

1 **Resolving embarrassing medical conditions with online health information**

2 ***Sarah Redston^a, Sharon de Botte^a, Carl Smith^{b,c,*}***

3 *^aHealth Press Ltd, Elizabeth House, Queen Street, Abingdon, UK*

4 *^bSchool of Biology, University of St Andrews, St Andrews, UK*

5 *^cDepartment of Ecology & Vertebrate Zoology, University of Łódź, Łódź, Poland*

6 **Corresponding author. Carl Smith, School of Biology, Bute Building, University of St.*

7 *Andrews, St. Andrews, Fife KY16 9TS, UK. email: cs101@st-andrews.ac.uk*

8

9 ABSTRACT

10 *Purpose:* Reliance on online health information is proliferating and the Internet has the
11 potential to revolutionize the provision of public health information. The anonymity of
12 online health information may be particularly appealing to people seeking advice on
13 'embarrassing' health problems. The purpose of this study was to investigate (1)
14 whether data generated by the *embarrassingproblems.com* health information site
15 showed any temporal patterns in problem resolution, and (2) whether successful
16 resolution of a medical problem using online information varied with the type of
17 medical problem.

18 *Methods:* We analyzed the responses of visitors to the *embarrassingproblems.com*
19 website on the resolution of their problems. The dataset comprised 100,561 responses
20 to information provided on 77 different embarrassing problems grouped into 9 classes
21 of medical problem over an 82-month period. Data were analyzed with a Bernoulli
22 Generalized Linear Model using Bayesian inference.

23 *Results:* We detected a statistically important interaction between embarrassing
24 problem type and the time period in which data were collected, with an improvement in
25 problem resolution over time for all of the classes of medical problem on the website but
26 with a lower rate of increase in resolution for urinary health problems and medical
27 problems associated with the mouth and face. As far as we are aware, this is the first
28 analysis of data of this nature.

29 *Conclusions:* Findings support the growing recognition that online health information
30 can contribute to the resolution of embarrassing medical problems, but demonstrate
31 that outcomes may vary with medical problem type. The results indicate that building
32 data collection into online information provision can help to refine and focus health
33 information for online users.

34 *Keywords:* Internet; Health communication; Online information; Quantitative results

35 **1. Introduction**

36 The Internet increasingly serves as a secondary, and even primary, source of
37 health and medical information to the public [1-3]. The role that online health
38 information will come to play in the future is unclear, but it is a potentially important
39 source of information for promoting good public health [3], and may come to have as
40 great an impact as that seen in banking, education, leisure and social relationships [2].
41 The advantages of online health information are that it is almost limitless in scope,
42 unregulated, accessible from anywhere with an internet connection, available at any
43 time, potentially interactive and typically free to access, as well as having the potential
44 to combine expert advice with user-generated experience [4]. The anonymity of online
45 health information may be particularly appealing to people seeking information on
46 'embarrassing' health problems, for example those associated with sexual or mental
47 health [5,6] or body image issues, but potentially a wide range of medical conditions [7].
48 Some of the drawbacks of online health information are the potential for
49 misinformation, social isolation of users and the undermining of, and distancing from,
50 health professionals [4,8].

51 While the proliferation and impact of online health information is clearly
52 significant, its success in resolving health problems is more equivocal. Thus, while
53 variables such as the demographic profile, health status, educational attainment and
54 familiarity with social media of online users of health information have been well
55 characterized [3,9,10], no attempt has been made to quantify the outcome of seeking
56 health information online.

57 In this study, we analyzed data gathered over an 82-month period from
58 *embarrassingproblems.com*, a website specifically designed to provide information on
59 health problems that are often perceived as 'embarrassing' or difficult to discuss. During
60 the study period the website received up to 1.2 million hits each month, providing

61 information on 77 discrete medical problems. At the bottom of each problem page on
62 the website, users are asked to anonymously record whether the information provided
63 has enabled them to tackle the given problem or whether it remains unresolved. Our
64 goal was to investigate whether the data generated by the site shows any temporal
65 patterns in problem resolution, and whether successful outcome varies with the type of
66 medical problem.

67

68 **2. Methods**

69 *2.1 Background*

70 Established in January 2000 by the healthcare publisher Health Press Ltd, the
71 *embarrassingproblems.com* online health site (<http://www.embarrassingproblems.com>)
72 provides expert information on personal health that might be considered difficult to
73 discuss. The service is free to access. The information on the website is provided by
74 recognized medical practitioners who write for Health Press Ltd. Individual health
75 problems are grouped into nine broad classes of medical problem (Table 1).

76 Since February 2010, website content users have been invited to check one of
77 three boxes at the bottom of each page to indicate whether the information provided by
78 the site has enabled them to tackle the problem (“Tackled it, moved on”), or whether they
79 feel the problem remains unresolved (“Tackling it” or ‘Still struggling’) (Fig. 1).
80 Responses were collected over an 82-month period, from February 2010 to November
81 2016. Responses were anonymous, but were problem specific. In addition, while the
82 temporal order of responses was known, the exact date of response was not recorded on
83 the site. Thus, temporal patterns in the data could be examined statistically, but could
84 not be linked to specific dates.

85

86 *2.2 Data analysis*

87 Online responses were treated as binomial data with problems scored as having
88 been tackled or not tackled. Data for responses to specific medical problems were highly
89 unbalanced and were subsequently analyzed by class of problem [11] (Table 1). After
90 grouping medical problems, one class of problem ('cognitive') still showed imbalance
91 and was subsequently dropped from the analysis.

92 Data were modeled using a Bernoulli Generalized Linear Model (GLM), which
93 took the form:

94

$$95 \text{ Outcome}_i \sim \text{Binomial}(\pi_i)$$

$$96 E(\text{Outcome}_i) = \pi_i$$

$$97 \eta_i = \beta + \text{Problem}_i * \text{Time}_i$$

$$98 \text{logit}(\pi_i) = \eta_i$$

99

100 Outcome_i is the probability of a positive outcome (i.e. problem tackled) for respondent i
101 assuming a Bernoulli distribution with mean π_i and variance $\pi_i \times (1 - \pi_i)$. Problem_i is the
102 class of medical problem (Table 1) experienced by respondent i . Time_i is the time period
103 in which respondent i submitted their online response. Time periods comprised 13
104 discrete ordinal categories, broadly taken to represent 6-month intervals over the 82
105 months of data collection. While the temporal order of these data is reliable, the precise
106 timing of responses reflected variation in site traffic and was not interpreted as
107 representing discrete Julian time periods.

108 To make inferences about the parameters in the model, a Bayesian approach was
109 used. A Bayesian GLM is robust in dealing with complex datasets, unbalanced data, an
110 inherent lack of dependency due to repeated measures, and a highly varied non-normal
111 response variable. Bayesian models are flexible in allowing the estimation of a posterior
112 distribution of differences between parameters and across levels of factors. These are

113 relatively straightforward procedures using Bayesian inference, but problematic in a
114 frequentist framework [12,13], notwithstanding more general reservations in using
115 frequentist analyses [12-15].

116 Diffuse or non-informative univariate priors were put on all parameters. The
117 model was fitted in a Bayesian framework using Markov Chain Monte Carlo (MCMC)
118 with the *R2jags* package [16] in the R statistical environment [17]. Three independent
119 Markov chains were run simultaneously with a burn-in of 50,000 iterations and then
120 500,000 iterations for estimates of parameter and 95% credibility intervals. Chains
121 were thinned every 10th iteration, resulting in 50,000 Markov Chain samples for each
122 estimated parameter. Mixing and autocorrelation of chains were checked visually using
123 trace plots and the Gelman-Rubin statistic [15]. Autocorrelation was low and good
124 mixing was achieved in each case. The Gelman-Rubin statistic was estimated to be less
125 than 1.002 in all cases, indicating good convergence. Model validation showed no
126 evidence of overdispersion, heterogeneity or non-linear patterns in the model residuals
127 [18]. As part of the model-fitting process, the model was used to simulate an alternative
128 dataset. This procedure allowed the fitted values to be compared with the simulated
129 data, with probability values for each data point used to assess model fit. A probability of
130 0.49 indicated the model complied closely with the data [12]. All data from this article
131 will be made available in the Dryad Digital Repository.

132

133 **3. Results**

134 Overall there was a statistically important increase in the probability of problem
135 resolution among time periods across all classes of embarrassing medical problem
136 (Table 2, Fig. 2). In addition, there was a significant interaction of medical problem with
137 time period. There was a lower increase in the resolution of problems classed as
138 'urinary' and 'mouth and face' compared with the baseline class of problem ('anal')

139 (Table 2, Fig. 2). Improvement in resolution of problems classed as ‘breast and nipple’,
140 ‘gut’, ‘hair’, ‘hands, legs and feet’, ‘sex and genital’ and ‘skin’ did not differ from the
141 improvement seen with baseline (Table 2, Fig. 2).

142

143 **4. Discussion**

144 The proliferation of online health information presents an opportunity to
145 distribute reliable, authoritative public health information. The anonymity that online
146 information allows may be especially useful in disseminating advice on health problems
147 perceived to be embarrassing. Using data generated by the online health information
148 site *embarrassingproblems.com* we showed an improvement over time in the successful
149 resolution of problems for users and an interaction between embarrassing problem type
150 with time. ‘Urinary’ health problems and problems associated with the ‘mouth and face’
151 demonstrated less improvement over time than the baseline comparator class of ‘anal’
152 problems. These results provide supporting evidence for the efficacy of online health
153 information, but demonstrate that outcomes may vary with medical problem type.

154 In the present study, a temporal improvement in problem resolution was
155 observed across all medical problem types. While the reason for this trend is opaque,
156 this finding ostensibly implies that the impact of this online health information
157 improved over time. Over the 82-month time period of this study, the
158 *embarrassingproblems.com* site editors implemented a ‘listen and respond’ approach to
159 content provision. All user-generated comments were manually moderated before being
160 published on each problem page below the expert advice. The site’s editors then
161 ensured that these comments were taken into consideration in content updates made by
162 expert authors and in other materials highlighted on the site. Although the extent to
163 which user-generated content has contributed to the resolution of any given health

164 problem has not been ascertained, its use in continually improving the editorial offering
165 on the site is likely to have had a positive effect on visitor engagement.

166 It is also possible that the improvement in problem resolution reflects a general
167 increase in confidence in online health information, particularly from sources that are
168 perceived as reliable and trustworthy. The emergence of the internet as a source of
169 health information is increasingly recognized [2]. Furthermore, the
170 *embarrassingproblems.com* site has had a strong authoritative presence online for 16
171 years, and the expert authorship of the information is clearly signposted. However,
172 without further data collection from the site, the underlying reason for this positive
173 trend is unclear.

174 Notably, this trend mirrors that in mainstream healthcare, where information
175 technology plays an increasing role in supporting healthcare professionals [19,20]. The
176 role of online technology in other aspects of life has been profound. The rapid growth in
177 online shopping has been attributed to the expansion in the types of goods available,
178 combined with the trust in the security and reliability of service [21]. Growth in online
179 banking, education, leisure and social relationships has similarly expanded rapidly. In
180 this context, the growth in self-directed access to online health information, and
181 increasing trust in its content, is not unexpected.

182 The interaction of temporal changes in health problem resolution with problem
183 type implies variation in the tractability of healthcare problems, but also that successful
184 resolution of specific problem types has increased faster than others. Given the wide
185 variety of health problems addressed by the *embarrassingproblems.com* site, it is not
186 unexpected that its success in tackling different problems might vary among problem
187 types. For example, the successful resolution of piles, crab lice, and chronic diarrhea are
188 high in our dataset, reflecting the relative ease with which these problems can be tackled
189 with readily available proprietary medical products. Others, such as profuse sweating,

190 blushing and flushing, and bed-wetting in adults are medically less tractable problems
191 and, consequently, show much lower rates of resolution. In our analysis, urinary
192 problems and medical problems associated with the mouth and face showed the
193 weakest improvement over time. While these findings are unambiguous in our data, an
194 explanation for the observed variance in the temporal improvement of different
195 problem types is opaque and will require further data collection.

196 The relationship between online medical information and professional medical
197 providers is a potentially contentious one. It has been proposed that online advice may
198 undermine the work of healthcare professionals, though evidence for this is equivocal
199 [4,8]. It is also unclear whether online consumers of medical information are more or
200 less likely to seek medical advice from healthcare professionals, though some evidence
201 indicates that online information is sought in combination with more conventional
202 medical support [1,22]. Indeed, since its inception, the *embarrassingproblems.com* site
203 has encouraged content users to seek professional medical advice when appropriate in
204 conjunction with the information it provides. Thus, the role of online medical
205 information appears to be in tackling occasional health problems [1], with the
206 personalized nature of online medical health information contrasting with more diffuse
207 general health information, but complementing that from medical professionals
208 [1,22,23].

209 A distinction must also be made between information provision and active
210 engagement in online discussion of health. Hitherto, studies of online health information
211 have tended to conflate these two, yet they clearly differ. How forum content is used and
212 how reliable and trusted information from this source is in comparison with more
213 conventional sources is not clear.

214 The growth of online medical information has enormous potential to inform and
215 empower patients and thereby facilitate improvements in healthcare outcomes [24]. In

216 the context of the present study, online health information that addresses problems that
217 are perceived as embarrassing may be particularly effective, and the provision of online
218 health information more generally may benefit from refinement and focus. Young adults
219 are a well-represented demographic group that readily uses online health information
220 [22], as are well-educated high-income women [10,25]. Online health information that
221 targets these groups is evidently in demand, though the needs of online users of
222 different racial, ethnic, and socioeconomic backgrounds is clearly also necessary.

223 Ultimately it is the degree of trust users place on information sources that will
224 determine whether the potential of online health information is realized. Studies that
225 have addressed this question indicate that information originating from public sources,
226 presented in the English language and that adopt 'scientific' language garners most trust
227 [22], and these are considerations in designing and distributing online health
228 information.

229 A caveat to the present study is that the *embarrassingproblems.com* site was
230 created as a vehicle for the dissemination of medical information and not for data
231 collection. Thus, while there are clear temporal trends in the data, these cannot be
232 pinned to specific dates, which might be informative in fully interpreting patterns of
233 problem resolution. Similarly, data collection cannot be categorized by the sex, age, race
234 or social background of respondents, variables that might be instructive in assessing the
235 effectiveness of the site in providing appropriate medical information. A clear
236 implication of the findings of this study, for the *embarrassingproblems.com* site and
237 others like it, is to better tailor the site for data collection and analysis, to identify user
238 groups and their problems and, thereby, better target information provision. Indeed, the
239 specific date of each data entry on the *embarrassingproblems.com* site is now recorded,
240 with the aim of improving interpretation of problem resolution in future analyses.
241 Further options for improving data collection on the site will need to consider the

242 balance between collection of more user-specific information and the respondents'
243 desire for anonymity.

244 In conclusion, we present an analysis of data gathered from an online
245 information website, which demonstrates that online medical information, across
246 multiple types of health problem, can contribute to the successful resolution of medical
247 problems that are perceived as embarrassing. The probability of problem resolution
248 showed a statistically important temporal improvement on the
249 *embarrassingproblems.com* site, which may reflect both improvement of content during
250 the study period and increasing trust in online resources. The lower increase in the
251 resolution of problems classed as 'urinary' and 'mouth and face' is useful information
252 that will enable the site's editors to continue to refine and focus content provision in
253 these areas, showing a clear benefit to building data collection into the provision of
254 online health information.

255 **Acknowledgements**

256 Thanks to Rowena Spence and Mark Warren for statistical advice, and to Rhiannon
257 Klyne, David Spiegelhalter, Chris Sowden, Rose Thompson and Robert West for
258 comments on the manuscript. *Funding*: This research did not receive any specific grant
259 from funding agencies in the public, commercial, or not-for-profit sectors.

260 REFERENCES

- 261 [1] J. Kivits, Everyday health and the internet: a mediated health perspective on health
262 information seeking. *Sociol. Health Illn.* 31 (5) (2009) 673-687.
- 263 [2] S. Reis, A. Visser, R. Frankel, Health information and communication technology in
264 healthcare communication: The good, the bad, and the transformative. *Patient*
265 *Educ. Couns.* 93 (3) (2013) 359-362.
- 266 [3] L. Kelly, S. Ziebland, C. Jenkinson, Measuring the effects of online health
267 information: Scale validation for the e-Health Impact Questionnaire. *Patient Educ.*
268 *Couns.* 98 (11) (2015) 1418-1424.
- 269 [4] J. Powell, N. Inglis, J. Ronnie, S. Large, The characteristics and motivations of online
270 health information seekers: cross-sectional survey and qualitative interview study.
271 *J. Med. Internet Res.* 13 (1) (2011) e20.
- 272 [5] J.E. Chung, Social networking in online support groups for health: how online social
273 networking benefits patients. *J. Health Commun.* 19 (6) (2014) 639-659.
- 274 [6] I.J. Rowlands, D. Loxton, A. Dobson, G.D. Mishra, Seeking health information online:
275 association with young Australian women's physical, mental, and reproductive
276 health. *J. Med. Internet Res.* 17 (5) (2015) e120.
- 277 [7] M. White, S.M. Dorman, Receiving social support online: implications for health
278 education. *Health Educ. Res.* 16 (6) (2001) 693-707.
- 279 [8] S. Dean, J. Lewis, C. Ferguson, Editorial: Is technology responsible for nurses losing
280 touch? *J. Clin. Nurs.* 26 (5-6) (2016) 583-585.
- 281 [9] M. Koo, M.C. Lu, S.C. Lin, Predictors of Internet use for health information among
282 male and female Internet users: Findings from the 2009 Taiwan National Health
283 Interview Survey. *Int. J. Med. Inform.* 94 (2016) 155-163.

- 284 [10] I.A. Nikoloudakis, C. Vandelanotte, A.L. Rebar, S. Schoeppe, S. Alley, M.J. Duncan,
285 C.E. Short, Examining the correlates of online health information-seeking behavior
286 among men compared with women. *Am. J. Mens Health* 18 (2016)
287 1557988316650625.
- 288 [11] E.N. Ieno, A.F. Zuur, *Data Exploration and Visualisation with R*. Highland Statistics
289 Ltd, Newburgh, UK, 2015.
- 290 [12] A.F. Zuur, A.A. Saveliev, E.N. Ieno, *A Beginner's Guide to Generalised Additive*
291 *Mixed Models with R*. Highland Statistics Ltd, Newburgh, UK, 2014.
- 292 [13] F. Korner-Nievergelt, T. Roth, S. von Felten, J. Guélat, B. Almasi, P. Korner-
293 Nievergelt, *Bayesian Data Analysis in Ecology using Linear Models with R, BUGS,*
294 *and Stan*. Academic Press, London, 2015.
- 295 [14] K.P. Burnham, D.R. Anderson, P values are only an index to evidence: 20th-vs. 21st-
296 century statistical science. *Ecology* 95 (3) (2014) 627-630.
- 297 [15] J.K. Kruschke, *Doing Bayesian Data Analysis*. Academic Press, London, 2015.
- 298 [16] Y.-S. Su, M. Yajima, *R2jags: A Package for Running JAGS from R*. [http://CRAN.R-](http://CRAN.R-project.org/package=R2jags)
299 [project.org/package=R2jags](http://CRAN.R-project.org/package=R2jags), 2012.
- 300 [17] R Development Core Team, *R: A Language and Environment for Statistical*
301 *Computing*. R Foundation for Statistical Computing, Vienna, 2016.
- 302 [18] A. Zuur, E.N. Ieno, N. Walker, A.A. Saveliev, G.M. Smith, *Mixed Effects Models and*
303 *Extensions in Ecology with R*. Springer, New York, 2009.
- 304 [19] A.D. Black, J. Car, C. Pagliari, C. Anandan, K. Cresswell, T. Bokun, B. McKinstry, R.
305 Procter, A. Majeed, A. Sheikh, The impact of eHealth on the quality and safety of
306 health care: a systematic overview. *PLoS Med.* 8 (1) (2011) e1000387.
- 307 [20] J. Sligo, R. Gauld, V. Roberts, L. Villa, A literature review for large-scale health
308 information system project planning, implementation and evaluation. *Int. J. Med.*
309 *Inform.* 97 (2017) 86-97.
- 310 [21] I.B. Hong, H.S. Cha, The mediating role of consumer trust in an online merchant in
311 predicting purchase intention. *Int. J. Inform. Manag.* 33 (6) (2013) 927-939.
- 312 [22] Á. Mendes, L. Abreu, M.R. Vilar-Correia, J. Borlido-Santos, "That should be left to
313 doctors, that's what they are there for!" - Exploring the reflexivity and trust of
314 young adults when seeking health information. *Health Commun.* 32 (9) (2017)
315 1076-1081.
- 316 [23] S. Nettleton, The emergence of e-scaped medicine? *Sociology* 38 (4) (2004) 661-
317 679.

- 318 [24] K. Dobransky, E. Hargittai, Inquiring minds acquiring wellness: Uses of online and
319 offline sources for health information. *Health Commun.* 27 (4) (2012) 31-343.
- 320 [25] M.J. Manierre, Gaps in knowledge: tracking and explaining gender differences in
321 health information seeking. *Soc. Sci. Med.* 128 (2015) 151-158.

322 **Figure legends**

323 **Fig 1.** Screenshot of the boxes that visitors to the *embarrassingproblems.com* site are
324 asked to check to indicate successful or non-successful resolution of their problem after
325 reading the information provided. The binomial data used in this analysis (tackled or not
326 tackled) were generated from these three boxes.

327 **Fig. 2.** Mean fitted probability (solid line) of medical problem resolution and 95%
328 credible intervals (dashed lines) for different classes of medical problem for a Bernoulli
329 GLM estimated by MCMC as a function of time period. Time periods comprised 13
330 discrete ordinal categories, broadly representing 6-month intervals over the 82-month
331 study period.

Tell us your thoughts

Did you find what you were looking for?

Tackled it, moved on

Tackling it

Still struggling

11524 people have tackled this problem!

Submit 

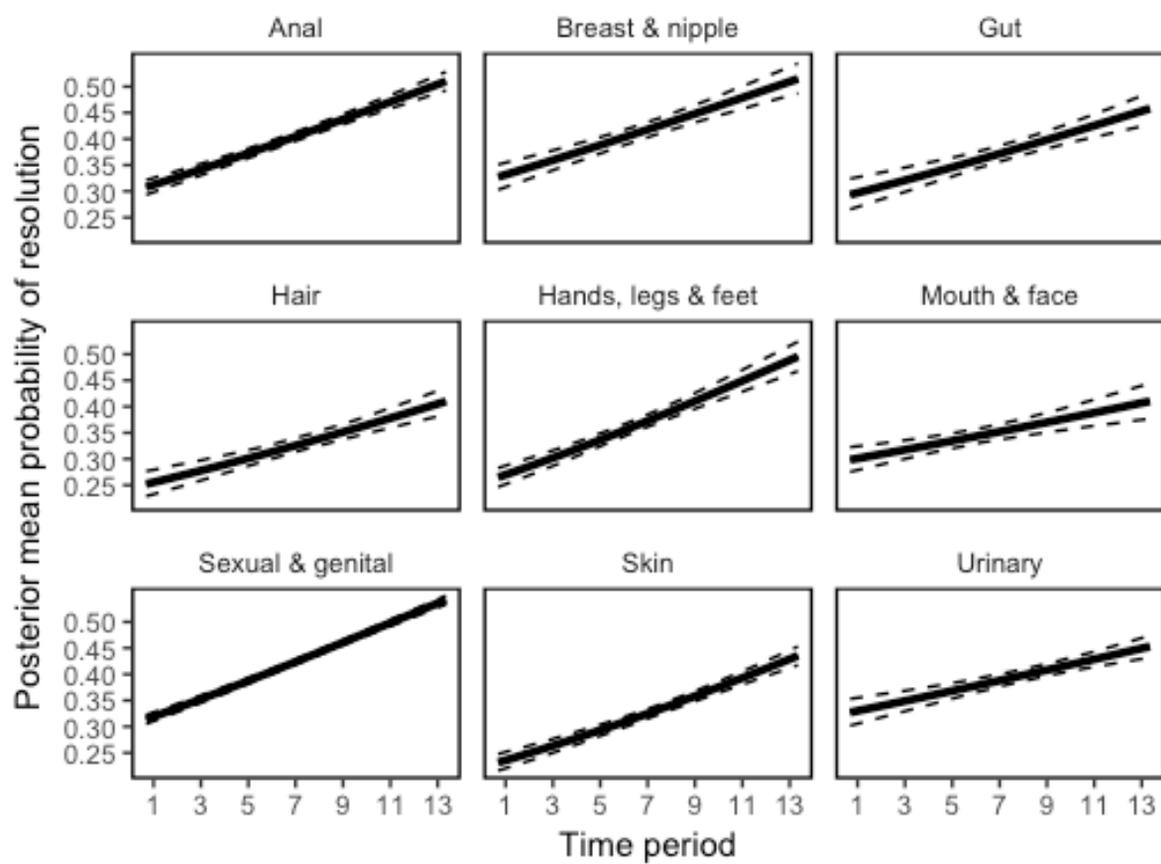


Table 1

Number of responses either not tackled or tackled by health information provided on the *embarrassingproblems.com* online health site. Due to unbalanced results, data for 'cognitive' problems were dropped from the final analysis.

Class of problem	not tackled	tackled	total	proportion tackled (%)
Sexual and genital	25712	19847	45559	43.6
Breast and nipple	2959	2065	5024	41.1
Anal	7705	5097	12802	39.8
Urinary	4112	2724	6836	39.8
Gut	2304	1411	3715	38.0
Hands, legs and feet	4202	2291	6493	35.3
Mouth and face	2675	1398	4073	34.3
Skin	7060	3566	10626	33.6
Hair	3188	1589	4777	33.3
Cognitive	472	184	656	28.0
Total	60389	40172	100561	39.9

Table 2

Parameter estimates of probability of successful resolution of embarrassing medical problems modelled using a Bernoulli GLM. CrI is the 95% Bayesian credible interval. Credible intervals that do not contain zero are in bold to indicate statistical importance. Parameter estimates are presented for each medical problem with anal problems as the baseline category.

Model parameter	Posterior mean	Lower CrI	Upper CrI
Fixed intercept _(anal)	-0.37	-0.41	-0.34
<i>time</i>	0.26	0.23	0.29
<i>problem class</i> _(breast,nipple)	0.05	-0.01	0.12
<i>problem class</i> _(gut)	-0.14	-0.22	-0.06
<i>problem class</i> _(hair)	-0.34	-0.42	-0.27
<i>problem class</i> _(hands,legs,feet)	-0.13	-0.19	-0.07
<i>problem class</i> _(mouth,face)	-0.23	-0.31	-0.15
<i>problem class</i> _(sexual,genital)	0.08	0.04	0.12
<i>problem class</i> _(skin)	-0.34	-0.40	-0.29
<i>problem class</i> _(urinary)	-0.07	-0.13	-0.01
<i>time x problem class</i> _(breast,nipple)	-0.02	-0.09	0.04
<i>time x problem class</i> _(gut)	-0.04	-0.12	0.04
<i>time x problem class</i> _(hair)	-0.04	-0.12	0.03
<i>time x problem class</i> _(hands,legs,feet)	0.05	-0.02	0.11
<i>time x problem class</i> _(mouth,face)	-0.11	-0.18	-0.04
<i>time x problem class</i> _(sexual,genital)	0.03	-0.01	0.07
<i>time x problem class</i> _(skin)	0.03	-0.03	0.09
<i>time x problem class</i> _(urinary)	-0.10	-0.15	-0.04

Summary Table

What was already known on the topic

- The Internet is an important source of online health information and is proliferating
- The anonymity of online health information may be particularly appealing in the case of medical problems perceived as embarrassing
- The success of online health information in resolving health problems is equivocal

What this study has added to our knowledge

- Online health information can make a significant contribution to resolving medical problems that are perceived as embarrassing
- The resolution of embarrassing medical problems showed a statistically important temporal improvement
- The probability of resolving a medical problem varied with medical problem type