

## Supplementary file S1

### METHODS

#### Steps in developing the virtual patient application

##### Step 1. Case concept: developing the vignette design and content

The factorial experimental design, informed by reviews of the literature on non-clinical factors affecting GP decision making and lung cancer diagnosis and survival, covered four experimental factors (**Table 1**) known to be associated with variation in lung cancer survival, but whose effect on inequalities in GPs' rates of referral for investigation or to secondary care is uncertain:<sup>1</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, medium 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>2</sup>

We constructed six clinical profiles, two for each risk level using different combinations of symptoms, age, and smoking status. (**Table 2**) The clinical profiles and risk levels were based on positive predictive values (PPV) of lung cancer from the CAPER case-control dataset and the latest available NICE guidance on investigation of suspected cancer.<sup>2,3</sup> This 2005 guidance recommended investigation for symptoms present for >3 weeks - which equates to PPVs >1.2%, though the guidance did not specifically use PPV thresholds. The low-risk vignettes (Profiles 1 and 2) therefore reflected a PPV < 1.2% with symptom duration of 1-2 weeks, such that investigation would not be indicated (**Table 2**). Investigation would be clearly indicated for 'high-risk vignettes (Profiles 5 and 6), which reflected a PPV > 3% with symptom duration more than 5 weeks. In the medium-risk vignettes (Profiles 3 and 4, PPV = 1.7-2.5%), investigation would be consistent with guidance but information on symptom duration was kept intentionally vague so "safety-netting" (i.e. a back-up plan if symptoms persist or escalate) without active investigation could also be appropriate. The first symptom was volunteered by the 'patient,' the second only elicited if GPs specifically asked. In Profile 1 (low risk), symptoms and co-morbidities unrelated to lung cancer, to deflect GPs from the primary purpose of the study.

For each clinical profile a comprehensive set of additional information was developed to include:

- Medical records for each of the 'patients'; similar to what GPs would find in their own clinical system. These incorporated information on socio-demographic and lifestyle

characteristics, past medical history and medication, and a recent consultation history. Consistent with what would be expected for patients of their age, many 'patients' had co-morbidities; but these were selected so in most cases they did not alter the patient's likelihood of lung cancer.

- Results of examinations and tests that GPs might perform: including tests unrelated to the risk profile symptoms to avoid priming GPs' behaviour. In most cases results were the same for all 'patients' with that profile, although some varied according to 'patient' gender. The respiratory and cardiovascular examinations were unremarkable for all six profiles to ensure we were studying GPs' responses to the presence/absence of symptoms, rather than to positive examination findings.

### ***Step 2. Review and Revision by Content Experts***

To maximise the clinical authenticity of the cases, GPs specializing in cancer diagnosis and non-academic GPs reviewed the proposed vignettes, which were then revised following their comments. The website content and functionality were also informed by patient representatives' comments. This led to the inclusion of non-smokers because of the risk of diagnostic delay if GPs are less likely to suspect cancer in this group. It also directly informed the types of responses 'patients' provided, where patient representatives corroborated previous research that patients may well not disclose certain symptoms with their doctors without being directly asked about them.

### ***Step 3. Outline and Flow Development: A typical consultation in the study***

We developed the outline flow of a typical consultation on the application, the duration of which would be determined by what the GP sought to find out:

- GP enters "waiting room" and clicks on "Patient" link
- GP selects question "What seems to be the trouble?"
- Video plays where the 'patient' volunteers their first symptom; the view is a head shot of 'patient' in GP consulting room.
- GPs can find out additional information through:
  - asking the 'patient' questions (e.g. on the nature of a symptom, presence of other symptoms). 'Patient' videos then play giving the GP requisite information. If the system was unable to provide a meaningful response, users receive an error message.
  - consulting medical records (e.g. on previous consultations, medications),
  - performing examinations (e.g. blood pressure, with findings provided as text).
- GP selects "Make the final note" where they enter ideas about diagnosis (main, possible, possible but unlikely) and their management plan.
- The GP completes six such consultations over ~3 weeks.

**Table 1: Each of the 36 vignettes combinations representing the four experimental factors: gender, socioeconomic circumstances, ethnicity, and risk level (across two clinical profiles)**

		<b>Low risk (PPV &lt;1.2%)</b>		<b>Medium risk (PPV = 1.7-2.5%)</b>		<b>High risk (PPV ≥ 3%)</b>	
		Clinical Profile 1 PPV=0.4 [0.1-3.1]	Clinical Profile 2 PPV=0.4 [0.1-3.1]	Clinical Profile 3 PPV=0.4 [0.1-3.1]	Clinical Profile 4 PPV=0.4 [0.1-3.1]	Clinical Profile 5 PPV=0.4 [0.1-3.1]	Clinical Profile 6 PPV=0.4 [0.1-3.1]
<b>Disadvantaged</b>		1. South Asian	2. South Asian	3. South Asian	4. South Asian	5. South Asian	6. South Asian
		7. Black	8. Black	9. Black	10. Black	11. Black	12. Black
		13. White	14. White	15. White	16. White	17. White	18. White
<b>Advantaged</b>		19. South Asian	20. South Asian	21. South Asian	22. South Asian	23. South Asian	24. South Asian
		25. Black	26. Black	27. Black	28. Black	29. Black	30. Black
		31. White	32. White	33. White	34. White	35. White	36. White

<b>Key:</b>	Male	Female
-------------	------	--------

**Table 2. 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		
<b>Low risk: Expected action = no active investigation (safety netting appropriate)</b>							
1	Younger (Late fifties)	Non smoker	Breathlessness	Fatigue	1-2 weeks	0.40%	Patient has swollen ankles, possibly due to heart failure
2	Younger (Late fifties)	Smoker	Chest pain	Cough	1-2 weeks	1.10%	
<b>Medium risk: Expected action = either investigation (e.g. order chest x-ray) or safety netting</b>							
3	Older (Late seventies)	Smoker	Chest pain	Cough	Uncertain (approx 3 weeks)	1.70%	
4	Older (Late seventies)	Non-smoker	Cough	Appetite loss	Uncertain (approx 3 weeks)	2.50%	
<b>High risk: Expected action = lung cancer investigation</b>							
5	Younger (Late fifties)	Smoker	Breathlessness	Fatigue	>5 weeks	3-4%	Chronic obstructive pulmonary disease (COPD) present
6	Older (Late seventies)	Smoker	Chest pain	Weight loss	>5 weeks	14%	

#### ***Step 4. Translation of vignette content into simulation platform: the virtual patient application website***

The translation of content into the virtual patient application website required filming actors portraying patients, creating and populating the website with that content. Twelve actors with medical role-playing experience were each filmed playing three 'patients' in a studio resembling a GP's consulting room. Actors were selected to fulfil the 'patient' template of the factorial design, i.e. every combination of three ethnicities and male/female across the two age groups (58/59 year olds, and 78/79 years). Each actor received a detailed brief for three 'patients'. This contained profile information (e.g. symptom presentation and features) plus details relating to the specific character (e.g. occupation). Actors represented the socio-economic circumstances of their 'patient' profiles through appearance (e.g. clothing, hairstyle, makeup), accent and information about their occupation/lifestyle. In each case actors started with an introduction to their presenting symptom - how one might answer a GP's initial question, "What seems to be the trouble?" – and continued with responses to additional questions about specific features of the presenting symptom, additional symptoms and their features, and other relevant subjects (e.g. smoking status). Actors were asked to describe these symptoms in their own words but had example scripts provided by patient representatives of how real patients might describe their experiences and sensations. To ensure consistency in content across all the vignettes, the researchers used checklists to ensure the actors had mentioned all the details relevant to their profiles.

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 Doan et al 2014 for an explanation of the principles).<sup>4</sup> This system was underpinned by databases on symptoms or risk factors and the features those symptoms (e.g. what exacerbates or relieves the symptom or how long it has been present). Each database was populated with a set of key words (including common typographical errors) which GPs might use to ask about the existence and features of these symptoms, see Table 3 for an example. The key words were initially developed by the research team in consultation with content experts and subsequently extended to enable the system to respond to the language and content of questions asked by GPs in pilots. Finally, the website was populated for each vignette with:

- Over 300 videos of the 'patient' actors describing symptoms and main risk factor responses to provide answers to GPs' typed in questions. This included a generic "No" or "Don't understand" where there was no clinically relevant information available.
- Text (available as drop-down menu) for all other aspects.

**Table 3. Database example: Key words for Dyspnoea (symptom) and onset (feature)**

symptom or risk factor	shortness breathless breathlessness breathe dyspnoea puff short of breath lost breath lose breath catch breath breatlessness breatless breathing difficulty breathing trouble breathing out of breath	Onset	what brings exacerbates what triggers makes it happen start to happen causes exacerbate aggravate aggravates agrivate aggrivate agrivates aggrivates especially bad aggrivate aggrivates makes it worse exacerbation pleuritic plueritic deep breath taking a breath take a breath breathing in breathe in breath in
------------------------	---	-------	---

***Step 5. Pilot testing***

Three pilot stages were conducted to identify changes needed to content, functionality and design. In stage 1, researchers were present whilst three GPs tried up to four online consultations to identify any problems in using the application and where additional vignette content was needed. In stage 2, GPs (n=7) conducted up to four online consultations remotely. After their pilot, they participated in a telephone interview with a researcher to provide feedback on the intuitiveness of the application, credibility of the vignettes, the consultation process and the extent to which they were able to use similar reasoning as in their day-to-day practice. In stages 1 and 2, researchers (JMc, RS, JS) reviewed participants' log forms to identify where GPs' questions led to an error message or an inappropriate video response. Revisions to the website databases and functionality were revised in response. In stage 3, researchers not connected with the study (n=10) conducted up to four consultations

to check whether errors from pilots 1 and 2 had been corrected. See S1 for details of revisions made as a result of the pilots.

We made changes following pilots to the appearance, content, and functionality of the application:

- **Appearance:** We altered the design after the first pilots to ensure GPs could see all the opportunities to find out information on onscreen without scrolling down and changed the colour scheme in response to pilot feedback.

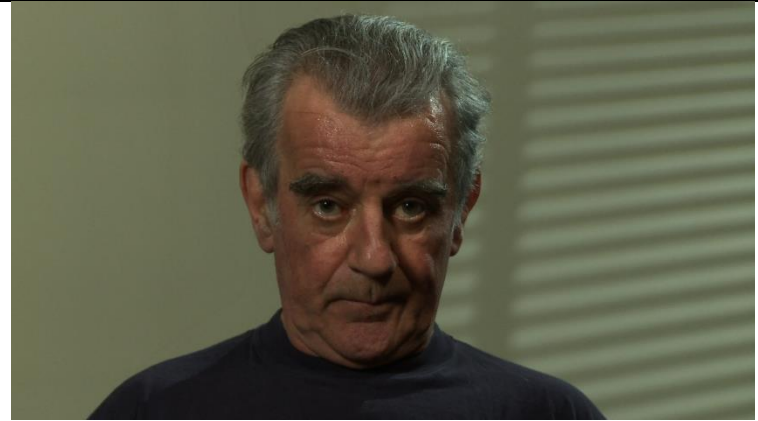
**Content:** In response to GP feedback in the early pilots, we filmed longer 'patient' clips describing symptoms (from 15-30 secs to 45-60 secs) with less relevant clinical information (from describing the nature and frequency of symptoms to just reporting presence of symptom and instead recounting effects on daily life). In addition, we added more content for each profile was developed and filmed to provide answers to a wider range of questions.

**Function:** Using the log file data from the pilots, the symptom and features databases were extended and refined to enable the website to provide more meaningful answers to GPs' questions.

There were limitations in natural language function that could not be further overcome. For example the application required GPs to repeat the name of the symptom they were asking about in all their questions (e.g. 'how long have you had chest pain' or 'what makes the breathlessness worse') which does not realistically mimic spoken conversation. We used data from the log files on where these caused GPs to get error responses in the pilots to inform development of a help video and PDF that GPs could access whilst using the application. We also provided GPs with feedback after their first 'consultation' to reduce the likelihood that they missed key information in future 'consultations' because of repeated error.

**Profile 1: complete vignette**

The description below illustrates the information available for each profile and how GP participants could access it. It is shown for the first vignette profile, the ‘deflecting’ vignette, where the risk of lung cancer is lowest and the most probable diagnosis is heart failure

		<p>Thumbnail: ‘Jack Jones’ White British man, aged ~60 years, works on security in a block of offices. Non-smoker, has diabetes and a history of depression,</p>
---	--	--

Data item	Accessed by	Information	Format
Presenting symptom	Video – Displayed when participant clicked on default question on screen	<u>Breathless</u>	Patient account: <i>Never felt like this before and he is not sure what’s going on. It’s interfering with his life (e.g. had to get the bus into work rather than walking) and so wife suggested he come and check it out.</i>
Second symptom	Video – Displayed if participant used text box to ask a direct question about presence of symptom.  <i>Synonyms recognised included: tiredness, tired Energy, lethargic, lethargy,</i>	<u>Fatigue</u>	Patient account: <i>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 doesn’t feel stressed.</i>



	<i>drained, exhaustion Exhausted, fatigue, fatigued, sluggish, knackered pooped</i>		
Further information on symptom characteristics	<p>Video – Displayed if participant used text box to ask a direct question about characteristics of the symptom.</p> <p><i>For breathlessness, questions that could be addressed included: How long have you been breathless? What makes it better? What makes it worse? How far can you walk? Is it worse on exercise? Is it worse when you lie down? Does it stop your normal activities? Can you carry things? Have you ever had this before? Do you have chest pain? Do you have swollen ankles? Have you had calf swelling? Do you have asthma? Do you have COPD? Are you a smoker? Do you have heavy periods?)</i></p>	<ul style="list-style-type: none"> <li>• Duration</li> <li>• Onset</li> <li>• Offset</li> <li>• Frequency</li> <li>• Effect of: exercise, lying down</li> <li>• life changes</li> <li>• Diet, bowel</li> <li>• Position (of pain)</li> <li>• Illness ideas</li> <li>• Family history</li> <li>• Medication</li> <li>• Related symptoms</li> </ul>	<p>Patient account: <i>Notice it particularly when active (e.g. 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's been happening for 1-2 weeks (e.g. trains young boxers at the local gym but hasn't been able to make boxing training for the last week and a half because of it).</i></p>

**Information available from drop-down menus**

<b>Patient information</b>		<b>Comments</b>
Name	Jack Jones	Consistent with ethnicity and socioeconomic circumstances
Date of birth	19.05.1954	
Gender	Male	Also evident from patient video
Address (first line)	xxx	Consistent with socio-economic circumstance as far as possible...
Ethnicity	White	Also evident from patient video (dress, accent)
Occupation	Security guard	Also available as patient video (in response to questions about occupation, job etc)
<b>Lifestyle factors</b>		
Smoking status	Never smoked	Also available as video if patient asked through text box
Units of alcohol per week	25 units	Consistent with socioeconomic and ethnicity profile and set so as not to raise suspicion that alcohol misuse caused symptoms.
BMI	xx kg/m <sup>2</sup>	See weight
Family history	None recorded	

<b>Systems Examinations</b>	<b>Information</b>
Abdomen (including rectal)	Soft and non tender. No abnormalities detected.
Breast	Normal.
Cardiovascular system (note to us, include heart rate/rhythm)	Heart rate 72 beats/minute. Regular rhythm. Normal hearts sounds. No sacral or peripheral oedema.
ENT examination	No abnormality detected.
Eye examination (including fundoscopy)	No abnormality detected.
Foot examination	Pulses palpable. Sensation normal.
Genitalia examination	No abnormality detected.
Heart rate/rhythm	Heart rate 72 beats/minute. Regular rhythm.
Nail examination	All nails appear normal.

Neurological examination, central (including cranial nerves)	No abnormality detected.
Neurological examination, peripheral	No abnormality detected.
Peripheral pulses	All pulses palpable. No abnormality detected.
Respiratory rate	14 breaths/minute.
Respiratory system	Rate: 14 breaths/minute. No peripheral or central cyanosis. Good chest movement. Chest clear.
Joint examination, cervical spine	Good range of pain-free movement.
Joint examination, shoulder	Both joints normal in appearance and movement.
Joint examination, elbow	Both joints normal in appearance and movement.
Joint examination, wrist	Both joints normal in appearance and movement.
Joint examination, hand	Joints normal in appearance and movement.
Joint examination, thoraco-lumbar spine	Normal gait. Good range of pain-free movement.
Joint examination, hip	Both joints normal in appearance and movement.
Joint examination, knee	Both joints normal in appearance and movement.
Joint examination, ankle	Both joints normal in appearance and movement.
Joint examination, foot	Joints normal in appearance and movement.
<b>Bedside tests</b>	<b>Information</b>
Blood glucose	6.7 mmol/L
Blood pressure	140/80 mmHg
Cultures	Sputum sample provided and sent to laboratory
Height	180cm (men); 163cm (women)
Peak flow	575l/min (men); 390l/min (women)
Swabs	Swabs taken and sent to laboratory
Temperature	36.5 <sup>0</sup> C
Urinalysis	Urinalysis normal
Weight	What seems reasonable for actor/actress

<b>Significant medical history</b>	<b>Information</b>
Co-morbidities & date of diagnosis	Diabetes mellitus 24.11.09
	Depression 05.01.11
	Allergies None recorded
<b>Current medication</b>	<b>Information</b>
Drug name, dose, instructions for use	Metformin 500mg bd
	Fluoxetine 20mg od
<b>Recent appointment history</b>	
	01.08.12 Diabetes Review Taking metformin 500mg bd, no problems. BP: 140/90 (on ramipril, amlodipine). HbA1c: 7.1. Normal FBC, renal function, cholesterol. Foot check: normal sensation.
	25.10.11. Cellulitis. Cellulitis L great toe (following cut). Apyrexial, does not appear unwell. Rx: flucloxacillin 250mg and penicillin (V) 250mg qds (7 day course). Advised to return if not resolving in 5 days. 09.08.11 Diabetes Review Taking metformin 500mg bd, no problems. HbA1c: 7.5. Foot check: normal. Discussed dietary compliance.

## References

1. 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-s101.
2. 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.
3. NICE. Referral guidelines for suspected cancer: NICE guidelines CG27, 2005.
4. 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.