PREHOSPITAL MEDICINE

NHS Direct: consistency of triage outcomes

A O'Cathain, E Webber, J Nicholl, J Munro, E Knowles

....

Objectives: To examine the consistency of triage outcomes by nurses using four types of computerised decision support software in NHS Direct.

Methods: 119 scenarios were constructed based on calls to ambulance services that had been assigned the lowest priority category by the emergency medical dispatch systems in use. These scenarios were presented to nurses working in four NHS Direct call centres using different computerised decision support software, including the NHS Clinical Assessment System.

Results: The overall level of agreement between the nurses using the four systems was "fair" rather than "moderate" or "good" (κ=0.375, 95% CI: 0.34 to 0.41). For example, the proportion of calls triaged to accident and emergency departments varied from 22% (26 of 119) to 44% (53 of 119). Between 21% (25 of 119) and 31% (37 of 119) of these low priority ambulance calls were triaged back to the 999 ambulance service. No system had both high sensitivity and specificity for referral to accident and emergency services.

Conclusions: There were large differences in outcome between nurses using different software systems to triage the same calls. If the variation is primarily attributable to the software then standardising on a single system will obviously eliminate this. As the calls were originally made to ambulance services and given the lowest priority, this study also suggests that if, in the future, ambulance services pass such calls to NHS Direct then at least a fifth of these may be passed back unless greater sensitivity in the selection of calls can be achieved.

See end of article for authors' affiliations
....

Correspondence to:
Alicia O'Cathain, Medical Care Research Unit, University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK;
a.ocathain@sheffield.ac.uk

Accepted for publication 29 July 2002
....

NHS Direct, the 24 hour telephone advice line staffed by nurses, was introduced in three pilot sites in 1998 and extended to 22 sites, covering the whole of England, during 2001. Nurse advisors use computerised decision support software to triage callers to self care, to contact their general practitioner immediately or later, or to attend accident and emergency (A&E) departments urgently or as an emergency via a 999 ambulance. In addition, they offer self care advice and health information to callers on a wide range of health problems. Until 2001, three software systems were in use: TAS (Plain Software), Personal Health Adviser (McKesson HBOC), and Centramax (McKesson HBOC). During 2001 a fourth—the NHS Clinical Assessment System (AXA Assistance)—was implemented as the national standard system across all sites in England.[1] This system is also being used for triaging calls to out of hours primary care as part of the Exemplar Programme,[2] is being piloted for use in general practice in normal hours,[3] is being used for face to face consultations in walk in centres,[4] and there are plans to pilot a face to face version of the software to triage people attending A&E departments.[5]

There is a growing body of knowledge about this new service. There is evidence that callers find NHS Direct's advice helpful and reassuring,[6] that it has halted the upward trend in demand for out of hours general practice, and that there has been no measurable impact on demand for A&E departments and 999 ambulance services,[7] although it may have reduced the number of telephone calls for advice made to A&E departments.[8] As well as evidence on how users react to the service and its impact on demand for other services, evidence is also emerging around the quality of the triage outcome. Concerns have been raised that NHS Direct triage may be no better than self referral to 999,[9] and that there is considerable variation in the triage outcome and self care advice given to callers.[10] Therefore we have attempted to explore further the consistency of triage outcomes by NHS Direct in relation to the software system used, taking advantage of the temporary coexistence of four different software systems in NHS Direct.

METHODS

We constructed scenarios, or case vignettes, of calls to NHS Direct. We did not base these scenarios on actual calls made to NHS Direct, as the call information available might have been influenced by the software used. Instead, we used 119 calls about minor problems to three ambulance services, collected as part of two reviews of priority dispatch system performance.[11] The 119 calls had all been assigned the lowest priority category by the priority dispatch systems used by the services, and all of the patients had been conveyed, but not admitted, to hospital. For each call, the patient report forms completed by ambulance service crews, the A&E department notes, and anonymised transcripts of the calls that had been made from the recorded call logs, were collated into a single scenario from which all personal and place names, locations, and dates were omitted. For example, one scenario was a young woman, calling from her workplace at midday on a Monday, with a swollen throat caused by a wasp sting; she had had the symptoms for three hours and had already contacted a GP about the incident. Another was for a 4 year old child who had fallen down some concrete steps on a Friday afternoon and had a number of cuts and bruises.

A single researcher presented these 119 scenarios in person to one NHS Direct nurse in each of the four call centres. The nurses had been working in A&E departments before joining NHS Direct and each had worked with their respective triage software for at least three months. The calls were presented in person and not on the telephone. Each of the calls was introduced in the same way by telling the nurse the day and time of day of the call (for example, ... "this call is being made on Monday at 7 30 pm"). Next the researcher introduced the problem about which they were calling using the description used by the original caller to the ambulance service. The researcher then answered factually each question asked by the NHS Direct nurse, if information was available. If a question was asked for which no factual information was available the researcher invented an answer consistent with the overall history, and recorded this question and answer so that the same

Table 1 Characteristics of the 119 scenarios compared with NHS Direct calls

Characteristic	Level	% scenarios (n=119)	% NHS Direct calls* (n=49219)
Time and day of week	In hours (0800–1759 weekdays)	47	29
	Out of hours	53	71
Caller	Patient	9	44
	Other	91	55
Patient age	Child (0–14)	14	38
	Adult	86	62
Patient sex	Male	40	43
	Female	60	57
Necessity of attendance at A&E†	Necessary	77	Not available
	Unnecessary	23	

*Based on calls to three first wave NHS Direct sites.[14 15] †Based on 118 calls for which classification available.

Table 2 NHS Direct triage outcomes for the 119 scenarios by software*

Advice	TAS		Personal Health Adviser		Centramax		NHS Clinical Assessment System	
	n	(%)	n	(%)	n	(%)	n	(%)
Emergency ambulance	32	(26.9)	33	(27.7)	37	(31.1)	25	(21.0)
A&E department	50	(42.0)	26	(21.8)	53	(44.5)	31	(26.1)
GP immediately	6	(5.0)	25	(21.0)	9	(7.6)	18	(15.1)
GP later/other professional	9	(7.6)	16	(13.4)	9	(7.6)	11	(9.2)
Self care	22	(18.5)	19	(16.0)	11	(9.2)	34	(28.6)

*The pattern of outcomes for NHS Direct triaged calls is 2%–3% to 999 ambulance, 15%–27% to A&E, 35%–49% to general practice and 33%–35% to self care.[15]

Table 3 Sensitivity and specificity of systems in predicting necessity of attendance at A&E departments (percentages)

	TAS	Personal Health Adviser	Centramax	NHS Clinical Assessment System
Sensitivity	70	51	78	49
Specificity	33	52	33	59

reply could be given to the same question when presenting the call to other NHS Direct call centres. At the end of the "call" the researcher recorded the outcome given by the nurse using the categories 999 ambulance, A&E department, GP immediately (up to four hours), GP later, other service, and self care. The length of calls was not recorded, but the researcher perception was that they lasted between 2 and 10 minutes and tended to be shorter where the outcome was for a 999 ambulance.

The patients who were the subjects of each of the 119 calls had been coded as a necessary or unnecessary attendance at an emergency department as part of a previous study.[11] The classification used to determine the necessity of attendance at A&E was based on a validated measure of processes of care rather than diagnosis, and has been described elsewhere.[12]

Analysis

Agreement between the triage systems was measured as the proportion of all calls that were managed in the same way, corrected for chance agreement using Cohen's κ statistic. There is a generally accepted view that values of κ between 0.2 and 0.4 represent only "fair" agreement, with values of 0.4 to

0.6 being "moderate", and values above 0.6 being "good".[13] The sensitivity and specificity of each system in predicting necessary attendance at an A&E department was calculated using the classification system referred to above as the "gold standard".

RESULTS

The scenarios were similar to the types of calls received by NHS Direct, but they were not representative of NHS Direct calls. They were less likely to be made out of hours, less likely to be made by the patient, and less likely to be made on behalf of children, than calls to NHS Direct (table 1). About a quarter of the scenarios (27 of 118) were classified as unnecessary attendance at A&E, similar to the 24% unnecessary attendance found at eight A&E departments in another study.[12]

The triage outcomes for each system were very different (table 2). For example, the proportion of scenarios triaged to accident and emergency varied between 22% and 44%, and the proportion disposed to self care between 9% and 29%. The overall level of agreement between the nurses using the four systems was "fair" (κ=0.375, 95% CI: 0.34 to 0.41). Between 21% (25 of 119) and 31% (37 of 119) of these low priority

ambulance calls were triaged back to the 999 ambulance service. The sensitivity of the systems, that is the proportion of necessary attendances triaged to A&E by each system, was high for two of the systems and low for two of the systems (table 3). Systems with higher sensitivity tended to have lower specificity—that is, they sent more unnecessary calls to A&E, than the less sensitive systems. None of the systems had both high sensitivity and specificity.

DISCUSSION

We observed large differences in triage outcomes between different nurses using different software systems to triage the same calls. The level of agreement was higher than that reported in a previous study of inexperienced nurses using software to triage paediatric calls, where the κ statistic was poor at 0.11.[16] The nurses in our study were experienced and dealt with few paediatric calls—which may be more difficult to assess than adult calls—and therefore we might expect a better performance in our study. The sensitivity of NHS Direct to necessary attendance at A&E, of between 49% and 78%, compared well with a sensitivity of 54% found for 50 emergency calls triaged by telephone by nurses.[17] However, our study has illustrated the difficulty in devising a software system that has both high sensitivity and specificity. The differences we have found were not subtle: for example, more than three times as many of our "callers" were advised to self care by the new NHS Clinical Assessment System than by one of the older systems. If this is the case in routine use then the NHS Clinical Assessment system may have more success in reducing demand on immediate care services than previous systems, but may also carry a greater risk of under-triage.

Strengths and limitations

Scenarios or case vignettes are widely used when assessing consistency of diagnosis and treatment in health services,[18 19] and developing clinical guidelines using consensus methods.[20] Cues included in the scenarios need to be selected with care, contextual cues should be made explicit, and it may not be best practice to include all possible scenarios.[20] The cues in our scenarios were taken from real events and were consistent across the four presentations; we included the contextual cues of day and time; and we did not attempt to include all possible scenarios. The scenarios reflected the types of calls undertaken by NHS Direct—indeed none of the four nurses participating in our study expressed any concerns about the nature of the 119 calls. We did not attempt to simulate the process of NHS Direct calls but we feel that this did not adversely affect the study. The "calls" were undertaken face to face rather than by telephone, but the patient was not visible to the nurse during our scenarios in the same way they are not visible to NHS Direct nurses. The calls were shorter than NHS Direct calls because they did not include the checking of personal details, such as address and general practitioner, nor the conversation involved in "wrapping up" a call,[3] issues that were not central to the outcome of the call. In fact, the main concern of the nurses in the study was that they could not probe the "caller" for enough information, a limitation of all scenarios that prevents them fully reflecting real life. This concern was expressed by all four nurses in our study.

The purpose of the study was to compare the four software systems rather than assess the appropriateness of the triage outcome. Thus consistency between all four presentations of the scenarios was the important issue. To this end, one researcher undertook all four presentations and made extra notes for each scenario at the first presentation, which she adhered to for further presentations. The first three systems were tested within two weeks of one another. However, the fourth system was tested several months later, after the same researcher re-familiarised herself with the process and used the same scenarios and notes and did not deviate from those

notes. No external checks were made on consistency but we feel that there was a good level of consistency imposed by the adherence to written notes.

Finally, the variation found in our study may be attributable to variation between nurses as well as between software. TAS is an interpretative software allowing nurses to decide from available options the triage outcome they will recommend to the caller. Both Access and Centramax are more prescriptive and indicate the triage outcome for the nurse. In practice, nurses can over-ride the triage outcome offered by the software and we allowed them to do so in our study to reflect real life. Therefore our study compares nurses using the software rather than the software in isolation. Thus we are unable to disentangle the effect of different nurses and the effect of different software on triage outcomes in NHS Direct.

Implications

The variation we have observed is clearly not attributable to case mix, which was held constant. If the variation is mainly attributable to the nurse, then NHS Direct callers may expect quite different advice depending on who answers their call, raising a question about the experience and training needed by nurses to enable them to answer calls appropriately. If the variation is primarily attributable to the software, then standardising on a single system will obviously eliminate this. However, none of the four NHS Direct software systems tested to date seem to perform well on both sensitivity and specificity for necessity of attendance at A&E services, illustrating the difficulty in balancing the risks of over-triage and under-triage in such a system. It is clearly important that more detailed studies of the appropriateness of NHS Direct triage decisions are undertaken, with the aim of highlighting areas where improvements might be made.

As the calls used in this study were originally made to ambulance services and given the lowest priority by the emergency medical dispatch systems they used, this study also gives an indication of what may happen if, in the future, ambulance services pass such calls to NHS Direct instead of sending an emergency ambulance.[21] At least a fifth of these may be passed back, potentially leading to delay in accessing care. Given that the desired end point is the appropriate use of emergency ambulances, it is important that further refinements are made to the types of low priority calls to be referred to NHS Direct and that operational protocols for such a system consider the possibility of calls being returned to 999 ambulance services so that the risks to the patient are minimised.

ACKNOWLEDGEMENTS

Contributors

Alicia O'Cathain analysed the data, wrote the paper, and acts as guarantor. Elizabeth Webber presented the scenarios to nurses and contributed to interpretation of the findings and writing of the paper. Jon Nicholl designed the study and contributed to interpretation of the findings and writing of the paper. James Munro contributed to the design of the study, interpretation of the findings, and writing of the paper. Emma Knowles contributed to interpretation of the findings and writing of the paper. Brigitte Colwell undertook data administration.

. . . .

Authors' affiliations

A O'Cathain, E Webber, J Nicholl, J Munro, E Knowles, Medical Care Research Unit, University of Sheffield, Sheffield, UK

REFERENCES

1 **Department of Health**. *New national NHS Direct computer system to benefit patients*. Press release 2000/0509, 11 September 2000.
2 **Department of Health**. *Raising standards for patients. New partnerships in out-of-hours care*. Independent Review of General Practice out-of-hours services in England. London: Department of Health, 2000.
3 **National Audit Office**. *NHS Direct in England*. Report by the Comptroller and Auditor General. London: The Stationery Office, 2002.
4 **Salisbury C**, Chalder M, Manku Scott T, *et al*. What is the role of walk-in centres in the NHS? *BMJ* 2002;**324**:399–402.
5 **Department of Health**. *Reforming emergency care: first steps to a new approach*. London: Department of Health, 2001.
6 **O'Cathain A**, Munro JF, Knowles E, *et al*. How helpful is NHS Direct? Postal survey of callers. *BMJ* 2000;**320**:1035.
7 **Munro JF**, Nicholl JP, O'Cathain A, *et al*. Impact of NHS Direct on demand for immediate care: observational study. *BMJ* 2000;**321**:150–3.
8 **Jones J**, Playforth MJ. The effect of the introduction of NHS Direct on requests for telephone advice from an accident and emergency department. *Emerg Med J* 2001;**18**:300–1.
9 **Gaffney P**, Crane S, Johnson G, *et al*. An analysis of calls referred to the emergency 999 service by NHS Direct. *Emerg Med J* 2001;**18**:302–4.
10 **Williams S**. NHS Direct investigated. *Health Which?* August 2000:12–16.
11 **Hartley-Sharpe C**, Snooks H, Dixon S, *et al*. The implications of using a modified priority dispatch system to achieve the new ambulance performance standards in London. London: London Ambulance Service, August 1997.
12 **Lowy A**, Kohler B, Nicholl J. Attendance at accident and emergency departments: unnecessary or inappropriate? *J Public Health Med* 1994;**16**:134–40.
13 **Brennan P**, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1999;**304**:1491–4.
14 **Munro JF**, Nicholl JP, O'Cathain A, *et al*. *Evaluation of NHS Direct first wave sites*. First interim report to the Department of Health. Sheffield: Medical Care Research Unit, University of Sheffield, 1998.
15 **Munro JF**, Nicholl JP, O'Cathain A, *et al*. *Evaluation of NHS Direct first wave sites*. Second interim report to the Department of Health. Sheffield: Medical Care Research Unit, University of Sheffield, 2000.
16 **Wachter DA**, Brillman JC, Lewis J, *et al*. Pediatric telephone triage protocols: standardised decision making or a false sense of security? *Ann Emerg Med* 1999;**33**:388–94.
17 **Leprohon J**, Patel VL. Decision-making strategies for telephone triage in emergency medical services. *Med Decis Making* 1995;**15**:240–53.
18 **Lee R**, MacFarlane T, O'Brien K. Consistency of orthodontic treatment planning decisions. *Clin Orthod Res* 1999;**2**:79–84.
19 **Kalf AJ**, Spruijt-Metz D. Variation in diagnoses: influence of specialists' training on selecting and ranking relevant information in geriatric case vignettes. *Soc Sci Med* 1996;**42**:705–12.
20 **Murphy MK**, Black NA, Lamping DL, *et al*. Consensus development methods, and their use in clinical guideline development. *Health Technol Assessment* 1998;**2**:25–30.
21 **Nicholl J**, Turner J, Martin D. *The future of ambulance services in the United Kingdom*. London: The Ambulance Service Association, 1999.