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

Luiza Siqueira Do Prado, Samuel Allemann, Marie Viprey, Anne-Marie Schott, Dan Dediu, Alexandra Dima

BMJ Open

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

Luiza Siqueira do Prado ,[1] Samuel Allemann,[1,2] Marie Viprey,[1,3] Anne-Marie Schott,[1,3] Dan Dediu,[4] Alexandra L Dima[1]

To cite: Siqueira do Prado L, Allemann S, Viprey M, *et al*. Quantification and visualisation methods of data-driven chronic care delivery pathways: protocol for a systematic review and content analysis. *BMJ Open* 2020;**10**:e033573. doi:10.1136/bmjopen-2019-033573

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-033573).

Received 11 August 2019
Revised 10 February 2020
Accepted 11 February 2020

[1]Health Services and Performance Research EA 7425, Université Claude Bernard Lyon 1, Lyon, France
[2]Pharmaceutical Care Research Group, University of Basel, Basel, Switzerland
[3]Pôle Santé Publique, Hospices Civils de Lyon, Lyon, France
[4]Laboratoire Dynamique du Langage UMR 5596, Université Lumière Lyon 2, Lyon, France

Correspondence to
Luiza Siqueira do Prado;
luiza.siqueira-do-prado@univ-lyon1.fr

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 utilisation 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 visualise 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. 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 visualisation methods. We will describe them using several characteristics relevant for EHD use in long-term care, grouped into three domains: (1) clinical (what clinical information does the method use and how was it considered relevant?), (2) data science (what are the method's development and implementation characteristics?) and (3) behavioural (which behaviours and interactions does the method aim to promote among users and how?). 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.

Strengths and limitations of this study

- 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.

PROSPERO registration number CRD42019140494.

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] Centralising 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 utilisation events and related medical data.[8–17] For example, Zhang *et al* have produced longitudinal trajectories using electronic health records 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* have reported on the development of a patient-centric database from multiple Hospital Information Systems[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 analysing 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 utilisation (eg, quantifying) and graphically showing the temporal elements of chronic CDP (eg, visualising), we will assess how they addressed the following three domains.

The selection of relevant clinical and health-related events

This domain will examine how the methods define health status and evaluate disease progression or stabilisation, 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]

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. Digitalisation 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]

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, visualisations 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 hospitalisations. This domain will examine what behaviours 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 visualise data-driven CDPs 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 PRISMA.[33] PRISMA-P checklist is presented in online supplementary file 1. The review will be performed by one primary reviewer (LSP)

and three secondary reviewers (ALD, MV and SA) and will follow six 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 analysed in this work will likely be descriptive and not follow standard methodology (ie, 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 modelling 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 online 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 analysed 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 (eg, CDP may cover periods from days or few months to several years).

We will consider peer-reviewed publications (1) reporting methods for visualisation 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 utilisation 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 analyse 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*[34] to describe the technical characteristics of the proposed solutions and on the Action, Actor, Context, Target, Time (AACTT) framework[29 30], to describe the behavioural domain and inductive when additional relevant characteristics need to be described.

Data from included studies will be extracted using a customised 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 (eg, involving experts, final users or other stakeholders).
2. For method development and data processing, we will analyse and compare to what has been proposed by Moreno-Conde *et al*.[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 meets 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 organisation responsible for developing and maintaining the tool.

Other information extracted from studies regarding this domain will be healthcare utilisation characteristics (type of event, for example, 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, organisational 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 organisations 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 synthesised using the content analysis described above and the studies will be categorised and described using the three domains, depending on study type and reporting. We will present the results in tables along with method and study identification and summarise via descriptive statistics. We will compare the different characteristics within the three 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 online supplementary materials and open trusted repositories.

Contributors LSP, ALD 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 ALD conceived the content analysis stages and conceptual framework. LSP wrote the first version of the manuscript, ALD extensively reviewed it, SA, DD, A-MS and MV revised it critically for important intellectual content. All authors have approved the publication of this protocol and contributed to the final manuscript.

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, ALD 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).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Luiza Siqueira do Prado http://orcid.org/0000-0001-8134-6246

REFERENCES

1 Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, part 2. *JAMA* 2002;288:1909–14.
2 Ouwens M, Wollersheim H, Hermens R, *et al*. Integrated care programmes for chronically ill patients: a review of systematic reviews. *Int J Qual Health Care* 2005;17:141–6.
3 Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. *JAMA* 2002;288:1775–9.
4 Pham HH, Schrag D, O'Malley AS, *et al*. Care patterns in Medicare and their implications for pay for performance. *N Engl J Med* 2007;356:1130–9.
5 Wagner EH. Meeting the needs of chronically ill people: socioeconomic factors, disabilities, and comorbid conditions are obstacles. *BMJ* 2001;323:945–6.
6 Campbell H, Hotchkiss R, Bradshaw N, *et al*. Integrated care pathways. *BMJ* 1998;316:133–7.
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* 2011;18:738–48.
8 Huang Z, Dong W, Ji L, *et al*. Discovery of clinical pathway patterns from event logs using probabilistic topic models. *J Biomed Inform* 2014;47:39–57.
9 Perer A, Wang F, Hu J. Mining and exploring care pathways from electronic medical records with visual analytics. *J Biomed Inform* 2015;56:369–78.
10 Gotz D, Wang F, Perer A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *J Biomed Inform* 2014;48:148–59.
11 Huang Z, Lu X, Duan H, *et al*. Summarizing clinical pathways from event logs. *J Biomed Inform* 2013;46:111–27.
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*, 2014: 1–5.
13 Jensen AB, Moseley PL, Oprea TI, *et al*. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun* 2014;5:4022.
14 Zhang Y, Padman R, Patel N. Paving the COWpath: learning and visualizing clinical pathways from electronic health record data. *J Biomed Inform* 2015;58:186–97.
15 Bettencourt-Silva JH, Clark J, Cooper CS, *et al*. Building data-driven pathways from routinely collected hospital data: a case study on prostate cancer. *JMIR Med Inform* 2015;3:e26.
16 Zhang Y, Padman R, Wasserman L, *et al*. On clinical pathway discovery from electronic health record data. *IEEE Intell Syst* 2015;30:70–5.
17 Zhang Y, Padman R. Data-Driven clinical and cost pathways for chronic care delivery. *Am J Manag Care* 2016;22:816–20.
18 Bettencourt-Silva J, De La Iglesia B, Donell S, *et al*. On creating a patient-centric database from multiple Hospital information systems. *Methods Inf Med* 2012;51:210–20.
19 Denaxas SC, Morley KI. Big biomedical data and cardiovascular disease research: opportunities and challenges. *Eur Heart J Qual Care Clin Outcomes* 2015;1:9–16.
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* 2018;157:53–61.
21 De Bleser L, Depreitere R, De Waele K, *et al*. Defining pathways. *J Nurs Manag* 2006;14:553–63.
22 Kinsman L, Rotter T, James E, *et al*. What is a clinical pathway? Development of a definition to inform the debate. *BMC Med* 2010;8:31.
23 Adler-Milstein J, DesRoches CM, Kralovec P, *et al*. Electronic health record adoption in US hospitals: progress continues, but challenges persist. *Health Aff* 2015;34:2174–80.
24 Ben-Assuli O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. *Health Policy* 2015;119:287–97.
25 Zhang Y, Padman R. Innovations in chronic care delivery using data-driven clinical pathways. *Am J Manag Care* 2015;21:e661–8.
26 Liaw ST, Rahimi A, Ray P, *et al*. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int J Med Inform* 2013;82:10–24.
27 Peikes D, Chen A, Schore J, *et al*. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries: 15 randomized trials. *JAMA* 2009;301:603–18.
28 Stille CJ, Jerant A, Bell D, *et al*. Coordinating care across diseases, settings, and clinicians: a key role for the generalist in practice. *Ann Intern Med* 2005;142:700–8.
29 Francis JJ, Presseau J. Health care practitioner behaviour. In: Llewellyn C, ed. *Cambridge handbook of psychology, health and medicine*. Cambridge University Press, 2019: 324.
30 Presseau J, McCleary N, Lorencatto F, *et al*. Action, actor, context, target, time (AACTT): a framework for specifying behaviour. *Implement Sci* 2019;14:1–13.
31 Higgins JPT, Green S. Cochrane Handbook for systematic reviews of interventions. Available: http://handbook-5-1.cochrane.org/
32 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
33 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
34 Moreno-Conde A, Moner D, Cruz WDda, *et al*. Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. *J Am Med Inform Assoc* 2015;22:925–34.
35 Elo S, Kyngäs H. The qualitative content analysis process. *J Adv Nurs* 2008;62:107–15.
36 Goodman RA, Posner SF, Huang ES, *et al*. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. *Prev Chronic Dis* 2013;10:E66.
37 Staley K. 'Is it worth doing?' measuring the impact of patient and public involvement in research. *Res Involv Engagem* 2015;1:6.
38 Hoffman A, Montgomery R, Aubry W, *et al*. How best to engage patients, doctors, and other stakeholders in designing comparative effectiveness studies. *Health Aff* 2010;29:1834–41.
39 Selby JV, Beal AC, Frank L. The patient-centered outcomes research Institute (PCORI) national priorities for research and initial research agenda. *JAMA* 2012;307:1583–4.
40 Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf* 2016;25:626–32.
41 Als-Nielsen B, Chen W, Gluud C, *et al*. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921–8.