Resolving embarrassing medical conditions with online health information

Sarah Redston\textsuperscript{a}, Sharon de Botte\textsuperscript{a}, Carl Smith\textsuperscript{b,c,*}

\textsuperscript{a}Health Press Ltd, Elizabeth House, Queen Street, Abingdon, UK
\textsuperscript{b}School of Biology, University of St Andrews, St Andrews, UK
\textsuperscript{c}Department of Ecology & Vertebrate Zoology, University of Łódz, Łodz, Poland

*Corresponding author. Carl Smith, School of Biology, Bute Building, University of St. Andrews, St. Andrews, Fife KY16 9TS, UK. email: cs101@st-andrews.ac.uk
ABSTRACT

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

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

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

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

Keywords: Internet; Health communication; Online information; Quantitative results
1. Introduction

The Internet increasingly serves as a secondary, and even primary, source of health and medical information to the public [1-3]. The role that online health information will come to play in the future is unclear, but it is a potentially important source of information for promoting good public health [3], and may come to have as great an impact as that seen in banking, education, leisure and social relationships [2].

The advantages of online health information are that it is almost limitless in scope, unregulated, accessible from anywhere with an internet connection, available at any time, potentially interactive and typically free to access, as well as having the potential to combine expert advice with user-generated experience [4]. The anonymity of online health information may be particularly appealing to people seeking information on 'embarrassing' health problems, for example those associated with sexual or mental health [5,6] or body image issues, but potentially a wide range of medical conditions [7].

Some of the drawbacks of online health information are the potential for misinformation, social isolation of users and the undermining of, and distancing from, health professionals [4,8].

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

In this study, we analyzed data gathered over an 82-month period from embarrassingproblems.com, a website specifically designed to provide information on health problems that are often perceived as ‘embarrassing’ or difficult to discuss. During the study period the website received up to 1.2 million hits each month, providing
information on 77 discrete medical problems. At the bottom of each problem page on
the website, users are asked to anonymously record whether the information provided
has enabled them to tackle the given problem or whether it remains unresolved. Our
goal was to investigate whether the data generated by the site shows any temporal
patterns in problem resolution, and whether successful outcome varies with the type of
medical problem.

2. Methods

2.1 Background

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

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

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

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

\[ \text{Outcome}_i \sim \text{Binomial}(\pi_i) \]

\[ E(\text{Outcome}_i) = \pi_i \]

\[ \eta_i = \beta + \text{Problem}_i \times \text{Time}_i \]

\[ \logit(\pi_i) = \eta_i \]

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

To make inferences about the parameters in the model, a Bayesian approach was used. A Bayesian GLM is robust in dealing with complex datasets, unbalanced data, an inherent lack of dependency due to repeated measures, and a highly varied non-normal response variable. Bayesian models are flexible in allowing the estimation of a posterior distribution of differences between parameters and across levels of factors. These are
relatively straightforward procedures using Bayesian inference, but problematic in a frequentist framework [12,13], notwithstanding more general reservations in using frequentist analyses [12-15].

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

3. Results

Overall there was a statistically important increase in the probability of problem resolution among time periods across all classes of embarrassing medical problem (Table 2, Fig. 2). In addition, there was a significant interaction of medical problem with time period. There was a lower increase in the resolution of problems classed as ‘urinary’ and ‘mouth and face’ compared with the baseline class of problem (‘anal’).
(Table 2, Fig. 2). Improvement in resolution of problems classed as ‘breast and nipple’, ‘gut’, ‘hair’, ‘hands, legs and feet’, ‘sex and genital’ and ‘skin’ did not differ from the improvement seen with baseline (Table 2, Fig. 2).

4. Discussion

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

In the present study, a temporal improvement in problem resolution was observed across all medical problem types. While the reason for this trend is opaque, this finding ostensibly implies that the impact of this online health information improved over time. Over the 82-month time period of this study, the embarrassingproblems.com site editors implemented a ‘listen and respond’ approach to content provision. All user-generated comments were manually moderated before being published on each problem page below the expert advice. The site’s editors then ensured that these comments were taken into consideration in content updates made by expert authors and in other materials highlighted on the site. Although the extent to which user-generated content has contributed to the resolution of any given health
problem has not been ascertained, its use in continually improving the editorial offering on the site is likely to have had a positive effect on visitor engagement.

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

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

The interaction of temporal changes in health problem resolution with problem type implies variation in the tractability of healthcare problems, but also that successful resolution of specific problem types has increased faster than others. Given the wide variety of health problems addressed by the embarrassingproblems.com site, it is not unexpected that its success in tackling different problems might vary among problem types. For example, the successful resolution of piles, crab lice, and chronic diarrhea are high in our dataset, reflecting the relative ease with which these problems can be tackled with readily available proprietary medical products. Others, such as profuse sweating,
blushing and flushing, and bed-wetting in adults are medically less tractable problems
and, consequently, show much lower rates of resolution. In our analysis, urinary
problems and medical problems associated with the mouth and face showed the
weakest improvement over time. While these findings are unambiguous in our data, an
explanation for the observed variance in the temporal improvement of different
problem types is opaque and will require further data collection.

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

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

The growth of online medical information has enormous potential to inform and
empower patients and thereby facilitate improvements in healthcare outcomes [24]. In
the context of the present study, online health information that addresses problems that are perceived as embarrassing may be particularly effective, and the provision of online health information more generally may benefit from refinement and focus. Young adults are a well-represented demographic group that readily uses online health information [22], as are well-educated high-income women [10,25]. Online health information that targets these groups is evidently in demand, though the needs of online users of different racial, ethnic, and socioeconomic backgrounds is clearly also necessary.

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

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

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

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

REFERENCES


13

Figure legends

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

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

The figure illustrates the posterior mean probability of resolution over time for various body parts and conditions. The x-axis represents time periods, and the y-axis shows the posterior mean probability. Each graph represents different body parts: Anal, Breast & nipple, Gut, Hair, Hands, legs & feet, Mouth & face, Sexual & genital, Skin, Urinary. The trend lines indicate an increasing probability over time.
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.

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

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