

# Identifying early warning signs for diagnostic errors in primary care: a qualitative study

John Balla,<sup>1</sup> Carl Heneghan,<sup>1</sup> Clare Goyder,<sup>1</sup> Matthew Thompson<sup>2</sup>

**To cite:** Balla J, Heneghan C, Goyder C, *et al*. Identifying early warning signs for diagnostic errors in primary care: a qualitative study. *BMJ Open* 2012;**2**:e001539. doi:10.1136/bmjopen-2012-001539

► Prepublication history and additional material for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-001539>).

Received 22 May 2012  
Accepted 9 August 2012

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see <http://bmjopen.bmj.com>

<sup>1</sup>Centre for Evidence-based Medicine, University of Oxford, Oxford, UK

<sup>2</sup>Department of Primary Health Care, University of Oxford, Oxford, UK

## Correspondence to

Dr John Balla;  
[john.balla@phc.ox.ac.uk](mailto:john.balla@phc.ox.ac.uk)

## ABSTRACT

**Objective:** We investigate the mechanisms of diagnostic error in primary care consultations to detect warning signs for possible error. We aim to identify places in the diagnostic reasoning process associated with major risk indicators.

**Design:** A qualitative study using semistructured interviews with open-ended questions.

**Setting:** A 2-month study in primary care conducted in Oxfordshire, UK.

**Participants:** We approached about 25 experienced general practitioners by email or word of mouth, 15 volunteered for the interviews and were available at a convenient time.

**Intervention:** Interview transcripts provided 45 cases of error. Three researchers searched these independently for underlying themes in relation to our conceptual framework.

**Outcome measures:** Locating steps in the diagnostic reasoning process associated with major risk of error and detecting warning signs that can alert clinicians to increased risk of error.

**Results:** Initiation and closure of the cognitive process are most exposed to risk of error. Cognitive biases developed early in the process lead to errors at the end. These warning signs can be used to alert clinicians to the increased risk of diagnostic error. Ignoring red flags or critical cues was related to processes being biased through the initial frame, but equally well, it could be explained by knowledge gaps.

**Conclusions:** Cognitive biases developed at the initial framing of the problem relate to errors at the end of the process. We refer to these biases as warning signs that can alert clinicians to the increased risk of diagnostic error. We conclude that lack of knowledge is likely to be an important factor in diagnostic error. Reducing diagnostic errors in primary care should focus on early and systematic recognition of errors including near misses, and a continuing professional development environment that promotes reflection in action to highlight possible causes of process bias and of knowledge gaps.

## INTRODUCTION

Diagnostic error has been defined ‘as a diagnosis that was unintentionally delayed (sufficient information was available earlier),

## ARTICLE SUMMARY

### Article focus

- Investigation of the relation of clinical reasoning to diagnostic error.
- Identifying places in the diagnostic reasoning process associated with major risk of error.
- Detection of warning signs that can alert clinicians to the potential of increased risk of error.

### Key messages

- Initiation and closure of the cognitive process are most exposed to the risk of diagnostic error in primary care.
- Cognitive biases developed at initiation may be used as warning signs to alert clinicians to the increased risk of diagnostic error.
- Deliberate practice of looking for warning signs may be a potential method of professional development to reduce error through reflection in action.

### Strengths and limitations of this study

- The design of the study is based on a strong theoretical framework provided by the dual theory of cognition and is based on real-life situations.
- Hindsight bias may lead to unjustified causal explanations of errors.
- Limitations concern generalisability, participants consisting of self-selected individuals and lack of replication in other cohorts.

wrong (another diagnosis was made before the correct one), or missed (no diagnosis was ever made)’.<sup>1–3</sup>

These definitions have also been refined to emphasise the ‘longitudinal aspect of diagnosis’, where a precise diagnosis need not be made, or it can wait, ‘because other decisions may take priority’.<sup>4</sup> The relative importance of diagnostic errors to the widespread health problem of medical errors in general is unclear, yet several US sources suggest that they are a major contributor.<sup>5</sup> A systematic review of diagnostic error in primary care in the UK quoted diagnostic error accounts for the greatest proportion of medical malpractice

claims against general practitioners (GPs) (63%).<sup>6</sup> An analysis of 1000 claims of GPs in the UK identified 631 alleged delayed diagnoses.<sup>7</sup> A New Zealand review of primary care claims showed that delays in diagnosis were few, but associated with a disproportionate number of serious and sentinel injuries (16%) and deaths (50%).<sup>8</sup> Studies of diagnostic error outside of insurance claims generally focus on hospital populations and specialties.<sup>9</sup> A recent review quotes a diagnostic error rate of between 5% and 15%, depending on the specialty and methods of data collection.<sup>10</sup>

Several commentators have noted the importance of attempting to reduce diagnostic error by understanding in more depth the cognitive reasoning processes underlying diagnostic decision-making.<sup>11</sup> Winters *et al*<sup>12</sup> noted that any attempts to improve diagnostic safety must 'intuitively support how our brains work rather than how we would like them to work.' Norman<sup>13</sup> noted the potential importance of reflecting on the clinician's 'own performances and identify places where their reasoning may have failed'. Indeed, Elstein<sup>11</sup> suggested that a major objective for improving cognitive reasoning in future research should include delineating how feedback on clinical reasoning could be used to 'guide clinicians to identify priority tasks for reducing diagnostic errors'. Ely *et al*<sup>4</sup> propose that one method for doing this could be to identify those red flags 'that should prompt a time-out' as a priority for the study of diagnostic error.

We have previously suggested a model of the Dual Theory of Cognition (DTC) as a framework to explore diagnostic reasoning in primary care. This identifies high-risk places in the diagnostic reasoning process where the clinician is most likely to commit an error is consistent with our objective.<sup>14</sup> In this study, we investigate the mechanisms of diagnostic error in primary care clinical practice using qualitative methods with a well-defined cohort and a strong theoretical base.

## METHODS

As indicated in the introduction, the majority of studies of diagnostic error rely on retrospective reviews of data collected for malpractice claims. A lesser number of reports deal with hospital incidents and consist of ill-defined cohorts. In these reports, a mix of poorly defined methodologies is used. What happened during the actual consultation is rarely understood and by the nature of the data unavailable. Instead, chart reviewers tend to make personal judgments. A further confounding problem relates to the unavoidable fact that clinicians may never find out about their errors and if they do, this could be weeks or months after the event. These issues will be discussed in more detail in the section on strengths and limitations of the study reported here. Our choice of method, relying on interviews for data collection was made with the aim of getting a little closer to the truth of what happens when a consultation effectively fails. Knowing the limitations of the method, our

expectation was that this may lead us, and hopefully others, to further fruitful research. We leave the discussion on these matters to the section on Implications for further research, at the end of this paper.

## Design

Semistructured interviews using open-ended questions were based on a model of clinical reasoning derived from the DCT. This is a generally accepted model of human cognition that encompasses an initial fast response to a problem (System 1) that occurs at a sub-conscious level and is associated with or followed by a slow analytic, reflective response (System 2).<sup>15</sup> The initial response is based on<sup>16</sup> a mental construct made up of the individual's interpretation of past experience, understanding of theories of professional knowledge, value judgments and the social context. Initial judgments based on these pre-existing constructs in relation to the context of the new case, lead to the construction of a frame<sup>17</sup> that limits and gives direction to the rest of the cognitive process.<sup>18</sup>

Closure of the process may occur with minimal or no exposure to the analytic System 2. Just how much analysis occurs depends on the relative dominance of System 1. The use of closure rules<sup>19</sup> is consistent with System 2. Diagnostic error frequently occurs at initiation, but most often at closure and both Systems 1 and 2 may be involved.<sup>4</sup> It is the congruence of the DTC with what we know about clinical reasoning that leads us to use it as the model for this study.

## Participants

Primary care clinicians (known as general practitioners or GPs in the UK) were asked to first define diagnostic error and then discuss such cases from their practice. Fifteen GPs (>5 years in practice) working in primary care in Oxfordshire, provided 45 cases. The model highlights key areas of knowledge and critical value judgments that are used in the clinical encounter that provide the questions for our interviews. These are described in appendix 1. We agreed that data saturation had been reached about two-thirds of the way and did not recruit more participants.

## ANALYSIS

Three researchers searched the text independently for underlying themes in relation to our conceptual framework. The steps in the dual theory provided the structure for this, so that it started with themes for the initial salient features of the case and ended with closure. Categories were coded according to emerging themes and added to or changed as new concepts emerged.<sup>20</sup> The researchers compared their findings in person, on Skype and through emails. This occurred up to once or twice a week over a 2–3 month period. Differences in interpretation were few and were discussed. There were no major differences and consensus was reached on

clustering to common themes. Direct quotes from the interviews appear in italics, numerics preceded by G are interview identifiers.

## RESULTS

Participants indicated that our questions were easy to follow and most commented that they enjoyed the interview as a good way to reflect on practice.

### Definitions of diagnostic error

The definitions of diagnostic error provided by respondents were consistent with previous definitions. Over half of the 45 cases analysed were associated with the clinician focusing on a single diagnosis from presentation to closure of the cognitive process. Respondents divided errors into *two different categories (G12)*: first, the wrong diagnostic label consisting of a *diagnosis that's wrong or proven to be wrong by yourself or someone else at a later date (G3)* and second, delayed diagnosis described as *missed the boat you should have done something but you didn't (G2)*.

GPs raised two issues that caused difficulties in defining an error, namely variation in how to deal with the severity of the impact of the error: *I rarely give someone a firm diagnosis ... it would be an error if there was something serious and I had told someone [it] wasn't (G8)*, and also what constitutes unacceptable delay given that *lots of what we see is at low prevalence and evolving, so at the very front end it's very vague so actually most of that by definition should be delayed (G1)*.

### Process of clinical reasoning in relation to diagnostic error

Results are reported with reference to the 45 cases and not individual GPs. The cohort provided sufficient data for a case-by-case analysis of errors, but not for comparing across the 15 GPs, though there were no indications of differing themes between individuals. The themes identified divided into two main groups: initiation and setting the initial diagnostic frame, followed by stopping the search for further clinical information and achieving diagnostic closure. Additional themes which emerged are also discussed.

### Initiation and setting the initial diagnostic frame

Salient features link the individual clinician's personal knowledge of similar cases to the new presentation. A number of themes emerged from our analysis to shape the frame for the new case in its specific context. In most cases, GPs formed instantaneous diagnoses. For example, the patient's appearance—*thought actually looked OK for a first child (G19)*, or, *she didn't sound too unwell over the phone (G9)*—before other information was available, provided a powerful bias for framing the case. In over 2/3 (31/45) of cases, the focus was on a single diagnosis, in about half of these based on presentation

with a pre-existing diagnostic label (16/31). Box 1 reconstructs such a case.

The patient's history of previous psychosocial problems or abnormal behaviour, were predominant at this stage of the presentation: *previous consultation which had set her up as a particular kind of person (4)*. Other salient features led GPs to instant recognition of a diagnosis or a limited number of differentials. For instance, *just focusing on the vomiting (39)*, made the GP think of a gastric problem and delayed the diagnosis of an obstructed hernia. Wrong localisation framed the cognitive process and biased further information gathering, directly impacting on diagnostic closure thresholds.

Participants made repeated references to needing to focus on the natural history of disease and expected response to treatment. For instance, in referring to a case of missed cancer: *people that have haemorrhoids that respond beautifully to treatment and have no other symptoms we don't tend to think, oh have they got a colonic cancer (14)*. Or, the need for experience: *lacked experience at that time [to] potential of this case.... and work on the possible diagnosis (9)*. Most of these references to experience were suggestive of knowledge gaps rather than cognitive error only. Further examples of biases arising from the initial framing appear in [table 1](#).

### Conditions and thresholds for diagnostic closure

Participants did not use numerical criteria to describe the thresholds they used to decide when to stop searching for more clinical information. Nor did they express confidence in their decisions this way. When pressed, some responded in terms such as: *the test would exclude X in 70% of cases*, or *I was more than 80% certain that I excluded Y*, but the basis for these numerical values was very unclear. Since we felt that a number of participants found these questions judgmental, we therefore dropped them as the interview schedule progressed.

A number of GPs raised safety netting spontaneously, or in response to our questions, related to diagnostic closure. It soon became apparent that recollections were hazy and they were unsure as to whether they actually used safety netting, or just thought that they should have. Therefore, we do not include safety netting in our analysis, recognising its importance as a potential cause of error.

The decisions made at closure were affected by biases from the setting of the initial frame, effectively impeding the reflective System 2 review expected at this stage. Box 2 is a reconstruction of such a case to provide insights into how the relationship between biases formed at initiation may affect decisions at closure.

Other themes related to diagnostic errors after ignoring or misinterpreting the predictive value of critical information coming from the patient, as did ignoring 'gut feelings'.<sup>21</sup> Other respondents noted the need to be circumspect when responding to patient needs, including poor outcome with a patient who did not wish to follow advice: *had a cruise booked and he chose to cancel the appointment and go on the cruise (15)*. Some GPs raised



## Warning signs for diagnostic errors

### Box 1 Case 27: illustration of initiation of the process and setting the initial diagnostic frame

Reconstruction	Analysis
<p>Presentation</p> <p><i>Patient in 70's came with breathlessness ... the first thing was he kept saying to me 'this is exactly like it was about 6 months previously' ... looked back in his notes and 6 months previously he'd been diagnosed with heart failure and so I thought 'well, you know' and he was so insistent that it was the same thing .... he was a bit breathless but there wasn't anything really obvious going on ... so I thought maybe he was anaemic as well and that had got worse. And it was a Saturday morning so I couldn't easily get any tests straight away so I booked for him to come back first thing on Monday morning for blood tests and ECG and I sent him up to the hospital for a chest x-ray. He was so insistent that it was the same thing and in retrospect that was really misleading for me</i></p> <p>Context issues</p> <p><i>It was a Saturday morning so I couldn't easily get any tests straight away</i></p> <p>Outcome</p> <p><i>Next day contacted by one of his friends ... to say 'actually, he's had a pulmonary embolism ... he'd got quite a lot worse that afternoon and been admitted to hospital and ...CT showed multiple pulmonary emboli</i></p> <p>Summary</p> <p>System 1 single diagnosis based on the existing label immediately jumps to the diagnosis. Weak System 2 affected by context issues, delaying diagnosis.</p>	<p>Salient feature was patient's insistence that the diagnosis was the same as previously, seemingly confirmed by looking at case notes of his previous presentation</p> <p>System 2 in action as tests ordered, largely to rule in cardiac failure and rule out possible complicating factor of anaemia</p> <p>Management affected by practice environment—routine blood tests not immediately available</p> <p>Delay in diagnosis likely to System 1 overpowering System 2, raising closure threshold enough to be affected by context issues (no blood tests available at weekend)</p>

**Table 1** Biases arising from salient features of presentation which initiate the diagnostic process and frame the direction of subsequent information gathering

Previous diagnosis label	<i>Because somebody had wrote down that he had bell's palsy and he'd been seen in hospital .... I immediately thought that's what he had (1)</i>
Pre-existing psychosocial problems	<i>Story of the insect bite and that was what we were sort of using as our diagnostic tool really (6)</i>
Reassurance from initial appearance	<i>All thought some of the bleeding might be from sexual abuse (31)</i> <i>Sick notes, and prescriptions and whatever and I thought that that was probably the main reason behind the um sort of um consultations (37)</i>
Similarity to a recent case or similarity to representative case built from experience	<i>When I called the patient back I got hold of the granny who said oh yes mum's in the shower that as a clue to me meant that maybe the child wasn't that ill (11)</i> <i>She wasn't terribly unwell (33)</i>
Incorrect localisation of salient features	<i>My diagnosis was fed by a patient the previous week who'd presented with an ischemic foot (40)</i> <i>And I thought he had cancer because of the mass and the weight loss and the paleness (44)</i>
Common things occurring commonly (probabilistic reasoning)	<i>Vomiting and sweating and diarrhoea ... epigastric pain (10)</i> <i>Epigastric discomfort . ...must be indigestion (20)</i> <i>Viral infections are common (16)</i>
Ignoring as well as over or under estimating red flags or critical cues	<i>My preconception at the time was that a young &lt;30 year old is very, very unlikely to have bowel cancer (32)</i> <i>He came in hopping, which is quite unusual. Not weight bearing at all is quite unusual (30)</i>
Vague presenting symptoms, no salient features recognised	<i>Normal chest on examination (24)</i> <i>Fatigue from whatever cause (3); it was all very vague (28)</i> <i>Atypical leg pain couldn't work out what was going on (21)</i>

**Box 2** Case 14: illustration of dominant System 1 impeding System 2 review at closure, leading to error**Reconstruction****Presentation**

*Elderly patient seen 6 years ago for what appeared to be resolving haemorrhoidal bleed ... 6 months prior [to the most recent visit] described narrow stools like a snake ... [At the present visit] bowel frequency and some bleeding with examination of clear external piles no rectal masses on rectal examination. [Also] did some bloods but wasn't anaemic [I ignored] the older the patient the lower the threshold for colorectal cancer that we would have for referring ... red flag that's there for a reason therefore it would be foolish to sort of dismiss...*

**Salience**

*External piles with a normal PR ... [6 years ago] with haemorrhoides seen by a colleague.*

**Outcome**

*2 months after last visit ... change in bowel habit with rectal bleeding and as part of investigation had a sigmoidoscopy and biopsy which found a malignant colonic tumour*

**Summary**

System 1 single diagnosis based on the label immediately jumps to the diagnosis. Ignored expected natural history, and the presence of a red flag. Diagnosis was delayed until a new critical cue emerged.

**Analysis**

System 1 dominance may explain the high threshold for vigilance in this age group

No significant attempt to rule out and normal Hb wrongly used for rule out. Another example of the power of a perceived label in biasing process

Salient feature was a normal examination 6 years earlier

Delay in diagnosis likely to System1 overpowering System 2, raising closure threshold

issues about their own behaviours: *one's own state of confidence or call it what you want competence confidence arrogance or risk taking or not risk caution all play in the actual what you decide to do* (12); *I'm right at one end of low referrers* (43). Contextual factors were often raised as contributing to faulty decisions: *[knowing how busy it is before a week end] do I want to send a frail, elderly lady up to the hospital on a Friday afternoon when it would be mayhem [there]* (34); explaining a missed diagnosis: *we were really busy and I think they came in as an emergency* (39). Table 2 provides examples of biases affecting thresholds for ruling disease in or out.

**DISCUSSION****Main findings**

The findings of this study identify the initiation and closure of the cognitive process in the clinical consultation as those most exposed to risk of diagnostic error. Initiation is a critical step as it sets the frame for subsequent information search, whereas closure occurs when thresholds for stopping the search have been met. We show that cognitive biases developed at framing appear to relate directly to errors at the end of the process. We refer to these as warning signs (table 2), as we believe they can be used to alert the clinician to the increased risk of diagnostic error. Previous studies<sup>4</sup> have also highlighted these two steps as the points where most cognitive errors occur. However, our findings build on these, providing insights into the underlying sources of the

biases that made the process go wrong. This is consistent with Wearth and Nemeth's<sup>17</sup> observation that '[we] do not learn much by asking why the way a practitioner framed the problem turned out to be wrong. We do learn when we discover why that framing seemed so reasonable at the time'.

The initial process of framing a new case is mostly subconscious, and occurs within the domain of System 1 of the DTC (instant, fast response). The salient feature of the case leads to recognition of similar cases, and framing is modulated by constructs from experience, knowledge, values and social context.<sup>22</sup> If there is no instant recognition, System 2 (slow, analytic thinking) may be engaged, but this may consist of no more than gut feelings.<sup>21</sup> In the cases analysed here, there was a dominant focus on a single diagnosis. It is likely that other options were also entertained but forgotten by interviewees. The most significant biases occurring at this stage related to patients with pre-existing diagnostic labels and those with underlying psychosocial problems (table 1). Ignoring red flags or critical cues may have been related to processes being biased through the frame, but equally well, they could be explained by lack of knowledge of the significance of these clinical features.

We have previously suggested<sup>23</sup> that the informal 'rules' that clinicians use to cease their search for further clinical features during a consultation (ie, stopping rules for diagnostic closure) involved three criteria: (1) high-risk conditions have been excluded and other

**Table 2** Effect of framing biases on closure thresholds for ruling disease in or out

Presents with diagnosis label	<i>I'd keyed in too quickly and then just ignored any of the sort of differential information (1)</i>
	<i>When your brain immediately jumps to the obvious diagnosis its worth just having in the back of your mind what else it could be (6)</i>
Psychosocial label/behavioural	<i>I closed it before she came in ... I think hadn't really thought out the differential diagnosis (4)</i>
	<i>Not appreciating the seriousness of the, of the problem, coupled with not really wanting to think about it because the patient was so difficult. (31)</i>
Ignores red flag	<i>[did not] take a step back and consider what we call the sort of red flagged ones, are there any flags in front of you that are presenting information of other serious diseases that might kill or harm? (2)</i>
	<i>Think I ought to have thought this severe pain which isn't improving I ought to go back to cancer but so I was put off by the negative investigations and that kind of prior assessment and err ... level of pain which was not otherwise explained (15)</i>
Ignores possibility of serious disease with low probability	<i>[ignored] older the patient the lower the threshold for particularly for colorectal cancer that we would have for referring ... red flag that's there for a reason therefore it would be foolish to sort of dismiss (14)</i>
	<i>My preconception at the time was that a young 28 year old is very, very unlikely to have bowel cancer ... slightly raised C-reactive protein...it wasn't dramatically raised ... I certainly didn't act on it because I think I was confused by the fact he'd got better the second consultation (32)</i>
Used wrong clinical features to rule-out a condition	<i>[ignored] new onset quite severe headache in a (40) something year old is a red flag in itself (22)</i>
	<i>We think of ectopic pregnancy as being bleeding and pain and this was painless bleeding (17)</i>
Ignored gut feelings	<i>it's a sixth sense ... that I think as you gain more experience you really hone and fine tune ... it's invaluable particularly with children 19; was not terribly unwell ... obviously needed more investigations ... wasn't happy with my decision even though it wasn't a conscious process. (33)</i>

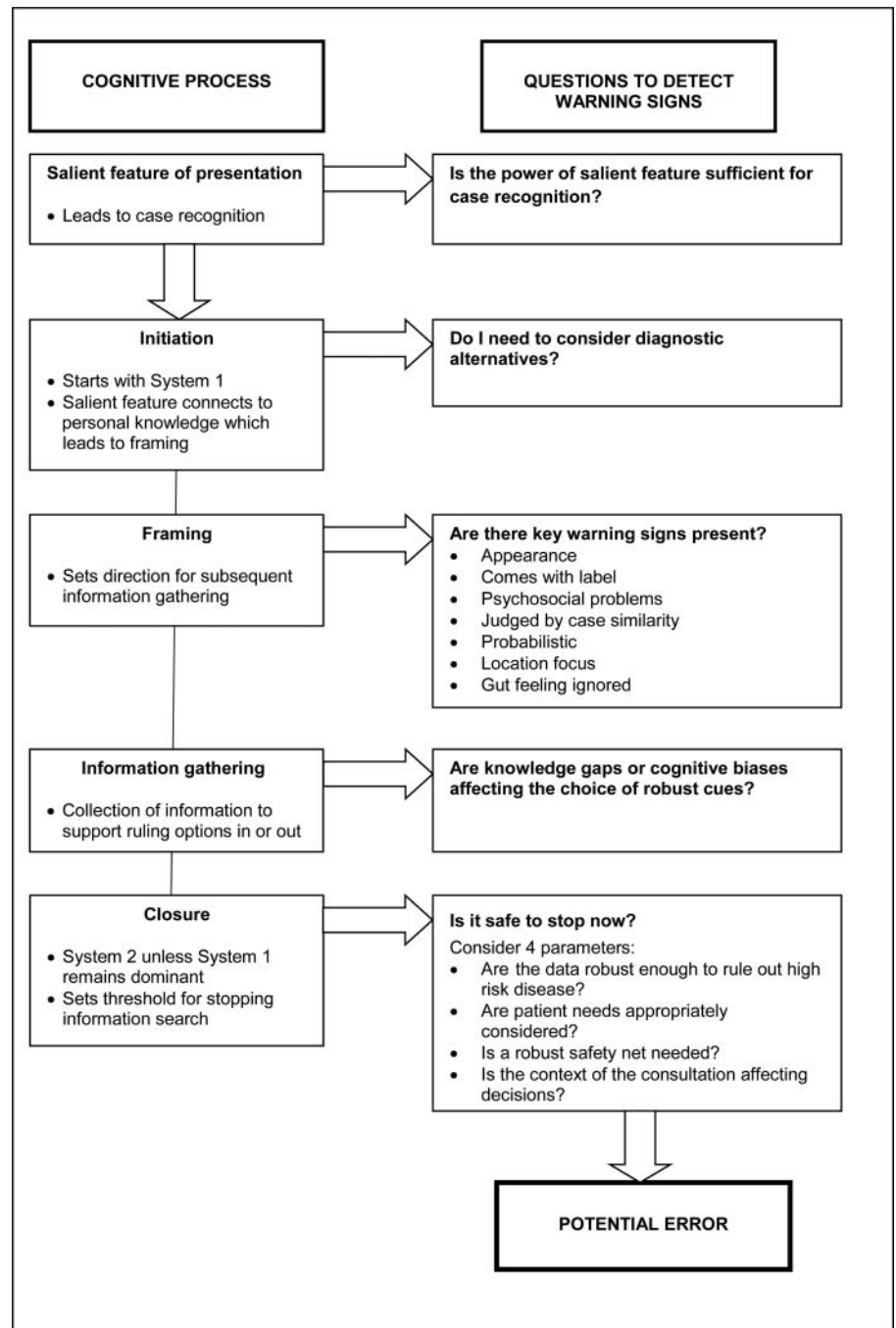
options appropriately ruled in; (2) there was a direct response to the patient's needs; and (3) there was a reliable safety net in place. As a result of the biases we describe based on our current findings, it appears that in some cases red flags and critical cues suggesting alternative diagnoses were ignored or misinterpreted. The focus was on ruling in early diagnoses, rather than the usually preferred option of ruling hazardous conditions out first. This is consistent with System 1 being dominant, ignoring the reflective System 2. We found that ignoring or failing to search for important cues may be due to knowledge gaps or biased reasoning processes. Participants' frequent reference to *not being experienced* suggests knowledge gaps contributing to a number of errors. For these reasons, and contrary to much of the literature (with the exception of Norman and Eva<sup>10</sup>), we therefore conclude that lack of knowledge is indeed likely to be an important factor in diagnostic error.

Finally, we found that participants did not use numerical values as thresholds for stopping the collection of more clinical data. We hypothesise that this relates to the complexity of these judgments: needing to satisfy decision rules based largely on subjective inputs. First, the clinician needs to take into account the basis of ruling disease in or out. Most early formulations are at least partly based on System 1 knowledge and have a tendency to bias unless one takes time for reflection

through the influence of System 2. Robust data for predictive values may not exist or not be powerful enough to lead to safe closure. Second, assessing the values and individual needs of patients in the context of an illness is subjective. Third, confidence levels in a safety net may not be reliable, yet confidence in safe closure must be closely dependent on confidence in the quality of the net.<sup>23</sup>

If we add to these issues the impact of the context or environment on the clinician's decision-making<sup>24</sup> and willingness or ability to involve System 2, it becomes apparent why Bayesian and similar 'rational' approaches are not the norm. It comes as no surprise when one of our GPs refers to closure as it *feels like a nebulous thing*.<sup>23</sup> We believe that clinicians use individualised and often tacit guides for how to deal with the problem and other 'multifarious factors that come into reckoning when making decisions'.<sup>22</sup> With this package they then make judgments, described by Polanyi<sup>16</sup> writing about personal knowledge: *in respect of choices made in the exercise of personal judgment .... there is always a range of discretion open in a choice.... In view of the unspecifiability of the particulars on which such a decision will be based, it is heavily affected by the participation of the person pouring himself into these particulars and may in fact represent a major feat of his originality, and concludes that valid choices can be made by submitting to one's own sense of responsibility*. This last statement sums

**Figure 1** Questions to relate potential sources of error to cognitive process: anatomy of diagnostic error.



up the core of professionalism, where we expect a deep sense of responsibility from the decision maker.

So, what can we offer to reduce error? In an ideal world clinicians would have robust clinical data with high true positive and negative characteristics. This could then be used to develop decision aids, as suggested by Buntinx *et al.*<sup>25</sup> However, such information is scarce and there would still be a great deal of subjective judgment to be made. Providing indicators of early warning signs where errors may occur in the diagnostic process may help to prompt a reflective review of the cognitive process. We need to look at ways of presenting

this as an integral part of the clinical process. This approach may then become part of the ongoing continual professional development for clinicians. Figure 1, based on our results is a summary of how warning signs may be incorporated in a practice environment to constructively promote reflection in practice.

### Strengths and limitations of the study

We designed the study to be consistent with desired future directions as outlined in the introduction. It is based on a strong theoretical framework provided by the DTC. The strength of this model is not only that it



'intuitively supports how our brains work',<sup>12</sup> but particularly that the initiation of the process is consistent with System 1 working at a subconscious level. Our interviews explore what is salient in a presentation<sup>26</sup> and show how initial biases influence the diagnostic process. We identify warning signs on the way, alerting clinicians to risk of error.<sup>4</sup> These may have the potential to reduce diagnostic error.<sup>11</sup>

Limitations relate to three main issues. First, sampling: participants were not chosen at random but consisted of self-selected individuals made up of more than half of the cohort we approached. This is an almost universal problem with studies that involve busy and senior clinicians. Our research needs to be replicated in other settings in the UK or in countries with different health systems, or with other specialties. However, given that our findings are consistent with other studies with different methodologies and use a strong theoretical base the results are likely to have a degree of generalisability. Second, hindsight bias: clinicians generally become aware of errors after an event and the information may be fragmentary. They construct a narrative based on hindsight, ending up with 'illusions that one has understood the past'.<sup>15</sup> In the construction of the narrative, hindsight bias converts events 'into a coherent causal framework'.<sup>17</sup> Occurring at a subconscious level, the bias is inevitable and all previous studies of error were tainted by it. We therefore attempted to minimise these limitations by examining the clinicians in real-life situations rather than a laboratory, and avoiding data that were collected for other reasons (eg, litigation),<sup>17</sup> interviews tried to reconstruct the context in which error took place. Third, we do not have evidence that alerts such as the warning signs we propose will reduce diagnostic errors in clinical practice. Similarly, deliberate practice<sup>27</sup> has not been shown to be successful in non-procedural specialties but in view of its success in procedural clinical settings, would be a strong contender to be explored.<sup>4</sup>

### Implications for practice

There are two prerequisites for change. (1) Early and systematic recognition of errors including near misses: this could be achieved through regular, non-threatening, in-practice audits or significant event analyses. Without this it will not be possible to reduce hindsight bias. (2) Provision of a clinical environment that promotes reflection in action to detect the causes of process bias and knowledge gaps. This may be more feasible when working in a group of trusted peers and using methods such as incident reviews and journal meetings focused on recent errors as a way of reviewing the diagnostic reasoning processes as well as knowledge gaps. Since reflection will need to focus both on the cognitive process and evidence-based diagnosis resources, it would be important to have meetings facilitated by a trained, preferably internal, GP.

### Implications for future research

We do not know if the practice changes that we propose will lead to better clinical care and this needs separate evaluation. To explore the generalisability of our findings we need to replicate this study in groups of clinicians in other specialties and settings. Future studies will need to deal with cases closer to the event and place even more emphasis on context. The model of deliberate practice<sup>27</sup> is likely to be suitable for ongoing professional development and training to reduce error. Some of these have led to improved clinical care, albeit in interventional specialties, which suggests that similar improvements might be feasible in primary care, but this needs evaluation.

### CONCLUSIONS

Initiation and closure of the cognitive process are most exposed to the risk of diagnostic error in primary care. Cognitive biases developed at framing directly influence errors at the end of the process. We refer to these as warning signs that can alert clinicians to the increased risk of diagnostic error. The most significant reasoning biases we observed related to patients presenting with pre-existing diagnostic labels and psychosocial problems. Others included the use of heuristics, patient's appearance, wrong initial localisation of the problem and probabilistic bias. Subsequently ignored red flags or critical cues may have been related to biased process through the frame, but could also be explained by knowledge gaps.

We conclude that lack of knowledge is likely to be an important factor in diagnostic error. Reducing diagnostic errors in primary care should focus on early and systematic recognition of errors including near misses, and a continuing professional development environment that promotes reflection in action to highlight possible causes of process bias and of knowledge gaps. Alerting clinicians to warning signs of where there is an increased risk of error may be one way to prompt a reflective review of the cognitive process. For this to become an integral part of the clinical process, we may need to experiment with deliberate practice of looking for warning signs as a potential method of professional development to reduce error.

**Acknowledgements** We wish to acknowledge a number of colleagues who helped us: Margaret Balla participated in the interview analysis and produced figure 1; Nia Roberts assisted with the literature review; Mary Hodgkinson and Dawn Fraser had the arduous task of transcribing the interviews. We particularly thank our willing participants who freely discussed the sensitive issues relating to their experiences with diagnostic error.

**Contributors** JB designed the project, carried out interviews and analysis as well as writing the first draft, was involved in all drafts and was involved in the final manuscript. CH and MT conceived the project and were involved in the design, in all drafts and were involved in the final manuscript. CG was involved in the literature search, the analysis and final manuscript. All authors contributed to the final document.

**Competing interests** None.

**Data sharing statement** There are no additional data available.



## REFERENCES

1. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
2. Schiff GD, Hasan O, Kim S, *et al*. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;169:1881–7.
3. Singh H, Thomas EJ, Khan MM, *et al*. Identifying diagnostic errors in primary care using an electronic screening algorithm. *Arch Intern Med* 2007;167:302–8.
4. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med* 2011;86:307–13.
5. Jena AB, Seabury S, Lakdawalla D, *et al*. Malpractice risk according to physician specialty. *N Engl J Med* 2011;365:629–36.
6. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care—a systematic review. *Fam Pract* 2008;25:400–13.
7. Silk N. An analysis of 1000 consecutive general practice negligence claims. *Med Protect Soc* 2000;1–18.
8. Wallis K, Dovey S. No-fault compensation for treatment injury in New Zealand: identifying threats to patient safety in primary care. *BMJ Qual Saf* 2011;20:87–91.
9. Schiff GD, Kim S, Abrams R, *et al*. Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project and methodology 2005; <http://www.ahrq.gov/downloads/pub/advances/vol2/schiff.pdf> (sighted 20 August 2011)
10. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* 2010;44:94–100.
11. Elstein AS. Thinking about diagnostic thinking: a 30-year perspective. *Adv Health Sci Educ Theory Pract* 2009;14(S1):7–18.
12. Winters BD, Aswani MS, Pronovost PJ. Commentary: reducing diagnostic errors: another role for checklists? *Acad Med* 2011;86:279–81.
13. Norman G. Dual processing and diagnostic errors. *Adv Health Sci Educ Theory Pract* 2009;14(S1):37–49.
14. Balla JI, Heneghan C, Glasziou P, *et al*. A model for reflection for good clinical practice. *J Eval Clin Pract* 2009;15:964–9.
15. Kahneman D. *Thinking fast and slow*. London: Allen Lane, 2011: 21–2.
16. Polanyi M. *The study of man*. London: The University of Chicago Press, 1959.
17. Wears RL, Nemeth CP. Replacing hindsight with insight: toward better understanding of diagnostic failures. *Ann Emerg Med* 2007;49:206–9.
18. Lucchiari C, Pravettoni G. Cognitive balanced model: a conceptual scheme of diagnostic decision making. *J Eval Clin Pract* 2012;18:82–8.
19. Gigerenzer G, Goldstein DG. Betting on one good reason: take the best heuristic. In: Gigerenzer G, Todd PM, eds. *Simple heuristics that make us smart*. ABC Research Group. New York: Oxford University Press, 1999:75–95.
20. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess RG. *Analysing Qualitative Data*. London: Routledge, 2011:173–94.
21. Stolper E, van de Wiell M, van Royen P, *et al*. Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Intern Med* 2011;26:197–203.
22. Gabbay J, LeMay A. *Practice-based evidence for healthcare*. Abingdon: Routledge, 2011.
23. Balla J, Heneghan C, Thompson M, *et al*. Clinical decision making in a high-risk primary care environment: a qualitative study in the UK. *BMJ Open* 2012;2:e000414.
24. Berner ES. Mind wandering and medical errors. *Med Educ* 2011;45:1068–9.
25. Buntinx F, Mant D, van den Bruel A, *et al*. Dealing with low-incidence serious diseases in general practice. *Br J Gen Pract* 2011;61:43–6.
26. Kahneman D. Maps of bounded rationality: a perspective on intuitive judgment and choice 2002. [http://nobelprize.org/nobel\\_prizes/economics/laureates/2002/kahnemann-lecture.pdf](http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemann-lecture.pdf) (sighted 8 May 2012).
27. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;79(Suppl 10):S70–81.

## **Appendix 1: List of interview questions**

Before commencing the interview participants were reassured about confidentiality and advised that tapes would be destroyed so that voices could not be recognised. Further questions about the process were invited. After Question 1 they were again asked if they were comfortable to proceed.

**Q 1: to gain insights into participants' conceptualisations of a diagnostic error.**  
*What do you regard as a diagnostic error?*

**Q 2: to describe themes of errors in relation to the cognitive model we used.**  
*If you are comfortable, can you tell me about some errors that you have made?*

GPs would then describe an error they made or were closely involved with. This spontaneous response generally covered a clear sequence of the salient features of the case, the context in which it occurred, their analysis of why and where they went wrong and the outcome. They often added remarks about what they learned from it and how they thought they changed their practice as a consequence, with more experience or working in a different environment.

### **Q 3: Clarifying questions**

These were used only if more details were needed and were open-ended. Occasional questions used to clarify terminology.

#### **Q 3.1. : Examples of clarifying questions about framing**

*How do you deal with this sort of undifferentiated, I know there's something but I don't know what it is?*

*So what did you sort of think was the problem? Why do you think it happened?*

#### **Q 3.2. : Examples of clarifying questions about effect of biases**

*People often talk about getting stuck on the first thing which is what you just said, can you tell me how you manage that because it must be a common issue?*

*So why do you think that happened?*

*You often use the term 'red flag' which you just used, what do you mean by that?*

#### **Q 3.3. : Examples of clarifying questions about closure thresholds**

*So this issue of confidence, how do you deal with that .. so how do you have a threshold of confidence that you say "stop now",*

*But what makes you feel, do you think, that "I'm not worried you"? What is it?*

*Can I just ask you a question about finishing the consultation? What do you hope to achieve before you're willing to do it.*

## Appendix 1

*Could you sort of speculate on the idea of a level of confidence that you have in your diagnoses?*

*One of the problems people describe is when to stop looking for things like and where do you stop?*

**COREQ guidelines table**

<b>Domain 1: Research team and reflexivity</b>			<b>Comment</b>
Personal Characteristics			
1.	Interviewer/facilitator	Which author/s conducted the interview?	JB
2.	Credentials	What were the researcher's credentials? <i>E.g. PhD, MD</i>	JB:FRCPE, FRACP, MA; CG:BM,BS; MB*: MEduc
3.	Occupation	What was their occupation at the time of the study?	Research Fellows
4.	Gender	Was the researcher male or female?	1 male/2 female for analysis
5.	Experience and training	What experience or training did the researcher have?	>30 years qualitative research JB and MB*, trainee CG
Relationship with participants			Nil
6.	Relationship established	Was a relationship established prior to study commencement?	With some of them
7.	Participant knowledge of the interviewer	What did the participants know about the researcher? <i>e.g. personal goals, reasons for doing the research</i>	Broad outlines given .
8.	Interviewer characteristics	What characteristics were reported about the interviewer/facilitator? <i>e.g. Bias, assumptions, reasons and interests in the research topic</i>	Reasons for research and interest in training
<b>Domain 2: study design</b>			
Theoretical framework			Dual theory of cognition



COREQ guidelines table

9.	Methodological orientation and Theory	What methodological orientation was stated to underpin the study? <i>e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis</i>	Content analysis
Participant selection			
10.	Sampling	How were participants selected? <i>e.g. purposive, convenience, consecutive, snowball</i>	Convenience
11.	Method of approach	How were participants approached? <i>e.g. face-to-face, telephone, mail, email</i>	Email and face-to-face
12.	Sample size	How many participants were in the study?	15
13.	Non-participation	How many people refused to participate or dropped out? Reasons?	60% of those approached not interviewed for lack of time or interest
Setting			
14.	Setting of data collection	Where was the data collected? <i>e.g. home, clinic, workplace</i>	Clinic for most, 32at home
15.	Presence of non-participants	Was anyone else present besides the participants and researchers?	No
16.	Description of sample	What are the important characteristics of the sample? <i>e.g. demographic data, date</i>	All experienced GPs in active clinical practice
Data collection			
17.	Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	Pilot tested. Semi-structured interview
18.	Repeat interviews	Were repeat interviews carried out? If yes, how many?	No

COREQ guidelines table

19.	Audio/visual recording	Did the research use audio or visual recording to collect the data?	Audiotaped
20.	Field notes	Were field notes made during and/or after the interview or focus group?	Yes
21.	Duration	What was the duration of the interviews or focus group?	30 minutes
22.	Data saturation	Was data saturation discussed?	Yes and reached at about 2/3 of way
23.	Transcripts returned	Were transcripts returned to participants for comment and/or correction?	No
<b>Domain 3: analysis and findings</b>			
Data analysis			
24.	Number of data coders	How many data coders coded the data?	3: JB, JG, MB
25.	Description of the coding tree	Did authors provide a description of the coding tree?	yes
26.	Derivation of themes	Were themes identified in advance or derived from the data?	Both, as we responded to the data
27.	Software	What software, if applicable, was used to manage the data?	NVivo
28.	Participant checking	Did participants provide feedback on the findings?	yes
Reporting			
29.	Quotations presented	Were participant quotations presented to illustrate the themes / findings? Was each quotation identified? e.g. <i>participant number</i>	yes

**COREQ guidelines table**

30.	Data and findings consistent	Was there consistency between the data presented and the findings?	yes
31.	Clarity of major themes	Were major themes clearly presented in the findings?	yes
32.	Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	yes

\* MB refers to Margaret Balla identified in Acknowledgements

**Box 1. Case 27: Illustration of initiation of the process and setting the initial diagnostic frame.**

Reconstruction	Analysis
<p><b>Presentation</b></p> <p><i>Patient in 70's came with breathlessness .. the first thing was he kept saying to me "this is exactly like it was about 6 months previously".. looked back in his notes and 6 months previously he'd been diagnosed with heart failure and so I thought "well, you know" and he was so insistent that it was the same thing... he was a bit breathless but there wasn't anything really obvious going on .. so I thought maybe he was anaemic as well and that had got worse. And it was a Saturday morning so I couldn't easily get any tests straight away so I booked for him to come back first thing on Monday morning for blood tests and ECG and I sent him up to the hospital for a chest x-ray. He was so insistent that it was the same thing and in retrospect that was really misleading for me</i></p>	<p>Salient feature was patient's insistence that the diagnosis was the same as previously, seemingly confirmed by looking at case notes of his previous presentation.</p> <p>System 2 in action as tests ordered, largely to rule in cardiac failure and rule out possible complicating factor of anaemia.</p>
<p><b>Context issues</b></p> <p><i>It was a Saturday morning so I couldn't easily get any tests straight away.</i></p>	<p>Management affected by practice environment - routine blood tests not immediately available</p>
<p><b>Outcome</b></p> <p><i>Next day contacted by one of his friends .. to say "actually, he's had a pulmonary embolism .. he'd got quite a lot worse that afternoon and been admitted to hospital and ..CT showed multiple pulmonary emboli.</i></p>	<p>Delay in diagnosis likely to System 1 overpowering System 2, raising closure threshold enough to be affected by context issues (no blood tests available at weekend).</p>
<p><b>Summary</b></p> <p>System 1 single diagnosis based on existing label, immediately jumps to the diagnosis. Weak System 2 affected by context issues, delaying diagnosis.</p>	



**Table 1. Biases arising from salient features of presentation which initiate the diagnostic process and frame the direction of subsequent information gathering.**

<b>Previous diagnosis label</b>	<p><i>Because somebody had wrote down that he had bell's palsy and he'd been seen in hospital .. I immediately thought that's what he had (1);</i></p> <p><i>Story of the insect bite and that was what we were sort of using as our diagnostic tool really (6)</i></p>
<b>Pre existing psychosocial problems</b>	<p><i>all thought some of the bleeding might be from sexual abuse (31);</i></p> <p><i>sick notes, and prescriptions and whatever and I thought that that was probably the main reason behind the um sort of um consultations (37)</i></p>
<b>Reassurance from initial appearance</b>	<p><i>when I called the patient back I got hold of the granny who said oh yes mum's in the shower that as a clue to me meant that maybe the child wasn't that ill (11);</i></p> <p><i>She wasn't terribly unwell (33)</i></p>
<b>Similarity to a recent case or similarity to representative case built from experience</b>	<p><i>My diagnosis was fed by a patient the previous week who'd presented with an ischemic foot (40);</i></p> <p><i>And I thought he had cancer because of the mass and the weight loss and the paleness (44).</i></p>
<b>Incorrect localisation of salient features</b>	<p><i>vomiting and sweating and diarrhoea .. epigastric pain (10);</i></p> <p><i>epigastric discomfort . . . must be indigestion (20).</i></p>
<b>Common things occurring commonly (Probabilistic reasoning)</b>	<p><i>viral infections are common (16);</i></p> <p><i>my preconception at the time was that a young &lt;30 year old is very, very unlikely to have bowel cancer (32)</i></p>
<b>Ignoring as well as over or under estimating red flags or critical cues</b>	<p><i>he came in hopping, which is quite unusual. Not weight bearing at all is quite unusual (30);</i></p> <p><i>normal chest on examination (24)</i></p>
<b>Vague presenting symptoms, no salient features recognised</b>	<p><i>fatigue from whatever cause (3); it was all very vague (28);</i></p> <p><i>atypical leg pain couldn't work out what was going on (21)</i></p>

--	--

**Box 2. Case 14: Illustration of dominant System 1 impeding System 2 review at closure, leading to error.**

Reconstruction	Analysis
<p><b>Presentation</b></p> <p><i>Elderly patient seen 6 years ago for what appeared to be resolving haemorrhoidal bleed .. 6 months prior [to the most recent visit] described narrow stools like a snake ..[At the present visit] bowel frequency and some bleeding with examination of clear external piles no rectal masses on PR. Did some bloods but wasn't anaemic.</i></p> <p><i>[I ignored] the older the patient the lower the threshold for colorectal cancer that we would have for referring .. red flag that's there for a reason therefore it would be foolish to sort of dismiss..</i></p>	<p>System 1 dominance may explain the high threshold for vigilance in this age group.</p> <p>No significant attempt to rule out and normal Hb wrongly used for rule in. Another example of the power of a perceived label in biasing process.</p>
<p><b>Salience</b></p> <p><i>External piles with a normal PR .. [6 years ago] with haemorrhoides seen by a colleague.</i></p>	<p>Salient feature was a normal examination 6 years earlier.</p>
<p><b>Outcome</b></p> <p><i>2 months after last visit ..change in bowel habit with rectal bleeding and as part of investigation had a sigmoidoscopy and biopsy which found a malignant colonic tumour</i></p>	<p>Delay in diagnosis likely to System1 overpowering System 2, raising closure threshold.</p>
<p><b>Summary</b></p> <p>System 1 single diagnosis based on label immediately jumps to the diagnosis. Ignored expected natural history, and the presence of a red flag. Diagnosis was delayed until new critical cue emerged.</p>	

**Table 2. Effect of framing biases on closure thresholds for ruling disease in or out.**

<b>Presents with diagnosis label</b>	<p><i>I'd keyed in too quickly and then just ignored any of the sort of differential information (1);</i></p> <p><i>When your brain immediately jumps to the obvious diagnosis its worth just having in the back of your mind what else it could be (6)</i></p>
<b>Psychosocial label/behavioural</b>	<p><i>I closed it before she came in .. I think hadn't really thought out the differential diagnosis (4);</i></p> <p><i>Not appreciating the seriousness of the, of the problem, coupled with not really wanting to think about it because the patient was so difficult. (31)</i></p>
<b>Ignores red flag</b>	<p><i>[did not] take a step back and consider what we call the sort of red flagged ones, are there any flags in front of you that are presenting information of other serious diseases that might kill or harm? (2);</i></p> <p><i>Think I ought to have thought this severe pain which isn't improving I ought to go back to cancer but so I was put off by the negative investigations and that kind of prior assessment and err .. level of pain which was not otherwise explained (15)</i></p>
<b>Ignores possibility of serious disease with low probability</b>	<p><i>[ignored] older the patient the lower the threshold for particularly for colorectal cancer that we would have for referring .. red flag that's there for a reason therefore it would be foolish to sort of dismiss (14);</i></p> <p><i>My preconception at the time was that a young 28 year old is very, very unlikely to have bowel cancer .... slightly raised C-reactive protein...it wasn't dramatically raised ... I certainly didn't act on it because I think I was confused by the fact he'd got better the second consultation (32)</i></p>
<b>Used wrong clinical features to rule-out a condition</b>	<p><i>[ignored] new onset quite severe headache in a (40) something year old is a red flag in itself (22);</i></p> <p><i>We think of ectopic pregnancy as being bleeding and pain and this was painless bleeding (17)</i></p>

<b>Ignored gut feelings</b>	<i>it's a sixth sense .. that I think as you gain more experience you really hone and fine tune .. it's invaluable particularly with children 19; was not terribly unwell ... obviously needed more investigations.. wasn't happy with my decision even though it wasn't a conscious process. (33)</i>
-----------------------------	--