

# Variations in GPs' decisions to investigate suspected lung cancer: a factorial experiment using multimedia vignettes

Jessica Sheringham,<sup>1</sup> Rachel Sequeira,<sup>1</sup> Jonathan Myles,<sup>2</sup> William Hamilton,<sup>3</sup> Joe McDonnell,<sup>4</sup> Judith Offman,<sup>2</sup> Stephen Duffy,<sup>2</sup> Rosalind Raine<sup>1</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2016-005679>).

<sup>1</sup>Department of Applied Health Research, UCL, London, UK

<sup>2</sup>Queen Mary University of London, Centre for Cancer Prevention, London, UK

<sup>3</sup>University of Exeter, Peninsula College of Medicine and Dentistry, Exeter, UK

<sup>4</sup>Department of Public Health, London Borough of Waltham Forest, London, UK

## Correspondence to

Dr Jessica, Sheringham, Department of Applied Health Research, UCL, 1-19 Torrington Place, London WC1E 7HB, UK; [j.sheringham@ucl.ac.uk](mailto:j.sheringham@ucl.ac.uk)

Received 12 May 2016

Revised 24 July 2016

Accepted 18 August 2016

Published Online First

20 September 2016



► <http://dx.doi.org/10.1136/bmjqs-2016-006071>



CrossMark

**To cite:** Sheringham J, Sequeira R, Myles J, et al. *BMJ Qual Saf* 2017;**26**:449–459.

## ABSTRACT

**Introduction** Lung cancer survival is low and comparatively poor in the UK. Patients with symptoms suggestive of lung cancer commonly consult primary care, but it is unclear how general practitioners (GPs) distinguish which patients require further investigation. This study examined how patients' clinical and sociodemographic characteristics influence GPs' decisions to initiate lung cancer investigations.

**Methods** A factorial experiment was conducted among a national sample of 227 English GPs using vignettes presented as simulated consultations. A multimedia-interactive website simulated key features of consultations using actors ('patients'). GP participants made management decisions online for six 'patients', whose sociodemographic characteristics systematically varied across three levels of cancer risk. In low-risk vignettes, investigation (ie, chest X-ray ordered, computerised tomography scan or respiratory consultant referral) was not indicated; in medium-risk vignettes, investigation could be appropriate; in high-risk vignettes, investigation was definitely indicated. Each 'patient' had two lung cancer-related symptoms: one volunteered and another elicited if GPs asked. Variations in investigation likelihood were examined using multilevel logistic regression.

**Results** GPs decided to investigate lung cancer in 74% (1000/1348) of vignettes. Investigation likelihood did not increase with cancer risk. Investigations were more likely when GPs requested information on symptoms that 'patients' had but did not volunteer (adjusted OR (AOR)=3.18; 95% CI 2.27 to 4.70). However, GPs omitted to seek this information in 42% (570/1348) of cases. GPs were less likely to investigate older than younger 'patients' (AOR=0.52; 95% CI 0.39 to 0.7) and black

'patients' than white (AOR=0.68; 95% CI 0.48 to 0.95).

**Conclusions** GPs were not more likely to investigate 'patients' with high-risk than low-risk cancer symptoms. Furthermore, they did not investigate everyone with the same symptoms equally. Insufficient data gathering could be responsible for missed opportunities in diagnosis.

## INTRODUCTION

Lung cancer, the most common cancer worldwide, has comparatively poor survival in the UK.<sup>1</sup> Most patients with lung cancer first present to primary care, but diagnostic delays are well documented: patients with lung cancer have more consultations in primary care before investigation than those with many other cancers.<sup>2</sup> In addition, while intervals from presentation to diagnosis have reduced for other common cancers over time, they remain unchanged for lung cancer.<sup>3</sup> It has been suggested that missed opportunities for lung cancer diagnosis in primary care may contribute to poor lung cancer survival.<sup>4</sup>

Primary care physicians, referred to throughout this paper as general practitioners (GPs), have direct access to lung cancer diagnostic tools, including chest X-ray. GPs may not consider lung cancer as a differential diagnosis because patients with lung cancer commonly present to primary care with non-specific symptoms that are more often due to benign causes.<sup>5</sup> Non-specific symptoms and rare disease occurrence therefore present diagnostic difficulty for GPs.<sup>6</sup> Reducing diagnostic delays requires an understanding of how GPs decide which patients with

common, non-specific symptoms to investigate for lung cancer. It is unclear how GPs decide who requires further investigation by chest X-ray or by specialist referral, and inequalities by patient age, gender and socioeconomic circumstances have been identified in retrospective analyses of routine data.<sup>1 2 7 8</sup> Most previous research has examined the diagnostic process using retrospective data in patients with cancer only,<sup>5</sup> thus missing a key dimension, that is, how GPs decide which patients with symptoms do not require investigation.

Examining decision-making in a standardised way in clinical practice presents substantial methodological challenges.<sup>9 10</sup> Direct observation of real physician–patient encounters offers no opportunity to control patients' clinical and sociodemographic characteristics, and so requires observation of very large numbers of consultations to obtain the necessary numbers in specific risk or demographic categories. The use of fictional patient profiles (vignettes) can provide a valid and efficient approach to examining clinician behaviour,<sup>11</sup> and studies have already produced useful insights into sources of error in clinicians' decision-making processes, due to both patient factors (eg, symptom characteristics)<sup>12</sup> and physician factors (eg, cognitive biases).<sup>12 13</sup> As Blumenthal-Barby and others recognise, however, there are limits to the applicability of written vignettes and other vignette designs that do not simulate key features of real consultations.<sup>14</sup> In particular, when vignettes offer little or no opportunity for physicians to seek information from or about the vignette patient, they can inappropriately frame the decision for the physician by cueing what they should notice about the patient or by offering participants only a limited selection of response options. This risks priming participating physicians to consider certain actions and biasing their responses.

In this vignette study, we therefore sought to simulate key features of consultations. We designed a website using interactive multimedia vignettes with videos of actor 'patients', which enabled participating GPs to ask questions in their own words and receive real-time responses. We used this intervention in a factorial randomised experimental study to examine GPs' decisions to initiate lung cancer investigation across different combinations of patient clinical and sociodemographic characteristics.

## METHODS

### Design

We constructed 36 simulated consultations comprising video vignettes of actor 'patients' and comprehensive clinical information, including previous medical history, comorbidities and examination findings, and sociodemographic characteristics. The symptomatic information provided adhered to material in the latest

available National Institute for Health and Care Excellence (NICE) referral guidelines for suspected cancer (published in 2005),<sup>15</sup> with cancer risk based on data from the CAPER case–control study.<sup>16</sup> Each consultation was designed to take participating GPs approximately 10 min to complete, so that it mirrored the length of a 'real' clinical encounter in primary care in the UK National Health Service.

At the start of each 'consultation', a video was shown where the actor 'patient' volunteered a description of their presenting symptom. Participants could then elicit further information in real time on the presenting symptom, other symptoms and risk factors by typing in questions to which they received the 'patient's' video response. They could also, if they wished, click on a drop-down menu to obtain information on behavioural and familial risk factors, previous medical history, family history, sociodemographic information and examination findings (figure 1). A demonstration is available at: <http://www.ucl.ac.uk/stream/media/swatch?v=c22f1a2b58b8>.

We applied a factorial experimental design, where GPs undertook one consultation from each of six clinical profiles across three lung cancer risk levels (table 1); no GP saw the same actor twice. Within these constraints, allocation of GPs to vignettes was random. This achieved approximate balance of patient characteristics by clinical profile, gender, ethnicity and socioeconomic circumstances. The study protocol is available at: [http://www.ucl.ac.uk/dahr/research-pages/gp\\_study](http://www.ucl.ac.uk/dahr/research-pages/gp_study)

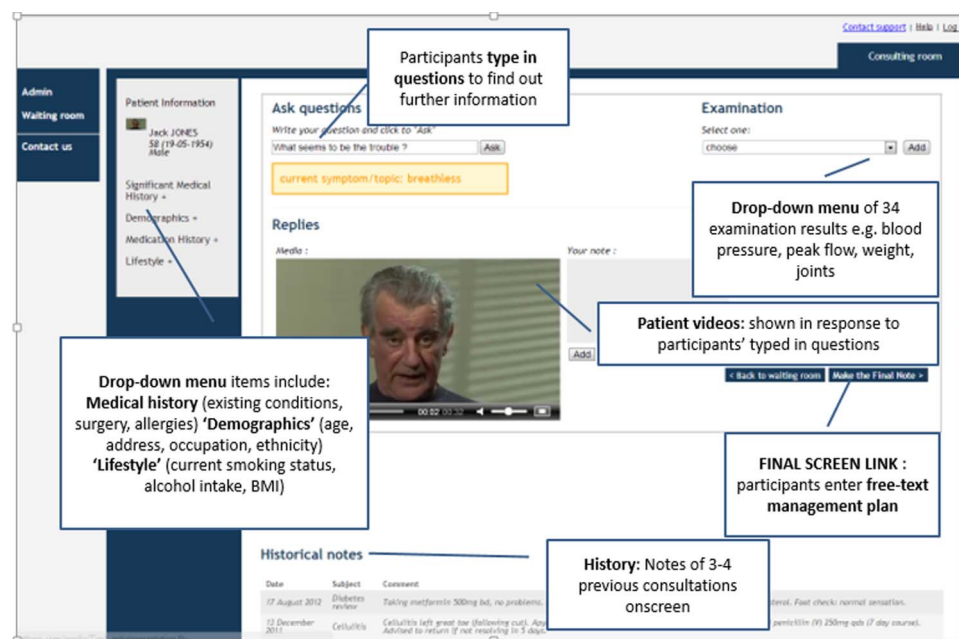
### Recruitment and participation

Qualified GPs and registrars nearing the end of their specialist GP training were invited through nine primary care research networks across England in 2012 and 2013 to participate in a study of decision-making (without explicit reference to lung cancer). Those that returned an expression of interest were sent further information. For GPs that wished to take part, their internet browsers were checked for compatibility with the study software.

GP participants were first trained to use the online simulated consultations. This was done using a web-based video in advance of the study, with access to support from the research team during or between study consultations. Each participating GP used the study website to 'consult' with six 'patients', and at the end of the 'consultation', entered their management plan. GPs also completed a brief questionnaire about their practice characteristics and years since qualifying.

### Application development

The application's development followed the steps recommended by Adler *et al*<sup>17</sup> for developing simulations: 1. Case concept: developing the vignette design and content



**Figure 1** Annotated screen grab from the study website showing an example 'patient', Jack Jones' (profile 1) thumbnail sketch: White British man, aged ~60 years, works on security in a block of offices. Non-smoker, has a history of diabetes (*available through drop-down menu and shown in video if asked*).

Social characteristics: Socioeconomic circumstances—not stated directly, but indicated in the accent and dress of 'patients' and occupation provided on a drop-down menu. Ethnicity—listed in a drop-down menu and reflected in vignette patients' accent and dress, with racial characteristics apparent in the facial features of the 'patient'. All actors spoke fluent English to avoid linguistic barriers.

Presenting symptom (*shown to all*): Breathless. Never felt like this before, and he is not sure what is going on. It is interfering with his life (eg, had to get the bus into work rather than walking), and so wife suggested he come and check it out.

Second symptom (*available if asked*): Fatigue. Presumed this is because of breathlessness, but it is more severe than normal. Not sure why: work is the same as normal; things are no different at home, and he does not feel stressed.

Further information on core profile characteristics (*available if asked*): Notice it particularly when active (eg, struggle playing with the grandchildren). Also, notice it when lying down in bed, and has had to start using one of his wife's pillows. It has been happening for 1–2 weeks (eg, trains young boxers at the local gym, but has not been able to make boxing training for the last week and a half because of it).

Examples of additional profile information (*available through drop-down menus*): medications, heart rate, blood pressure, weight, height, joints (all OK).

**Table 1** Components of the six different clinical profiles by risk level

Clinical profile	Information volunteered by 'patient' or available onscreen			Information only available if participant GPs asked		Positive predictive value (PPV) of lung cancer (%)	Other relevant information	
	Age range	Smoking status	Symptom 1	Symptom 2	Duration			
Low risk: Expected action=no active investigation (safety netting appropriate)								
1	Younger (late 50s)	Non-smoker	Breathlessness	Fatigue	1–2 weeks	0.40	Patient has swollen ankles, possibly due to heart failure	
2	Younger (late 50s)	Smoker	Chest pain	Cough	1–2 weeks	1.10		
Medium risk: Expected action=either investigation (eg, order chest X-ray) or safety netting								
3	Older (late 70s)	Smoker	Chest pain	Cough	Uncertain (~3 weeks)	1.70	Chronic obstructive pulmonary disease present	
4	Older (late 70s)	Non-smoker	Cough	Appetite loss	Uncertain (~3 weeks)	2.50		
High risk: Expected action=lung cancer investigation								
5	Younger (late 50s)	Smoker	Breathlessness	Fatigue	>5 weeks	3–4		
6	Older (late 70s)	Smoker	Chest pain	Weight loss	>5 weeks	14		

GPs, general practitioners.

2. Review and revision by content experts
3. Outline and flow development: a typical online consultation in the study
4. Translation of content into simulation platform: vignette interactive website
5. Pilot testing and revisions

A detailed description of each step is given in online supplementary file S1. In brief, the structure of the factorial experiment required 36 unique vignette combinations to cover the four experimental factors: known to be associated with variations in lung cancer survival, but whose effect on inequalities in GPs' rates of referral for investigation or to secondary care is uncertain.<sup>8</sup>

- *Ethnicity*: three variations (white, black Caribbean, South Asian)
- *Gender*: two variations (male, female)
- *Socioeconomic circumstances*: two variations (advantaged or disadvantaged)
- *Clinical risk of lung cancer*: three variations (low-risk, medium-risk and high-risk), with two profiles for each level of risk. Age was not included as a separate experimental factor, but was instead incorporated into profiles because older age increases the risk of cancer associated with most symptom combinations.<sup>16</sup> We constructed six clinical profiles, two for each risk level, using different combinations of symptoms, age and smoking status (table 1). The positive predictive values (PPV) of lung cancer were drawn from PPVs generated by analysis of symptom combinations in the CAPER case-control dataset and interpretation of these symptoms and their characteristics informed by the latest available NICE guidance on investigation of suspected cancer<sup>15 16</sup> (described further in online supplementary data).

To maximise the clinical authenticity of the cases, GPs specialising in cancer diagnosis and non-academic GPs reviewed the proposed vignettes. The website content and functionality were also informed by patient representatives' comments. For example, these influenced the types of responses 'patients' provided, because patient representatives corroborated previous research that patients may well not disclose certain symptoms with their doctors without being directly asked about them.<sup>18</sup>

The translation of content into the online study application website (virtual patient application) required filming actors portraying patients, creating and populating the website with that content. The website architecture and application software was produced by Athenaeum Educational Technologies. It involved the development of a bespoke system using natural language-processing principles to recognise GPs' free text questions and play a video clip in response (see ref. 19 for an explanation of the principles). This system was underpinned by databases on symptoms or risk factors and the features of those symptoms (eg, what exacerbates or relieves the symptom or how long it has been present).

## Analysis

Every action performed by GPs on the website (ie, all the questions asked of 'patients', drop-down menus accessed, free text entered in management plans) was captured by the study website. This information was used to measure the duration of each consultation and to generate three indicators about GPs' information requests in each consultation and the capacity of the research application to respond to these requests:

- *data sought*: average number of data items sought (questions asked or drop-down menu items accessed) by GP and by individual vignette
- *errors*: error messages displayed as a proportion of all data items sought, calculated for all consultations, consultation 1 and consultations 2–6 only, assuming that in the first consultation GPs were familiarising themselves with the application
- *key information elicited*: proportion of GPs that elicited information on the vignette's second, but unvolunteered, lung cancer symptom.

GPs also had the opportunity to provide free text comments on any aspect of the application in an online survey after all the consultations were completed. These comments were not treated as a representative survey of all participants' experiences, but were examined to provide insights into GPs' experiences of the application and their perceptions of its utility as a research tool for eliciting the decision-making process.

The primary outcome was the proportion of 'patients' for whom lung cancer investigation was included in the management plan. This included ordering appropriate imaging or referral for a specialist opinion, for example, from a respiratory consultant, whether participants' management plan stated this investigation was for lung cancer or not. This outcome variable was constructed from free text responses entered by participants in their management plan, according to predefined criteria. A clinician confirmed the validity of every constructed primary outcome.

Data were analysed by fitting multilevel logistic regression models using Markov chain Monte Carlo for estimation,<sup>20</sup> allowing variation between participants and between vignettes within participants. This allowed for a correlation between outcomes within a given GP but independent outcomes for two vignettes viewed by different GPs. Estimation of ORs and 95% credible intervals was carried out using the RStan library in R V.3.0.2 (Stan Development Team. RStan: the R interface to Stan. V2.5; 2014. <http://mc-stan.org/rstan.html>). Significance testing was carried out using Wald tests, based on the means and posterior variances of the estimates.

Variations in outcome were examined by 'patient' gender, ethnicity, socioeconomic circumstances and risk profile, an indicator variable for whether participants sought the second symptom and GP characteristics (demographics, experience and region). Two



models were built in order to examine differences by (a) clinical profile and (b) by age. A supplementary analysis was conducted to examine whether findings were difficulties in obtaining information sought from the application, by including the indicator on errors as another covariate in each model. To examine selection bias, the gender and age of participating GPs and their practices' cancer referral characteristics were compared with national data.<sup>21 22</sup>

The required sample size was calculated on the basis that a minimum difference in investigations of 10% was considered of clinical importance and realistic, given variations in cancer investigations in other studies.<sup>23</sup> A response from 216 participants was sought to give 1296 vignettes (ie, each of the 36 vignettes viewed 36 times). Each risk and ethnic group would therefore be viewed 432 times, and each gender and socioeconomic group 648 times. Assuming a 20% variance inflation factor for clustering of GPs/patients', 432 in each risk and ethnic group would give 95% power to detect a difference of 10%. For differences between gender and socioeconomic groups, 648 in each group would give 85% power for a difference of 5%.

## RESULTS

### Sample characteristics

Two hundred and twenty-seven GPs completed the study. This was 76% of the 300 GPs who registered for the study and 41% of the 556 GPs in total that initially expressed an interest in taking part (see online supplementary file S2A). There were no demographic differences between registered GPs who did and did not complete the study, but GP participants were younger than the national GP population, and practices had higher cancer referrals than non-participating practices (see online supplementary file S2B).

Out of 1362 vignettes, 14 (1%) were excluded due to missing participant demographic data in one GP ( $n=6$ , 0.4%), when participants asked about second symptoms but did not receive a response ( $n=4$ , 0.3%) or when participants did not enter a management plan ( $n=4$ , 0.3%).

### Consultation process

GPs spent on average 13 min on the first consultation and 11 min on consultations 2–6, and sought 47 items of information per consultation (by asking text questions of the patient, looking up patient history or personal information, conducting 'examinations' or 'bedside tests'). GPs received error messages in response to an average of 4.6% of data sought for consultations 2–6 (range 4%–22%) (see online supplementary file S2C).

### Lung cancer investigations

Participants initiated investigations in 1000 (74%) vignettes. There was little difference in investigation

between low-risk, medium-risk and high-risk levels (72%–75%) but large variation between clinical profiles (59%–86%). There were no variations by 'patient' gender or socioeconomic circumstances, but there was a gradient in investigation by ethnicity, with 'patients' of black ethnicities least likely and white ethnicities most likely to be investigated (71% vs 77%) (table 2).

GPs asked for additional, relevant information about second symptoms in 778 (58%) of cases overall, with marked variation by clinical profile, ranging from 48 (21%) in profile 1 to 214 and 216 (95%) in profiles 2 and 3, respectively. There was a significant interaction between seeking a relevant second symptom and clinical profile ( $p<0.001$ ). Ninety-one per cent of GPs who discovered the presence of weight loss initiating investigation were compared with just 46% who did not seek this information. In contrast, knowing 'patients' experienced fatigue did not significantly change the likelihood of investigation (table 3).

While obtaining second symptom information was associated with more investigation (adjusted OR (AOR): 3.18 (2.27 to 4.70),  $p<0.001$ ), there was still underinvestigation in 'patients' with appetite or weight loss (profiles 4 and 6) compared with 'patients' with chest pain and cough (profile 3) (AORs: 0.25 (0.14 to 0.42),  $p<0.001$  and 0.5 (0.29 to 0.91),  $p=0.02$ , respectively) (table 4a). GPs were less likely to investigate older than younger 'patients' (AOR: 0.52 (0.39 to 0.70),  $p<0.001$ ) and less likely to investigate 'patients' of black compared with white ethnicities (AOR: 0.68 (0.48 to 0.95),  $p=0.03$ ) (table 4b).

Associations were similar when the variable for errors received was included (see online supplementary file S2D).

Comments volunteered by GP participants on their experiences of the application and their perceptions of its utility as a research tool for eliciting the decision-making process are summarised in online supplementary file S3.

## DISCUSSION

### Summary

In this factorial experiment using vignettes in simulated consultations, GPs' decisions to investigate lung cancer were influenced by whether they sought out additional, relevant clinical information about the presence of common symptoms. Even when participating GPs elicited sufficient information about symptoms, inequalities by age and ethnicity in investigation decisions remained.

### Comparisons with existing literature

Our data were collected during 2012–2013, and our finding that GPs investigated a high proportion (72%–75%) of cases is in line with literature from 2013.<sup>23</sup> However, it is higher than that might have been

**Table 2** Frequency of lung cancer investigation

	Investigation		N (vignettes)
	n	Per cent	
Total	1000	74.18	1348
<i>(a) By 'patient' characteristic</i>			
Risk level			
Low	339	75.00	452
Medium	327	72.35	452
High	334	75.23	444
Clinical profile*			
1: PPV=0.4% (younger; ns; 1–2 weeks breathless (and fatigue))	152	66.96	227
2: PPV=1.1% (younger; s; 1–2 weeks chest pain (and cough))	187	83.11	225
3: PPV=1.7% (older; s; ~3 weeks chest pain (and cough))	195	85.90	227
4: PPV=2.5% (older; ns; ~3 weeks cough (and appetite loss))	132	58.67	225
5: PPV=3–4% (younger; s; >5 weeks breathless (and fatigue))	185	82.59	224
6: PPV~14% (older; s; >5 weeks chest pain (and weight loss))	149	67.73	220
Gender			
Female	489	74.09	660
Male	511	74.27	688
Socioeconomic circumstances			
Disadvantaged	508	74.49	682
Advantaged	492	73.87	666
Ethnicity			
White	369	76.56	482
Black	306	71.50	428
South Asian	325	74.20	438
Second symptom elicited			
No	361	63.33	570
Yes	639	82.13	778
<i>(b) By GP participant characteristic</i>			
GP gender			
Female	425	70.48	603
Male	573	77.12	743
GP age range			
25–34 years	227	70.06	324
35–44 years	336	72.89	461
45–54 years	325	78.69	413
55–64 years	102	75.00	136
65 years or over/missing	8	66.67	12
Years since qualifying			
0–2 years ago	120	71.43	168
2–5 years ago	186	69.14	269
5–10 years ago	177	73.75	240
10–20 years ago	256	77.58	330
20+ years ago	259	76.40	339
Ethnicity			
White	583	73.89	789
Black	34	80.95	42
South Asian	296	73.63	402
Other/missing	90	75.63	119
Region			
London	365	73.44	497
East of England	341	74.95	455

Continued

Table 2 Continued

	Investigation		N (vignettes)
	n	Per cent	
North-West	131	76.16	172
West Midlands	96	72.73	132
Surrey and Sussex	41	75.93	54
Locum GP	24	66.67	36

\*Younger=late 50s; older=late seventies; s=smoker; ns=non-smoker; (symptom)=not volunteered by patient.  
GP, general practitioner; PPV, positive predictive values.

expected if GPs were following the latest national guidance for suspected cancer investigation available during the study period.<sup>15</sup> Participants may have proposed more tests for vignette 'patients' than they would in reality, because they were not subject to the resource constraints of clinical practice or may have ordered X-rays primarily to investigate diagnoses other than cancer. Alternatively, they may have been aware of, and, responding to epidemiological evidence, presumed patient preferences and policy published since the 2005 NICE guidance, all of which support a lower threshold for cancer investigation.<sup>24–27</sup> Indeed, updated NICE guidance on referral of suspected cancer, published in 2015 (after our data were collected), includes a substantially lower investigation threshold than that recommended in their earlier guideline,<sup>28</sup> such that all our vignettes would now suggest investigation.

We found that in 42% of cases, GPs did not seek additional information that would help to make an informed decision regarding referral and that was available on request. This accords to some extent with international studies of missed opportunities in cancer diagnosis.<sup>29–30</sup> In the UK, the updated NICE guidance explicitly recognises that patients with combinations of common symptoms may be more likely to have lung cancer than patients with any one of these symptoms alone,<sup>28–31</sup> but patients may not volunteer all the symptoms they experience in consultations, perhaps due to real or perceived time constraints in the

consultation.<sup>31</sup> The importance of data gathering for reaching a timely diagnosis was highlighted in the recent Institute of Medicine report into improving diagnosis in healthcare.<sup>32</sup> The study by Zwaan *et al*<sup>33</sup> of breathlessness using expert review of medical records found evidence of inappropriately selective information gathering in a third of cases, with some evidence that diagnostic error and patient harm occurred in a proportion of these cases. Our study extends the field by providing objective evidence of non-clinical variations in data gathering by physicians in a large vignette study and demonstrates associations between gathering sufficient data and appropriate decision-making.

We also found that the effect of eliciting this second symptom on decision-making varied by symptom. It made little difference whether participants knew that patients had a cough or fatigue, but made significant difference to decision-making if participants knew of appetite and weight loss. For weight loss in particular (a key question when clinicians are considering whether cancer is a possible diagnosis), in 91% of cases where GP participants had elicited information about weight loss, they initiated investigation, compared with just 46% where GPs were unaware the patient had lost weight. It is important to acknowledge that neither in real life nor in the vignettes are the factors (symptom, age and smoking) that constituted each profile independent of one another. Therefore, while we contend the results are

Table 3 Lung cancer investigation by profile according to whether GPs did or did not elicit symptom information

Clinical profile (second symptom)*	Second symptom		Lung cancer investigation				
	Not elicited n (%)	Elicited n (%)	Symptom not elicited		Symptom elicited		Total n (%)
			n (%)	OR (95% CI)	n (%)	OR (95% CI)	
1 (fatigue)	179 (78.85)	48 (21.15)	120 (66.67)	1.00 (–)	31 (65.96)	0.94 (0.43 to 2.09)	152 (66.96)
2 (cough)	11 (4.89)	214 (95.11)	7 (63.64)	0.73 (0.16 to 3.18)	181 (84.19)	2.83 (1.82 to 4.40)	187 (83.11)
3 (cough)	11 (4.85)	216 (95.15)	7 (63.64)	0.93 (0.19 to 4.39)	189 (87.1)	3.67 (2.13 to 6.30)	195 (85.90)
4 (appetite loss)	89 (39.56)	136 (60.44)	42 (46.67)	0.38 (0.21 to 0.69)	91 (66.91)	0.98 (0.59 to 1.62)	132 (58.67)
5 (fatigue)	168 (75.00)	56 (25.00)	136 (80.47)	2.21 (1.31 to 3.72)	50 (89.29)	4.59 (2.86 to 7.37)	185 (82.59)
6 (weight loss)	112 (50.91)	108 (49.09)	52 (46.02)	0.36 (0.20 to 0.62)	99 (90.83)	5.69 (2.07 to 15.63)	149 (67.73)
Total	570 (42.28)	778 (57.72)	364 (63.41)		641 (82.18)		1000 (74.18)

\*Clinical profile is formed from symptoms, smoking status and patient age.  
GPs, general practitioners.

**Table 4** Multilevel logistic regression of cancer investigation by 'patient' characteristic

	Adjusted† OR (95% CI)
<i>(a) By clinical profile</i>	
Clinical profile (second symptom)	
1 (fatigue)	0.62 (0.35 to 1.10)
2 (cough)	0.65 (0.38 to 1.15)
3 (cough)	1
4 (weight loss)	0.25 (0.14 to 0.42)*
5 (fatigue)	1.64 (0.90 to 3.11)
6 (appetite loss)	0.50 (0.29 to 0.91)*
Ethnicity	
White	1
South Asian	0.86 (0.62 to 1.20)
Black	0.67 (0.47 to 0.96)*
Second symptom elicited	
No	1
Yes	3.18 (2.27 to 4.70)*
<i>(b) By age</i>	
Age	
Younger (late 50s)	1
Older (late 70s)	0.52 (0.39 to 0.70)*
Ethnicity	
White	1
South Asian	0.88 (0.63 to 1.27)
Black	0.68 (0.48 to 0.95)*
Smoking status	
Non-smoker	1
Smoker	2.24 (1.64 to 3.02)*
Second symptom elicited	
No	1
Yes	2.83 (2.09 to 3.83)*

\*Significant at  $p \leq 0.05$ .  
†(a) Adjusted for all other factors associated ( $p < 0.1$ ) with investigation in univariate analysis (ie, 'patient' profile and ethnicity, GP gender and age), and whether second symptom was elicited. (b) Adjusted for 'patient' profile, ethnicity, GP gender and age and whether second symptom was elicited.

interpretable and reliable, they are not as definitive as a randomised controlled trial result; so, this finding has to be treated with some caution. However, the finding accords with the recent 'think aloud' study by Kostopoulou *et al*,<sup>34</sup> which suggests that when physicians have an idea of cancer early in the consultation, they ask pertinent questions and initiate appropriate investigations to ensure a cancer diagnosis is reached. Therefore, it still seems likely that routinely questioning patients with ongoing respiratory symptoms about weight loss would expedite the diagnosis of some lung cancers.

Our finding that GPs were less likely to investigate older 'patients' is consistent with several observational studies of primary care cancer referral and investigation.<sup>35 36</sup> Scott *et al*<sup>37</sup> propose that as patients grow older, they are increasingly likely to attribute bodily changes to normal ageing processes than to disease. If

clinicians also apply this 'normal ageing' heuristic, it may explain why GPs in this study were less likely to investigate older patients, despite knowing their symptoms. In contrast, patient experience survey data indicate more referral delays in younger (aged 55–64 years) than older patients (over 75 years). However, survey data may be biased if older patients (with lower overall survival) were underrepresented because they had died or were too ill to participate in the survey (which was undertaken 6–12 months after diagnosis).<sup>2</sup>

We also found smaller ethnic variations in GPs' investigation behaviour, with fewer investigations initiated in black (and to some extent) South Asian 'patients' than white patients. This is consistent with survey data where non-white patients with cancer report more referral delays than white patients.<sup>2</sup> One possible explanation is that GPs were less ready to consider a lung cancer diagnosis in individual non-white 'patients' who presented with high-risk clinical profiles, because they placed weight on knowledge that lung cancer risk factors and prevalence are lower in black and South Asian than white populations.<sup>38</sup> However, there is no evidence that patients of different ethnicities exposed to the same risk factors with similar symptoms are at different risk of lung cancer; so, differential investigation by ethnicity is not clinically warranted. Another possible explanation is that investigation likelihood is influenced by GPs' ethnicity. In this study, there were only seven GPs identified as black; so, it was not possible to examine this, but the mechanism by which observed ethnic variations in decision-making occur remains an important question to address.

**Strengths and limitations**

Our novel approach, using vignettes in an interactive website that delivered real-time responses, obtained comprehensive information on decision-making in over 99% of consultations and in a timeframe comparable with a typical consultation. The method simulated more components of the decision-making process in real time than that has been achieved in previous studies.<sup>39–41</sup>

Of equal importance is the fact that we applied a randomised, factorial, experimental design, with exact balance on profile and risk, and approximate balance, with random allocation, to GPs, on sociodemographic factors. This allowed us to examine the effects of patients' sociodemographic and clinical characteristics on GPs' decision-making. We were not able to achieve total orthogonality in design of every patient characteristic, but the randomisation and approximate balance give some confidence in the general applicability of our results.

Despite the advances we achieved in simulating real consultations, the online vignettes were limited mainly due to the constraints of the natural language system. These constraints meant the website was



unable to provide responses to all GPs' information requests. In the postconsultation survey, 12 GP participants (5%) reported difficulty in obtaining information, which caused some of them frustration, and a small number ( $n=4$ , 1.8%) observed may have altered their decision-making behaviour. The process itself of typing in questions may also have prompted GP participants to consider their clinical reasoning more than they would in their routine clinical practice. Conversely, the opportunity to select from the extensive drop-down selections of examinations without facing any of the logistical constraints faced in a real consultation (eg, time required to measure weight) may have led them to seek more information with less consideration than they would do in routine clinical practice. However, it is important to note that all approaches to simulating consultations have some drawbacks. For example, while other vignette studies have enabled physicians to 'ask' questions of the patient, this has required a researcher to type responses online as 'the patient', sometimes resulting in longer 'consultations' than real consultations.<sup>39–41</sup> Moreover, there are several reasons why these simulations still provide valuable insights into GPs' decision-making. First, our sensitivity analysis indicates that results were very close to the main analysis even after taking into account GPs' difficulties in obtaining responses from the application. Second, shortcomings in doctor–patient communication during the clinical encounter are well recognised, such that patients in real consultations do not volunteer all the information clinicians would need to make informed decisions.<sup>18</sup> Third, it is the divergence from reality that makes simulated consultations useful for studying phenomena or circumstances not possible to observe or investigate in real life.<sup>42</sup> In this study, this divergence enabled the systematic manipulation of patient characteristics to examine their effects on GPs' decisions in isolation of the complex range of patient expectations and comorbidities that might explain variations in decision-making in real life. The divergence also meant GPs were not faced with the logistical and system/organisational constraints that affect referral decisions in practice. As a result, the findings provide insight into the cognitive processes underlying GPs' decision-making when the variation in system and patient factors present in real life are removed.

There was some bias in the GP sample registering for the study in that GP participants' practices had higher cancer referrals than non-participating practices; so, they may be more ready than GPs nationally to investigate symptoms suggestive of cancer. However, there was no evidence to suggest participating GPs would have greater or smaller variation in decision-making than non-participants.

Another possible limitation is that the risk levels were based on positive predictive values from the CAPER symptom case–control dataset, which had

wide and overlapping CIs (as shown in online supplementary data S1). Therefore, the PPVs alone are not sufficient to conclude that clinical risk and therefore decision-making should have varied by profile. However, even where the PPV point estimates are most disparate and CIs overlap minimally, GPs investigated similar proportions of patients. In addition, the risk profiles had additional information other than PPV which should have guided decision-making if GPs were acting in line with the latest available clinical guidance (eg, symptom duration). Furthermore, our three broad categories align well with the 2015 NICE guidance. These equate to: risk below 1%, safety netting; 1%–3%, test in primary care if possible; over 3%, refer for specialist testing.<sup>28</sup>

### Conclusions and implications for research and practice

This study demonstrates that GPs were not more likely to initiate cancer investigations for 'patients' with higher risk symptoms. Furthermore, they do not investigate everyone with the same symptoms equally. It also indicates that insufficient data gathering could be responsible for diagnostic errors. It is not that GPs are doing a bad job: the average GP sees one patient with new lung cancer a year.<sup>16</sup> Distinguishing symptoms indicating possible cancer from self-limiting illness that GPs see daily, therefore, is challenging. However, non-clinical variations in investigation could contribute to the sociodemographic inequalities in the timeliness of diagnosis and survival of lung cancer seen in the UK. It also marks a departure from the National Health Service commitment to promote equality through its services.<sup>43</sup> The findings also have wider implications for quality and safety in healthcare internationally. According to the Institute of Medicine, diagnostic errors contribute to approximately 10% of patient deaths, and sufficient data gathering is an essential part of reaching a timely diagnosis.<sup>32</sup>

It is therefore incumbent on health systems to consider strategies that can be implemented in practice, such as clinician education,<sup>32 44</sup> decision-support tools<sup>24</sup> and the assessment of equity in clinical practice.

**Twitter** Follow Jessica Sheringham at @CLAHRC\_N\_Thames

**Acknowledgements** We are grateful to the 227 GPs that took part in this study and the GP and medical colleagues that piloted the study website in various stages of development. We also acknowledge the following for their essential input in the study:

- Dr Anjali Bajekal provided clinical advice throughout the study and advised on successive rounds of piloting.
- Dr Stephanie Meats recruited, followed up and supported GPs to participate in the study during a placement as a GP registrar.
- Ms Lucy McCann, Mr Barnaby Raine, Ms Anastasia Tillman, Mr Eddy Wax recruited, followed up and supported GPs to participate in the study, working as administrators at UCL.
- Ms Rachael Dodd, Mr Steven Marcos and Ms Jessie Porter coded participant data to support the analysis.
- Mr Dave Ardrone and Mr Tom Haswell advised on patients' experience and styles of reporting lung cancer symptoms to the GP.

- Athenaeum Educational Technologies programmed the tool.
- UCL Media filmed the patient actors.
- Colleagues in the Policy Research Unit and the Department of Applied Health Research provided valuable advice in the development and interpretation of the vignette results.

We acknowledge the support of the National Institute for Health Research, through the Primary Care Research Network for its effective recruitment of study participants.

**Contributors** RR had the initial idea of the study; SD and RR determined the factorial design; JS and JO drafted the study protocol; RS, JO, JMcD and JS developed the data-collection tools; WH advised on the creation of clinical profiles, and analysis of the study with respect to its clinical context; RS, JMcD and JS recruited participants in collaboration with the PCRN and with the support of temporary staff at UCL; SD and JM conducted the analysis; JS managed the study and drafted the initial version of the manuscript; All authors commented on the drafts of the manuscript and agreed the decision to submit for publication.

**Funding** Department of Health Policy Research Programme.

**Competing interests** None declared.

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

**Data sharing statement** Anonymised raw data generated by this study will be available from UCL after a 6-month embargo from the date of publication. Further information is available from the corresponding author.

## REFERENCES

- Coleman MP, Forman D, Bryant H, *et al.* Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011;377:127–38.
- Lyratzopoulos G, Neal RD, Barbiere JM, *et al.* Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012;13:353–65.
- Neal RD, Din NU, Hamilton W, *et al.* Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *Br J Cancer* 2014;110:584–92.
- O'Dowd EL, McKeever TM, Baldwin DR, *et al.* What characteristics of primary care and patients are associated with early death in patients with lung cancer in the UK? *Thorax* 2015;70:161–8.
- Walter FM, Rubin G, Bankhead C, *et al.* Symptoms and other factors associated with time to diagnosis and stage of lung cancer: a prospective cohort study. *Br J Cancer* 2015;112(suppl 1):S6–13.
- Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care—a systematic review. *Fam Pract* 2008;25:400–13.
- Raine R, Wong W, Scholes S, *et al.* Social variations in access to hospital care for patients with colorectal, breast, and lung cancer between 1999 and 2006: retrospective analysis of hospital episode statistics. *BMJ* 2010;340:b5479.
- Macleod U, Mitchell ED, Burgess C, *et al.* Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *Br J Cancer* 2009;101(Suppl 2): S92–101.
- Singh H, Sittig DE. Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. *BMJ Qual Saf* 2015;24:103–10.
- Hrisos S, Eccles MP, Francis JJ, *et al.* Are there valid proxy measures of clinical behaviour? A systematic review. *Implement Sci* 2009;4:37.
- Peabody JW, Luck J, Glassman P, *et al.* Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715–22.
- Kostopoulou O, Oudhoff J, Nath R, *et al.* Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in family medicine. *Med Decis Making* 2008;28:668–80.
- Zwaan L, Monteiro S, Sherbino J, *et al.* Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workshops. *BMJ Qual Saf* 2016 Published Online First 29 Jan 2016. doi:10.1136/bmjqs-2015-005014
- Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making* 2015;35:539–57.
- National Institute for Health and Care Excellence. Referral guidelines for suspected cancer: NICE guidelines CG27, 2005. <https://www.nice.org.uk/Guidance/CG27> (accessed 10 May 2016).
- Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer* 2009;101 (Suppl 2):S80–6.
- Adler MD, Trainor JL, Siddall VJ, *et al.* Development and evaluation of high-fidelity simulation case scenarios for pediatric resident education. *Ambul Pediatr* 2007;7:182–6.
- Shim J, Brindle L, Simon M, *et al.* A systematic review of symptomatic diagnosis of lung cancer. *Fam Pract* 2014;31:137–48.
- Doan S, Conway M, Phuong TM, *et al.* Natural language processing in biomedicine: a unified system architecture overview. In: Trent R, ed. *Clinical bioinformatics*. New York, NY: Springer New York, 2014:275–94.
- Spiegelhalter D, Abrams K, Myles J. *Bayesian approaches to clinical trials and health care evaluation*. Chichester: Wiley, 2004.
- Health and Social Care Information Centre. Workforce. 2014. <http://www.hscic.gov.uk/workforce> (accessed 10 May 2016).
- National Cancer Intelligence Network. Practice Profiles. 2013. [http://www.ncin.org.uk/cancer\\_information\\_tools/profiles/gp\\_profiles](http://www.ncin.org.uk/cancer_information_tools/profiles/gp_profiles) (accessed 10 May 2016).
- Hamilton W, Green T, Martins T, *et al.* Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study. *Br J Gen Pract* 2013;63:e30–6.
- Hamilton W, Peters TJ, Round A, *et al.* What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax* 2005;60:1059–65.
- Banks J, Hollinghurst S, Bigwood L, *et al.* Preferences for cancer investigation: a vignette-based study of primary-care attendees. *Lancet Oncol* 2014;15:232–40.
- Department of Health. *Improving outcomes: a strategy for cancer*. London: Department of Health, 2011. <http://www.gov.uk> (accessed 10 May 2016).
- Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2011;61:e715–23.
- National Institute for Health and Care Excellence. Suspected cancer: recognition and referral. Secondary Suspected cancer: recognition and referral. 2015. [nice.org.uk/guidance/ng12](https://www.nice.org.uk/guidance/ng12) (accessed 10 May 2016).
- Jensen H, Nissen A, Vedsted P. Quality deviations in cancer diagnosis: prevalence and time to diagnosis in general practice. *Br J Gen Pract* 2014;64:e92–8.

- 30 Singh H, Daci K, Petersen LA, *et al.* Missed opportunities to initiate endoscopic evaluation for colorectal cancer diagnosis. *Am J Gastroenterol* 2009;104:2543–54.
- 31 Lyratzopoulos G, Vedsted P, Singh H. Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *Br J Cancer* 2015;112 (Suppl):S84–91.
- 32 Balogh E, Miller B, Ball JE, eds. *Improving diagnosis in health care/Committee on diagnostic error in health care*. Washington DC: The National Academies Press, 2015.
- 33 Zwaan L, Thijs A, Wagner C, *et al.* Does inappropriate selectivity in information use relate to diagnostic errors and patient harm? The diagnosis of patients with dyspnea. *Soc Sci Med* 2013;91:32–8.
- 34 Kostopoulou O, Sirota M, Round T, *et al.* The role of physicians' first impressions in the diagnosis of possible cancers without alarm symptoms. *Med Decis Making* 2016. [Epub ahead of print 25 Apr 2016] <http://dx.doi.org/10.1177/0272989X16644563>
- 35 McBride D, Hardoon S, Walters K, *et al.* Explaining variation in referral from primary to secondary care: cohort study. *BMJ* 2010;341:c6267.
- 36 Tate AR, Nicholson A, Cassell JA. Are GPs under-investigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010;102:947–51.
- 37 Scott SE, Walter FM, Webster A, *et al.* The model of pathways to treatment: conceptualization and integration with existing theory. *Br J Health Psychol* 2013;18:45–65.
- 38 Jack RH, Davies EA, Møller H. Lung cancer incidence and survival in different ethnic groups in South East England. *Br J Cancer* 2011;105:1049–53.
- 39 Kostopoulou O, Rosen A, Round T, *et al.* Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. *Br J Gen Pract* 2015;65:e49–54.
- 40 Hooper LM, Weinfurt KP, Cooper LA, *et al.* Virtual standardized patients: an interactive method to examine variation in depression care among primary care physicians. *Prim Health Care Res Dev* 2008;9:257–68.
- 41 Harries C, Forrest D, Harvey N, *et al.* Which doctors are influenced by a patient's age? A multi-method study of angina treatment in general practice, cardiology and gerontology. *Qual Saf Health Care* 2007;16:23–7.
- 42 Hughes R. Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociol Health Ill* 1998;20:381–400.
- 43 Department of Health. Equity and Excellence: Liberating the NHS. 2010. <http://www.gov.uk> (accessed 10 Jun 2016).
- 44 Reilly JB, Ogdie AR, Von Feldt JM, *et al.* Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ Qual Saf* 2013;22:1044–50.