



Exploring Patient Information Needs in Type 2 Diabetes: A Cross Sectional Study of Questions

Crangle, C., Bradley, C., Carlin, P., Esterhay, P., Harper, R., Kearney, P., McCarthy, V., Mc Tear, M., Savage, E., Wallace, J. G., & Tuttle, M. (2018). Exploring Patient Information Needs in Type 2 Diabetes: A Cross Sectional Study of Questions. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0203429>

[Link to publication record in Ulster University Research Portal](#)

Published in:
PLoS ONE

Publication Status:
Published (in print/issue): 16/11/2018

DOI:
[10.1371/journal.pone.0203429](https://doi.org/10.1371/journal.pone.0203429)

Document Version
Author Accepted version

General rights

The copyright and moral rights to the output are retained by the output author(s), unless otherwise stated by the document licence.

Unless otherwise stated, users are permitted to download a copy of the output for personal study or non-commercial research and are permitted to freely distribute the URL of the output. They are not permitted to alter, reproduce, distribute or make any commercial use of the output without obtaining the permission of the author(s).

If the document is licenced under Creative Commons, the rights of users of the documents can be found at <https://creativecommons.org/share-your-work/licenses/>.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk

1 **Exploring Patient Information Needs in Type 2 Diabetes: A Cross Sectional Study of**
2 **Questions**

3
4 Colleen E. Crangle^{1*}, Colin Bradley², Paul F. Carlin³, Robert J Esterhay¹, Roy Harper⁴, Patricia
5 M. Kearney⁵, Vera J.C. McCarthy⁶, Michael F. McTear⁷, Eileen Savage⁶, Mark S. Tuttle⁸,
6 Jonathan G. Wallace⁷

7
8 ¹Department of Health Management and Systems Sciences, School of Public Health and
9 Information Sciences, University of Louisville, Kentucky, United States of America

10
11 ² Department of General Practice, University College Cork, Cork, Ireland

12
13 ³South Eastern Health and Social Care Trust, Research Office, Northern Ireland, United
14 Kingdom

15
16 ⁴ Department of Endocrinology & Diabetes, The Ulster Hospital, South Eastern Health and
17 Social Care Trust, Northern Ireland, United Kingdom.

18
19 ⁵ Department of Epidemiology & Public Health, University College Cork, Cork, Ireland

20
21 ⁶ School of Nursing and Midwifery, University College Cork, Cork, Ireland

22
23 ⁷School of Computing and Mathematics, University of Ulster at Jordanstown, Northern Ireland,
24 United Kingdom

25
26 ⁸ Board of Directors, Apelon Inc., Hartford, Connecticut, United States of America

27
28 * Corresponding author

29 E-mail: crangle@alumni.stanford.edu

30 **Abstract**

31 This study set out to analyze questions about type 2 diabetes mellitus (T2DM) from patients and
32 the public. The aim was to better understand people's information needs by starting with what
33 they do not know, discovered through their own questions, rather than starting with what we
34 know about T2DM and subsequently finding ways to communicate that information to people
35 affected by or at risk of the disease. One hundred and sixty-four questions were collected from
36 120 patients attending outpatient diabetes clinics and 300 questions from 100 members of the
37 public through the Amazon Mechanical Turk crowdsourcing platform. Twenty-three general and
38 diabetes-specific topics and five phases of disease progression were identified; these were used
39 to manually categorize the questions. Analyses were performed to determine which topics, if
40 any, were significant predictors of a question's being asked by a patient or the public, and
41 similarly for questions from a woman or a man. Further analysis identified the individual topics
42 that were assigned significantly more often to the crowdsourced or clinic questions. These were
43 CAUSES (CI: [-0.07, -0.03], $p < .001$), RISK FACTORS ([-0.08, -0.03], $p < .001$), PREVENTION ([-
44 0.06, -0.02], $p < .001$), DIAGNOSIS ([-0.05, -0.02], $p < .001$), and DISTRIBUTION OF A DISEASE IN
45 A POPULATION ([-0.05, -0.01], $p = .0016$) for the crowdsourced questions and TREATMENT ([0.03,
46 0.01], $p = .0019$), DISEASE COMPLICATIONS ([0.02, 0.07], $p < .001$), and PSYCHOSOCIAL ([0.05,
47 0.1], $p < .001$) for the clinic questions. No highly significant gender-specific topics emerged in
48 our study, but questions about WEIGHT were more likely to come from women and
49 PSYCHOSOCIAL questions from men. There were significantly more crowdsourced questions
50 about the time PRIOR TO ANY DIAGNOSIS ([-0.11, -0.04], $p = .0013$) and significantly more clinic
51 questions about HEALTH MAINTENANCE AND PREVENTION after diagnosis ([0.07, 0.17], $p < .001$).
52 A descriptive analysis pointed to the value provided by the specificity of questions, their

53 potential to disclose emotions behind questions, and the as-yet unrecognized information needs
54 they can reveal. Large-scale collection of questions from patients across the spectrum of T2DM
55 progression and from the public – a significant percentage of whom are likely to be as yet
56 undiagnosed – is expected to yield further valuable insights.

57

58 **Introduction**

59 Diabetes is a major health problem worldwide. The prevalence of global, age-standardized diabetes
60 is 9% in men and 7.9% in women, with the number having risen around the globe from 108 million
61 in 1980 to 422 million in 2014 [1]. Type 2 diabetes mellitus (T2DM) is a main driver of the
62 increase, accounting for approximately 90% of all diabetes cases [2-4]. Diabetes is a complex
63 condition and people with diabetes have a diverse range of information needs [5-8]. Large-scale
64 investigations such as the DAWN studies on the attitudes, wishes and needs of patients and
65 caregivers [9, 10] have told us much, but research to date has paid little attention to exploring the
66 information needs of patients as expressed in the questions they have about diabetes. Questions
67 convey information needs in the patient's own voice and permit the individual and subjective
68 experience of illness to be captured [11]. To our knowledge, no-one to date has investigated on a
69 large scale what T2DM patients want to know at different stages of diagnosis and treatment by
70 asking them directly what their questions are, nor have questions from the public been solicited
71 and examined.

72 Our study concerns a new way of thinking about patient information needs in diabetes,
73 starting not with what we know about T2DM and finding ways to communicate that information
74 to patients but starting with what patients do not know, discovered through their own questions.
75 Soliciting, and then responding, to patient questions on a large scale has the potential to create a

76 new information resource for T2DM, both in terms of content and organization. A questions-
77 based approach to patient knowledge is distinct from active information seeking through which
78 the patient searches extant information resources [5], and it is distinct from passive information
79 receipt in which the patient is exposed either accidentally or deliberately to extant information
80 resources [12]. A questions-based approach has the potential to create a dynamic, continually
81 updated resource that will capture patient information needs as they evolve over time.

82 It is estimated that more than half of American adults have either T2DM or prediabetes
83 (as measured by blood sugar levels or determined by diagnosis) and of those more than one-third
84 are unaware they have the disease [13]. Consequently, it is crucial that we understand the
85 information needs and voice of those who do not have diabetes, or do not know they have
86 diabetes, but still have questions whether out of curiosity or concern for themselves or a loved
87 one. In this paper, we report on the first stage of our work soliciting questions directly from both
88 patients and the general public and analyzing the questions to see what they reveal.

89 Questions play a vital role in health care. Patient questions foster good communication
90 with health professionals, resulting in better care and the right care at the right time [14-17].
91 However, poor bi-directional flow of information between the diabetes health professional and the
92 patient has been documented. Discrepancies have also been noted between information provided
93 by health care providers and what patients with diabetes need [18]. Patients often cannot get as
94 much detail as they need during office visits [19]. Time constraints, whether actual or perceived,
95 prevent some patients from asking questions during the consultation [5, 20]. Patients also find it
96 difficult to retain much of what they have been told by a health professional, and what they do
97 remember is incorrect almost half the time [21-23].

98 Clinical information needs have been extensively studied by collecting questions from

99 physicians and analyzing them [24-44]. For patients and the general population, the situation is
100 very different. Only recently have their health questions been studied in any depth [45-51], with
101 few studies, to our knowledge, focusing on diabetes or investigating differences between questions
102 from patients and those not in a patient setting. Our recent study has shown that available online
103 sources of information do not provide answers to patient questions about diabetes and that there is
104 an urgent need to better understand these information needs [52]. In this study, therefore, we set
105 out to collect and investigate questions about diabetes from two sources, namely, patients attending
106 a diabetes clinic and the public through crowdsourcing. We hypothesized that an analysis of the
107 questions in terms of the topics they cover and the phases of disease progression they concern
108 would provide important insights, potentially also revealing differences in information needs
109 between patients and those outside the patient setting, who may or may not have diabetes or may
110 be unaware they have the disease.

111

112 **Methods**

113 **Ethics statement**

114 This study makes secondary use of anonymized data. A prior service evaluation had been
115 approved within the South Eastern Health and Social Care Trust, Northern Ireland, to assess
116 patient information needs by approaching patients attending the diabetes clinic and asking them
117 to provide questions. They were free to refuse if they wished to. No participant consent was
118 needed for the service evaluation and none was sought. Questions were recorded on a sheet
119 provided to each patient if interested, with no identifiers such as clinic time, clinician or personal
120 information collected. No ethics committee approval was needed for our secondary analysis of

121 the collected questions. This practice conforms to the guidelines of the Health Research
122 Authority of the UK National Health Service and current UK legislative and good practice
123 arrangements. The authors had no direct contact with the participants and there were no minors
124 among the participants.

125 **Question collection**

126 As part of a prior service evaluation, all patients attending the weekly diabetes outpatient clinic at
127 the Ulster Hospital in Northern Ireland during February to April, 2014 had been invited to submit
128 questions by responding to the following: *What are the one or two most pressing questions about*
129 *your diabetes that you would like answered?* Patients were provided with a blank page to record
130 their questions and questions from the same individual were marked as such.

131 We obtained additional questions using the crowdsourcing platform of Amazon
132 Mechanical Turk (AMT). Crowdsourcing has become an important part of many clinical studies
133 [53], with new platforms emerging to meet the particular requirements of research [54]. One
134 hundred AMT participants were asked to each enter three questions s/he had about diabetes.
135 Each participant was asked to specify age, sex/gender and if s/he had a diagnosis of type 2 or
136 type 1 diabetes, or a diagnosis of diabetes but did not know the type, and if s/he had a friend or
137 family member with a diagnosis of type 2, type 1 or unknown type. Crowdsourced question
138 collection took place on July 8th -11th, 2015.

139 The clinic and crowdsourced question corpora are given in S1 File and S2 File. All
140 questions are presented as written by the participants, with spelling and punctuation intact.

141 **Categorization by topic and phase**

142 Question content was determined through fine-grained manual categorization of the topics and the
143 phases of diabetes progression the question referred to. Such detailed assessment of need is part of

144 the move towards better disease management through understanding the likely information needs
145 of different subgroups of people at different phases of the disease, at the onset of diabetes, for
146 example, or later when a new complication has developed.

147 An initial set of 13 topics, based on known concerns of patients with diabetes [5-10, 55-
148 57] and our prior work on consumer questions [16], was compiled and used to conduct a
149 preliminary categorization of the crowdsourced questions. This undertaking led to an expanded
150 set of 23 topics for use in this study. We additionally compiled a five-part patient-oriented
151 classification of the phases of T2DM drawing on prior work and our clinical experience [58-
152 67].

153 Two researchers independently categorized each question by topic and phase. A question
154 could fall under more than one topic and more than one phase, but the phases had to be
155 consecutive, as in the range 3-5, for instance. There were therefore more question-topic
156 assignments and more question-phase assignments than there were questions.

157 Coding was performed by CEC (all crowdsourced questions), PK (half the crowdsourced
158 questions), VMC (half the crowdsourced questions), and PC and RH (all clinic questions each).
159 PK and RH are clinicians, VMC and PC healthcare researchers, and CEC a non-clinical
160 bioinformatics researcher. For each question and topic, a score of 1 indicated that the question
161 fell under that topic and a score of 0 that it did not. If both coders scored 1 or 0 for a question and
162 topic, it was counted as agreement. Agreement for phase was determined by an overlap between
163 one coder and the other. Intercoder reliability was computed using Cohen's kappa with the
164 following guidelines from [68]: slight agreement (0-0.2); fair (0.21 – 0.4); moderate (0.41 - 0.6);
165 substantial (0.61 – 0.8); almost perfect (0.81 – 1). Disagreement between coders was resolved
166 through consensus review by the coders and members of the project team.

167 **Statistical analysis**

168 The following analyses were performed for topics and stratified by sex for the crowdsourced
169 questions. The significance threshold was set at .05 except where indicated.

170 Because consecutive questions are more likely to stem from the same questioner in each
171 corpus, the samples cannot be assumed to be independent. We therefore determined which, if
172 any, individuals had highly correlated questions in terms of their topic assignments using the
173 Pearson correlation coefficient. Then, following the guideline that multicollinearity may be a
174 problem in a data set if any pairwise $|r| > 0.7$ [69], we removed the questions from any individual
175 who had strongly correlated questions ($|r| > 0.7$ for any pair of his/her questions). For each
176 corpus, we also examined all pairwise correlations between topics, in terms of the questions
177 assigned to them, removing those topics, if any, that were strongly correlated.

178 To determine which topics, if any, were significant predictors of a question's coming
179 from a patient in the clinic or from the public through crowdsourcing, we used Lasso regression
180 with the Least Angle Regression (LARS) algorithm [70,71], Lasso-LARS is a model selection
181 algorithm that uses repeated internal cross-validation to select variables and estimate coefficients
182 in the presence of collinearity. We applied Lasso-LARS both before and after removal of highly
183 correlated questions and topics. Computations were performed using the *LassoLarsCV* function
184 from the *scikit-learn* python package with 10-fold cross validation and default parameters [72].
185 Lasso-LARS regression was also performed on the crowdsourced questions to determine which,
186 if any, topics were significant predictors of a question's coming from a woman or a man.

187 We also examined each topic individually to determine if it was assigned significantly
188 more often to the clinic or the crowdsourced questions, correcting for multiple comparisons
189 using the Benjamini–Hochberg false discovery rate (FDR) [73-75]. For the crowdsourced

190 questions only, we similarly asked for each topic if it was assigned significantly more often to
191 the questions asked by men or those asked by women. The 2-tailed z-test provided 95%
192 confidence intervals (CI) for these estimates. This analysis told us something about the topics, in
193 contrast to the Lasso-LARS analysis that told us something about the questions and the people
194 asking them. The z-tests were performed after confirmation that the distribution of questions
195 over topics was approximately normal. That is, we confirmed that the number of questions per
196 topic was approximately normally distributed for both the crowdsourced and clinic questions
197 under the Shapiro-Wilk test, both before and after removal of the correlated questions, and
198 similarly for the female and male questions [76]. For the phases of disease progression, a similar
199 analysis was done to determine which phases, if any, were assigned significantly more often to
200 the clinic or the crowdsourced questions.

201 To gain additional understanding of the differences between the clinic and crowdsourced
202 questions, the top three (85th percentile) and top five (75% percentile) topics in terms of the
203 number of questions to which they were assigned were identified for each corpus. Those that
204 were top in one corpus and not the other were recognized as characteristic of that corpus. A
205 similar analysis was done for the phases of disease progression.

206

207 **Descriptive analysis**

208 In addition to topic analysis and the analysis by phase of disease progression, the combined
209 corpus of questions was reviewed from a holistic and descriptive perspective to ascertain any
210 inferences implicit in the questions that might reveal underlying concerns or issues for the person
211 generating the question. It was apparent that the questioners, not all of whom had diabetes, were
212 seeking more than just factual information. A limited qualitative analysis of the combined corpus

213 was therefore undertaken to address this need for a broader interpretation of the questions beyond
214 their literal content. This analysis was not exhaustive but illustrative, identifying themes that
215 might inform a detailed analysis of a larger collection of questions.

216

217 **Results**

218 **The topics**

219 A preliminary categorization of the crowdsourced questions using a core set of categories
220 derived from earlier work [5-10, 16,55-57] produced a Cohen’s kappa score of 0.61 overall,
221 which represents moderate to substantial agreement [68]. A subsequent round-table discussion by
222 members of the project team (CEC, PK, VMC, MFM, ES, JGW and PC) led to the formulation of
223 the 23 categories described in Table 1. Several diabetes-specific categories, namely LIFESTYLE /
224 BEHAVIOR CHANGE (hereafter abbreviated simply to LIFESTYLE), EXERCISE, DIET, WEIGHT, and
225 CURE OR REVERSAL, were added to the core categories. For this last topic, we note that a more
226 clinically oriented topic descriptor would be CONTROL OR REMISSION. However, our experience to
227 date with patient and general-public questions is that the lay perception centers on the idea of
228 completely getting rid of a disease and for this reason we use the descriptor CURE OR REVERSAL.
229 The topic of COMPLICATIONS derived from earlier work was split into DISEASE COMPLICATIONS
230 and TREATMENT COMPLICATIONS to properly represent the types of questions found for T2DM.

231

232 Table 1. Topic Categories for T2DM Questions.

233

CAUSES

Questions about the causes of diabetes or one of its complications. Includes

	causal factors that might increase risk and causes of symptoms.
RISK FACTORS	Questions about factors that raise the risk of developing diabetes or any of its complications (not necessarily causal factors, for example, gender).
PREVENTION	Questions about the prevention of diabetes or the prevention of complications arising from diabetes.
DIAGNOSIS	Questions about diagnostic tests for diabetes or any of its complications. Includes questions about signs or symptoms that might lead to a diagnosis. Includes methods for determining the difference between pre-diabetes, type 1 and type 2.
MANIFESTATIONS	Questions about signs or symptoms of diabetes or any of its complications.
TREATMENT	Questions about treatments for diabetes. Includes medication and self-management behaviors that could be part of a treatment plan.
ANATOMY	Questions that make reference to any particular part of the body, such as questions about a location affected by diabetes.
CURE / REVERSAL	Questions about a cure for diabetes or about the reversal of symptoms to the point where one could be considered condition free or in remission.
DIET / NUTRITION	Questions about the role of diet or nutrition in the prevention, development or management of diabetes and its complications.
EXERCISE	Questions about the role of exercise in the prevention, development or management of diabetes and its complications.
WEIGHT	Questions about the role of weight in the prevention, development or management of diabetes and its complications.
LIFESTYLE	Questions about things a person can or must do to prevent or manage diabetes or its complications (including diet, exercise, or weight).
DISEASE COMPLICATIONS	Questions about the problems diabetes causes. This includes the risks faced by patients with diabetes and the nature and experience of the complications.
TREATMENT COMPLICATIONS	Questions about problems arising from specific treatments for diabetes or one of its complications.
PERSON OR ORGANIZATIONS	Questions about a person or organization involved with a disease. This can include medical specialists, hospitals, research teams, insurance payments, or support groups for a particular disease.
PROGNOSIS	Questions asking about life expectancy, quality of life, or the probability of success of a given treatment.
DISTRIBUTION OF A DISEASE IN A POPULATION	Questions about the occurrence of diabetes in a population and questions about the distribution of complications in the population of people with diabetes.
INHERITANCE PATTERNS	Questions about inheritance patterns in diabetes.

TRANSMISSION PATTERNS	Questions about transmission patterns for diabetes (when conceived of as an infectious disease).
RESEARCH	Questions about research on diabetes. Includes questions about clinical trials.
PSYCHOSOCIAL	Questions about the social-emotional ramifications of diabetes.
OWN HEALTH RECORD RELATED	Questions that relate specifically to the questioner's own health or that reference information in the person's health record. Includes questions about "my" medication, etc.
OTHER	Questions that do not belong to any of the above. Includes non-medical questions about a disease, such as policy decisions, for example.

234
235

236 **The questions**

237 One hundred and sixty-four questions were collected from 120 patients during 12 outpatient
238 clinics. Most of the questions were about diabetes (N=155) with the remainder related to clinic
239 operation (N=9). Of the questions on diabetes, 152 from 101 patients were about T2DM and
240 these questions were retained for our analysis. Although only 1 to 2 questions were asked for, 2
241 patients gave 3 questions each. Most of the patients attending the clinic had T2DM (95%).
242 These questions are given in S1 File.

243 For the crowdsourced questions, 100 AMT participants each contributed three questions
244 about diabetes (N=300). Most of the questions were about T2DM (N=284) with a smaller
245 number related to type 1 diabetes (N=15) and 1 question duplicated by one of the questioners. Of
246 the 100 questioners (F 34, M 66), 9 had diabetes (6 type 2, 2 type 1, one unknown type) and 91 a
247 friend or family member with diabetes (30 type 2, 17 type 1, 44 unknown type). The 284
248 questions about T2DM were retained for analysis. These questions are given in S2 File . For the
249 clinic questions overall, agreement between the coders was substantial (Cohen's kappa=0.77, SD
250 = 0.18). For the crowdsourced questions overall, agreement between the coders was almost
251 perfect (Cohen's kappa = 0.86 SD = 0.1). Disagreements were resolved by consensus between

252 the coders.

253 We found that for the crowdsourced questions, 16 of the 100 individuals had strongly
254 correlated questions ($|r| > 0.7$ for any pair of their 2 or 3 questions) and, for the clinic questions,
255 3 of the 101 patients. After removing the questions from the identified individuals there were 236
256 crowdsourced questions from 84 individuals and 147 clinic questions from 98 patients
257 remaining. We did not find the topics in the crowdsourced questions to be correlated at the 0.7
258 criterion value, but for the clinic questions, TRANSMISSION PATTERNS was correlated with
259 INHERITANCE PATTERNS at the 0.7 level. We therefore dropped the category TRANSMISSION
260 PATTERNS, which had only 2 questions in the crowdsourced corpus and 5 in the clinic corpus, all
261 of which also fell under other topic categories and consequently did not need to be removed.

262 **Topics and the clinic and crowdsourced questions**

263 The clinic questions had an average of 2.8 topics per question (min 1, max 7) and the
264 crowdsourced questions had an average of 2.1 topics per question (min 1, max 5). The results of
265 the Lasso-LARS regression on all questions showed slightly higher odds ratios in favor of
266 questions that were OWN HEALTH RECORD RELATED and about TREATMENT coming from the
267 clinic patients (1.143 and 1.114 respectively). The odds ratios for all other questions were less
268 than 1.062. The optimum alpha value found was 0.0009 with a mean squared error of 0.151 for
269 both the training and test data. Lasso-LARS regression on only the non-correlated questions
270 revealed similar slightly higher odds ratios in favor of the clinic questions for the same two
271 topics (1.173 and 1.156 for OWN HEALTH RECORD RELATED and TREATMENT, respectively), with
272 all other odds ratios less than 1.1. The optimum alpha value was 0.0006 with a mean squared
273 error for the training data of 0.132 and 0.187 for the test data.

274 In terms of the individual topics, topics that were assigned significantly more often to the

275 crowdsourced than the clinic questions were CAUSES (CI: [-0.07, -0.03], $p < .001$), RISK
 276 FACTORS ([-0.08, -0.03], $p < .001$), PREVENTION ([-0.06, -0.02], $p < .001$), DIAGNOSIS ([-0.05, -
 277 0.02], $p < .001$), and DISTRIBUTION OF A DISEASE IN A POPULATION ([-0.05, -0.01], $p = .0016$). In
 278 contrast, the topics TREATMENT ([0.03, 0.01], $p = .0019$), DISEASE COMPLICATIONS ([0.02, 0.07],
 279 $p < .001$), and PSYCHOSOCIAL ([0.05, 0.1], $p < .001$) were assigned significantly more often to the
 280 clinic questions. See Table 2.

281
 282 **Table 2. Topics Assigned to the Crowdsourced and Clinic Questions.**
 283

	Crowdsourced (N=236)	Clinic (N=147)	CI*	FDR- adjusted p- values, 2- tailed z-test
Causes	49 / 675**	14 / 598**	(-0.07, -0.03)	< .001
Risk Factors	64 / 675	23 / 598	(-0.08, -0.03)	< .001
Prevention	33 / 675	7 / 598	(-0.06, -0.02)	< .001
Diagnosis	31 / 675	6 / 598	(-0.05, -0.02)	< .001
Manifestations	45 / 675	43 / 598	(-0.02, 0.03)	.747
Treatment	61 / 675	91 / 598	(0.03, 0.01)	.0019
Anatomy	20 / 675	29 / 598	(-0.0, 0.04)	.1383
Cure / Reversal	47 / 675	28 / 598	(-0.05, 0.0)	.1383
Diet / Nutrition	8 / 675	5 / 598	(-0.01, 0.01)	.6942
Exercise	12 / 675	9 / 598	(-0.02, 0.01)	.747
Weight	65 / 675	56 / 598	(-0.03, 0.03)	.8721
Lifestyle	33 / 675	43 / 598	(-0.0, 0.05)	.1383
Disease Complications	16 / 675	40 / 598	(0.02, 0.07)	< .001

Treatment Complications	7 / 675	16 / 598	(0.0, 0.03)	.0696
Person or Organization	41 / 675	41 / 598	(-0.02, 0.03)	.6973
Prognosis	18 / 675	7 / 598	(-0.03, -0.0)	.1207
Distribution of a Disease in a Population	26 / 675	5 / 598	(-0.05, -0.01)	.0016
Inheritance Patterns	15 / 675	20 / 598	(-0.01, 0.03)	.3251
Research	20 / 675	20 / 598	(-0.02, 0.02)	.747
Psychosocial	4 / 675	47 / 598	(0.05, 0.1)	< .001
Own Health Record Related	27 / 675	30 / 598	(-0.01, 0.03)	.5244

284 *Confidence intervals at the .05 level

285 **The denominator in each column is the number of topic assignments in total for the corpus.

286

287 The three most frequent clinic topics (Table 2) were TREATMENT (91 questions), WEIGHT
288 (56) and PSYCHOSOCIAL (47). The three most frequent crowdsourced topics also included
289 WEIGHT (65) and TREATMENT (61), but included RISK (64) rather than PSYCHOSOCIAL. The topic
290 PSYCHOSOCIAL therefore characterizes the clinic questions and the topic RISK FACTORS the
291 crowdsourced questions. The next two clinic topics were MANIFESTATIONS (43) and LIFESTYLE
292 (43), which were not in the top crowdsourced topics, and therefore further characterize the clinic
293 questions. The next two crowdsourced topics were CAUSES (49) and CURE / REVERSAL (47),
294 which were not in the top clinic questions, and therefore further characterize the crowdsourced
295 questions.

296 For the crowdsourced questions, Lasso-LARS regression showed a slightly higher odds
297 ratio (1.122) for a question about WEIGHT coming from a woman rather than a man. The

298 optimum alpha value found was 0.0024 and the mean squared error for the training data was
299 0.193 and 0.235 for the test data.

300 In terms of the individual topics, the only topic that was more strongly associated with
301 one gender over another was PSYCHOSOCIAL (CI: [0.02, 0.05], $p = .1497$), which was more
302 strongly represented in the male questions, but only at the 0.2 level. For both men and women,
303 the topics LIFESTYLE (women 24 questions; men 41 questions), RISK FACTORS (24; 41), and DIET
304 (19; 28) were most frequently assigned. There was one topic appearing uniquely in the top five
305 for men and one for women that might in addition be thought of as characterizing the two
306 groups. These were MANIFESTATIONS (women 18 question) and PROGNOSIS (men 24 questions).

307 **The phases of disease progression**

308 The five phases of T2DM progression we identified from the literature and our experience
309 were: PRIOR TO ANY DIAGNOSIS; PRE-DIABETIC (DIAGNOSED); ONSET OF T2DM; HEALTH
310 MAINTENANCE AND PREVENTION; COMPLICATIONS – MINOR (ONSET) OR MAJOR
311 (DOMINANCE). These are listed in Table 3 along with a description of each phase.

312
313 **Table 3. Phases of Disease Progression for T2DM.**

PHASE 1	PRIOR TO ANY DIAGNOSIS	Given that 27.8% (8.1 million) of those estimated to have diabetes in the USA are undiagnosed, it is important to understand the questions people may have prior to a diagnosis. Given the increasing prevalence of T2DM, many people have family members or friends with a diagnosis. Questions may therefore not be about their own susceptibility but about giving support to others with the condition.
PHASE 2	PRE-DIABETIC (DIAGNOSED)	On being diagnosed as pre-diabetic, a person's questions may reveal a response of confusion, denial or fear. In coming to terms with the fact that s/he may be facing a serious chronic illness, the person may have questions about the choices s/he has to make.

PHASE 3	ONSET OF T2DM	With the onset of T2DM, questions may reveal a response of anger or denial or a wait-and-see attitude. In accepting that s/he has to live with a serious chronic illness, the person may have questions about new knowledge that must be acquired and a new and possibly demanding self-care regimen that must be adjusted to. If the onset of diabetes is abrupt, the adjustment may be particularly difficult and questions may reflect this struggle.
PHASE 4	HEALTH MAINTENANCE AND PREVENTION	After initial diagnosis, questions may reveal a person energized to manage his/her diabetes. A lack of questions may indicate a person trying to ignore his/her diabetes. As treatment focuses on the prevention of diabetes complications, questions may focus on the new self-care behaviors that must be maintained.
PHASE 5	COMPLICATIONS – MINOR (ONSET) OR MAJOR (DOMINANCE)	With the onset of complications, a person may be energized by the complications to manage his/her diabetes. Or s/he may respond with fatalism or increased distress. Questions may reflect attempts to understand and accept a new condition trajectory. Questions may focus on maximizing quality of life, especially as complications come to dominate the person’s life. Questions may seek emotional support, not just knowledge, as the person’s self-image as a functioning, healthy adult undergoes possible change. The person must learn to live with challenges that may affect his/her activity levels, functional abilities, and emotional and social well-being. As the person’s condition progresses, new questions will arise.

314

315

316

For both the clinic and consumer questions, intercoder reliability for categorization by phase was moderate ($k = 0.64$ clinic; $k = 0.67$ crowdsourced). Given the exploratory nature of this categorization by phase, by consensus the coders agreed to assign a phase category to a question if either one of the coders did so. In this way, the judgements of all coders (clinical and non-clinical) could be taken into account.

320

321

Phases and the clinic and crowdsourced questions

322

There were significantly more crowdsourced questions that concerned Phase 1, the time PRIOR

323

TO ANY DIAGNOSIS (CI: [-0.11, -0.04], $p = .0013$), and significantly more clinic questions about

324 Phase 4, HEALTH MAINTENANCE AND PREVENTION (CI: [0.07, 0.17], $p < .001$). See Table 4.

325

326 **Table 4. Phases Assigned to the Crowdsourced and Clinic Questions.**

327

	Crowdsourced (N=236)	Clinic (N=147)	CI*	FDR- adjusted p- values, 2- tailed z-test
Phase 1: Prior to any Diagnosis	118 / 756**	32 / 386**	(-0.11, -0.04)	.0013
Phase 2: Pre-diabetic (diagnosed)	127 / 756	49 / 386	(-0.08, 0.0)	.0692
Phase 3: Onset of T2DM	220 / 756	91 / 386	(-0.11, -0.0)	.0647
Phase 4: Health Maintenance and Prevention	148 / 756	122 / 386	(0.07, 0.17)	< .001
Phase 5: Complications – Minor (onset) or Major (dominance)	143 / 675	92 / 386	(-0.0, 0.1)	.0647

328 *Confidence intervals at the .05 level

329 **The denominator in each column is the number of phase assignments in total for the corpus.

330

331 The most frequently applied phase in the clinic questions was Phase 4 (HEALTH
332 MAINTENANCE AND PREVENTION, 122 questions). This was followed by Phases 5 and 3, then
333 Phases 2 and 1. The most frequently applied phase in the crowdsourced questions was Phase 3
334 (ONSET OF T2DM, 220 questions) followed by Phases 4 and 5, then Phases 2 and 1. Phase
335 assignment numbers for clinic and crowdsourced questions are shown in Fig 1.

336

337 Fig 1. Clinic and Crowdsourced Questions by Phase.

338

339 **Descriptive analysis**

340 Our descriptive analysis identified four themes to pursue in future studies: (1) the specificity
341 found in questions; (2) questions revealing the emotion behind an information need; (3)
342 questions disclosing information needs not yet recognized in standard patient information
343 resources; and (4) the potential for questions to identify specific constituent groups with their
344 own information needs.

345 **(1) The specificity of questions**

346 Questions encourage specificity. The topic “diabetes and prognosis,” for instance, does not
347 capture the specificity of the following four prognosis-related questions taken from our corpora:

- 348 • Is diabetes a death sentence?
- 349 • Will all Type 2 eventually go on to insulin?
- 350 • Is there any potential for a cure within the next few years, according to current research?

351 The first concerns a worst outcome, the second the inevitability of a treatment, and the third a
352 best outcome. Numerous questions contemplated a decline in health, for example:

- 353 • Will it get worse [?]
- 354 • Will my condition only worsen [?]

355 Many asked about the likelihood and hoped-for outcome of specific treatments, for example:

- 356 • Can I ever reduce insulin & meds and feel good [?]
- 357 • Can diabetes be cured or rendered almost gone overtime through medicine and nutrition
358 [?]
- 359 • Could a pancreas transplant cure diabetes in a person?
- 360 • What dictates the type of treatment needed/required for diabetes, and is directly injecting
361 insulin ever avoidable?

362 Prognosis questions that were about a possible cure for diabetes were prominent. See S1 Table.
363 All questions contained the word “cure,” or similar, such as “reverse,” “heal,” “[fully] go away,”
364 “[completely] get over.” Many questions looked to scientific research for a cure and
365 acknowledged it as a matter for the future. A smaller number specifically referred to diet or
366 lifestyle changes, something an individual can do to affect the course of diabetes. Those asking
367 such questions may be more receptive to taking action on their own behalf.

368 **(2) Questions revealing the emotion behind an information need**

369 Consider the following two questions, both ostensibly seeking to understand why the questioner
370 has diabetes.

- 371 • How on earth I ever got diabetes in the first place. Never over weight blood pressure
372 always fine never eat sweet food [?]
- 373 • Why me?

374 The first question shows some understanding of risk factors for the condition without, it seems,
375 fully understanding genetic risks. Puzzlement and frustration are expressed. The second question
376 is less a plea for information and more an expression of frustration and defeat. Its meaning, and
377 what counts as an adequate answer, will differ depending on when it is asked – at diagnosis, at a
378 periodic review of the patient’s care, when a new complicating factor has arisen that will affect
379 self-management, or when there is a transition in care, such as with age-related changes or a
380 change in the care team [77].

381 **(3) Questions disclosing as yet unrecognized information needs**

382 It is crucial that information on diabetes covers not only what health professionals consider
383 important for people to know but also what the different constituent groups want to know,
384 whether considered important to health professionals or not. Directly solicited, open-ended

385 invitations to ask questions are a way to reveal information needs that may not be anticipated by
386 health educators. Take the following questions that in effect ask for a severity index for diabetes.

- 387 • Are there variations in severity to diabetes and what determines severity?
- 388 • To what extent does it exist on a spectrum, such that people may be classified according
389 to the degree to which they are diabetic, even if they are not diagnosable as diabetic
390 according to present criteria?
- 391 • Are there variations in severity to diabetes and what determines severity?

392 This topic is covered in the literature [78] but not prominently or not at all in the trusted and
393 vetted sources of patient-oriented diabetes information resources. It may be important to some
394 patients' needs to fully understand their condition. A related set of questions reveals a similar and
395 important wish, whether feasible or not, to be able to monitor one's health before it gets to a
396 point of no return [79], as explicitly stated in this question:

- 397 • If you suspect you have Type 2 diabetes, at which point will it become impossible for you
398 to reverse it by only changing your diet and exercise habits (and without requiring
399 medication or the need to see a doctor?)

400 **(4) Questions from specific constituent groups**

401 Diabetes information is important not only for people with diabetes but also friends and family of
402 people diagnosed with diabetes and for caregivers, those family members, neighbors, friends or
403 paid persons who regularly look after someone with diabetes. Our corpora included several such
404 questions.

- 405 • What are some ways to help a family member accept a diagnosis of diabetes?
- 406 • How hard is it to treat when the person who needs help isn't very receptive to their
407 condition?

408 • What are some things you can do to help a family member better manage an appropriate
409 diet for type 2 diabetes?

410 • What is the best way I can help my friends and family members with controlling their
411 diabetes?

412 For other chronic conditions such as mental-health disorders, for instance, the role of family and
413 friends is broadly acknowledged and discussed in education and information resources. Question
414 collection on a massive scale may suggest a more prominent place for this topic in diabetes
415 education.

416

417 **Discussion**

418 **Principal results**

419 The topics associated with the clinic questions (OWN HEALTH RECORD RELATED, DISEASE
420 COMPLICATIONS, TREATMENT, and PSYCHOSOCIAL) confirm what might be expected, namely that
421 patients whose condition is actively being managed are most concerned about complications of
422 the condition specific to their medical history, with a primary concern being about psychosocial
423 matters related to their disease. T2DM is a complex condition that has different disease
424 progressions for different people and for the same person over time and as life circumstances
425 change [69]. Significant effort has to go into making sense of the experience. A recent study
426 comparing people seeking online health information for their own problem against those seeking
427 information for someone else's showed that the first group in contrast to the second focused
428 primarily on symptoms and matters related to their own disease history [80].

429 The crowdsourced questions' focus on CAUSES, RISK FACTORS. PREVENTION, DIAGNOSIS,

430 and DISTRIBUTION OF A DISEASE IN A POPULATION most likely reflects the fact that the
431 crowdsourced questioners were in the main (91%) not themselves diagnosed but knew someone
432 who was and so likely sought to understand what leads to diabetes and who among their family
433 members may be at risk. Those seeking online information for someone else's health problem
434 have been shown to focus primarily on causes of a disease and disease terminology [80].

435 The stronger representation of PSYCHOSOCIAL questions from men warrants further
436 investigation. Gender-based notions of masculinity have been shown for some people to be in
437 conflict with effective self-management of T2DM, a central component in the treatment of
438 diabetes [81]. The stronger representation of questions on WEIGHT from women is perhaps not
439 unexpected, but with recent research showing that men are developing T2DM at lower levels of
440 adiposity than women, this may change [82].

441 The clinic questions, not surprisingly, predominantly concerned post-diagnosis issues
442 whereas the onset of diabetes dominated crowdsourced questions. The number of crowdsourced
443 questions asking, in effect, how you know if you have diabetes accounts for the high number of
444 questions categorized under Onset of T2DM. Such a concern is consistent with the fact that over
445 30 percent of those with diabetes in the United States are unaware they have the disease [13]. It
446 also perhaps indicates that the public health message about the prevalence of diabetes is being
447 heard and people are wondering about their own health status.

448 **Related studies**

449 There is a long and extensive record of questions being collected from health professionals and
450 analyzed. Questions have been collected at the point of care, from email consultation with
451 specialists, and through queries to information systems [26, 31, 33, 35-38]. Clinical questions
452 have been categorized as to the kind of knowledge they sought and the kind of answers they

453 needed, with taxonomic and other organizing structures proposed for them [24, 27, 35, 37]. The
454 questions of family-medicine, elder-care, and rural-health physicians have been explored [25, 29,
455 30, 32, 34, 44]. Experiments have been done on different ways of capturing clinical questions
456 through voice and other input media [28, 39-43]. Clinical questions associated with specific
457 disorders have been evaluated, most notably cancer [42], and T2DM [31]. A systematic review
458 of three decades of studies on clinical information needs [30] found that roughly 30% of the
459 question types accounted for 80% of the questions clinicians asked, where a question's type was
460 relative to a 64-item taxonomy [24].

461 Studies of questions from healthcare consumers are relatively recent. In [45], 276 health-
462 related questions posted on a social media question-answer website were subjected to qualitative
463 content analysis, focusing on meta-characteristics of the questions such as the users' motivations
464 for asking the questions. In [46] and [47] a manual topic-based analyses of consumer questions
465 was done using topics from the UMLS. In [49], 365 questions from a mailing list were analyzed
466 in terms of topics and the type of question. In [50] and [51] smaller question collections (72 and
467 12) were subjected to detailed semantic, attitudinal or linguistic analysis. An increasing number
468 of studies concern the development of question–answering technology for consumer health
469 questions [83-87]. Patients have different information needs about T2DM at different points as
470 their disease progresses. However little is known about these needs and how they change over
471 time or across varying health or life circumstances [77] — even though there has been a
472 significant amount of research on what the different phases of T2DM are [58-67]. It is in cancer
473 care that the needs of patients at different stages of their disease have been most thoroughly
474 studied [88-~~90~~91]. These studies, show, for instance, that while most (91%) female breast-cancer
475 patients wanted to know their prognosis before beginning adjuvant treatment [91], after the first

476 consultation, their needs often shifted to matters of support, with 59-63% primarily wanting
477 reassurance and hope and patients with advanced disease often desiring less information about
478 their illness [88]. It is important that we develop a similar understanding of the changing needs
479 of people with T2DM.

480 **Further research**

481 The urgent need for resources allowing patients with T2DM to find answers to their questions
482 has recently been documented [52]. One longer-term goal of this study is to develop a question-
483 answer system, informed by the analysis of a very large number of questions and vetted answers
484 and based on the automated identification of topics in questions. The twenty-three categories we
485 devised for this study will almost certainly need further refinement, with a hierarchy of topics or
486 an ontology possibly providing a better representation. In addition, *answer* topics as well as
487 question topics need to be defined. For example, suppose a patient's question is "I'm 44 and
488 recently diagnosed with Type 2 diabetes, and now I am having difficulty reading fine print. Is
489 this related to my diabetes?" This question falls into four possible answer categories. The first
490 relates to temporary changes in eyesight when blood glucose fluctuates. The second concerns a
491 side effect of the drug pioglitazone. The third is about diabetic retinopathy that leads to
492 blindness. And the fourth concerns normal age-related changes in eyesight.

493 Finer-grained characteristics that are important in the management of diabetes are also
494 needed. For example, the capacity of a person to act in any given environment (known as
495 agency) seems to be expressed differentially in our questions [92]. The following question about
496 a cure for diabetes appears to locate agency within the patient: "What stuff do you have to do to
497 cure diabetes?" This is in contrast to a question that appears to locate agency within the broader
498 society: "How close is science to finding a cure for diabetes?" If patients over time asked

499 questions that differed in the location of agency, that would be of interest and possible clinical
500 significance. In our follow-up studies when new questions are collected from patients, we will be
501 labeling each question by the stage the questioner is in relative to his or her own disease
502 progression. In this way a record of the questions asked in the aggregate by patients at each
503 phase of the disease can be compiled along with the progression of questions for each patient
504 individually, providing a broader and deeper perspective on the complex needs of those affected
505 by or at risk of T2DM.

506

507 **References**

- 508 1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled
509 analysis of 751 population-based studies with 4.4 million participants. *Lancet*.
510 2016;387(10027):1513-1530. doi:10.1016/S0140-6736(16)00618-8.
- 511 2. Zimmet P, Alberti K, Shaw J. Global and societal implications of the diabetes epidemic. *Nature*.
512 2001;414(6865):782-787. doi:10.1038/414782a.
- 513 3. Danaei G, Finucane M, Lu Y, Singh G, Cowan M, Paciorek C. National, regional, and global trends
514 in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health
515 examination surveys and epidemiological studies with 370 country-years and 2.7 million
516 participants. *Lancet*. 2011;378(9785):31-40. doi:10.1016/S0140-6736(11)60679-X.
- 517 4. Gregg E, Li Y, Wang J, Burrows N, Ali M, Rolka D, et al. Changes in diabetes-related
518 complications in the United States, 1990-2010. *New England Journal of Medicine*. 2014;370:1514-
519 1523. doi:10.1056/NEJMoa1310799.
- 520 5. Whitford D, Paul G, Smith S. Patient generated “frequently asked questions”: Identifying
521 informational needs in a RCT of peer support in type 2 diabetes. *Primary Care Diabetes*.
522 2013;7(2):103-109. doi:10.1016/j.pcd.2013.01.006.

- 523 6. Browne J, Scibilia R, Speight J. The needs, concerns, and characteristics of younger Australian
524 adults with Type 2 diabetes. *Diabetic Medicine*. 2013;30(5):620-626. doi:10.1111/dme.12078.
- 525 7. Weymann N, Harter M, Dirmaier J. Information and decision support needs in patients with type 2
526 diabetes. *Health Informatics Journal*. 2014;22(1):46-59. doi:10.1177/1460458214534090.
- 527 8. Barnes P, Kelly C, Connelly K, Siek K. Understanding the needs of low SES patients with type 2
528 diabetes. 7th International Conference on Pervasive Computing Technologies for Healthcare and
529 Workshops; Venice, Italy 2013.
- 530 9. Rubin R, Peyrot M, Siminerio L. Health care and patient-reported outcomes: results of the cross-
531 national Diabetes Attitudes, Wishes and Needs (DAWN) study. *Diabetes Care*. 2006;29(6):1249-
532 1255. doi:10.2337/dc05-2494.
- 533 10. Bootle S, Skovlund S. Proceedings of the 5th International DAWN Summit 2014: Acting together
534 to make person-centred diabetes care a reality. *Diabetes Research and Clinical Practice*.
535 2015;109(1):6-18. doi:10.1016/j.diabres.2015.04.004.
- 536 11. Atkinson S, Rubinelli S. Narrative in cancer research and policy: voice, knowledge and context.
537 *Critical Reviews in Oncology/Hematology*. 2012;84(2):S11-S16. doi:10.1016/S1040-
538 8428(13)70004-0.
- 539 12. Longo D, Schubert S, Wright B, LeMaster J, Williams C, Clore J. Health Information Seeking,
540 Receipt, and Use in Diabetes Self-Management. *The Annals of Family Medicine*. 2010;8(4):334-
541 340. doi:10.1370/afm.1115.
- 542 13. Menke A, Casagrande S, Geiss L, Cowie C. Prevalence of and Trends in Diabetes Among Adults
543 in the United States, 1988-2012. *JAMA* 2015;314(10):1021. doi:10.1001/jama.2015.10029.
- 544 14. Questions To Ask Your Doctor. Content last reviewed May 2018. Agency for Healthcare Research
545 and Quality, Rockville, MD. [http://www.ahrq.gov/patients-consumers/patient-involvement/ask-](http://www.ahrq.gov/patients-consumers/patient-involvement/ask-your-doctor/index.html)
546 [your-doctor/index.html](http://www.ahrq.gov/patients-consumers/patient-involvement/ask-your-doctor/index.html). Accessed August 16, 2018
- 547 15. Powers MA, Bardsley J, Cypress M, et al.: Diabetes self-management education and support in
548 type 2 diabetes: a joint position statement of the American Diabetes Association, the American

- 549 Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. *Diabetes*
550 *Care* 2015; 38: 1372-1382.
- 551 16. Crangle C, Brothers Kart J. A questions-based investigation of consumer mental-health
552 information. *PeerJ* 2015;3:e867. doi:10.7717/peerj.867.
- 553 17. [Sleath B, Sayner R, Blalock SJ, Carpenter DM, Muir KW, Hartnett ME, Sleath B](#) et al.
554 Patient question-asking about glaucoma and glaucoma medications during videotaped medical
555 visits. *Health Commun.* 2015;30(7):660-8. doi: 10.1080/10410236.2014.888387. Epub 2014 Jul 25.
556 DOI:10.1080/10410236.2014.888387
- 557 18. Stuckey H, Olmsted T, Mincemoyer R, Gabbay R. What Do People with Diabetes Talk about on a
558 Diabetes Social Networking Web Site? . *Journal of Diabetes Science and Technology.*
559 2012;6(3):716-717.
- 560 19. Hilliard M, Sparling K, Hitchcock J, Oser T, Hood K. The Emerging Diabetes Online Community.
561 *Current Diabetes Reviews.* 2015;11(4):261-272.
- 562 20. Mann D. Diabetes Risk Communication Tool To Improve Lifestyle Behaviors. 2013.
563 <http://grantome.com/grant/NIH/K23-DK081665-03>. Accessed August 16, 2018
- 564 21. Kessels R. Patients' memory for medical information. *Journal of the Royal Society of Medicine.*
565 2003;96(5):219-222.
- 566 22. Selic P, Svab I, Repolusk M, Gucek N. What factors affect patients' recall of general practitioners'
567 advice? *BMC Family Practice.* 2011;12:141. doi:10.1186/1471-2296-12-141.
- 568 23. Patocka C, Lin M, Voros J, Chan T. Point-of-care Resource Use in the Emergency Department: A
569 Developmental Model. *AEM Educ Train.* 2018 May 30;2(3):221-228. doi: 10.1002/aet2.10101.
570 eCollection 2018 Jul.
- 571 24. Ely J, Osheroff J, Gorman P, Ebell M, Chambliss M, Pifer E, et al. A taxonomy of generic clinical
572 questions: classification study. *BMJ.* 2000;321(7258):429-432. doi:10.1136/bmj.321.7258.429.
- 573 25. Ely J, Osheroff J, Ebell M, Bergus G, Levy B, Chambliss M, et al. Analysis of questions asked by
574 family physicians regarding consumer care. *Western Journal of Medicine.* 2000;172(5):315-319.

- 575 26. Jerome R, Giuse N, Gish K, Sathe N, Dietrich M. Information needs of clinical teams: analysis of
576 questions received by the Clinical Informatics Consult Service. *Bulletin of the Medical Library*
577 *Association*. 2001;89(2):177-184.
- 578 27. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for
579 clinical questions. *AMIA Annual Symposium Proceedings*. 2006:359-363.
- 580 28. Chase H, Kaufman D, Johnson S, Mendonca E. Voice capture of medical residents' clinical
581 information needs during an inpatient rotation. *Journal of the American Medical Informatics*
582 *Association* 2009;16(3):387-394. doi:10.1197/jamia.M2940.
- 583 29. Hung P, Johnson S, Kaufman D, Mendonça E. A multi-level model of information seeking in the
584 clinical domain. *Journal of Biomedical Informatics* 2008;41(2):357-370.
585 doi:10.1016/j.jbi.2007.09.005.
- 586 30. Del Fiol G, Workman T, Gorman P. Clinical Questions Raised by Clinicians at the Point of Care: A
587 Systematic Review. *American Medical Association's Journal of Internal Medicine*.
588 2014;174(5):710-718. doi:10.1001/jamainternmed.2014.368.
- 589 31. Hübner-Bloder G, Duftschmid G, Kohler M, Rinner C, Saboor S, Ammenwerth E. Clinical
590 situations and information needs of physicians during treatment of diabetes mellitus patients: a
591 triangulation study. *Studies in Health Technology and Informatics*. 2011;169:369-373.
- 592 32. Dorsch J. Information needs of rural health professionals: a review of the literature. *Bulletin of the*
593 *Medical Library Association* 2000;88(4):346-354.
- 594 33. Hersh W, Crabtree M, Hickam D, Sacherek L, Rose L, Friedman C. Factors associated with
595 successful answering of clinical questions using an information retrieval system. *Bulletin of the*
596 *Medical Library Association*. 2000;88(4):323-331.
- 597 34. Green M, Ciampi M, Ellis P. Residents' medical information needs in clinic: are they being met?
598 *American Journal of Medicine* 2000;109(3):218-223.
- 599 35. Bergus G, Randall C, Sinift S, Rosenthal D. Does the structure of clinical questions affect the
600 outcome of curbside consultations with specialty colleagues? *Archives of Family Medicine*

- 601 2000;9(6):541-547.
- 602 36. Florance V. Medical knowledge for clinical problem solving: a structural analysis of clinical
603 questions. *Bulletin of the Medical Library Association*. 1992;80(2):140-149.
- 604 37. Osheroff J, Forsythe D, Buchanan B, Bankowitz R, Blumenfeld B, Miller R. Physicians'
605 information needs: analysis of questions posed during clinical teaching. *Annals of Internal
606 Medicine* 1991;114(7):576-581.
- 607 38. Cimino C, Barnett G. Analysis of physician questions in an ambulatory care setting. *Proceedings of
608 the Annual Symposium on Computer Application in Medical Care*. 1991:995-999.
- 609 39. Crangle C, Carlson R, Fagan L, Davis A, Erlbaum M, Keck K, et al. Conversational access to on-
610 line cancer information: An adaptable speech interface. *Proceedings of the AMIA (American
611 Medical Informatics Association) Fall Symposium: Beyond the Superhighway: Exploiting the
612 Internet with Medical Informatics; 1996 October 26-30; Washington D.C., Philadelphia, USA*.
- 613 40. Crangle C, Carlson R, Fagan L, Erlbaum M, Sherertz D, Tuttle M. Conversational access to on-line
614 medical information. *15th Annual Conference of the American Voice Input/Output Society
615 (AVIOS); 1996 September 10-12; San Jose, California, USA*.
- 616 41. Crangle C, Carlson R, Fagan L, Erlbaum M, Olson N, Sherertz D, et al. Meeting clinical
617 information needs: An interface that provides uniform access across diverse information sources.
618 In: Press A, editor. *AAAI-96 Spring Symposium on Artificial Intelligence in Medicine; March 25-
619 27; Stanford University, California, USA 1996*. p. 21-25.
- 620 42. Sherertz D, Tuttle M, Olson N, Hsu G, Carlson R, Fagan L, et al. Accessing oncology information
621 at the point of care: experience using speech, pen, and 3-D interfaces with a knowledge server.
622 *Medinfo*. 1995;8(1):792-795.
- 623 43. Tuttle M, Sherertz D, Olson N, Nelson S, Erlbaum M, Keck K. Toward reusable software
624 components at the point of care. *Proceedings of the AMIA Fall Symposium; 1996*.
- 625 44. Del Fiol G, Weber A, Bruncker C, Weir C. Clinical questions raised by providers in the care of older
626 adults: a prospective observational study. *BMJ Open*. 2014;4(7):e005315. doi:10.1136/bmjopen-

- 627 2014-005315.
- 628 45. Zhang Y. Contextualizing Consumer Health Information Searching: an Analysis of Questions in a
629 Social Q&A Community. ACM International Health Informatics Symposium; 2010.
- 630 46. Roberts K, Masterton K, Fisman M, Kilicoglu H, Demner-Fushman D. Annotating Question
631 Types for Consumer Health Questions. LREC 2014.
- 632 47. Cui L, Tao S, Zhang G-Q. A Semantic-based Approach for Exploring Consumer Health Questions
633 Using UMLS. AMIA Annual Symposium Proceedings. 2014:432-441.
- 634 48. Roberts K, Demner-Fushman D. Interactive use of online health resources: a comparison of
635 consumer and professional questions. Journal of the American Medical Informatics Association
636 2016;23(4):802-811. doi:10.1093/jamia/ocw024.
- 637 49. White M. Questioning Behavior on a Consumer Health Electronic List. The Library Quarterly.
638 2000;70(3):302-334.
- 639 50. Oh J, He D, Jeng W. Linguistic characteristics of eating disorder questions on Yahoo! Answers:
640 content, style, and emotion. Proceedings of the American Society for Information Science and
641 Technology. 2013;50:1-10.
- 642 51. Slaughter L, Soergel D, Rindfleisch T. Semantic representation of consumer questions and physician
643 answers. International Journal of Medical Informatics. 2006;75:513-529.
644 doi:10.1016/j.ijmedinf.2005.07.025.
- 645 52. Crangle CE, Bradley C, Carlin PF, Esterhay RJ, Harper R, Kearney PM, ~~Lorig K, McCarthy VJC,~~
646 ~~McTear M, Savage E, Tuttle MS, Wallace JG-et al.~~ (2017) Soliciting and Responding to Patients'
647 Questions about Diabetes through Online Sources. Diabetes Technology & Therapeutics, Volume
648 19, Number 3, 2017. DOI: 10.1089/dia.2016.0291
- 649 53. Swan M. Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical
650 Trials in the Public Health Research Ecosystem. Journal of Medical Internet Research.
651 2012;14(2):e46. doi:10.2196/jmir.1988.
- 652 54. Litman L, Robinson J, Abberbock T. TurkPrime.com: A versatile crowdsourcing data acquisition

- 653 platform for the behavioral sciences. *Behavior Research Methods*. 2016. doi:10.3758/s13428-016-
654 0727-z.
- 655 55. Peyrot M, Burns K, Davies M, Forbes A, Hermanns N, Holt R, et al. Diabetes Attitudes Wishes and
656 Needs 2 (DAWN2): a multinational, multi-stakeholder study of psychosocial issues in diabetes and
657 person-centred diabetes care. *Diabetes Research and Clinical Practice*. 2013;99(2):174-184.
658 doi:10.1016/j.diabres.2012.11.016.
- 659 56. Skovlund S, Peyrot M. The diabetes attitudes, wishes and needs (DAWN) program: A new
660 approach to improving outcomes of diabetes care. *Diabetes Spectrum* 2005;18(3):136-142.
- 661 57. Nicolucci A, Kovacs Burns K, Holt R, Comaschi M, Hermanns N, Ishii H, et al. Diabetes Attitudes,
662 Wishes and Needs second study (DAWN2): cross-national benchmarking of diabetes-related
663 psychosocial outcomes for people with diabetes. *Diabetic Medicine*. 2013;30:767-777.
664 doi:10.1111/dme.12245
- 665 58. Weinger K, Welch G, Jacobson A. Psychological and psychiatric issues in diabetes mellitus. In:
666 Poretsky L, editor. *Principles of Diabetes Mellitus*. Norwell, Massachusetts: Kluwer Academic
667 Publishers; 2002. p. 639-654.
- 668 59. Welch G, Jacobson A, Polonsky W. The Problem Areas in Diabetes Scale: An evaluation of its
669 clinical utility. *Diabetes Care* 1997;20:760-766. doi:10.2337/diacare.20.5.760.
- 670 60. Gonzalez J, Safren S, Delahanty L, Cagliero E, Wexler D, Meigs J, et al. Symptoms of depression
671 prospectively predict poorer self-care in patients with Type 2 diabetes. *Diabetic Medicine*.
672 2008;25:1102-1107. doi:10.1111/j.1464-5491.2008.02535.x.
- 673 61. Walker R, Smalls B, Hernandez-Tejada M, Campbell J, Davis K, Egede L. Effect of diabetes
674 fatalism on medication adherence and self-care behaviors in adults with diabetes. *General Hospital
675 Psychiatry*. 2012;34:598-603. doi:10.1016/j.genhosppsy.2012.07.005.
- 676 62. Yi J, Vitaliano P, Smith R, Yi J, Weinger K. The role of resilience on psychological adjustment and
677 physical health in patients with diabetes. *British Journal of Health Psychology* 2008;13:311-325.
678 doi:10.1348/135910707X186994.

- 679 63. Yi J, Yi J, Vitaliano P, Weinger K. How does anger coping style affect glycemic control in diabetes
680 patients? . International Journal of Behavioral Medicine 2008;15:167-172.
681 doi:10.1080/10705500802219481.
- 682 64. Yi-Frazier J, Smith R, Vitaliano P, Yi J, Mai S, Hillman M, et al. A Person-Focused analysis of
683 resilience resources and coping in diabetes patients. Stress Health. 2010;26:51-60.
- 684 65. Rucker J, McDowd J, Kluding P. Executive function and type 2 diabetes: putting the pieces
685 together. Physical Therapy 2012;92:454-462. doi:10.2522/ptj.20100397.
- 686 66. Schiotz M, Bogelund M, Almdal T, Jensen B, Willaing I. Social support and self-management
687 behaviour among patients with Type 2 diabetes. Diabetic Medicine 2012;29(654-661).
688 doi:10.1111/j.1464-5491.2011.03485.x.
- 689 67. Diapedia. Coping with diabetes in adults. Diapedia 2016. doi:10.14496/dia.61047161185.10.
- 690 68. Warrens MJ (2015) Five Ways to Look at Cohen’s Kappa. J Psychol Psychother
691 5:197.doi:10.4172/2161-0487.100019759.
- 692 69. Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ~~Marquéz, J. R. G., Gruber,~~
693 ~~B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B.,~~
694 ~~Schröder, B., Skidmore, A. K., Zurell, D. and Lautenbach, S et al.~~ 2013, Collinearity: a review of
695 methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36: 27–46.
696 doi:10.1111/j.1600-0587.2012.07348.x
- 697 70. Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal*
698 *Statistical Society (Series B)*, 58, 267-288.
- 699 71. Efron, Bradley; Hastie, Trevor; Johnstone, Iain; Tibshirani, Robert. Least angle regression. *Ann.*
700 *Statist.* 32 (2004), no. 2, 407--499. doi:10.1214/009053604000000067.
701 <https://projecteuclid.org/euclid.aos/108317893562>.
- 702 72. Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, *JMLR* 12, pp. 2825-2830, 2011.
- 703 73. Benjamini, ~~Yoav;~~ Hochberg, ~~Yoel~~ (1995). Controlling the false discovery rate: a practical and
704 powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B.* 57 (1):

- 705 | 289–300.
- 706 | 74. Stevens JR, Al Masud A, Suyundikov A. (2017) A comparison of multiple testing adjustment
707 | methods with block-correlation positively-dependent tests. Zaykin D, ed. PLoS ONE.
708 | 2017;12(4):e017
- 709 | 75. Benjamini, Yoav; Yekutieli, Daniel. The control of the false discovery rate in multiple testing under
710 | dependency. Ann. Statist. 29 (2001), no. 4, 1165--1188. Yoav Benjamini Y, Yekutieli D.(2001) The
711 | control of the false discovery rate in multiple testing under dependency. Ann. Statist. Volume 29,
712 | Number 4 (2001), 1165-1188,doi:10.1214/aos/1013699998.
713 | <https://projecteuclid.org/euclid.aos/1013699998>
- 714 | 76. Rochon J, Gondan M, Kieser M. To test or not to test: Preliminary assessment of normality when
715 | comparing two independent samples. BMC Medical Research Methodology. 2012;12:81.
716 | doi:10.1186/1471-2288-12-81.
- 717 | 77. Zare-Farashbandi F, Lalazaryan A, Rahimi A, Zadeh A. How health information is received by
718 | diabetic patients? Advanced Biomedical Research 2015;4:126. doi:10.4103/2277-9175.158047.
- 719 | 78. ~~William P-Glasheen, Andrew Renda A, Yanting-Dong Y,~~ Diabetes Complications Severity Index
720 | (DCSI)—Update and ICD-10 translation, Journal of Diabetes and its Complications, Volume 31,
721 | Issue 6, 2017, Pages 1007-1013, ISSN 1056-8727, <https://doi.org/10.1016/j.jdiacomp.2017.02.018>.
- 722 | 79. Peyrot M, McMurry JJ, Kruger D. A biopsychosocial model of glycemic control in diabetes: stress,
723 | coping and regimen adherence. Journal of Health and Social Behavior. 1999;40:141-158.
- 724 | 80. Pian W, Khoo C, Chang Y-K, Moorhead A. The Criteria People Use in Relevance Decisions on
725 | Health Information: An Analysis of User Eye Movements When Browsing a Health Discussion
726 | Forum. Journal of Medical Internet Research. 2016;18(6):e136. doi:10.2196/jmir.5513.
- 727 | 81. Liburd L, Namageyo-Funa A, Jack LJ. Understanding "masculinity" and the challenges of
728 | managing type-2 diabetes among African-American men. Journal of the National Medical
- 729 | 82. Peters S, Huxley R, Sattar N, Woodward M. Sex Differences in the Excess Risk of Cardiovascular
730 | Diseases Associated with Type 2 Diabetes: Potential Explanations and Clinical Implications.

- 731 Current Cardiovascular Risk Reports 2015;9(7):36. doi:10.1007/s12170-015-0462-5.
- 732 83. Bickmore T, Utami D, Matsuyama R, Paasche-Orlow M. Improving Access to Online Health
733 Information With Conversational Agents: A Randomized Controlled Experiment. *Journal of*
734 *Medical Internet Research*. 2016;18(1):e1. doi:10.2196/jmir.5239.
- 735 84. Bates M. Subject access in online catalogs: a design model. *Journal of the American Society for*
736 *Information Science*. 1986;37:357-376.
- 737 85. Anderson J, Perez-Carballo J. The nature of indexing: how humans and machines analyze messages
738 and texts for retrieval, Part I: Research, and the nature of human indexing. *Information Processing*
739 *and Management* 2001;37:231-254.
- 740 86. Van Der Volgen J, Harris B, Demner-Fushman D. Analysis of consumer health questions for
741 development of question–answering technology. In: Mitchell N, editor. *Proceedings of one*
742 *HEALTH: Information in an interdependent world, the 2013 annual meeting and exhibition of the*
743 *Medical Library Association (MLA)*. Chicago 2013.
- 744 87. Cronin R, Fabbri D, Denny J, Jackson G. Automated Classification of Consumer Health
745 Information Needs in Patient Portal Messages. *AMIA Annual Symposium Proceedings*. 2015:1861-
746 1870.
- 747 88. Butow P, Maclean M, Dunn S, Tattersall M, Boyer M. The dynamics of change: cancer patients'
748 preferences for information, involvement and support. *Annals of Oncology*. 1997;8(9):857-863.
- 749 89. Mills M, Sullivan K. The importance of information giving for patients newly diagnosed with
750 cancer: a review of the literature. *Journal of Clinical Nursing*. 1999;8(6):631-642.
- 751 90. Siminoff L, Ravdin P, Colabianchi N, Sturm C. Doctor-patient communication patterns in breast
752 cancer adjuvant therapy discussions. *Health Expect*. 2000;3(1):26-36.
- 753 91. Lobb E, Butow P, Meiser B, Barratt A, Gaff C, Young M, et al. Tailoring communication in
754 consultations with women from high risk breast cancer families. *British Journal of Cancer*.
755 2002;87(5):502-508. doi:10.1038/sj.bjc.6600484.
- 756 92. Armstrong, David. Actors, patients and agency: a recent history. *Sociology of Health & Illness* Vol.

757 36 No. 2 2014 ISSN 0141-9889, pp. 163–174 doi: 10.1111/1467-9566.12100

758

759 **Supporting information**

760 **S1 Table. Questions about Cure or Reversal.**

761 **S1 File. Clinic Questions.**

762 **S2 File. Crowdsourced Questions.**

763