



# BMJ Open Factors associated with undertriage and overtriage in telephone triage in Danish out-of-hours primary care: a natural quasi-experimental cross-sectional study of randomly selected and high-risk calls

Dennis Schou Graversen <sup>1,2</sup>, Anette Fischer Pedersen <sup>1,3</sup>,  
Morten Bondo Christensen,<sup>1,2</sup> Fredrik Folke,<sup>4,5</sup> L Huibers<sup>6</sup>

**To cite:** Graversen DS, Pedersen AF, Christensen MB, *et al*. Factors associated with undertriage and overtriage in telephone triage in Danish out-of-hours primary care: a natural quasi-experimental cross-sectional study of randomly selected and high-risk calls. *BMJ Open* 2023;**13**:e064999. doi:10.1136/bmjopen-2022-064999

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-064999>).

Received 26 May 2022  
Accepted 27 February 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dennis Schou Graversen;  
d.graversen@ph.au.dk

## ABSTRACT

**Objectives** We aim to explore undertriage and overtriage in a high-risk patient population and explore patient characteristics and call characteristics associated with undertriage and overtriage in both randomly selected and in high-risk telephone calls to out-of-hours primary care (OOH-PC).

**Design** Natural quasi-experimental cross-sectional study.  
**Setting** Two Danish OOH-PC services using different telephone triage models: a general practitioner cooperative with GP-led triage and the medical helpline 1813 with computerised decision support system-guided nurse-led triage.

**Participants** We included audio-recorded telephone triage calls from 2016: 806 random calls and 405 high-risk calls (defined as patients  $\geq 30$  years calling with abdominal pain).

**Main outcome measures** Twenty-four experienced physicians used a validated assessment tool to assess the accuracy of triage. We calculated the relative risk (RR) for *clinically relevant* undertriage and overtriage for a range of patient characteristics and call characteristics.

**Results** We included 806 randomly selected calls (44 *clinically relevant* undertriaged and 54 *clinically relevant* overtriaged) and 405 high-risk calls (32 undertriaged and 24 overtriaged). In high-risk calls, nurse-led triage was associated with significantly less undertriage (RR: 0.47, 95% CI 0.23 to 0.97) and more overtriage (RR: 3.93, 95% CI 1.50 to 10.33) compared with GP-led triage. In high-risk calls, the risk of undertriage was significantly higher for calls during nighttime (RR: 2.1, 95% CI 1.05 to 4.07). Undertriage tended to be more likely for calls concerning patients  $\geq 60$  years compared with 30–59 years (11.3% vs 6.3%) in high-risk calls. However, this result was not significant.

**Conclusion** Nurse-led triage was associated with less undertriage and more overtriage compared with GP-led triage in high-risk calls. This study may suggest that to minimise undertriage, the triage professionals should pay extra attention when a call occurs during nighttime or concerns elderly. However, this needs confirmation in future studies.

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ A strength of our study is the natural quasi-experimental design using real-life calls as opposed to a constructed setup.
- ⇒ To our knowledge, this is the first study to explore factors associated with undertriage and overtriage in real audio-recorded calls.
- ⇒ A limitation is the use of only one assessor per call, but acceptable interrater agreement was previously found.

## INTRODUCTION

In out-of-hours primary care (OOH-PC), telephone triage plays a pivotal role in managing patient flows and workload.<sup>1–3</sup> Telephone triage aims to ensure a safe and efficient delivery of healthcare, avoiding undertriage and minimising overtriage.<sup>4</sup> However, accurate telephone triage is difficult due to the lack of visual cues of the patient, challenges in telecommunication and time pressure.<sup>5–8</sup>

Safety and efficiency of telephone triage in OOH-PC have been explored in a range of studies that used varying outcome measures.<sup>4,9–13</sup> Studies have identified a range of risk factors for potentially unsafe telephone triage in out-of-hours (OOH) care: calls concerning abdominal pain,<sup>14–20</sup> chest pain<sup>16,17</sup> and shortness of breath,<sup>16,18</sup> calls for patients with increasing age<sup>14,17,21</sup> and calls during nighttime.<sup>17,22</sup> Thus, these calls can be seen as potential high-risk calls. A few studies have explored factors associated with overtriage as a measure of inefficient telephone triage. Nurse-led triage<sup>1,23</sup> and triage aided by computerised decision support system (CDSS)<sup>23–26</sup> have been associated with overtriage. However, studies that explored factors associated with undertriage and overtriage

used a range of study designs. None of these studies used real audio-recorded telephone calls to assess the risk of undertriage and overtriage and related risk factors.

Previously, we found that telephone triage was associated with significantly less undertriage and significantly more overtriage for nurse-led triage compared with GP-led triage in a sample of random calls.<sup>23</sup> However, it remains unclear whether this association also exists for high-risk calls. Therefore, we aim to investigate the risk of undertriage and overtriage in high-risk telephone calls to OOH-PC in Denmark. In addition, we aim to explore patient characteristics and call characteristics associated with undertriage and overtriage in random calls and in high-risk calls to OOH-PC.

## METHODS

### Design and setting

This paper presents secondary analyses of a natural quasi-experimental study in two regional OOH-PC services in Denmark.<sup>23 27</sup> In one prior paper,<sup>23</sup> we explored safety and efficiency in randomly selected calls, using undertriage and overtriage. In the present paper, we included potentially high-risk calls and studied undertriage and overtriage of these calls. Subsequently, we explored factors associated with undertriage and overtriage in randomly selected calls and in high-risk patient calls.

OOH-PC is provided by the general practitioner cooperative (GPC) in the Central Denmark Region since 1992 and by the medical helpline 1813 (MH-1813) in the Capital Region of Denmark since 2014. Both services are

open outside office hours (ie, on weekdays from 16:00 to 8:00, weekends and national holidays), offering telephone consultations, clinic consultations and home visits. The GPC and MH-1813 use different triage models. The GPC uses GP-led telephone triage without CDSS, whereas the MH-1813 uses nurse-led telephone triage (see [table 1](#)). At the MH-1813, the majority of incoming unselected calls (ie, 74%) are answered by registered nurses.<sup>28</sup> Nurses are obliged to use a locally developed CDSS and they can redirect calls to a physician on call (ie, 11% of all incoming unselected calls answered by a nurse).<sup>28</sup> These physicians, who have various medical specialties, also conduct telephone triage. In the present paper, we excluded these calls from our analyses (ie, triaged by physicians).

### Definition of groups of calls

We defined two groups of calls: (1) a group of randomly selected calls that was representative for all calls to OOH-PC and (2) a group of potentially high-risk patient calls (referred to as 'high risk'). To define the high-risk patient calls, we conducted a systematic literature search in 2016 to identify factors associated with unsafe or inaccurate telephone triage. Seven identified studies described factors associated with a risk of unsafe telephone triage: infants,<sup>22</sup> increasing patient age,<sup>14 17 21</sup> calls during nighttime,<sup>17 22</sup> abdominal pain,<sup>14-18</sup> chest pain<sup>16 17</sup> and breathing difficulties.<sup>16 18</sup> In a consensus meeting, the authors (DSG—medical doctor, AFP—psychologist, MBC—general practitioner, LH—medical doctor) defined criteria for high-risk calls as calls concerning patients above 30 years who suffered from abdominal

**Table 1** Description of the organisation and telephone triage of the two included models for out-of-hours primary care

	GP cooperative (GPC)	Medical helpline 1813 (MH-1813)
Region	Central Denmark Region	Capital Region of Denmark
Population <sup>33</sup>	1.3 m citizens	1.8 m citizens
Telephone calls in 2016 <sup>28</sup>	701 000	806 000
Organiser	GPs in the region	Regional administration
Organisation and services	<ul style="list-style-type: none"> <li>▶ Telephone consultations, home visits, and clinic consultations at the GPC</li> <li>▶ GPs are obliged to take part in the service</li> </ul>	<ul style="list-style-type: none"> <li>▶ Telephone consultations and home visits run by MH-1813</li> <li>▶ Clinic consultations are located in hospital facilities and managed by EDs</li> </ul>
Remuneration of professionals	Fee for service	Payment by the hour
Triage model	<p>GPs or GP trainees in their final year of specialty; no CDSS available</p> <p>GPs typically work 8 hour shifts and have approx. 1–4 shifts per month besides their daytime work</p>	<p>Nurses who are obliged to use a CDSS and have option to redirect calls to a physician</p> <p>Physicians with different medical specialties (a minority being GPs)</p> <p>Nurses work 8 hour shifts and are mostly fully employed at MH-1813.</p> <p>Physicians work 8 hour shifts and are employed besides their daytime job</p>
The triage models and organisations remain largely unchanged since 2014.		
CDSS, computerised decision support system; EDs, Emergency Departments; GP, general practitioner.		

pain. We added the age criterion, as we aimed to include potentially dangerous conditions that present with vague, indistinctly or greatly differing symptomatology (such as dissecting aorta aneurism, myocardial infarction, cholecystitis, pyelonephritis, acute pancreatitis, gastrointestinal ulcer and gynaecological causes). Due to pragmatic considerations, we focused on one reason for encounter; selecting high-risk calls had to be done by manual development of an algorithm (see further).

### Selection of calls

All calls to the GPC and MH-1813 outside office hours during the inclusion period (GPC: 23 November to 7 December 2016; MH-1813: 23 November to 8 December 2016) were potentially eligible. For calls redirected by a nurse to a physician at MH-1813, only the part conducted by the nurse was eligible and later assessed as described below. Randomly selected calls were selected from all eligible calls; these calls were the same as we studied in a prior paper.<sup>23</sup> The aim was to include 435 calls by GPs and 435 by nurses, based on a power calculation to detect a 5% difference in undertriage between nurse-led and GP-led undertriage in line with the aim of the main study.<sup>23</sup> Based on expected exclusion, we randomly selected 525 random calls from GPC and 500 calls from MH-1813.

To identify potential eligible high-risk calls, we received electronic patient journal records from the GPC and MH-1813. We defined a list of wordings and abbreviations associated with abdominal pain, which was used to search the patient records and identify all eligible calls concerning abdominal pain. DSG (medical doctor) assessed these marked patient records to check whether inclusion for abdominal pain was correct. Calls were excluded when the triage professional in the patient record dismissed the presence of abdominal pain (ie, triagist noted *no abdominal pain*) or when the complaint was clearly outside the thoracic, abdominal or pelvic region. Thus, we identified 846 (GPC) and 884 (MH-1813) eligible high-risk calls for potential inclusion. A power calculation was designed to detect a significant difference between nurse-led and GP-led undertriage for high-risk calls, without detecting associations between risk factors and over- and undertriage. As this calculation revealed the need of 206 calls per triage model, we randomly selected 252 calls from the GPC and 240 calls from MH-1813. Selection of both randomly selected and high-risk calls was done, matching the overall distribution on day of week (ie, weekend, not weekend) and time of day (ie, day, evening, night).

Each selected call had a unique identification number that was used to identify the corresponding audio recording. Three master students of medicine masked the audio-recordings using beep tones to cover information revealing triage profession, OOH organisation and patient identification information. In addition, the students screened all calls for fulfilling exclusion criteria as shown in online supplemental appendix 1. DSG reviewed all calls that fulfilled exclusion criteria and those

that were unclear. In case of doubt, DSG and AFP reached consensus regarding exclusion.

### Assessment of accuracy of triage

All included calls have been assessed as described in prior papers,<sup>23 29</sup> using the tool 'Assessment of Quality in Telephone Triage' (AQTT). The AQTT comprises 24 items assessing the health-related quality and the quality of communication.<sup>29</sup> AQTT has a rating manual, describing when to assign the different ratings for each item.<sup>29</sup> In the present study, the outcome accuracy of triage was measured by one item that used a 7-point scale to differentiate between levels of undertriage (ratings 1 to 3), optimal triage (rating 4) and overtriage (ratings 5 to 7). The AQTT showed inter-rater disagreement when using the entire 7-point scale for assessment of the accuracy of triage but revealed satisfactory inter-rater agreement when using dichotomous scales for *clinically relevant* undertriage (ratings '1' and '2' vs ratings '3' to '7') and *clinically relevant* overtriage (ratings '6' and '7' vs ratings '1' to '5').<sup>29</sup>

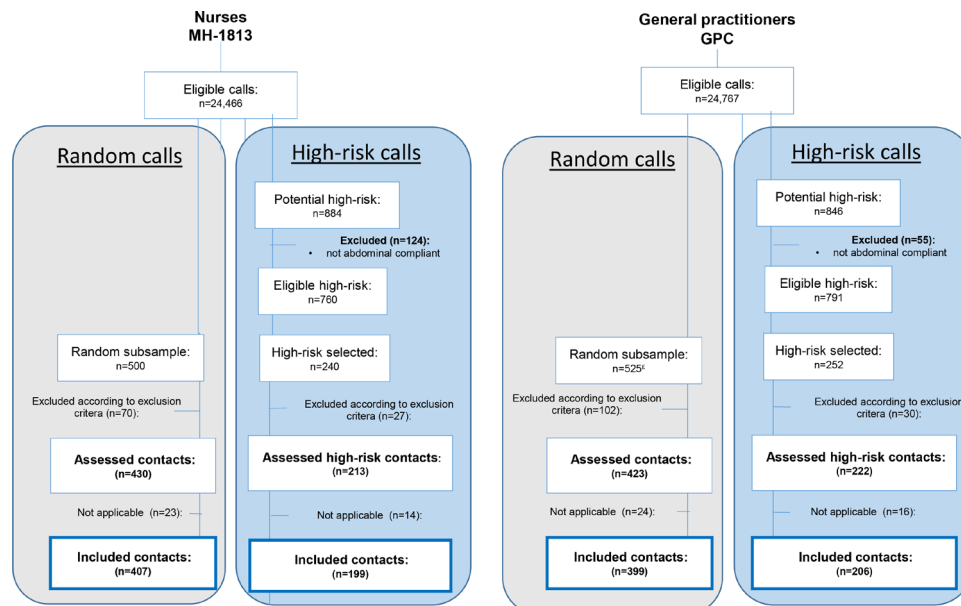
For calls that triage nurses redirected to a physician, only the part conducted by the nurse was available for assessment. These calls could be assessed as optimal, if the decision to redirect the call was what would be expected by a nurse.

For the assessment panel, we recruited 24 physicians among triage professionals from the GPC and MH-1813 using two inclusion criteria: (1) >1 year experience and (2) active in telephone triage in OOH-PC at time of study. We randomly selected 16 GPs from the 56 interested GPs from the GPC. At MH-1813, we included all eight physicians fulfilling our inclusion criteria from the 10 interested physicians. All assessors followed a 2-day training course providing knowledge on telephone triage and communication, introducing the AQTT and rating manual, and assessing triage calls individually and in plenary, focusing on achieving consistency. We randomly distributed calls to all assessors, so each member of the assessment panel assessed random and high-risk calls triaged with both the GP-led model and the nurse-led model. Assessors were blinded for the type of call and triage model. Information on age and sex of the patient, day of week and the time of each call was available for the assessor, extracted from the registration systems from the GPC and MH-1813.

### Statistical analyses

Accuracy of triage decision was categorised into *clinically relevant* undertriage (rated '1' or '2') and *clinically relevant* overtriage (rated '6' or '7'). We conducted separate analyses for randomly selected calls and for high-risk calls. We used descriptive analyses to describe patient characteristics and call characteristics as well as the risk of *clinically relevant* undertriage and *clinically relevant* overtriage for both type of calls. We explored the association of individual patient characteristics (ie, age, patient sex) and call characteristics (ie, weekend, time of day, triage model) with inaccurate telephone triage, by calculating





**Figure 1** Flowchart of inclusion and exclusion of randomly selected and high-risk calls from the GPC and MH-1813. Not applicable: accuracy of triage could not be assessed due to insufficient information in the call or assessment of accuracy of triage was deemed not relevant. GPC, general practitioner cooperative.

the risk ratio (RR) of having *clinically relevant* undertriage (vs no *clinically relevant* undertriage) and *clinically relevant* overtriage (vs no *clinically relevant* overtriage), using binomial regression. 95% CIs were calculated. All analyses were performed in Stata V.14.2 (StataCorp. 2015. *Stata Statistical Software: Release 14.2*. College Station, Texas: StataCorp LP).

### Public and patient involvement statement

We explored patients' perspective in a focus group interview concerning the development of the AQTT and incorporated received input in the rating manual of the AQTT.<sup>29</sup>

## RESULTS

### Description of calls

Figure 1 shows the flowchart of selection and exclusion of calls. We excluded 47 randomly selected calls and 30 high-risk calls assigned 'not applicable', as assessing accuracy of triage was not possible (eg, insufficient information was available) or not relevant. Thus, our final study population included 806 randomly selected calls and 405 high-risk calls (table 2). In high-risk calls, the risk of *clinically relevant* undertriage was 7.9%, whereas the risk of *clinically relevant* overtriage was 5.9%.

### Risk factors for inaccurate triage in randomly selected calls

For randomly selected calls, we found no significant association between patient characteristics (ie, age and sex), call characteristics (ie, weekend and time of call) and the risk of clinically relevant undertriage or overtriage (table 3).

### Risk factors for inaccurate triage in high-risk calls

Nurse-led triage was associated with significantly less *clinically relevant* undertriage than GP-led triage (RR=0.47, 95% CI 0.23 to 0.97) in high-risk calls (table 4). Nurse-led triage had significantly more *clinically relevant* overtriage (RR=3.93, 95% CI 1.50 to 10.53) compared with GP-led triage. High-risk calls conducted during nighttime had a significantly higher risk of being *clinically relevant* undertriaged (13.2%) compared with calls conducted during day or evening (6.4%) (RR=2.1, 95% CI 1.05 to 4.07). We found no significant association for the other patient characteristics and call characteristics. However, a trend was seen for patient age, as the risk of *clinically relevant* undertriage was lower in patients aged 30–59 years (6.3%) in comparison with elderly patients ≥60 years (11.3%) (RR=0.55, 95% CI 0.29 to 1.07).

## DISCUSSION

### Principal findings

In high-risk calls, nurse-led triage was associated with significantly lower risk of *clinically relevant* undertriage and significant higher risk of *clinically relevant* overtriage compared with GP-led triage. For high-risk calls, the risk of *clinically relevant* undertriage was significantly higher if the call was conducted during nighttime compared with day and evening. For randomly selected calls, we found no significant association between defined risk factors and the risk of *clinically relevant* undertriage or overtriage.

### Interpretation and comparison of results

Our study revealed that the triage model and time of call had an effect on accuracy of triage. In a prior study, we found that nurse-led triage had less *clinically relevant*

**Table 2** Baseline description of patient and call characteristics in randomly selected calls and high-risk calls

	Randomly selected calls	High-risk calls*
	N=806†	N=405†
	% (N)	% (N)
Patient characteristics		
Age groups (in years)		
<18	36.5 (294)	N.a.*
18–29	21.0 (169)	N.a.*
30–59	27.5 (222)	67.2 (272)
≥60	15.0 (121)	32.8 (133)
Sex		
Male	40.1 (323)	40.3 (163)
Female	59.9 (483)	59.8 (242)
Call characteristics		
Weekend		
Not weekend	46.4 (374)	51.4 (208)
Weekend	53.6 (432)	48.6 (197)
Time of day		
Day or evening	83.6 (674)	77.5 (314)
Nighttime	16.4 (132)	22.5 (91)
Assessed accuracy of triage‡		
Clinically relevant undertriage		
1: Severe undertriage	1.5 (12)	3.0 (12)
2: Moderate undertriage	4.0 (32)	4.9 (20)
Satisfactory triage		
3: Mild undertriage	8.6 (69)	11.6 (47)
4: Optimal triage	69.0 (556)	66.2 (268)
5: Mild overtriage	10.3 (83)	8.4 (34)
Clinically relevant overtriage		
6: Moderate overtriage	4.6 (37)	4.4 (18)
7: Severe overtriage	2.1 (17)	1.5 (6)
Triage model		
GP-led triage	49.5 (399)	50.9 (206)
Nurse-led triage with CDSS	50.5 (407)	49.1 (199)

Definitions: Weekend=Friday 16:00 to Monday 8:00 and holidays. Day=8:00 to 16:00, Evening=16:00 to midnight, Night=midnight to 8:00.

\*High-risk calls only included patients aged ≥30 years calling OOH-PC with abdominal pain.

†Not applicable: Calls where the assessment of accuracy of triage was not possible or not relevant were excluded (randomly selected: n=47, high-risk; n=30).

‡Accuracy of triage was categorised into three categories: (1) satisfactory triage (including optimal triage ('4'), mild undertriage ('3'), and mild overtriage ('5')), (2) clinically relevant undertriage ('1' or '2') and (3) clinically relevant overtriage ('6' or '7').

CDSS, computerised decision support system; GP, general practitioner; OOH-PC, out-of-hours primary care.

undertriage and more *clinically relevant* overtriage than GP-led triage in randomly selected calls.<sup>23</sup> In this study, we found the same tendencies in a selection of potential high-risk calls. Our study could not elicit which factors of the triage models influence the difference between nurse-led and GP-led triage in high-risk calls, such as the role of CDSS, professional background, working conditions and organisational conditions. Nurses at the MH-1813 were obligated to use CDSS, whereas GPs did not use CDSS. CDSSs aim to ensure consistency<sup>30,31</sup> and to have a high degree of safety (ie, low level of undertriage), which consequently leads to a higher level of overtriage. Furthermore, telephone triage in OOH-PC serves as a form of gatekeeping to acute healthcare in Denmark. A qualitative study described that nurses did not consider themselves as 'gatekeepers', but as 'service providers'.<sup>32</sup> Hence, perceptions of the task at hand may differ between nurses and GPs, thereby affecting overtriage, which could be seen as a service to the callers.

We found that high-risk calls during nighttime were significantly more likely to be undertriaged than calls during day or evening. This corresponds to a study by Hayward *et al*, finding that patients calling during low call volume (eg, nighttime) had a higher risk of requiring secondary care within 3 days after the OOH-PC contact.<sup>17</sup> Our study cannot elicit the mechanisms behind the increased risk of undertriage during nighttime. Fatigue of the triage professional could play a role. Moreover, a stricter triage and gatekeeping function may be conducted during nighttime due to different organisational setup with less capacity of staff and consultations.

Our study may suggest that being elderly could influence the risk of undertriage in high-risk calls. This non-significant trend corresponds to prior studies, which found that increasing age was associated with unsafe triage.<sup>14,17,21</sup> One could hypothesise that elderly may tend to wait longer before contacting OOH-PC, which could result in calls more at risk of being urgent.

### Implications for future research and clinical practice

Although the risk of inaccurate triage differed between nurses and GPs, knowledge of mechanisms behind this difference are lacking and need further exploration. Calling during nighttime was associated with higher risk of being undertriaged for high-risk patient calls, which may be related to a change in available resources and an urge to increase gatekeeping. Future studies should explore the effect of using a CDSS, working in different working conditions and organisational conditions or having different professional background on the level of undertriage and overtriage. Moreover, the influence of other patient characteristics (eg, socioeconomically factors and age) and of health complaints presented in the call are relevant themes to study in relation to the accuracy of triage. From a clinical perspective, this study suggests that triage professionals preferably should pay extra attention when making a triage decision in calls concerning abdominal pain conducted during nighttime

**Table 3** Patient characteristics and call characteristics associated with clinically relevant undertriage and overtriage in randomly selected calls\*

	Undertriage†		Overtriage‡	
	N=44		N=54	
	% (N)	RR† (95% CI)	% (N)	RR† (95% CI)
Patient characteristics				
Age groups (in years)				
<18 (n=294)	3.7 (11)	0.57 (0.23 to 1.37)	7.5 (22)	1.51 (0.63 to 3.63)
18–29 (n=169)	4.1 (7)	0.62 (0.23 to 1.68)	7.7 (13)	1.55 (0.61 to 3.97)
30–59 (n=222)	8.1 (18)	1.23 (0.55 to 2.74)	5.9 (13)	1.18 (0.46 to 3.03)
≥60 (n=121)	6.6 (8)	1	5.0 (6)	1
Sex				
Male (n=323)	5.3 (17)	1	4.6 (15)	1
Female (n=483)	5.6 (27)	1.06 (0.59 to 1.92)	8.1 (39)	1.74 (0.97 to 3.10)
Call characteristics				
Weekend				
Not weekend (n=374)	5.4 (20)	1	6.7 (25)	1
Weekend (n=432)	5.6 (24)	1.04 (0.58 to 1.85)	6.7 (29)	1.00 (0.60 to 1.68)
Time of day				
Day or evening (n=674)	5.3 (36)	1	7.0 (47)	1
Nighttime (n=132)	6.1 (8)	1.13 (0.54 to 2.39)	5.3 (7)	0.76 (0.35 to 1.65)

\*We excluded 47 calls, as assessing accuracy of triage was not possible or not relevant.  
†Undertriage: triage decision assessed as clinically relevant undertriage (rated '1' or '2').  
‡Overtriage: clinically relevant overtriage (rated '6' or '7').  
§Significant difference in RR, using binomial regression analyses; p<0.05: Definitions: Weekend=Friday 16:00 to Monday 8:00 and holidays. Day=8:00 to 16:00, Evening=16:00 to midnight, Night=midnight to 8:00.

and that extra attention may be focused when call is concerning elderly.

### Strengths and limitations of the study

A major strength was the quasi-experimental design using real-life calls as opposed to the constructed setup used in previous studies.<sup>18–20 30 31</sup> An additional strength was the meticulous assessment process using the validated AQTT tool combined with a comprehensive rating manual.<sup>29</sup>

Our study had the following limitations. Due to the thorough assessment process, each call was assessed by one assessor. Consequently, bias due to misclassification cannot be rejected. However, we took several precautions to ensure consistency of assessments. The assessors followed a comprehensive training course and assessments followed the meticulously developed and validated AQTT.<sup>29</sup> Furthermore, we attempted to mask the audio-recordings for information about organisation and triage model. Also, assessors were not aware of the design with both randomly selected and high-risk calls. Moreover, we dichotomised accuracy of triage into *clinically relevant* undertriage and overtriage, which had a satisfactory inter-rater agreement of the AQTT.<sup>29</sup> We carefully decided on our definition of high-risk calls (ie, calls for adults >30 years with abdominal pain). The cut-off point for age was reached through meticulous discussions among

the authors, but could be too low, thereby including a larger group of low-risk calls. Another limitation was our small sample size. In line with the main study, our power calculation was made to identify a significant difference between nurse-led and GP-led undertriage and overtriage. As only a selection of these calls was assessed as *clinically relevant* undertriage or overtriage, the present study lacked power to identify risk factors for undertriage and overtriage. Therefore, we explored clinically relevant associations of patient characteristics and call characteristics with inaccurate triage. Furthermore, we chose physicians as assessors of the accuracy of triage, as they were most frequently used in other studies. The decision to include only physicians in the assessment panel may have induced similar-to-me cognitive bias when assessing nurse-led triage, leading to underassessment of the accuracy of nurses' triage decisions. Additionally, knowledge about the time of the triage call could have resulted in bias, as the assessors may have a different threshold for assessing a decision as accurate during nighttime versus daytime. However, as we used experienced triage physicians as assessors, who assessed a call using their clinical experience, we expect this bias to have minor influence. Also, our study was underpowered to perform multivariate analyses, so we cannot ignore potential confounding

**Table 4** Patient characteristics and call characteristics associated with clinically relevant undertriage and overtriage in high-risk calls

	Undertriage*		Overtriage†	
	N=32		N=24	
	% (N)	RR (95% CI)	% (N)	RR (95% CI)
Patient characteristics				
Age groups (in years)				
30–59 (n=272)	6.3 (17)	0.55 (0.29 to 1.07)	7.4 (20)	2.44 (0.85 to 7.01)
>60 (n=133)	11.3 (15)	1	3.0 (4)	1
Sex				
Male (n=163)	8.6 (14)	1	6.8 (11)	1
Female (n=242)	7.4 (18)	0.87 (0.44 to 1.69)	5.4 (13)	0.80 (0.40 to 1.73)
Call characteristics				
Weekend				
Not weekend (n=208)	6.7 (14)	1	6.3 (13)	1
Weekend (n=197)	9.1 (18)	1.36 (0.69 to 2.65)	5.6 (11)	0.89 (0.41 to 1.95)
Time of day				
Day or evening (n=314)	6.4 (20)	1	6.1 (19)	1
Nighttime (n=91)	13.2 (12)	2.1 (1.05 to 4.07)*	5.5 (5)	0.91 (0.35 to 2.36)
Triage model				
GP-led triage (n=206)	10.7 (22)	1	2.4 (5)	1
Nurse-led triage with CDSS (n=199)	5.0 (10)	0.47 (0.23 to 0.97)‡	10.0 (19)	3.93 (1.50 to 10.33)‡

We excluded 30 calls as assessing accuracy of triage was not possible or not relevant.

\*Undertriage: triage decision assessed as clinically relevant undertriage (rated '1' or '2').

†Overtriage: clinically relevant overtriage (rated '6' or '7').

‡Significant difference in RR, using binomial regression analyses;  $p < 0.05$ : Definitions: Weekend=Friday 16:00 to Monday 8:00 and holidays.

Day=8:00 to 16:00, Evening=16:00 to midnight, Night=midnight to 8:00.

CDSS, computerised decision support system; GP, general practitioner.

concerning the associations found. Finally, we needed to exclude calls for which the level of accuracy was assessed as not applicable, as it could both reflect a correct performance (ie, *this item correctly found not relevant*) or potentially cover a poor performance (ie, *available information is insufficient for assessment*).

## CONCLUSION

This study found that high-risk calls triaged by nurses were less likely to be undertriage and more likely to be overtriage compared with calls triaged by GPs in OOH-PC. High-risk calls conducted during nighttime were significantly more likely to be undertriaged than those during day and evening.

### Author affiliations

<sup>1</sup>Research Unit General Practice, Aarhus University, Aarhus, Denmark

<sup>2</sup>Department of Public Health, Aarhus University, Aarhus, Denmark

<sup>3</sup>Department of Clinical Medicine, Aarhus University, Aarhus, Midtjylland, Denmark

<sup>4</sup>Emergency Medical Services Copenhagen, Copenhagen University Hospital, Ballerup, Denmark

<sup>5</sup>Department of Cardiology—Herlev and Gentofte, Copenhagen University Hospital, Copenhagen, Denmark

<sup>6</sup>Research Unit for General Practice, Aarhus University Research Unit General Practice, Aarhus, Midtjylland, Denmark

**Acknowledgements** The authors would like to thank the 24 assessors who participated in the assessment process and the patients who gave valuable feedback during the development of the AQTT. The authors thank the MH-1813 and the GPC organisations. A special acknowledgement to statistician Claus H Vestergaard for providing valuable feedback in analyses.

**Contributors** All authors contributed to the development of the study protocol and design. DSG (guarantor) collected the calls, conducted the statistical analyses, and produced the first draft of the manuscript. DSG, AFP, LH, MBC, and FF (medical doctor) contributed to the interpretation of data and critically revised the manuscript. All authors contributed with proofreading of the manuscript.

**Funding** This study was supported by the Danish foundation TrygFonden (104893), Primary Health Care Research Foundation of the Central Denmark Region (*Praksisforskningsfonden*—1-15-1-72-13-09), the Committee for Quality Improvement and Continuing Medical Education in general practice in the Central Denmark Region (*Kvalitets- og Efteruddannelsesudvalget* - 1-30-72-227-15) and the Committee of Multipractice Studies in General Practice (*Multipraksisudvalget* - 15/1880).

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants but The National Committee on Health Research Ethics in the Central Denmark Region was consulted and found that no approval was required for this study [personal communication—ref.number:125/2015]. The study was approved by the Danish Patient Safety Authority (reference number: 3-3013-1274/1). Consent was not





obtained as approved by the Danish Patient Safety Authority (reference number: 3-3013-1274/1). The project (ID: 200) has been approved and is registered in the Record of Processing Activities at the Research Unit for General Practice in Aarhus in accordance with the provisions of the General Data Protection Regulation (GDPR).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Dennis Schou Graversen <http://orcid.org/0000-0002-7784-9403>

Anette Fischer Pedersen <http://orcid.org/0000-0003-0253-3671>

#### REFERENCES

- Smits M, Rutten M, Keizer E, *et al*. The development and performance of after-hours primary care in the Netherlands: a narrative review. *Ann Intern Med* 2017;166:737–42.
- Christensen MB, Olesen F. Out of hours service in Denmark: evaluation five years after reform. *BMJ* 1998;316:1502–5.
- Anderson A, Roland M. Potential for advice from doctors to reduce the number of patients referred to emergency departments by NHS 111 call handlers: observational study. *BMJ Open* 2015;5:e009444.
- Bunn F, Byrne G, Kendall S. Telephone consultation and triage: effects on health care use and patient satisfaction. *Cochrane Database Syst Rev* 2004;CD004180.
- Salk ED, Schriger DL, Hubbell KA, *et al*. Effect of visual cues, vital signs, and protocols on triage: a prospective randomized crossover trial. *Ann Emerg Med* 1998;32:655–64.
- Morony S, Weir K, Duncan G, *et al*. Enhancing communication skills for telehealth: development and implementation of a teach-back intervention for a national maternal and child health helpline in Australia. *BMC Health Serv Res* 2018;18:162.
- Agha Z, Roter DL, Schapira RM. An evaluation of patient-physician communication style during telemedicine consultations. *J Med Internet Res* 2009;11:e36.
- Ernesäter A, Winblad U, Engström M, *et al*. Malpractice claims regarding calls to Swedish telephone advice nursing: what went wrong and why? *J Telemed Telecare* 2012;18:379–83.
- Huibers L, Smits M, Renaud V, *et al*. Safety of telephone triage in out-of-hours care: a systematic review. *Scand J Prim Health Care* 2011;29:198–209.
- Morreel S, Colliers A, Remmen R, *et al*. How accurate is telephone triage in out-of-hours care? An observational trial in real patients. *Acta Clin Belg* 2022;77:301–6.
- Hansen EH, 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.
- Smits M, Hanssen S, Huibers L, *et al*. Telephone triage in general practices: a written case scenario study in the Netherlands. *Scand J Prim Health Care* 2016;34:28–36.
- Montalto M, Dunt DR, Day SE, *et al*. Testing the safety of after-hours telephone triage: patient simulations with validated scenarios. *Australasian Emergency Nursing Journal* 2010;13:7–16.
- Rørtveit S, Meland E, Hunskaar S. Changes of triage by GPs during the course of prehospital emergency situations in a Norwegian rural community. *Scand J Trauma Resusc Emerg Med* 2013;21:89.
- Kristoffersen JE. Out-Of-Hours primary care and the patients who die. A survey of deaths after contact with a suburban primary care out-of-hours service. *Scand J Prim Health Care* 2000;18:139–42.
- Philips H, Van Bergen J, Huibers L, *et al*. Agreement on urgency assessment between secretaries and general practitioners: an observational study in out-of-hours general practice service in Belgium. *Acta Clin Belg* 2015;70:309–14.
- Hayward GN, Vincent C, Lasserson DS. Predicting clinical deterioration after initial assessment in out-of-hours primary care: a retrospective service evaluation. *Br J Gen Pract* 2017;67:e78–85.
- Gamst-Jensen H, Lippert FK, Egerod I. Under-triage in telephone consultation is related to non-normative symptom description and interpersonal communication: a mixed methods study. *Scand J Trauma Resusc Emerg Med* 2017;25:52.
- Elliott AM, McAteer A, Heaney D, *et al*. Examining the role of Scotland's telephone advice service (NHS 24) for managing health in the community: analysis of routinely collected NHS 24 data. *BMJ Open* 2015;5:e007293.
- North F, Odunukan O, Varkey P. The value of telephone triage for patients with appendicitis. *J Telemed Telecare* 2011;17:417–20.
- Smits M, Huibers L, Kerssemeijer B, *et al*. Patient safety in out-of-hours primary care: a review of patient records. *BMC Health Serv Res* 2010;10:335.
- Kempe A, Bunik M, Ellis J, *et al*. How safe is triage by an after-hours telephone call center? *Pediatrics* 2006;118:457–63.
- Graversen DS, Christensen MB, Pedersen AF, *et al*. Safety, efficiency and health-related quality of telephone triage conducted by general practitioners, nurses, or physicians in out-of-hours primary care: a quasi-experimental study using the assessment of quality in telephone triage (AQTT) to assess audio-recorded telephone calls. *BMC Fam Pract* 2020;21:84.
- Ernesäter A, Holmström I, Engström M. Telenurses' experiences of working with computerized decision support: supporting, inhibiting and quality improving. *J Adv Nurs* 2009;65:1074–83.
- Holmström IK. Decision aid software programs in telenursing: not used as intended? Experiences of Swedish telenurses. *Nurs Health Sci* 2007;9:23–8.
- Holmström IK, Kaminsky E, Lindberg Y, *et al*. Registered nurses' experiences of using a clinical decision support system for triage of emergency calls: a qualitative interview study. *J Adv Nurs* 2020;76:3104–12.
- Graversen DS, Huibers L, Christensen MB, *et al*. Communication quality in telephone triage conducted by general practitioners, nurses or physicians: a quasi-experimental study using the AQTT to assess audio-recorded telephone calls to out-of-hours primary care in Denmark. *BMJ Open* 2020;10:e033528.
- VIVE. Regional GP cooperatives and medical helpline 1813. A mapping focusing on organisation, activity and economy. 2018. Available: <https://www.vive.dk/da/udgivelser/regionale-laegevagter-og-akuttefon-1813-10796/>
- Graversen DS, Pedersen AF, Carlsen AH, *et al*. Quality of out-of-hours telephone triage by general practitioners and nurses: development and testing of the AQTT—an assessment tool measuring communication, patient safety and efficiency. *Scand J Prim Health Care* 2019;37:18–29.
- Murdoch J, Barnes R, Pooler J, *et al*. The impact of using computer decision-support software in primary care nurse-led telephone triage: interactional dilemmas and conversational consequences. *Soc Sci Med* 2015;126:36–47.
- Wouters LT, Zwart DL, Erkelens DC, *et al*. Tinkering and overruling the computer decision support system: working strategies of telephone triage nurses who assess the urgency of callers suspected of having an acute cardiac event. *J Clin Nurs* 2020;29:1175–86.
- Lindberg BH, Rebnord IK, Høye S. Phone triage nurses' assessment of respiratory tract infections - the tightrope walk between gatekeeping and service providing. A qualitative study. *Scand J Prim Health Care* 2021;39:139–47.
- Statistics Denmark. StatBank Denmark. n.d. Available: <http://www.statbank.dk/statbank5a/default.asp?w=1920>



**Appendix 1.** Exclusion criteria

Type of call	Definition/clarification
Frequent callers	Patients with $\geq 7$ calls during the two-week inclusion period (the patient's medical record from the OOH service that could include important information on these patients was only available to the triage professional and not to the assessor)
Calls by mistake	Calls with no caller answering the triage professional
Daytime calls	Calls performed during daytime (the telephone triage service at MH-1813 was available during daytime)
Calls by other health professionals	The caller was another healthcare professional, e.g. from a nursing home
Administrative calls	The reason for calling was administrative, e.g. calling to get the number for the acute dentist
Calls regarding simple drug prescriptions	The patient called for renewal of a prescription that required little information sharing
Preterm termination	Calls that were ended too early, e.g. calls made by error, no sound on call, or sound interrupted in the middle of call
Calls from other localisation	Calls from a caller who was not in the same location as the patient, e.g. parent on the way to pick up a sick child from day care
Calls with poor sound quality	Calls with poor sound quality making assessment difficult
Language issues	Calls in which language issues challenged the triage, i.e. caller did not speak Danish or English

Not able to identify call | Calls where an exact linkage to the corresponding audio recording could not be established