Patient narratives – a still undervalued resource for healthcare improvement

Giovanni Spitale

0000-0002-6812-0979

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

giovanni.spitale@ibme.uzh.ch

Andrea Glässel

0000-0002-9199-3604

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

ZHAW School of Health Sciences, Institute of Public Health (IPH), Winterthur, Switzerland

andrea.glaessel@ibme.uzh.ch

Mirriam Tyebally-Fang

0000-0002-2187-0641

mirriam.tyebally@ibme.uzh.ch

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

Corine Mouton Dorey

0000-0001-5513-1655

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

corine.moutondorey@ibme.uzh.ch

Nikola Biller-Andorno

0000-0001-7661-1324

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

biller-andorno@ibme.uzh.ch

Abstract

In recent years, patient narratives have gained increasing attention as a valuable source of insights into the subjective experience of healthcare. This paper outlines a best-practice approach to the collection, analysis, and use of patient narratives, based on current literature and on the experience of developing the Swiss Database of Individual Patient Experiences (DIPEx). The DIPEx project aims to provide a systematic and methodologically rigorous collection of patient narratives on various health situations and topics. This paper presents and details the DIPEx approach as a current standard in the field, offering a comprehensive overview and discussing the potential uses and benefits of patient narratives: improve healthcare practice, empower patients and caregivers, help structure better communication in healthcare, and contribute to medical teaching and learning.

Keywords

Introduction

Research on patient narratives has steadily gained momentum over the last decades. For good reasons: patient narratives are unique windows on the subjective aspects of the healthcare experience – which is a very relevant aspect of healthcare. Work on patient narratives can build empirical foundations for better patient-centred care; inform the development of Patient Reported Experience Measures (PREMs) and of Patient Reported Outcome Measures (PROMs); empower patients and caregivers; help structure better communication in healthcare; contribute to medical teaching and learning. In a nutshell, narratives help us understand the very meaning of health, illness, and care for the key stakeholders of healthcare: patients (Pallai and Tran 2019). Patient narratives can bring pack people's voices in patient centred care; they are evidence, as in 'evidence-based medicine', complementary – not alternative – to quantitative evidence. Nevertheless, they need to be collected and studied with method and rigour to contribute to designing better healthcare.

Individual patient accounts that appear more or less randomly in the medical literature and media may be appealing and insightful, but come with obvious limitations. They cannot capture variations, nor do they routinely provide detailed information about disease, diagnosis and treatment. A systematic collection, however, provides a clear methodology and transparent analysis that is amenable to critical scrutiny. Results are contextualized and discussed, including other qualitative studies as well as quantitative work, to increase validity and reduce potential bias.

In this article, based on current literature and on our hands-on experience of developing a Swiss Database of Individual Patient Experiences (DIPEx), we detail a standardized, best-practice approach to the collection, analysis, and use of patient narratives, in the form of semi-structured interviews – keeping in view that standards are currently evolving (Deutsches Netzwerk Gesundheitskompetenz 2022). The Swiss DIPEx project aims to provide a systematic and methodologically rigorous collection of patient narratives on various health situations, as diseases like dementia, chronic pain, Parkinson's, COVID-19, Multiple Sclerosis, rare diseases and others, but also on health-related topics, e.g. Intensive Care Management and Coercion, Pregnancy and prenatal testing, or possible future topics as adolescent obesity, risk of falls in older age, addictive behaviour among adolescents, or vocational reintegration at workplace.

Background

The value of qualitative research has been long recognized in public health and health services research. The focus on methods that encourage participation and provide an deep insight into the subjective experience, which is one of the strength of qualitative research, is now increasingly published and accepted in the mainstream medical literature: Lancet commissions incorporate patient voices (Han et al. 2016), and JAMA has a specific section on narratives (JAMA 2022). The Equator Network, an international initiative seeking "to improve the reliability and value of published health research literature by promoting transparent and accurate reporting", lists qualitative research as one of the main study types in health research (Equator Network 2022; O'Brien et al. 2014; Tong, Sainsbury, and Craig 2007).

Patient narratives can help close the gap between what really matters to the individual patient and what healthcare professionals perceive (Kleinman 1989). Beyond patient centeredness, other dimensions of healthcare quality can be improved. Targeted care, based on patients' needs and priorities, can be more effective and efficient, avoiding unnecessary expenses that do not add value for patients (Porter 2010). The saved resources can in turn be invested to deliver more timely and equitable care. Better understanding of patients' perspectives contributes also to patient safety, e.g., through adapting information procedures to patients' emotional state and current cognitive receptivity. By sharing their experiences and by learning from others with similar conditions, patients and their relatives can feel supported and more able to cope with their disease (Ziebland, Lavie-Ajayi, and Lucius-Hoene 2015).

Patient narratives can also contribute to medical education and teaching, fostering understanding and improving communication. In fact, the creation of a database for a learning system has been identified as one of the core challenges that Swiss healthcare faces today (Biller-Andorno and Zeltner 2015).

Distant reading

To understand how patient narratives have been incorporated in healthcare scholarly work we adopted a 'distant reading' strategy (Franzini et al. 2015; Moretti 2013, 49). We used TopicTracker (Spitale and Biller-Andorno 2021), a Python pipeline to retrieve PubMed entries and to perform Natural Language Processing analyses on the corpus. Our query is fairly simple – but rather specific:

> "1975/01/01"[Date - Publication] : "2021/12/31"[Date - Publication] AND ("patient s"[All Fields] OR "patients"[MeSH Terms] OR "patients"[All Fields] OR "patient"[All Fields] OR "patients s"[All Fields]) AND ("narration"[MeSH Terms] OR "narration"[All Fields] OR "narrative"[All Fields] OR "narratives"[All Fields] OR "narrative s"[All Fields] OR "narratively"[All Fields])

The query captures everything indexed in PubMed from 1975 (i.e. when the combination 'patient narratives' appeared for the first time in a paper's keywords) to 2021.

Our analysis is focused 1. on normalized keywords and MeSH terms to describe the evolution of the field and the main topics; and 2. on normalized journals (i.e: the normalized count of journals publishing this literature), to describe the impact, the target, and the typical audience of these publications. Normalization is performed in the same way for each entity, i.e: normalized entity = count of entity / number of papers. The original dataset is available for replication and further exploration (Spitale 2022).

Our query captures a total of 26739 papers (after duplicate removal).

Papers on 'patient narratives', 1975 - 2021, normalized
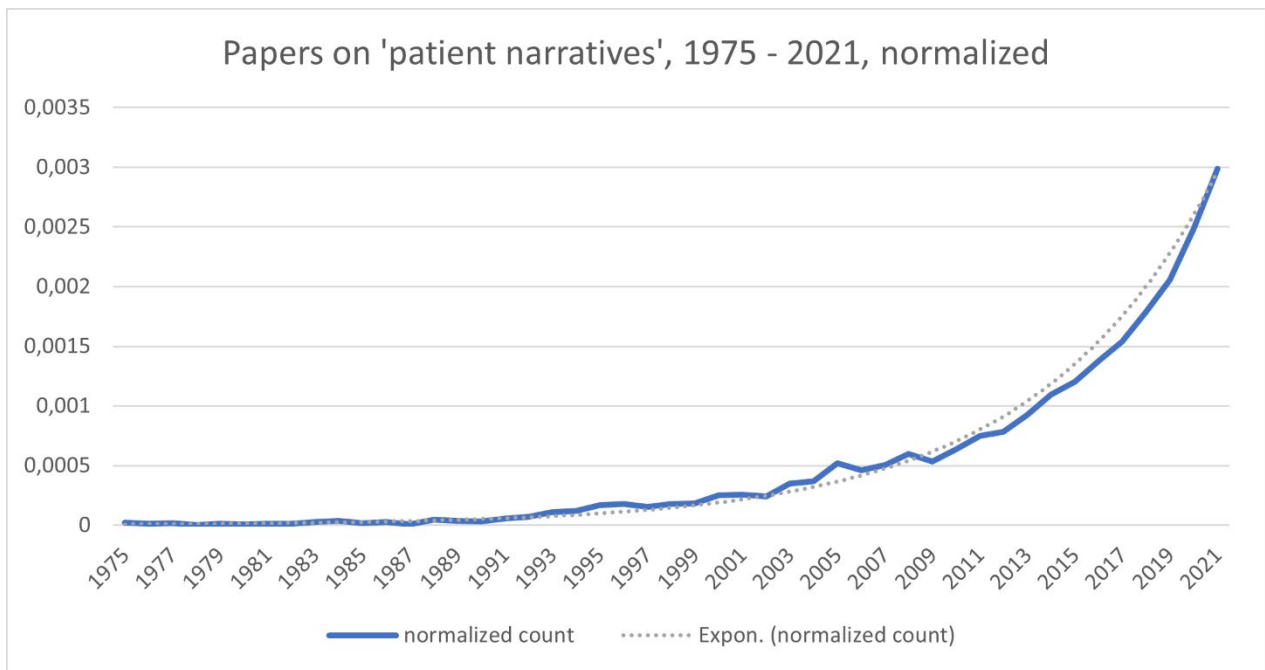
[figure 1, number of articles captured by our query on 'patient narratives', 1975 - 2021, normalized to total articles indexed in PubMed in the same year]

It is clear that the last decade saw the start of an exponential growth in publications dealing with 'patient narratives'. Of note, in 2021 0.3% of all the literature indexed in PubMed touches this topic.

Keywords

The analysis of the most frequent 50 normalized keywords offers a first overview of what this field is about in terms of methodologies, conditions targeted, and context of application of the findings. Keywords focus on:

1. The methodology and the approach, including: analytical approach, empirical approach, philosophical approach, religious approach, qualitative research, natural language processing, patient experiences.

2. The target conditions, including: death and euthanasia, mental health, cancer, genetics and reproduction, depression, dementia, stroke, pain / chronic pain, schizophrenia, diabetes, obesity, breast cancer. The target conditions remain rather stable over time.

3. The context in which this research can have a substantial impact, including: professional patient relationship, palliative care, quality of life, nursing, rehabilitation, communication, narrative medicine, patient safety, primary care, diagnosis, surgery, prevention, medical education, and quality improvement. Some changes occurred over time in the 'context' keywords, with some relevant new entries, becoming more popular in the last decade, namely: education, patient education, adherence. This suggests a gradual but steady switch of focus towards what we could call 'patient oriented research'.

MeSH terms

The analysis of the 50 most frequent normalized MeSH terms provides confirmation and further insight into what highlighted by the keyword analysis. MeSH terms focus on:

1. the methodology and the approach, including: narration, surveys and questionnaires, qualitative research, nursing methodology research, interviews, psychometrics, evaluation studies, retrospective studies, neuropsychological tests, and randomized control trials.

2. The context in which this research can have a substantial impact, including: interpersonal relations, attitude to health, self-concept, decision making, communication, medical history taking, attitude of health personnel, treatment outcome, social isolation, personal autonomy, verbal behaviour, social support, quality of life, and grief. Of note, starting from 2011 'treatment outcome' becomes the most frequent 'contextual' MeSH term in the corpus, followed by 'quality of life' and 'communication', providing a very clear indication on how patient narratives are mostly used now – informing the development of PROMs.

Journals

Journal trends suggests that although scholarly work on patient narratives tends to be published by discipline-specific journals (e.g: 'Brain and language', 'Social science and medicine'), it is gaining momentum and attention from top-notch medical journals, traditionally less inclined to publish this type of research. Although it is possible to speculate that the COVID-19 pandemic had a negative impact on the field, the most influential, high impact medical publications are accepting more research on patient narratives than before.
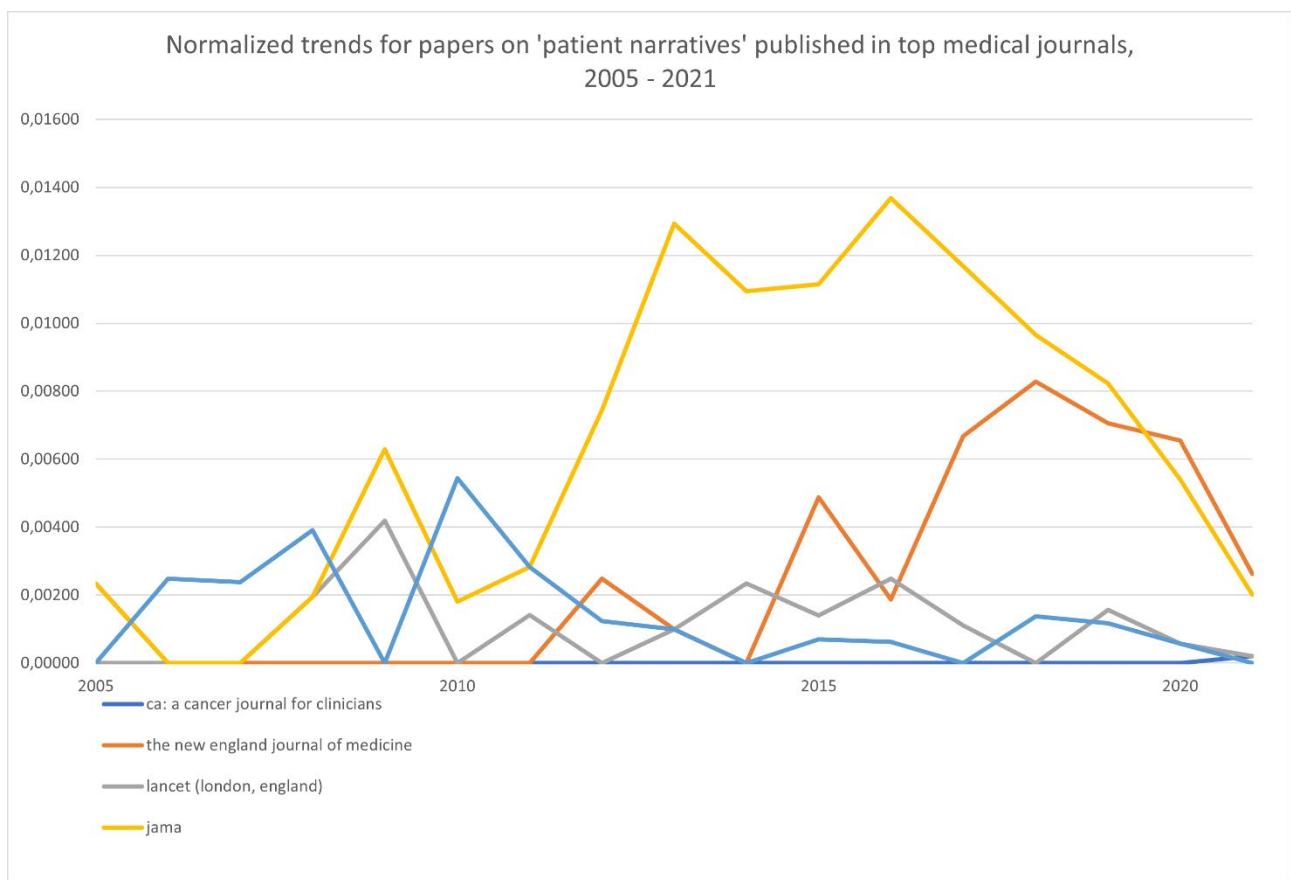
[figure 2, Normalized trends for papers on 'patient narratives' published in top medical journals, 2005 – 2021. Top medical journals have been identified and ordered by their reported impact factor in 2020: CA: A Cancer Journal for Clinicians - 120,83; the New England Journal of Medicine - 74,699; Lancet - 60,392; Journal of the American Medical Association - 51,273; Journal of Clinical Oncology - 32,956]

DIPEx International

Distant reading shows how participated and polyphonic the work on patient narratives has become, while maintaining some internal consistency. This richness is benefitting the field, nevertheless we believe that some form of standardization is needed. Standards ensure consistency, comparability, inter-compatibility, and quality of the data, needs that have emerged even more strongly during the COVID-19 crisis: comparing narrative data from different countries allows to compare the influence of structural factors like

the healthcare system or response strategies on individual experiences, hence, to identify good practices, and improve quality of care.

Some work in the sense of defining solid standards for research on patient narratives, more specifically on patient interviews already exists, resulting from the joint efforts of the DIPEx International research community. DIPEx is a multi-media approach to collecting, analysing, and disseminating patient interviews, emerged in the early 2000s from the personal healthcare experience of Ann McPherson, a general practitioner, and Andrew Herxheimer, a clinical pharmacologist. The original idea of 'a systematic collection and analysis of interviews with people about their experience of illness with evidence of the effects of treatments, and information about support groups and other resource materials' (Herxheimer et al. 2000) grew over the years, as well as the research community.

DIPEx aims to 'identify the questions that matter to people when they are ill' (Herxheimer et al. 2000). It collects interviews with patients with different diseases that are made available on a website as audio, video, or text, in accordance with the preferences of the interviewee. The idea is to inform patients, to educate healthcare professionals, and to provide a 'patient centred perspective to researchers and those who manage health services' (Herxheimer et al. 2000).

The methodology has been further codified (Ziebland and McPherson 2006; Medical Sociology and Health Experiences Research Group (MS & HERG) 2020), as well as the use and impact of the data for researchers, healthcare professionals, patients, and caregivers (Ziebland and Herxheimer 2008; Ziebland, Lavie-Ajayi, and Lucius-Hoene 2015; Drewniak et al. 2020; Christensen, Parker, and Cottrell 2021).

In the last two decades, DIPEx international has grown into a solid network with research groups in 14 countries – including Switzerland (Ziebland, Grob, and Schlesinger 2021).

The DIPEx.ch approach

In this section we present the current state of the art, the best practices, and the main limitations of this kind of approach in developing our database of patient narratives, an open structure that allows the addition of further modules, providing data that can be used for future secondary analyses in other national or international research contexts, as done for COVID-19. This way, the project and database aim to contribute not only to a smarter but to a wiser health care that is well anchored in patients' needs, values and priorities.

Project definition

A DIPEx project starts defining a 'module'. Each module targets a specific health condition. The choice of the target depends on several factors, including evidence of knowledge gaps, suggestions by healthcare institutions, or requests from patient organizations.

Modules are conducted by a team of expert researchers, with experience in qualitative research, who received specific formal training (Medical Sociology and Health Experiences Research Group (MS & HERG) 2020), and with a good command of multiple languages (a factor of crucial importance in a multilingual country such as Switzerland).

DIPEx Switzerland obtained a clarification of responsibility on the methodology from Zurich's Cantonal Ethics Committee, stating that 'it does not fall within the scope of the Human Research Act' (BASEC-Nr. 2017-00678). This was later extended to the entire country, allowing to 'carry out the project throughout Switzerland without ethics committee approval' (BASEC-Nr. 2018-00050). In case new modules include substantial methodological changes, they are subject to a second round of ethical approval.

After a review of the available literature, the research team defines a research question and a topic-informed interview guide. It comprises an open narrative section and follow-up questions that allow understanding specific topics, stemming from the research question. Each interview guide receives feedback from a module-specific advisory board, composed of healthcare professionals, other researchers, and patients. The feedback addresses both structural factors (e.g: appropriateness of the questions, salience) and linguistical factors (e.g: use of language-specific, non-translatable words, sentences or constructs). Our team of researchers comprises English, German, French and Italian speakers, with English being the common language. The interview guides are typically drafted in English, pilot-tested, and subsequently translated. The translation is focused on maintaining an accurate interpretation of the questions, as opposed to an accurate linguistic reproduction (Schaffner 1997).

An example of interview guide is provided as multimedia appendix 1.

Data collection

As typically qualitative research aims for transferability and not representativity (Morse 2000; Marshall 1996), the sampling strategy follows a maximum variation approach (Coyne 1997). The idea is to understand which variables could influence the experience (age, gender, living arrangement, family background, condition-specific factors, …) and to represent these possible variations in the sample (Herxheimer et al. 2000). Participants are recruited through patients' organizations, healthcare institutions, registries, although the strategy may vary depending on the target condition. The interviews are performed in the language participants are most comfortable in. When participants are not able to speak for themselves (e.g., due to dementia or speech and cognitive disorders after a stroke), relatives are invited for an interview. In these cases, due consideration is given to specific ethical issues that may arise regarding informed consent procedures; authorization and advice are sought from the local IRB. Interviews are conducted either in person or online, depending on the participant's preference and context. Depending on the participant's preferences, interviews are recorded as audio or video. Files are transcribed either manually or using GDPR (General Data Protection Regulation) compliant software (HappyScribe 2022), and proofread. Solid data protection protocols are of utmost importance in this phase, as the audio files may contain personal identifiers.

Coding and analysis

In order to determine theoretical saturation (Hennink and Kaiser 2020), the coding and analysis proceed in parallel with data collection. Anonymized transcripts are loaded in computer assisted qualitative data analysis software (Kuckartz and Rädiker 2019; Woolf and Silver 2017), assigned the appropriate variables, and coded. The preferential methodology is thematic analysis, a very flexible method 'for identifying, analysing, and interpreting patterns of meaning ('themes') within qualitative data' (Braun et al. 2018; Braun and Clarke 2012). The approach is hybrid and iterative, with both top-down components (i.e: theory-driven codes, strongly connected with the research question and interview guide) and bottom-up components (i.e: data-driven codes, useful to capture emerging themes). Coding aims to identify, organize, describe, and explain themes emerging in the corpus, coming to a shared interpretation, and identifying opportunities for improving the quality and efficiency of healthcare.

Coding is a collaborative effort due to professional background and language. As the healthcare experiences encompasses medical, psychological, and social factors, a multidisciplinary team can come to a more complete understanding of the corpus. To guarantee intersubjective consistency, every code is complemented by a memo, explaining its exact meaning and its intended use. When theoretical saturation is considered reached, recruitment is stopped, and the coding work is consolidated and finalized.

To ensure that implicit meaning specific to a language are not lost, interviews are coded in the original language. The codes/coding tree remains in the common language of the research team. Particular words and phrases that are difficult to directly translate are highlighted by the researcher fluent in that language,

and the cultural significance and meaning in the context of the interview transcript extensively memo-ed. To ensure all researchers have an understanding of the interview and an overview of the corpus of interview material, interview transcripts are translated into the common language using professional translation software. These translations are used only as a reference for the benefit of researchers outside of that language group and are not part of the formal coding and analysis process. Coded segments are translated to English only in the final step of preparing quotes for publication (Ferguson, Pérez-Llantada, and Plo 2011).

The goal is to enable the switch from 'vertical reading' (i.e: traditional reading of the transcripts, line by line) to 'horizontal reading' (i.e: retrieving all the fragments in which the participants talk about specific topics, defined by codes or groups of codes). This is the basis to go beyond individual experiences and understand topics, distilling the 'collective experience' (i.e: the common traits), with an eye to the variations (i.e: the discrepancies). This mapping of themes emerging from the topics, their interplay and their interconnection informs the development of 'OSOPs' (One Sheet of Paper), in which the evidence emerging from the narratives is contextualized, connected, and used to inform ethical reflection, either on specific topics or on the general management of the condition. This approach is based on thematic analysis, which in turn is derived from grounded theory (Ziebland and McPherson 2006).

Dissemination of the results

Once the analysis is finished, the data are prepared for dissemination and future use through a database. Based on the coding structure, the research team defines in the database a module-specific two-level taxonomy, cogent to report and portray the significant aspects of the 'collective experience'. Each taxa is accompanied with a text describing the general lines of the content, again, focussing both on trends and on variations. After completing the creation of the taxonomy, the research team identifies the most relevant coded fragments to attribute to the structure, and creates in the database pseudonymized profiles for the participants from whose interviews the fragments come. The creation of 'experiences' (i.e: selected, curated, and classified fragments in which one interviewee tells a significant portion of their story) knits everything together. Experiences contain a title, a short description, pseudonymized transcript, and (depending on the interviewee's preferences) a link to the original audio or video. Experiences are thus the minimum viable entity of the database, and the database allows to store (and retrieve) them using their ontological properties, classification properties, descriptive properties, or additional metadata. Properties are summarized in figure 3.
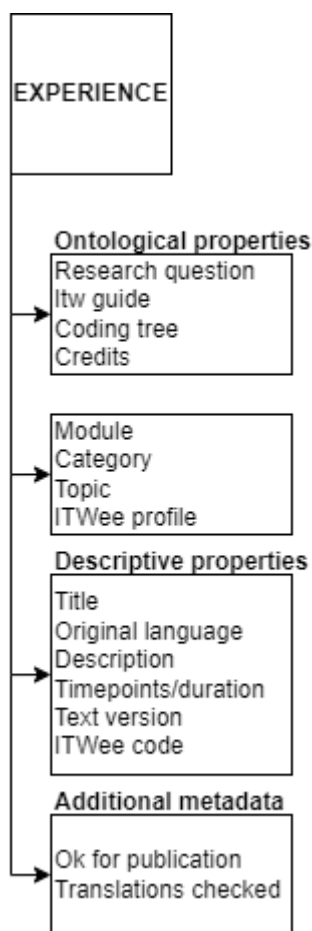
[figure 3, properties of experiences in the DIPEx database]

All the content of the database is automatically translated to English, German, French, and Italian, and proofread. The database is a MariaDB system, hosted in the data centre of the University of Zurich, ensuring high data protection standards and adherence to FAIR (findable, accessible, interoperable, reusable) data principles.

Our first completed module contains a total of 334 experiences, belonging to 28 interviewees and taxonomized in 14 categories and 61 topics. The data are fed into the system using a dedicated interface, and can be pulled using anything that supports MySQL queries. This allows multiple data-out interfaces, ranging from multimedia websites (as in the original DIPEx concept) to application programming interfaces (APIs). While the website is mainly intended for lay users, APIs will allow a more specialistic use of the data (e.g: education, training, secondary research); for instance, defining 'meta queries' or allowing bulk download of tabular data. These functions maximise the (re)use potential of the dataset, increasing the value of the entire project, and providing valuable resources for teaching, current, and future research lines.

Modules ready for the public are made available via www.dipex.ch and launched via a symposium including expert patients and healthcare professionals, providing excellent opportunities to present and discuss findings as done in Excellence in patient care in November 2021 (IBME 2021). The website is fully translated into German, French, Italian and English, and constantly reviewed with qualitative and quantitative methods with a view to usability, accessibility, and individual benefit, aiming to fulfil the requirements of the Health On the Net (HON) certificate (HON 2017), to which DIPEx subscribed. Participants can request to have their materials withdrawn from the website at any time.

Quality control and auditing

To ensure consistent quality of our data, alongside the application of the standards recognized by the Equator Network (O'Brien et al. 2014; Tong, Sainsbury, and Craig 2007), we developed a quality control checklist specific to our processes, applied as a self-assessment tool for the researchers working on a module, and as the basis for peer review of the module. Quality control considers every step of the module production. The checklist is provided as supplementary material (appendix 2).

Data protection and data management

Interview data can contain personal sensitive information and are therefore to be treated with utmost care in terms of data protection. To protect our participants' privacy, our process comprises a 2-level informed consent system, and a set of data protection strategies.

Participants are required to sign the first informed consent before recording the interview. Signing this document, they declare they have read and understood the information sheet (appendix 3); that they were given the opportunity to ask questions and clarifications; that they accept the interview to be recorded (in video or audio); and that they consent to the use of the full pseudonymized interview for research and training. The first informed consent is provided as appendix 4.

After transcription and pseudonymization, participants receive a copy of the transcript and a second informed consent form (appendix 5). They can allow the publication of the entire interview, of specific passages only, or completely opt out from publication. Moreover, they can specify if they prefer to make their interview available as video, audio, or text only.

In case we need to send non-pseudonymized interviews to third parties (e.g: for transcription) we use SHA (Secure Hash Algorithm) encrypted containers, and third parties sign a non-disclosure agreement.

Upon data collection, identifiers are decoupled and an individual code is assigned to each participant. Data are stored in a password-protected folder in a server located at the University of Zurich, subject to incremental backup. The file containing the identities and contact details of the participants is SHA encrypted and stored on a different volume.

Data selected for publication are double-checked for complete pseudonymization and saved in the database. The data-in interface has a two-factor multi-user authentication system, allowing granular permission management. Proof-reading can be done either directly via the data-in interface, or exporting and re-importing translation files.

If a participant decides to opt-out from the study and requires data deletion, their code is retrieved, and the material attributed to that code is deleted from the server and from the database. Finally, the participant's entry is deleted from the key file.

Ethics

Any sustainable, future-oriented healthcare system needs to be conceived as a learning system. Learning, of course, requires feedback loops – such as patient experiences. Healthcare systems have been notoriously slow in developing smart ways of systematically and meaningfully taking patient experiences on board. For a long time, data collection has been limited to "patient satisfaction", focusing on items such as food or convenient parking. Patient experiences have evolved conceptually, covering outcome measures relevant to patients, but their collection has remained sketchy in many instances. Narrative data can complement this puzzle in a relevant way: by not imposing topics, approaches such as DIPEx can reveal what genuinely matters to patients, and to what extent patient expectations are met by healthcare services. Listening to patients is indispensable if the goal is patient-oriented care. Translating insights into action is the next step that must follow. Researchers can contribute through the preparation of material

(such as trigger films) based on patients' voices that can start interprofessional discussions of opportunities for improvement.

Another imperative regards the dissemination of findings. If participants dedicate their time to research, it is researchers' responsibility as good stewards of the entrusted data to ensure the results are accessible, not only to a narrow community of experts, but also to a broader range of academics and the interested public. Clearly, the use of data and results – particularly if re-identification cannot be excluded – is acceptable only with participants' written informed consent. Interviewees are not to be manipulated into revealing information they do not like to share. Treating participants with respect and heeding the limits they set is an obvious ethical rule that helps avoid exploitation.

Ethical issues may come up when participants reveal delicate information, such as suicidal ideations, drug abuse, or involvement in illegal activities. In such cases, the complex interplay between the interviewer's duty to confidentiality, the participant's right to privacy, and the safety of both is carefully assessed to define a course of action – e.g. putting the participant in contact with organizations or professionals that can offer qualified help.

Another caveat regards the groups and themes that receive scholarly attention. Some groups of patients may seem more attractive from a research point of view, and more promising with a view to citations of future publications. However, from an ethical perspective, all groups of patients should have a chance to be heard, particularly those whose voice might be fainter, such as those who are marginalized, e.g. sans-papiers. At the same time, it is important to acknowledge that reaching these groups is not easy and often requires significant efforts. There is also an attention bias for certain health conditions, which may seem advantageous with a view to recruitment, funding or publication. This should not deter researchers to insist on also covering conditions that are less easy to approach.

Finally, it is a matter of fairness to involve participants rather than just treating them as a data source. We recognize citizen science as a marker of good (open) science, and therefore we try to include participants in every phase of the process, from the definition and testing of the interview guide, to the analysis and interpretation of the results (Participatory Science Academy 2020).

Limitations

Research on patient narratives is powerful and versatile, but it has some important limitations to keep in mind.

The approach we detailed is resource intensive. Researchers in charge of a module need solid training and must be very familiar with each aspect, from sample definition to interview techniques, to qualitative data analysis. While some division of labour is possible, it is inadvisable to split a module in isolated and self-standing work packages. Some aspects can be automated: transcriptions and translations can be produced by software and proof-read by humans; we also automated the production and upload of the clips. This is an important innovation, as manual cutting and upload requires approximately 60 – 70 hours of human labour and about 7 hours of machine time per module, while programmatic cutting and uploading allows to achieve the same results in about 8 hours of machine time. We are currently exploring the possibility to automate (to some extent) also the coding, but it is unclear whether it is possible to achieve a level of quality comparable with human coding (Spitale, Biller-Andorno, and Germani 2022). In addition, after some years modules might require updates to remain current, adding new interviews and re-conducting the analysis.

This research can be rigorous and systematic. Nevertheless, personal bias can skew the coding and the interpretation. This risk can be mitigated by adopting a shared and multidisciplinary approach, by explaining codes with memos, and by exploring and challenging the preconceptions of the research team. Assessing theoretical saturation is an issue connected to what above. On paper, theoretical saturation is

considered reached when no new codes are emerging (Nascimento et al. 2018). In practice, if the coding team didn't approach coding with enough rigour, it could declare theoretical saturation without actually reaching it.

Results from research on patient narratives are particularly vulnerable to 'data torturing'. They make sense when considered in their unity, complexity, and context; in order to build reflections on the entire movie, rather than on a snapshot, hosting a DIPEx project in a research unit natively offering interdisciplinary expertise (such as a biomedical ethics unit) is a good mitigation strategy.

Finally, results must be operationalized and integrated into patient care. This requires a certain predisposition to integrate this evidence on the part of the healthcare system – which currently can depend a lot on personal sensitivities and inclinations.

Conclusion

When compared to other current and emerging collections of patient narratives (e.g: social media pages/groups), the approach we detailed offers significant advantages: conflict of interest is assessed and avoided; data collection and analysis are rigorous and systematic; public-facing content is carefully curated, and the whole process is supported by a lively research community.

A recent systematic review on risks and benefits of patient narratives concluded that 'patient narratives seem to be a promising means to support users in improving their understanding of certain health conditions and possibly to provide emotional support and have an impact on behavioral changes' (Drewniak et al. 2020). This approach goes beyond informing the development of evidence-based PROMs and PREMs, decision aids, or trigger films. Our data structure is built with an eye to the future: curated collections hosted on FAIR databases will provide a fundamental infrastructure for natural language processing approaches to patient narratives, for the training of conversational AIs, for medical education, and for data sharing in the context of international studies.

Finally, this approach can help integrate the contributions of different disciplines sitting at the interdisciplinary table of biomedical ethics – medicine, nursing studies, philosophy, sociology, psychology, anthropology, and others – in the interest of a joint goal: patient-centred care. At the same time, making patients' stories publicly available endorses important cultural changes currently under way, reducing knowledge and power asymmetries between healthcare practitioners and patients (and their relatives), and fostering genuinely personalized healthcare by providing opportunities to listen to what matters to patients.

References

Biller-Andorno, Nikola, and Thomas Zeltner. 2015. 'Individual Responsibility and Community Solidarity--The Swiss Health Care System'. *The New England Journal of Medicine* 373 (23): 2193–97. https://doi.org/10.1056/NEJMp1508256.

Braun, Virginia, and Victoria Clarke. 2012. 'Thematic Analysis'. In *APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological*, 57–71. APA Handbooks in Psychology®. Washington, DC, US: American Psychological Association. https://doi.org/10.1037/13620-004.

Braun, Virginia, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2018. 'Thematic Analysis'. In *Handbook of Research Methods in Health Social Sciences*, edited by Pranee Liamputtong, 1–18. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-2779-6_103-1.

Christensen, Vivian, Kellee Parker, and Erika Cottrell. 2021. 'Leveraging a Qualitative Data Repository to Integrate Patient and Caregiver Perspectives into Clinical Research'. *Journal of Clinical and Translational Science* 5 (1): e155. https://doi.org/10.1017/cts.2021.822.

Coyne, IT. 1997. 'Sampling in Qualitative Research. Purposeful and Theoretical Sampling; Merging or Clear Boundaries?' *Journal of Advanced Nursing* 26 (3): 623–30. https://doi.org/10.1046/j.1365-2648.1997.t01-25-00999.x.

Deutsches Netzwerk Gesundheitskompetenz. 2022. 'Erfahrungsberichte Im Gesundheitswesen'. Deutsches Netzwerk Gesundheitskompetenz DNGK. 20 February 2022. https://dngk.de/erfahrungsberichte/.

Drewniak, Daniel, Andrea Glässel, Martina Hodel, and Nikola Biller-Andorno. 2020. 'Risks and Benefits of Web-Based Patient Narratives: Systematic Review'. *Journal of Medical Internet Research* 22 (3): e15772. https://doi.org/10.2196/15772.

Equator Network. 2022. 'About Us | The EQUATOR Network'. Equator-Network.Org. 2022. https://www.equator-network.org/about-us/.

Ferguson, Gibson, Carmen Pérez-Llantada, and Ramón Plo. 2011. 'English as an International Language of Scientific Publication: A Study of Attitudes'. *World Englishes* 30 (1): 41–59. https://doi.org/10.1111/j.1467-971X.2010.01656.x.

Franzini, Greta, Stefan Jänicke, Gerik Scheuermann, and Muhammad Cheema. 2015. *On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. A State-of-the-Art (STAR) Report.*

Han, MeiLan K, Carlos H Martinez, David H Au, Jean Bourbeau, Cynthia M Boyd, Richard Branson, Gerard J Criner, et al. 2016. 'Meeting the Challenge of COPD Care Delivery in the USA: A Multiprovider Perspective'. *The Lancet Respiratory Medicine* 4 (6): 473–526. https://doi.org/10.1016/S2213-2600(16)00094-1.

HappyScribe. 2022. 'Security and Data Protection'. Www.Happyscribe.Com. 2022. https://www.happyscribe.com/security.

Hennink, Monique M., and Bonnie N. Kaiser. 2020. 'Saturation in Qualitative Research'. In *SAGE Research Methods Foundations*. London: SAGE Publications Ltd. https://doi.org/10.4135/9781526421036822322.

Herxheimer, A., A. McPherson, R. Miller, S. Shepperd, J. Yaphe, and S. Ziebland. 2000. 'Database of Patients' Experiences (DIPEx): A Multi-Media Approach to Sharing Experiences and Information'. *Lancet (London, England)* 355 (9214): 1540–43. https://doi.org/10.1016/S0140-6736(00)02174-7.

HON. 2017. 'Our Commitment to Reliable Health and Medical Information'. 2 May 2017. https://www.hon.ch/HONcode/Patients/Visitor/visitor.html.

IBME. 2021. 'Excellence in Patient Care Symposium 2021 – The Swiss Database of Individual Patient Experiences (DIPEx.Ch)'. University of Zurich. 26 November 2021. http://www.ibme.uzh.ch/en/Biomedical-Ethics/Agenda/Previous-Events/Excellence-in-Patient-Care-Symposium-2021.html.

JAMA. 2022. 'Narrative Medicine'. Jamanetwork.Com. 2022. https://jamanetwork.com/collections/5766/narrative-medicine.

Kleinman, Arthur. 1989. *The Illness Narratives: Suffering, Healing, And The Human Condition*. Reprint edizione. New York: Basic Books.

Kuckartz, Udo, and Stefan Rädiker. 2019. *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*. Cham, Switzerland: Springer.

Marshall, M.N. 1996. 'Sampling for Qualitative Research'. *Family Practice* 13 (6): 522–25. https://doi.org/10.1093/fampra/13.6.522.

Medical Sociology and Health Experiences Research Group (MS & HERG). 2020. 'Researcher's Handbook - Healthtalk Modules'. University of Oxford.

Moretti, Franco. 2013. *Distant Reading*. Verso Books.

Morse, J.M. 2000. 'Determining Sample Size'. *Qualitative Health Research* 10 (1): 3–5. https://doi.org/10.1177/104973200129118183.

Nascimento, Luciana de Cassia Nunes, Tania Vignuda de Souza, Isabel Cristina dos Santos Oliveira, Juliana Rezende Montenegro Medeiros de Moraes, Rosane Cordeiro Burla de Aguiar, and Liliane Faria da Silva. 2018. 'Theoretical Saturation in Qualitative Research: An Experience Report in Interview with

Schoolchildren'. *Revista Brasileira de Enfermagem* 71 (1): 228–33. https://doi.org/10.1590/0034-7167-2016-0616.

O'Brien, Bridget C., Ilene B. Harris, Thomas J. Beckman, Darcy A. Reed, and David A. Cook. 2014. 'Standards for Reporting Qualitative Research: A Synthesis of Recommendations'. *Academic Medicine: Journal of the Association of American Medical Colleges* 89 (9): 1245–51. https://doi.org/10.1097/ACM.0000000000000388.

Pallai, EmmaLee, and Kim Tran. 2019. 'Narrative Health: Using Story to Explore Definitions of Health and Address Bias in Health Care'. *The Permanente Journal* 23 (January): 18–052. https://doi.org/10.7812/TPP/18-052.

Participatory Science Academy. 2020. 'What Is Participatory Research?' Universität Zürich. 30 May 2020. http://www.pwa.uzh.ch/en/aboutus/What-is-participatory-research-.html.

Porter, Michael E. 2010. 'What Is Value in Health Care?' *New England Journal of Medicine* 363 (26): 2477–81. https://doi.org/10.1056/NEJMp1011024.

Schaffner, Christina. 1997. 'From "Good" to "Functionally Appropriate": Assessing Translation Quality'. *Current Issues In Language and Society* 4 (1): 1–5. https://doi.org/10.1080/13520529709615476.

Spitale, Giovanni. 2022. 'PubMed Literature on Patient Narratives 1975 - 2021'. Zenodo. https://doi.org/10.5281/zenodo.6534889.

Spitale, Giovanni, and Nikola Biller-Andorno. 2021. *TopicTracker: A Python Pipeline to Search, Download and Explore PubMed Entries*. Zenodo. https://doi.org/10.5281/zenodo.4643876.

Spitale, Giovanni, Nikola Biller-Andorno, and Federico Germani. 2022. 'Concerns Around Opposition to the Green Pass in Italy: Social Listening Analysis by Using a Mixed Methods Approach'. *Journal of Medical Internet Research* 24 (2): e34385. https://doi.org/10.2196/34385.

Tong, Allison, Peter Sainsbury, and Jonathan Craig. 2007. 'Consolidated Criteria for Reporting Qualitative Research (COREQ): A 32-Item Checklist for Interviews and Focus Groups'. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care* 19 (6): 349–57. https://doi.org/10.1093/intqhc/mzm042.

Woolf, Nicholas H., and Christina Silver. 2017. *Qualitative Analysis Using MAXQDA: The Five-Level QDATM Method*. Routledge.

Ziebland, Sue, Rachel Grob, and Mark Schlesinger. 2021. 'Polyphonic Perspectives on Health and Care: Reflections from Two Decades of the DIPEx Project'. *Journal of Health Services Research & Policy* 26 (2): 133–40. https://doi.org/10.1177/1355819620948909.

Ziebland, Sue, and Andrew Herxheimer. 2008. 'How Patients' Experiences Contribute to Decision Making: Illustrations from DIPEx (Personal Experiences of Health and Illness)'. *Journal of Nursing Management* 16 (4): 433–39. https://doi.org/10.1111/j.1365-2834.2008.00863.x.

Ziebland, Sue, Maya Lavie-Ajayi, and Gabriele Lucius-Hoene. 2015. 'The Role of the Internet for People with Chronic Pain: Examples from the DIPEx International Project'. *British Journal of Pain* 9 (1): 62–64. https://doi.org/10.1177/2049463714555438.

Ziebland, Sue, and Ann McPherson. 2006. 'Making Sense of Qualitative Data Analysis: An Introduction with Illustrations from DIPEx (Personal Experiences of Health and Illness)'. *Medical Education* 40 (5): 405–14. https://doi.org/10.1111/j.1365-2929.2006.02467.x.