**BMJ Health &
Care Informatics**

Categorising patient concerns using natural language processing techniques

Paul Fairie,[1,2] Zilong Zhang,[3,4] Adam G D'Souza,[3,5] Tara Walsh,[5] Hude Quan,[1,3]
Maria J Santana 🄳 [1,2,6]

[1]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[2]Alberta Strategy for Patient-Oriented Research Patient Engagement Platform, Calgary, Alberta, Canada
[3]Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[4]Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[5]Alberta Health Services, Calgary, Alberta, Canada
[6]Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

Correspondence to
Dr Maria J Santana;
mjsantan@ucalgary.ca

ABSTRACT

Objectives Patient feedback is critical to identify and resolve patient safety and experience issues in healthcare systems. However, large volumes of unstructured text data can pose problems for manual (human) analysis. This study reports the results of using a semiautomated, computational topic-modelling approach to analyse a corpus of patient feedback.

Methods Patient concerns were received by Alberta Health Services between 2011 and 2018 (n=76 163), regarding 806 care facilities in 163 municipalities, including hospitals, clinics, community care centres and retirement homes, in a province of 4.4 million. Their existing framework requires manual labelling of pre-defined categories. We applied an automated latent Dirichlet allocation (LDA)-based topic modelling algorithm to identify the topics present in these concerns, and thereby produce a framework-free categorisation.

Results The LDA model produced 40 topics which, following manual interpretation by researchers, were reduced to 28 coherent topics. The most frequent topics identified were communication issues causing delays (frequency: 10.58%), community care for elderly patients (8.82%), interactions with nurses (8.80%) and emergency department care (7.52%). Many patient concerns were categorised into multiple topics. Some were more specific versions of categories from the existing framework (eg, communication issues causing delays), while others were novel (eg, smoking in inappropriate settings).

Discussion LDA-generated topics were more nuanced than the manually labelled categories. For example, LDA found that concerns with community care were related to concerns about nursing for seniors, providing opportunities for insight and action.

Conclusion Our findings outline the range of concerns patients share in a large health system and demonstrate the usefulness of using LDA to identify categories of patient concerns.

INTRODUCTION

In an effort to create learning health systems, quality improvement staff continually seek feedback from patients on the quality of care they receive.[1–4] While conventional patient experience surveys provide patients with validated and highly structured ways to evaluate their experiences which have been shown to correlate with outcomes such as 30-day morbidity, unplanned readmission

Summary

What is already known?
► Free-text feedback is commonly received by health systems, and with the rise of online patient portals the volume of such data is rapidly increasing.
► Unstructured data (such as free text) can be more difficult to rapidly mine than structured data (such as validated surveys), which may make it difficult to rapidly respond to patient safety and experience concerns.
► Latent Dirichlet allocation (LDA) is an approach that can classify documents into topics in a data-driven rather than framework-driven way.

What does this paper add?
► This study reports the results of using a novel topic-modelling approach (label-enhanced LDA (LabEL)) to analyse a corpus of patient feedback, combining LDA with a custom model selection process that leverages the pre-existing category labels.
► Following an interpretation process by researchers (including patient-centred care experts), the model produced 28 coherent topics, some comparable to those in a pre-existing framework, and others more novel.
► The results also serve as a demonstration of the LabEL approach, which can be used to provide rapid, framework-free analytics of free-text patient concerns data.

and unplanned reoperation,[5] unstructured feedback in the form of free text (such as telephone calls or online patient portals) can allow them to be more expressive and can play an important role in improving patient safety and quality of care.[1] Addressing this feedback in a timely fashion has been demonstrated to reduce patient mortality,[6] and patient narratives can also have powerful impacts on health policy-makers.[7]

As a result, unstructured patient feedback is being increasingly collected. For example, the UK's National Health Service receives more than 100 000 feedback messages per year related to inpatient care.[8] In Alberta, Canada, Alberta Health Services (AHS), the single health authority in the province (population:

4.4 million) receives more than 10 000 patient concerns over the phone per year, and stores them in a dataset called the Feedback and Concerns Tracking (FACT) database.[9 10] Given this volume, being able to rapidly theme concerns can help quality improvement professionals and health services researchers monitor changes in the nature of the concerns and better understand the patient experience.

Additionally, it is critical for systems to begin to prepare for an increased volume due to the ease of submitting concerns online.[11] Advances in natural language processing (NLP) have made alternatives to hand-coded, framework-driven categorisations possible, allowing the development of automated tools to assist coders in classifying concerns by topic, and to give decision makers big-picture summaries of themes as they arrive.[12] For example, a recent study, focusing specifically on categorising patient survey responses regarding their care, used an unsupervised topic modelling method called non-negative matrix factorisation (NMF) to derive their discussion topics.[13] Another study also applied NMF to model topics in online reviews of doctor performance posted by patients.[14]

One particular unsupervised topic modelling method, latent Dirichlet allocation (LDA),[15] has proven particularly popular and successful. LDA has been used for topic mining in studies of health data across an array of data sources, including discussions from condition-specific online support groups[16–20] and more general online discussion platforms,[21–29] data about adverse medical events,[30] interview transcripts of patients,[31 32] media articles[33] and survey data.[34 35] Other studies have used LDA to analyse topics in patient-reported concerns as well, in situations where no existing topic information is available. For example, a recent study performed fully unsupervised topic analysis on patient-reported experiences from a British database using LDA.[36]

The FACT database used in the present study does provide manual topic labels for each of the concerns, but the framework used for the categorisation is not directly data-driven, and the labels are not very detailed. Furthermore, only one label is provided to each concern, even though in reality multiple topics may have been discussed. Variations on LDA have been proposed to deal with supervised topic modelling in instances with existing labels, such as for example labelled LDA.[37] However, labelled LDA assumes that the provided set of topic labels is the complete and final set to be used, which is not an appropriate assumption when the given categorisation framework does not have all of the desired properties.

The objective of this paper is to identify common topics of concern for patients as reported in the FACT database, using an approach that can report results in a timely fashion and handle any volume increases anticipated with the upcoming introduction of an easy-to-use province-wide patient portal. The method used leverages the existing coarse-grained topic labels to guide the process of assigning more specific labels that can be potentially actionable.

METHODS

To classify these patient concerns, our analytical approach consisted of four steps. A schematic depiction of the entire analytical approach (called label-enhanced LDA (LabEL)) can be found in figure 1, and the following text outlines the procedure for using LDA, combined with expert consensus, to identify topics from a corpus of a text-based patient concerns data. LabEL employs standard LDA to model the topics, but with a novel methodology to select the optimal LDA model leveraging the existing manually labelled categories. Some existing proposals extend standard LDA by incorporating expert knowledge to produce more relevant and interpretable topics.[38 39] Other proposals use co-occurrence statistics of LDA topic words to evaluate the model,[9] but the resulting metrics do not always correspond with intuitive human judgements.[10] LabEL, inspired by other previous work,[40 41] was designed to mitigate both of these issues; it uses existing coarse-grained manual labels to derive a corpus-specific evaluation metric based on idealised 'dummy documents' that are specially constructed to be about a single topic, which rewards highly specific topics while penalising overlap between different topics. Additional details can be found in a forthcoming publication.
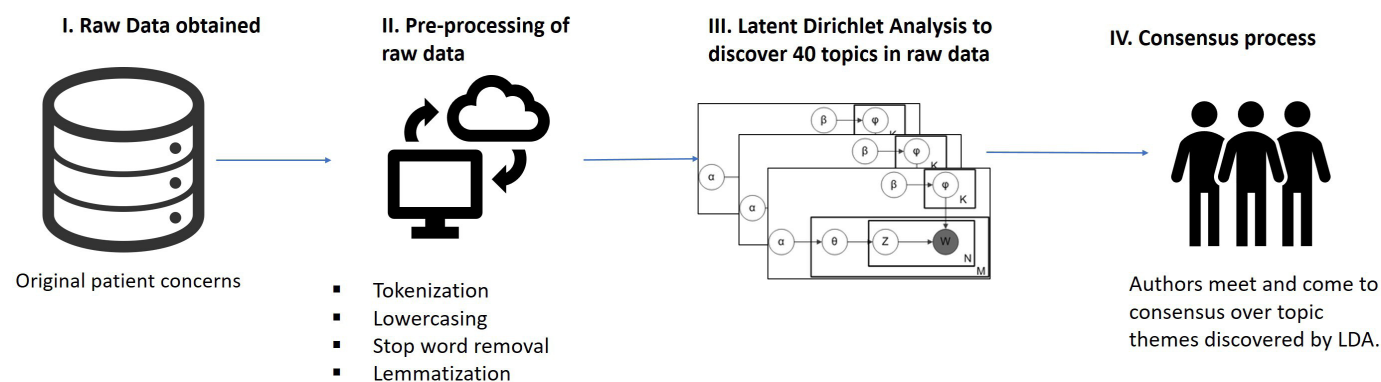

Figure 1 Illustration of the LabEL analytic pipeline used to combine existing manual labels with LDA modelling to produce topics discussed in patient concerns corpus. LabEL, label-enhanced LDA; LDA, latent Dirichlet allocation.

I. Raw Data obtained

Original patient concerns

II. Pre-processing of raw data

- Tokenization
- Lowercasing
- Stop word removal
- Lemmatization

III. Latent Dirichlet Analysis to discover 40 topics in raw data

IV. Consensus process

Authors meet and come to consensus over topic themes discovered by LDA.

Step 1: raw data obtained

The data used for this study were abstracts of patient concerns (n=76163) from the FACT database, collected via telephone calls to the AHS patient relations department between 2011 and 2018 and abstracted by patient concerns consultants (PCCs). Calls were made by patients, family and other caregivers. PCCs summarised the calls in free-text narrative style, and categorised each one into one of four predetermined primary category labels, and one of 26 secondary category labels.

To first understand the important characteristics of the data, we conducted a descriptive analysis. We computed the proportions of comments that fell into each of the primary and secondary categories provided by AHS, as well as the proportion of comments that were provided by patients themselves and by other people.

Step 2: preprocessing of raw data

Text was preprocessed using the spaCy Python package.[42] Records were removed when either the primary and secondary category labels or the full-text description of the concern were missing, which resulted in 74 260 documents remaining. All text was converted to lower-case and stop words (standard and customised) were removed. Text was then lemmatised and vectorised using a bag-of-words representation. After removing duplicates and concerns consisting entirely of stopwords, the final dataset contained 73 093 documents.

Step 3: LDA

LDA models documents as distributions over topics, and topics as distributions over words.[15] The underlying assumption of LDA is that several different topics compose each document, and if a proportion of a text is about a certain topic, then vocabulary related to it will appear in that document. For example, if a concern is about care wait times, then the words time, wait, appointment and urgent will appear with high frequency, and if a concern is related to community care for the elderly, words such as nursing, long, term, plan and facilities will similarly appear with high frequency.

LDA determines which words are strongly related to each other by counting the frequency with which words appear together in the one document. Finally, the LDA model decomposes the input text data into two parts. The first part is which topics each article is composed of and in what percentage. The second part is the occurrence probability of words for each topic.[15] We can interpret important words of topics to form thematically meaningful issues. With the first part, the proportion of each topic in the text, we can have a comprehensive understanding of the factors that worry patients in the medical system.[43]

Two features make LDA an attractive method for analysing patient concerns data. First, in contrast to a one-time, expert-driven manual development of a framework, LDA can produce a data-driven framework reflecting the empirical patterns that exist in the data, and can potentially identify more specific categories than manual analysis completed by public-facing patient relations officials, rather than qualitative researchers. Second, LDA can also detect novel and emerging concerns that do not belong to an existing topic, responding to new and emerging concerns (like a new care procedure), or a contextual change (like a global pandemic) without needing to re-design the framework.

The analysis was conducted using the online LDA[44] algorithm from scikit-learn,[45] a variant on standard LDA that splits the data into mini-batches and updates the estimators according to a specified learning rate after each mini-batch, to speed up convergence. We fit the model to the bag-of-words representation of the processed, de-duplicated 73 093 patient concerns, the keywords for topics found by LDA were compared with the semantic valid keywords explored in previous supervised learning to determine the optimal LDA model.

More technical details of the LDA-specific steps of this analysis will be provided in a forthcoming publication.

Step 4: thematic analysis consensus process

After the LDA analysis, three of the authors met several times to assign labels to each LDA topic. For each LDA topic, they examined the list of the 40 words with the highest weights from LDA, as well as the distribution of weights, and defined one or more candidate labels based on these words when possible. The results of this process were presented and discussed with the rest of the research team.

To check the consistency of the candidate labels with the data, they looked at a list of ten example documents for which LDA had assigned a high weight (>0.8) to that topic (or all such documents if there were fewer than ten in the corpus), to select and refine the candidate labels that made the most sense.

If this list of examples was deemed insufficient to achieve consensus, they further examined a random sample of examples with weight 0.5 or more for that topic in order to attempt to achieve consensus. If it was still impossible to achieve consensus after this process, the topic was discarded. Otherwise, the consensus labels were selected and recorded in a table. The proportions of the remaining topics were then calculated, and compared against the proportions from the original framework. This is consistent with other studies where consensus processes were used to give a group name to a collection of themed texts.[46]

RESULTS

Characteristics of participants and concerns data

Concerns were received regarding care from 806 different institutions across the province during the nine calendar years, 2010–2018 inclusive. 49.9% of concerns were reported by patients, 48.1% reported by other people such as caregivers or relatives and 2.0% were anonymous, unknown or otherwise unclassified. The abstracted
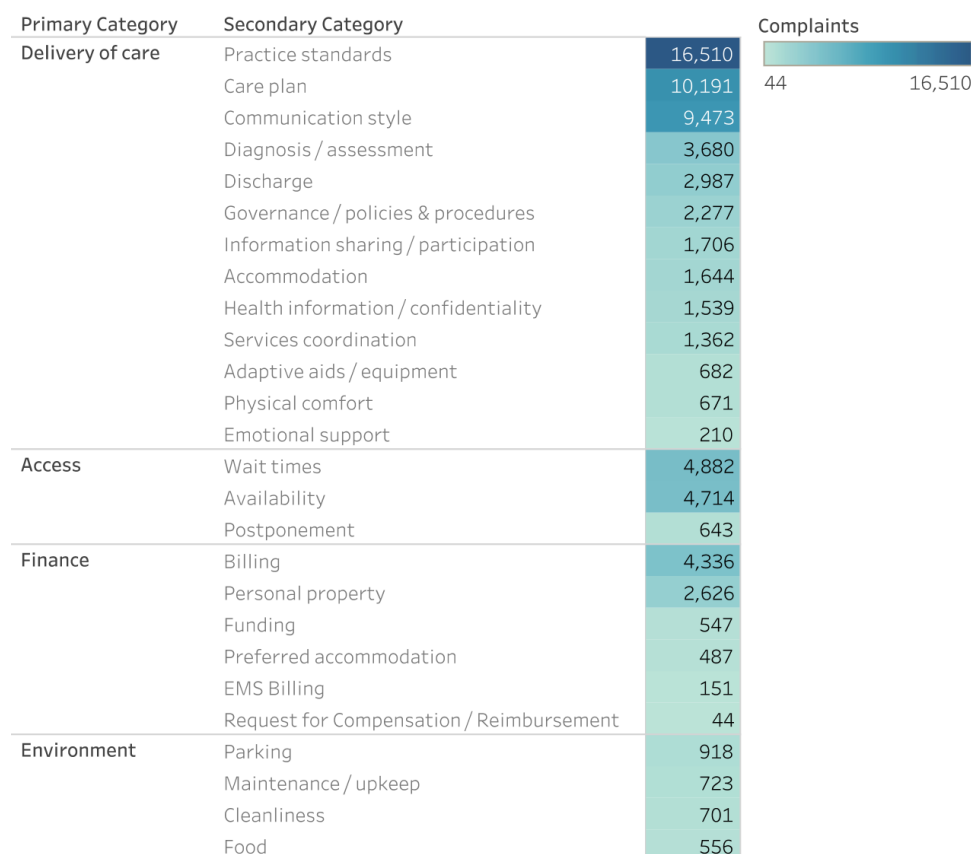
Figure 2 Breakdown of concerns by original manually assigned primary and secondary categories. EMS, emergency medical services.

concerns ranged in length from 3 to 28 649 characters, with a median of 456 characters (IQR: 170 to 994.5).

Figure 2 lists the breakdown of concerns by their original primary and secondary categories. Most concerns (71%) fell into the Delivery of care primary category, while Access (14%), Finance (11%) and Environment (4%) made up the balance. The three largest secondary categories by far were all within the Delivery of care primary category: practice standards (22% of all concerns), care plan (14%) and communication style (13%), with the other half split between the 23 other secondary categories. Some of the secondary categories within Delivery of care are larger than the other primary categories.

LDA topic modeling and thematic analysis

The topics discovered by the process outlined in figure 1 (LDA, followed by a consensus process) are listed in table 1, in descending order of their total weights in the corpus. Of the original 40 LDA, thematic analysis revealed that 28 of them coherently captured subjects that were mentioned in the abstracted patient concerns data, while 12 of them did not appear to be coherent, and were removed. While 12 of the 40 were removed, the total weight of the concerns that were removed was 3.7%, leaving 96.3% of the concerns with assigned topic labels. The most heavily weighted complaint was 'communication issues causing delays', which had a total weight of

10.6%, and the top 7 concern topics constituted more than half of the proportion of concerns.

DISCUSSION

Our findings revealed many common topics of patient concerns, such as communication, access to care and coordination of care. On consideration, these themes are closely aligned with the six domains of healthcare quality from the Institute of Medicine (IOM) framework[47]: safe (eg, abuse in long-term care facilities), effective (eg, medication concerns), patient-centred (eg, communication), timely (eg, problems with test results), efficient (eg, billing accuracy) and equitable. No themes were obviously related to equity, but equity is more of a group-level concept and patient concerns exist more at the individual patient level. Overall, however, the alignment between the IOM framework and real-life free-text patient concerns suggests that these six domains are well-aligned with the concerns of patients, as themed by an LDA-consensus approach, rather than a pre-established framework.

Alignment of topics identified by LabEL with existing manually assigned categories

Most of the LabEL-identified topics align well with the existing manually assigned categories. However, LabEL identified more detailed versions of many of the manually assigned labels, and also identified completely novel

Table 1 Topics identified using analytical approach

Candidate labels	Topic words	Proportion (%)
▶ Communication issues causing delays ▶ Communication-related issues with booking	Appointment, wait, time, clinic, speak, clerk	10.58
▶ Community care (particularly for elderly patients) ▶ Treatment and nursing issues for elderly people	Care, home, staff, facility, provide, manager	8.82
▶ Interactions with nurses	Nurse, come, triage, ask, leave, sit, rude,	8.80
▶ Emergency department care	ED, physician, pain, CT, scan, symptom	7.52
▶ Policies, procedures or practice standards ▶ Very short concerns	Health, centre, practice, standard, emergency, social	7.20
▶ Medication and mental health	Medication, psychiatrist, mental, health, mg, remand	4.92
▶ Injury and wound care	Antibiotic, injury, ray, infection, left, fracture	4.85
▶ Surgical issues (especially cancer and childbirth)	Surgery, surgeon, baby, post, delivery, procedure	4.64
▶ Lost items	Discharge, miss, belonging, item, lose, glass	4.27
Mixture of two topics: ▶ Care of patients who died ▶ Physician communication style	Medication, relay, physician, resident, ask, oxygen	4.08
▶ Errors in forwarding of documents	Report, dialysis, outcome, denture, document, forward	3.88
▶ Emergency medical services	EMS, ambulance, crew, transport, paramedic, attendant	3.58
▶ In-hospital hospitality	Bed, food, unit, floor, room, meal	3.52
▶ Tests results (esp. patients who are pregnant and patients with cancer)	Blood, oncologist, test, CCI, lab, fort	3.38
▶ Parking and accessibility	Programme, parking, access, service, wheelchair, machine	3.14
▶ Accommodation-related billing	Private, room, charge, semi, accommodation, sign	3.11
▶ Billing accuracy and affordability	Bill, pay, invoice, tech, cost, receive	3.07
Mixture of two topics: ▶ Smoking in inappropriate settings ▶ Care of children	Child, security, parent, stollery, guard, children	2.74
Mixture of two topics: ▶ Care in Medicine Hat (a name of a city) ▶ Pain management medication	Knee, medicine, hat, physiotherapy, specialist, GI, ERC	2.11
Mixture of two topics: ▶ Long-term care placement ▶ Abuse	Stroke, placement, hearing, guardian, LTC, legal	1.56
▶ File closed	Urgent, file, chumir, sheldon, cardiologist, PCC, close	1.05
Mixture of two topics: ▶ Left message ▶ Urinary catheterisation	Catheter, voice, mail, message, law, crutch	0.76
▶ Aircast boots and splints	Cast, air, throat, ACH, boot, sore	0.59
▶ Issues with MRIs	MRI, chinook, CRH, stocking, kin, reduction, compression	0.45
▶ Lost jewellery	Grandmother, ring, NP, scope, gold, practitioner	0.44
▶ Issues with intubation ▶ Treatment of vehicular accident patients	Tube, vehicle, bariatric, CPAP, nose, motor	0.37
▶ Dietary concerns	Diet, special, dietician, PCO, restriction, dietary	0.33
▶ Skin care	Rash, cream, innisfail, INCC, soap, haemorrhoid	0.26

topics. Table 2 explicates the alignment between the LabEL topics and the manually assigned categories.

Because the descriptions of the pre-defined secondary categories are vague, most (21/28) LabEL topics matched or partially overlapped with the secondary categories. Among the 21 overlapping topics, LabEL offers 16 with finer granularity. We can regard those 16 topics as subcategories of the original secondary classification. Additionally, LabEL identified 7 novel topics, 6 of which were clear and 1 of which was a mixture of two semantically unrelated topics.

By comparing the LabEL-identified topics and categories in the previous classification framework, we can see that the manually defined categorisation focuses more on matters directly related to medical institutions, such as issues encountered in hospitalisation (accommodation, food and so on) and treatment (practice standard, care plan and so on). It also considers the time cost (wait time)

Table 2 Comparison of the topics identified by LabEL with existing manually assigned categories

	LDA topics	Matched manually defined category
More detailed topics	Communication issues causing delays	Wait time
	Interactions with nurses	Communication style
	Physician communication style	
	Surgical issues	Practice standards
	Emergency medical services	
	Issue with intubation	
	MRI	
	Boots and splints	
	Injury and wound care	Care plan
	Medication and mental health	
	Emergency department care	
	Skin care	
	Lost items	Personal property
	Lost jewellery	
	Smoking and children care	Maintenance/upkeep
	Parking and accessibility	Parking
Old topics	Policies, procedures or practice standards	Governance/policies and procedures
	In-hospital hospitality	Maintenance/upkeep
	Billing accuracy and affordability	Billing
	Accommodation-related billing	Preferred accommodation
	Dietary concerns	Food
New topics	Errors in forwarding of documents	
	Test result	
	Pain management medication	
	Long-term care placement and abuse	
	File closed	
	Community care and treatment and nursing issues for elderly people	
	Mixture of two topics: Left message and urinary catheterisation	

LabEL, label-enhanced LDA; LDA, latent Dirichlet allocation.

and financial cost (funding, billing) required to access medical services. However, other important elements of the healthcare system such as community care (community care, long-term care placement and abuse), testing agencies and drug use are not included. The original framework also does not include any aspects of the concern-sharing process itself (issues of forwarding files and closing files). For the concerns considered in the original framework, the LabEL-identified topics provide a more detailed description of the original framework categories, and make more specific, sometimes more actionable topics.

Alignment of topics identified by LabEL with other existing frameworks

Many LabEL-identified topics are aligned with person-centred care frameworks.[48] In particular, communication appears as a common theme in many topics—not only communication between providers and patients and their families, but also communication between providers across healthcare sectors, which suggests improvements in coordination could be achieved, which could improve timely access to care. Using this approach itself could also help to achieve person-centred care since it allows patients to communicate their concerns in their own voice (as free text is often more expressive than survey responses, and responses can be coded without needing an expert-developed framework).

Implications for patient safety and quality improvement

Further, many topics identified using this approach also addressed patient safety issues, such as interaction with nurses addressing the needs of elderly patients, particularly at the community care level and also concerns

directly relevant to abuse and long-term care placement. The flexibility of this approach can help to identify emergent patient safety issues, rather than force the system to imagine how care might go wrong.

Additionally, the results provide deeper insight beyond topic proportions. In several instances, it identifies specific, actionable topics. For example, it identifies coherent clusters of concerns around aircast boots & splints and lost jewellery & other items. While these topics would not be normally present in most broad classification schemas, they are specific enough to be likely resolvable by quality improvement professionals.

The mixed nature of certain topics also reveals how certain classes of concerns are related. For example, concerns around patients who died and provider communication styles might reveal a relationship between these two sub-themes that is worthy of further exploration. While the findings from this specific study are not robust enough to alone suggest that a direct causal link between provider communication issues and death, they do suggest that patient concerns about provider communication might be related to patient concerns about adverse patient outcomes. This, by itself, may then suggest future avenues for study and investigation.

Applications of LabEL to other data
LabEL can be applied to other topic modelling tasks on free-text corpora, in the medical domain and elsewhere. The general purpose of the model is to leverage any existing labels that may exist, and which may provide clues about which documents are semantically related to each other, instead of relying solely on statistical co-occurrence of words in the corpus. The use case for this appears to be reasonably common. For instance, Tapi Nzali et al[49] apply LDA to social media posts from patients with breast cancer to identify key topics of interest to this population. The corpora include a health forum with 16 868 posts in 1050 discussion threads, and Facebook groups with 70 092 posts distributed over 11 013 discussion threads. Knowledge that two posts have come from the same discussion thread constitutes valuable domain information that can inform topic modelling, and LabEL is perfectly suited for this kind of analysis. As another example, Liu et al[33] analyse the topics present in articles about thirdhand smoke extracted from new databases, including Factiva.[50] Factiva uses a proprietary taxonomy called Dow Jones Intelligent Identifiers to provide labels to articles, including data elements such as region, topic and company (among others). LabEL can potentially leverage these additional taxonomic identifiers to inform the topic modelling process and produce more coherent topics.

Benefits of LDA and LabEL compared with manual analysis
Additionally, LabEL still confers all the benefits of traditional LDA compared with manual analysis. For the FACT dataset, we used to develop LabEL, the proportions of the different LDA topics can be valuable as a supplement to the proportions of manual categories already available to AHS to aid them in understanding the nature of patient feedback they have received. This is partially because LDA assigns fractional labels for multiple topics to each document. This is important because the concerns are documented and classified for a whole conversation. Consequently, for instance, concerns about both communication issues and inadequate parking facilities no longer need to be classified in just one category.

LDA can also be used to process new concerns as they arrive. If LDA strongly classifies the concerns into a few topics, then that can probably be trusted, but if LDA assigns a relatively uniform distribution of topics to a patient concern, it can be flagged for manual review, possibly to detect a previously unidentified topic. This can certainly help to flag new and rare types of events, but can also help to identify rare safety events. LDA has already shown promise in identifying patient safety concerns[43] in a corpus of text obtained via keyword searches from web forums. These keywords themselves could be leveraged as domain information for LabEL (the equivalent of the manual category labels from AHS in our dataset), as can other information from the dataset such as relevant stakeholder group and community name.

In the context of the present study, the multiclass LDA labelling allows for the identification of concerns not well-captured by the LDA classification by searching for concerns with relatively uniform topic weight distributions. For example, consider the following patient concern:

> Complaint … states… the patient's private room where her newborn infant was in a crib by the doorway… was entered by strangers. Complainant states the strangers did not knock nor announce themselves upon entry and pulled back the curtains and were standing by the patient's bed. Complainant states that this incident could have led to criminal activity and/or kidnapping of the baby by strangers who walked into her private room and exited without being stopped by any nursing staff.

LDA does not classify this example well; the most significant topic only has weight 0.18. Concerns about security issues related to children do not form a clear LDA topic because they occur infrequently, and were not part of the original AHS classification framework, but this may nevertheless be a category that would warrant a specific action plan to address it. Combining LabEL with a process to flag concerns with high-entropy topic weight distributions can make LabEL into one method for flagging rare safety events in near-real time.

Strengths
One major strength of this study is the volume of data available, comprising tens of thousands of patient concerns over a time span of 8 years. This volume of data cannot easily be acquired via traditional survey methods. Another strength is the consistency of data capture; because the

reports are recorded by trained personnel, the summaries of the reports tend to be written in a consistent style. While patient concerns provided entirely in their own words can also be handled using this general approach, and provide advantages of their own, the consistent style adopted by trained personnel minimises spurious detection of topics that might be caused by natural variations in language as opposed to true variations in patient experience.

The usage of LDA in the domain of patient concerns received by a large healthcare system is also new, and combined with the consensus process, can provide a novel and practical way of classifying free-text patient concerns data, even for large health systems, that is responsive to concerns, without the need for a pre-existing framework.

Limitations

One limitation of our data is that health system workers provide the summarised free text used in this analysis after listening to the patient; it is not the text directly spoken by patients or caregivers. Many of the concerns are highly summarised, and did not provide any thematic information. Some of the text may reflect stylistic choices of specific PCCs, rather than the patients themselves. The actual text of patient concerns captured directly, for example, through a patient portal, would work well with this LDA approach, especially for more routine concerns that are likely to be highly summarised by the PCCs. AHS is presently in the process of deploying a province-wide patient portal in Alberta.

Another limitation is that the concerns data are not unambiguously linked to specific individuals, so it is not straightforward to determine if several comments are actually about the same issue, as reported by the same patient, which could potentially bias the LDA results. This would be more of a concern with a smaller dataset, and the upcoming AHS patient portal will also alleviate this concern going forward.

A specific issue we noted in our analysis of the FACT dataset was that the LDA model conflated 'Medicine Hat', the name of a city in Alberta, with concerns related to medication. This mistake is easily avoided by humans. NLP models based on bag-of-words representations of text are not good at dealing with polysemous words and special uses of words. Users of LabEL, LDA and other such methods should bear this in mind. Modern NLP analyses of health-related text[51] are often based on language models that leverage deep learning techniques to capture the context in which words are used, thereby mitigating this problem. An exploration of deep-learning-based topic models for this application could be a fruitful future avenue of research.

Finally, the keyword list is generated from FACT database entries so it may be overfitted to the current data, which may limit generalisability to unseen data. However, LabEL can be easily retrained periodically to update the classification framework over time. LabEL has the same model complexity as standard online LDA, and should have similar generalisability.

CONCLUSION

Free-text patient concerns are a critical source of data for healthcare quality improvement professionals. With patient portals launching all around the world, their volume will also certainly increase. While other health informatics studies have used LDA, this is the first study to analyse such a large corpus of general patient concerns from a large health system. Our study demonstrates that specific topic modelling of patient concerns can be performed using LDA, and that an increased volume of patient concerns data submitted through emerging patient portals need not create additional work for patient concerns consultants, if they are armed with these new tools. Moving forward, we will apply NLP methods for understanding the patient experience and capturing the patient voice to allow patients to provide feedback in manners most suitable to them and on topics not influenced by data collection tools.

The code we used is available at: https://github.com/theLongLab/manual_label_enhanced_LDA.

ORCID iD
Maria J Santana http://orcid.org/0000-0002-0202-5952

REFERENCES

1 Griffiths A, Leaver MP. Wisdom of patients: predicting the quality of care using aggregated patient feedback. *BMJ Qual Saf* 2018;27:110–8.
2 Coulter A. Patient feedback for quality improvement in general practice. *BMJ* 2016;i:i913.
3 Berger S, Saut AM, Berssaneti FT. Using patient feedback to drive quality improvement in hospitals: a qualitative study. *BMJ Open* 2020;10:e037641.
4 Kuluski K, Guilcher SJT. Toward a person-centred learning health system: understanding value from the perspectives of patients and caregivers. *Healthc Pap* 2019;18:36–46.
5 Liu JB, Pusic AL, Gibbons CJ, *et al*. Association of patient-reported experiences and surgical outcomes among group practices: retrospective cohort study. *Ann Surg* 2020;271:475–83.

6 Francis R. *Report of the mid-Staffordshire NHS Foundation trust public inquiry*. London, UK: The Stationary Office, 2013.

7 Clwyd A, Hart T. *A review of the NHS hospitals complaints system: putting patients back in the picture*. London, England: Department of Health, 2013.

8 Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. *BMJ Qual Saf* 2014;23:678–89.

9 Liddle AC. Understanding the experiences of nurses managing Querulous Complainants: what does health care know? 2020.

10 Servics AH. Alberta health services annual report, 2019-2020, 2020. Available: https://www.albertahealthservices.ca/assets/about/publications/2019-20-annual-report-web-version.pdf

11 Boylan A-M, Williams V, Powell J. Online patient feedback: a scoping review and stakeholder consultation to guide health policy. *J Health Serv Res Policy* 2020;25:122–9.

12 Gonzalez-Hernandez G, Sarker A, O'Connor K, *et al*. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform* 2017;26:214–27.

13 Cammel SA, De Vos MS, van Soest D, *et al*. How to automatically turn patient experience free-text responses into actionable insights: a natural language programming (Nlp) approach. *BMC Med Inform Decis Mak* 2020;20:97.

14 Shah AM, Yan X, Shah SJ. Use of sentiment mining and online NMF for topic modeling through the analysis of patients online unstructured comments. In: Chen H, Fang Q, Zeng D, eds. *Lecture notes in computer science*. Springer, 2018: 191–203.

15 Blei DM, AY N, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3:993–1022.

16 Lee YJ, Kamen C, Margolies L, *et al*. Online health community experiences of sexual minority women with cancer. *J Am Med Inform Assoc* 2019;26:759–66.

17 Zhao Y, Zang J, Wu M. Finding users' voice on social media: An investigation of online support groups for autism-affected users on facebook. *Int J Environ Res Public Health* 2019;16:4804.

18 Feldhege J, Moessner M, Bauer S. Who says what? content and participation characteristics in an online depression community. *J Affect Disord* 2020;263:521–7.

19 Moessner M, Feldhege J, Wolf M, *et al*. Analyzing big data in social media: text and network analyses of an eating disorder forum. *Int J Eat Disord* 2018;51:656–67.

20 Westmaas JL, McDonald BR, Portier KM. Topic modeling of Smoking- and Cessation-Related posts to the American cancer Society's cancer Survivor network (CSN): implications for cessation treatment for cancer survivors who smoke. *Nicotine Tob Res* 2017;19:952–9.

21 Okon E, Rachakonda V, Hong HJ, *et al*. Natural language processing of Reddit data to evaluate dermatology patient experiences and therapeutics. *J Am Acad Dermatol* 2020;83:803–8.

22 Lama Y, Hu D, Jamison A, *et al*. Characterizing trends in human papillomavirus vaccine discourse on Reddit (2007-2015): an observational study. *JMIR Public Health Surveill* 2019;5:e12480.

23 Adnan MM, Yin J, Jackson AM, *et al*. World pneumonia day 2011-2016: Twitter contents and retweets. *Int Health* 2019;11:297–305.

24 Dzubur E, Khalil C, Almario CV, *et al*. Patient concerns and perceptions regarding biologic therapies in ankylosing spondylitis: insights from a Large-Scale survey of social media platforms. *Arthritis Care Res* 2019;71:323–30.

25 Miller M, Banerjee T, Muppalla R, *et al*. What are people Tweeting about Zika? an exploratory study concerning its symptoms, treatment, transmission, and prevention. *JMIR Public Health Surveill* 2017;3:e38.

26 Zhan Y, Liu R, Li Q, *et al*. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. *J Med Internet Res* 2017;19:e24.

27 Pappa GL, Cunha TO, Bicalho PV, *et al*. Factors associated with weight change in online weight management communities: a case study in the loseit reddit community. *J Med Internet Res* 2017;19:e17.

28 Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. *J Med Internet Res* 2016;18:e108.

29 Yang M, Kiang M, Shang W. Filtering big data from social media--Building an early warning system for adverse drug reactions. *J Biomed Inform* 2015;54:230–40.

30 Zhu L, Reychav I, McHaney R, *et al*. Combined SNA and LDA methods to understand adverse medical events. *Int J Risk Saf Med* 2019;30:129–53.

31 Valenti AP, Chita-Tegmark M, Tickle-Degnen L. Using topic modeling to infer the emotional state of people living with Parkinson's disease. *Assist Technol* 2019. doi:10.1080/10400435.2019.1623342

32 Li Y, Rapkin B, Atkinson TM, *et al*. Leveraging latent Dirichlet allocation in processing free-text personal goals among patients undergoing bladder cancer surgery. *Qual Life Res* 2019;28:1441–55.

33 Liu Q, Chen Q, Shen J, *et al*. Data analysis and visualization of newspaper articles on Thirdhand smoke: a topic modeling approach. *JMIR Med Inform* 2019;7:e12414.

34 Hatakeyama Y, Miyano I, Kataoka H, *et al*. Use of a latent topic model for characteristic extraction from health checkup questionnaire data. *Methods Inf Med* 2015;54:515–21.

35 Wagland R, Recio-Saucedo A, Simon M, *et al*. Development and testing of a text-mining approach to analyse patients' comments on their experiences of colorectal cancer care. *BMJ Qual Saf* 2016;25:604–14.

36 Bahja M, Lycett M. Identifying patient experience from online resources via Sentiment analysis and topic modelling. in: proceedings of the 3rd IEEE/ACM International Conference on big data computing, applications and technologies 2016:94–9.

37 Ramage D, Hall D, Nallapati R. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In: EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009 2009:248–56.

38 Andrzejewski D, Zhu X, Craven M. *Incorporating domain knowledge into topic modeling via Dirichlet forest priors*. In: ACM International Conference Proceeding Series, 2009.

39 Gallagher RJ, Reing K, Kale D. *Anchored correlation explanation: topic modeling with minimal domain knowledge*. In: Transactions of the Association for Computational Linguistics, 2017: 529–42.

40 Meng Y, Shen J, Zhang C. Weakly-supervised neural text classification. in: proceedings of the 27th ACM International Conference on information and knowledge management 2018:983–92.

41 Zhou Z-H. A brief introduction to weakly supervised learning. *Natl Sci Rev* 2018;5:44–53.

42 Honnibal M, Montani I. spaCy2: natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing 2017.

43 Cho I, Lee M, Kim Y. What are the main patient safety concerns of healthcare stakeholders: a mixed-method study of web-based text. *Int J Med Inform* 2020;140:104162.

44 Hoffman MD, Blei DM, Bach F. Online learning for latent Dirichlet allocation. AdV neural Inf process Syst 23 24th Annu Conf neural Inf process Syst 2010. *NIPS* 2010.

45 Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30 http://jmlr.org/papers/v12/pedregosa11a.html

46 Nowell LS, Norris JM, White DE. Thematic analysis: Striving to meet the Trustworthiness criteria. *Int J Qual Methods* 2017;16. doi:10.1177/1609406917733847Article%20information

47 Institute of Medicine (IOM). *Crossing the quality chasm: a new health system for the 21st century*. Washington, D.C: National Academy Press, 2001.

48 Santana MJ, Manalili K, Jolley RJ, *et al*. How to practice person-centred care: a conceptual framework. *Health Expect* 2018;21:429–40.

49 Tapi Nzali MD, Bringay S, Lavergne C, *et al*. What patients can tell us: topic analysis for social media on breast cancer. *JMIR Med Inform* 2017;5:e23.

50 Factiva DJ. Factiva - Global News Monitoring & Search Engine. Available: https://professional.dowjones.com/factiva/ [Accessed 4 Feb 2021].

51 Shickel B, Tighe PJ, Bihorac A, *et al*. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform* 2018;22:1589–604.