BMJ Global Health

Optimising mHealth helpdesk responsiveness in South Africa: towards automated message triage

Matthew Engelhard,[1] Charles Copley,[2] Jacqui Watson,[2] Yogan Pillay,[3] Peter Barron,[4] Amnesty Elizabeth LeFevre[5]

To cite: Engelhard M, Copley C, Watson J, *et al*. Optimising mHealth helpdesk responsiveness in South Africa: towards automated message triage. *BMJ Glob Health* 2018;**3**:e000567. doi:10.1136/bmjgh-2017-000567

Handling editor Seye Abimbola

Received 12 September 2017
Revised 9 January 2018
Accepted 10 January 2018

[1]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, North Carolina, USA
[2]Praekelt Foundation, Cape Town, South Africa
[3]Department of HIV, TB and MCWH, National Department of Health, Pretoria, South Africa
[4]School of Public Health, University of the Witwatersrand, Johannesburg, South Africa
[5]Department of International Health, Johns Hopkins University Global mHealth Initiative, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Correspondence to
Dr Matthew Engelhard;
mme9n@virginia.edu

ABSTRACT

In South Africa, a national-level helpdesk was established in August 2014 as a social accountability mechanism for improving governance, allowing recipients of public sector services to send complaints, compliments and questions directly to a team of National Department of Health (NDoH) staff members via text message. As demand increases, mechanisms to streamline and improve the helpdesk must be explored. This work aims to evaluate the need for and feasibility of automated message triage to improve helpdesk responsiveness to high-priority messages. Drawing from 65 768 messages submitted between October 2016 and July 2017, the quality of helpdesk message handling was evaluated via detailed inspection of (1) a random sample of 481 messages and (2) messages reporting mistreatment of women, as identified using expert-curated keywords. Automated triage was explored by training a naïve Bayes classifier to replicate message labels assigned by NDoH staff. Classifier performance was evaluated on 12 526 messages withheld from the training set. 90 of 481 (18.7%) NDoH responses were scored as suboptimal or incorrect, with median response time of 4.0 hours. 32 reports of facility-based mistreatment and 39 of partner and family violence were identified; NDoH response time and appropriateness for these messages were not superior to the random sample (P>0.05). The naïve Bayes classifier had average accuracy of 85.4%, with ≥98% specificity for infrequently appearing (<50%) labels. These results show that helpdesk handling of mistreatment of women could be improved. Keyword matching and naïve Bayes effectively identified uncommon messages of interest and could support automated triage to improve handling of high-priority messages.

Key questions

What is already known?

- South Africa has the highest use of health services across the continuum of care in sub-Saharan Africa.
- An mHealth helpdesk staffed by the National Department of Health has received nearly 250 000 messages since its launch in August of 2014.
- Disrespect and abuse of women during childbirth has emerged as a key indicator of the overall quality of care and a barrier to improving maternal and child health outcomes.

What are the new findings?

- 32 incidents of facility-based mistreatment and 39 of partner and family violence were reported to the helpdesk over a 9-month period.
- National Department of Health responsiveness to these messages, quantified via response time and appropriateness, was not superior to random sample.
- Labels used by staff to categorise messages were predicted by machine learning (naïve Bayes classifier) with 85.4% accuracy, including ≥98% specificity for infrequently appearing labels.

What do the new findings imply?

- Automated triage of incoming helpdesk messages could be used to prioritise reports of mistreatment and other urgent messages, leading to improved responsiveness.
- The naïve Bayes classifier is a promising means to categorise incoming messages and could form the basis of automated triage.
- Reports of mistreatment sent to the helpdesk add to a complex picture of neglect at health facilities and could support or inform targeted interventions.

INTRODUCTION

South Africa has the highest use of health services across the continuum of care in sub-Saharan Africa. In the 2016 Demographic and Health Survey, 94% of women attended antenatal care from a skilled provider, 96% delivered in a health facility and 84% attended postnatal care within 2 days following birth.[1] Despite increasing uptake of public sector health services, maternal and child mortality rates remain well above the Millennium Development Goal targets[2] raising important questions about the underlying content and quality of care received.

Disrespect and abuse of women during childbirth—a subset of violence against women—has emerged as a key indicator of the overall quality of care and a barrier to improving maternal and child health outcomes.[3] In South

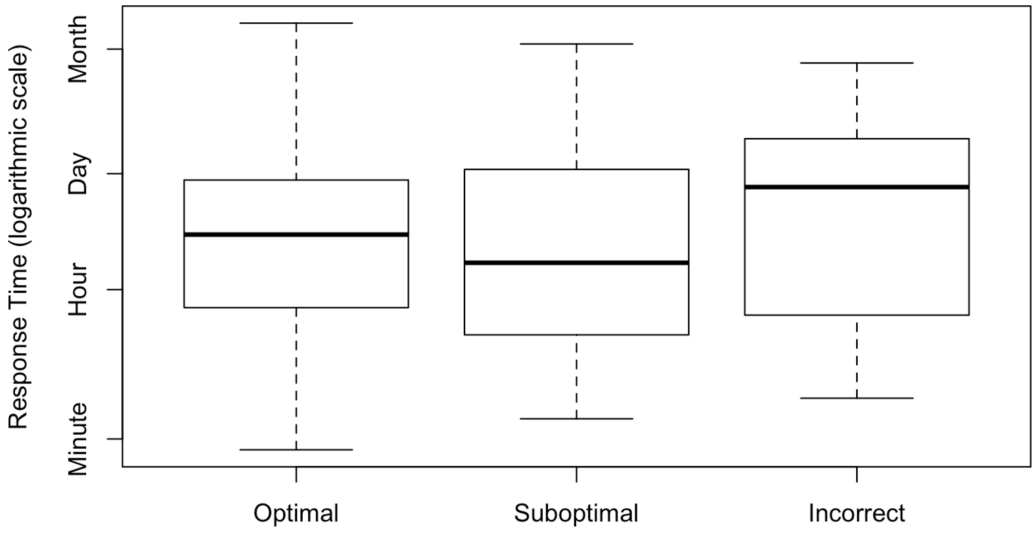

Figure 1 Distribution of helpdesk response times between responses scored as optimal, suboptimal and incorrect.

Africa, the importance of provider–patient relationships first gained traction over two decades ago with the 1998 publication of a qualitative study led by Dr Rachel Jewkes and colleagues, exploring the question of "why do nurses abuse patients?".[4] Through interviews with patients and staff, a complex picture of clinical neglect, verbal and physical abuse emerged, which suggested that mistreatment of women during childbirth had become commonplace in South Africa.[4] Subsequent studies reinforce these findings, attributing mistreatment, in part, to a lack of accountability and action on the part of managers.[3 5 6]

In August of 2014, the National Department of Health (NDoH) launched the helpdesk as a social accountability mechanism for improving governance, allowing recipients of public sector services to hold providers and the NDoH accountable for the content and quality of care provided. Any individual attending public health facilities in South Africa can SMS (short message service) the helpdesk with complaints, compliments or questions. Importantly, the helpdesk is tied to MomConnect, a maternal messaging platform designed to support pregnancy and motherhood, leading to improved outcomes for South African women and their children.[7] Thus, in addition to its broader functionality, the helpdesk is used by MomConnect users to opt out of messages or communicate important updates, such as the birth of a baby.

All incoming messages are sent directly to the NDoH, where the helpdesk, staffed by nurses, is physically located. Helpdesk messages are labelled and assigned to one of four full-time personnel for handling. Based on their content, responses to questions use one of 114 custom responses derived from frequently asked questions. In the event that none of the custom responses are appropriate, the woman is given a customised response, which often includes referral to her health facility. Processes for responding to complaints follow a lengthier process, which typically includes a response from the helpdesk to provincial representatives who then follow up on a case-by-case basis with district and facility authorities.

Since its launch in August of 2014, the helpdesk has received nearly 250 000 messages.[7] However, little is known about the use of the helpdesk for reporting instances of violence against women, including mistreatment during pregnancy or childbirth. Understanding user engagement, coupled with the timeliness and appropriateness of the helpdesk's response, to reported instances of violence that require an urgent or escalated response is vital for ensuring that it is responsive to population needs. At present, the management of messages has largely been manual, necessitating a gradual increase in the central-level personnel required to manage responses. As the helpdesk continues to expand its user base and move into new programme areas, efforts are needed to optimise response times and content and also to accommodate increases in message volume without overburdening the support staff. Where the prior paper in this series sought to describe user engagement with the helpdesk (ref series paper 5 on helpdesk), this paper outlines early efforts to understand and potentially enhance helpdesk performance through automated message triage using the handling of messages on mistreatment of women as a case study.

We begin by characterising the helpdesk response to incoming messages, with a focus on potential areas of improvement. Then, we explore the feasibility of an automated triage system, one which would sort and prioritise incoming messages, as a possible *mechanism* of improvement. Message handling is assessed in terms of response quality and timeliness among (1) a random sample of messages and (2) messages relating to mistreatment of women, as identified via expert-defined keywords. While all messages would ideally receive a prompt and appropriate response, the mistreatment of women is a high-importance topic warranting prioritisation should a triage system be implemented. After determining whether mistreatment can be identified by keyword, we further explore automated triage by training a naïve Bayes classifier to assign the message labels already used by NDoH

Table 1 Keywords used to identify typologies of mistreatment

Typology of mistreatment	Keywords
Verbal abuse	shout, scream, yell, insult
Physical abuse	hit, beat, slap, push, pinch, grab
Violations of confidentiality or privacy	confidential, private, secret
Discrimination	discriminate, deny, refuse, racist, sexist
Politeness	rude, mean, angry, abrupt, hostile
Abandonment	attend, abandon, alone, myself
Autonomy	permission, touch, consent, scare
Birth companion	companion, visitor, parent, friend, family
Bribes	bribe, pay, money

staff. We hypothesise that (1) mistreatment of women can indeed be identified with keywords, (2) handling of messages reporting mistreatment is no better than the handling of other messages and (3) naïve Bayes will replicate less-common labels with high specificity, allowing them to be selectively identified to avoid overburdening NDoH staff. If confirmed, (1) and (2) would establish the need for triage, and (1) and (3) would demonstrate its feasibility.

CHARACTERISING THE HELPDESK RESPONSE

Acquisition and demographics of incoming messages

Individuals attending public health facilities in South Africa can send a text message via SMS to the helpdesk at any time to communicate directly with NDoH staff. Messages received between 27 October 2016 and 17 July 2017 were downloaded from the District Health Information System (DHIS2) for this analysis. Responses to these messages by helpdesk staff were also downloaded, and incoming messages were paired to the corresponding responses for subsequent analysis.

A total of 65 768 messages acquired during the aforementioned window were available for analysis. In total, 49 300 (75.1%) of these were questions, 9189 (14.0%) were message switch requests, 3121 (4.8%) were compliments, 2561 (3.9%) pertained to prevention of mother-to-child transmission of HIV ('PMTCT'), 560 (0.9%) were opt-out requests, 361 (0.5%) were complaints, 351 (0.5%) were language switch requests, 109 (0.2%) were spam and 126 (0.2%) could not be classified.

The helpdesk users sending these messages ranged from 13 to 52 years old, with an average age of 27.8±6.15 years and an IQR of 9.2 years. The largest number of messages were received from KwaZulu-Natal (28.1%) followed by Gauteng (22.1%), Limpopo (12.9%), Mpumalanga (9.4%), Eastern Cape (8.8%), North West (7.5%), Western Cape (5.7%), Free State (4.5%) and lastly Northern Cape (1.0%). The majority of registered users (50.2%) were recorded as possessing a South African National ID card, and approximately a third (33.0%) opted to receive messages in English. Gestational age averaged 19.5±8.4 weeks with an IQR of 12.1 weeks.

Message handling: random sample

To characterise the response to these messages, a random sample of 481 English-language messages was drawn from the complete message set. This sample size was chosen to obtain desirable (<4%) margins of error for the scoring categories described in the next section.

Helpdesk responses were characterised by two measures: the appropriateness of the response and the time delay between the incoming message and its associate response. Response appropriateness was manually scored as follows: a score of 0 (none) indicates no reply was given, 1 (incorrect) indicates the reply did not make sense or otherwise seemed to be in error, 2 (suboptimal) indicates an inappropriate or incomplete response, and all other responses were scored as 3 (satisfactory). Time delay (or response time) was measured as the difference between timestamps associated with a query and its corresponding reply. If more than one reply was sent, the earliest timestamp was used in the calculation.

Among the randomly sampled messages (n=481), the age of users ranged from 15 to 44 years, with an average age of 26.6. While all provinces were represented, the majority of messages were sent from users in Gauteng (116 messages, 24.1%) and the least from Northern Cape (4 messages, 0.8%).

A total of 391 messages (81.3%±3.5%) were scored as optimal, 80 (16.6%±3.3%) as suboptimal and 10 (2.1%±1.3%) as incorrect. Median response time was 4.0 hours among all sampled responses, 4.5 hours among those graded as optimal, 2.1 hours among those graded as suboptimal and 17.2 hours among those graded as incorrect. Figure 1 shows response times for these categories; a logarithmic scale has been used due to the presence of extreme outliers. The difference in response time between optimal and suboptimal responses was not statistically significant (P=0.07). With only 10 messages graded as incorrect, differences in response time between this group and the others were not statistically significant (P>0.3).

CASE STUDY: RESPONSES TO REPORTS OF MISTREATMENT

Identifying mistreatment of women

Messages were screened for the presence of keywords pertaining to mistreatment (table 1). Keywords were identified based on frequent typologies of abuse reported in the literature.[3 5 6 8–10] Those matching any of the keywords were flagged for examination, and the presence of one of the typologies listed in the table was manually confirmed. Messages were further divided according to whether the mistreatment occurred at a health facility or at home; the

Table 2 Typologies of mistreatment identified among helpdesk responses October 2016 to July 2017

Typologies of abuse	n	Illustrative question	Helpdesk response
Partner and family violence	39	"I have a problem with my boyfriend. always when he get drunked He always hurting me, so what i must do plz" (38 y.o. registered user from Eastern Cape)	"It is NOT OK if your partner or anyone hits you or shouts at you. You have the right to seek help. Talk to a friend or a health worker for advice. You need to put your health and the health of your baby first. Call 0800 150 150. It's a 24 hours Stop Gender Violence helpline and it's free to call this number from a landline. (Normal cell phone rates apply)"
Facility-based mistreatment	32		
Discrimination	9	"hospital are refusing me my right to have my baby treated somewhere. Can i get help?" (29 y.o. registered user from Gauteng)	"Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility."
Verbal abuse	8	"The nurse at the clinic yells at me" (22 y.o. registered user from Limpopo)	"Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility."
Politeness	7	"[removed] clinic nurses are rude" (21 y.o. registered user from Limpopo)	"Thank you for sending in your complaint, we have taken note of it and will log the complaint with the Department of Health and your facility."
Violations of confidentiality or privacy	2	"I JUST THANK ALL THE HARDWORK THEY HAVE BEING DOING, BUT SOME OF THEM ARE IMPATIENT, THEY USE PAINFUL WORDS, LIKE TELLING PEOPLE ABOUT MY STATUS." (24 y.o. registered user from Limpopo)	"Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility."
Abandonment	2	"My baby died hours after delivery. Because I was left in Labour for three days my baby got tired and died I asked for C-section doctors refused" (Unregistered user)	"Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility."
Autonomy	2	"[the nurses] ddnt respect us they harass us and force us to do things we don't want to do" (Unregistered user)	"Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility."
Birth companion	1	"MAY I ASK WHY ARE GOVERNMENT HOSPITALS NOT ALLOWING FAMILY MEMBERS DURING LABOUR" (40 y.o. registered user from Gauteng)	"It depends on the structure, if there is other people in labour the same time it poses a challenge for the privacy of the next patient. In general women are allowed to have one family member with them during labour."
Unknown	1	"I was mistreated before, during and after delivering my baby by [removed]" (29 y.o. registered user from Eastern Cape)	"Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility."
Physical abuse	0	NA	NA
Bribes	0	NA	NA
Poor service	19	"The clinic is too small we don't have enough room for pregnant, and new born babies & family planning. We don't exercise, no enough nurses to assist"	"Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility."

NA, not applicable.

Table 3 Observed response appropriateness ratings for facility-based mistreatment, partner and family violence, and the random sample along with their expected values assuming independence of factors

	Facility-based mistreatment and service	Partner and family violence	Random sample
Optimal, observed (expected)	38 (41.4)	34 (31.6)	391 (390.0)
Suboptimal, observed (expected)	13 (8.6)	3 (6.6)	80 (80.9)
Incorrect, observed (expected)	0 (1.1)	2 (0.8)	10 (10.1)

former were categorised as facility-based mistreatment and the latter as partner and family violence. All messages identified could be reliably assigned to one of these categories, so no other categories were needed. Additionally, all messages labelled as a complaint by helpdesk staff were manually examined for the presence of facility-based mistreatment or partner and family violence. Those categorised as facility-based mistreatment were assigned to one of the typologies in table 1.

Out of the full message set (n=65 768), 32 reports of facility-based mistreatment and 39 reports of partner or family violence were identified. A further breakdown of these counts along with illustrative messages and responses is given in table 2. Additionally, 19 messages reporting poor service at facilities were found.

Our ability to detect these reports of mistreatment and partner and family violence via keyword supports the feasibility of flagging high-priority messages. When the keywords in table 1 are present, it is likely that the message pertains to mistreatment. This finding could not be more rigorously evaluated using a classifier in the current work, partly because rates of mistreatment reporting were low, but more importantly because there was no 'ground truth' regarding mistreatment. In other words, we had no independent method of verifying whether mistreatment was reported in the 65 768 messages available. Nevertheless, this result provides preliminary evidence that keyword-matching can form the basis of a message triage system.

Message handling: mistreatment of women

In the majority of the facility-based mistreatment cases (72%; 23 of 32), the message was acknowledged as a complaint using the following reply: "Thank you for sending in your complaint. We have taken note of it and will log the complaint with the Department of Health and your facility." In three cases, the helpdesk directed the woman back to the clinic where she had been mistreated.

Response appropriateness was compared between both facility-based mistreatment and partner and family violence and the random sample (n=481) via χ^2 test; both message sets were found to be similar to the random sample. There appears to be a slight but not significant trend towards better handling of partner and family violence, with 34 of 39 scored as optimal (87.2%) compared with 391 of 481 (81.3%) in the random sample (P=0.182); and poorer handling of complaints about facilities (poor service or facility-based mistreatment), with 38 of 51 (74.5%) scored as optimal (P=0.183). Table 3 provides counts of the scores in each message set, along with their expected values calculated during χ^2 testing. The latter shows counts that would be expected if response handling were the same between sets.

Response time was also compared between both facility-based mistreatment and partner and family violence and the random sample via Kolmogorov-Smirnov test. Figure 2 shows the distribution of response time between these three groups. Median response times were

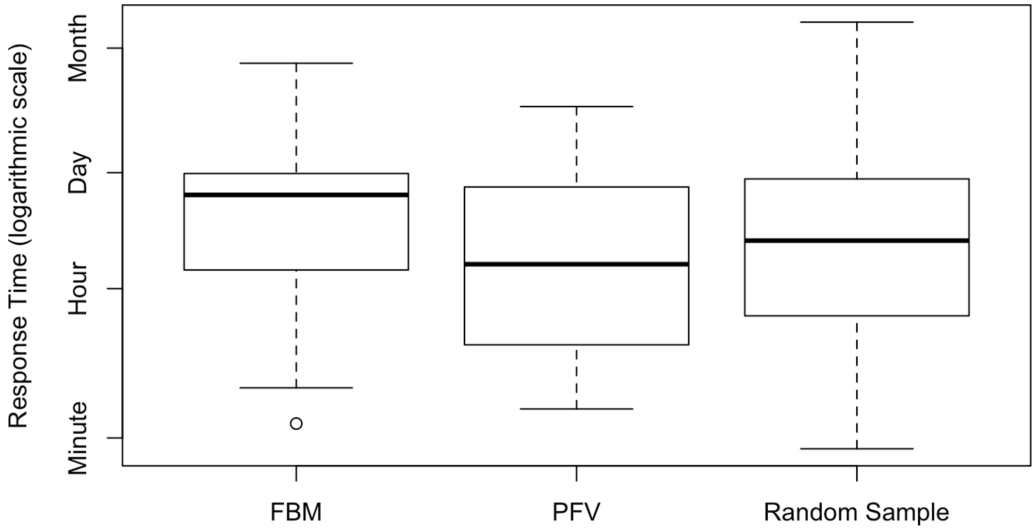

Figure 2 Distribution of helpdesk response time between three groups: facility-based mistreatment (FBM), partner and family violence (PFV), and the random sample.

13.1 hours for facility-based mistreatment, 2.0 hours for partner and family violence, and 3.7 hours in the random sample. The differences in response time between facility-based complaints and the random sample did not reach statistical significance (P=0.074). Response time was similar between partner and family violence and the random sample (P=0.566).

Age was tested for normality by Shapiro-Wilk test, then compared pairwise between groups by Mann-Whitney U test. This analysis showed that age was not normally distributed in our sample (P<0.001) and was similar between mothers from the random sample and those with facility-based complaints (P=0.267) and experiencing partner and family violence (P=0.516). Differences among provinces were compared by χ^2 test, but no statistically significant association between mistreatment reporting and province was detected (P=0.169).

The need for triage is supported by the presence of these reports of mistreatment as well as their handling by helpdesk staff. While these messages are uncommon, proper identification and handling is critical to provide the best possible support for helpdesk users and also for the purposes of health sector oversight and accountability. Results did not conclusively demonstrate that handling of these messages is inferior to the handling of others, but neither is it superior. With a median response time of 8.6 hours and 18 non-optimal responses among the 90 total reports of mistreatment, there is certainly room to improve. Given the large volume of messages received by the limited number of helpdesk staff, triage is needed to improve this performance.

Limitations

Our keyword-based approach, while appropriate under the circumstances, is not guaranteed to identify all messages reporting mistreatment. The list of keywords is not exhaustive and does not account for misspellings and other typos or character errors. Manual inspection could have been used to more reliably identify mistreatment, but was prohibitively labour intensive on a message set of this size. Further, the keywords are all in English; a major limitation was our inability to identify mistreatment not reported in English in this analysis.

Messages reporting mistreatment were limited in number, with 32 facility-based mistreatment and 39 partner and family violence messages identified. This limited sample size prevented us from drawing definitive conclusions about the handling of these messages (response time and appropriateness) as well as the demographic determinants of mistreatment (age and province). While several trends have been identified, they did not reach statistical significance.

EXPLORING AUTOMATED MESSAGE TRIAGE

Classifying text via naïve Bayes

To explore the feasibility of automated triage, a naïve Bayes classifier was trained to label incoming messages. Naïve Bayes is a simple, scalable and robust classifier with a long history of use in text classification.[11] In brief, naïve Bayes is a probabilistic classifier that simplifies the likelihood of a given class using the naïve Bayes assumption, which stipulates that features—in this case, individual word occurrences—are conditionally independent given the class label. This assumption may be written as follows:

$$P(W_1, \ldots, W_n | C) = \prod_{i=1}^{n} P(W_i | C)$$

In this equation, the W_i represent counts of the n possible words of interest, and C is the class label—in our case, the message label. Having made this assumption, the terms $P(W_i|C)$ may be calculated simply as the frequencies of word occurrences for each of the labels. New messages may then be classified based on Bayes' rule as:

$$\underset{i}{\operatorname{argmax}} \left(P(C = c_i) \prod_{i=1}^{n} P(W_i | C = c_i) \right)$$

For this application, the message labels C belong to one of the following 10 categories: 'Question', 'Message Switch', 'Compliment', 'PMTCT', 'Opt Out', 'Complaint', 'Language Switch', 'Spam' and 'Unable to Assist'. Words are counted in each message, making ours a multinomial model.[11] In addition to its simplicity and history of success in similar applications, the naïve Bayes approach is equally simple to implement in a multiple-language setting, giving it an edge over newer, more sophisticated classifiers for the current application. Indeed, naïve Bayes should perform equally well even when multiple languages are present in a single message, which is common in this dataset.

The 65 768 available, labelled helpdesk messages were divided into a training set (n=53 232) used to train the classifier and a test set (n=12 536) used to evaluate its performance. Individual words were identified by their word stems; for example, 'training' and 'train' count as the same word. The classifier was trained to replicate the labels used by NDoH staff based on the word stems present in a message. Following training, performance was evaluated on the test set using a confusion matrix of message labels correctly and incorrectly predicted by the classifier. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were then calculated for each label.

Classifier performance

Overall accuracy for the naïve Bayes classifier was 85.4%, with 95% CI ranging from 84.8% to 86.0%, and Cohen's kappa coefficient equal to 0.556. Classification accuracy was conclusively superior to the no-information rate of 82.4% ($P<10^{-10}$). The confusion matrix for the naïve Bayes classifier is presented in table 4; this matrix summarises performance by cross-tabulating NDoH-assigned labels with our classifier's predictions. Sensitivity, specificity, PPV and NPV for each label are presented in table 5.

Table 4 Confusion matrix for naïve Bayes classifier applied to helpdesk queries

		Predicted label									
		Question	Message Switch	Compliment	PMTCT	Opt Out	Complaint	Language Switch	Spam	Unable to Assist	False negatives
Assigned label	Question	9374	202	125	270	108	60	90	54	37	946
	Message Switch	304	489	26	2	6	1	0	0	0	339
	Compliment	103	43	438	5	8	4	5	1	0	169
	PMTCT	164	7	3	307	3	2	1	2	1	183
	Opt Out	51	18	4	3	29	0	1	0	1	78
	Complaint	43	2	11	0	1	14	0	1	0	58
	Language Switch	24	0	5	0	0	0	48	0	0	29
	Spam	17	0	0	0	1	0	0	0	0	18
	Unable to Assist	6	0	1	0	0	0	0	0	0	7
	False positives	712	272	175	280	127	67	97	58	39	

PMTCT, prevention of mother-to-child transmission of HIV.

Table 5 Sensitivity, specificity, and positive and negative predictive values of naïve Bayes classifier on test queries (25% of all queries fielded between November 2016 and June 2017)

Category	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Question	0.91	0.68	0.93	0.61
Message Switch	0.59	0.98	0.64	0.97
Compliment	0.72	0.99	0.71	0.99
PMTCT	0.63	0.98	0.52	0.98
Opt Out	0.27	0.99	0.19	0.99
Complaint	0.19	0.99	0.17	1.00
Language Switch	0.62	0.99	0.33	1.00
Spam	0.00	1.00	0.00	1.00
Unable to Assist	0.00	1.00	0.00	1.00

Specificity was lowest for 'Question', the most common label, but ≥98% for all other labels. Similarly, NPV was ≥97% for all labels other than 'Question'. Sensitivity and PPV were less consistent. Questions, message switch requests, compliments and PMTCT had sensitivity and PPV both >50% and as high as 93%. On the other hand, none of the small number of spam and unable-to-assist messages were correctly identified by the classifier—most were labelled as 'Question', as shown in table 4—resulting in sensitivity and PPV of 0 for those labels.

These results—along with successful identification of mistreatment—strongly suggest that helpdesk triage is feasible. As hypothesised, the naïve Bayes classifier had high sensitivity for all labels except the most common one, 'Question'. While higher sensitivity is desirable, high specificity is most essential for the purposes of triage: it ensures that staff would not be inundated with messages that have been incorrectly labelled as high priority. When specificity is low, high-priority labels would become less useful to helpdesk staff due to comparatively larger number of false positives. Varying sensitivity between labels implies that in some cases, messages deserving high-priority status will not be successfully identified; however, this is no worse than the current system, under which no messages are flagged. PPV is fairly high for several labels other than 'Question', showing that when these labels are assigned, they are meaningful.

More generally, these results add to a large body of evidence supporting use of the naïve Bayes classifier in text classification[11] and demonstrate that a 'bag of words' approach is effective for SMS labelling and triage. As the phrase 'bag of words' implies, this approach considers messages only as a set of their constituent words—or in this case, word stems—without considering syntax or word order. Similar to the keyword-matching used to identify mistreatment, naïve Bayes relies on the presence or absence of specific indicator words to guide classification. If a message contains the word 'beat' or 'thank', for instance, it is likely to be a report of violence or a compliment, respectively.

Because it does not rely on word order or syntax, this method of classification can perform well in any language or multiple languages. This is an important advantage in the current setting. However, good performance across the 11 languages supported by the helpdesk would require adequate training data in each of them, which could prove difficult to obtain. More generally, this method can be used in any two-way communication platform able to consistently identify individual words.

Taken together, these analyses show that the handling of high-priority messages could be improved, and that a simple classifier is capable of automatically flagging important but infrequently occurring messages. Thus, automated triage could feasibly be used to identify such messages for priority handling by NDoH staff, likely leading to meaningful helpdesk improvement.

Limitations

Our method does not take advantage of word order or syntax, but instead understands each message as a 'bag of words'. While we have cited this as an advantage in the current application, particularly given the multiple languages present in our message set, a syntax-aware approach might yield better results. The conditional independence assumption of naïve Bayes is another well-known limitation, but one that has not hindered its effective use in many text classification applications.

Importantly, while our results support naïve Bayes as the basis for automated triage, training such a system would first require manual identification of high-priority topics among a large set of messages. This burden could be reduced to some degree using expert knowledge and/or active learning.

As an exploratory analysis, this work is only a first step in the development of automated triage, which would require continued development of a triage system; its integration into the helpdesk platform; usability testing to integrate the triage process into existing helpdesk workflows; and ultimately an evaluation of its effectiveness via prospective trial. Nevertheless, the current analysis reveals that automated triage is a promising means to streamline and improve the helpdesk even as the user-base continues to grow. This in turn helps to ensure the helpdesk remains an effective mechanism for

empowering and mobilising women and for improving the quality of care provided and promoting health systems' accountability.

CONCLUSION

The helpdesk response to a specific, high-priority topic—the mistreatment of women—is no better than its response to the average incoming message. This is understandable given the high volume of messages faced by NDoH staff on a daily basis, but nevertheless, response time and content stand to be improved, particularly for uncommon but important topics. An automated triage system, one that sorts incoming messages by priority, is one method for targeted improvement of helpdesk responses. This work provides preliminary but important support for such a system by demonstrating that (1) mistreatment can be identified by keyword and (2) less-common message labels can be automatically identified with high specificity. Thus, automated triage appears to be feasible, and it could be effective. As the helpdesk continues to grow, it is important to benchmark its performance and explore opportunities for improvement. Automated triage could improve overall quality of message handling yet reduce burden on NDoH staff, boosting its overall effectiveness as an accountability mechanism, an information gathering platform and a resource for women receiving health services in South Africa.

Acknowledgements The authors would like to acknowledge Jane Sebidi, Kate Wilson, Debbie Rogers and Marcha Bekker for their ongoing support of MomConnect, the helpdesk and associated research. The support provided by John Snow, Inc. (JSI) in the President's Emergency Plan for AIDS Relief (PEPFAR) and United States Agency for International Development (USAID)-funded MEASURE Evaluation Strategic Information for South Africa (MEval-SIFSA) project, and the Bill & Melinda Gates Foundation to enable this publication is acknowledged with gratitude.

Contributors All authors conceived and designed the analysis. ME, CC and AEL drafted the manuscript. JW and CC oversaw the data collection and processing. ME and CC analysed the data. YP and PB oversaw the project. All authors revised the manuscript.

Funding This work, along with the MomConnect programme, was funded primarily by the Department of Health, Republic of South Africa.

Competing interests None declared.

Ethics approval Please refer to other MomConnect supplement papers for details of ethical approval.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. National Department of Health (NDoH), Statistics South Africa (Stats SA), South African Medical Research Council (SAMRC), and ICF. *South Africa demographic and health survey 2016: key indicators*. Pretoria, South Africa, and Rockville, Maryland, USA, 2017.
2. Institute for International Programs. *Countdown to 2015*: South Africa Country Profile Baltimore, Maryland Johns Hopkins School of Public Health, 2015.
3. Bohren MA, Vogel JP, Hunter EC, *et al*. The mistreatment of women during childbirth in health facilities globally: a mixed-methods systematic review. *PLoS Med* 2015;12:e1001847.
4. Jewkes R, Abrahams N, Mvo Z. Why do nurses abuse patients? Reflections from South African obstetric services. *Soc Sci Med* 1998;47:1781–95.
5. Brown H, Hofmeyr GJ, Nikodem VC, *et al*. Promoting childbirth companions in South Africa: a randomised pilot study. *BMC Med* 2007;5:7.
6. Chadwick RJ, Cooper D, Harries J. Narratives of distress about birth in South African public maternity settings: a qualitative study. *Midwifery* 2014;30:862–8.
7. Barron P, Pillay Y, Fernandes A, *et al*. The MomConnect mHealth initiative in South Africa: early impact on the supply side of MCH services. *J Public Health Policy* 2016;37(Suppl 2):201–12.
8. Freedman LP, Ramsey K, Abuya T, *et al*. Defining disrespect and abuse of women in childbirth: a research, policy and rights agenda. *Bull World Health Organ* 2014;92:915–7.
9. Jewkes R, Penn-Kekana L. Mistreatment of women in childbirth: time for action on this important dimension of violence against women. *PLoS Med* 2015;12:e1001849.
10. Lambert J, Etsane E, van den Broek N, *et al*. *'Think they are going to handle me like a queen but they didn't': key questions regarding identified regarding quality of care in urban and rural setting in RSA*. Pretoria, South Africa: Centre for Maternal and Newborn Health, Liverpool School of Tropical Medicine, SAMRC Unit for Maternal and Infant Health Care Strategies, University of Pretoria.
11. McCallum A, Nigam K. *A comparison of event models for naive Bayes text classification*, 1998.