (0) 7 83 45 23 66

Authors

Luiza Siqueira do Prado (luiza.siqueira-do-prado@univ-lyon1.fr)[1],

Samuel Allemann (s.allemann@unibas.ch)[1,2],

Marie Viprey (marie.viprey@chu-lyon.fr)[1,3],

Anne-Marie Schott (anne-marie.schott-pethelaz@chu-lyon.fr)[1,3],

Dan Dediu (dan.dediu@univ-lyon2.fr)[4],

Alexandra Dima (alexandra.dima@univ-lyon1.fr)[1]

Affiliations

[1] EA 7425 Health Services and Performance Research HeSPeR – Université Claude Bernard Lyon 1, Lyon, France

[2] Pharmaceutical Care Research Group, Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland

[3] Pôle de Santé Publique, Hospices Civils de Lyon, Lyon, France

[4] Laboratoire Dynamique du Langage UMR 5596, Université Lumière Lyon 2, Lyon, France

Abstract

Introduction Chronic conditions require long periods of care and often involve repeated interactions with multiple healthcare providers. Faced with increasing illness burden and costs, healthcare systems are currently working towards integrated care to streamline these interactions and improve efficiency. To support this, one promising resource is the information on routine care delivery stored in various electronic healthcare databases (EHD). In chronic conditions, care delivery pathways (CDPs) can be constructed by linking multiple data sources and extracting time-stamped healthcare utilization events and other medical data related to individual or groups of patients over specific time periods; CDPs may provide insights into current practice and ways of improving it. Several methods

have been proposed in recent years to quantify and visualize CDPs. We present the protocol for a systematic review aiming to describe the content and development of CDP methods, to derive common recommendations for CDP construction.

Methods and analysis This protocol followed the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P). A literature search will be performed in PubMed (MEDLINE), Scopus, IEEE, CINAHL and EMBASE, without date restrictions, to review published papers reporting data-driven chronic CDPs quantification and visualization methods. We will describe them using several characteristics relevant for EHD use in long-term care, grouped into three domains: 1) clinical (what health-related events it includes and for what clinical aims?), 2) data science (how is the method developed and what data infrastructure it relies on?), and 3) behavioral (what behaviors and interactions does it promote in users and through what methods?). Data extraction will be performed via deductive content analysis using previously defined characteristics and accompanied by an inductive analysis to identify and code additional relevant features. Results will be presented in descriptive format and used to compare current CDPs and generate recommendations for future CDP development initiatives.

Ethics and dissemination Ethical approval is not required for this review. Results will be disseminated in peer-reviewed journals and conference presentations.

PROSPERO registration CRD42019140494

Keywords clinical decision support systems; medical informatics application; data visualization; clinical pathway; delivery of health care, integrated; electronic healthcare databases.

Word count 2612

Strengths and Limitations

- While most reviews of health technology tools focus on clinical objectives and technical characteristics, we will also consider behaviours of and interactions between users to describe the selected methods.
- We will perform both deductive and inductive content analysis to fully describe the methods.
- We will focus on methods described in peer-reviewed papers and exclude conference proceedings and other types of reports, to obtain detailed validated descriptions; this may limit our access to more recent studies due to the fast-paced development in the field.
- Lack of completeness in methods descriptions may limit our ability to assess all characteristics, such as the stages of development, the involvement of stakeholders or experts prior to data acquisition and analysis.
- As this is a relatively new field of health technology, there are no guidelines for reporting and no consensus on quality criteria for the studies we will evaluate; our work will also contribute to the development of such recommendations.

Introduction

Effective delivery of integrated care is a priority for healthcare systems worldwide and has been the focus of considerable efforts in recent years, particularly in response to the increasing demands of chronic care[1,2]. Long-term conditions may require lifetime care, which may consist of multiple interactions with a variety of healthcare providers at variable time intervals[3,4]. When service delivery is fragmented, the overall effectiveness of these interactions in terms of long-term quality of life and health-related outcomes is reduced, and risk of harm is increased[5,6]. Centralizing patient information

produced by different providers in electronic healthcare databases (EHD) has the potential to help implementing new ways of service delivery to improve outcomes[7]. Several attempts have been made to link multiple data sources to generate comprehensive descriptions of patients' healthcare journeys in long-term conditions. These descriptions are produced by constructing longitudinal trajectories from various time-stamped healthcare utilization events and related medical data[8–17]. For example, Zhang et al. have produced longitudinal trajectories using electronic health records (EHR) and cost pathways [14,16,17] of people living with chronic kidney disease to inform patient engagement and to detect common pathways. Bettencourt-Silva et al. (2015) have reported on the development of a patient-centric database from multiple Hospital Information Systems (HIS)[18] and on building data-driven pathways from routine hospital data on people living with prostate cancer to explore their potential use in biomedical research[15]. However, generating these informative trajectories from disparate and often incompatible data sources proves challenging[18,19]. As various initiatives have been developed independently, with distinct methodologies and objectives, it is essential to examine systematically the proposed solutions in order to derive principles of action to stimulate convergence of methods.

In the context of chronic conditions, the way patient trajectories are established may be subject to multiple influences and analyzing routine care data can provide insights on how they have been drawn over time and their potential sources of variation[14,20]. In the literature, trajectories within healthcare systems have been described using many terms, which makes it challenging to build consensus on terminology and practical meaning[21,22]. We will use the term data-driven 'care delivery pathway' (CDP) to group several terms we will find in the selected studies to designate retrospective trajectories obtained from EHD. To describe the methods proposed for synthetically displaying objective measures or assessments of health status or healthcare utilization (e.g., quantifying) and graphically showing the temporal elements of chronic CDP (e.g., visualizing), we will assess how they addressed three domains:

1) The selection of relevant clinical and health-related events.

This domain will examine how the methods define health status and evaluate disease progression or stabilization, and how they show transitions between health status and acute manifestations[13]. Usually, the trajectory timeline begins at diagnosis and involves more than one provider[13–15]. Treatment decisions are generally based on health status (indicated by biomarkers, clinical examination, self-declared levels of quality of life, etc.), care units and settings, treatment availability (medication, procedures, etc.), and patient-provider preferences[20].

2) The technological development itself and considering issues related to data quality and exchange.

This domain aims to describe how the method is built, which data sources and analyses are used, and the necessary infrastructure surrounding its implementation. Digitalization of health-related data is a global trend[23,24] and highly detailed data are being collected daily in diverse settings and healthcare services. Such methods may apply a range of techniques from basic algorithms to advanced statistical and machine learning models[25], which can provide useful insights into care delivery processes. Technological developments in this field also need to meet strict criteria of data security, accuracy of models and predictions, openness of development and validation processes, among others[15,26].

3) Considering behaviours of actors and interactions between them with the aim of effectively improving care delivery.

Integrated care depends on multiple actions and decisions made collaboratively by patients, healthcare providers, administrative staff and other actors concerning patients' course of treatment[27]. To inform these decisions, technological solutions must have access to clinical exams and provide key actors with relevant information, such as the patients' past interactions with other providers, the medical procedures performed, the medications prescribed[28]. To have a positive impact on improving care delivery, visualizations and quantitative indicators of the patient's prior care need to be adapted to the user's needs at specific points in the trajectory, like after acute events or hospitalizations. This domain will examine what behaviors and interactions the methods promote (who are its target individuals, what actions need to be performed, in what context, when, and by whom)[29,30], and what strategies are proposed to encourage this performance.

Aims and objectives

We aim to identify and describe the methods that have been proposed to quantify and/or visualize data-driven CPDs of people living with chronic conditions. Given the complexity of their context of use, more than only reviewing technical methods, we aim to investigate how these tools have considered the three domains described above.

For this end, we propose the following research questions:

1. What clinical information does the method use and how was it considered relevant?
2. What are the method's development and implementation characteristics?
3. Which behaviours and interactions does the method aim to promote among users and how?

Methods

The Cochrane Handbook[31] and the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P)[32] were used to write this protocol and the systematic review will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[33]. PRISMA-P checklist is in Supplementary File 1. The review will be performed by one primary reviewer (LSP) and three secondary reviewers (AD, MV and SA) and will follow 6 steps: literature search, records screening and pre-selection (title and abstract), full-text screening and final selection, extraction of data, quality assessment, analysis and synthesis of data.

The studies expected to be analyzed in this work will likely be descriptive and not follow standard methodology (i.e., experimental or observational, method validation), yet considering the manuscripts as a qualitative corpus allows for coding the narratives according to the conceptual structure we propose[7,34]. Content analysis has been used in many studies in health sciences[35] and an inductive content analysis applied in a systematic review of clinical information modeling processes[34] has developed descriptive categories in a context similar to the one we propose here. As we consider them relevant to the studies we will review, they will be included in our coding framework, as detailed below.

Searches

A literature search will be performed in the following electronic databases: PubMed (MEDLINE), Scopus, IEEE, CINAHL and EMBASE. The search will be adapted to each database, and the resulting search strategies are provided as Supplementary File 2. The terms searched will be related to three main categories, connected by the AND operator: "data-driven" (MeSH terms like "Electronic health record", "data mining", etc.), "clinical pathways" (MeSH terms like "clinical pathway", "disease management", etc.), and "chronic conditions" (MeSH term "chronic diseases"). Searches will be

performed with MeSH terms or with keywords in Title/Abstract in PubMed; MeSH terms will be adapted for the databases that do not permit their usage or use different indexed terms. Bibliographies and citation tracking of relevant literature will be hand searched to identify additional relevant studies. A first selection will be performed using abstracts and titles, followed by full-text examination of entries selected.

Types of publications/studies and eligibility criteria

We will consider CDPs to be a series of time-stamped events describing the sequence of care of users with a diagnosed chronic condition (conditions requiring medical attention for a period longer than 12 months)[36]. These events can be the diagnosis itself, routine, non-scheduled or emergency consultations with a general practitioner and/or specialist, therapeutic education sessions and other health-related interventions. These can result in prescriptions of medications, medical procedures and tests, which may also appear in the trajectory. Data-driven CDP analyzed here will need to be composed of at least two time-stamped events recorded in EHD from people with the diagnosis of a chronic condition, with no duration restrictions (e.g., CDP may cover periods from days or few months to several years).

We will consider peer-reviewed publications (1) reporting methods for visualization or quantification of data-driven chronic CDP (including protocols and reports of study results), (2) using data from people living with chronic conditions retrieved from EHD and (3) published in English. No restrictions on publication date, study design, population characteristics, type of healthcare facility and level of care will be applied.

We will exclude studies that aim only to assess healthcare utilization over a specific period as part of a single research study, for example as an outcome to evaluate health-related interventions, to describe populations or disease prevalence, or as a proxy measure of disease aggravation risk. We will also exclude studies that do not mention population or data characteristics or do not state they analyze data from people living with chronic conditions, papers with full-text not written in English, conference abstracts, systematic or narrative reviews, meta-analyses and grey literature.

Screening

We will use Covidence, an online systematic review management software, for records screening. After duplicates removal, titles and abstracts in the remaining records will be screened independently by two reviewers for full text appraisal. If reviewer discordance arises, consensus will be reached through discussion and arbitration with one of the secondary reviewers not involved in the selection of the record. Studies selected in the first step will go through full text screening using the same process to establish eligibility. Inter-rater reliability (Cohen's Kappa) between primary and secondary reviewers will be computed and reported.

Data management

We will report the number of included and excluded articles as well as the number of full-text papers obtained and assessed. Reasons for exclusion of screened full-text studies will also be stated in the final review. The data will be managed using Covidence and Microsoft Excel spreadsheets.

Data extraction and analysis

We will use both deductive and inductive content analysis[35] to appraise the selected studies: deductive when relying on pre-defined frameworks such as the categories previously described by Moreno-Conde et al. (2015)[34] to describe the technical characteristics of the proposed solutions and

on the AACTT framework[29,30] (action, actor, context, target, time), to describe the behavioural domain, and inductive when additional relevant characteristics need to be described.

Data from included studies will be extracted using a customized electronic data extraction form. Information on study characteristics (authors, title, type of study, year and country of study, objective and research questions); population characteristics (number of patients, age, gender, condition) will be extracted directly from the included studies.

Deductive-Inductive content analysis

We will perform a deductive content analysis following existing theories, as described below, and inductive analysis for observed relevant characteristics not yet covered by existing literature. If more than one selected record describe development, validation and/or implementation of the same method, we will extract basic paper characteristics, as described above, but the content analysis will be performed per method.

1) For the clinical domain, we will extract information on clinical or cost outcomes the method might target (if reported and which ones) and on how the outcomes were considered relevant (e.g., involving experts, final users or other stakeholders).

2) For method development and data processing, we will analyze and compare to what has been proposed by Moreno-Conde et al. (2015)[34]. The categories detailed in the study are briefly described below.

- Scope definition leading to selection of the domain and selecting relevant experts: identifying the domain and expected uses of the method through the creation of a group of experts.
- Analysis of the information covered in the specific domain: creation of definitions, identification of clinical scenarios, workflows, users, guidelines, literature, etc., so the method meet the requirements of clinical practice or other intended usages.
- Design of the tool: detailing the set of attributes associated with the method, domain terminologies, ensuring compatibility across domains.
- Definition of implementable tool specifications: description of implementable technical specification.
- Validation: use of techniques to validate the method, such as peer-review validation or creation of prototype screens.
- Publishing and maintenance: availability in public repositories.
- Governance: description of the organization responsible for developing and maintaining the tool.

Other information extracted from studies regarding this domain will be healthcare utilization characteristics (type of event, e.g., consultation, test, procedure) and data characteristics (sources of data, data preparation, data analysis).

3) To describe behaviour and interactions the method might promote or facilitate, we will apply the AACTT[29,30] framework. Other information extracted from studies will be output characteristics like intended final users, purpose and use scenarios. We will also code the presence of strategies planned or performed to achieve these behavioural change objectives, such as training, organizational changes, evaluation of the performance of the method in routine care, if implemented, and other initiatives studies might present.

The primary reviewer and one secondary reviewer will pilot data extraction independently for a subset of 10% of selected records to compare and discuss data extraction process. If necessary, we will repeat the pilot extraction process (outlined above) until agreement is reached. Disagreements will be solved with the help of a third reviewer and piloting may consist of several interactions between reviewers to compare and reach consensus regarding relevant information to be extracted from full-text analysis. After this first step, a codebook will be developed, and data extraction of the remaining records will be performed by the primary reviewer.

Quality and bias assessment

As most quality assessment tools are developed for commonly-used study designs and there is no consensus regarding tools for generic use, we propose to evaluate quality from a different perspective. We will evaluate if main stakeholders (patients and/or family, healthcare professionals, administrative personnel) were involved at any stage of the development of the method. Research shows the importance of involving patients, the public and other stakeholders in health-related research to obtain experiential knowledge, setting research priorities and focus on practical questions[37–40]. Also, it has been shown that trials funded by for-profit organizations can positively bias interpretation of trial results[41], and research in data usage can be funded by companies interested in selling their own methods. To assess potential bias, we will evaluate declared conflicts of interest and sources of funding. Quality assessment will be discussed in the review, but no study will be excluded from the analysis based on quality criteria.

Data synthesis

The technical methods will be synthesized using the content analysis described above and the studies will be categorized and described using the 3 domains, depending on study type and reporting. We will present the results in tables along with method and study identification and summarize via descriptive statistics. We will compare the different characteristics within the 3 domains to identify common, infrequent, or missing features of these tools, and extract recommendations for future initiatives.

Patient and public involvement

A representant of a patients' association was involved in reading and approving of this protocol. This systematic review is part of a larger project that will be developed closely with patients and healthcare providers.

Ethics and dissemination

Ethical approval is not required. Results will be disseminated in peer-reviewed journals and/or conference presentations. Data used in this review will be made available through supplementary materials and open trusted repositories.

Funding

DD was supported by an IDEXLYON (16-IDEX-0005) Fellowship grant (2018-2021), LSP was supported by a PhD funding within the same grant, AD by a Marie Curie Individual Fellowship from the European Commission (MCRA-IF n°706028) during the preparation of this review protocol, and SA by the Swiss Science Foundation (P2BSP3_ 178648).

Conflicts of interest

None declared.

Author statement

LSP, AD and SA designed the protocol and planned data extraction and quality assessment. LSP put together the search strategy and SA helped adapt it to the different databases. LSP and AD conceived the content analysis stages and conceptual framework. LSP wrote the first version of the manuscript, AD extensively reviewed it, SA, DD, AMS and MV revised it critically for important intellectual content. All authors have approved the publication of this protocol and contributed to the final manuscript.

References

1. Bodenheimer, T., Wagner, E. H. & Grumbach, K. Improving Primary Care for Patients With Chronic Illness: The Chronic Care Model, Part 2. *JAMA* **288**, 1909–1914 (2002).

2. Ouwens, M., Wollersheim, H., Hermens, R., Hulscher, M. & Grol, R. Integrated care programmes for chronically ill patients: a review of systematic reviews. *Int J Qual Health Care* **17**, 141–146 (2005).

3. Bodenheimer, T., Wagner, E. H. & Grumbach, K. Improving Primary Care for Patients With Chronic Illness. *JAMA* **288**, 1775–1779 (2002).

4. Pham, H. H., Schrag, D., O'Malley, A. S., Wu, B. & Bach, P. B. Care Patterns in Medicare and Their Implications for Pay for Performance. doi:10.1056/NEJMsa063979.

5. Wagner, E. H. Meeting the needs of chronically ill people: Socioeconomic factors, disabilities, and comorbid conditions are obstacles. *BMJ* **323**, 945–946 (2001).

6. Campbell, H., Hotchkiss, R., Bradshaw, N. & Porteous, M. Integrated care pathways. *BMJ* **316**, 133–137 (1998).

7. Gooch, P. & Roudsari, A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *J Am Med Inform Assoc* **18**, 738–748 (2011).

8. Huang, Z. *et al.* Discovery of clinical pathway patterns from event logs using probabilistic topic models. *Journal of Biomedical Informatics* **47**, 39–57 (2014).

9. Perer, A., Wang, F. & Hu, J. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics* **56**, 369–378 (2015).

10. Gotz, D., Wang, F. & Perer, A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *Journal of Biomedical Informatics* **48**, 148–159 (2014).

11. Huang, Z., Lu, X., Duan, H. & Fan, W. Summarizing clinical pathways from event logs. *Journal of Biomedical Informatics* **46**, 111–127 (2013).

12. Yang, W. & Su, Q. Process mining for clinical pathway: Literature review and future directions. in *2014 11th International Conference on Service Systems and Service Management (ICSSSM)* 1–5 (2014). doi:10.1109/ICSSSM.2014.6943412.

13. Jensen, A. B. *et al.* Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature Communications* **5**, 4022 (2014).

14. Zhang, Y., Padman, R. & Patel, N. Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. *J Biomed Inform* **58**, 186–197 (2015).

15. Bettencourt-Silva, J. H. *et al.* Building Data-Driven Pathways From Routinely Collected Hospital Data: A Case Study on Prostate Cancer. *JMIR Med Inform* **3**, e26 (2015).

16. Y. Zhang *et al.* On Clinical Pathway Discovery from Electronic Health Record Data. *IEEE Intelligent Systems* **30**, 70–75 (2015).

17. Zhang, Y. & Padman, R. Data-driven clinical and cost pathways for chronic care delivery. *Am J Manag Care* **22**, 816–820 (2016).

18. Bettencourt-Silva, J., De La Iglesia, B., Donell, S. & Rayward-Smith, V. On creating a patient-centric database from multiple Hospital Information Systems. *Methods Inf Med* **51**, 210–220 (2012).

19. Denaxas, S. C. & Morley, K. I. Big biomedical data and cardiovascular disease research: opportunities and challenges. *Eur Heart J Qual Care Clin Outcomes* **1**, 9–16 (2015).

20. Vanasse, A., Courteau, M. & Ethier, J.-F. The '6W' multidimensional model of care trajectories for patients with chronic ambulatory care sensitive conditions and hospital readmissions. *Public Health* **157**, 53–61 (2018).

21. Bleser, L. D. *et al.* Defining pathways. *Journal of Nursing Management* **14**, 553–563 (2006).

22. Kinsman, L., Rotter, T., James, E., Snow, P. & Willis, J. What is a clinical pathway? Development of a definition to inform the debate. *BMC Medicine* **8**, 31 (2010).

23. Adler-Milstein, J. *et al.* Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. *Health Affairs* **34**, 2174–2180 (2015).

24. Ben-Assuli, O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. *Health Policy* **119**, 287–297 (2015).

25. Zhang, Y. & Padman, R. Innovations in chronic care delivery using data-driven clinical pathways. *Am J Manag Care* **21**, e661-668 (2015).

26. Liaw, S. T. *et al.* Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int J Med Inform* **82**, 10–24 (2013).

27. Peikes, D., Chen, A., Schore, J. & Brown, R. Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures Among Medicare Beneficiaries: 15 Randomized Trials. *JAMA* **301**, 603–618 (2009).

28. Stille, C. J., Jerant, A., Bell, D., Meltzer, D. & Elmore, J. G. Coordinating care across diseases, settings, and clinicians: a key role for the generalist in practice. *Ann. Intern. Med.* **142**, 700–708 (2005).

29. Francis, J. J. & Presseau, J. Health Care Practitioner Behaviour. in *Cambridge Handbook of Psychology, Health and Medicine* (eds. Llewellyn, C. et al.) 324–324 (Cambridge University Press, 2019).

30. Presseau, J. *et al.* Action, actor, context, target, time (AACTT): a framework for specifying behaviour. *Implementation Sci* **14**, 1–13 (2019).

31. Cochrane Handbook for Systematic Reviews of Interventions. http://handbook-5-1.cochrane.org/.

32. Moher, D. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* **4**, 1 (2015).

33. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **6**, e1000097 (2009).

34. Moreno-Conde, A. *et al.* Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. *J Am Med Inform Assoc* **22**, 925–934 (2015).

35. Elo, S. & Kyngäs, H. The qualitative content analysis process. *Journal of Advanced Nursing* **62**, 107–115 (2008).

36. Goodman, R. A. Defining and Measuring Chronic Conditions: Imperatives for Research, Policy, Program, and Practice. *Prev Chronic Dis* **10**, (2013).

37. Staley, K. 'Is it worth doing?' Measuring the impact of patient and public involvement in research. *Research Involvement and Engagement* **1**, 6 (2015).

38. Hoffman, A., Montgomery, R., Aubry, W. & Tunis, S. R. How Best To Engage Patients, Doctors, And Other Stakeholders In Designing Comparative Effectiveness Studies. *Health Affairs* **29**, 1834–1841 (2010).

39. Selby, J. V., Beal, A. C. & Frank, L. The Patient-Centered Outcomes Research Institute (PCORI) National Priorities for Research and Initial Research Agenda. *JAMA* **307**, 1583–1584 (2012).

40. Ocloo, J. & Matthews, R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf* **25**, 626–632 (2016).

41. Als-Nielsen, B., Chen, W., Gluud, C. & Kjaergard, L. L. Association of Funding and Conclusions in Randomized Drug Trials: A Reflection of Treatment Effect or Adverse Events? *JAMA* **290**, 921–928 (2003).

Reporting checklist for protocol of a systematic review.

Based on the PRISMA-P guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the PRISMA-P reporting guidelines, and cite them as:

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1.

		Reporting Item	Page Number
Title			
Identification	#1a	Identify the report as a protocol of a systematic review	1
Update	#1b	If the protocol is for an update of a previous systematic review, identify as such	n/a
Registration			
	#2	If registered, provide the name of the registry (such as PROSPERO) and registration number	1
Authors			
Contact	#3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contribution	#3b	Describe contributions of protocol authors and identify the	7

		guarantor of the review	
Amendments			
	#4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support			
Sources	#5a	Indicate sources of financial or other support for the review	7
Sponsor	#5b	Provide name for the review funder and / or sponsor	7
Role of sponsor or funder	#5c	Describe roles of funder(s), sponsor(s), and / or institution(s), if any, in developing the protocol	n/a
Introduction			
Rationale	#6	Describe the rationale for the review in the context of what is already known	2,3
Objectives	#7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	4
Methods			
Eligibility criteria	#8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	4,5
Information sources	#9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	4
Search strategy	#10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	4
Study records - data management	#11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	5
Study records -	#11b	State the process that will be used for selecting studies (such	5

selection process		as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	
Study records - data collection process	#11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	5,6
Data items	#12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	5,6
Outcomes and prioritization	#13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	n/a
Risk of bias in individual studies	#14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	6
Data synthesis	#15a	Describe criteria under which study data will be quantitatively synthesised	6
Data synthesis	#15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I2, Kendall's τ)	n/a
Data synthesis	#15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
Data synthesis	#15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	#16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	n/a
Confidence in cumulative evidence	#17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

COMPLETE SEARCH STRATEGY

Category	Medical Subject headings (MeSH)	CINAHL Search	EMBASE	Keywords
Data-driven	Electronic health record; data mining; machine learning; clinical decision support systems; analysis, cluster; medical informatics application	MH Electronic Health Records OR MH Data Mining OR MH Nursing Informatics OR MH Machine Learning OR MH Decision Support Systems, Clinical OR MH Cluster Analysis OR MH Medical Informatics OR MH Computer Graphics OR MH Algorithms OR TI Data-driven OR TI visualisation OR TI computer graphics OR TI process mining OR TI data mining OR TI visualization OR TI supervised learning OR TI unsupervised learning OR TI practice based OR TI modelling OR TI mapping OR TI cluster* OR TI data analys* OR AB Data-driven OR AB visualisation OR AB computer graphics OR AB process mining OR AB data mining OR AB visualization OR AB supervised learning	(Electronic health record or data mining or machine learning or clinical decision support systems or analysis, cluster or medical informatics application).sh. or (Data-driven or visualisation or computer graphics or process mining or data mining or visualization or supervised learning or unsupervised learning or practice based or modelling or mapping or cluster* or data analys*).ti. or (Data-driven or visualisation or computer graphics or process mining or data mining or visualization or supervised learning or unsupervised learning or practice based or modelling or mapping or cluster* or data analys*).ab.	Data-driven OR visualisation OR computer graphics OR process mining OR data mining OR visualization OR supervised learning OR unsupervised learning OR practice based OR modelling OR mapping OR cluster* OR data analys*

		OR AB unsupervised learning OR AB practice based OR AB modelling OR AB mapping OR AB cluster* OR AB data analys*		
Clinical pathways	Clinical pathways; delivery of health care, integrated; clinical practice pattern; disease management; care management, patient	MH Critical Path OR MH Health Care Delivery, Integrated OR MH Practice Patterns OR MH Disease Management OR MH Patient Care Plans OR TI Clinical course OR TI integrated care OR TI care map OR TI care pathway OR TI care plan OR TI treatment plan OR TI patient journey OR TI patient flow OR TI clinical redesign OR TI integrated care OR AB Clinical course OR AB integrated care OR AB care map OR AB care pathway OR AB care plan OR AB treatment plan OR AB patient journey OR AB patient flow OR AB clinical redesign OR AB integrated care	(Clinical pathways or delivery of health care, integrated or clinical practice pattern or disease management or care management, patient).sh. or (Clinical course or integrated care or care map or care pathway or care plan or treatment plan or patient journey or patient flow or clinical redesign or integrated care).ti. or (Clinical course or integrated care or care map or care pathway or care plan or treatment plan or patient journey or patient flow or clinical redesign or integrated care).ab.	Clinical course OR integrated care OR care map OR care pathway OR care plan OR treatment plan OR patient journey OR patient flow OR clinical redesign OR integrated care

Chronic conditions	Chronic diseases; chronic illness	MH Chronic Disease OR TI Integrated chronic care OR AB Integrated chronic care		Integrated chronic care

EQUATIONS

MEDLINE

((data-driven[Title/Abstract] OR health information[Title/Abstract] OR data analys*[Title/Abstract] OR computer graphics[MeSH Terms] OR visualization[Title/Abstract] OR machine learning[MeSH Terms] OR data mining[MeSH Terms] OR clinical decision support systems[MeSH Terms] OR medical informatics application[MeSH Terms] OR algorithm[MeSH Terms] OR supervised learning[Title/Abstract] OR unsupervised learning[Title/Abstract] OR analysis, cluster[MeSH Terms] OR practice-based[Title/Abstract] OR electronic health record[MeSH Terms] OR clinical decision support systems[Title/Abstract] OR process mining[Title/Abstract] OR data mining [Title/Abstract] OR machine learning [Title/Abstract] OR medical informatics application[Title/Abstract] OR cluster*[Title/Abstract] OR modeling[Title/Abstract] OR mapping[Title/Abstract])

AND

(chronic diseases[MeSH Terms] OR chronic illness[MeSH Terms] OR integrated chronic care[Title/Abstract])

AND

(delivery of health care, integrated[MeSH Terms] OR clinical practice pattern[MeSH Terms] OR clinical pathway[MeSH Terms] OR critical pathway[MeSH Terms] OR clinical course[Title/Abstract] OR integrated care[Title/Abstract] OR care map[Title/Abstract] OR care pathway[Title/Abstract] OR care plan[Title/Abstract] OR treatment plan[Title/Abstract] OR disease management[MeSH Terms] OR disease management[Title/Abstract] OR care management, patient[MeSH Terms] OR patient journey[Title/Abstract] OR patient flow[Title/Abstract] OR clinical redesign[Title/Abstract] OR integrated care[Title/Abstract]))

SCOPUS

(TITLE-ABS-KEY(“data-driven”) OR TITLE-ABS-KEY("health information") OR TITLE-ABS-KEY("data analys*") OR INDEXTERMS("computer graphics") OR TITLE-ABS-KEY(“visuali*ation”) OR INDEXTERMS("machine learning") OR INDEXTERMS("data mining") OR INDEXTERMS("clinical decision support systems") OR INDEXTERMS("medical informatics application") OR INDEXTERMS("algorithm") OR TITLE-ABS-KEY("supervised learning") OR TITLE-ABS-KEY("unsupervised learning") OR INDEXTERMS("cluster analysis") OR INDEXTERMS(“practice-based”) OR INDEXTERMS("electronic health record") OR TITLE-ABS-KEY("clinical decision support systems") OR TITLE-ABS-KEY("process mining") OR TITLE-ABS-KEY("data mining") OR TITLE-ABS-KEY("machine learning") OR TITLE-ABS-KEY("medical informatics application") OR TITLE-ABS-KEY(cluster*) OR TITLE-ABS-KEY("modelling") OR TITLE-ABS-KEY("mapping"))

AND (INDEXTERMS("chronic diseases") OR INDEXTERMS("chronic illness") OR TITLE-ABS-KEY("integrated chronic care"))

AND (INDEXTERMS("integrated delivery of health care") OR TITLE-ABS-KEY("clinical practice pattern") OR INDEXTERMS("clinical pathway") OR INDEXTERMS("critical pathway") OR TITLE-ABS-KEY("clinical course") OR TITLE-ABS-KEY("integrated care") OR TITLE-ABS-KEY("care map") OR TITLE-ABS-KEY("care pathway") OR TITLE-ABS-KEY("care plan") OR TITLE-ABS-KEY("treatment plan") OR INDEXTERMS ("disease management") OR TITLE-ABS-KEY("disease management") OR INDEXTERMS("patient care management") OR TITLE-ABS-KEY("patient journey") OR TITLE-ABS-KEY("patient flow") OR TITLE-ABS-KEY("clinical redesign") OR TITLE-ABS-KEY("integrated care"))

IEEE

((((((((((((((((((((((((("Index Terms":electronic health records) OR "Index Terms":data mining) OR "Index Terms":machine learning) OR "Index Terms":clinical decision support systems) OR "Index Terms":cluster analysis) OR "Index Terms":medical informatics) OR "Index Terms":computer graphics) OR "Index Terms":algorithm) OR "IEEE Terms":medical information systems) OR "IEEE Terms":electronic medical records) OR "Author Keywords":healthcare practices) OR "Publication Title":data-driven) OR "Abstract":data-driven) OR "Publication Title":machine learning) OR "Abstract":machine learning) OR "Publication Title":cluster analys*) OR "Abstract":cluster analys*) OR "Publication Title":data mining) OR "Abstract":data mining) OR "Author Keywords":electronic health record) OR "IEEE Terms":Guidelines) OR "IEEE Terms":Data mining) OR "IEEE Terms":Algorithm design and analysis) OR data mining) OR data-driven) OR electronic health record) OR algorithm) OR visualization) OR clustering) OR algorithm)

AND ((((((((((("Index Terms":clinical pathway) OR "Publication Title":clinical pathway) OR "Abstract":clinical pathway) OR"Author Keywords ":clinical pathway) OR "Author Keywords":healthcare practices) OR"Author Keywords ":Pathway) OR clinical path*) OR care pattern) OR care plan) OR care map) OR critical path*)

AND (((((("Index Terms":chronic disease) OR "Publication Title":chronic disease) OR "Abstract":chronic disease) OR "IEEE Terms":Diseases) OR chronic*)

BMJ Open

Quantification and visualization methods of data-driven chronic care delivery pathways: protocol for a systematic review and content analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033573.R2
Article Type:	Protocol
Date Submitted by the Author:	10-Feb-2020
Complete List of Authors:	Siqueira do Prado, Luiza; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research Allemann, Samuel; University of Basel, Pharmaceutical Sciences Viprey, Marie; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research; Hospices Civils de Lyon, Pôle Santé Publique Schott, Anne-Marie; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research; Hospices Civils de Lyon, Pôle Santé Publique Dediu, Dan; Université Lumière Lyon 2, Laboratoire Dynamique du Langage UMR 5596 Dima, Alexandra; Université Claude Bernard Lyon 1, HeSPeR - Health Services and Performance Research
Primary Subject Heading:	Health informatics
Secondary Subject Heading:	Health services research, Public health
Keywords:	clinical decision support systems, medical informatics application, data visualization, clinical pathway, delivery of health care, integrated, electronic healthcare databases

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title

Quantification and visualization methods of data-driven chronic care delivery pathways: protocol for a systematic review and content analysis

Corresponding author

Luiza Siqueira do Prado

luiza.siqueira-do-prado@univ-lyon1.fr

EA 7425 HESPER
Health Services and Performance Research
Université Claude Bernard Lyon 1
Domaine Rockefeller- 2eme étage (aile CD)
8 avenue Rockefeller
69373 Lyon 8
+33 (0) 7 83 45 23 66

Authors

Luiza Siqueira do Prado (luiza.siqueira-do-prado@univ-lyon1.fr)[1],

Samuel Allemann (s.allemann@unibas.ch)[1,2],

Marie Viprey (marie.viprey@chu-lyon.fr)[1,3],

Anne-Marie Schott (anne-marie.schott-pethelaz@chu-lyon.fr)[1,3],

Dan Dediu (dan.dediu@univ-lyon2.fr)[4],

Alexandra Dima (alexandra.dima@univ-lyon1.fr)[1]

Affiliations

[1] EA 7425 Health Services and Performance Research HeSPeR – Université Claude Bernard Lyon 1, Lyon, France

[2] Pharmaceutical Care Research Group, Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland

[3] Pôle de Santé Publique, Hospices Civils de Lyon, Lyon, France

[4] Laboratoire Dynamique du Langage UMR 5596, Université Lumière Lyon 2, Lyon, France

Abstract

Introduction Chronic conditions require long periods of care and often involve repeated interactions with multiple healthcare providers. Faced with increasing illness burden and costs, healthcare systems are currently working towards integrated care to streamline these interactions and improve efficiency. To support this, one promising resource is the information on routine care delivery stored in various electronic healthcare databases (EHD). In chronic conditions, care delivery pathways (CDPs) can be constructed by linking multiple data sources and extracting time-stamped healthcare utilization events and other medical data related to individual or groups of patients over specific time periods; CDPs may provide insights into current practice and ways of improving it. Several methods

have been proposed in recent years to quantify and visualize CDPs. We present the protocol for a systematic review aiming to describe the content and development of CDP methods, to derive common recommendations for CDP construction.

Methods and analysis This protocol followed the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P). A literature search will be performed in PubMed (MEDLINE), Scopus, IEEE, CINAHL and EMBASE, without date restrictions, to review published papers reporting data-driven chronic CDPs quantification and visualization methods. We will describe them using several characteristics relevant for EHD use in long-term care, grouped into three domains: 1) clinical (what health-related events it includes and for what clinical aims?), 2) data science (how is the method developed and what data infrastructure it relies on?), and 3) behavioral (what behaviors and interactions does it promote in users and through what methods?). Data extraction will be performed via deductive content analysis using previously defined characteristics and accompanied by an inductive analysis to identify and code additional relevant features. Results will be presented in descriptive format and used to compare current CDPs and generate recommendations for future CDP development initiatives.

Ethics and dissemination Database searches will be initiated in May 2019. The review is expected to be completed by February 2020. Ethical approval is not required for this review. Results will be disseminated in peer-reviewed journals and conference presentations.

PROSPERO registration CRD42019140494

Keywords clinical decision support systems; medical informatics application; data visualization; clinical pathway; delivery of health care, integrated; electronic healthcare databases.

Word count 2612

Strengths and Limitations

- While most reviews of health technology tools focus on clinical objectives and technical characteristics, we will also consider behaviours of and interactions between users to describe the selected methods.
- We will perform both deductive and inductive content analysis to fully describe the methods.
- We will focus on methods described in peer-reviewed papers and exclude conference proceedings and other types of reports, to obtain detailed validated descriptions; this may limit our access to more recent studies due to the fast-paced development in the field.
- Lack of completeness in methods descriptions may limit our ability to assess all characteristics, such as the stages of development, the involvement of stakeholders or experts prior to data acquisition and analysis.
- As this is a relatively new field of health technology, there are no guidelines for reporting and no consensus on quality criteria for the studies we will evaluate; our work will also contribute to the development of such recommendations.

Introduction

Effective delivery of integrated care is a priority for healthcare systems worldwide and has been the focus of considerable efforts in recent years, particularly in response to the increasing demands of chronic care[1,2]. Long-term conditions may require lifetime care, which may consist of multiple interactions with a variety of healthcare providers at variable time intervals[3,4]. When service delivery is fragmented, the overall effectiveness of these interactions in terms of long-term quality of life and

health-related outcomes is reduced, and risk of harm is increased[5,6]. Centralizing patient information produced by different providers in electronic healthcare databases (EHD) has the potential to help implementing new ways of service delivery to improve outcomes[7]. Several attempts have been made to link multiple data sources to generate comprehensive descriptions of patients' healthcare journeys in long-term conditions. These descriptions are produced by constructing longitudinal trajectories from various time-stamped healthcare utilization events and related medical data[8–17]. For example, Zhang et al. have produced longitudinal trajectories using electronic health records (EHR) and cost pathways [14,16,17] of people living with chronic kidney disease to inform patient engagement and to detect common pathways. Bettencourt-Silva et al. (2015) have reported on the development of a patient-centric database from multiple Hospital Information Systems (HIS)[18] and on building data-driven pathways from routine hospital data on people living with prostate cancer to explore their potential use in biomedical research[15]. However, generating these informative trajectories from disparate and often incompatible data sources proves challenging[18,19]. As various initiatives have been developed independently, with distinct methodologies and objectives, it is essential to examine systematically the proposed solutions in order to derive principles of action to stimulate convergence of methods.

In the context of chronic conditions, the way patient trajectories are established may be subject to multiple influences and analyzing routine care data can provide insights on how they have been drawn over time and their potential sources of variation[14,20]. In the literature, trajectories within healthcare systems have been described using many terms, which makes it challenging to build consensus on terminology and practical meaning[21,22]. We will use the term data-driven 'care delivery pathway' (CDP) to group several terms we will find in the selected studies to designate retrospective trajectories obtained from EHD. To describe the methods proposed for synthetically displaying objective measures or assessments of health status or healthcare utilization (e.g., quantifying) and graphically showing the temporal elements of chronic CDP (e.g., visualizing), we will assess how they addressed three domains:

1) The selection of relevant clinical and health-related events.

This domain will examine how the methods define health status and evaluate disease progression or stabilization, and how they show transitions between health status and acute manifestations[13]. Usually, the trajectory timeline begins at diagnosis and involves more than one provider[13–15]. Treatment decisions are generally based on health status (indicated by biomarkers, clinical examination, self-declared levels of quality of life, etc.), care units and settings, treatment availability (medication, procedures, etc.), and patient-provider preferences[20].

2) The technological development itself and considering issues related to data quality and exchange.

This domain aims to describe how the method is built, which data sources and analyses are used, and the necessary infrastructure surrounding its implementation. Digitalization of health-related data is a global trend[23,24] and highly detailed data are being collected daily in diverse settings and healthcare services. Such methods may apply a range of techniques from basic algorithms to advanced statistical and machine learning models[25], which can provide useful insights into care delivery processes. Technological developments in this field also need to meet strict criteria of data security, accuracy of models and predictions, openness of development and validation processes, among others[15,26].

3) Considering behaviours of actors and interactions between them with the aim of effectively improving care delivery.

Integrated care depends on multiple actions and decisions made collaboratively by patients, healthcare providers, administrative staff and other actors concerning patients' course of treatment[27]. To inform these decisions, technological solutions must have access to clinical exams and provide key actors with relevant information, such as the patients' past interactions with other providers, the medical procedures performed, the medications prescribed[28]. To have a positive impact on improving care delivery, visualizations and quantitative indicators of the patient's prior care need to be adapted to the user's needs at specific points in the trajectory, like after acute events or hospitalizations. This domain will examine what behaviors and interactions the methods promote (who are its target individuals, what actions need to be performed, in what context, when, and by whom)[29,30], and what strategies are proposed to encourage this performance.

Aims and objectives

We aim to identify and describe the methods that have been proposed to quantify and/or visualize data-driven CPDs of people living with chronic conditions. Given the complexity of their context of use, more than only reviewing technical methods, we aim to investigate how these tools have considered the three domains described above.

For this end, we propose the following research questions:

1. What clinical information does the method use and how was it considered relevant?
2. What are the method's development and implementation characteristics?
3. Which behaviours and interactions does the method aim to promote among users and how?

Methods

The Cochrane Handbook[31] and the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P)[32] were used to write this protocol and the systematic review will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[33]. PRISMA-P checklist is presented in Supplementary File 1. The review will be performed by one primary reviewer (LSP) and three secondary reviewers (AD, MV and SA) and will follow 6 steps: literature search, records screening and pre-selection (title and abstract), full-text screening and final selection, extraction of data, quality assessment, analysis and synthesis of data.

The studies expected to be analyzed in this work will likely be descriptive and not follow standard methodology (i.e., experimental or observational, method validation), yet considering the manuscripts as a qualitative corpus allows for coding the narratives according to the conceptual structure we propose[7,34]. Content analysis has been used in many studies in health sciences[35] and an inductive content analysis applied in a systematic review of clinical information modeling processes[34] has developed descriptive categories in a context similar to the one we propose here. As we consider them relevant to the studies we will review, they will be included in our coding framework, as detailed below.

Searches

A literature search will be performed in the following electronic databases: PubMed (MEDLINE), Scopus, IEEE, CINAHL and EMBASE. The search will be adapted to each database, and the resulting search strategies are provided as Supplementary File 2. The terms searched will be related to three main categories, connected by the AND operator: "data-driven" (MeSH terms like "Electronic health record", "data mining", etc.), "clinical pathways" (MeSH terms like "clinical pathway", "disease management", etc.), and "chronic conditions" (MeSH term "chronic diseases"). Searches will be

performed with MeSH terms or with keywords in Title/Abstract in PubMed; MeSH terms will be adapted for the databases that do not permit their usage or use different indexed terms. Bibliographies and citation tracking of relevant literature will be hand searched to identify additional relevant studies. A first selection will be performed using abstracts and titles, followed by full-text examination of entries selected.

Types of publications/studies and eligibility criteria

We will consider CDPs to be a series of time-stamped events describing the sequence of care of users with a diagnosed chronic condition (conditions requiring medical attention for a period longer than 12 months)[36]. These events can be the diagnosis itself, routine, non-scheduled or emergency consultations with a general practitioner and/or specialist, therapeutic education sessions and other health-related interventions. These can result in prescriptions of medications, medical procedures and tests, which may also appear in the trajectory. Data-driven CDP analyzed here will need to be composed of at least two time-stamped events recorded in EHD from people with the diagnosis of a chronic condition, with no duration restrictions (e.g., CDP may cover periods from days or few months to several years).

We will consider peer-reviewed publications (1) reporting methods for visualization or quantification of data-driven chronic CDP (including protocols and reports of study results), (2) using data from people living with chronic conditions retrieved from EHD and (3) published in English. No restrictions on publication date, study design, population characteristics, type of healthcare facility and level of care will be applied.

We will exclude studies that aim only to assess healthcare utilization over a specific period as part of a single research study, for example as an outcome to evaluate health-related interventions, to describe populations or disease prevalence, or as a proxy measure of disease aggravation risk. We will also exclude studies that do not mention population or data characteristics or do not state they analyze data from people living with chronic conditions, unavailable full texts, papers not written in English, conference abstracts or abstract-only papers, systematic or narrative reviews, meta-analyses and grey literature.

Screening

We will use Covidence, an online systematic review management software, for records screening. After duplicates removal, titles and abstracts in the remaining records will be screened independently by two reviewers for full text appraisal. If reviewer discordance arises, consensus will be reached through discussion and arbitration with one of the secondary reviewers not involved in the selection of the record. Studies selected in the first step will go through full text screening using the same process to establish eligibility. Inter-rater reliability (Cohen's Kappa) between primary and secondary reviewers will be computed and reported, values greater than 0.80 will be considered adequate.

Data management

We will report the number of included and excluded articles as well as the number of full-text papers obtained and assessed. Reasons for exclusion of screened full-text studies will also be stated in the final review. The data will be managed using Covidence and Microsoft Excel spreadsheets.

Data extraction and analysis

We will use both deductive and inductive content analysis[35] to appraise the selected studies: deductive when relying on pre-defined frameworks such as the categories previously described by Moreno-Conde et al. (2015)[34] to describe the technical characteristics of the proposed solutions and on the AACTT framework[29,30] (action, actor, context, target, time), to describe the behavioural domain, and inductive when additional relevant characteristics need to be described.

Data from included studies will be extracted using a customized electronic data extraction form. Information on study characteristics (authors, title, type of study, year and country of study, objective and research questions); population characteristics (number of patients, age, gender, condition) will be extracted directly from the included studies.

Deductive-Inductive content analysis

We will perform a deductive content analysis following existing theories, as described below, and inductive analysis for observed relevant characteristics not yet covered by existing literature. If more than one selected record describe development, validation and/or implementation of the same method, we will extract basic paper characteristics, as described above, but the content analysis will be performed per method.

1) For the clinical domain, we will extract information on clinical or cost outcomes the method might target (if reported and which ones) and on how the outcomes were considered relevant (e.g., involving experts, final users or other stakeholders).

2) For method development and data processing, we will analyze and compare to what has been proposed by Moreno-Conde et al. (2015)[34]. The categories detailed in the study are briefly described below.

- Scope definition leading to selection of the domain and selecting relevant experts: identifying the domain and expected uses of the method through the creation of a group of experts.
- Analysis of the information covered in the specific domain: creation of definitions, identification of clinical scenarios, workflows, users, guidelines, literature, etc., so the method meet the requirements of clinical practice or other intended usages.
- Design of the tool: detailing the set of attributes associated with the method, domain terminologies, ensuring compatibility across domains.
- Definition of implementable tool specifications: description of implementable technical specification.
- Validation: use of techniques to validate the method, such as peer-review validation or creation of prototype screens.
- Publishing and maintenance: availability in public repositories.
- Governance: description of the organization responsible for developing and maintaining the tool.

Other information extracted from studies regarding this domain will be healthcare utilization characteristics (type of event, e.g., consultation, test, procedure) and data characteristics (sources of data, data preparation, data analysis).

3) To describe behaviour and interactions the method might promote or facilitate, we will apply the AACTT[29,30] framework. Other information extracted from studies will be output characteristics like intended final users, purpose and use scenarios. We will also code the presence of strategies planned

or performed to achieve these behavioural change objectives, such as training, organizational changes, evaluation of the performance of the method in routine care, if implemented, and other initiatives studies might present.

The primary reviewer and one secondary reviewer will pilot data extraction independently for a subset of 10% of selected records to compare and discuss data extraction process. If necessary, we will repeat the pilot extraction process (outlined above) until agreement is reached. Inter-rater reliability (Cohen's Kappa) will be computed, and values greater than 0.80 will be considered adequate. Disagreements will be solved with the help of a third reviewer and piloting may consist of several interactions between reviewers to compare and reach consensus regarding relevant information to be extracted from full-text analysis. After this first step, a codebook will be developed, and data extraction of the remaining records will be performed by the primary reviewer.

Quality and bias assessment

As most quality assessment tools are developed for commonly-used study designs and there is no consensus regarding tools for generic use, we propose to evaluate quality from a different perspective. We will evaluate if main stakeholders (patients and/or family, healthcare professionals, administrative personnel) were involved at any stage of the development of the method. Research shows the importance of involving patients, the public and other stakeholders in health-related research to obtain experiential knowledge, setting research priorities and focus on practical questions[37–40]. Also, it has been shown that trials funded by for-profit organizations can positively bias interpretation of trial results[41], and research in data usage can be funded by companies interested in selling their own methods. To assess potential bias, we will evaluate declared conflicts of interest and sources of funding. Quality assessment will be discussed in the review, but no study will be excluded from the analysis based on quality criteria.

Data synthesis

The technical methods will be synthesized using the content analysis described above and the studies will be categorized and described using the 3 domains, depending on study type and reporting. We will present the results in tables along with method and study identification and summarize via descriptive statistics. We will compare the different characteristics within the 3 domains to identify common, infrequent, or missing features of these tools, and extract recommendations for future initiatives.

Patient and public involvement

A representant of a patients' association was involved in reading and approving of this protocol. This systematic review is part of a larger project that will be developed closely with patients and healthcare providers.

Ethics and dissemination

The search strategy was developed in collaboration with health sciences librarian services in early 2019. Database searches will be initiated in May 2019. The review is expected to be completed by February 2020. Ethical approval is not required. Results will be disseminated in peer-reviewed journals and/or conference presentations. Data used in this review will be made available through supplementary materials and open trusted repositories.

Funding

DD was supported by an IDEXLYON (16-IDEX-0005) Fellowship grant (2018-2021), LSP was supported by a PhD funding within the same grant, AD by a Marie Curie Individual Fellowship from the European Commission (MCRA-IF n°706028) during the preparation of this review protocol, and SA by the Swiss Science Foundation (P2BSP3_ 178648).

Conflicts of interest

None declared.

Author statement

LSP, AD and SA designed the protocol and planned data extraction and quality assessment. LSP put together the search strategy and SA helped adapt it to the different databases. LSP and AD conceived the content analysis stages and conceptual framework. LSP wrote the first version of the manuscript, AD extensively reviewed it, SA, DD, AMS and MV revised it critically for important intellectual content. All authors have approved the publication of this protocol and contributed to the final manuscript.

References

1. Bodenheimer, T., Wagner, E. H. & Grumbach, K. Improving Primary Care for Patients With Chronic Illness: The Chronic Care Model, Part 2. *JAMA* **288**, 1909–1914 (2002).

2. Ouwens, M., Wollersheim, H., Hermens, R., Hulscher, M. & Grol, R. Integrated care programmes for chronically ill patients: a review of systematic reviews. *Int J Qual Health Care* **17**, 141–146 (2005).

3. Bodenheimer, T., Wagner, E. H. & Grumbach, K. Improving Primary Care for Patients With Chronic Illness. *JAMA* **288**, 1775–1779 (2002).

4. Pham, H. H., Schrag, D., O'Malley, A. S., Wu, B. & Bach, P. B. Care Patterns in Medicare and Their Implications for Pay for Performance. doi:10.1056/NEJMsa063979.

5. Wagner, E. H. Meeting the needs of chronically ill people: Socioeconomic factors, disabilities, and comorbid conditions are obstacles. *BMJ* **323**, 945–946 (2001).

6. Campbell, H., Hotchkiss, R., Bradshaw, N. & Porteous, M. Integrated care pathways. *BMJ* **316**, 133–137 (1998).

7. Gooch, P. & Roudsari, A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *J Am Med Inform Assoc* **18**, 738–748 (2011).

8. Huang, Z. *et al.* Discovery of clinical pathway patterns from event logs using probabilistic topic models. *Journal of Biomedical Informatics* **47**, 39–57 (2014).

9. Perer, A., Wang, F. & Hu, J. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics* **56**, 369–378 (2015).

10. Gotz, D., Wang, F. & Perer, A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *Journal of Biomedical Informatics* **48**, 148–159 (2014).

11. Huang, Z., Lu, X., Duan, H. & Fan, W. Summarizing clinical pathways from event logs. *Journal of Biomedical Informatics* **46**, 111–127 (2013).

12. Yang, W. & Su, Q. Process mining for clinical pathway: Literature review and future directions. in *2014 11th International Conference on Service Systems and Service Management (ICSSSM)* 1–5 (2014). doi:10.1109/ICSSSM.2014.6943412.

13. Jensen, A. B. *et al.* Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature Communications* **5**, 4022 (2014).

14. Zhang, Y., Padman, R. & Patel, N. Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. *J Biomed Inform* **58**, 186–197 (2015).

15. Bettencourt-Silva, J. H. *et al.* Building Data-Driven Pathways From Routinely Collected Hospital Data: A Case Study on Prostate Cancer. *JMIR Med Inform* **3**, e26 (2015).

16. Y. Zhang *et al.* On Clinical Pathway Discovery from Electronic Health Record Data. *IEEE Intelligent Systems* **30**, 70–75 (2015).

17. Zhang, Y. & Padman, R. Data-driven clinical and cost pathways for chronic care delivery. *Am J Manag Care* **22**, 816–820 (2016).

18. Bettencourt-Silva, J., De La Iglesia, B., Donell, S. & Rayward-Smith, V. On creating a patient-centric database from multiple Hospital Information Systems. *Methods Inf Med* **51**, 210–220 (2012).

19. Denaxas, S. C. & Morley, K. I. Big biomedical data and cardiovascular disease research: opportunities and challenges. *Eur Heart J Qual Care Clin Outcomes* **1**, 9–16 (2015).

20. Vanasse, A., Courteau, M. & Ethier, J.-F. The '6W' multidimensional model of care trajectories for patients with chronic ambulatory care sensitive conditions and hospital readmissions. *Public Health* **157**, 53–61 (2018).

21. Bleser, L. D. *et al.* Defining pathways. *Journal of Nursing Management* **14**, 553–563 (2006).

22. Kinsman, L., Rotter, T., James, E., Snow, P. & Willis, J. What is a clinical pathway? Development of a definition to inform the debate. *BMC Medicine* **8**, 31 (2010).

23. Adler-Milstein, J. *et al.* Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. *Health Affairs* **34**, 2174–2180 (2015).

24. Ben-Assuli, O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. *Health Policy* **119**, 287–297 (2015).

25. Zhang, Y. & Padman, R. Innovations in chronic care delivery using data-driven clinical pathways. *Am J Manag Care* **21**, e661-668 (2015).

26. Liaw, S. T. *et al.* Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int J Med Inform* **82**, 10–24 (2013).

27. Peikes, D., Chen, A., Schore, J. & Brown, R. Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures Among Medicare Beneficiaries: 15 Randomized Trials. *JAMA* **301**, 603–618 (2009).

28. Stille, C. J., Jerant, A., Bell, D., Meltzer, D. & Elmore, J. G. Coordinating care across diseases, settings, and clinicians: a key role for the generalist in practice. *Ann. Intern. Med.* **142**, 700–708 (2005).

29. Francis, J. J. & Presseau, J. Health Care Practitioner Behaviour. in *Cambridge Handbook of Psychology, Health and Medicine* (eds. Llewellyn, C. et al.) 324–324 (Cambridge University Press, 2019).

30. Presseau, J. *et al.* Action, actor, context, target, time (AACTT): a framework for specifying behaviour. *Implementation Sci* **14**, 1–13 (2019).

31. Cochrane Handbook for Systematic Reviews of Interventions. http://handbook-5-1.cochrane.org/.

32. Moher, D. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* **4**, 1 (2015).

33. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **6**, e1000097 (2009).

34. Moreno-Conde, A. *et al.* Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. *J Am Med Inform Assoc* **22**, 925–934 (2015).

35. Elo, S. & Kyngäs, H. The qualitative content analysis process. *Journal of Advanced Nursing* **62**, 107–115 (2008).

36. Goodman, R. A. Defining and Measuring Chronic Conditions: Imperatives for Research, Policy, Program, and Practice. *Prev Chronic Dis* **10**, (2013).

37. Staley, K. 'Is it worth doing?' Measuring the impact of patient and public involvement in research. *Research Involvement and Engagement* **1**, 6 (2015).

38. Hoffman, A., Montgomery, R., Aubry, W. & Tunis, S. R. How Best To Engage Patients, Doctors, And Other Stakeholders In Designing Comparative Effectiveness Studies. *Health Affairs* **29**, 1834–1841 (2010).

39. Selby, J. V., Beal, A. C. & Frank, L. The Patient-Centered Outcomes Research Institute (PCORI) National Priorities for Research and Initial Research Agenda. *JAMA* **307**, 1583–1584 (2012).

40. Ocloo, J. & Matthews, R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf* **25**, 626–632 (2016).

41. Als-Nielsen, B., Chen, W., Gluud, C. & Kjaergard, L. L. Association of Funding and Conclusions in Randomized Drug Trials: A Reflection of Treatment Effect or Adverse Events? *JAMA* **290**, 921–928 (2003).

Reporting checklist for protocol of a systematic review.

Based on the PRISMA-P guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the PRISMA-P reporting guidelines, and cite them as:

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1.

		Reporting Item	Page Number
Title			
Identification	#1a	Identify the report as a protocol of a systematic review	1
Update	#1b	If the protocol is for an update of a previous systematic review, identify as such	n/a
Registration			
	#2	If registered, provide the name of the registry (such as PROSPERO) and registration number	1
Authors			
Contact	#3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contribution	#3b	Describe contributions of protocol authors and identify the	7

		guarantor of the review	
Amendments			
	#4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support			
Sources	#5a	Indicate sources of financial or other support for the review	7
Sponsor	#5b	Provide name for the review funder and / or sponsor	7
Role of sponsor or funder	#5c	Describe roles of funder(s), sponsor(s), and / or institution(s), if any, in developing the protocol	n/a
Introduction			
Rationale	#6	Describe the rationale for the review in the context of what is already known	2,3
Objectives	#7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	4
Methods			
Eligibility criteria	#8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	4,5
Information sources	#9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	4
Search strategy	#10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	4
Study records - data management	#11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	5
Study records -	#11b	State the process that will be used for selecting studies (such	5

selection process		as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	
Study records - data collection process	#11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	5,6
Data items	#12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	5,6
Outcomes and prioritization	#13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	n/a
Risk of bias in individual studies	#14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	6
Data synthesis	#15a	Describe criteria under which study data will be quantitatively synthesised	6
Data synthesis	#15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I2, Kendall's τ)	n/a
Data synthesis	#15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
Data synthesis	#15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	#16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	n/a
Confidence in cumulative evidence	#17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

COMPLETE SEARCH STRATEGY

Category	Medical Subject headings (MeSH)	CINAHL Search	EMBASE	Keywords
Data-driven	Electronic health record; data mining; machine learning; clinical decision support systems; analysis, cluster; medical informatics application	MH Electronic Health Records OR MH Data Mining OR MH Nursing Informatics OR MH Machine Learning OR MH Decision Support Systems, Clinical OR MH Cluster Analysis OR MH Medical Informatics OR MH Computer Graphics OR MH Algorithms OR TI Data-driven OR TI visualisation OR TI computer graphics OR TI process mining OR TI data mining OR TI visualization OR TI supervised learning OR TI unsupervised learning OR TI practice based OR TI modelling OR TI mapping OR TI cluster* OR TI data analys* OR AB Data-driven OR AB visualisation OR AB computer graphics OR AB process mining OR AB data mining OR AB visualization OR AB supervised learning	(Electronic health record or data mining or machine learning or clinical decision support systems or analysis, cluster or medical informatics application).sh. or (Data-driven or visualisation or computer graphics or process mining or data mining or visualization or supervised learning or unsupervised learning or practice based or modelling or mapping or cluster* or data analys*).ti. or (Data-driven or visualisation or computer graphics or process mining or data mining or visualization or supervised learning or unsupervised learning or practice based or modelling or mapping or cluster* or data analys*).ab.	Data-driven OR visualisation OR computer graphics OR process mining OR data mining OR visualization OR supervised learning OR unsupervised learning OR practice based OR modelling OR mapping OR cluster* OR data analys*

		OR AB unsupervised learning OR AB practice based OR AB modelling OR AB mapping OR AB cluster* OR AB data analys*		
Clinical pathways	Clinical pathways; delivery of health care, integrated; clinical practice pattern; disease management; care management, patient	MH Critical Path OR MH Health Care Delivery, Integrated OR MH Practice Patterns OR MH Disease Management OR MH Patient Care Plans OR TI Clinical course OR TI integrated care OR TI care map OR TI care pathway OR TI care plan OR TI treatment plan OR TI patient journey OR TI patient flow OR TI clinical redesign OR TI integrated care OR AB Clinical course OR AB integrated care OR AB care map OR AB care pathway OR AB care plan OR AB treatment plan OR AB patient journey OR AB patient flow OR AB clinical redesign OR AB integrated care	(Clinical pathways or delivery of health care, integrated or clinical practice pattern or disease management or care management, patient).sh. or (Clinical course or integrated care or care map or care pathway or care plan or treatment plan or patient journey or patient flow or clinical redesign or integrated care).ti. or (Clinical course or integrated care or care map or care pathway or care plan or treatment plan or patient journey or patient flow or clinical redesign or integrated care).ab.	Clinical course OR integrated care OR care map OR care pathway OR care plan OR treatment plan OR patient journey OR patient flow OR clinical redesign OR integrated care

Chronic conditions	Chronic diseases; chronic illness	MH Chronic Disease OR TI Integrated chronic care OR AB Integrated chronic care		Integrated chronic care

EQUATIONS

MEDLINE

((data-driven[Title/Abstract] OR health information[Title/Abstract] OR data analys*[Title/Abstract] OR computer graphics[MeSH Terms] OR visualization[Title/Abstract] OR machine learning[MeSH Terms] OR data mining[MeSH Terms] OR clinical decision support systems[MeSH Terms] OR medical informatics application[MeSH Terms] OR algorithm[MeSH Terms] OR supervised learning[Title/Abstract] OR unsupervised learning[Title/Abstract] OR analysis, cluster[MeSH Terms] OR practice-based[Title/Abstract] OR electronic health record[MeSH Terms] OR clinical decision support systems[Title/Abstract] OR process mining[Title/Abstract] OR data mining [Title/Abstract] OR machine learning [Title/Abstract] OR medical informatics application[Title/Abstract] OR cluster*[Title/Abstract] OR modeling[Title/Abstract] OR mapping[Title/Abstract])

AND

(chronic diseases[MeSH Terms] OR chronic illness[MeSH Terms] OR integrated chronic care[Title/Abstract])

AND

(delivery of health care, integrated[MeSH Terms] OR clinical practice pattern[MeSH Terms] OR clinical pathway[MeSH Terms] OR critical pathway[MeSH Terms] OR clinical course[Title/Abstract] OR integrated care[Title/Abstract] OR care map[Title/Abstract] OR care pathway[Title/Abstract] OR care plan[Title/Abstract] OR treatment plan[Title/Abstract] OR disease management[MeSH Terms] OR disease management[Title/Abstract] OR care management, patient[MeSH Terms] OR patient journey[Title/Abstract] OR patient flow[Title/Abstract] OR clinical redesign[Title/Abstract] OR integrated care[Title/Abstract]))

SCOPUS

(TITLE-ABS-KEY(“data-driven”) OR TITLE-ABS-KEY("health information") OR TITLE-ABS-KEY("data analys*") OR INDEXTERMS("computer graphics") OR TITLE-ABS-KEY(“visuali*ation”) OR INDEXTERMS("machine learning") OR INDEXTERMS("data mining") OR INDEXTERMS("clinical decision support systems") OR INDEXTERMS("medical informatics application") OR INDEXTERMS("algorithm") OR TITLE-ABS-KEY("supervised learning") OR TITLE-ABS-KEY("unsupervised learning") OR INDEXTERMS("cluster analysis") OR INDEXTERMS(“practice-based”) OR INDEXTERMS("electronic health record") OR TITLE-ABS-KEY("clinical decision support systems") OR TITLE-ABS-KEY("process mining") OR TITLE-ABS-KEY("data mining") OR TITLE-ABS-KEY("machine learning") OR TITLE-ABS-KEY("medical informatics application") OR TITLE-ABS-KEY(cluster*) OR TITLE-ABS-KEY("modelling") OR TITLE-ABS-KEY("mapping"))

AND (INDEXTERMS("chronic diseases") OR INDEXTERMS("chronic illness") OR TITLE-ABS-KEY("integrated chronic care"))

AND (INDEXTERMS("integrated delivery of health care") OR TITLE-ABS-KEY("clinical practice pattern") OR INDEXTERMS("clinical pathway") OR INDEXTERMS("critical pathway") OR TITLE-ABS-KEY("clinical course") OR TITLE-ABS-KEY("integrated care") OR TITLE-ABS-KEY("care map") OR TITLE-ABS-KEY("care pathway") OR TITLE-ABS-KEY("care plan") OR TITLE-ABS-KEY("treatment plan") OR INDEXTERMS ("disease management") OR TITLE-ABS-KEY("disease management") OR INDEXTERMS("patient care management") OR TITLE-ABS-KEY("patient journey") OR TITLE-ABS-KEY("patient flow") OR TITLE-ABS-KEY("clinical redesign") OR TITLE-ABS-KEY("integrated care"))

IEEE

((((((((((((((((((((((((((("Index Terms":electronic health records) OR "Index Terms":data mining) OR "Index Terms":machine learning) OR "Index Terms":clinical decision support systems) OR "Index Terms":cluster analysis) OR "Index Terms":medical informatics) OR "Index Terms":computer graphics) OR "Index Terms":algorithm) OR "IEEE Terms":medical information systems) OR "IEEE Terms":electronic medical records) OR "Author Keywords":healthcare practices) OR "Publication Title":data-driven) OR "Abstract":data-driven) OR "Publication Title":machine learning) OR "Abstract":machine learning) OR "Publication Title":cluster analys*) OR "Abstract":cluster analys*) OR "Publication Title":data mining) OR "Abstract":data mining) OR "Author Keywords":electronic health record) OR "IEEE Terms":Guidelines) OR "IEEE Terms":Data mining) OR "IEEE Terms":Algorithm design and analysis) OR data mining) OR data-driven) OR electronic health record) OR algorithm) OR visualization) OR clustering) OR algorithm)

AND ((((((((((("Index Terms":clinical pathway) OR "Publication Title":clinical pathway) OR "Abstract":clinical pathway) OR"Author Keywords ":clinical pathway) OR "Author Keywords":healthcare practices) OR"Author Keywords ":Pathway) OR clinical path*) OR care pattern) OR care plan) OR care map) OR critical path*)

AND (((((("Index Terms":chronic disease) OR "Publication Title":chronic disease) OR "Abstract":chronic disease) OR "IEEE Terms":Diseases) OR chronic*)