Dec 12th, 12:00 AM

What matters most to patients? On the Core Determinants of Patient Experience from Free Text Feedback

Adegboyega Ojo
Maynooth University, adegboyega.ojo@mu.ie

Nina Rizun
Gdansk University of Technology, rizunnina68@gmail.com

What matters most to patients? On the Core Determinants of Patient Experience from Free Text Feedback

Completed Research Paper

Adegboyega Ojo
Maynooth University
adegboyega.ojo@mu.ie

Nina Rizun
Gdansk University of Technology
nina.rizun@pg.edu.pl

Abstract

Free-text feedback from patients is increasingly used for improving the quality of healthcare services and systems. A major reason for the growing interest in harnessing free-text feedback is the belief that it provides richer information about what patients want and care about. The use of computational approaches such as structural topic modelling for analysing large unstructured textual data such as free-text feedback from patients has also been gain traction lately. However, its use for generating insights is constrained by the apparent lack of statistical rigour and explanatory capability required for credible evidence in decision making. From the theoretical perspective, theory-building from unstructured textual data is also currently problematic in IS and health service research. This study presents an approach to address this challenge by integrating text analytics, predictive and quantitative models as part of a computational grounded theory approach to determine factors that significantly determine overall patient experience.

Keywords: patient experience, healthcare services, free-text feedback, structural topic modelling, computational grounded theory

Introduction

Patient feedback is one of the important mechanisms for determining the quality of healthcare services and systems. The quality of healthcare is measured by the extent to which provided care is person-centred, i.e. “respectful of and responsive to individual preferences, needs, and values" (Larson et al. 2019). While most of the feedback on healthcare service are collected through carefully designed and psychometrically sound national survey instruments (Price et al. 2014) with a focus on quantitative data, free-text responses are increasingly captured in surveys (Cunningham and Wells 2017). In general, analysis of free-text feedback often shed light on the experiences of patients that closed questions might not reveal (Ranard et al. 2016; Wiseman et al. 2015). Free-text comments about patient experience are also increasingly available on different publicly accessible platforms. However, this freely available resource is yet to be meaningfully exploited for novel actionable insights to complement findings from official surveys (Ojo and Rizun 2020) for different reasons. The barriers *inter-alia* include the cost and resources required to meaningfully process the text beyond automated thematic analysis (Maramba et al. 2015) commonly employed in practice to produce the deep insights characteristic of traditional qualitative analysis at scale. In this vein, structural topic modelling (STM) has emerged as a rigorous (semi) automated approach to the analysis of unstructured textual information with the possibility of generating rich insights (Kar and Dwivedi 2020; Müller et al. 2016; Schmiedel et al. 2019). Despite the promise and potentials of techniques such as STM for analysing patient feedback, their adoption and use for driving healthcare service improvement is constrained by the need for providers and policymakers to base their decisions and actions on sound statistical measures not provided in text analytics techniques. Moreover, care providers and policymakers are also faced with the dilemma of how to prioritise the required improvements and actions resulting from patient feedback. Some studies have already attempted to determine the predictor of patient satisfaction such as (Carcamo and Lledo 2001; Fiala 2012; Schoenfelder et al. 2011). Although these studies focused on different care contexts, e.g., in-patients, surgical treatment and plastic surgery, they are all quantitative. A recent study focused on the hospitality industry attempted to use topic proportions in comments as some measure of the importance of the topics (Korfiatis et al. 2019). Beyond this, we are unaware of studies aiming to provide a rigorous estimation of the importance of specific dimensions or themes regarding their effect on overall service quality.

In addition, studies involving the use of computational techniques and big data analytics related methods (such as STM) in IS research have been criticised for weak theoretical contributions (Kar and Dwivedi 2020). To address this challenge, inductive theory development combining human expertise and hermeneutic skills with the use of computational techniques—Computational Grounded Theory (CGT), has been proposed as a possible solution (Berente and Seidel 2014; Nelson 2020). CGT enables the derivation of emergent substantive theory grounded in the dataset by developing local or mid-range theories or by extending general theory in a domain (Berente and Seidel 2014). Nelson (2020) described a three-step methodological framework for CGT. A similar three-step approach albeit more abstract was earlier provided by Berente & Seidel (2014).

This study addresses both the knowledge gap of how to prioritise emergent themes from free-text feedback and the methodological challenge of how to build substantive theories inductively from large textual data. Specifically, our study aims to determine the latent topics in the free-text feedback on hospital care experience and which of these topics are the core determinant of the overall patient experience using a CGT approach. Our CGT approach operationalises the abstract process described in Berente & Seidel (2014) in four steps—(1) extraction of latent topics from the free-text feedback as "patterns" of interest; (2) mapping the topics to SERVQUAL constructs as the lexicon of the domain; (3) building a predictive model to generate plausible hypotheses of salient factors that strongly predict overall service ratings as the first step in substantive theory building; (4) validating the generated hypotheses using structural/regression models to refine emergent substantive theory. The rest of the paper is organized as follows. The next section explores the extant literature on patient experience and is followed by the description of the methodology employed in our work. We then present our results and discuss our methodological contributions as well as how our findings contribute to both theoretical and practical knowledge of factors that matter most in improving the overall hospital service quality rating and consequently, patient experience.

Literature Background

Patient experience is mainly associated with objective events including what happened in healthcare facilities or how often it happened (AHRQ 2017). In most cases, patient experience is related to the patients' impressions about the quality of medical care received (Lee 2017); timeliness of access to services (Pakdil and Harwood 2005); quality of tests and diagnostics (Shafei et al. 2015); explanation of the patient's current situation and perspectives (Fiala 2012); the emotional support and attitude of medical staff towards the patients and their relatives (Mourad et al. 2010); patient safety (Rathert et al. 2012), patient-provider trust (Shan et al. 2016) and the physical conditions of the hospital (Al-Damen 2017). One of the main objectives of patients experience research is to identify the determinants of patients' experiences and satisfaction and to determine the degree of their influence on the overall service quality. Researchers typical rely on patient responses to open and closed-ended questionnaire' questions, interview scripts, and free-text comments provided by patients on both closed and open Internet platforms.

The most popular model for evaluating service quality is the SERVQUAL model (Parasuraman et al. 1988). The model comprises five dimensions namely Tangibles, Reliability, Responsiveness, Assurance and Empathy. Many studies have confirmed the validity of using the SERVQUAL scale for measuring healthcare services quality (Valencia-Arias et al. 2018). There have also been some attempts at using SERVQUAL-based quantitative methods to determine the relative importance of SERVQUAL dimensions. In most of these studies, the SERVQUAL dimensions are treated as independent factors and the relative *importance* of the individual factors is computed as a function of the absolute mean differences between perception and expectation (i.e. gap) (Kumar et al. 2010). At the same time, there is a representative number of the studies casting doubt on the adequacy of the SERVQUAL-dimensions for service quality assessment and offering extended and modified versions of this model. In those studies, the authors typically developed context-specific measurement instrument that is based on closed-ended questionnaires and takes into account specific cultural factors (Karami et al. 2016), organizational factors (Lee 2017). To establish the importance of service quality dimensions, this category of studies employ exploratory and confirmatory factor analysis and structural equation models to identify the specific structure of patient experience factors (dimensions) and their *importance* (loading) (Otalora et al. 2018); multiple linear regression analysis (Jenkinson et al. 2002); machine learning algorithms (Bari et al. 2020).

Free-text patient comments and opinions are intended to complement the quantitative measures by providing information about experiences not covered by pre-defined factors from a closed-ended questionnaire structure and thereby providing more detailed information that can help contextualize closed-ended questions. (Wagland et al. 2016). One of the most popular approaches for analysing free-text responses (unstructured text) is the thematic analysis through coding of contents. This involves the manual development of a coding framework by a group of researchers with iterative discussion and agreement of the results (Chan et al. 2018). The developed frameworks are used for the subsequent (i) text categorization (coding); (ii) entities (patterns) extraction and grouping, (iii) deep statistical analysis of extracted entities, and (iv) determination of the core factors characterizing the service quality (Cunningham and Wells 2017). The frequency of mentions of a particular entity, extracted from the patient responses, is used as some measure of importance for service quality factors. An approach to analysing free-text is *topic modelling,* which allows automatic extraction of latent topics characterizing main texts. The most common techniques for topic modelling are Latent Dirichlet Allocation (LDA) and Structural Topic Modeling (Schmiedel et al. 2019). After the topics extracting, they could be labelled automatically (Kozbagarov et al. 2021) or using the experts' knowledge (in this case, the latent topic extraction process can be considered semi-automatic). Then the results are statistically (descriptive) processed. A measure of the importance of the extracted topics is associated with their respective proportions in the comment corpus. A few studies like (Ding et al. 2020) have also attempted at mapping extracted topics to SERVQUAL constructs. Some of these studies have employed Sentiment analysis (more often in combination with Topics Modeling) to rank topic importance (Ojo and Rizun 2019; Bahja and Lycett 2016).

Methodology

The object of our research is to determine core factors and related service quality aspects that most strongly impact the overall patient experience from free-text feedback. To this end, the study seeks to answer the following research questions: *(R1) What are the latent topics in the free-text feedback and how do they related to the SERVQUAL measurement construct? (R2) What are the core service quality factors, that determine the patients' overall hospital care service quality rating?* We address these research questions using the Computational Grounded Theory method (Berente and Seidel 2014; Nelson 2020). CGT allows combining expert human knowledge and skills in interpretation with the use of computational techniques to analyse a large corpus of text to achieve a reproducible grounded theory approach (Nelson 2020). Grounded theory enables the development of themes inductively from data for a theoretical understanding of the phenomenon under study rather than follow the more popular hypothetical-deductive reasoning approach adopted in IS research. Berente & Seidel (2014) outlined a process for theorizing from big (textual) data to include three steps—(1) identification of patterns from data, (2) filtering the patterns through the lens of a theoretical lexicon and (3) generation of novel theory. However, no specific set of techniques were prescribed for operationalizing these steps. While Nelson (2020) also prescribed a three-step methodological framework for obtaining CGT, the specific techniques proposed are not sufficient to address our research question. Albeit, Berent & Seidel's process is abstract, it provides a more general framework suitable for answering our research questions. Our research design outlines how we operationalise CGT process of Berent & Seidel (2014).

Research Design

To realize the aim of our research, we adopt a research design that comprises four-step: *Pattern Identification—Step 1: Extraction of the latent topics from free-text comments*—we employed the *Structural topic modelling,* an extension of the LDA framework. LDA is one of the well-known unsupervised learning-based text analysis methods, which provides both a predictive and latent topic representation of the corpus (Blei et al. 2010). This method is widely adopted in customer experience studies (Schmiedel et al. 2019) for exploring the topics contained in the free-text feedback and patients comments (Ojo and Rizun 2020). In STM models the topic prevalence (content) is specified in the form of generalized linear models parameterized by document specific covariates X(Y) (Hu et al. 2019). These covariates inform either the topic prevalence (covariates X) or the topical content (covariates Y) latent variables with information "about the text" (metadata). Also, to the topics labelling and interpretation, four different types of word weightings are provided in STM: (i) Highest Prob: are the words within each topic with the highest probability (inferred directly from topic-word distribution parameter η); (ii) FREX: are the words that are both frequent and exclusive, identifying words that distinguish topics; (iii) Lift: give more weight to words that appear less frequently in other topics by dividing their frequency into other topics; (iv) Score: score words are weighted by dividing the log frequency of the word in the topic by the log frequency in other topics (Chang 2015).

Lexicon Filtering—Step 2: Mapping the latent topics to SERVQUAL model—to enable us associate the extracted topics from text with service quality dimensions, we carried out three steps, namely: (i) manual lifting of the latent topics to SERVQUAL dimensions based on definitions SERVQUAL scales in literature; (ii) machine learning-based clustering to identify the semantic similarity between latent topics; (iii) the alignment of the manual lifting and machine learning-based clustering results to improve SERVQUAL dimensions consistency. The K-means clustering algorithm is a typical unsupervised learning algorithm that classifies samples into k clusters (Rashid et al. 2019), which is successfully applied for textual data analysis (Rejito et al. 2021). By calculating the spatial similarity between each coordinate of the topic-words distribution vectors η and the centroid of each cluster, the whole set of topics were split into several groups (factors). We used the correlation coefficient as a measure of topic semantic similarity. The results from the clustering process were used to refine the manual lifting of topics to SERVQUAL.

Theory Building—Step 3a: Determining core service quality factors (latent topics)—we first build a random forest classifier with latent topics (described by document-topic proportions θ) as an independent variable and the overall service quality rating as the outcome variable. We then compute the Gini importance metrics of the resulting Random Forest classifier as a basis for determining the saliency of the topics as predictors of service quality ratings. As a classifier, Random Forest performs an implicit feature selection, using a subset of "strongest variables" for the classification only (Breiman 2001) to reach the best performance on high dimensional data. The outcome of this implicit feature selection of the random forest can be determined and visualized by the *Gini Importance* (Mean Decrease in Impurity, MDI) (Menze et al. 2009). In our work, we used the Random Forest classifier to determine the predictive power of the obtained latent topics regarding the comments rating; and used the Gini Importance measure to identify core service quality factors which most strongly affect overall patient experience or service quality rating.

Theory Building—Step 3b: Validation of core factors—we validate the core set of factors produced in Step 3 by building a structural equation model describing the effects of the core service quality factors expressed as SERVQUAL dimensions on the overall patient experience (Bhoomadevi et al. 2021). Structural equation modelling (SEM) is a powerful multivariate analysis technique that is widely used in information systems and social sciences research. It enables analyzing complex relationships between multiple variables, and also allows validation of theory using empirical models. Our SEM model comprised SERVQUAL latent constructs measured by the topic's proportions or distribution (observable variables). The SEM model was fitted using lavaan[1] R package. Our decision about the model's goodness of fit was based on relevant indices including the Comparative Fit Index (CFI), which compares the existing model with a null model; (ii) Root Mean Square Error of Approximation (RMSEA), which is the square root of mean differences between the estimate and the true value; (iii) Tucker-Lewis index (TLI), which is also called the non-normed fit index (NNFI). The TLI represents the distance (in terms of fit) between your model and the independence model as a proportion of the distance between the independence model and the saturated model) (Hu and Bentler 1998); (iv) Standardized root mean square residual (SRMR); the difference between the observed correlation and the model implied correlation matrix and allows assessing the average magnitude of the discrepancies between observed and expected correlations as an absolute measure of (model) fit criterion (Pavlov et al. 2021). A good "model–data" fit is indicated by RMSEA < .06, SRMR < .08, CFI > .95, and TLI > .95 (Xia and Yang 2019). Using these indices, we re-specified our SEM model by changing the number of model variables in the model to improve fit results. The final model was compared with the core factors obtained in step 3. A summary of our overall research design is presented in Figure 1.

[1] https://lavaan.ugent.be/

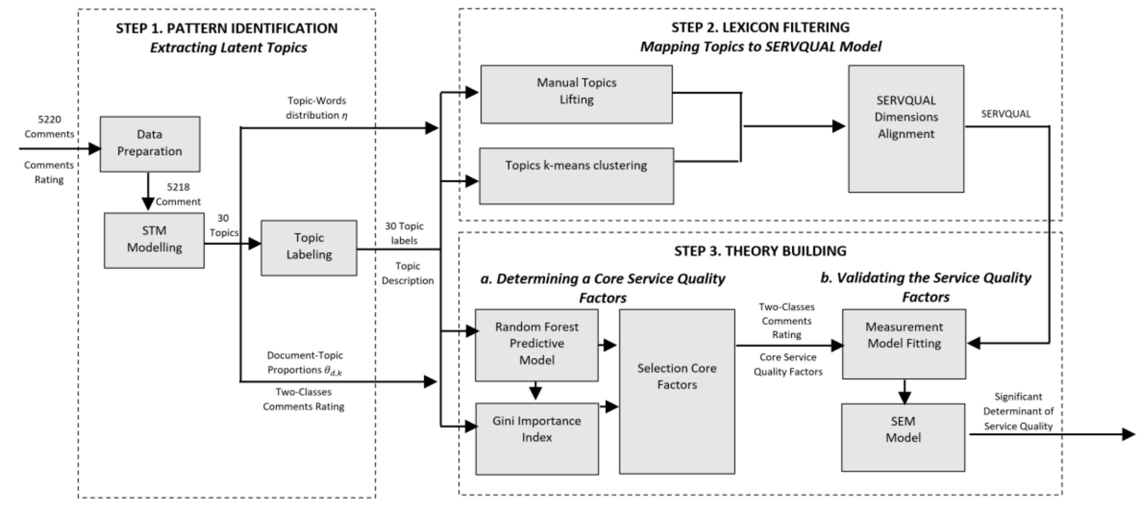

Figure 1. Research Design

Data Collection

The data source of our research is the free-text comments provided on the hospital review platform "ratemyhospital.ie"[2]. We collected a total of 5220 anonymous comments from the period 1st December 2010 till 31st December 2019. Each comment in the dataset contains (1) comment rating (from 0 to 5.5); and (2) comment date. Subsequent data preparation consists of (i) text pre-processing, (ii) converting data set into STM text corpus format, composed of three elements namely the document term matrix, vocabulary character vector and the metadata matrix containing document covariates; (iii) converting the Rating data into two main categories—Low (rating values are from 0 till 3) and High (rating values are above 3). This step produced our final dataset which contains 5218 comments referred to two categories of service quality rating. The summary of the dataset after preparation comprising the distribution of the Number of comments and percentage (%), as well as average comment length (words) are given in Table 1.

Data		Number	Percentage	Average comments length
Number of comments		5218	100	406.70
Distribution of Comments Rating	Low	2136	40.94	539.62
	High	3082	59.06	315.33

Table 1. Sample Summary

Data Analysis

Extracting Latent Topics

We employed STM for determining the latent topics in the feedback comments about hospital services and the distribution of these topics per comment. STM implements these tasks by building a generalized linear model of the influence of document-level covariates on the topical prevalence parameter μ, which in turn determines the θ document-topic proportions (DTP). In this step, *first*, the STM model was set up. To determine the optimal number of topics, STM models ranging from 10 to 100 topics were built as part of a model selection procedure. While a combination of the coherence and exclusivity scores of the models are prescribed for selecting models, there are specific guidelines on how to combine these metrics (Schmiedel et al. 2019) to identify the best model(s). Here, we adopted a simple but conceptually sound approach in identifying models with coherence and exclusivity metrics closest to a theoretically optimal one. By normalising the exclusivity and coherence scores (see Figure 2), we determined that Models with 50 and 30 topics have the shortest Euclidean distance to the *theoretical optimal model* (i.e. the model with

[2] http://www.ratemyhospital.ie

exclusivity and coherence scores of 1). Given the closeness of the distance scores of these two models (i.e. Model 50 & 30), we selected the model with 30 topics based on the principle of parsimony. The choice of a simpler topic model reduces the possibility of redundancy in topic space and the need for merging topics.

STM model	Exclusivity (normalized)	Coherence (normalized)	Distance to ideal STM Model
Model 100	1.000	0.000	1.000
Model 90	0.961	0.086	0.915
Model 75	0.745	0.254	0.789
Model 70	0.699	0.292	0.769
Model 80	0.587	0.370	0.753
Model 65	0.505	0.382	0.792
Model 60	0.495	0.434	0.759
Model 50	0.481	0.499	0.721
Model 45	0.371	0.546	0.775
Model 30	0.258	0.910	0.747
Model 40	0.222	0.593	0.878
Model 20	0.175	0.796	0.850
Model 25	0.130	0.746	0.906
Model 15	0.120	0.642	0.950
Model 35	0.058	0.654	1.004
Model 10	0.000	1.000	1.000

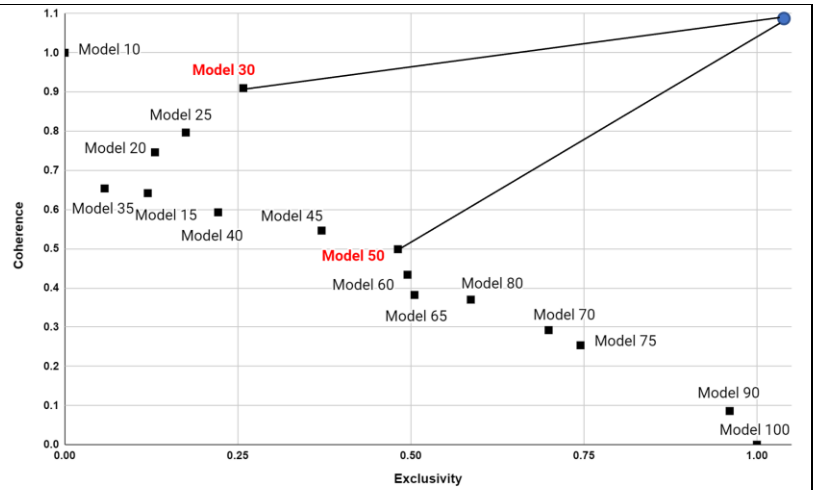

Figure 2. Optimal semantic coherence and exclusivity scores of STM models

Second, we built the STM model to produce the topic-words distribution η; document-topic proportions θ; list of Highest probability-, FREX-, Lift- and Score-keywords; and the set of documents, mostly associated with each topic. As a result, 30 Latent topics, which are described by the (i) top-weighted keywords and the (ii) set of documents, mostly associated with this topic (based on the modal estimate of the proportion of word tokens assigned to the topic under the model) were obtained. *Third*, the process of topics labelling was carried out iteratively: (1) two experts independently labelled the topics to produce the first version of labels based on top weighted keywords (STM model outcome); (2) experts discussed the labels and resolved differences in labelling; (3) the experts independently refined topic labels based on the computationally guided deep reading of 20 of the most representative (or exemplar) comments of the topics; (4) experts discussed to align the refinements done in the previous step by jointly reading and analysing most representative comments, and (5) experts agreed on the final set of topic labels and described the topics.

Mapping Topics to SERVQUAL Model

We realized this step in three stages: manually assigning the latent topics to SERVQUAL dimensions to obtain the best quality of interpretation; Machine Learning-based clustering to identify the semantic similarity between latent topics, and the alignment of the Manual assignment and Machine Learning-based clustering results to improve SERVQUAL dimensions consistency. To realize the first *stage,* the following analyses were carried out: (1) the definitions associated with each of the five SERVQUAL dimensions in the extant literature and related to healthcare services were reviewed; (2) two experts independently develop the first version of SERVQUAL dimension definitions; (3) experts integrated the SERVQUAL dimension definitions. Next (1) two experts independently mapped each of the topics to exactly one SERVQUAL dimension and subsequently met to integrate and agree on the mappings. (2) Checking the internal consistency of the SERVQUAL dimensions concerning their composite topics using Cronbach's alpha coefficient α using the topic-word distribution η. Cronbach's alpha coefficient is used to assess the reliability of a set of topics within the particular SERVQUAL dimension. Considering the bottom-up and inductive nature of our analysis, we adopted a threshold value of Cronbach's alpha α_s=0.5.

In *Stage 2*, to identify the semantic similarity between latent topics, first, we applied the k-means clustering algorithm, using correlation coefficient (r) between the vectors of topic-words distribution η as a measure of topics semantic similarity. The resulting clusters were proposed were used to refine the manual assignment of topics done in stage 1. Second, we identified topics that were semantically problematic for a particular SERVQUAL dimension. We consider a topic to be semantically problematic if: (i) the maximum value of the Correlation coefficients for a given topic with other topics in this dimension is insignificant (we accept threshold value $\xi_r \leq 0.3$), and (ii) results of removing this topic from the SERVQUAL dimension causes an increase in the internal consistency of this dimension.

In *Stage 3*, we attempt to improve the SERVQUAL dimensions' internal consistency and reliability, we iteratively carry out topic reclassification and exclusion of problematic topics based on the results from stages 1 and 2. For this purpose, specifically, we refine the manual mapping with results from the cluster analysis based on the following rule: a topic will be moved from one SERVQUAL dimension to another if the topic is not in the SERVQUAL dimension where most of its cluster members (as determined in stage 2) are. The reclassification of the topic is accepted if significantly increases the reliability value of the target SERVQUAL dimension but only slightly reduces the reliability of the source SERVQUAL dimension. We terminate the iterative process when most SERVQUAL dimensions have a value of Cronbach's alpha at least α=0.7 and no further improvement was obtained concerning Cronbach's alpha value of the dimensions. Topics that did not fit into any of the existing SERVQUAL dimensions were assigned to a new group.

Determining a Core Service Quality Factors

The supervised learning Random Forest (RF) classifier was applied to determine the predictive power of latent topics regarding the overall service quality rating by patients. The main classification model *is as follows*: (i) the overall service quality rating is the output variable. For classification purposes, the Rating variable was converted from numerical into a categorical scale (High and Low) in the data preparation step (Table 1); (ii) document-topic proportions $\theta_{d,n}$ (the topic proportion for topic n in document d)—was assigned as an input or predictor variable; (iii) training and test subsamples were selected from the targeted dataset with the proportion 70%:30%. The number of the selected features (latent topics) for the initial model is *N*=30. For RF analysis, RF classification tree methods (number of trees =300; the number of variables tried at each split =3) is used. For evaluation, an evaluation method the Confusion matrix was adopted, that commonly used to present performances of classifiers in classification tasks and shows the associations between real class attributes and that of predicted classes. For model performance, accuracy as a percentage of correct outcome among the test sets was used. *Second*, to determine the core set of service quality factors, the Gini Importance Index was adopted. *The main goal at this stage was to identify the top M factors (M<N), related to the latent topics, that most strongly determine the overall service quality rating without significantly reducing the accuracy of the initial model.* A stepwise elimination of the input variables of low importance was carried out. This involved: (1) using the Variable Importance Plot to select the group of factors, with relatively high Gini Importance; (2) checking the accuracy of the predictive model with the selected group of factors as an input variable; (3) checking if the drop in model accuracy relative to the initial model is not more than 2.5%.

Validating the Core Service Quality Factors

A Structural Equation Model was developed to examine the structural relationship between core latent topics, SERVQUAL dimensions and healthcare service quality patients' rating. The *first* step here includes developing hypotheses about the relationships among core service quality factors (expressed as SERVQUAL dimensions) and the overall service ratings. *Second*, to specify the model, a measurement model for the latent variables based on the latent topics was specified. When the full measurement model for the SERVQUAL dimensions as latent constructs was specified using all 30 latent topics as observed variables, however, the covariance matrix for the model was found not to be positive definite. This was followed by *reducing* the number and composition of independent variables using the GINI Importance index for the variables to establish the optimal structure of the model. Specifically, two independent experts interactively re-specified the model by eliminating and including individual factors (topics) and the corresponding latent variables (SERVQUAL dimensions) to improve Goodness of Fit (GOF) results (Zhang 2017). To coordinate the process of selection of factors to be included in the model, the experts agreed to start with the top-10 core most influential topics identified in the previous step of the study. Next, the experts discussed the resulting models' quality (CFI/TLI >.95, RMSEA <.06, SRMR <.08) and agreed on the final model structure. The final SEM model is based on 7 variables (topics) and four related SERVQUAL constructs.

Findings

Latent topics Associated with Patient Experience

A total of 30 topics were identified from the dataset and labelled as described in our methodology. Each labelled topic is characterized by (1) topic label, (2) topic description and (3) topic proportion (TP, %) and exemplar comments for the topics are provided in the repository[3]. The top five (5) most common topics account for 34.13% of the analysed comments and include *Care team friendliness* (7.9%); *Medics care and attention* (7.5%); *Organization of care processes* (7.4%); *Waiting time* (6.2%); and *Communication from doctor* (5.2%).

Mapping Topics to SERVQUAL Model

After mapping the latent topics described in the previous section to the SERVQUAL dimensions, we obtained the results presented in Table 2. At the end of the first round of manual mapping (Segment 1 of Supplemental material[4]), most of the topics could be assigned exclusively to one of the five SERVQUAL dimensions. A few topics that did not fit into the original SEVQUAL dimensions were assigned to a new dimension. Our experience here is consistent with the findings from existing studies such as (Lai et al. 2007) which revealed the need for additional dimensions to the SERVQUAL model to measure service quality in different domains. This stage of assignment produced reliability values of between 0.2 and 0.69 for the SERVQUAL dimensions. The Responsiveness dimension had the lowest reliability value of 0.2 while the Reliability dimension had a value of 0.69. Given that not all dimensions meet the threshold of Cronbach's alpha (α_s=0.5) adopted in our study for reasons discussed earlier, the Responsiveness and General Experience dimensions had to be further refined. After the clustering of the topics in the second step (Segment 2 of Supplemental material[4]) six clusters were obtained. We could only interpret 4 out of 6 obtained clusters. Despite the challenge of interpretation, the reliability values of the clusters were between 0.3 and 0.76, which were significantly better than the manually determined topics. Lastly, we aligned the manual mapping and the machine learning-based clustering to improve the internal consistency and reliability of the SERVQUAL dimensions. After iteratively reclassifying topics through the movement of the topics between dimensions as described earlier, we obtained the results in Table 2.

Given the high correlation between the topics in Assurance and Empathy, we decided to merge these two dimensions into one resulting in our final four SERVQUAL constructs—Tangibles, Reliability, Assurance & Empathy, Responsiveness—was built. This result is consistent with the findings from existing studies such as (Fiala 2012; Naidu 2009) which suggested merging SERVQUAL dimensions due to their high correlations. The final set of SERVQUAL dimensions all have acceptable Cronbach's alpha values ($\alpha \geq 0.5$). We can summarize our findings from the mapping follows, that patients are most concerned the: (1) readiness and willingness of hospital staff to offer timely services (Responsiveness, Total topic proportion (TTP)=32.0%, 7 topics); (2) the capacity of healthcare service to engender trust, support patients through courteous services; as well as guarantee the hospital staff politeness, kindness, attention and respect to patient needs (Assurance & Empathy, TTP=29.0%, 8 topics) and (3) that patients are less likely to discuss the degree of issues related to the conditions such as staffing conditions and equipment and physical environments at the hospital (Tangible, TTP=11.10%, 5 topics). The last point may be related to the fact most of the feedback analysed appeared to be related to out-patient care.

Determining Salient Service Quality Factors

While patients may provide more feedback about some aspects of care, the aspects of their feedback that most strongly predicts their overall service quality rating and experience is very important. These core factors (a subset of the 30 latent topics) have high predictive power regarding the overall service rating. For this, *first*, the initial Random Forest predictive model with 30 input variables (document-topic proportions $\theta_{d,n}$) and overall service quality Rating (Table 1) as an output variable, with D=5218 observations (comments), was built.

The accuracy of the initial Random Forest classifier was 83.59%. *Second*, we generated the variable importance plot (Figure 3), showing by Mean Decrease Gini Index (MDGI) for all 30 topics. The variable importance plot expresses how much accuracy the predictive model loses by excluding each latent topic. The MDGI can be interpreted as a measure of how document-topic proportions $\theta_{d,n}$ of each latent topic contributes to the homogeneity of the nodes and leaves in the resulting Random Forest.

[3] Topics Labels and Comments Examples

[4] Steps of Mapping Topics to SERVQUAL Model

Topic Labels	Topic Description	Topic Labels	Topic Description
TANGIBLE	*Cronbach's Alpha= 0.70, TTP=11.10%*	*RESPONSIVENESS*	*Cronbach's Alpha= 0.56, TTP=32.00%*
Hospital hygiene	Hygiene of hospital's premises and equipment	Organization of care process	Patient care in the ward, the organization and sufficient number of the medical staff
Hospital facilities (beds)	Patient conditions on the ward, privacy and comfort of beds, patients in trolley	Hospital environment standards	Level of hospitals standards (cleanliness, living conditions, equipment)
Hospital management & care	Coordination of care, treatment, diagnosis. Procedure's clearness	Waiting time	Problems with the time and conditions of waiting for a doctor's appointment
Nursing staff attitude	Nurses attitude towards patients	Care team responsiveness	Hospital staff kindness and friendliness
Professional practices	Overall experience in medical professional practices	Hospital service experience	General experience of hospital staying (efficiency, quality and speed of service)
RELIABILITY	*Cronbach's Alpha= 0.71, TTP=28.10%*	Timely service	Care and follow up rapidness
Care team professionalism	Staff professionalism and working in teams	Hospital care experience	General experience of hospital staying (patient care, treatment, communication)
Maternity Mishaps	Patient's experience associated with maternity mishaps	*ASSURANCE & EMPATHY*	*Cronbach's Alpha= 0.70, TTP=29.00%*
Procedure & Surgery	The organization and holding of medical operations and follow-up care	Nursing care & attention	Medical staff care and personal attention to patients
Medical tests	The quality of the conduct and interpretation of medical tests	Communication from doctor	The level of doctor's respect and openness, politeness to patients
Administrative side of the hospital, costs	Administrative issues in the hospital, additional fees (for treatment, parking)	Communication with patient/family	Problems with communication between doctor and a patient's family
Care & treatment in emergency	The level of professionalism, speed of decision-making and care in emergency	Maternity ward care	Midwives' professionalism and care in the maternity ward
Staffing situation	The quality and organization of staff work, staff overloading, staffing issues	Infant care assurance	Communication issues, a sense of confidence and security of patient children
Dementia care	Experience in the care and treatment of patients with brain diseases	Staff attitude	The attitude of the hospital staff towards the patients, the degree of respect and attention
Older patient care	Elderly care issues	Patient care and safety	Patient safety and compassion during care
Acute care	Quality of decision making and treatment outcomes for acute diseases	Procedure and treatment	Procedure and treatment quality

Table 2. Results of Mapping Topics to SERVQUAL Model

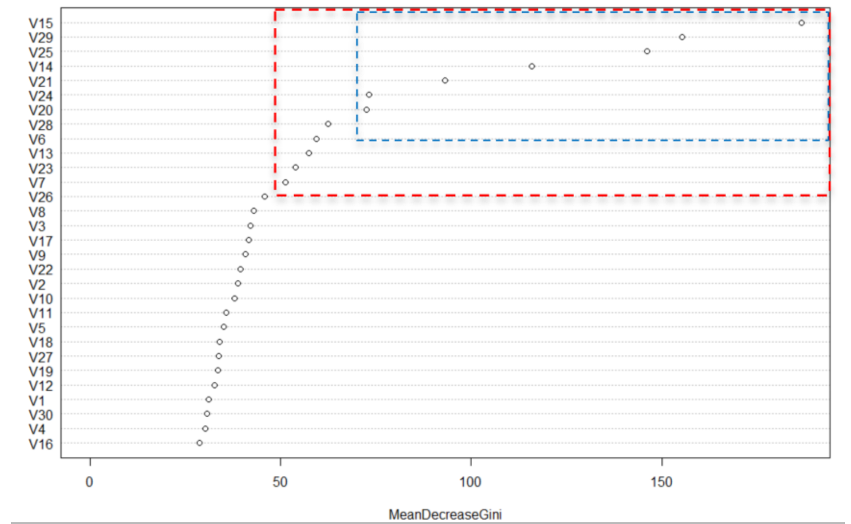

Figure 3. Service quality Factors Importance Plot

Based on the approach described earlier in our methodology, we identified topics (or factors) with high values of the Gini Importance visually (indicated by the *red* dotted line in Figure 3). This group contains 12 latent topics. Checking the accuracy of the predictive model with this group of 12 factors as an input variable yielded a model with 0.1% less accuracy compared to the initial full model of 30 topics. The next iteration involving only the top *seven* factors, (indicated by the *blue* dotted line in Figure 3) produced a predictive accuracy of which was 1.3% lower than the initial model. The summary of the results of the predictive models' performance is presented in Table 3. Thus, the top *seven* factors are sufficient to determine the overall patient experience. These seven factors are associated with (i) *Assurance & Empathy* dimension—*Medics care and attention* (Topic 15), *Infant care assurance* (Topic 29), *Staff attitude* (Topic 20); (ii) *Responsiveness* dimension—*Staffing situation* (Topic 26), *Timeliness of Service* (Topic 14), *Organization of care process* (Topic 21); and (iii) *Reliability* dimension—*Care team professionalism* (Topic 24).

RF- Performance measures	30-factors RF model	12-factors RF model	7-factors RF model
Accuracy	0.8359	0.835	0.8249
P-Value	2.20E-16	2.20E-16	2.20E-16
Kappa	0.6607	0.6587	0.6366
Sensitivity	0.8896	0.886	0.8746
Specificity	0.7636	0.7657	0.7558
Precision	0.8351	0.8383	0.833
Recall	0.8896	0.886	0.8746
Prevalence	0.5736	0.5782	0.582
F1	0.8615	0.8615	0.8533

Table 3. Comparative RF-performance measures

Validation of Core Service Quality Factors

Following the results above, a structural equation model was developed to validate the core factors by testing the statistical significance of the effect of the related SERVQUAL dimensions on the overall service quality. Thus, we proposed the following hypotheses: *H1*: "Tangible factors significantly influence overall service quality rating"; *H2*: "Reliability dimension-related patient experience significantly influences overall healthcare service rating"; *H3*: "Responsiveness factors significantly influence overall service quality rating"; *H4*: "Assurance & Empathy factors significantly influence overall service quality rating". Based on the procedure described earlier for this step of our approach, we obtain the model with the GOF indices shown in Table 4. As shown below, the model has a satisfactory fit across all indices, except for the χ2 test (χ2 [*df*] = 6.076; p = 0.000).

	χ^2	df	χ^2/df	CFI	TLI	NFI	RMSEA	SRMR
Recommended values	N/A	N/A	< 3.0	> 0.9	> 0.9	> 0.9	< 0.08	<0.08
Fitted Structural Model	139.750	23	6.076	0.969	0.932	0.966	0.043	0.025

Table 4. SEM Model Reliability Test Summary

A reflective measurement model was adopted in relating the constructs to the topics as shown in Table 5. Specifically, the measurement model includes (i) *seven* selected observed variables: Topic 13. *Nurses Professionalism*, Topic 6. *Maternity Mishaps*, Topic 21. *Organization of care*, Topic 14. *Timeliness of Service*, Topic 25. *Care team responsiveness*, Topic 15. *Medics care and attention*, and Topic 29. *Infant care assurance;* which are reflectively associated with the *four* latent endogenous SERVQUAL variables or constructs.

	Variables in the measurement model	Std. path coefficient	Estimate	SV	SE	z	p
1	Tangible (Tng) ← Topic_13	1	1		Fixed		
2	Reliability (Rlb) ←Topic_6	1	1		Fixed		
3	Responsiveness (Rsp) ←Topic_21	0.395	1	0.844	Fixed		
4	Responsiveness (Rsp) ←Topic_14	0.344	0.944	0.881	0.050	18.946	***
5	Responsiveness (Rsp) ←Topic_25	0.440	1.423	0.807	0.065	21.916	***
6	Assurance & Empathy (A_E) ← Topic_15	0.378	1	0.857	Fixed		
7	Assurance & Empathy (A_E) ←Topic_29	0.230	0.421	0.959	0.030	14.275	***

*** $p < 0.05$ **Table 5. Measurement Model Summary**

The summary of standardized path coefficients of the best-fit measurement model with information about the estimation of unstandardized factor loadings (Estimate), standardized variance (SV), Standard Error (SE), z-value (Wald test), p-value (p), is presented in Table 6. All coefficients are statistically significant at p<0.05, thus their significance to the model is amplified. Table 6 and Fig. 4 show the standardized path coefficients of the structural model that indicates the strength of the direct relationship between constructs. The results from the SEM modelling show that *Assurance & Empathy* significant affect the overall service rating or patients experience ($\beta_4 = 0.966$, $p < 0.05$), supporting H4.

	Variables in the structural model	Std. coefficient	Estimate	SE	z	p
H1	Tangible (Tng) → Service Quality Rating	0.270	4.403	0.975	4.518	***
H2	Reliability (Rlb) → Service Quality Rating	0.361	4.709	1.014	4.649	***
H3	Responsiveness (Rsp) → Service Quality Rating	0.411	17.551	10.816	1.623	0.105
H4	Assurance & Empathy (A_E) → Service Quality Rating	0.966	30.962	12.293	2.59	***

*** $p < 0.05$ **Table 6. Relationships with Core Service Quality Factors and Overall Healthcare Service Quality Rating**

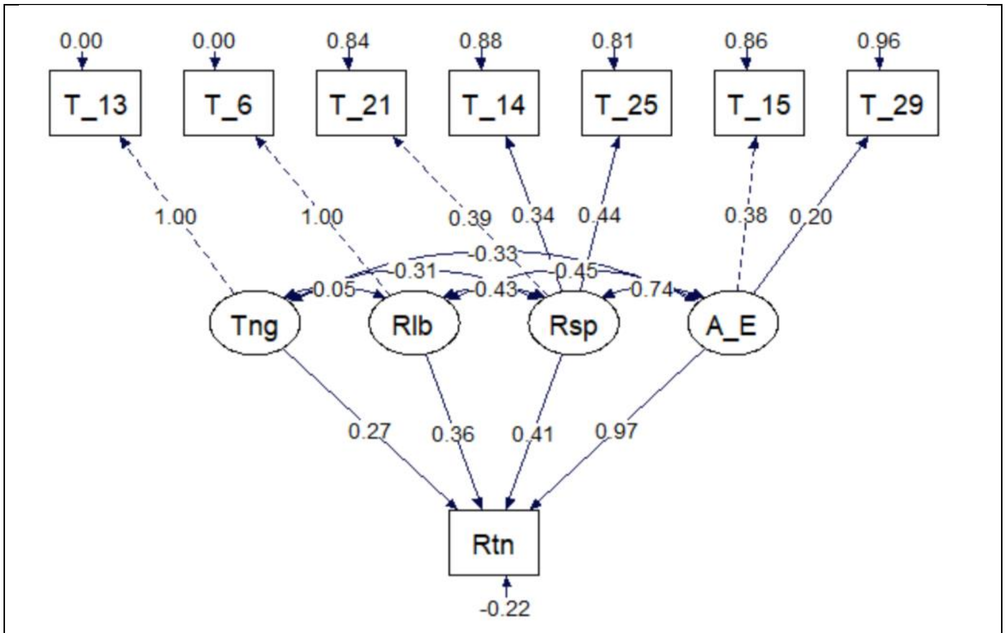

Figure 4. Final Structural Equation Model

In addition, *Reliability* ($\beta 2 = 0.361$, $p < 0.05$) and *Tangible* ($\beta 1 = 0.270$, $p < 0.05$) related factors (as posited in H2 and H1 correspondingly) significantly influence overall service quality rating. However, there was no support for our hypothesis that Responsiveness-related factors affect overall service quality ratings or customer experience $\beta 3 = 0.411$, $p > 0.05$). So, H3 was not confirmed in our data. The Responsiveness construct has strong co-variance or correlation with Assurance and Empathy (0.74). *We thus conclude that from our analysis that Assurance and Empathy is the most influential factor for service quality rating followed by Reliability and Tangible related factors.*

Discussion and Conclusion

Free-text feedback on services (e.g. hospital care services) from publicly accessible platforms is a valuable source of information on what service users value and do not value in the service experience. The use of computational techniques for processing large unstructured textual datasets continues to receive attention in the IS research community. However, the are increasing calls for greater rigour in studies employing these computational and big data analytics methods to enable the development of novel and important theories. One of the emerging approaches for theory building from large datasets of textual information is the CGT approach (Berente and Seidel 2014; Nelson 2020).

In this study, we have operationalized the CGT method to develop important insights into service quality factors that determine the overall service quality ratings from the perspectives of patients in the context of hospital care services. Specifically, *we established that the assurance and empathy dimension (including* medics care and attention, infant care assurance and staff attitude*) of hospital care services have the strongest influence on the overall service ratings and consequently patient experience.* We also established that to a relatively lesser degree, the reliability of hospital care and the tangible dimension of care (care team professionalism in particular) do influence the service ratings. However, even though our predictive model identified the responsiveness of hospital personnel in providing the needed care as a strong influential factor, we could not confirm this relationship in our structural model. The assurance and empathy and tangible dimensions are directly linked with the experience of the patients in their interaction with nursing and medical staff. While there were no specific metadata in our dataset that indicated the specific care context for each comment analysed, our exploration of the comments suggest that comments mostly related to outpatient care. These results constitute the substantive (mid-range) theory generated inductively from our dataset.

Other studies such as (Fiala 2012) have shown that patients assess the quality of their hospital experience based on functional quality of care service, that is the process by which care service is delivered and not the actual procedure itself. Whereas the providers tend to focus more on the technical quality of care, can be considered as a level of compliance with technical standards (for example, in surgery, return of function, absence of mortality, or lack of perioperative complications). This finding is consistent with ours as assurance and empathy are primarily associated with the functional aspect of care. In (Jenkinson et al. 2002), based on regression analysis, the authors found significant determinants of service quality were identified to be related to physical comfort, emotional support, and respect for patient preferences. In (Mourad et al. 2010), the five aspects of care were ranked by patients for relative importance with "Doctor's attitude" as the most important aspect of care (36%), followed by "information provision" (29%), "organisation of diagnostics" (24%), "waiting time" (8%), and "emotional support" (4%). According to (Rehaman and Husnain 2018), assurance, empathy and tangible aspects of care were found to be significant predictors of patient satisfaction level, but reliability and responsiveness were insignificant. Furthermore, authors of (Alkuwaiti et al. 2020) reported that patients' satisfaction is mostly explained by doctor competency and professional care (*Care team professionalism,* Topic 24), waiting time (*Timeliness of Service*, Topic 14); also (Rathert et al. 2012), that identify staffing issues (*Staffing situation,* Topic 26) and medication administration (*Organization of care process,* Topic 21) as the main factor of patient trust and satisfaction. (Shan et al. 2016) reported that good staff attitude (*Staff attitude,* Topic 20) was mostly associated with patient satisfaction. Similarly, it was reported in (Carcamo and Lledo 2001) that in addition to treatment outcome (in the context of surgical care), other factors affecting patient satisfaction include patient-doctor relationship, personal attention, communication with medical staff, and courtesy. Nursing kindness was found to follow treatment outcome as the most salient predictor of patient satisfaction (Schoenfelder et al. 2011). These results provide support for our substantive mid-range theory. Our work reinforces the view in (Ranard et al. 2016) that concrete actionable insights could be produced from unstructured or free text data such as those on Yelp or the "ratemyhospital.ie" platform; the source of the data analysed for this research.

Our work makes three important *contributions* to IS research and Service Quality studies. *Firstly*, it makes a *methodological* contribution by demonstrating how the major steps of the computational grounded theory approach could be operationalized to generate substantive theory grounded in the data. The efficacy of the different methods employed within our research design has been established. Thus, our CGT approach as a whole could be applied in similar studies in different domains. *Secondly*, the work makes two *theoretical* contributions: (a) it offers a novel and highly contextualized measurement model for SERVQUAL based on latent topics uncovered from our datasets. This could also be seen as the adaptation of SERVQUAL lexicons or constructs for the domain of hospital care services (Berente and Seidel 2014); (b) it provides empirical evidence on the relative importance of the SERVQUAL dimensions in the context of hospital care. *Thirdly*, it contributes to *practice* by providing concrete information that could guide the prioritization of interventions to improve patients' experience and lead to higher service quality ratings.

Our work also has a few *limitations*. The *first* issue is associated with our dataset. We note that demographic factors such as age, gender, race, social and education levels which have been reported to affect the degree of satisfaction of patients were not considered in our study. Other limitations of our study include the lack of precise information on the proportion of comments associated with in-patient to out-patient making it difficult to more strongly contextualise our findings. In addition, the detection of highly specific topics such as Dementia care, in particular, may indicate some bias in our dataset. While systematically administered surveys may be designed to avoid this type of bias, it is difficult to avoid bias in datasets comprising of public comments with little or no demographic information on contributors. The *second* is linked to the procedure for expert human labelling of topics. Only two experts (although with deep knowledge of the domain) were used for the labelling of latent topics. The reliability of the labelling can be increased with more expert labellers. Another limitation is directly linked to the nature of the CGT method which produces substantive (mid-range) theory. To move towards more formal theories, further validation and triangulation of results using a more traditional quantitative or qualitative approach in both general and specific hospital care settings is required. Despite these limitations, we believe that our approach is novel and illustrates how theory-building (in the information systems and health informatics) could be achieved using CGT for semi-automatically analysing a large collection of unstructured text.

References

AHRQ. 2017. "What Is Patient Experience?," *Rockville, MD*. (https://www.ahrq.gov/cahps/about-cahps/patient-experience/index.html).

Al-Damen, R. 2017. "Health Care Service Quality and Its Impact on Patient Satisfaction 'Case of Al-Bashir Hospital,'" *International Journal of Business and Management* (12:9), p. 136. (https://doi.org/10.5539/ijbm.v12n9p136).

Alkuwaiti, A., Maruthamuthu, T., and Akgun, S. 2020. "Factors Associated with the Quality of Outpatient Service: The Application of Factor Analysis – A Case Study," *International Journal of Healthcare Management* (13:sup1), Taylor & Francis, pp. 88–93. (https://doi.org/10.1080/20479700.2017.1422910).

Bahja, M., and Lycett, M. 2016. "Identifying Patient Experience from Online Resources via Sentiment Analysis and Topic Modelling," in *BDCAT'16*, pp. 94–99. (https://doi.org/10.1145/3006299.3006335).

Bari, V., Hirsch, J. S., Narvaez, J., Sardinia, R., Bock, K. R., Oppenheim, M. I., and Meytlis, M. 2020. "An *Approach* to Predicting Patient Experience through Machine Learning and Social Network Analysis," *Journal of the American Medical Informatics Association* (27:12), pp. 1834–1843. (https://doi.org/10.1093/jamia/ocaa194).

Berente, N., and Seidel, S. 2014. "Big Data & Inductive Theory Development: Towards Computational *Grounded* Theory?," in *20th Americas Conference on Information Systems, AMCIS 2014*, pp. 1–11.

Bhoomadevi, A., Ganesh, M., and Panchanatham, N. 2021. "Improving the Healthcare Using Perception of Health Professional and Patients: Need to Develop a Patients Centered Structural Equation Model," *International Journal of Healthcare Management*, pp. 42–49. (https://doi.org/10.1080/20479700.2019.1607014).

Blei, D., Carin, L., and Dunson, D. 2010. "Probabilistic Topic Models," *IEEE Signal Processing Magazine* (27:6), pp. 55–65. (https://doi.org/10.1109/MSP.2010.938079).

Breiman, L. 2001. "Random Forests," *Machine Learning* (45:1), pp. 5–32. (https://doi.org/10.1023/A:1010933404324).

Carcamo, C. R., and Lledo, R. 2001. "Predictors of Satisfaction with Surgical Treatment," *International Journal for Quality in Health Care* (13:3), pp. 267–269. (https://doi.org/10.1093/intqhc/13.3.267).

Chan, E. A., Wong, F., Cheung, M. Y., and Lam, W. 2018. "Patients' Perceptions of Their Experiences with Nurse-Patient Communication in Oncology Settings: A Focused Ethnographic Study," *PLoS ONE* (13:6), pp. 1–17. (https://doi.org/10.1371/journal.pone.0199183).

Chang, J. 2015. "Package 'Lda'. Collapsed Gibbs Sampling Methods for Topic Models," *R Package Version 1.4.2*. (https://doi.org/10.1109/TCOM.1984.1096090).

Cunningham, M., and Wells, M. 2017. "Qualitative Analysis of 6961 Free-Text Comments from the First National Cancer Patient Experience Survey in Scotland," *BMJ Open* (7:6), pp. 1–16. (https://doi.org/10.1136/bmjopen-2016-015726).

Ding, K., Chong, W., Yap, K., and Imm, S. 2020. "Employing Structural Topic Modelling to Explore Perceived Service Quality Attributes in Airbnb Accommodation," *International Journal of Hospitality Management* (91:August), Elsevier Ltd, p. 102676. (https://doi.org/10.1016/j.ijhm.2020.102676).

Fiala, T. G. 2012. "What Do Patients Want? Technical Quality versus Functional Quality: A Literature Review for Plastic Surgeons.," *Aesthetic Surgery Journal / the American Society for Aesthetic Plastic Surgery* (32:6), pp. 751–759. (https://doi.org/10.1177/1090820X12452555).

Hu, L., and Bentler, P. 1998. "Fit Indices in Covariance Structure Modeling : Sensitivity to Underparameterized Model Misspecification," *Psychological Methods* (3), pp. 424–453.

Hu, N., Zhang, T., Gao, B., and Bose, I. 2019. "What Do Hotel Customers Complain about? Text Analysis Using Structural Topic Model," *Tourism Management* (72:December 2018), Elsevier, pp. 417–426. (https://doi.org/10.1016/j.tourman.2019.01.002).

Jenkinson, C., Coulter, A., Bruster, S., Richards, N., and Chandola, T. 2002. "Patients' Experiences and Satisfaction with Health Care: Results of a Questionnaire Study of Specific Aspects of Care," *Quality and Safety in Health Care* (11:4), pp. 335–339. (https://doi.org/10.1136/qhc.11.4.335).

Kar, A. K., and Dwivedi, Y. K. 2020. "Theory Building with Big Data-Driven Research – Moving Away from the 'What' towards the 'Why,'" *International Journal of Information Management* (54:June), Elsevier, p. 102205. (https://doi.org/10.1016/j.ijinfomgt.2020.102205).

Karami, M., Maleki, M. M., and Dubinsky, A. J. 2016. "Cultural Values and Consumers' Expectations and Perceptions of Service Encounter Quality," *International Journal of Pharmaceutical and Healthcare*

Marketing (10:1), pp. 2–26. (https://doi.org/10.1108/IJPHM-09-2014-0051).
Korfiatis, N., Stamolampros, P., Kourouthanassis, P., and Sagiadinos, V. 2019. "Measuring Service Quality from Unstructured Data: A Topic Modeling Application on Airline Passengers' Online Reviews," *Expert Systems with Applications* (116:September), pp. 472–486. (https://doi.org/10.1016/j.eswa.2018.09.037).
Kozbagarov, O., Mussabayev, R., and Mladenovic, N. 2021. "A New Sentence-Based Interpretative Topic Modeling and Automatic Topic Labeling," *Symmetry 2021, Vol. 13, Page 837* (13:5), Multidisciplinary Digital Publishing Institute, p. 837. (https://doi.org/10.3390/SYM13050837).
Kumar, M., Kee, F. T., and Charles, V. 2010. "Comparative Evaluation of Critical Factors in Delivering Service Quality of Banks: An Application of Dominance Analysis in Modified SERVQUAL Model," *International Journal of Quality and Reliability Management* (27:3), pp. 352–378. (https://doi.org/10.1108/02656711011023320).
Lai, F., Hutchinson, J., Li, D., and Bai, C. 2007. "An Empirical Assessment and Application of SERVQUAL in Mainland China's Mobile Communications Industry," *International Journal of Quality and Reliability Management* (24:3), pp. 244–262. (https://doi.org/10.1108/02656710710730852).
Larson, E., Sharma, J., Bohren, M. A., and Tunçalp, Ö. 2019. "When the Patient Is the Expert: Measuring Patient Experience and Satisfaction with Care," *Bulletin of the World Health Organization* (97:8), pp. 563–569. (https://doi.org/10.2471/BLT.18.225201).
Lee, D. H. 2017. "HEALTHQUAL: A Multi-Item Scale for Assessing Healthcare Service Quality," *Service Business* (11:3), Springer Berlin Heidelberg, pp. 491–516. (https://doi.org/10.1007/s11628-016-0317-2).
Maramba, I., Davey, A., Elliott, M., Roberts, M., Roland, M., Brown, F., Burt, J., Boiko, O., and Campbell, J. 2015. "Web-Based Textual Analysis of Free-Text Patient Experience Comments from a Survey in Primary Care," *JMIR Med Inform* (3(2)). (https://doi.org/doi: 10.2196/medinform.3783).
Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., and Hamprecht, F. A. 2009. "A Comparison of Random Forest and Its Gini Importance with Standard Chemometric Methods for the Feature Selection and Classification of Spectral Data," *BMC Bioinformatics* (10:1), p. 213. (https://doi.org/10.1186/1471-2105-10-213).
Mourad, S. M., Nelen, W. L. D. M., Akkermans, R. P., Vollebergh, J. H. A., Grol, R. P. T. M., Hermens, R. P. M. G., and Kremer, J. A. M. 2010. "Determinants of Patients' Experiences and Satisfaction with Fertility Care," *Fertility and Sterility* (94:4), Elsevier Ltd, pp. 1254–1260. (https://doi.org/10.1016/j.fertnstert.2009.07.990).
Müller, O., Junglas, I., vom Brocke, J., and Debortoli, S. 2016. "Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines," *European Journal of Information Systems* (forthcomin:November 2014), Taylor & Francis Online, pp. 1–14. (https://doi.org/10.1057/ejis.2016.2).
Naidu, A. 2009. "Factors Affecting Patient Satisfaction and Healthcare Quality," *International Journal of Health Care Quality Assurance* (22:4), pp. 366–381. (https://doi.org/10.1108/09526860910964834).
Nelson, L. K. 2020. "Computational Grounded Theory: A Methodological Framework," *Sociological Methods and Research* (49:1), pp. 3–42. (https://doi.org/10.1177/0049124117729703).
Ojo, A., and Rizun, N. 2019. "Enabling Deeper Linguistic-Based Text Analytics - Construct Development for the Criticality of Negative Service Experience," *IEEE Access* (7), pp. 169217–169256. (https://doi.org/10.1109/ACCESS.2019.2947593).
Ojo, A., and Rizun, N. 2020. "Structural and Temporal Topic Models of Feedbacks on Service Quality-A Path to Theory Development?," in *AMCIS 2020 Proceedings. 15.*
Otalora, M. L., Rosenbaum, M. S., and Orejula, A. R. 2018. "Understanding Health Care Service Quality in Developing Latin America," *Health Marketing Quarterly* (35:3), Routledge, pp. 167–185. (https://doi.org/10.1080/07359683.2018.1514733).
Pakdil, F., and Harwood, T. N. 2005. "Patient Satisfaction in a Preoperative Assessment Clinic: An Analysis Using SERVQUAL Dimensions," *Total Quality Management and Business Excellence* (16:1), pp. 15–30. (https://doi.org/10.1080/1478336042000255622).
Parasuraman, A., Zeithaml, V., and Berry, L. L. 1988. "SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality," *Journal of Retailing* (64), pp. 12–40. (https://doi.org/10.1002/9781118785317.weom090654).
Pavlov, G., Maydeu-Olivares, A., and Shi, D. 2021. "Using the Standardized Root Mean Squared Residual (SRMR) to Assess Exact Fit in Structural Equation Models," *Educational and Psychological Measurement* (81:1), pp. 110–130. (https://doi.org/10.1177/0013164420926231).

Price, R. A., Elliott, M. N., Zaslavsky, A. M., Hays, R. D., Lehrman, W. G., Rybowski, L., Edgman-Levitan, S., and Cleary, P. D. 2014. "Examining the Role of Patient Experience Surveys in Measuring Health Care Quality," *Medical Care Research and Review* (71:5), pp. 522–554. (https://doi.org/10.1177/1077558714541480).

Ranard, B. L., Werner, R. M., Antanavicius, T., Andrew Schwartz, H., Smith, R. J., Meisel, Z. F., Asch, D. A., Ungar, L. H., and Merchant, R. M. 2016. "Yelp Reviews of Hospital Care Can Supplement and Inform Traditional Surveys of the Patient Experience of Care," *Health Affairs* (35:4), pp. 697–705. (https://doi.org/10.1377/hlthaff.2015.1030).

Rashid, J., Adnan Shah, S. M., Irtaza, A., Mahmood, T., Nisar, M. W., Shafiq, M., and Gardezi, A. 2019. "Topic Modeling Technique for Text Mining Over Biomedical Text Corpora Through Hybrid Inverse Documents Frequency and Fuzzy K-Means Clustering," *IEEE Access* (7), pp. 146070–146080. (https://doi.org/10.1109/ACCESS.2019.2944973).

Rathert, C., Brandt, J., and Williams, E. S. 2012. "Putting the 'patient' in Patient Safety: A Qualitative Study of Consumer Experiences," *Health Expectations* (15:3), pp. 327–336. (https://doi.org/10.1111/j.1369-7625.2011.00685.x).

Rehaman, B., and Husnain, M. 2018. "The Impact of Service Quality Dimensions on Patient Satisfaction in the Private Healthcare Industry in Pakistan," *Journal of Hospital & Medical Management* (04:03), pp. 1–8. (https://doi.org/10.4172/2471-9781.100048).

Rejito, J., Atthariq, A., and Abdullah, A. S. 2021. "Application of Text Mining Employing K-Means Algorithms for Clustering Tweets of Tokopedia," *Journal of Physics: Conference Series* (1722), {IOP} Publishing, p. 12019. (https://doi.org/10.1088/1742-6596/1722/1/012019).

Schmiedel, T., Müller, O., and vom Brocke, J. 2019. "Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial With an Application Example on Organizational Culture," *Organizational Research Methods* (22:4), pp. 941–968. (https://doi.org/10.1177/1094428118773858).

Schoenfelder, T., Klewer, J., and Kugler, J. 2011. "Determinants of Patient Satisfaction: A Study among 39 Hospitals in an in-Patient Setting in Germany," *International Journal for Quality in Health Care* (23:5), pp. 503–509. (https://doi.org/10.1093/intqhc/mzr038).

Shafei, I., Walburg, J. A., and Taher, A. F. 2015. "Healthcare Service Quality: What Really Matters to the Female Patient?," *International Journal of Pharmaceutical and Healthcare Marketing* (9:4), pp. 369–391. (https://doi.org/10.1108/IJPHM-05-2014-0028).

Shan, L., Li, Y., Ding, D., Wu, Q., Liu, C., Jiao, M., Hao, Y., Han, Y., Gao, L., Hao, J., Wang, L., Xu, W., and Ren, J. 2016. "Patient Satisfaction with Hospital Inpatient Care: Effects of Trust, Medical Insurance and Perceived Quality of Care," *PLoS ONE* (11:10), pp. 1–18. (https://doi.org/10.1371/journal.pone.0164366).

Valencia-Arias, A., Piedrahita, L. B., Zapata, A. B., Benjumea, M., and Moya, L. P. 2018. "Mapping the Healthcare Service Quality Domain: A Bibliometric Analysis," *Journal of Clinical and Diagnostic Research* (12:8), pp. IC01–IC05. (https://doi.org/10.7860/JCDR/2018/30361.11863).

Wagland, R., Recio-Saucedo, A., Simon, M., Bracher, M., Hunt, K., Foster, C., Downing, A., Glaser, A., and Corner, J. 2016. "Development and Testing of a Text-Mining Approach to Analyse Patients' Comments on Their Experiences of Colorectal Cancer Care," *BMJ Quality and Safety* (25:8), pp. 604–614. (https://doi.org/10.1136/bmjqs-2015-004063).

Wiseman, T., Lucas, G., Sangha, A., Randolph, A., Stapleton, S., Pattison, N., O'Gara, G., Harris, K., Pritchard-Jones, K., and Dolan, S. 2015. "Insights into the Experiences of Patients with Cancer in London: Framework Analysis of Free-Text Data from the National Cancer Patient Experience Survey 2012/2013 from the Two London Integrated Cancer Systems," *BMJ Open* (5:10), pp. 1–10. (https://doi.org/10.1136/bmjopen-2015-007792).

Xia, Y., and Yang, Y. 2019. "RMSEA, CFI, and TLI in Structural Equation Modeling with Ordered Categorical Data: The Story They Tell Depends on the Estimation Methods," *Behavior Research Methods* (51:1), pp. 409–428. (https://doi.org/10.3758/s13428-018-1055-2).

Zhang, Z. 2017. "Structural Equation Modeling in the Context of Clinical Research," *Ann Transl Med* (5:102), pp. 539–547. (https://doi.org/http://dx.doi.org/10.21037/atm.2016.09.25).