

The Quality of Online Answers to Parents Who Suspect That Their Child Has An Autism Spectrum Disorder

Ayelet Ben-Sasson

Haifa University
199 Aba Khoushy Ave., Haifa, Israel

Dan Pelleg

Yahoo Research
Haifa, Israel

Elad Yom-Tov

Microsoft Research
13 Shenkar st., Herzeliya, Israel

Abstract

The growing diagnosis and public awareness of Autism Spectrum Disorders (ASD) leads more parents to seek answers to their suspicions for ASD in their child on Internet forums. This study describes an analysis of the quality of content of 371 answers on Yahoo Answers (YA), a social question-and-answer forum, to parents querying whether their child has ASD. We contrasted the perceived quality of answers by clinicians with that of parents. The study tested the feasibility of automatically assisting parents in selecting answers with higher quality using a predictive model based on the text of answers and the attributes of answerers.

Introduction

There is growing awareness of autism spectrum disorder (ASD) and its increasing incidence (French et al. 2013), resulting in more parents suspecting their child has ASD. Many of these parents search for information and answers online. Parents approach online communities to query about their child's developmental and health status; with about a third sharing search results with their healthcare provider (Tuffrey and Finlay 2002). Online communities offer parents accessible, immediate, personal, and anonymous communication (Pelleg, Yom-Tov, and Maarek 2012), as well as an ability to obtain a wide range of perspectives. At the same time, health information seekers report challenges in dealing with the vast amount of online answers and their variability (Worrall, Oh, and Yi 2012). Some online health information is inaccurate, misleading, and biased (Rice 2006; Scullard, Peacock, and Davies 2010) which can lead to dangerous actions, increase parental anxiety, or build false hopes (Stvilia, Mon, and Yi 2009).

Parental concerns have been shown to predict an ASD diagnosis (Ozonoff et al. 2009). There is a lag between first concern and diagnosis (Siklos and Kerns 2007). During this lag parents may turn online to search for answers, but quality and content of online answers can further delay or prevent a parent from approaching a professional, consequently delaying diagnosis. In addition, answerers may recommend non-validated solutions (e.g., dairy free diet), public health hazards (e.g., avoiding vaccines), and costly actions.

The present study focused on the questions of parents suspecting their child has ASD that were posted on Yahoo Answers (YA), a social question-and-answer (Q&A) forum. YA is one of the most popular websites on the Internet (at the time of writing, it was ranked fifth by Alexa¹, and arguably the most popular Q&A website). Askers on YA post detailed questions using natural language, usually a paragraph or more in length. Other users of the site can submit answers to these questions. A question can elicit multiple answers, one of which can be rated as best by the asker or, if none is selected by the asker, by the community². YA supports a reputation system which involves earning points based on the number of answers, and number of answers rated as best. But the intrinsic reputation system is driven by many conflicting factors (Pelleg, Yom-Tov, and Maarek 2012). Therefore, there is a need for an embedded objective clinical quality metric for health answers.

The quality of online health information is known to be problematic (Eysenbach et al. 2002; Scullard, Peacock, and Davies 2010). Most studies on the quality of health information online and quality evaluation tools reviewed deal with websites that provide static health information. Some of these quality assessment tools cite popularity as a quality measure rather than clinical quality parameters (Bernstam et al. 2005; Eysenbach et al. 2002; Scullard, Peacock, and Davies 2010; Wilson 2002). Q&A sites comprise of ever changing health information, hence pose a unique challenge for assessing their quality. Specifically, surveys indicate that 95% of questions on YA received at least one answer; however, the quality of the answers ranged from correct, partial, unhelpful, to spam (Shah and Pomerantz 2010). Users are challenged in determining the credibility of online information (Metzger 2007). Investigating quality assessment of ten answers from 25 health categories in YA showed that librarians and nurses are more concerned with the subjectivity and the source of the answer, while users (however not the askers) are more concerned with the style of the answer and the social-emotional support it provides (Worrall, Oh, and Yi 2012). Some studies have pointed to users seeking other cues

¹<http://www.alexa.com/siteinfo/yahoo.com>

²In early 2014, the rules on Yahoo Answers changed, disallowing community voting. The effect on the current research is negligible as most data collected precedes this date.

of quality. For example, parents reported checking whether someone has an economic incentive in the health website, the authority of the source, professionalism, readability, and whether it provides good resources (Eysenbach et al. 2002; Impicciatore et al. 1997). Establishing evidence regarding the quality of online information transferred to worried parents is critical for understanding the benefits and risks of the growing reliance upon the web for diagnostic information. Such research can assist in developing tools to guide parents' critical evaluation of online clinical information.

Methods

All English language queries from YA which contained the words "autism", "Asperger", "ASD", or "PDD" were extracted. A total of 8,681 queries met these criteria. Crowdsourcing (using CrowdFlower) was used to classify the queries into queries of parents of a child who was diagnosed with ASD (n=2,412), parents who have a child diagnosed with ASD and another child who they suspect has ASD (n=41), parents who suspect their child has ASD (n=1,081), and none of the above (n=5,147).

Of the 1,081 queries of suspecting parents, 371 answers to 50 queries were randomly selected. The mean number of answers per query was 7.46 (SD=234, range=2-10). In 80% of queries that mentioned the age of the child, it was, on average, 1.4 years (SD=1.4, range=0.06-8). Fifty-four percent of queries were about males, 22% about females, and the rest unspecified. Twenty-four percent of parents indicated they wanted a first-hand answer, of which 42% received a first-hand answer. Another 4.5% received an answer from an expert in the field. Of the 50 coded queries, askers selected within the YA platform the best answer for 29 queries. Another 20 queries were selected within the YA platform by the community as best.

A set of 16 quality variables were defined for coding the clinical quality of answers: first-hand, expert, contact information, courteous, support, signs, actions, referral to professional, resources, responsible, asking the asker a question, evidence, no risk, differential diagnosis, no advertisement, and relevance. A detailed coding manual with a definition and examples for each variable was developed. Two ASD clinical experts first coded the answers for the presence of each of the 16 quality variables. Experts also selected the best answer relative to the other answers to that query, as a judgment of their overall quality impression. Experts were blind to the best answer selected by the users. Inter-rater reliability was tested for 20% of the answers that were coded by 2 clinical experts. For this subset of data, kappa was above 0.70 for the majority of quality variables. Variables that remained low in agreement (i.e., best answer and relevance) were assigned a consensus rating. A total quality summary score of the 16 variables was computed. Two composite scores were created to investigate answer attributes that may drive 'best answer' selection by users. These included:

1. Information: signs, actions, referral, resources, evidence, and differential diagnosis.
2. Global support: contact, courteous, and support.

This study was approved by the Haifa University IRB.

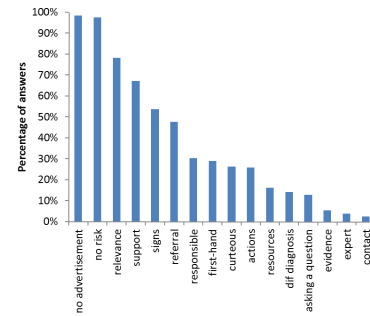


Figure 1: Percentage of answers for each quality variable.

	Clinicians			Parents		
	Best	Not Best	P	Best	Not Best	P
N=	50	321		29	185	
First-hand	18 (36%)	89 (28%)	0.24	17 (59%)	38 (21%)	*
Signs	34 (68%)	165 (51%)	0.02	21 (72%)	90 (49%)	0.01
Actions	19 (38%)	77 (24%)	0.03	11 (38%)	46 (25%)	0.17
Resources	8 (16%)	52 (16%)	1.00	10 (34%)	25 (13%)	0.01
Referral	37 (74%)	140 (44%)	*	14 (48%)	96 (52%)	0.84
Evidence	6 (12%)	14 (4%)	0.04	6 (21%)	8 (4%)	0.01
Support	40 (80%)	209 (65%)	0.02	21 (72%)	112 (60%)	0.30
Responsible	27 (54%)	85 (26%)	*	11 (38%)	58 (31%)	0.52
Relevance	48 (96%)	242 (75%)	*	22 (76%)	145 (78%)	0.81
Total quality score (of 16)	8 (2%)	6 (2%)	*	7 (2%)	6 (2%)	0.001

Table 1: Quality variables significantly different between best versus not best answers. Stars denote $P \leq 0.001$.

Results

Figure 1 shows the percentage of answers rated by clinicians as meeting criteria for each clinical quality variable. Over 50% of the answers had no advertisement, no risk, were relevant, provided support, and described early ASD signs. There was low agreement between the 'best answers' selected by clinicians (n=50) versus those selected by parents (n=29) or all users (n=49), kappa=0.19 ($p = 0.005$), and 0.17 ($p = 0.001$), respectively. Of answers rated best by parents, 31% were rated as best by clinicians; of user-rated best answers 28.6% were rated as best by clinicians.

Fisher's Exact tests indicated that best answers selected by clinicians had significantly ($p < 0.05$) higher frequencies of signs, actions, referral to professional, evidence, support, responsibility, and relevance (see Table 1).

Features of Best Answers Selected by Askers

Fisher's Exact tests indicated that best answers as selected by parents had significantly ($p < 0.05$) higher frequencies of signs and of evidence, which were also favored by clinicians. Additional variables that differentiated parent-rated best answers were first-hand answers, and resources (see Table 1). The degree to which the findings related to the parents perspective versus a non-expert was tested using all 49 queries which had a best answer selected by the community and/or parents. The above findings were replicated in this sample.

To determine whether parents sought support versus information, Mann-Whitney tests were conducted. Tests showed that best answers selected by parents versus not had significantly higher information scores ($M = 0.39$ [$SD = 0.21$], $M = 0.26$ [$SD = 0.19$], respectively, $Z = -2.89$, $p = 0.004$) but did not differ in global support ($p = 0.08$). To isolate the two constructs, the 214 answers (for the 29 queries selected by parents for 'best answer') were classified into answers providing information and support (57.9%), only information (26.2%), only support (10.7%), and neither (5.1%). Of the best answers selected by parents, 72.4% included both information and support, 24.1% included only information and 3.4% included only support.

Prediction of Clinical Quality of Answers

In this section we tested the feasibility of automatically identifying the clinical quality variables of each answer and whether it would be marked as the best answer by a clinician. Each answer was modeled using four classes of attributes: Answer text, technical attributes of the answer, attributes of the answerer, and answerer history. We trained a decision tree classifier to predict the output variables and estimated its performance using 10-fold cross validation (Duda, Hart, and Stork 2012).

The model attempted to predict both individual clinical attributes and (separately) the best answer chosen by clinicians. For the latter, new attribute vectors were constructed and comprised of the difference between individual answer vectors of the same query (following (Joachims 2002)). Testing all combinations of attribute families showed that, on average, the technical attributes of answers provide the best prediction for the clinical variables. Best answer was predicted most accurately by adding the textual information to the technical attributes of the answer with an Area Under the Receiver Operating Curve (AUC) of 0.85. Classifier performance, as measured using the AUC is given in Table 2. Some attributes (asking a question, providing resources) were easily identified. However, most attributes were difficult to identify automatically from the answer using the attributes extracted. Interestingly, it was possible to identify the best answer that would have been chosen by clinicians.

Using sequential forward feature selection indicated that the most important attributes for prediction of the best answer were (in descending order):

1. The number of sentences in an answer.
2. The word "name" (e.g., "if the child knows their name")
3. The word "answer" (e.g., "This is hard to answer.")

Variable	AUC (s.d.)	90% CI
Resources	0.84 (0.02)	0.80 - 0.88
Asking a question	0.82 (0.02)	0.81 - 0.90
Referral to professional	0.76 (0.02)	0.73 - 0.80
First-hand	0.75 (0.03)	0.68 - 0.77
Courteous	0.73 (0.02)	0.69 - 0.77
Differential diagnosis	0.71 (0.04)	0.61 - 0.73
Signs	0.67 (0.02)	0.67 - 0.74
Support	0.65 (0.03)	0.60 - 0.69

Table 2: Area Under the ROC (AUC) for predicting quality variables with AUCs above 0.60.

4. The word "repetitive" (e.g., "Repetitive movement.")
5. The word "give" (e.g., "give a list of symptoms")

Discussion

The internet offers worried parents immediate judgment regarding their child's likelihood of having ASD. Although medical guidelines call for universal ASD screening as early as 18 months (Johnson, Myers, and others 2007) in practice many parents remain worried until a final diagnosis is made, sometimes three years or more later (Siklos and Kerns 2007). Parents approach online communities for information regarding the signs of ASD and for verifying their concern for their child being on the Autism spectrum.

This study highlights the variability in the quality of answers parents receive. Our results show a difference in the best answer selected by parents compared to that chosen by clinicians. Using machine learning methods, we demonstrated the possibility to use answer features to predict the clinical quality of answers. These findings underscore both the current need for concerned parents to obtain a clinical perspective regarding the child's risk for ASD, and for the possibility of designing computerized tools to provide automatic evaluation of online health answers so as to highlight more authoritative answers for use by parents.

There was low agreement between clinicians and asking parents in their selection of the best answer although the mean quality scores of best answers selected by parents versus those of clinicians were similar. Unique features of answers preferred by clinicians were suggestions for actions, referral, and support; while, parents favored answers providing resources and first-hand accounts. This is similar to previous research showing user preference for first-hand accounts and resources (Eysenbach et al. 2002; Impicciatore et al. 1997). Our analysis reveals that clinicians valued answers with factual information regarding ASD signs, evidence, recommendations for action, referral to a professional, relevance of answer to the query, answers reflecting responsibility, and answers explicitly supporting the askers. This is in contrast with experts in a previous study of health answers who were primarily concerned with the subjective nature of the answers (Worrall, Oh, and Yi 2012). Of the features that characterized best answers selected by clinicians, signs and evidence also characterized queries preferred by the asking parents.

Based on Worrall et al.s (Worrall, Oh, and Yi 2012) study the working hypothesis in our study was that parents will prefer answers that provide support rather than information. This hypothesis was not fully confirmed, possibly due to methodological differences between studies. While the present study examined askers in-vivo selection of best answer, Worrall et al. asked a group of users (not askers) to rate the best answer and quality features of an existing set of queries which covered a broad range of health issues. It is plausible that in the case of a parent suspecting their child has a clinical condition the major incentive for querying is for information exchange as opposed to a parent dealing with a chronic diagnosed condition who is looking for support. Nonetheless, best answers selected by clinicians did have a higher rate of providing the asker support (i.e., calming, empowering, expressing empathy). Cautious interpretation of these differences is required given the small sample size and the fact that parents rated an absolute best answer whereas clinicians rated the relative best answer.

Although it was encouraging that the global clinical quality of an answer (i.e., best answer) was automatically predicted, most individual clinical quality features identified as important by clinicians could not be automatically predicted. Clinicians are trained to provide information that is balanced in presenting all possibilities. It is not surprising that those features, which reflect this clinical reasoning, such as a responsible answer, referral to a professional, and mentioning a differential diagnosis could not be predicted. The findings also underscore the role of healthcare providers in filling the information gap associated with relying on online answers. Clinicians can help parents critically appraise online answers in terms of their relevance to their individual case.

When predicting the clinical quality of health answers the content of the answer is the primary player. Answerer features were not predictive of the clinical quality of the answer. This differs from research of questions on YA for non-health topics showing that automatically extracted features of the answerer were better predictors of quality ('best answer' by asker) than the human coded dimensions of quality (Shah and Pomerantz 2010).

Early diagnosis of ASD can be a challenging task for an expert and even more so for a non-specialist Internet user. Parents approaching the online community for answers to their worries face the risk of obtaining inaccurate and irrelevant information that lacks the consideration of alternative explanations for their child's state. Healthcare providers play an important role in facilitating parents appraisal of health information transmitted on social media. Since most parents do not consult their healthcare professional regarding online information (Tuffrey and Finlay 2002), there is a need for systems that support parents to critically appraise of online health answers, and our results show that building such systems is feasible.

Acknowledgment: This research was partially funded by a grant from the Caesarea Rothschild Institute Foundation.

References

- Bernstam, E. V.; Shelton, D. M.; Walji, M.; and Meric-Bernstam, F. 2005. Instruments to assess the quality of health information on the world wide web: what can our patients actually use? *International journal of medical informatics* 74(1):13–19.
- Duda, R. O.; Hart, P. E.; and Stork, D. G. 2012. *Pattern classification*. John Wiley & Sons.
- Eysenbach, G.; Powell, J.; Kuss, O.; and Sa, E.-R. 2002. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA* 287(20):2691–2700.
- French, L. R.; Bertone, A.; Hyde, K. L.; and Fombonne, E. 2013. Epidemiology of autism spectrum disorders. *The neuroscience of autism spectrum disorders* 3–24.
- Impicciatore, P.; Pandolfini, C.; Casella, N.; Bonati, M.; et al. 1997. Reliability of health information for the public on the world wide web: systematic survey of advice on managing fever in children at home. *BMJ* 314(7098):1875.
- Joachims, T. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 133–142.
- Johnson, C. P.; Myers, S. M.; et al. 2007. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 120(5):1183–1215.
- Metzger, M. J. 2007. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology* 58(13):2078–2091.
- Ozonoff, S.; Young, G. S.; Steinfeld, M. B.; Hill, M. M.; Cook, I.; Hutman, T.; Macari, S.; Rogers, S. J.; and Sigman, M. 2009. How early do parent concerns predict later autism diagnosis? *Journal of developmental and behavioral pediatrics: JDBP* 30(5):367.
- Pelleg, D.; Yom-Tov, E.; and Maarek, Y. 2012. Can you believe an anonymous contributor? on truthfulness in yahoo! answers. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*, 411–420. IEEE.
- Rice, R. E. 2006. Influences, usage, and outcomes of internet health information searching: multivariate results from the pew surveys. *International journal of medical informatics* 75(1):8–28.
- Scullard, P.; Peacock, C.; and Davies, P. 2010. Googling children's health: reliability of medical advice on the internet. *Archives of Disease in Childhood* 95(8):580–582.
- Shah, C., and Pomerantz, J. 2010. Evaluating and predicting answer quality in community qa. In *Proceedings of the 33rd International ACM SIGIR conference*, 411–418.
- Siklos, S., and Kerns, K. A. 2007. Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. *Research in developmental disabilities* 28(1):9–22.
- Stvilia, B.; Mon, L.; and Yi, Y. J. 2009. A model for online consumer health information quality. *Journal of the American Society for Information Science and Technology* 60(9):1781–1791.
- Tuffrey, C., and Finlay, F. 2002. Use of the internet by parents of paediatric outpatients. *Archives of Disease in Childhood* 87(6):534–536.
- Wilson, P. 2002. How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet. *BMJ* 324(7337):598.
- Worrall, A.; Oh, S.; and Yi, Y. J. 2012. Quality evaluation of health answers in social q&a: Socio-emotional support and evaluation criteria. *Proceedings of the American Society for Information Science and Technology* 49(1):1–5.