Performance of a new guideline for telephone triage in out-of-hours services in Belgium : a pilot study using simulated patients

Performance of a new guideline for telephone triage in out-of-hours services in Belgium: a pilot study using simulated patients

1 Abstract

Background

Patients in Belgium needing out-of-hours care have two options: the emergency department (ED) or the general practitioner (GP) on call often organised in a general practitioner cooperative (GPC). Currently there is no triage system so patients do not know where to go.

Methods

The 1733-guideline is a new version of an existing telephone guideline containing protocols for 40 common health problems. This study aims to be a first approach in its validation. Ten clinical vignettes were presented to 12 operators in a simulated phone call. The operators had to assign a protocol, urgency level and resource to dispatch (ambulance, GP house visit, etc.) to each case.

Results

Hundred-twenty phone calls were analysed. The operators chose the right protocol in 69% and the correct urgency level in 35% of the cases. The proportion of under- and over-triage was 26% and 39% respectively. There was important variation in between the operators. The sensitivity for detecting high urgent cases was 0.42, the specificity 0.92.

Conclusion

Using the new Belgian 1733 guideline for telephone triage, operators mostly chose the appropriate protocol but only chose the correct urgency in 1 out of 3 cases. In this phase of development, the studied telephone guideline is not ready for implementation.

Word count of the abstract only: 200

2 Introduction

From the nineties, general practitioner cooperatives (GPC) were established in many European countries, as a new alternative for the organisation of out-of-hours (OOH) medical care by general practitioners (GPs). Although we do not have a clear-cut definition of 'appropriate use' or, 'inappropriate use' of the emergency department (ED), it has been argued that many medical problems presented at the ED could easily be managed in a primary care setting.[1-3] One objective therefore may be to redirect patients from secondary to primary care. In Belgium, patients have a free choice to go to the ED or GPC. Previous research revealed that patients find it difficult to make this decision themselves.[4] Co-location of a GPC and an ED does not necessarily decrease the caseload at the ED.[5] Triage is the sorting out and classification of patients or casualties to determine priority of need and proper place of treatment. In physical triage, a health care professional at the ED or at the common gate of an ED and GPC makes a quick assessment of the patient and then allocates the patient to the most appropriate service. This can lead to efficient redirection of self-referrals.[6] In telephone triage, the patient calls a hotline and gets advice about the most appropriate service by a nurse, paramedic or doctor. When a patient's urgency is underestimated, the decision is considered under-triage. This might lead to increased morbidity and mortality. Overestimation of the urgency is called over-triage and leads to inefficiency because expensive resources are inappropriately dispatched. In other countries, research after implementation of telephone triage showed it is efficient but possibly not safe especially for high urgent cases.[7] Consequently, research is necessary before implementing new triage tools.

Our primary objective was to examine the ability of a newly developed telephone guideline, called the 1733 guideline, to adequately estimate the urgency of health problems presented by simulated patients. An adequate telephone triage decision consists of three steps: choosing the correct protocol (e.g. headache), assigning the right urgency level (e.g. very urgent) and dispatching the according resource (e.g. ambulance). The interrater agreement when using the protocols was the secondary objective.

3 Methods

Used materials

We study the 1733 guideline, named after the telephone number patients will have to call in the near future to have access to OOH care. It is based upon the historically used guideline for ambulance dispatch in Belgium, which has never been validated, but are being used for many years now. A working group consisting of GPs, emergency physicians and staff of the ambulance dispatch services adjusted this old guideline for the incorporation of primary care. Three new urgency levels were introduced: urgent GP, standard GP and standard non-urgent care. The development procedure of the 1733 guideline is described elsewhere.[8] It consists of 40 protocols, each for a specific patient presentation. A protocol consists of a table with six urgency levels (U1-U6). Each urgency level contains a number of discriminators. The operator is supposed to check the presence of these discriminators in a top down order. Each urgency level has a corresponding resource (see table 1). Dispatchers are allowed to deploy another resource for reasons not described in the protocol (mostly psychosocial concerns or practical considerations).

Pairs of a GP and an emergency physician wrote 28 clinical vignettes. Afterwards the head of the dispatch centre reviewed and adjusted the vignettes when necessary. The entire working group unanimously decided on the correct outcome for each vignette (gold standard). We piloted all of

these vignettes. Those leading to unclear answers or ambiguous interpretations were left out, leading to 10 selected vignettes (see table 2). Operators received a hard copy of the manual and a brief training on how to use it.

Study design

We performed a single centre prospective study using simulated patients, on April 20th, 2017. Three GPs (including author HP) simulated each three to four cases. All 13 operators working at the studied dispatch centre and available during the study period participated. They sat at their normal working station and were informed about the test situation. They noted the chosen protocol, urgency level and resource to dispatch on a spreadsheet with drop down menus (Microsoft Excel 2016). For each answer they added how confident they were in there answers on a scale from one to ten.

Ethics

The medical research ethics committee of the University of Leuven (number S59385) approved this study as part of a larger approval. All participating operators signed an informed consent.

Analysis

Concordance with the golden standard, standard deviations and sensitivity/specificity were calculated using Microsoft Excel 2016.

We calculated Interrater agreement in R with the Various Coefficients of Interrater Reliability and Agreement (IRR) package.

4 Results

We excluded one out of thirteen operators because of a lack of cooperation with the researchers. The simulated patients made 120 telephone calls. Table 3 shows the characteristics of the operators.

The operators chose in 83 (69%) cases the right protocol. In six out of 10 vignettes, at least three out of every four operators chose correctly. For the remaining four vignettes, the correct percentage ranged from zero to 69%. In case of a wrong answer, the operators mostly chose logical alternatives. For the case of a chemical burn to the forearm for example, the working group proposed the protocol “problems of the extremities” while all operators choose “exposure to chemical substances”. There were only four entirely wrong answers (<1%), all of them in case four (Intoxication with mushrooms and leaves).

The operators estimated the urgency correctly in 42 (35%) cases. Among the 77 wrong estimates, there was a difference of one urgency level in 49 cases (42%), two categories in twenty cases (17%) and three categories in eight cases (7%). In total, they overestimated the urgency in 46 cases (39%) and underestimated it in 33 cases (26%). There was significant variation between the operators (see table 4): one operator did not make any underestimation whereas two others made an underestimation of the urgency in four out of eight cases. The variation among the cases was similar: from zero to 58% of correct triage. The interrater agreement was moderate (Kendall’s W 0.57).

The capacity to discriminate potentially life-threatening cases from less urgent cases was examined by creating a dichotomy between U1 and U2 and the other categories. In the two vignettes with a correct solution of U1 or U2, the operators chose U1 or U2 in ten out of twenty-four answers (sensitivity 0.42). In the other vignettes, the operators chose U1 or U2 in eight out of ninety-five cases (specificity 0.92).

Finally, the chosen resource to deploy was studied. Compared to the gold standard the operators deploy the correct resource in 45-five cases (38%). Among the 75 wrong estimates, there was a difference of one category in 47 cases (39%), two categories in 22 cases (18%), three categories in five cases (4%) and four categories in one case (1%). There was over-triage in 40% and under-triage in the remaining 23%. The interrater agreement was moderate (Kendall's W 0.59).

In 101 cases (84%), the operators picked the resource corresponding to the chosen urgency level. In 13 cases (10%) this resource was higher (more urgent or upscaling) and in six cases (5%) it was lower (less urgent or downscaling).

The average certainty scores per operator were 8.5 (SD 0.93) for protocol, 7.5 (SD 0.77) for urgency level and 7.3 (SD 1.50) for resource. The scores per vignette showed similar averages and standard deviations.

5 Discussion

In this study, we examined whether the newly developed 1733 guideline used by operators adequately estimates the urgency of health problems presented by simulated patients. The operators mostly chose the appropriate protocol or an acceptable alternative. A low concordance with the gold standard was found for the urgency level. For the detection of the most urgent cases, this study reveals a low sensitivity and an acceptable specificity. This lower accuracy for high urgent cases is similar to current literature.[9]

Although operators often deployed another resource than the one corresponding to the chosen urgency level, this did not improve their triage accuracy.

In a systematic review the percentage of safe performance in high urgent cases was 46% which was considered unsafe[7] , similar to the 42% in this study. A summary of this review and two more recent studies[10, 11] can be found in table 5. The proportion of correct decisions in this study is the lowest among these comparable studies, mainly because of the high proportion of overtriage. In only one similar article, the authors made a positive conclusion about safety with better results than in the this study.[10] None of these comparable studies used the same methodology as we did: all either used mystery patients (simulated patients who call the operators unexpectedly during routine clinical work) or written case scenarios.

A recent study about a Dutch guideline for daytime practice used by practice assistants tested written case scenarios. The authors found correct triage in 64%, under-triage in 17% and over-triage in 19%.[9] These figures are significantly better than the results of this study (38%, 40% and 23% respectively). Most telephone triage systems use a four or five point scale whereas we have studied a six point scale. The more points, the more options an operator will have to make a mistake.

Although currently there is no agreement on the most appropriate statistics to assess interrater agreement in triage, the moderate interrater agreement presented here is not good enough.[12] We found high certainty scores indicating that the operators do not experience much doubt in their decisions. Possibly they feel acquainted to the new protocols as they are build up in a same chart as they are used to and for that reason, they might not be aware of the huge importance of adding new resources like the general practitioner. Alternatively, the protocols might be clearly written but not correctly in terms of content.

We believe a combination of several factors causes our unsatisfactory results. Firstly, the operators might not have been prepared well enough for their complex task: they did not have any experience with the 1733 guideline and only received a brief training of half a day. Secondly, the protocols might

be insufficient. They are written in an ambiguous way: not all discriminators clearly lead to a specific urgency level and the same presentation will give different urgency categories depending on the chosen protocol. Protocols often contain discriminators that are difficult to interpret (e.g. structures with “and”, “or”, “not”). Finally, the way the vignettes were designed or played by the simulated patients might have confused the operators. The extent of the contribution of all these factors cannot be determined using the current study.

This study was the first in Belgium to validate a triage instrument and thus provides interesting insights. Unique in this study is that it we did not only study the urgency level but also the dispatched resource and certainty scores. The results are very important for the organisation of OOH triage in Belgium but also in other countries using not yet validated telephone triage guidelines. When interpreting the results of this study it is important to consider that most researchers validate triage systems after implementation. This study proves it is also possible to study a new triage system in laboratory circumstances.

This study has some limitations. The operators, simulated patients and researchers were not blinded. We obtained a relatively small sample of operators, all but one from the same dispatch centre. The current sample was too small to assess the importance of training, experience and educational level. In further research, we need to assess these characteristics as they might explain the moderate interrater variability. The protocols were developed in an unstandardized way.[8] GPs took the role of simulated patients because they have experience with the clinical presentation of the chosen vignettes. They might cause a bias by trying to direct the operator in the right direction.

This version of the 1733 guideline is not yet ready for clinical implementation. Ideally, it should be improved by using the current study and afterwards perform a new study on simulated patients. Keeping in mind its shortcomings, the Belgian government has chosen to start implementation with an improved but unvalidated guideline, as happened in most countries. Permanent evaluation of the 1733 guideline and performance of the operators will be deployed.

6 Conclusion

It is feasible and useful to study telephone triage guidelines before implementation. Using the Belgian 1733 guideline for telephone triage, operators mostly chose the appropriate protocol but only dispatch the correct resource in 38% of the cases, which is lower than in similar studies. The studied telephone guidelines are not ready for implementation in this phase of the development.

7 Literature

1. Kraaijvanger N, Rijpsma D, van Leeuwen H, van Dijk N and Edwards M. Self-referrals in a Dutch Emergency Department: how appropriate are they? *Eur J Emerg Med*. 2016; 23: 194-202.
2. Derlet RW and Ledesma A. How do prudent laypeople define an emergency medical condition? *The Journal of emergency medicine*. 1999; 17: 413-8.
3. Carret ML, Fassa AG and Kawachi I. Demand for emergency health service: factors associated with inappropriate use. *BMC health services research*. 2007; 7: 131.
4. Philips H, Mahr D, Remmen R, Weverbergh M, De Graeve D and Van Royen P. Experience: the most critical factor in choosing after-hours medical care. *Quality & safety in health care*. 2010; 19: e3.
5. Philips H, Remmen R, Van Royen P, et al. What's the effect of the implementation of general practitioner cooperatives on caseload? Prospective intervention study on primary and secondary care. *BMC health services research*. 2010; 10: 222.
6. van Veelen MJ, van den Brand CL, Reijnen R and van der Linden MC. Effects of a general practitioner cooperative co-located with an emergency department on patient throughput. *World J Emerg Med*. 2016; 7: 270-3.
7. Huibers L, Smits M, Renaud V, Giesen P and Wensing M. Safety of telephone triage in out-of-hours care: a systematic review. *Scand J Prim Health Care*. 2011; 29: 198-209.
8. Van der Mullen C, Quintens H, Van Baelen S, Crits T, Wuyts J and Sabbe M. De patiënt met een niet-planbare zorgvraag naar het gepaste zorgniveau verwijzen: nieuwe 112-1733 geïntegreerde telefonische triage- en regulatieprotocollen. *Tijdschirft voor geneeskunde*. 2017; 73: 7.
9. Smits M, Hanssen S, Huibers L and Giesen P. Telephone triage in general practices: A written case scenario study in the Netherlands. *Scand J Prim Health Care*. 2016; 34: 28-36.
10. Hansen EH and Hunskaar S. Telephone triage by nurses in primary care out-of-hours services in Norway: an evaluation study based on written case scenarios. *BMJ Qual Saf*. 2011; 20: 390-6.
11. Pasini A, Rigon G and Vaona A. A cross-sectional study of the quality of telephone triage in a primary care out-of-hours service. *J Telemed Telecare*. 2015; 21: 68-72.
12. van der Wulp I and van Stel HF. Calculating kappas from adjusted data improved the comparability of the reliability of triage systems: a comparative study. *Journal of clinical epidemiology*. 2010; 63: 1256-63.
13. Moriarty H, McLeod D and Dowell A. Mystery shopping in health service evaluation. *Br J Gen Pract*. 2003; 53: 942-6.
14. Giesen P, Ferwerda R, Tijssen R, et al. Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency? *Quality & safety in health care*. 2007; 16: 181-4.
15. Derkx HP, Rethans JJ, Muijtjens AM, et al. Quality of clinical aspects of call handling at Dutch out of hours centres: cross sectional national study. *BMJ*. 2008; 337: a1264.

Table 1. Urgency levels and corresponding resources

Level	Description	Recommended resource
U1	Immediate life threatening	Ambulance with a doctor and nurse
U2	Possibly life threatening, fast evolution expected	Ambulance with nurse
U3	Not life threatening, admission probably necessary	Ambulance with paramedics
U4	Not life threatening, admission probably not necessary	Urgent GP (within one hour)
U5	In need of less urgent care	Standard GP on call (within 12 hours)
U6	Does not need medical care at this moment	Refer patient to standard non urgent care

Table 2. Vignettes

Caller	Complaint	Correct urgency class	Correct protocol (%)	Correct urgency level (%)
1. Granddaughter of older female	My grandmother remains in bed and can no longer speak.	U4	100	58
2. Spouse of middle aged male	Severe abdominal pain	U1	92	42
3. Mother of baby	Heat stroke with dehydration	U2	31	17
4. Mother of toddler	Intoxication with mushrooms and leaves	U4	69	58
5. Adult male	Chest pain, hyperventilation	U6	77	9
6. Middle aged male	Trauma with short term loss of consciousness	U5	54	50
7. Adult male	Chemical burn of forearms	U4	0	0
8. Middle aged female	Shortness of breath, long term psychosocial problems	U5	92	8
9. Adolescent female	Contusion of the head	U6	77	58
10. Adult female	Suicidal thoughts	U4	100	50

Table 3. Characteristics of the operators

Average age	37 years (range 30-51)
Average experience as an operator	7 years (range 1-19)
Gender	7 (58%) female, 5 (42%) male
Educational level*	6 (55%) bachelor or master, 5 (45%) high school

*: one missing value

Table 4. Variation among the operators for urgency level (N cases: 120): Number of cases per operator triaged correctly, over- or under triaged.

Operator:	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Correct triage	2	5	7	3	5	3	3	3	3	5	3	3
Possible over-triage	7	2	3	6	2	3	3	4	4	4	4	6
Possible under-triage	1	3	0	1	3	4	4	3	3	1	3	1

Table 5. Comparison to the current literature

First Author	Year of publication	setting	Study design	Triagist's background	Number of scale levels	Total number of cases	% Correct triage	% Undertriage	% Overtriage	Author's conlusion regarding safety
Morreel	2019	Out-of-hours call center	10 simulated patients	paramedics	6	120	35	26	39	Unsafe
Moriarty[13]	2003	Primary care telephone triage pilot service	4 mystery patients with need to refer (feasibility study)	nurses	referral or not	85	51	49	N/A	None
Giesen[14]	2007	GPC (regional telephone number)	5 mystery patients, 20 cases	nurses	4	352	69	19	12,5	Potentially unsafe
Derkx[15]	2008	GPC (regional telephone number)	7 mystery patients, no U1 or U5	nurses	3	357	58	41	1	Unsafe
Hansen[10]	2011	OOH center	20 written scenario's	nurses	3	1620	70	12	18	Safe
Pasini[11]	2015	OOH call center	4 mystery patients	GPs	2 (delay of care yes or no)	40	93	N/A	8	None