Quality Measurement of Consumer Health Questions: Content and Language Perspectives

Ashwag Alasmari, Lina Zhou

Submitted to: Journal of Medical Internet Research
on: April 17, 2023

Table of Contents

Quality Measurement of Consumer Health Questions: Content and Language Perspectives

Ashwag Alasmari[1] PhD; Lina Zhou[2] PhD

[1]National Institutes of Health (NIH) Bethesda US
[2]University of Maryland, Baltimore County Baltimore US

Corresponding Author:
Ashwag Alasmari PhD
National Institutes of Health (NIH)
8600 Rockville Pike
Bethesda
US

Abstract

Background: Health information consumers find question and answer (Q&A) communities as an alternative way of meeting their information needs. However, the quality of their questions can directly impact what answers they receive from the community and whether they receive an answer at all.

Objective: This study aims to improve our understanding of quality of questions from health information consumers in online Q&A communities.

Methods: To this end, we develop a research construct of question quality and its operationalization for the first time, then empirically identify the content and language determinants of question quality. Moreover, we propose a novel approach to measuring the overall quality of questions by adapting the k-means clustering algorithm and automated measures of indicators of question quality. To validate our proposed constructs and determinants of question quality, we collected and analyzed questions for kidney disease from both expert-curated and community-based Q&A communities. The research construct was validated by manual assessment of health experts, and the determinants were tested using regression analysis.

Results: High-quality questions are more likely to include demographic and medical information than lower-level quality questions. At the same time, asking questions at the various stages of disease development was more likely to reflect high quality than lower-level quality questions. Low quality questions were generally shorter with lengthier sentences than high quality questions.

Conclusions: The findings of this research can guide health consumers in better formulating their health information questions and support online Q&A communities in managing health questions effectively.

(JMIR Preprints 17/04/2023:48257)
DOI: https://doi.org/10.2196/preprints.48257

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
   Please make my preprint PDF available to anyone at any time (recommended).
   Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
   ✓ **Only make the preprint title and abstract visible.**
   No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
   ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
   Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
   Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Quality Measurement of Consumer Health Questions: Content and Language Perspectives

Abstract

Background: Health information consumers find question and answer (Q&A) communities as an alternative way of meeting their information needs. However, the quality of their questions can directly impact what answers they receive from the community and whether they receive an answer at all.

Objective: This study aims to improve our understanding of quality of questions from health information consumers in online Q&A communities.

Methods: We develop a novel framework for defining and measuring question quality within online health communities, incorporating both content-based and language-based variables. This framework leverages k-means clustering and established automated metrics to assess overall question quality. To validate our framework, we analyze questions related to kidney disease from expert-curated and community-based Q&A platforms. Expert evaluations confirm the validity of our quality construct, while regression analysis helps identify key variables.

Results: High-quality questions are more likely to include demographic and medical information than lower-level quality questions ($p < .001$). At the same time, asking questions at the various stages of disease development was less likely to reflect high quality ($p < .001$). Low quality questions were generally shorter with lengthier sentences than high quality questions ($p < .01$).

Conclusions: This research empowers consumers to formulate more effective health information questions, ultimately leading to better engagement and more valuable insights within online Q&A communities. Additionally, our findings provide valuable insights for platform developers and moderators seeking to enhance the quality of user interactions and foster a more trustworthy and informative environment for health information exchange.

Keywords: Question quality; Quality measurement; Health questions; Q&A platforms; Information needs; Information behavior; information sharing.

Introduction

Health Information Consumers (HIC) are increasingly involved in their healthcare decisions and feel empowered to leverage a myriad of online resources to fulfill their information needs. Online question and answer (Q&A) communities are one of these valuable resources, which allow HIC to post questions to seek answers from other members [1]. Online Q&A communities serve as a direct and alternative means of acquiring information compared to search engines, especially when the latter fails to fulfill the user needs [2].

Articulating health information needs in a concise and understandable question is essential because the relevancy, quality, and nature of the information acquired are substantially linked to the nature and quality of the question representing the information needs [3,4]. Questions serve as the starting point in the Q&A setting and the primary driver of what might happen next - who responds, and how good and relevant the provided answers are. However, users may experience significant uncertainty in conceiving information needs and subsequent difficulty in articulating these needs [5].

Writing high-quality questions can bring many potential benefits. Well-formed questions attract more high-quality answers than poorly formed ones, as subject-matter experts are more likely to assist users who already put in some effort [6]. The quality of the question itself can significantly impact the probability of receiving helpful answers [7], which eventually drives the popularity of a Q&A community. Previous studies [8–12] show that the features of the questions and the responsiveness to these questions are correlated. More specifically, there are language determinants of the quantity and quality of responses in online communities and social networks [11]. For instance, stating the information needs in a question format, explicitly scoping the audience, and using only one sentence lead to more, better, and faster responses [11]. Other studies indicate that a good question should contain details and examples [12]. Thus, it is important for HIC to formulate their information needs in high-quality health questions to receive timely, relevant, and comprehensive answers in online Q&A communities.

There are three streams of research that are related to question quality. The first one is focused on the characteristics of answers with the assumption that high-quality questions generate answers, and the most straightforward measure of a good quality question is whether it has received an acceptable response [13]. In contrast, a poor-quality question likely fails to receive any answers. Some studies have further examined the quality of answers in Q&A communities [14]. This stream of research relies on the outcomes of Q&A interaction, which may not be available. Even if they are available, a lag is expected between the times when an HIC posts a question and when an answer is provided in response. The second stream of research examines the characteristics of the askers, such as reputation or expertise [15], on the basis that community recognition reflects the ability to construct appropriate and valuable questions. However, the asker's information is generally not publicly available due to the platform's privacy policy and/or the HIC's privacy concerns. The third stream is on analyzing question characteristics, such as the topics, number of views, and content and language variables that correlate with high-quality answers [16]. Examining question quality based on the question's intrinsic features holds promise for overcoming the limitations of the first two research streams. However, few studies have focused on question quality and its measurements.

This study seeks to enhance our understanding of the intrinsic variables influencing question quality posed in online Q&A communities from the content and language perspectives. Specifically, we address the following primary research question: What are the content and language variables of high-quality questions in online health Q&A communities? Answering the question can not only assist HIC in soliciting answers from peer community members and actively engaging themselves in community discussions but also promote the success of online Q&A communities.

Our Proposed Constructs for Health Question Quality

In this section, we first introduce our proposed construct of the health question quality. We then develop quantitative measures for the construct using a clustering approach, and finally validate the measurements through human assessment. Figure 1 illustrates the architecture of health question quality measurement, including data collection, feature extraction, clustering and question quality validation.

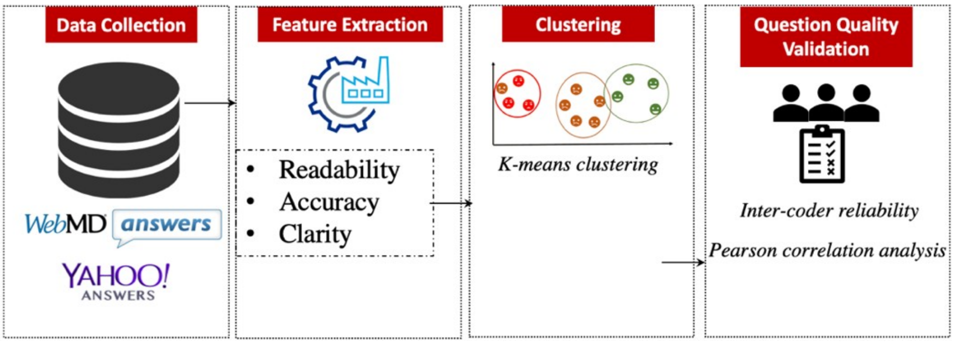

Figure 1. Architecture of Health Question Quality Assessment

Construct Design

We introduce a conceptualization of question quality that centers on two keys indicators: readability and clarity. We chose these two indicators because they represent the variety of measures that impact how individuals perceive questions Quality questions should be easy to read to make them accessible to a large and diverse audience, convey the correct information without any misspellings, and be clear about what is expected from an answer. To the best of our knowledge, these indicators have not yet been explored in the context of health-related questions, let alone their collective impacts. These proposed indicators will enable us to uncover fundamental aspects of question quality that would otherwise be missed through solely examining answers. Therefore, our proposed construct will facilitate the timely assessment of question quality prior to receiving any answer.

Readability relates to how well a question is written, how understandable it is, and the extent to which it is free of unnecessary complexity. The concept of readability is often studied in natural language processing and used as an indicator of article content quality in the digital library [18]. For online Q&A communities, we assume that most community members communicate at an average reading level [17]. Accordingly, questions with a high reading level tend to attract fewer potential answers because fewer community members would comprehend the question [17].

Clarity assesses how easily understood and can be readily answered by other HICs. This translates to a higher likelihood of receiving high-quality responses.

Construct Measurement

We propose measures for each of the two indicators of health question quality construct.

Readability

Automated measures have been abundant for readability, such as the Flesch-Kincaid Grade Level, Gunning Fog Index, SMOG Index, Coleman-Liau Index, and the Automated Readability Index [25]. Among them, the Flesch-Kincaid Reading Ease metric (see Equation (1)) is the most widely used [26] and has been adopted in question-answering in various significant contexts [8]. Similarly, we have incorporated it into our study.

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right) \quad (1)$$

The Flesch-Kincaid readability score uses the average length of sentences and the average number of syllables per word to calculate the reading ease. The output is on a scale, which typically ranges from 1 to 100, with lower scores indicating a more difficult text to read. As laypersons typically compose health questions without a medical background, these questions are non-scientific and share similarities with informal discourses encountered in a high-school education setting. Hence, the

Flesch-Kincaid Reading Ease method is applicable in this context.

Clarity

Questions that convey a clear message are easily understood and can be responded to by other HIC with greater ease, Moreover, questions with clarity increase their chance of receiving high quality answers. For a question to be considered explicit, it should include a minimum of one interrogative word [8]. Interrogative words (e.g., who, what, where, when, why, how) can indicate the clarity of health questions. Following the work of Kitzie, Choi and Shah [17], we devised a scale of clarity based on the number of interrogative words (e.g., who, what, where, when, why, how, should) normalized by the question length. Specifically, a higher ratio of interrogative words to the total number of words suggests a clearer formulation of the question. We illustrate the measure of clarity with two sample questions as the following:

- *Q1: "What's the first thing doctors do when a child is diagnosed with chronic kidney disease? And what happens next. Like when you first go in what do they do? What do they tell you? what happens?"*
- *Q2: "I have been early diagnosed with stage 3 kidney failure. Let me know thoughts on treatments? The condition stems from many years of high blood pressure. I am 55 and hope this doesn't lead to dialysis"*.

Question Q1 contains multiple occurrences of interrogative words, such as what and when, accounting for 19.44 percent words in the question. They clearly indicate the user's informational needs. In contrast, question Q2 provides more background information about the health information consumer's condition but does not explicitly state his/her question with any interrogative word, which results in a ratio of zero for clarity. Thus, the information needs expressed in Q2 are not so clear as those in Q1. We used linguistic inquiry and word count (LIWC) [27] to extract interrogative words.

Overall Question Measurement

One straightforward method to derive an overall quality measure is by averaging the measures of individual indicators, namely readability and clarity. However, assigning equal weight to all indicators may overlook the nuanced differences among individual constructs within questions. In this study, we propose a clustering analysis approach complemented with a human assessment to validate the measurement . The study employed the k-means clustering technique [28] to group health questions to a small number of distinct clusters based on question similarities. The parameter *k* was empirically determined based on the number of clusters sets and the proportion of betweenness. We selected the elbow method to determine the number of clusters, which runs the k-means clustering algorithm on the dataset for a range of values and then calculate the sum of squared error for each value [28]. Accordingly, we set k to 4. To validate the clustering results, we used the method of human validation.

Human Validation

We recruited three human judges to assess the quality of each question manually. All judges were experienced with searching for health information online and were familiar with seeking and sharing information in online Q&A communities. Their primary language was English, and they either developed competence in disease management or had access to it through a family member. We opted not to include human judges with medical backgrounds because online Q&A communities tend to attract individuals without medical expertise, as observed in prior studies [29–31].

We performed stratified sampling by randomly selecting 10% of the questions from each quality cluster. The order of the questions was randomized for each of the human judges. The judges were asked to rate each question independently at one of the four quality levels: high, average, and low.

The final quality rating of the health questions was determined based on a joint discussion of the three judges. Inter-coder reliability was assessed with Cohen's kappa statistics, following the convention of $\kappa > 0.70$ indicating 'substantial' agreement [32]. In addition, we also performed a Pearson correlation analysis between ratings of the human judges and results of k-means clustering using the following coding scheme: very low=1; low=2; average=3, and high=4. The results show that r>0.5, which indicates a very high correlation.

Methods

To answer the research question about content and language variables of the quality of health-related questions in online Q&A communities, we first describe the dataset and then introduce content-based and language-based variables and analysis methods.

Data Collection

For this study, we collected health questions from Yahoo! Answers, a community-based Q&A site, and WebMD, an expert-based Q&A site. WebMD is one of the most influential online health sites, and users were able to post questions for certified health experts to answer in the Q&A section, covering more than 900 health topics. We collected all posts made from August 2008 until the closure of WebMD Answers in 2018. Yahoo! Answers features health as one of the top-level categories. Since its release in December 2005, Yahoo! Answers has become a popular Internet reference site worldwide and the most frequented Q&A community in the U.S. As of June 2019, the site ranked ninth in global internet traffic and engagement over the past three months and seventh in the United States.

Following the prior work [33–36], we treated postings in a question section of Q&A site a question regardless of whether they ended with a question mark or not. We selected questions related to kidney diseases for our case study. For example, chronic kidney disease is a non-communicable health disease, often accompanied by multimorbidity (frequent co-occurrences with other diseases) [37]. According to the National Kidney Foundation, chronic kidney disease (CKD) affects 14% of adults in the United States [38]. Given its prevalence and complexity of disease management, kidney disease served as a valuable test case for our investigation into question quality in online health communities. In these platforms, patients navigate and manage complex health information, making this study particularly relevant. We screened for queries based on pre-selected key terms that directly refer to kidney conditions or early signs of chronic kidney disease [39]. Using an Application Programming Interface, we sampled 400 random questions from Yahoo! Answers and another 400 questions from WebMD relating to the above kidney-related vital terms. In other words, the questions included in our dataset must be related to human kidney diseases. We eliminated repeated questions using a combination of user identification and question similarity techniques. In addition, we manually filtered irrelevant questions, including advertisements, those unrelated to kidney-related topics, not involving human patients or associated with student research. For each of the relevant questions, we extracted its title, descriptions, date of posting, topics where the questions were posted, and number of answers.

Content-based Variables

We contextualized the health questions by drawing on relevant findings from a previous study on health information seeking [33]. The two main areas of interest in our analyses included health stages of disease development and information shared (i.e., demographic and medical information).

Two coders, with medical background, manually coded the stage of disease development and type of information shared. To ensure coding consistency, a well-defined coding guideline was followed. Additionally, any discrepancies between the coders' initial assessments were resolved through adjudication by a third coder to minimize subjective bias and ensure the accuracy of the coded data. Variables were coded as binary variables were (1 indicating presence, 0 indicating absence) of a specific information.

- **Health Stages of Disease Developments.** We followed two systems in representing the stages of disease development: stages of health questioning [33] and chronic disease stages [40]
    - **Stages of Health Questioning.** Managing a disease or condition is an ongoing process, and HIC at different stages often has different levels of information needs [41]. Also, HIC may display other information-seeking behaviors based on the nature and extent of their needs. To identify at which stage of their disease HIC asked questions (stage of health questioning), we used a model proposed by Zhang [33] consisting of eight stages: 1) being healthy – at this stage, questions related to disease prevention and health promotion; 2) self-diagnosed as being ill; 3) before having a medical test or checkup; 4) after being diagnosed or self-diagnosed as ill; 5) before treatment (such as surgery or medications); 6) during treatment (including medications or exercise); 7) after treatment, and 8) when the disease becomes chronic or reaches the terminal stage. To address data sparsity, we merged stages (2~4) as stage 2 and stages (5~7) as stage 3.
    - **Chronic Disease Stages.** We also drew on the stages of chronic illness to understand the questions of HIC who are chronically ill. To this end, we used Corbin's Chronic Illness Trajectory Framework because it includes all stages of chronic disease [40] and is used by clinicians in nursing care and in chronic illness management [42]. This trajectory framework is built around the idea that chronic conditions vary and change over time, which consists of nine stages [40]: 1) pre-trajectory: before the disease onset; 2) trajectory onset: the appearance of symptoms and diagnosis; 3) stable: condition and symptoms are under control, everyday life is unaffected, illness management is home-centered, and hospitalization not required; 4) unstable: condition and symptoms are not under control, everyday life is disrupted, but care remains centered in the home; 5) acute: symptoms or complications require hospitalization or other measures, everyday life activities are cut back or severely curtailed; 6) crisis: a life-threatening situation that requires emergency care, everyday life is placed on hold; 7) comeback: a return to everyday life activities, possibly with changed ability for everyday life activities; 8) downward: decline associated with increased disability and trouble controlling symptoms, requires adaptation in everyday life activities; and 9) dying: death of the patient. To address data sparsity, we grouped the nine stages into two: stable and non-stable stages, with the latter covering stages 4 through 9 and former covering the rest.
- **Type of Information Shared.** When HICs communicate their information needs by asking questions, they often include some demographic and medical information related to diagnosis, treatment, and prevention that represent their understanding of their disease [33]. Demographic information includes age, gender, ethnicity, weight, location, and profession. Medical information may pertain to diagnosis, treatment, or prevention. Diagnostic medical information includes symptoms, medical tests, and personal and family medical history, etc. Treatment and prevention medical information include treatment options, duration of treatment, lifestyle, prescribed medications, and length of hospitalization, etc.

Language-based Variables

We leveraged Linguistic Inquiry and Word Count (LIWC) [27] to extract language-based variables due to the tool's ability to measure language in multiple dimensions. The value of each dimension is calculated based on the percentage of words related to this particular dimension [27]. For this research, we selected the following variables that can be used to characterize the language and writing style of a health question:

- **The total number of words and the number of words per phrase.**
- **Total number of pronouns.** These include personal pronouns and impersonal pronouns.
- **Social processes.** This includes references to family or friends, as well as biological sex (male and female).
- **Time orientation.** This includes references to the past, present, and future, which reflect a general time orientation.
- **Biological process.** This includes references to the body and health.
- **Affect.** This includes five-word variables: overall affect, positive and negative emotion, anxiety, anger, and sadness.

Analysis Methods

A multinomial logistic regression model was used to analyze the content and language variables determining the health question quality. The dependent variable was the overall question quality, and the independent variables were content-based and language-based variables.

Results

Descriptive Statistics

Table 2 reports the descriptive statistics and the statistical test for the content-based variables. The table shows that the proportion of medical information included in WebMD questions is significantly higher ($p<.01$) than that in Yahoo! Answers questions. More specifically, medical treatment and medical diagnosis information in WebMD are significantly higher ($p<.05$) than that in Yahoo! Answers. In contrast, demographic information is higher in Yahoo! Answers ($p<.01$) than in WebMD. The comparison results of the stages of general health questioning show that the proportions of all stages except for stage 1 are significantly higher in WebMD ($p<.01$) as compared to Yahoo! Answers. In contrast, the proportions of questions regarding the three chronic disease stages are higher in Yahoo! Answers ($p<.01$) than WebMD.
Table 3 reports the descriptive statistics and the statistical test for language-based variables. It shows that the average word counts per question, the total number of pronouns, personal pronouns, and affect in WebMD is significantly lower ($p<.01$) than Yahoo! Answers. In contrast, social process is lower in questions from Yahoo! Answers ($p<.01$) than those from WebMD. However, no statistically significant differences were detected in word per sentence and time orientation between the two platforms ($p=.964$).

Table 2. *Descriptive statistics for content-based variables*

		Yahoo! Answers		**WebMD**		***p*-value**
		Frequency	Percent	Frequency	Percent	

Types of Information Shared	**Medical Information**	147	46.52	182	59.09	0.000^{***}
	Demographic Information	100	31.65	58	18.83	0.000^{***}
	Medical Diagnosis Information	111	35.13	154	50	0.000^{***}
	Medical Treatment Information	78	24.68	96	31.17	0.035^{*}
Stages of Health Questioning	1	12	3.8	11	3.57	0.88
	2	55	17.41	87	28.25	0.000^{**}
	3	4	1.27	36	11.69	0.000^{**}
	4	191	60.44	64	20.78	0.000^{**}
	5	54	17.09	110	35.71	0.000^{**}
Chronic Disease Stage	1	119	37.66	40	12.99	0.000^{**}
	2	50	15.82	18	5.84	0.000^{**}
	3	147	46.52	250	81.17	0.000^{**}

Notes: $^{*}p<0.05$; $^{**}p<0.01$

Table 3. *Descriptive statistics for language-based variables*

	WebMD		Yahoo! Answers		***p*-value**
	Mean	SD	Mean	SD	
Word Count	42.1	36.9	60.2	59.4	0.000^{**}
Word/Sentence	13.1	7.9	13.1	7.1	0.964
Total Pronouns	12.1	7.1	14.3	7.2	0.000^{**}
Personal Pronouns	6.6	6.3	7.9	5.6	0.003^{**}
Affect	1.0	2.5	1.5	2.2	0.014^{*}
Social Process	3.3	5.1	6.6	6.0	0.000^{**}
Time Orientation	3.0	3.8	2.6	3.1	0.086

Notes: $^{*}p<0.05$; $^{**}p<0.01$

Based on the results of k-means clustering, the coders manually analyzed the indicators of the question quality within each of the four clusters and provided a quality rating for each cluster separately, ranging from low to high quality. We discussed the method for determining the number of clusters in the second section (Our Proposed Constructs for Health Question Quality).

- Cluster 1 (High quality) comprised 101 questions with readability ranging between 39.01% and 80.5%, and clarity 33.33% or above.
- Cluster 2 (Medium quality) had 169 questions with readability ranging between 0.7 and 49.3%, and clarity 12.5% or lower.
- Cluster 3 (Very low quality) contained 354 questions with readability ranging between 1.5% and 14.7% and clarity being 9.09% or lower. While the clarity range of the current cluster is somehow similar to that of cluster 2, the latter is rated lower than the former on range of readability.

Table 4. *Descriptive Statistics of Quality Indicators in the Question Clusters of Different Quality Levels*

Cluster	Indicators	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
1 (high quality)	**Readability**	9.88	8.95	39.01	80.5
	Clarity	2.28	2.35	9.09	33.33
2 (Medium quality)	**Readability**	9.60	6.90	0.7	49.3
	Clarity	1.85	2.72	0.0	12.5
3 (Low quality)	**Readability**	6.67	2.62	1.5	14.7
	Clarity	15.35	5.40	0	9.09

The majority of questions fell into the low and very low-quality clusters, highlighting the critical need to improve the quality of online health questions, a pursuit central to this research.

Regression Analysis Results

We chose high-quality questions as the reference cluster because it is the expectations of all health questions. Accordingly, the regression coefficients indicate which independent variables significantly discriminate high-quality questions from low, and medium quality questions, respectively.

Tables 5 and 6 provide the coefficient estimates for all outcome comparisons, along with the standard error for each independent variables. We found statistically significant differences between the reference quality cluster and the other clusters in some of the stages of health questioning, most types of shared information, and the language variables of the health questions. However, there is no evidence that expressions of anxiety differ among the different quality clusters of the health questions.

Three of the stages of questioning (before/after diagnosis, when chronically ill, and when unstable condition) were strong predictors of high quality among all comparisons with the reference cluster (high-quality questions) (p<.001) (Table 5). This reached statistical significance only when comparing high quality to low-quality questions for unstable condition (b= 0.87, SD = 0.34, p<.005). Thus, questions relating to diagnosis, chronically or unstable condition are less likely to be of high quality. Questions relating to preventative reasons, treatment and stable stages of chronic disease did not significantly affect the prediction of question quality (p>0.1).

Demographic information and diagnostic medical information are significant predictors across the two comparisons (p<.001). High quality questions are more likely to include more demographic information than average quality (b= -0.84, SD = 0.21, p<.001), and low quality (b= -3.39, SD = 1.02, p<.001). In addition, high quality questions are more likely to include more diagnostic medical information than low quality (b= -2.96, SD = 0.73, p<.001), but least comparing to average quality (b= 1.08, SD = 0.19, p<.001). Finally, high quality questions are more likely to include more treatment and prevention information than low quality questions (b= -3.20, SD = 1.02, p<.001).

Table 5. *Multinomial logistic regression analysis of the effect of content variables on question quality*

		Medium versus High	***Low versus High***
	Explanatory variable	***Coefficient (SD)***	***Coefficients (SD)***
Stages of Health Questioning	*Preventative reasons*	0.82 (0.51)	-1.24 (0.83)
	Before/after diagnosis	2.83*** (0.78)	0.18 (1.33)
	Before/after/during treatment	-0.64 (0.50)	-0.80 (0.69)
Chronic Disease Stages	*When chronic*	0.54 (0.53)	2.08*** (0.68)
	Stable	-0.41 (0.40)	0.26 (0.52)
	Unstable	1.16*** (0.24)	0.87** (0.34)
Demographic Information	*Demographic information*	-0.84*** (0.21)	-3.39*** (1.02)
Medical Information	Diagnostic information	1.08*** (0.19)	-2.96*** (0.73)
	Treatment & Prevention	0.34* (0.20)	-3.20*** (1.02)

Notes: *$^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$, High quality (Cluster 1) serving as a reference cluster*

Regarding language variables of health questions (Table 6), we found that the number of words ($p<.001$) and the number of words per sentence ($p<.001$) tend to differ significantly between the two average and high-quality questions. Longer questions are more likely to be classified as high-quality (b= -0.01, SD = 0.003, $p<.001$) than low quality questions. However, high-quality questions are less likely to be associated with a higher word count per sentence than average quality questions (b= 0.04, SD = 0.01, $p<.001$). That means average quality questions tend to be overall shorter, but with more words per sentence than high quality questions.
Regarding affect and measures of emotions, the statistically significant difference was found between low/average and high-quality questions. High quality questions were more likely to be affective comparing to low quality (b= -2.12, SD = 0.03, $p<.001$), and yet less likely to include positive (b= 2.05, SD = 0.04, $p<.001$) or negative (b= 2.08, SD = 0.03, $p<.001$) emotions than low quality questions (Table 6). Expression of anger were also associated with high quality questions (b= -0.07, SD = 0.00, $p<.001$). In addition, expression of sadness was associated with high quality questions (b=-0.25, SD=0.06, $p<.001$). However, there is no evidence that expressions of anxiety in health questions differ among the different quality clusters.
Questions that focused on the past were more likely to be high quality compared to low quality (b= -0.30, SD = 0.14, $p<.001$). In comparison to low-quality questions, high-quality questions were more likely to focus on the future (b= -0.16, SD = 0.06, $p<.001$). Questions focusing on the present were equally likely to be of any quality.
In addition, we found that among the social processes, only mentions of friends (b= -1.74, SD = 0.00, $p<.001$) or male references (b= -5.85, SD = 0.00, $p<.001$) tend to be strong predictors of high-quality questions compared to low-quality ones. No other differences were detected between any of the other social processes (including family and female references) and other pairs of comparisons ($p>.05$).

Table 6. *Multinomial logistic regression analysis of the effect of language variables on question quality*

	Medium vs. High	**Very low vs. High**
Explanatory variable	**Coefficient estimate (SD)**	**Coefficient estimate (SD)**
Word count	-0.01*** (0.003)	-0.14*** (0.03)
Word/sentence	0.04*** (0.01)	0.22** (0.08)
Total pronouns	-4.30*** (0.01)	7.54*** (0.01)
Personal pronouns	4.254*** (0.01)	-7.53*** (0.02)
Impersonal pronouns	4.160*** (0.02)	-7.36*** (0.02)
Social processes	-0.11*** (0.03)	-0.07** (0.03)
Family	0.02 (0.09)	0.00 (0.28)
Friend	-0.29 (0.22)	-1.74*** (0.00)
Female references	0.05 (0.06)	-0.04 (0.25)
Male references	0.11** (0.053)	-5.85*** (0.00)
Past focus	0.11*** (0.03)	-0.30** (0.14)
Present focus	0.01 (0.02)	0.02 (0.03)
Future focus	-0.16** (0.06)	-0.06 (0.09)
Body	0.04* (0.02)	0.024 (0.03)
Health	-0.11*** (0.02)	-0.06*** (0.02)
Affect	0.37 (0.26)	-2.12*** (0.03)
Positive emotion	-0.37 (0.26)	2.05*** (0.04)

Negative emotion	-0.33 (0.26)	2.08*** (0.03)
Anxiety	-0.07 *(*0.00*)*	0.12 *(*0.11*)*
Anger	0.09 *(*0.15*)*	-0.71*** *(*0.00*)*
Sadness	-0.25*** *(*0.06*)*	-0.12* *(*0.07*)*

*Notes: *p<.1; **p<.05; ***p<.01, High quality (Cluster 1) serving as a reference cluster*

Discussion

Principal Results

Formulating high-quality questions can potentially bring many benefits, directly impacting the relevancy, quality, and nature of the information acquired [4]. Advancing HICs' understanding of question content will contribute significantly to identifying quality questions and facilitate HIC in formulating their questions to solicit better answers. This study highlights several content and language variables of high-quality questions. For example, asking questions at various stages of disease development is more likely to be associated with lower quality questions. On the other hand, high-quality questions are more likely to include demographic and medical information than lower-level quality questions. These results are consistent with a previous study that reported that the chances of an inquiry being answered are higher when background and personal information about the health situation are provided [8] because this motivates potential respondents to provide more detailed answers. Health experts will provide relevant and accurate information when the HIC include more contextual information about them, such as age, gender, and other background information about the health condition in question.

While high-quality questions conveyed more information using shorter sentences, low-quality questions were shorter overall but contained lengthier sentences. This suggests that clear and concise language expressing specific information needs is crucial for perceived question quality. Complex questions, often with long sentences, may require excessive effort from potential respondents, hindering their willingness to answer [17]. This aligns with our observation that individuals with more chronic conditions (potentially having more experience formulating clear questions) tend to ask higher-quality questions with shorter sentences.

The expression of positive or negative emotion was more likely to be found in low than high-quality questions, except for anger and sadness which was more likely to be seen in high-quality questions. One possible explanation is that HIC were seeking emotional support in low quality questions and hence disclosed their emotions more readily [43] than those who posted high quality questions and were seeking informational help.

Our analysis revealed significant discrepancies in the descriptive statistics between the WebMD Answers and Yahoo! Answers datasets. These differences likely stem from several factors. The platforms themselves cater to distinct audiences. WebMD Answers focused on health information, attracting users with specific medical inquiries and potentially greater health knowledge. Conversely, Yahoo! Answers served a broader audience with diverse interests, resulting in a wider range of question complexity and potentially lower health expertise. Despite these disparities, we believe merging the datasets offers valuable insights. The combined dataset encompasses a broader range of health-related questions and user experiences. This allows us to explore a more comprehensive picture of online health information seeking behavior, potentially revealing patterns not evident within a single platform.

Research Contribution

One of the main contributions of this study lies in quantifying health question quality. Although other quality measures already exist, such as receiving satisfactory answers and subjective judgment of quality[9], improving the interaction between users on health-related Q&A communities is of paramount importance. Any further insight into improving the presentation and expression of HIC needs is welcomed.

Unlike previous studies that focus on either the answers or the askers' profiles to measure the question quality in online Q&A communities, this study defined question quality as objective measures of the questions themselves based on their readability and clarity. More specifically, this research explores the idea that questions may share similar quality features. Rather than scoring each question on predefined measures, it utilized k-means clustering to facilitate the analysis and divided question quality into four clusters. Based on the means of each question quality measure, we identified high, average, very low, and low-quality clusters. The results of the clustering process, and the subsequent validation with human assessment support the idea that grouping questions is a viable analysis method because of shared characteristics.

One final contribution of this study is that the proposed framework of measuring question quality is independent of the application context. All two measures, readability and clarity can be applied or extended to other types of online Q&A communities beyond healthcare.

Research Implications

This study's results align with the theory that question quality can be seen from the lens of UGC rather than general IQ. When asking a question, the askers are users rather than consumers, and the question serves their purpose, rather than the purpose of potential respondents to receive further valuable information. In addition, the results confirm the theory that quality is an inherent characteristic of a question, irrespective of the content or desired outcome, which is for a HIC to receive an acceptable answer.

This new question of quality measurement has significant research implications. Firstly, compared to alternative quality measures (e.g., receiving satisfactory answers or based on the asker's profiles [8,13]), the textual features of questions can provide a timely assessment of quality based on characteristics of the questions themselves without relying on answers. This is particularly important for researchers aiming to build models that automatically predict or evaluate the quality of large volumes of health questions in online Q&A communities. Second, this study establishes a foundation for developing a set of objective quality criteria that can be utilized to create classifiers for identifying high-quality research questions in the future. Such work will require the identification of those content and language variables most relevant to the specific health subject.

This study also offers practical suggestions on improving the effectiveness of online Q&A communities. The findings of this study may guide designers and developers of Q&A systems to design community support systems that encourage user contributions and control quality. Based on the measurement methods of question quality introduced in this work, Q&A communities could develop automated systems for prescreening the quality of health questions before HIC submit them to the community and facilitate the asker in formulating high-quality questions. Such approaches could include employing an iterative feedback system, query expansion, or syntax checking correction. Quality questions are likely to lead to quality answers, critical to user engagement in those communities.

Employing iterative feedback systems, such as query reformulation in interactive information retrieval [20], has been shown to assist users in articulating their information needs, depending on what information they initially deem relevant and how the system processes this information to provide better answers. Within the Q&A context, archived answers to similar questions could be presented to the asker, and the asker could pick the questions that are most relevant to their question to check out their answers. The asker may not only learn from others how to write queries, but also find answers to their questions directly. In addition, by understanding the textual features of high-quality questions, consumers can tailor their inquiries to elicit optimal responses. Our research provides consumers with concrete steps to improve their questions. Highlighting details like demographics, medical history, and symptoms in a clear and concise manner can lead to more targeted and accurate answers. This empowers consumers to make informed healthcare decisions based on reliable information.

One long-term implication of this study lies in helping healthcare professionals or artificial intelligence systems improve their performance by adding a layer of quality assurance before the question is processed. Already, AI is finding its uses in healthcare such as diagnosis, personalized care, and treatment [44], and it relies on quality data to perform well. In addition, healthcare providers will also benefit from receiving high-quality questions that contain all relevant information at the outset when providing online health advice or telemedicine due to their limited availability. This is particularly important because health professionals rely more on the patient's information than the examination findings when managing a patient [45].

Limitations

One of the main limitations is the small number of health questions used in our study. The increasing diversity of analyzed health questions will strength the generalization of our conclusion. Future studies may include more health questions and use an automated approach to analyze the content variables of health questions. In addition to not fully exploring platform effects, we examined queries from only two distinct types of online Q&A communities. Health questions are posted at diverse online platforms with varying features and functionalities. This limited sample size and lack of platform-specific analysis restrict the generalizability of our findings to other types of Q&A communities. Future research could expand the scope by including a broader range of Q&A platforms and investigate how platform-specific factors like guidelines, moderation, and user demographics interact with question characteristics to influence response dynamics across diverse online communities. This comprehensive approach would provide a deeper understanding of how the online environment shapes user behavior and information exchange in different Q&A contexts. Furthermore, this study establishes a strong foundation for investigating question quality assessment in the context of kidney disease. Future research can then expand on this work by conducting empirical validations in other chronic diseases.

In addition, the validation of this new quality measurement was based on human judges who have experience in finding online information but not necessarily in answering other HIC questions. Ideally, further validation with other quality measures that have already been evaluated will add another layer of validity to the study. For example, answers can reveal additional insights about the quality of the questions, and future exploration of textual quality indicators should also consider the quality of the answers to the questions. Our new quality measures could potentially be used to analyze the peculiarities of high-profile HIC queries in online forums. While our current focus is on measures like readability and clarity, future research can delve into prompt engineering, exploring how users craft effective questions for healthcare information systems. The rise of platforms like ChatGPT underscores this need, as our findings provide a foundation for equipping users with the skills to formulate high-quality prompts regardless of the platform. Longitudinal studies can then assess the real-world impact on information access and communication success in AI-powered healthcare.

Conclusions

The increasing popularity and usage of online Q&A communities for seeking health information calls for an investigation into factors influencing the content quality and ways to improve the quality of questions. By identifying the content and language variables of health questions that affect the quality, this research contributes to a deeper understanding of how users can formulate effective questions that receiving accurate and relevant responses. This research would offer meaningful suggestions for platforms managers and users as well. This knowledge can empower consumers to become more active participants in their online health information seeking journeys. By formulating concise, specific, and focused questions, individuals can maximize the effectiveness of online Q&A platforms and increase the likelihood of receiving high-quality responses from healthcare professionals and other informed users. Furthermore, our findings provide valuable insights for platform managers seeking to enhance the quality of user interactions within their communities. By promoting best practices for question formulation through educational resources and user guidance, online Q&A platforms can foster a more efficient and trustworthy environment for the exchange of health information. This research paves the way for future investigations into the dynamic interplay between user behavior, platform characteristics, and response quality within online health communities. Through continued exploration, we can work towards optimizing the online environment for effective health information access and utilization.

Acknowledgements

Conflicts of Interest

There are no conflicts of interest to declare.

Abbreviations

HIC: health information consumers
LIWC: linguistic inquiry and word count
Q&A: online question and answer

References

1. Shah C, Oh S, Oh JS. Research agenda for social Q&A. Libr Inf Sci Res [Internet] Elsevier Inc.; 2010;31(4):205–209. [doi: 10.1016/j.lisr.2009.07.006]
2. Liu Q, Agichtein E, Dror G, Maarek Y, Szpektor I. When web search fails, searchers become askers. Proc 35th Int ACM SIGIR Conf Res Dev Inf Retr - SIGIR '12 [Internet] 2012;801. PMID:22352717
3. Barry CL. User-Defined Relevance Criteria: An Exploratory Study. J Am Soc Inf Sci Technol 1994;45(3):149–159.
4. Taylor RS. Question-Negotiation and Information Seeking in Libraries. Coll Res Libr 1968;29(3):178–194.
5. Belkin NJ. Anomalous states of knowledge as a basis for information retrieval. Can J Inf Libr Sci 1980;5:133–143. [doi: 10.1037/0735-7044.119.1.78]
6. Ameri F, Keeling K, Salehnejad R. You Get What You Pay for on Health Care Question and

Answer Platforms: Nonparticipant Observational Study. J Med Internet Res [Internet] 2020 Jan;22(1):e13534. [doi: 10.2196/13534]
7. Baltadzhieva A, Chrupa\la G. Question Quality in Community Question Answering Forums: A Survey. SIGKDD Explor Newsl [Internet] New York, NY, USA: Association for Computing Machinery; 2015 Sep;17(1):8–13. [doi: 10.1145/2830544.2830547]
8. Shah C, Radford ML, Connaway LS, Choi E, Kitzie V, Brunswick N. " How Much Change Do You Get from 40 $?" – Analyzing and Addressing Failed Questions on Social Q & A. 2012;
9. Treude C, Barzilay O, Storey M-A. How do programmers ask and answer questions on the web? 33rd Int Conf Softw Eng 2011. [doi: 10.1145/1985793.1985907]
10. Li B, Jin T, Lyu MR, King I, Mak B. Analyzing and Predicting Question Quality in Community Question Answering Services. Proc 21st Int Conf World Wide Web [Internet] New York, NY, USA: Association for Computing Machinery; 2012. p. 775–782. [doi: 10.1145/2187980.2188200]
11. Teevan J, Morris M, Panovich K. Factors Affecting Response Quantity, Quality, and Speed for Questions Asked Via Social Network Status Messages. Proc Fifth Int AAAI Conf Weblogs Soc Media [Internet] 2011;630–633. Available from: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2795/3219
12. Nasehi SM, Sillito J, Maurer F, Burns C. What makes a good code example?: A study of programming Q&A in StackOverflow. IEEE Int Conf Softw Maintenance, ICSM IEEE; 2012;25–34. [doi: 10.1109/ICSM.2012.6405249]
13. Chatterjee P, Damevski K, Kraft NA, Pollock L. Automatically Identifying the Quality of Developer Chats for Post Hoc Use. ACM Trans Softw Eng Methodol [Internet] New York, NY, USA: Association for Computing Machinery; 2021 Jul;30(4). [doi: 10.1145/3450503]
14. Fu H, Oh S. Quality assessment of answers with user-identified criteria and data-driven features in social Q&A. Inf Process Manag [Internet] Elsevier; 2019;56(1):14–28. [doi: 10.1016/j.ipm.2018.08.007]
15. Shi J, Du W, Xu W. Identifying Impact Factors of Question Quality in Online Health Q&A Communities: an Empirical Analysis on MedHelp. PACIS 2018 Proc 2018;173.
16. Zhao Y, Wu L, Zhang J, Le T. How Question Characteristics Impact Answer Outcomes on Social Question-and-Answer Websites. J Glob Inf Manag [Internet] Hershey, PA, USA: IGI Global; 2021;29(6):1–21. [doi: 10.4018/JGIM.20211101.oa20]
17. Kitzie V, Choi E, Shah C. Analyzing question quality through intersubjectivity: World views and objective assessments of questions on social question-answering. Proc ASIST Annu Meet 2013;50(1). [doi: 10.1002/meet.14505001052]
18. Rassbach L, Pincock T, Mingus B. Exploring the Feasibility of Automatically Rating Online Article Quality. Int Wikimedia Conf [Internet] 2007 [cited 2019 Dec 17]. Available from: https://pdfs.semanticscholar.org/789e/3b36a645bec2a99bc0c4d9760be3f22c40a9.pdf
19. Alfattni G, Belousov M, Peek N, Nenadic G. Extracting Drug Names and Associated Attributes From Discharge Summaries: Text Mining Study. JMIR Med informatics 2021 May;9(5):e24678. PMID:33949962
20. Hersh W. Information Retrieval. In: Shortliffe EH, Cimino JJ, editors. Biomed Informatics Comput Appl Heal Care Biomed [Internet] Cham: Springer International Publishing; 2021. p. 755–794. [doi: 10.1007/978-3-030-58721-5_23]
21. Chu Z, Chen M, Chen J, Wang M, Gimpel K, Faruqui M, Si X. How to Ask Better Questions? A Large-Scale Multi-Domain Dataset for Rewriting Ill-Formed Questions. Proc AAAI Conf Artif Intell [Internet] 2020;34(05):7586–7593. [doi: 10.1609/aaai.v34i05.6258]
22. Ghosh S. Predicting Question Deletion and Assessing Question Quality in Social Q{\&}A Sites using Weakly Supervised Deep Neural Networks. Proc 54th Hawaii Int Conf Syst Sci [Internet] Hawaii International Conference on System Sciences; 2021. [doi:

10.24251/hicss.2021.329]
23. Workman TE, Shao Y, Divita G, Zeng-Treitler Q. An efficient prototype method to identify and correct misspellings in clinical text. BMC Res Notes [Internet] BioMed Central; 2019;12(1):1–5. [doi: 10.1186/s13104-019-4073-y]
24. Ruch P, Baud RH, Geiddbühler A, Lovis C, Rassinoux AM, Rivière A. Looking back or looking all around: comparing two spell checking strategies for documents edition in an electronic patient record. Proceedings AMIA Symp [Internet] 2001 [cited 2020 Feb 4];568–72. PMID:11837217
25. Zhou S, Jeong H, Green PA. How consistent are the best-known readability equations in estimating the readability of design standards? IEEE Trans Prof Commun Institute of Electrical and Electronics Engineers Inc.; 2017 Mar 1;60(1):97–111. [doi: 10.1109/TPC.2016.2635720]
26. Kincaid J, Fishburne R, Rogers R, Chissom B. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. Inst Simul Train [Internet] 1975 Jan 1 [cited 2019 Dec 27]; Available from: https://stars.library.ucf.edu/istlibrary/56
27. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. The development and psychometric properties of LIWC2007. LIWC Inc, Austin, Texas 78703 USA conjunction with LIWC2007 Softw Progr LIWC Inc, Austin, Texas 78703 USA in conjunction with the LIWC2007 software program; 2007.
28. MacQueen J. Some methods for classification and analysis of multivariate observations. Proc 5th Berkeley Symp Math Stat Probab 1967;1(14):281–297.
29. Tavakoli L, Zamani H, Scholer F, Croft WB, Sanderson M. Analyzing clarification in asynchronous information-seeking conversations. J Assoc Inf Sci Technol [Internet] 2022;73(3):449–471. [doi: https://doi.org/10.1002/asi.24562]
30. Liu Y, Bian J, Agichtein E. Predicting information seeker satisfaction in community question answering. Proc 31st Annu Int ACM SIGIR Conf Res Dev Inf Retr - SIGIR '08 [Internet] New York, New York, USA: ACM Press; 2008 [cited 2019 Sep 2]. p. 483. [doi: 10.1145/1390334.1390417]
31. Rath M, Shah C, Floegel D. Identifying the reasons contributing to question deletion in educational Q\&A. Proc Assoc Inf Sci Technol [Internet] 2017;54(1):327–336. [doi: https://doi.org/10.1002/pra2.2017.14505401036]
32. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics [Internet] JSTOR; 1977 Mar [cited 2020 Jun 22];33(1):159. [doi: 10.2307/2529310]
33. Zhang Y. Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. Proc 1st ACM Int Heal [Internet] 2010;210–219. [doi: 10.1145/1882992.1883023]
34. Roberts K, Demner-Fushman D. Interactive use of online health resources: A comparison of consumer and professional questions. J Am Med Informatics Assoc 2016;23(4):802–811. PMID:27147494
35. Alasmari A, Zhou L. Share to Seek: The Effects of Disease Complexity on Health Information--Seeking Behavior. J Med Internet Res [Internet] 2021 Mar;23(3):e21642. [doi: 10.2196/21642]
36. Alasmari A, Zhou L. How Multimorbid Health Information Consumers Interact in Online Q&A Websites. Int J Med Inform 2019;
37. Daniel W. Coyne. Management of Chronic Kidney Disease Comorbidities [Internet]. CKD Medscape C Expert Column Ser Issue 3. 2011. Available from: https://www.medscape.org/viewarticle/736181

38. Prevention C for DC and. Chronic Kidney Disease in the United States, 2021 [Internet]. www.cdc.gov. 2021. Available from: https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html
39. American Kidney Fund. Kidney Problem Symptoms, Causes, and Types - American Kidney Fund (AKF). 2019.
40. Corbin J. The Corbin and Strauss Chronic Illness Trajectory model: an update. Sch Inq Nurs Pract 1998;12(1):33–41. PMID:9805470
41. Skeels MM. Sharing by Design: Understanding and Supporting Personal HealthInformation Sharing and Collaboration within Social Networks [Internet]. 2010 [cited 2019 Aug 12]. Available from: http://eds.a.ebscohost.com/eds/detail/detail?vid=1&sid=a99ddb31-080c-443c-9a7d-5be749ecd369%40sdc-v-sessmgr03&bdata=JkF1dGhUeXBlPWlwLHVybCx1aWQmc2l0ZT1lZHMtbGl2ZSZzY29wZT1zaXRl#AN=ED523577&db=eric
42. Hinkle JL, Cheever KH. Brunner and Suddarth's textbook of medical-surgical nursing. Wolters kluwer india Pvt Ltd; 2018.
43. Wang YC, Kraut RE, Levine JM. Eliciting and receiving online support: Using computer-aided content analysis to examine the dynamics of online social support. J Med Internet Res [Internet] Journal of Medical Internet Research; 2015 Apr 1 [cited 2020 Jul 26];17(4):e99. [doi: 10.2196/jmir.3558]
44. Horgan D, Romao M, Morré SA, Kalra D. Artificial Intelligence: Power for Civilisation – and for Better Healthcare. Public Health Genomics [Internet] S. Karger AG; 2019 Feb 1 [cited 2020 Jul 26];22(5–6):145–161. [doi: 10.1159/000504785]
45. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, Ito M. Role of Artificial Intelligence within the Telehealth Domain. Yearb Med Inform [Internet] NLM (Medline); 2019 Aug 1 [cited 2020 Jul 26];28(1):35–40. [doi: 10.1055/s-0039-1677897]

Supplementary Files

Figures

Architecture of health question quality measurement.