International Journal of Medical Informatics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Resolving embarrassing medical conditions with online health information

Sarah Redston[a], Sharon de Botte[a], Carl Smith[b,c,*]

[a] Health Press Ltd, Elizabeth House, Queen Street, Abingdon, UK
[b] School of Biology, University of St Andrews, St Andrews, UK
[c] Department of Ecology & Vertebrate Zoology, University of Łodz, Łodz, Poland

ARTICLE INFO

Keywords:
Internet
Health communication
Online information
Quantitative results

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.

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

* Corresponding author at: School of Biology, Bute Building, University of St. Andrews, St. Andrews, Fife KY16 9TS, UK.
E-mail address: cs101@st-andrews.ac.uk (C. Smith).

https://doi.org/10.1016/j.ijmedinf.2018.03.010
Received 13 May 2017; Received in revised form 31 October 2017; Accepted 20 March 2018

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

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

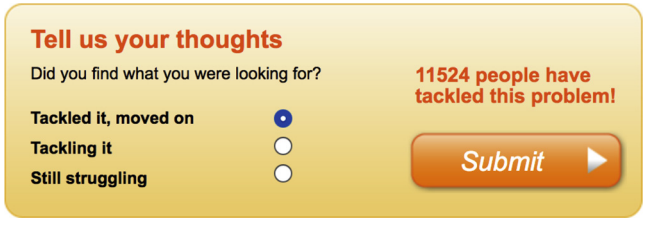

Fig. 1. Screenshot of the boxes that visitors to the embarrassingproblems.com site are asked to check to indicate successful or non-successful resolution of their problem after reading the information provided. The binomial data used in this analysis (tackled or not tackled) were generated from these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

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:

$$Outcome_i \sim Binomial(\pi_i)$$

$$E(Outcome_i) = \pi_i$$

$$\eta_i = \beta + Problem_i * Time_i$$

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

$Outcome_i$ is the probability of a positive outcome (i.e. problem tackled) for respondent i assuming a Bernoulli distribution with mean π_i and variance $\pi_i \times (1 - \pi_i)$. $Problem_i$ is the class of medical problem (Table 1) experienced by respondent i. $Time_i$ 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 nonlinear 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).