






OPEN ACCESS

# Is emergency doctors' tolerance of clinical uncertainty on a novel measure associated with doctor well-being, healthcare resource use and patient outcomes?

Luke Budworth <sup>1</sup>, Brad Wilson,<sup>1</sup> Joanna Sutton-Klein,<sup>2</sup> Subhashis Basu,<sup>2</sup> Colin O'Keeffe <sup>3</sup>, Suzanne M Mason <sup>2</sup>, Andrew Ang,<sup>4</sup> Sally Anne-Wilson,<sup>5</sup> Kevin Reynard,<sup>6</sup> Susan Croft,<sup>2</sup> Anoop D Shah,<sup>7</sup> Sakarias Bank,<sup>8</sup> Mark Conner,<sup>9</sup> Rebecca Lawton<sup>1,10</sup>

Handling editor Bhakti Hansoti

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/emered-2023-213256>).

<sup>1</sup>Bradford Institute for Health Research, Bradford, UK

<sup>2</sup>The University of Sheffield School of Health and Related Research, Sheffield, UK

<sup>3</sup>SCHARR, University of Sheffield, Sheffield, UK

<sup>4</sup>Barnsley District General Hospital, Barnsley, UK

<sup>5</sup>Airedale General Hospital, Keighley, UK

<sup>6</sup>Leeds Teaching Hospitals NHS Trust, Leeds, UK

<sup>7</sup>University College London Institute of Health Informatics, London, UK

<sup>8</sup>University of Skövde School of Health Sciences, Skövde, Västra Götaland, Sweden

<sup>9</sup>University of Leeds School of Psychology, Leeds, UK

<sup>10</sup>University of Leeds Institute of Psychological Sciences, Leeds, UK

## Correspondence to

Dr Luke Budworth;  
[l.w.budworth@leeds.ac.uk](mailto:l.w.budworth@leeds.ac.uk)

Received 31 March 2023  
Accepted 22 October 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Budworth L, Wilson B, Sutton-Klein J, et al. *Emerg Med J* Epub ahead of print: [please include Day Month Year]. doi:10.1136/emered-2023-213256

## ABSTRACT

**Introduction** Emergency doctors routinely face uncertainty—they work with limited patient information, under tight time constraints and receive minimal post-discharge feedback. While higher uncertainty tolerance (UT) among staff is linked with reduced resource use and improved well-being in various specialties, its impact in emergency settings is underexplored. We aimed to develop a UT measure and assess associations with doctor-related factors (eg, experience), patient outcomes (eg, reattendance) and resource use (eg, episode costs).

**Methods** From May 2021 to February 2022, emergency doctors (specialty trainee 3 and above) from five Yorkshire (UK) departments completed an online questionnaire. This included a novel UT measure—an adapted Physicians' Reaction to Uncertainty scale collaboratively modified within our team according to Hillen *et al's* (2017) UT model. The questionnaire also included well-being-related measures (eg, Brief Resilience Scale) and assessed factors like doctors' seniority. Patient encounters involving prespecified 'uncertainty-inducing' problems (eg, headache) were analysed. Multilevel regression explored associations between doctor-level factors, resource use and patient outcomes.

**Results** 39 doctors were matched with 384 patients. The UT measure demonstrated high reliability (Cronbach's  $\alpha=0.92$ ) and higher UT was significantly associated with better psychological well-being including greater resilience (Pearson's  $r=0.56$ ; 95% CI=0.30 to 0.74) and lower burnout (eg, Cohen's  $d=-2.98$ ;  $-4.62$  to  $-1.33$ ; mean UT difference for 'no' vs 'moderate/high' burnout). UT was not significantly associated with resource use (eg, episode costs:  $\beta=-0.07$ ;  $-0.32$  to  $0.18$ ) or patient outcomes including 30-day readmission (eg, OR=0.82; 0.28 to 2.35).

**Conclusions** We developed a reliable UT measure for emergency medicine. While higher UT was linked to doctor well-being, its impact on resource use and patient outcomes remains unclear. Further measure validation and additional research including intervention trials are necessary to confirm these findings and explore the implications of UT in emergency practice.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Emergency medicine demands rapid decision-making with limited information where doctors' tolerance for uncertainty may affect resource utilisation and clinical decisions.

## WHAT THIS STUDY ADDS

⇒ We introduce a measure of uncertainty tolerance (UT) showing significant associations with improved doctor well-being but inconclusive results for resource use and patient outcomes.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Enhancing UT could improve doctor well-being suggesting potential benefits from integrating tolerance training in emergency departments. Further research is needed to replicate findings related to resource use and patient outcomes.

## BACKGROUND

Emergency doctors routinely face uncertainty; they work with limited patient information under tight time constraints and receive minimal postdischarge feedback. Indeed, to do their jobs they need to tolerate this uncertainty. Uncertainty tolerance (UT) means adaptively balancing thoughts, feelings and actions in uncertain situations.<sup>1,2</sup> Research shows that people with higher versus lower UT tend to feel less anxious, think more clearly and make decisions less influenced by uncertainty.<sup>1,3-5</sup> While some see UT as a context-dependent 'state', we believe it's likely that while UT fluctuates, an underlying 'trait'-like stability shapes how individuals handle uncertainty.<sup>6</sup>

Across various specialities, higher UT is robustly associated with better psychological and work-life related well-being among clinicians (eg, reduced stress; increased resilience).<sup>2</sup> Supporting doctors to manage uncertainty may thus help reduce endemic burnout among healthcare staff<sup>7,8</sup>—particularly in specialties with 'high rates of undifferentiated illness and uncertainty' as a large recent survey



emphasised.<sup>9</sup> While mixed, a plurality of evidence also shows that clinicians with higher UT make decisions that lead to lower resource use (eg, admit fewer patients; order fewer tests).<sup>2</sup> This, combined with the possibility that higher burnout in low UT doctors could result in poor staff retention and/or more safety incidents, has led some to hypothesise that supporting staff management uncertainty could improve efficiency.<sup>6,7</sup>

Nevertheless, UT remains underexplored in the unique emergency medicine context. Most recent work has explored how emergency doctors navigate uncertainty rather than examining associations between UT and factors like clinical decision-making.<sup>6,10–12</sup> Past studies looking at UT associations<sup>13–17</sup> have had limitations. First, they were USA-based—poorly reflecting emergency department (ED) dynamics of resource-constrained universal healthcare systems.<sup>18</sup> Second, they focused on specific tests/treatments (eg, CT scans) and/or patient groups—limiting generalisability. Third, they used the Physicians' Reaction to Uncertainty Scale (PRUS) to assess UT which is not emergency medicine specific and primarily assesses emotional reactions to uncertainty (eg, anxiety) missing others (eg, behavioural) and undermining validity.<sup>1</sup> Lastly, they did not assess UT associations with indicators of safety (eg, postdischarge mortality) despite the possibility that higher UT may cause patient harm through risky decision-making (eg, inappropriate discharges).<sup>19</sup>

We developed a more comprehensive, modified PRUS for the emergency medicine context and conducted an initial evaluation to assess associations between UT, doctor factors (eg, experience; burnout), resource use (eg, episode costs) and patient outcomes (eg, readmission). Based on literature from other specialties, we predicted that higher UT would be associated with lower resource use and better well-being indicators (eg, lower burnout). The association between UT and patient outcomes was considered either way; higher tolerance could lead to more risky decisions or alternatively, higher tolerance may protect doctors from cognitive 'overload'—providing better judgement of costs/benefits of interventions which could have a beneficial impact on resource use and patient flow.<sup>20</sup>

## METHOD

### Study design

We developed an UT measure and tested it among emergency doctors, assessing its association with their demographics (eg, years of experience, gender), a chart review of their recent patient management decisions and outcomes (eg, episode costs, reattendance rates) and personality/work life factors (eg, self-perceived burnout, confidence, resilience). Online supplemental file 1 provides a study design diagram.

### Measures

#### UT measure

To measure doctors' UT, the PRUS<sup>21,22</sup> was adapted to better suit the emergency medicine context. Hillen *et al*<sup>1</sup> conducted a comprehensive analysis of UT in healthcare identifying 22 existing self-report measures including the most prominent: PRUS. They categorised items across these measures according to a taxonomy of uncertainty sources (eg, complexity, unpredictability, disorder) and three types of responses: cognitive (eg, denial), emotional (eg, stress) and behavioural (eg, avoidance).

The research team held discussions over a series of meetings to create a comprehensive and context-specific measure. Starting with the PRUS, we incorporated relevant items from other validated measures that addressed the sources and responses to uncertainty described in Hillen's model that were not fully

covered by the original PRUS. Items were also modified/excluded if they were not relevant to emergency medicine.

The measure was piloted with seven emergency doctors whose feedback led to further revisions enhancing clarity and ensuring the content was appropriate to the National Health Service. The measure asks respondents to rate agreement from strongly disagree (1) to strongly agree (5) on 34 items, including: 'I usually feel anxious when I am not sure of a diagnosis' and 'The hardest thing to say to patients or their families is: I don't know'. A score per doctor was derived by taking the mean (score range=1–5). It takes 7–9 min to complete. The measure can be seen in online supplemental file 2.

#### Personality and work-life measures

We captured other factors including resilience (6 items), risk aversion (1 item), burnout (1 item), confidence (5 items), work life well-being (10 items) and experience of an adverse event at work (1 item). For further details and specific items see online supplemental file 3.

#### Demographics

Doctors self-reported basic demographics including age, gender, hours worked a week, years of experience and seniority.

#### Patient selection criteria

The research team conducted an internal survey of 10 emergency doctors (from all sites except Leeds)—including members of the author team—ranking patient problems by greatest levels of uncertainty for the physician. These were used to guide the selection of patient charts to be reviewed for the association of a physician's UT and patient management.

#### Doctor questionnaire

The final questionnaire was constructed through discussion and debate among the research team with some additional piloting work conducted with seven emergency doctors. In response to extensive pilot feedback, we kept the battery of measures brief using short and/or novel scales to reduce participant burden (eg, 1-item risk aversion scale). This pragmatic decision prioritised minimising slow recruitment, attrition and missing data over the potential loss of validity or reliability from using such measures.

The resulting questionnaire (online supplemental file 2) consisted of the UT measure, items about basic demographics (eg, age, grade), and the aforementioned personality and work-life measures. In total, it was 65 items and took 10–15 min to complete.

#### Participants

Doctors from five EDs in Yorkshire, UK completed all measures using an online questionnaire. Doctors at specialty trainee (ST) (year) 3 or equivalent and above were recruited only as they operate more independently than more junior doctors and thus make more independent and potentially impactful patient management decisions. Doctors were approached by collaborating colleagues and completed consent online (before the questionnaire). Recruitment commenced 18 May 2021 to 16 February 2022.

#### Patient episode outcomes

To correlate doctor responses on the questionnaire with patient-level variables, we manually extracted data from the 10 most recent patient encounters with each doctor from health records

using a standardised form (online supplemental file 4). Eligible episodes met three criteria:

1. The episode occurred 30 days (minimum) before doctors completed the questionnaire to avoid any influence of the study on their decisions.
2. The patient's primary problem was abdominal pain, back pain, chest pain, headache or vomiting. These were the top five problems ranked as leading to uncertainty by the internal survey.
3. The patient was 18+ years old.

In addition to patient-level demographic and clinical variables, we extracted resource use and patient outcomes to assess their correlation with UT. Online supplemental file 5 provides details on all data items and our cost estimation procedures.

#### Resource use indicators

- ▶ Ordering of any test and/or treatment per patient episode.
- ▶ 24-hour or shorter stay length among admitted patients ('unnecessary admissions').
- ▶ Estimated patient episode costs.

Patient episode costs were estimated using resource use and patient outcome indicators. The aim was not to estimate costs exactly but rather generate measures to assess variance in costs between doctors.

#### Patient outcomes

- ▶ 30-day attendance after initial attendance ('reattendance').
- ▶ 30-day admission to hospital on reattendance.

Though we captured mortality, it was not assessed as an outcome given just six deaths).

#### Participants

##### Sample size

Multilevel regression power primarily depends on the sample size at the higher level (level 2) which in our study refers to the number of doctors versus patients (level 1). This was a challenge because we aimed to assess doctor-level variables as predictors of patient-level outcomes and doctor recruitment was much more resource-demanding than episode data extraction. According to simulation estimates derived by inputting a range of plausible values into a web app,<sup>23</sup> a sample of 500 patients across 50 doctors (achievable within our time constraints) provided sufficient power (70–95%) to detect small to medium effect sizes (Cohen's *d* of 0.30 or 0.50) when comparing median split UT differences on binary outcomes.

#### Analysis

UT score distributions were assessed for variability and skewness. Cronbach's  $\alpha$ /McDonald's  $\omega$  assessed internal consistency and Pearson's correlations, t-tests and analysis of variances (with Bonferroni-adjusted t-tests) were used to examine associations between UT scores and doctor-related factors.

Patient-level outcomes (ie, involving matched patient episode data) were assessed using multilevel logistics (eg, reattendance) or linear (costs) regression (random intercepts for doctors). Outcomes were regressed on doctor-level UT and adjusted for patient/doctor-level variables:

Model 1. Doctor UT + patient site

Model 2. Above + patient American Society of Anaesthesiologists (ASA) score.

Model 3. Above + patient age + patient gender

Model 4. Above + doctor experience (years) + doctor gender

Between-model fits were compared via Bayesian/Akaike Information Criteria. The within-model fit was assessed using  $R^2$  or pseudo- $R^2$  metrics. Intraclass correlation coefficients (ICC) quantified between-doctor outcome variance.

It was hypothesised that higher UT would be associated with lower resource use, thus one-tailed tests were used. Two-tailed tests were planned for readmission, reattendance and mortality as it was hypothesised that UT could be associated in either direction. 95% CIs are provided for all estimates.

Missing data were minimal for most variables. Triage scores (acuity) had high missingness (eg, National Early Warning Scores (NEWS)) so ASA physical status scores were used for covariate adjustment instead. Sensitivity analyses assessed multiply-imputed acuity scores using all variables included in adjusted models as covariates. Other analyses assessed the removal of multivariate outliers and associations with specific versus composite indicators (eg, ECGs only vs all tests).

#### Patient and public involvement

The research team held regular meetings with the Sheffield Emergency Care Forum and Yorkshire & Humber Patient Safety Translational Research Centres (PSTRC) patient panel. These groups were updated on the study progress and were provided feedback on the direction, management and findings throughout.

## RESULTS

### Recruitment and sample characteristics

41 doctors participated and matched patient episodes were obtained for 39 of these doctors (78% of the target 50) (table 1). The mean doctor age was 35.3 and a mean of 8.3 years working in EDs; 63.4% were women. Among all respondents, 12.5% reported having 'no' burnout.

Data for 384 patients were extracted and matched to the doctors (median=10, range=1–22 patients each). 54.7% of matched patients were women and their mean age was 46.6. The most prominent problems were chest (42.2%) and abdominal (25.8%) pain. Across episodes, doctors admitted 26% of patients. Tests ordered most were X-rays (29.4% episodes) and bloods (21.9%), the least MRI (0.3%) scans. 16% of patients reattended within 30 days of initial attendance, 7.6% were readmitted.

### UT and doctor factors

Doctors' mean UT score was 3.6/5 (range=2.6–4.8). The measure showed excellent internal reliability ( $\alpha=0.92/\omega=0.92^{24}$ ) indicating items consistently assessed one construct. Scores had low skewness (–0.01) and kurtosis (1.01) suggesting scores were evenly distributed around the mean with no extreme outliers or a sharp peak.

UT associations with categorical variables (eg, burnout) are shown in figure 1 and with continuous variables (eg, age in figure 2). Individuals with no burnout reported significantly higher tolerance compared with those with some (mean diff.=0.53; Cohen's *d*=1.53; 95% CI=0.23 to 2.83) and moderate/high (collapsed categories) burnout (mean diff.=1.03; Cohen's *d*=2.98; 95% CI=1.33 to 4.62). In turn, those with some reported significantly higher tolerance than those with moderate/high burnout (mean diff.=0.50; Cohen's *d*=1.45; 95% CI=0.39 to 2.51).

Higher resilience was also strongly associated with higher UT (Pearson's  $r=0.56$ ; 95% CI=0.30 to 0.74) as was higher work-life well-being (0.43; 0.13 to 0.65). Though higher confidence correlated with higher UT (0.27; –0.04 to 0.54), this was

**Table 1** Doctor and patient episode characteristics

Full doctor sample (n=41)			
Variable	Level	Value	n (%) miss.
Adverse event experienced? (n, %)	Yes	19 (47.5)	1 (2.4)
	No	21 (52.5)	
Age (M, SD)	–	35.3 (5.7)	0
Burnout (n, %)	None	5 (12.5)	1 (2.4)
	Some	26 (65)	
	Moderate	7 (17.5)	
	High	2 (5)	
Confidence (M, SD)	–	3.3 (0.4)	2 (4.8)
ED experience, years (M, SD)	–	8.3 (6.4)	0
Gender identity (n, %)	Male	15 (36.6)	0
	Female	26 (63.4)	
Grade* (n, %)	Specialty trainee year 3 (ST3)	7 (17.1)	1 (2.4)
	ST4	10 (24.4)	
	ST5	4 (9.8)	
	ST6	7 (17.1)	
	Consultant (highest level)	7 (17.1)	
	Other	6 (14.6)	
Resilience (M, SD)	–	3.3 (0.8)	1 (2.4)
Risk aversion (M, SD)	–	51.5 (21)	1 (2.4)
Site (n, %)	Airedale	6 (14.6)	–
	Barnsley	8 (19.5)	
	Bradford	15 (36.6)	
	Leeds	4 (9.8)	
	Sheffield	8 (19.5)	
Uncertainty tolerance (M, SD)	–	3.6 (0.4)	0
Weekly A&E hours (M, SD)	–	31.8 (10)	0
Work-related well-being (M, SD)	–	3.5 (0.4)	1 (2.4)
Patient episode sample (n=384)			
Variable	Level	Value	n (%) miss.
Admission (initial) (n, %)	Yes	100 (26)	0
	No	284 (74)	
Admission (return, 30 days) (n, %)	All	0	
	Yes	29 (7.6)	
Related to initial attendance?	No	355 (92.4)	
	Yes	26 (89.7)	
Related: number per patient (M, SD)	Yes	3 (10.3)	
	No	2 (6.6)	
Age (M, SD)	–	1.3 (0.6)	
	–	46.6 (18.8)	0
ASA score: continuous (M, SD)	–	1.9 (0.9)	2 (0.5)
	–	2.9 (0.8)	3 (0.8)
Problem type (primary) (n, %)	Ab pain	99 (25.8)	0
	Back pain	33 (11.7)	
	Chest pain	162 (42.2)	
	Headache	36 (9.4)	
	Vomiting	7 (1.8)	
	Combo of >1	47 (12.2)	
Cost index v1: tests/treat (M, SD)	–	£32.3 (48.8)	0
	–	£634.9 (1250.5)	0
Cost index v2: +admissions (M, SD)	–	£854.4 (1583.9)	0
	–	£854.4 (1583.9)	0
Death (30-day) (n, %)	All	0	
	Yes	6 (1.6)	
Related to initial attendance?	No	378 (98.4)	
	Yes	3 (50)	
Related to initial attendance?	No	3 (50)	
	Yes	3 (50)	

Continued

**Table 1** Continued

Patient episode sample (n=384)			
Variable	Level	Value	n (%) miss.
Diagnosis status discharge (n, %)	Confirmed	102 (29.6)	39 (10.2)
	Suspected	243 (70.4)	
Discharge location (n, %)	Home	357 (92.9)	0
	Other	27 (7.1)	
Gender identity (n, %)	Male	174 (45.3)	0
	Female	210 (54.7)	
Hospital stay length in days (M, SD)	–	1.2 (4.3)	0
If admitted (M, SD)	–	4.5 (7.4)	0
Mode of arrival (n, %)	Ambulance	104 (30.4)	42 (11)
	Other	238 (69.6)	
NEWS (M, SD)	–	0.8 (0.9)	265 (69)
NEWS2 (M, SD)	–	1.1 (1.4)	96 (25)
Ordered tests (by doctors) (n, %)	Blood tests	84 (21.9)	0
	CT	29 (7.6)	
	ECG	20 (5.2)	
	MRI	1 (0.3)	
	USS	6 (1.6)	
	X-ray	113 (29.4)	
	Other	36 (9.4)	
Possible decision deferral† (n, %)	Yes	100 (26)	0
	No	284 (74)	
Reattendance (return, 30-day) (n, %)	All	0	
	Yes	61 (15.9)	
Related to initial attendance?	No	323 (84.1)	
	Yes	54 (88.5)	
Related: number per patient (M, SD)	Yes	7 (11.5)	
	No	1.7 (1.9)	
Site (n, %)	Airedale	38 (9.9)	–
	Barnsley	86 (22.4)	
	Bradford	139 (36.2)	
	Leeds	41 (10.7)	
	Sheffield	80 (20.8)	
Waiting time hours (initial) (M, SD)	–	2.6 (2.1)	2 (0.5)

\*Note: In the UK: ST3–ST6: Specialty trainees in years 3–6 after foundational training equivalent to residents or registrars. Consultant: Fully qualified senior specialists similar to attending physicians. Other (eg, specialty doctors, trust grade doctors): For example, doctors not in formal training posts but with comparable experience.  
†Defined as any recorded evidence that the treating (ie, patient-matched) doctor conferred with a colleague about patient management decisions (admission, treatments and/or tests). Note, the full (41) and episode-matched (39) doctor samples showed minimal differences.  
Ab pain, abdomen pain; A&E, accident and emergency; ASA, American Society of Anaesthesiologists; ED, emergency department; % miss, % missing data; NEWS, National Early Warning Scores; Re-att./rtn admit, 30 day reattendance or admission on return; M, mean.

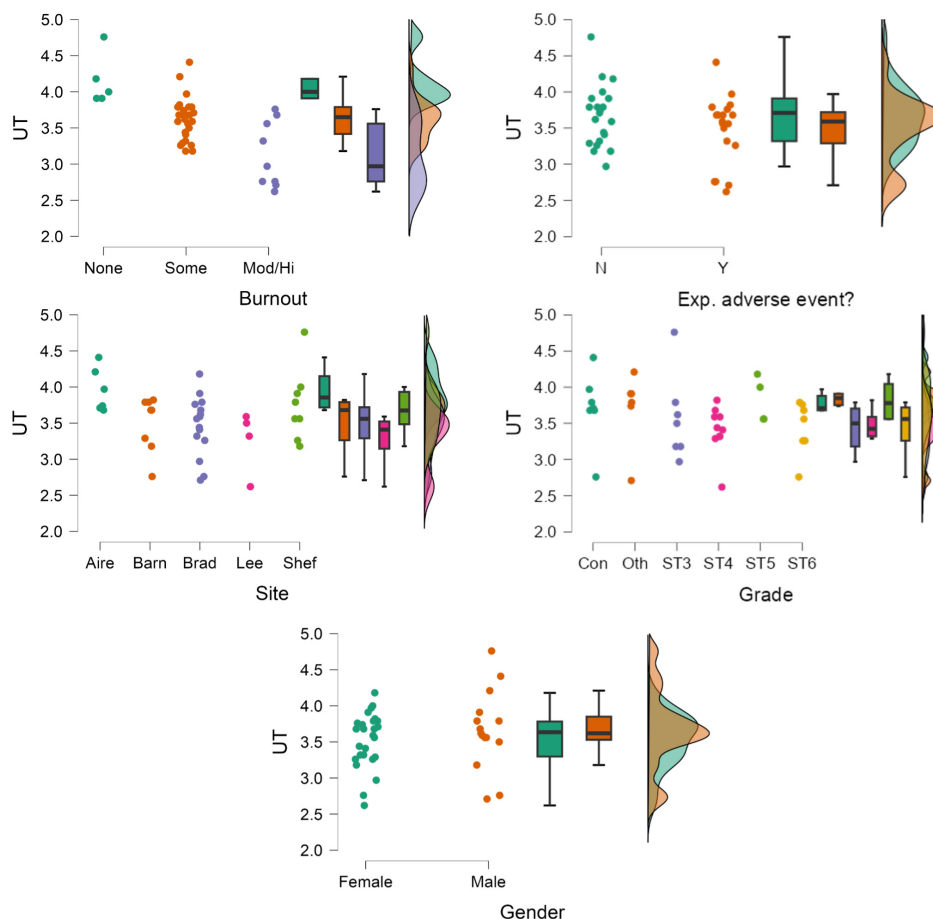
non-significant ( $p=0.09$ ). There were significant UT differences between sites.

Experience of adverse events, hours worked, gender and seniority was not notably associated with UT.

### UT and patient-level outcomes

**Table 2** provides patient-level outcomes results. Point estimates indicated that as UT increased, episode costs decreased (eg, in Model IV looking at total costs shows there was a 7% decrease in costs per one UT scale point increase though we could not statistically rule out estimated increases in costs (0.07, 95% CI: –0.32 to 0.18). ICC was generally low (meaning there was low between-doctor variation in this outcome) and models explained small-to-moderate variance.

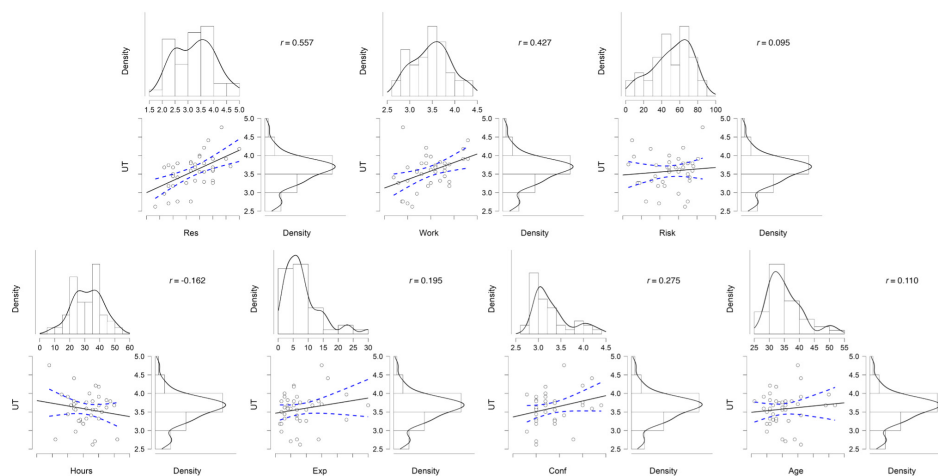
Tolerance was not significantly associated with the composite outcomes of rates of ordering ‘any test’, ‘any treatment’ or ‘any or test or treatment’. Models here explained small amounts of



**Figure 1** Jitter plots, box plots and density plots for associations between uncertainty tolerance (UT) (higher=more tolerant) and all categorical variables: burnout, experience of an adverse event, site, grade and gender.

variance and ICC was low. UT was also not significantly associated with the proportion of patients with  $\leq 24$ -hour hospital stay after admission (extremely high  $R^2$  values indicated overfitting, meaning estimates should be interpreted cautiously). When the composite primary outcome of 'any tests ordered' was broken down into singular test outcomes (eg, CT scan: yes vs no), higher

UT was significantly associated with less ECG ordering—but no other tests (see online supplemental file 6). Analyses showed a small decrease in reattendance and readmission rates with increased UT but estimates were again not significant. Models had low to moderate  $R^2$  values meaning other important variables were unaccounted for.



**Figure 2** Histograms and density plots for all continuous variables plus scatterplots (with Pearson's  $r$  estimates) between uncertainty tolerance (UT) (higher=more tolerant) and continuous variables: res (resilience), work (work life well-being), risk (risk aversion), hours (hours worked weekly), exp (years of experience), conf (confidence) and age. Could this be n (%)—clearer to readers fair point but it cannot be as the column sometimes references means and SD (eg, for age).

**Table 2** Uncertainty tolerance (UT) associations with patient-level outcomes

Model	UT estimate	-95% CI	+95% CI	P value*	R <sup>2</sup> <sub>M</sub> (R <sup>2</sup> <sub>C</sub> )	ICC	AIC   BIC
<b>Logistic models (ORs)</b>							
<b>Tests: any ordered?</b>							
Model I	0.75	0.40	1.41	0.18 <sup>1t</sup>	14% (18%)	5%	498   525
Model II	0.72	0.38	1.36	0.16 <sup>1t</sup>	14% (18%)	5%	496   528
Model III	0.72	0.38	1.35	0.16 <sup>1t</sup>	17% (21%)	5%	<b>491   530</b>
Model IV	0.74	0.40	1.36	0.17 <sup>1t</sup>	18% (21%)	4%	493   541
<b>Treatments: any ordered?</b>							
Model I	0.88	0.52	1.49	0.32 <sup>1t</sup>	6% (7%)	1%	515   543
Model II	0.87	0.52	1.45	0.30 <sup>1t</sup>	9%	<1%	<b>508   539</b>
Model III	0.90	0.54	1.52	0.35 <sup>1t</sup>	11%	<1%	506   545
Model IV	0.93	0.55	1.59	0.40 <sup>1t</sup>	14%	<1%	500   548
<b>Treatments or tests: any ordered?</b>							
Model I	0.65	0.29	1.47	0.15 <sup>1t</sup>	11% (22%)	13%	<b>451   478</b>
Model II	0.64	0.29	1.43	0.14 <sup>1t</sup>	11% (22%)	12%	451   483
Model III	0.66	0.30	1.46	0.15 <sup>1t</sup>	13% (23%)	11%	448   488
Model IV	0.71	0.35	1.45	0.16 <sup>1t</sup>	18% (23%)	7%	446   494
<b>Return attendance 30 days after presentation</b>							
Model I	0.95	0.47	1.92	0.89 <sup>2t</sup>	4%	<1%	320   348
Model II	0.97	0.47	2.00	0.94 <sup>2t</sup>	9%	<1%	<b>312   344</b>
Model III	0.95	0.46	1.95	0.88 <sup>2t</sup>	13%	<1%	311   351
Model IV	0.97	0.46	2.04	0.94 <sup>2t</sup>	13%	<1%	314   362
<b>Admission 30 days postpresentation</b>							
Model I	0.77	0.28	2.11	0.61 <sup>2t</sup>	18%	<1%	191   219
Model II	0.79	0.28	2.26	0.66 <sup>2t</sup>	24%	<1%	<b>185   217</b>
Model III	0.80	0.28	2.30	0.67 <sup>2t</sup>	24%	<1%	189   229
Model IV	0.82	0.28	2.35	0.71 <sup>2t</sup>	24%	<1%	192   240
<b>24-hour or shorter length of stay</b>							
Model I	0.91	0.35	2.34	0.43 <sup>1t</sup>	98%	<1%	<b>196   224</b>
Model II	0.91	0.35	2.34	0.42 <sup>1t</sup>	98%	<1%	198   230
Model III	0.97	0.36	2.59	0.48 <sup>1t</sup>	98%	<1%	198   238
Model IV	0.97	0.36	2.59	0.73 <sup>1t</sup>	98%	<1%	198   245
<b>Linear models (unstandardised β)†</b>							
<b>Test and treatment costs</b>							
Model I	-0.01	-0.06	0.05	0.37 <sup>1t</sup>	8% (12%)	4%	<b>-162   -130</b>
Model II	-0.01	-0.06	0.04	0.36 <sup>1t</sup>	9% (13%)	3%	-156   -120
Model III	-0.01	-0.06	0.04	0.31 <sup>1t</sup>	13% (16%)	3%	-150   -106
Model IV	-0.01	-0.06	0.04	0.32 <sup>1t</sup>	13% (16%)	4%	-130   -79
<b>+ admission costs</b>							
Model I	0.02	-0.24	0.28	0.56 <sup>1t</sup>	16% (17%)	1%	1122   1153
Model II	-0.002	-0.24	0.24	0.49 <sup>1t</sup>	23%	<1%	1088   1123
Model III	-0.005	-0.24	0.23	0.48 <sup>1t</sup>	27%	<1%	<b>1082   1126</b>
Model IV	-0.002	-0.23	0.23	0.48 <sup>1t</sup>	28%	<1%	1094   1145
<b>+ reattendance and return admission costs</b>							
Model I	-0.05	-0.34	0.24	0.37 <sup>1t</sup>	17% (19%)	2%	1182   1213
Model II	-0.07	-0.33	0.19	0.29 <sup>1t</sup>	25% (26%)	<1%	<b>1141   1177</b>
Model III	-0.08	-0.33	0.17	0.28 <sup>1t</sup>	27%	<1%	1146   1190
Model IV	-0.07	-0.32	0.18	0.28 <sup>1t</sup>	28%	<1%	1156   1208

R<sup>2</sup><sub>M</sub>=variance explained by fixed effects.R<sup>2</sup><sub>C</sub>=variance explained by fixed+random effects (where random effects variance minimal R<sup>2</sup><sub>C</sub>=R<sup>2</sup><sub>M</sub>).

Bold AIC/BIC 2=model with the lowest AIC+BIC.

\*1t/2t=one/two-tailed tests.

†Note the natural logarithm of all outcomes was used to mitigate skew after a constant of 182 was added to cost outcomes.

AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; ICC, intraclass correlation coefficients.

Coefficients were practically equivalent when NEWS2 values were imputed and used in place of ASA scores. Removing multivariate outliers did not produce notable changes.

## DISCUSSION

Our UT measure demonstrated a good range and balance of scores, high internal consistency and significant correlations with

self-reported doctor characteristics such as higher psychological resilience and lower burnout. There was no strong evidence of an association between tolerance and resource use; results indicated no significant associations and while coefficients mostly suggested an increase in costs with increased tolerance, CIs showed results consistent with the opposite. This pattern was also seen in patient outcomes.

While our UT measure addresses limitations with the PRUS including making it relevant to the UK emergency medicine context, it remains to be seen how the measure performs in further psychometric testing—for example, test–retesting—in larger samples. The measure appeared to be internally consistent and alpha/omega values were so high that some items may have been redundant and less burdensome measures could be derived. Future work within our team aims to conduct a more thorough validity assessment of our measure. We welcome future studies to adapt and assess the measure for UT assessment in other medical specialties and see whether scale-assessed tolerance is stable, modifiable and predictive of change in future behaviour.

The lack of significant associations between UT and resource use mirror findings from previous emergency medicine research,<sup>13–17</sup> though not all specialties.<sup>3</sup> One reason could be that patient pathways in emergency medicine are often protocolised, making significant practice variance difficult to capture. Homogenising the sample to ST3+ (ie, resident+level) doctors may have further reduced variation. Nonetheless, UT was linked with increased staff well-being suggesting it could be associated with resource use outcomes not captured here such as fewer staff sick days (eg, from burnout) and (eg, stress-induced) clinical errors.

Additional complexities in interpreting ED-based studies of UT arise from overlapping staff dynamics. In secondary analyses, we found that UT is significantly associated with lower ECG ordering rates but such relationships may be challenging to assess where several staff groups such as both nurses and doctors are empowered to order the same tests. Future studies could disentangle these subtleties, perhaps through path models and qualitative work.

For return admission outcomes, though small, regression coefficients indicated lower return rates with higher UT although it with CIs consistent with higher return rates. Thus, we could at least provide some evidence against positive associations of a medium to large size (ie, ORs above the upper estimate of our intervals which rarely exceeded 2). Future research needs to replicate our work with larger sample sizes to assess whether there are other outcomes not captured here that may be more sensitive or further contextual or other doctor factors which may moderate or mediate associations. For example, it could be that because the patients evaluated in this study all had high ‘uncertainty’ problems there was a ceiling effect; with more variation in problems, perhaps there would be more variation in decision-making processes sensitive to UT.

Consistent with previous studies,<sup>1–9</sup> the differences in UT across sites suggest that it may be influenced by contextual factors. Future research could identify and investigate departments where doctors exhibit relatively higher average UT, providing insights into contributing factors that could be applied in other organisations<sup>25</sup> to boost work-related well-being.

While we are unaware of trials of interventions to improve UT for doctors, recent research has identified uncertainty management strategies emergency doctors use already, providing a basis for intervention development.<sup>10–12 26</sup> Patel demonstrated how visual thinking strategies and interactions with horses can prepare medical students to manage uncertainty by exposing

them to a variety of clinical and non-clinical scenarios.<sup>27</sup> Ilgen *et al* also highlighted that UT may manifest not only as a personality trait but also as a situational response that varies with clinical context.<sup>6</sup> These insights suggest a need for interventions that are both adaptive and context-specific using real-world scenarios to effectively cultivate tolerance among healthcare staff.

### Limitations

Records data were taken retrospectively, so some doctors may have deferred decisions to colleagues without noting it, undermining the assumption that costs resulted from their decisions. Similarly, UT scores may not reflect past scores if tolerance-modifying events occurred between episodes and doctor participation.

Because the coefficients in our analysis cannot be interpreted causally, future studies should consider modifying UT through randomised trials to assess causal effects. We did not meet our sample size targets due to overestimating achievable recruitment rates; as a result, our CIs were wider and less precise, making it difficult to detect smaller effects suggested by our point estimates. Convenience sampling may have biased our doctor sample towards those who are more confident and relying on self-reported data could have introduced social desirability bias.<sup>28</sup> Additionally, the self-selection of sites may have led to the inclusion of locations with specific characteristics such as particular organisational cultures.

While no associations between UT and negative patient outcomes were found, the proxy measures chosen may have been too insensitive. Future research should capture broader outcomes such as 30-day patient general practitioner visits and provide more granular analyses of how UT relates to outcomes in specific contexts. For example, research could identify patient groups that are more often inappropriately dismissed or overlooked by doctors with higher UT and assess how this more targeted issue could be mitigated.

Studies have suggested UT can only be investigated as ‘state’<sup>11 12</sup> but this is disputed<sup>1 10</sup> not least because UT may be stable over time.<sup>29</sup> Further, some argue neither extreme tolerance nor intolerance may be inherently beneficial or detrimental as each could provide advantages and expose clinicians to risks in practice, undermining our assumptions that higher scores were ‘better’.<sup>30</sup>

While costs were taken from tariffs, some potentially unrealistic assumptions were made, for example, assuming a medication packet was opened for each given dose and some costs were missed such as tests on reattendance.

Some measures used only one question for example, risk aversion and burnout thus not very sensitive to the complexity of these concepts.

### CONCLUSION

Higher doctor UT on a novel measure was associated with higher self-reported psychological and work-related well-being but was not significantly associated with patient outcomes or most measures of resource use. Interventions to support emergency doctors to better tolerate uncertainty might have positive implications for them but our study is unable to answer the question as to whether there are positive or negative side effects on patient outcomes. It is necessary for future work to replicate and build on our study to resolve such ambiguities.

✗ Luke Budworth @lukewbudworth and Suzanne M Mason @ProfSueMason

**Acknowledgements** Thank you, Professor Tracy Young, for your advice and our data extractors, Amy Greaves, Kiran Bhanu, Shreya Agrawal, Alisa Morris, Hannah Wakefield and Jonathon Heron.

**Contributors** LB, BW, SBAs, CO'K, SMM, SA-W, SC, MC, RL designed the study, including selecting questionnaire measures, developing our uncertainty tolerance measure and deciding on outcome measures. LB managed the data sets, designed the statistical analysis plan and analysed the study data with support and supervision from ADS, BW, JS-K, SBAs, SMM, AA, SA-W, KR, SC all extracted data from patients medical records and contributed to the development of the patient data extraction tool in a pilot. BW, JS-K, SBAs, SMM, AA, SA-W, KR, SC approached and recruited colleagues into the study to complete our doctor questionnaire. SBan extracted cost values from online resources. LB, BW, JS-K, SBAs, CO'K, SMM, AA, SA-W, KR, SC, ADS, SBan, MC, RL all contributed to both the writing and editing of the manuscript. LB is the guarantor of the study.

**Funding** This report is independent research funded by the National Institute for Health and Care Research Yorkshire and Humber Applied Research Collaboration (ARC) (NIHR200166) and the National Institute for Health and Care Research Yorkshire and Humber Patient Safety Research Collaboration (PSRC). The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care. ADS is funded by a postdoctoral fellowship from THIS Institute, NIHR (AI\_AWARD01864 and COV-LT-0009), UKRI (Horizon Europe Guarantee for DataTools4Heart) and British Heart Foundation Accelerator Award (AA/18/6/24223).

**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** Our protocol and materials (accessible here: [github.com/Budlytics/uncertainty\\_tolerance](https://github.com/Budlytics/uncertainty_tolerance)) were approved by a Health Research Authority ethics committee (IRAS ID: 202830; East of Scotland REC).

**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 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Luke Budworth <http://orcid.org/0000-0002-2905-6011>

Colin O'Keeffe <http://orcid.org/0000-0002-4484-7401>

Suzanne M Mason <http://orcid.org/0000-0002-1701-0577>

#### REFERENCES

- Hillen MA, Gutheil CM, Strout TD, *et al*. Tolerance of uncertainty: Conceptual analysis, integrative model, and implications for healthcare. *Soc Sci Med* 2017;180:62–75.
- Alam R, Cheraghi-Sohi S, Panagioti M, *et al*. Managing diagnostic uncertainty in primary care: a systematic critical review. *BMC Fam Pract* 2017;18:79.
- Strout TD, Hillen M, Gutheil C, *et al*. Tolerance of uncertainty: A systematic review of health and healthcare-related outcomes. *Pat Educ Couns* 2018;101:1518–37.
- Rosen NO, Knäuper B, Sammut J. Do individual differences in intolerance of uncertainty affect health monitoring? *Psychol Health* 2007;22:413–30.
- Anderson EC, Carleton RN, Diefenbach M, *et al*. The Relationship Between Uncertainty and Affect. *Front Psychol* 2019;10:2504.
- Ilgén JS, Watsjold BK, Regehr G. Is uncertainty tolerance an epiphenomenon? *Med Educ* 2022;56:1150–2.
- Hall LH, Johnson J, Watt I, *et al*. Healthcare Staff Wellbeing, Burnout, and Patient Safety: A Systematic Review. *PLoS One* 2016;11:e0159015.
- West MC. The Effect of Medical Education on Clinical Decision-Making. *Med Educ* 2021.
- Begin AS, Hidrue M, Lehrhoff S, *et al*. Factors Associated with Physician Tolerance of Uncertainty: An Observational Study. *J Gen Intern Med* 2022;37:1415–21.
- Han PKJ, Strout TD, Gutheil C, *et al*. How Physicians Manage Medical Uncertainty: A Qualitative Study and Conceptual Taxonomy. *Med Decis Making* 2021;41:275–91.
- Ilgén JS, Regehr G, Teunissen PW, *et al*. Skeptical self-regulation: Resident experiences of uncertainty about uncertainty. *Med Educ* 2021;55:749–57.
- Ilgén JS, Teunissen PW, de Bruin ABH, *et al*. Warning bells: How clinicians leverage their discomfort to manage moments of uncertainty. *Med Educ* 2021;55:233–41.
- Pines JM, Hollander JE, Isserman JA, *et al*. The association between physician risk tolerance and imaging use in abdominal pain. *Am J Emerg Med* 2009;27:552–7.
- Pines JM, Isserman JA, Szyld D, *et al*. The effect of physician risk tolerance and the presence of an observation unit on decision making for ED patients with chest pain. *Am J Emerg Med* 2010;28:771–9.
- Baldwin RL, Green JW, Shaw JL, *et al*. Physician risk attitudes and hospitalization of infants with bronchiolitis. *Acad Emerg Med* 2005;12:142–6.
- Andruchow JE, Raja AS, Prevedello LM, *et al*. Variation in head computed tomography use for emergency department trauma patients and physician risk tolerance. *Arch Intern Med* 2012;172:660–1.
- Melnick ER, O'Brien EGJ, Kovalerchik O, *et al*. The Association Between Physician Empathy and Variation in Imaging Use. *Acad Emerg Med* 2016;23:895–904.
- Klein E. *In the UK's health system, rationing isn't a dirty word*. Vox 2020. Available: <https://www.vox.com/2020/11/28/21074386/health-care-rationing-britain-nhs-nice-medicare-for-all>
- Lawton R, Robinson O, Harrison R, *et al*. Are more experienced clinicians better able to tolerate uncertainty and manage risks? A vignette study of doctors in three NHS emergency departments in England. *BMJ Qual Saf* 2019;28:382–8.
- Burgess DJ. Are providers more likely to contribute to healthcare disparities under high levels of cognitive load? How features of the healthcare setting may lead to biases in medical decision making. *Med Decis Making* 2010;30:246–57.
- Gerrity MS, White KP, DeVellis RF, *et al*. Physicians' Reactions to Uncertainty: Refining the constructs and scales. *Motiv Emot* 1995;19:175–91.
- Gerrity MS, DeVellis RF, Earp JA. Physicians' reactions to uncertainty in patient care. A new measure and new insights. *Med Care* 1990;28:724–36.
- Olvera Astivia OL, Gadermann A, Guhn M. The relationship between statistical power and predictor distribution in multilevel logistic regression: a simulation-based approach. *BMC Med Res Methodol* 2019;19:1–20.
- Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011;2:53–5.
- Bradley EH, Curry LA, Ramanadhan S, *et al*. Research in action: using positive deviance to improve quality of health care. *Implement Sci* 2009;4:25.
- Parker E. Understanding and enhancing tolerance of uncertainty in emergency departments [PhD thesis]. Leeds: University of Leeds; 2024. <https://etheses.whiterose.ac.uk/35419/>.
- Patel P. Physician Uncertainty in Medical Settings: A Review. *Med Educ* 2021;5:234–56.
- Tyrer S, Heyman B. Sampling in epidemiological research: issues, hazards and pitfalls. *BJPsych Bull* 2016;40:57–60.
- Buhr K, Dugas MJ. The Intolerance of Uncertainty Scale: psychometric properties of the English version. *Behav Res Ther* 2002;40:931–45.
- Reis-Dennis S. Uncertainty and Decision-Making in Healthcare. *J Gen Intern Med* 2021;3:456–78.