Prevalence and alternative explanations influence cancer diagnosis: an experimental study with

physicians

Miroslav Sirota[1*], Olga Kostopoulou[2*], Thomas Round[3], Shyamalee Samaranayaka[4]

[1]Department of Psychology, University of Essex

[2]Department of Surgery and Cancer, Imperial College London

[3]Primary Care & Public Health Sciences, King's College London

[4]Department of Family Medicine, University of Sri Jayewardenepura, Sri Lanka

[*] These authors contributed equally to this work.

Author's Note

Correspondence concerning this paper should be addressed to either Miroslav Sirota, Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK (email: msirota@essex.ac.uk; phone: +44 (0)1206 874 229) or Olga Kostopoulou, Imperial College London, Department of Surgery and Cancer, 5th floor Medical School Building, St Mary's Campus, Norfolk Place, London W2 1PG, UK (email: o.kostopoulou@imperial.ac.uk; phone: +44 (0)2075949120).

Abstract

Objective: Cancer causes death to millions of people worldwide. Early detection of cancer in primary care may enhance patients' chances of survival. Physicians, however, often miss early cancers, which tend to present with undifferentiated symptoms. Within a theoretical framework of the hypothesis generation model (HyGene), and psychological literature, we studied how two factors – cancer prevalence and an alternative explanation for the patient's symptoms – impede early cancer detection, as well as prompt patient management. *Methods:* Three hundred family physicians diagnosed and managed two patient cases, where cancer was a possible diagnosis (one colorectal, the other lung cancer). We employed a 2 (cancer prevalence: low vs. high) $\times$ 2 (alternative explanation: present vs. absent) between-subjects design. Cancer prevalence was manipulated by changing either patient age or sex; the alternative explanation for the symptoms was manipulated by adding or removing a relevant clinical history. Each patient consulted twice. *Results:* In a series of random-intercept logistic models, both higher prevalence (*OR* 1.92, 95% CI [1.27, 2.92]) and absence of an alternative explanation (*OR* 1.70, 95% CI[1.11, 2.59]) increased the likelihood of a cancer diagnosis which, in turn, increased the likelihood of prompt referral (*OR* 22.84, 95% CI[16.14, 32.32]). *Conclusions:* These findings confirm the probabilistic nature of the diagnosis generation process and validate the application of the hypothesis generation model to early cancer detection. Increasing the salience of cancer – such as listing cancer as a diagnostic possibility – during the initial hypothesis generation phase may improve early cancer detection.

Keywords: cancer diagnosis, prevalence, alternative explanation, diagnostic reasoning, diagnostic support

Cancer is one of the most common causes of mortality in the world (Stewart & Wild, 2014). It is also a disease that often gets misdiagnosed (Kern, 1994; Singh et al., 2013). Since early detection of cancer increases the chances of successful treatment, minimizing diagnostic delays is crucial (Møller et al., 2015; Neal, 2009; Richards, 2009). Substantial variation in urgent cancer referrals in UK primary care has caused concern (Baughan, Keatings, & O'Neill, 2011; Meechan et al., 2012). Physician-related cognitive factors, such as failure to consider cancer as a possible diagnosis, might account for some of this variation (Richards, 2009). Therefore, the present paper focuses on studying some of the cognitive factors affecting cancer generation and the evaluation processes used by family physicians via an experimental vignette study.

Cognitive factors are a common cause of diagnostic errors (Graber, Franklin, & Gordon, 2005; Norman & Eva, 2010). Retrospective reviews of litigation cases and patient charts have inferred that cognitive factors are the commonest contributors to diagnostic error (Gandhi et al., 2006; Lyratzopoulos, Vedsted, & Singh, 2015). Psychological studies have indeed identified several cognitive factors that can influence the outcome of the diagnostic process: hypothesis generation (e.g., Kostopoulou, Mousoulis, & Delaney, 2009; Kostopoulou, Sirota, Round, Samaranayaka, & Delaney, 2016); information search and interpretation driven by the leading hypothesis (Kostopoulou et al., 2009; Nurek, Kostopoulou, & Hagmayer, 2014); and primacy effects (Chapman, Bergus, & Elstein, 1996; Rebitschek, Bocklisch, Scholz, Krems, & Jahn, 2015). Research therefore strongly indicates that having an appropriate hypothesis – which could be determined by the initial cues perceived, and which determines subsequent information search and interpretation – plays a crucial role in the diagnostic process.

Accounts from the clinical reasoning literature discuss hypothesis generation as matching of the presenting problem to a similar problem previously encountered (Norman & Brooks,

1997), abstraction from specific patient features to higher-order memory structures (Grant &

Marsden, 1987), and mapping specific patient cases to disease prototypes (Charlin, Tardif, &

Boshuizen, 2000). For a review, see Kostopoulou (2009). Accounts from the psychological

literature discuss the hypothesis generation as an employment of preconscious heuristic

processes, which form the most plausible epistemic mental models of the world (Evans, 2006,

2008; Evans, Venn, & Feeney, 2002). These accounts differ in the knowledge or memory

structures that they posit but are similar in that they underspecify the hypothesis generation

process.

In contrast to previous theoretical accounts, the Hypothesis Generation (HyGene) model

is a computational process model of hypothesis generation that attempts to specify this process

more precisely (e.g., Thomas, Dougherty, & Buttaccio, 2014; Thomas, Dougherty, Sprenger, &

Harbison, 2008). The model assumes three basic cognitive processes: 1/ retrieval from memory,

2/ maintenance in working memory, and 3/ judgment and decision-making. Specifically, the

patients' presented symptoms activate traces in episodic memory, representing similar past cases.

An "unspecified probe" extracts the hypotheses (diagnoses) that are most frequently and strongly

associated with the symptoms, and matches them against known hypotheses (e.g., prototypes of

diseases stored in semantic memory). Hence, the better the match between cancer prototype and

the patients' symptoms, the more likely it is that the cancer hypothesis will reach the minimum

activation threshold to enter working memory. This is an iterative process, so once an additional

hypothesis is sufficiently activated by the probe (i.e., achieves the dynamically updated

minimum activation threshold), it too enters working memory. However, due to memory

constraints, physicians would only keep a limited number of hypotheses in their working

memory; these would then guide subsequent information search and hypothesis testing. A

process comparing the relative strengths of the competing hypotheses in working memory determines the posterior probability of the hypotheses. A hypothesis-guided search starts and proceeds if more than one hypothesis is tested. The model assumes a consistency check of the hypothesis with the input data: a hypothesis is rejected if it is not consistent with the symptoms.

Several factors may affect the generation of cancer hypotheses (e.g., time constraints and working memory constraints) but two factors are especially likely to determine the degree of a match between patient symptoms and the cancer hypothesis, and therefore its strength of activation and maintenance in working memory: the a priori probability of cancer and the presence of another explanation (a more common disease that also matches the patients' symptoms). First, physicians perceiving a low probability of cancer in patients presenting with vague symptoms may fail to generate the cancer hypothesis or exclude it as an implausible explanation for the patients' symptoms in the hypothesis evaluation phase. Indeed, people tend to first generate the hypotheses with a high a priori probability (Dougherty & Hunter, 2003a; Weber, Böckenholt, Hilton, & Wallace, 1993). Physicians may also fail to diagnose cancer if there is an alternative, more common, explanation for the symptoms in an early stage of the diagnostic process. This might impair either the hypothesis generation process – the cancer diagnosis does not reach the dynamically updated activation threshold (Gettys & Fisher, 1979) – and/or the hypothesis evaluation process – the relative strength (and hence the posterior probability) of the cancer hypothesis is weakened due to a stronger plausible alternative (Dougherty & Hunter, 2003a; Windschitl & Wells, 1998).

Present research

In the experiment presented here, we tested the effect of prevalence and an alternative explanation on the diagnosis and subsequent management of early cancers. We constructed two patient cases that depicted common cancers: one colorectal and the other lung. Colorectal and lung cancers are some of the most common cancers (Maddams, Utley, & Møller, 2012), and are often missed in their early stages of presentation (Rubin, McPhail, & Elliott, 2011). We avoided well-known methodological pitfalls. First, we manipulated the two factors using a factorial design, while keeping confounding variables, such as disease severity, constant. Second, we designed the clinical vignettes and manipulated the study factors in an ecologically valid way: i) we manipulated patient characteristics to alter disease prevalence rather than provide descriptive base rates, and ii) we either provided or withheld a medical history that could offer a plausible, alternative explanation for the patients' symptoms. Finally, to overcome possible pitfalls of novice-experts differences (e.g., Shanteau, 1992), we recruited practicing physicians, rather than medical students, as study participants. We employed an experimental vignette study because it offers high internal and external validity of clinical judgments (Evans et al., 2015; Rose et al., 2015).

We tested three hypotheses. First, we hypothesized that a set of symptoms in patients with a higher prevalence of cancer will lead physicians to diagnose cancer more often than with a lower prevalence (Hypothesis 1). Second, we hypothesized that a set of symptoms in patients with a prior, alternative (i.e., non-cancerous) explanation of their symptoms will lead physicians to diagnose cancer less often than when an alternative explanation is not available (Hypothesis 2). Finally, as shown in previous research (e.g., Kostopoulou et al., 2008), we expected that diagnostic decisions would affect management decisions (Hypothesis 3).

Method

Participants

We determined our sample size in two steps. First, we defined the sample size stopping rule based on the a priori power analysis using G*Power (Cohen, 1988; Faul, Erdfelder, Lang, & Buchner, 2007). We required a minimum sample size of 237 participants for a logistic regression model with a binary predictor and adjusted for a covariate accounting for an additional 20% – to detect an odds ratio of 2.3, with 80% power at a 5% significance level. Such an effect size was assumed in other studies that used hypothetical vignettes to study the effect of patient demographics on diagnostic accuracy (Eva, Link, Lutfey, & McKinlay, 2010). Second, we increased the computed sample size by around 25% (i.e., $n = 297$) to account for the number of participants who would refer at least one patient on the first visit.

We invited, by e-mail, fully certified and currently practicing family physicians, who had participated in previous studies conducted by the lab. We also received help with recruitment from local clinical research networks around the UK, who contacted primary care clinics on our behalf. Potential participants were told that they would be taking part in a study of clinical reasoning. Cancer was not mentioned in the study information. After completing the study, each respondent was asked to provide an e-mail address in order to receive a £20 electronic Amazon voucher. To ensure that participants were actually physicians, and that each participant completed the study only once, we asked them to provide their professional NHS (National Health Service) e-mail address. Participants' e-mail addresses were not linked to their data.

We recorded 382 attempts to complete the questionnaire. Only participants who completed all the vignettes were entered into the analysis. We excluded 82 incomplete attempts, most of which only explore the first presented vignette. As a result, 300 family physicians

completed the questionnaire. Half of the sample (50.7%) were male and 36% of participants were currently practicing in an inner city, while 45.7% practiced in an urban area and 18.3% in a rural area. Experience in family medicine post-certification ranged from 0 to 40 years ($M = 10.7$, $SD = 9.2$ years, $Mdn = 7$).

Design and Materials

We employed a 2 (cancer prevalence: low vs. high) × 2 (alternative explanation: present vs. absent) between-subjects experimental design for both patient cases (colorectal cancer and lung cancer), respectively. Each patient consulted twice. We therefore constructed two vignettes per case; one per consultation. We constructed the vignettes with the help of two experienced family physicians (the last two co-authors), who systematically reviewed the clinical literature on predictive symptoms and signs of colorectal and lung cancers. Participants saw both vignettes per case, unless they decided to refer the patient at the first consultation. In the colorectal cancer case, we manipulated cancer prevalence by changing the patient's age (40 vs. 70 years for low vs. high prevalence respectively) to reflect increasing incidence with age (e.g., the increased incidence for UK males aged over 50, Cancer Research UK, 2012a). In the lung cancer case, we manipulated cancer prevalence by changing the patient's sex to reflect the past trend of higher prevalence of lung cancer in men vs. women (e.g., age-standardised incidence rates in the UK are higher for males compared with females: 95.7 per 100,000 males and 66.5 per 100,000 females, Cancer Research UK, 2012b). The alternative explanation in the colorectal cancer case was hemorrhoids; in the lung cancer case it was infective exacerbation of pre-existing COPD (chronic obstructive pulmonary disease). The alternative explanation was either present or absent. Thus, we created four vignette versions per consultation and patient case. The complete set of vignettes and their versions is presented in the Appendix.

Second consultation vignettes always followed first consultation vignettes. At the first consultation, participants were randomized to either the colorectal or the lung cancer case. They saw and responded to the first consultations of both cancer cases. At the second consultation, they were randomized again to either the colorectal or the lung cancer case. Between these second consultations they saw a single-consultation vignette, which was not about cancer (a child presenting with an ear infection). This served as a decoy, to minimize any suggestion that the study was about cancer, and was also used to test a different hypothesis, not related to this study. No data from the decoy case are presented here.

Each case included basic information about the patient (name, sex, age, smoking status, last recorded blood pressure, relevant past medical history, current medications, and the reason for the last consultation), the current health complaint, current symptoms, and the results of relevant physical examinations and investigations. At the second consultation, the patients were still suffering from the same or slightly worsened symptoms, and had some new symptoms; specifically, the colorectal cancer patient presented with two "alarm" symptoms (blood in stool and slightly reduced hemoglobin indicating possible anemia and therefore blood loss), while the lung cancer patient presented with additional vague symptoms (fatigue, loss of appetite), which are not cancer-specific.

Procedure

Once participants had accessed the study site and read some information about the study, they were asked to provide socio-demographic details: gender, number of years in family medicine and location of their practice (inner-city, urban, or rural). They then read a vignette about the first consultation of the first patient (either colorectal or lung cancer). At the end of the vignette, they confirmed that they had read it, before proceeding to the diagnosis and

management page. First, they were asked to enter their main diagnosis (*"What is your working diagnosis? (Please enter only one)"*). They could also enter any other diagnoses that they were considering (see Supplementary materials). They were then asked to select one or more of the following list of management options: "arrange follow-up", "order investigations", "prescribe", and "refer to secondary care". Some of these options could be specified further. For example, the option "refer to secondary care" required a reason for referral, the type of specialist and the type of referral: emergency (within 24 hours), urgent (within two weeks – as recommended for suspected cancer) or routine (typically four to eight weeks). The management options fully respected the UK healthcare system to make the diagnostic and management task as realistic as possible. The guidelines for urgent referral for suspected cancer are laid down by NICE (NICE, 2015). At the end of the study, participants entered their e-mail address in order to receive the Amazon voucher.

We created two diagnostic variables: i) cancer as the main diagnosis: whether cancer was the main diagnosis or not and ii) cancer as the diagnosis: whether cancer was recorded at all, either as the main diagnosis or as one of the other diagnoses considered (the "differential"). The first author coded whether participants provided a cancer diagnosis or not for both diagnostic variables on each occasion. One of the physician co-authors coded whether the management decisions were likely to lead to early, delayed or nil detection of cancer (e.g., a decision not to refer was coded as not leading to cancer detection; a decision to refer to the wrong specialty was coded as leading to delayed cancer detection). The other physician co-author coded the management decisions on a subset of responses ($n = 657$, 57.6%). Inter-rater agreement was very high: 98% agreement (Cohen's $\kappa = 0.96$, SE = 0.03) and the few disagreements were resolved in the discussion. We recoded the management decisions in two ways, one as early cancer

management (early cancer referral vs. late/no cancer referral) and one as cancer management (any cancer referral vs. no cancer referral).

We conducted the study in accordance with the ethical standards of the American Psychological Association, and obtained ethical approval from the Proportionate Review Sub-Committee of West London (REC 2).

Results

In total, the physicians considered cancer 70.2% of the time (800 out of 1,140 responses), while cancer was their main working diagnosis only 23.3% of the time (266 out of 1,140 responses). They did not refer for cancer at all 65.7% of the time, referred for cancer routinely 10.6% of the time, and urgently 23.7% of the time. Thus, although most responses included cancer as a diagnostic possibility, they did not always include a referral decision. The physicians recorded cancer as their main diagnosis more frequently in the colorectal cancer case than in the lung cancer case, and at the second rather than the first consultation (Figure 1, panel A). A similar pattern was seen in the urgent cancer referrals (Figure 2, panel A). In relation to our manipulated factors, the physicians recorded cancer as their main diagnosis more often in the vignettes where cancer prevalence was high (vs. low) and in the vignettes where there was no prior alternative explanation for the symptoms (vs. an alternative explanation) (Figure 1, panel B). Again, diagnostic rates were reflected in the urgent referral rates (Figure 2, panel B).

Insert Figure 1 around here

We measured the influence of cancer prevalence (Hypothesis 1) and an alternative explanation (Hypothesis 2) on diagnosis, using generalized estimating equations (GEE) – logistic

regression with random intercept, and vignette sequence as a repeated measure. In a model without covariates (Table 1, left column), higher cancer prevalence and absence of an alternative explanation almost doubled the frequency of cancer as the main diagnosis. There were no significant interactions between the two factors. In a model adjusted for covariates (Table 1, left column) – consultation (first and second) and physician experience (years in family medicine) – produced similar results: higher prevalence and absence of an alternative explanation increased cancer diagnosis, without a significant interaction. The physicians diagnosed cancer more than four times more often at the second than at the first consultation, and more experienced family physicians diagnosed cancer slightly more often than less experienced physicians. When we repeated these analyses for the more inclusive measure of diagnosis (cancer given as either the main diagnosis or in the differential, Table 1, right column), only prevalence remained significant. In terms of the covariates (Table 1, right column), consultation decreased and physician experience increased the odds of diagnosing cancer. Therefore, when consultation and physician experience were accounted for in the model, prevalence increased the odds of recording cancer, either as the main diagnosis or in the differential, whereas the alternative explanation increased the odds of cancer recorded as the main diagnosis only. Such evidence provides support for the two hypotheses derived from the hypothesis generation model about the effect of prevalence (Hypothesis 1) and an alternative explanation (Hypothesis 2) on cancer diagnosis.

Insert Table 1 around here

We had expected that, if cancer was the participants' main diagnosis, this would lead to

referral (Hypothesis 3). Indeed, a series of GEE logistic regression models with random intercept whether unadjusted or adjusted for consultation and physician experience, confirmed this expectation (Table 2). Recording cancer as the main diagnosis increased the odds of any cancer referral (vs. no referral) around 13 times in both unadjusted and adjusted models, and the odds of early cancer referral (vs. delayed or no referral) almost 23 times in the unadjusted and 20 in the adjusted model. A second consultation also increased the odds of both referral (vs. no referral) and urgent referral (vs. delayed or no referral), on average, by a factor of 5. Physician experience was not related to referral decisions (Table 2). Thus, Hypothesis 3 was also confirmed: having cancer as the main diagnosis led to a higher rate of referrals.

Insert Table 2 around here

Discussion

The goal of this study was to investigate two possible determinants of the prompt detection and management of early cancers by family physicians: disease prevalence and the availability of a prior, alternative explanation. Specifically, higher prevalence and no alternative explanation for the symptoms increased the frequency of cancer diagnoses and referrals to secondary care. In addition, we found that physicians gave cancer as their main working diagnosis and referred the patient urgently (i.e., within two weeks) much more often with repeated presentations that suggested a worsening of the condition, and more often for the colorectal cancer case, which included "alarm" symptoms, than for the lung cancer case, which included only vague symptoms.

Physicians' sensitivity to prevalence is consistent with the hypothesis generation model

and the literature in the field (Dougherty & Hunter, 2003a; Thomas et al., 2008; Weber et al.,

1993). It appears to contrast with the literature on base-rate neglect. According to this, people,

including physicians, neglect prevalence information (e.g., Elstein & Schwartz, 2000; Hamm,

1996; Hoffrage & Gigerenzer, 1998; Kahneman & Tversky, 1973; McNair & Feeney, 2015;

Sirota, Juanchich, & Hagmayer, 2014; Sirota, Kostovičová, & Vallée-Tourangeau, 2015). For

instance, only 5 out of 48 physicians integrated the disease base-rate information with the results

of cancer screening, when asked about its positive predictive value (Hoffrage & Gigerenzer,

1998). However, we wish to argue that this is only a superficial contradiction, because of

different contexts (screening vs. diagnosing), requiring different knowledge and engaging

different cognitive processes. Neglecting the base-rate information seems to be limited to

specific cognitive paradigms (e.g., textbook tasks) that require people to compute the

probabilities explicitly and provide the statistical information from description, not from

experience (Hertwig, Barron, Weber, & Erev, 2004; Sirota, Vallée-Tourangeau, Vallée-

Tourangeau, & Juanchich, 2015). In fact, some authors would argue that people, including

physicians, are sensitive to the base-rate information and integrate it successfully with other

pieces of information once they have experienced them (Christensenszalanski & Beach, 1982;

Gill, Sabin, & Schmid, 2005). In our study, we provided qualitative evidence of the updating that

is based on experience.

The effect of an alternative explanation fits well with the hypothesis generation model

(Thomas et al., 2014; Thomas et al., 2008). The immediate availability of an alternative

explanation creates a stronger link between the relevant diagnostic hypothesis and the observed

symptoms and enables it to reach the activation threshold for entering working memory. This

would also fit well with the general cognitive tendency of people to focus too narrowly on the

hypothesis currently under consideration, while ignoring other relevant hypotheses. People tend to neglect relevant alternative explanations (Fernbach, Darlow, & Sloman, 2010), favor information supporting the focal hypothesis (Doherty, Chadwick, Garavan, Barr, & Mynatt, 1996), direct a searching strategy that confirms their initial hypothesis (Klayman & Ha, 1987) and construct one mental model of a diagnostic problem at a time and then focus on it (see singularity principle, Evans, 2006). Future research should test further the specific psychological mechanisms of hypothesis activation and maintenance in medical diagnosis, using more real-world study designs, e.g. where physicians are allowed to gather information as they see fit, rather than being provided with information all at once or in a predetermined sequence. Physician experience increased the frequency of cancer provided as the main working diagnosis. This is, however, in direct contrast with the findings of a recent study (Kostopoulou et al., 2016), where family physicians diagnosed richer and more realistic cancer scenarios than the vignettes used in this study: more experienced physicians were significantly less likely to consider cancer as a possibility at the start of the clinical consultation, and to give it later as their main, working diagnosis. In this and the current study, nevertheless, the relationship between experience and diagnosis was weak, with odds ratios close to 1. These different findings exemplify the lack of consensus in the literature regarding the relationship between physician experience and diagnostic accuracy.

Our findings have theoretical implications. First, they provide some validation of the hypothesis generation model using a realistic task: physicians diagnosed and managed the hypothetical patients consistently with the probabilistic nature of this process, sensitive to changes in prevalence and to a more strongly-activated, competing diagnosis. These findings are also consistent with other theoretical models, such as the Extended Analytical-Heuristic

Processes Model (Evans, 2006). Prevalence of cancer likely plays some role in constructing the most plausible mental model of the symptoms; if the mental model of an alternative diagnosis is generated and meets the satisfaction principle threshold, then it is likely to be accepted; this, in turn, may eliminate the less likely mental model of cancer. However, these post-hoc explanations of our findings cannot corroborate Evans' model, as it does not postulate – in contrast with the HyGene model – these specific predictions a priori.

Furthermore, we did not study an exhaustive list of factors postulated by the HyGene model that could affect the hypotheses generation phase and, in turn, cancer detection. Other cognitive factors, e.g., working memory capacity and the number of diagnostic alternatives considered, as well as situational factors that directly impinge on cognition, such as time constraints and order of presented data are of equal theoretical importance for model validation on practicing physicians, and should be included in future studies (e.g., Doherty et al., 1996; Dougherty, Gettys, & Thomas, 1997; Dougherty & Hunter, 2003a, 2003b; Thomas et al., 2014; Thomas et al., 2008). The two factors that we chose to test in this study are theoretically important for the model validation and could be adopted in future interventions to improve cancer detection, as suggested below.

Finally, some physicians generated cancer even when its prevalence was low and an alternative, more common explanation was available. This may indicate concern with less common but more serious disease. Although we did not measure physicians' perceived probability of cancer, there is some evidence that the severity of the diagnostic outcome can influence hypothesis generation and inflate posterior probability judgments (Harris & Corner, 2011; Juanchich, Sirota, & Butler, 2012; Weber et al., 1993). This could be considered within the HyGene model – perhaps the traces encoding rare but costly diagnoses in episodic memory are

activated more easily and with minimal cue input.

Our findings have practical implications for medical education and training and the design of diagnostic support. In terms of education and training, if retrieval from episodic memory is crucial for early cancer diagnosis, then simulated experience rather than simple provision of statistical information may lead to better probability calibration (Hertwig et al., 2004). Appropriate calibration of probability may also be useful in avoiding over-diagnosis and over-referral for cancer. In terms of diagnostic support for hypotheses generation, offering a list of differential diagnoses at the beginning of the consultation, and no other help, was found to increase the final diagnostic accuracy of family physicians in two European countries (Kostopoulou, Lionis, et al., 2015; Kostopoulou, Rosen, et al., 2015). Specifically, family physicians in the UK and Greece diagnosed a series of simulated patients online. One group received early diagnostic support in the form of a list of diagnostic suggestions at the start of the consultation, after reading some limited information (a short patient description and the presenting problem) and before collecting further information to test their hypotheses. In both countries, the group who received early diagnostic support was significantly more accurate that the unaided control group (pooled *OR* 1.40, 95% CI [1.13-1.67]). Kostopoulou and colleagues subsequently designed a prototype for a Decision Support System, based on the principle of early diagnostic support, which they evaluated in a high-fidelity simulation (physicians consulting with actors-as-patients) and found significant improvement in diagnostic accuracy, without an associated increase in the number of investigations (Kostopoulou, Porat, Corrigan, S., & Delaney, in press). These studies provide evidence that early suggestions of diagnostic alternatives, before the physician engages in information gathering and hypothesis testing, can lead to improvements in diagnostic accuracy.

We acknowledge the tension between missing serious diseases, such as cancers, and over-diagnosis, leading to unnecessary investigations and sometimes treatment. Until there is more and better evidence to inform appropriate risk assessment, and until economic analyses are carried out to determine the most appropriate decision thresholds for referrals and investigations, such tensions will persist (Rubin et al., 2015). Although risk cancer tools are currently available, such as QCancer (http://www.qcancer.org) and The Risk Assessment Tool, which can calculate the patient's cancer risk, employing them requires that cancer has already been considered as a possibility. In fact, the argument for early support, i.e., support to influence a decision being made rather than change a decision already made, has been articulated in relation to other medical decision tasks, such as prescribing (Hayward et al., 2013).

Our study has several methodological limitations: two of which deserve more attention. First, one could question the external validity of the vignettes used in this study. We believe that the vignettes represented the diagnostic and management decisions well, since our approach to the patient vignettes' creation corresponded to the general recommendations for designing externally valid clinical judgment vignettes (e.g., Evans et al., 2015). For instance, the vignettes were constructed with the help of family physicians, after reviewing the relevant clinical literature, and included repeated patient presentations, as is common in patients with cancer. Furthermore, a recent multi-country study found that physicians' willingness to investigate possible cancer symptoms in vignettes was associated with national survival rates (Rose et al., 2015). Nevertheless, future research should corroborate our findings in more realistic conditions, as suggested earlier. Second, one could question to what extent our findings could be generalized to other cancers. Indeed, we simulated only two common and well-known cancers. We would expect qualitatively similar processes to be at play in other cancers and – in light of current

findings – an amplified effect of missing early presentations of less common and less-known cancers.

In summary, family physicians diagnosed cancer more frequently when the prior probability of cancer in patients presenting with vague symptoms was higher, when they were not offered an alternative explanation of the symptoms, and when the patient presented repeatedly with worsening symptoms. Physicians' working diagnoses guided their referral decisions. Our findings corroborate some of the assumptions of the hypothesis-generation model with practicing physicians, and, along with the other studies discussed, emphasize the importance of the initial diagnostic phase, when competing working diagnoses are generated. Future interventions aiming to improve cancer detection could target this phase, by offering possible differential diagnoses with severe consequences and/or with underestimated prior probabilities.

References

Baughan, P., Keatings, J., & O'Neill, B. (2011). Urgent suspected cancer referrals from general practice: audit of compliance with guidelines and referral outcomes. *The British Journal of General Practice, 61*(592), e700-e706. doi: 10.3399/bjgp11X606591

Cancer Research UK. (2012a). Bowel cancer incidence statistics.   Retrieved 21/7/2015, from http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence#heading-One

Cancer Research UK. (2012b). Lung cancer incidence statistics.   Retrieved 21/07/2015, from http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence#heading-Zero

Chapman, G. B., Bergus, G. R., & Elstein, A. S. (1996). Order of information affects clinical judgment. *Journal of Behavioral Decision Making, 9*(3), 201-211. doi: 10.1002/(SICI)1099-0771(199609)9:3<201::AID-BDM229>3.0.CO;2-J

Charlin, B., Tardif, J., & Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Academic Medicine, 75*(2), 182-190. doi: 10.1097/00001888-200002000-00020

Christensenszalanski, J. J. J., & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance, 29*(2), 270-278. doi: 10.1016/0030-5073(82)90260-4

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed. ed.). Hillsdale, NJ: Lawrence Erlbaum.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's

understanding of the diagnostic implications of probabilistic data. *Memory & Cognition,*

24(5), 644-654.

Dougherty, M. R. P., Gettys, C. F., & Thomas, R. (1997). The role of mental simulation in

judgments of likelihood. *Organizational Behavior and Human Decision Processes, 70*(2),

135-148.

Dougherty, M. R. P., & Hunter, J. E. (2003a). Hypothesis generation, probability judgment, and

individual differences in working memory capacity. *Acta Psychologica, 113*(3), 263-282.

doi: http://dx.doi.org/10.1016/S0001-6918(03)00033-7

Dougherty, M. R. P., & Hunter, J. E. (2003b). Probability judgment and subadditivity: The role

of working memory capacity and constraining retrieval. *Memory & Cognition, 31*(6),

968-982. doi: 10.3758/bf03196449

Elstein, A. S., & Schwartz, A. (2000). Clinical reasoning in medicine. In J. J. Higgs, Mark (Ed.),

Clinical Reasoning in the Health Professions (2nd ed., pp. 95-106). Oxford: Butterworth

Heinemann.

Eva, K. W., Link, C. L., Lutfey, K. E., & McKinlay, J. B. (2010). Swapping horses midstream:

Factors related to physicians' changing their minds about a diagnosis. *Academic*

Medicine, 85(7), 1112-1117. doi: 10.1097/ACM.0b013e3181e16103

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation.

Psychonomic Bulletin & Review, 13(3), 378-395. doi: 10.3758/BF03193858

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition.

Annual Review of Psychology, 59, 255-278. doi:

10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., Venn, S., & Feeney, A. (2002). Implicit and explicit processes in a hypothesis testing task. *British Journal of Psychology, 93*, 31-46. doi: 10.1348/000712602162436

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., . . . Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology, 15*(2), 160-170. doi: 10.1016/j.ijchp.2014.12.001

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191. doi: 10.3758/BF03193146

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science, 21*(3), 229-336. doi: 10.1177/0956797610361430

Gandhi, T. K., Kachalia, A., Thomas, E. J., Puopolo, A. L., Yoon, C., Brennan, T. A., & Studdert, D. M. (2006). Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. *Annals Of Internal Medicine, 145*(7), 488-496. doi: 10.7326/0003-4819-145-7-200610030-00006

Gettys, C. F., & Fisher, S. D. (1979). Hypothesis plausibility and hypothesis generation. *Organizational Behavior and Human Performance, 24*(1), 93-110. doi: http://dx.doi.org/10.1016/0030-5073(79)90018-7

Gill, C. J., Sabin, L., & Schmid, C. H. (2005). Why clinicians are natural bayesians. *British Medical Journal, 330*(7499), 1080-1083. doi: 10.1136/bmj.330.7499.1080

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165*(13), 1493-1499. doi: 10.1001/archinte.165.13.1493

Grant, J., & Marsden, P. (1987). The structure of memorized knowledge in students and

      clinicians: an explanation for diagnostic expertise. *Medical Education, 21*(2), 92-98.

Hamm, R. M. (1996). Physicians neglect base rates, and it matters. *Behavioral and Brain*

      Sciences, 19(1), 25-&. doi: 10.1017/S0140525X00041261

Harris, A. J. L., & Corner, A. (2011). Communicating environmental risks: Clarifying the

      severity effect in interpretations of verbal probability expressions. *Journal of*

      Experimental Psychology-Learning Memory and Cognition, 37(6), 1571-1578. doi:

      10.1037/a0024195

Hayward, J., Thomson, F., Milne, H., Buckingham, S., Sheikh, A., Fernando, B., . . . Pinnock, H.

      (2013). 'Too much, too late': mixed methods multi-channel video recording study of

      computerized decision support systems and GP prescribing. *Journal of the American*

      Medical Informatics Association, 20(e1), e76-84. doi: 10.1136/amiajnl-2012-001484

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the

      effect of rare events in risky choice. *Psychological Science, 15*(8), 534-539. doi:

      10.1111/j.0956-7976.2004.00715.x

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic

      inferences. *Academic Medicine, 73*(5), 538-540.

Juanchich, M., Sirota, M., & Butler, C. L. (2012). The perceived functions of linguistic risk

      quantifiers and their effect on risk, negativity perception and decision making.

      Organizational Behavior and Human Decision Processes, 118(1), 72-81.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review,*

      80(4), 237.

Kern, K. A. (1994). Medicolegal analysis of the delayed diagnosis of cancer in 338 cases in the

united states. *Archives of Surgery, 129*(4), 397-404. doi:

10.1001/archsurg.1994.01420280073009

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis-

testing. *Psychological Review, 94*(2), 211-228.

Kostopoulou, O. (2009). Diagnostic errors: Psychological theories and research implications. *In*

B. Hurwitz and A. Sheikh (Eds) *Health Care Errors and Patient Safety* (pp. 95-111):

Wiley-Blackwell. Chapter 8, p. 97-111.

Kostopoulou, O., Lionis, C., Angelaki, A., Ayis, S., Durbaba, S., & Delaney, B. C. (2015). Early

diagnostic suggestions improve accuracy of family physicians: A randomized controlled

trial in Greece. *Family Practice, 32*(3), 323-328. doi: 10.1093/fampra/cmv012

Kostopoulou, O., Mousoulis, C., & Delaney, B. (2009). Information search and information

distortion in the diagnosis of an ambiguous presentation. *Judgment and Decision Making,

4*(5), 408-418.

Kostopoulou, O., Oudhoff, J., Nath, R., Delaney, B. C., Munro, C. W., Harries, C., & Holder, R.

(2008). Predictors of diagnostic accuracy and safe management in difficult diagnostic

problems in family medicine. *Medical Decision Making, 28*(5), 668-680. doi:

10.1177/0272989x08319958

Kostopoulou, O., Porat, T., Corrigan, D., S., M., & Delaney, B. (in press). Supporting first

impressions reducses diagnostic error: evidence from a high-fidelity simulation. *British

Journal of General Practice*.

Kostopoulou, O., Rosen, A., Round, T., Wright, E., Douiri, A., & Delaney, B. (2015). Early

diagnostic suggestions improve accuracy of GPs: A randomised controlled trial using

computer-simulated patients. *British Journal of General Practice, 65*(630), e49-54. doi: 10.3399/bjgp15X683161

Kostopoulou, O., Sirota, M., Round, T., Samaranayaka, S., & Delaney, B. C. (2016). The role of physicians' first impressions in the diagnosis of possible cancers without alarm symptoms. *Medical Decision Making*. doi: 10.1177/0272989X16644563. [Epub ahead of print] http://mdm.sagepub.com/content/early/2016/04/25/0272989X16644563.full.pdf

Lyratzopoulos, G., Vedsted, P., & Singh, H. (2015). Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *British Journal of Cancer, 112*(s1), S84-S91. doi: 10.1038/bjc.2015.47

Maddams, J., Utley, M., & Møller, H. (2012). Projections of cancer prevalence in the United Kingdom, 2010–2040. *British Journal of Cancer, 107*(7), 1195-1202. doi: 10.1038/bjc.2012.366

McNair, S., & Feeney, A. (2015). Whose statistical reasoning is facilitated by a causal structure intervention? *Psychonomic Bulletin & Review, 22*(1), 258-264. doi: 10.3758/s13423-014-0645-y

Meechan, D., Gildea, C., Hollingworth, L., Richards, M. A., Riley, D., & Rubin, G. (2012). Variation in use of the 2-week referral pathway for suspected cancer: A cross-sectional analysis. *British Journal of General Practice, 62*(602), e590-e597. doi: 10.3399/bjgp12X654551

Møller, H., Gildea, C., Meechan, D., Rubin, G., Round, T., & Vedsted, P. (2015). Use of the English urgent referral pathway for suspected cancer and mortality in patients with cancer: cohort study. *BMJ, 351*. doi: 10.1136/bmj.h5102

Neal, R. D. (2009). Do diagnostic delays in cancer matter? *British Journal of Cancer, 101*(Suppl 2), S9-S12. doi: 10.1038/sj.bjc.6605384

NICE. (2015). *Suspected cancer: recognition and referral 2015*   Retrieved from https://www.nice.org.uk/guidance/ng12

Norman, G. R., & Brooks, L. R. (1997). The non-analytical basis of clinical reasoning. *Advances in Health Sciences Education, 2*(2), 173-184. doi: 10.1023/a:1009784330364

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical Education, 44*(1), 94-100.

Nurek, M., Kostopoulou, O., & Hagmayer, Y. (2014). Predecisional information distortion in physicians' diagnostic judgments: Strengthening a leading hypothesis or weakening its competitor? *Judgment and Decision Making, 9*(6), 572.

Rebitschek, F. G., Bocklisch, F., Scholz, A., Krems, J. F., & Jahn, G. (2015). Biased processing of ambiguous symptoms favors the initially leading hypothesis in sequential diagnostic reasoning. *Experimental psychology, 62*, 287-305.

Richards, M. A. (2009). The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *British Journal of Cancer, 101*(Suppl 2), S1-S4. doi: 10.1038/sj.bjc.6605382

Rose, P. W., Rubin, G., Perera-Salazar, R., Almberg, S. S., Barisic, A., Dawes, M., . . . Group*, T. I. M. W. (2015). Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. *BMJ Open, 5*(5). doi: 10.1136/bmjopen-2014-007212

Rubin, G., Berendsen, A., Crawford, S. M., Dommett, R., Earle, C., Emery, J., . . . Zimmermann,

C. (2015). The expanding role of primary care in cancer control. *The Lancet Oncology, 16*(12), 1231-1272. doi: 10.1016/S1470-2045(15)00205-3

Rubin, G., McPhail, S., & Elliott, K. (2011). National audit of cancer diagnosis in primary care. *Royal College of General Practitioners*.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*(2), 252-266. doi: http://dx.doi.org/10.1016/0749-5978(92)90064-E

Singh, H., Giardina, T., Meyer, A. D., Forjuoh, S. N., Reis, M. D., & Thomas, E. J. (2013). Types and origins of diagnostic errors in primary care settings. *JAMA Internal Medicine, 173*(6), 418-425. doi: 10.1001/jamainternmed.2013.2777

Sirota, M., Juanchich, M., & Hagmayer, Y. (2014). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review, 21*(1), 198-204.

Sirota, M., Kostovičová, L., & Vallée-Tourangeau, F. (2015). Now you Bayes, now you don't: effects of set-problem and frequency-format mental representations on statistical reasoning. *Psychonomic Bulletin & Review, 22*(5), 1465-1473. doi: 10.3758/s13423-015-0810-y

Sirota, M., Vallée-Tourangeau, G., Vallée-Tourangeau, F., & Juanchich, M. (2015). On Bayesian problem-solving: Helping Bayesians solve simple Bayesian word problems. *Frontiers in Psychology, 6*. doi: 10.3389/fpsyg.2015.01141

Stewart, B. W., & Wild, C. P. (2014). *World Cancer Report 2014*. Lyon, France: International Agency for Research on Cancer.

Thomas, R., Dougherty, M. R., & Buttaccio, D. R. (2014). Memory constraints on hypothesis generation and decision making. *Current Directions in Psychological Science, 23*(4), 264-270. doi: 10.1177/0963721414534853

Thomas, R., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115*(1), 155-185.

Weber, E. U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1151-1164. doi: 10.1037/0278-7393.19.5.1151

Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Journal Of Personality And Social Psychology, 75*(6), 1411.

Table 1. Effect of prevalence and a prior alternative explanation on cancer diagnosis.

	Outcomes			
	Cancer as the main diagnosis		Cancer as the main or differential diagnosis	
	Unadjusted model	Adjusted model*	Unadjusted model	Adjusted model*
Factors	*OR*, 95% CI	*OR*, 95% CI	*OR*, 95% CI	*OR*, 95% CI
Prevalence	1.92, [1.27, 2.92]	2.20, [1.41, 3.44]	2.04, [1.39, 3.01]	2.10, [1.42, 3.12]
	$z = 3.07, p = .002$	$z = 3.47, p = .001$	$z = 3.61, p < .001$	$z = 3.69, p < .001$
Alternative	1.70, [1.11, 2.59]	1.85, [1.18, 2.91]	0.97, [0.68, 1.39]	1.03, [0.72, 1.48]
Explanation	$z = 2.45, p = .014$	$z = 2.67, p = .008$	$z = -0.14, p = .885$	$z = 0.16, p = .870$
Prevalence*	0.62, [0.35, 1.09]	0.61, [0.33, 1.12]	0.89, [0.52, 1.53]	0.86, [0.50, 1.49]
Alt. Explanation	$z = -1.66, p = .096$	$z = -1.60, p = .109$	$z = -0.41, p = .679$	$z = -0.53, p = .596$
Covariates				
Consultation		4.45, [3.30, 6.01]		0.76, [0.60, 0.96]
	—	$z = 9.74, p < .001$		$z = -2.28, p = .022$
Experience		1.03, [1.02, 1.05]		1.04, [1.02, 1.06]
	—	$z = 3.71, p < .001$		$z = 3.98, p < .001$

Note. OR =odds ratio; 95% CI = 95% confidence interval; z = z-test, p = statistical significance
*Adjusted for consultation and experience

Table 2: Effect of reporting cancer as the main working diagnosis on urgent cancer referral (two-week referral for cancer) and on any referral leading to cancer detection.

|	Outcomes			
	Early cancer referral (vs. delay or no referral)		Cancer referral (vs no referral)	
	Unadjusted model	Adjusted model	Unadjusted model	Adjusted model
Factors	*OR*, 95% CI	*OR*, 95% CI	*OR*, 95% CI	*OR*, 95% CI
Cancer as the main diagnosis	22.84, [16.14, 32.32] $z = 17.66, p < .001$	20.16, [13.86, 29.33] $z = 15.71, p < .001$	13.67, [9.83, 19.01] $z = 15.54, p < .001$	12.56, [8.75, 18.02] $z = 13.72, p < .001$
Covariates				
Consultation		5.24, [3.57, 7.69] $z = 8.47, p < .001$	—	4.78, [3.49, 6.55] $z = 9.73, p < .001$
Experience		0.99, [0.97, 1.01] $z = -0.59, p = .555$	—	1.00, [0.98, 1.01] $z = -0.46, p = .646$

Note. OR =odds ratio; 95% CI = 95% confidence interval; z = z-test, p = statistical significance
*Adjusted for consultation and experience

Figure Captions

Figure 1. Effects of the main factors on cancer given as the main diagnosis (in %) and on urgent

referral for cancer (in %). Panel A: Effect of cancer type and patient consultation on generating

cancer as the main diagnosis; Panel B: Effect of prevalence and alternative explanation on

generating cancer as the main diagnosis; Panel C: Effect of cancer type and patient consultation

on urgent referral for cancer; Panel D: Effect of prevalence and alternative explanation on urgent

referral for cancer.

Note. The error bars represent 95% Agresti-Coull confidence intervals for a binomial proportion.
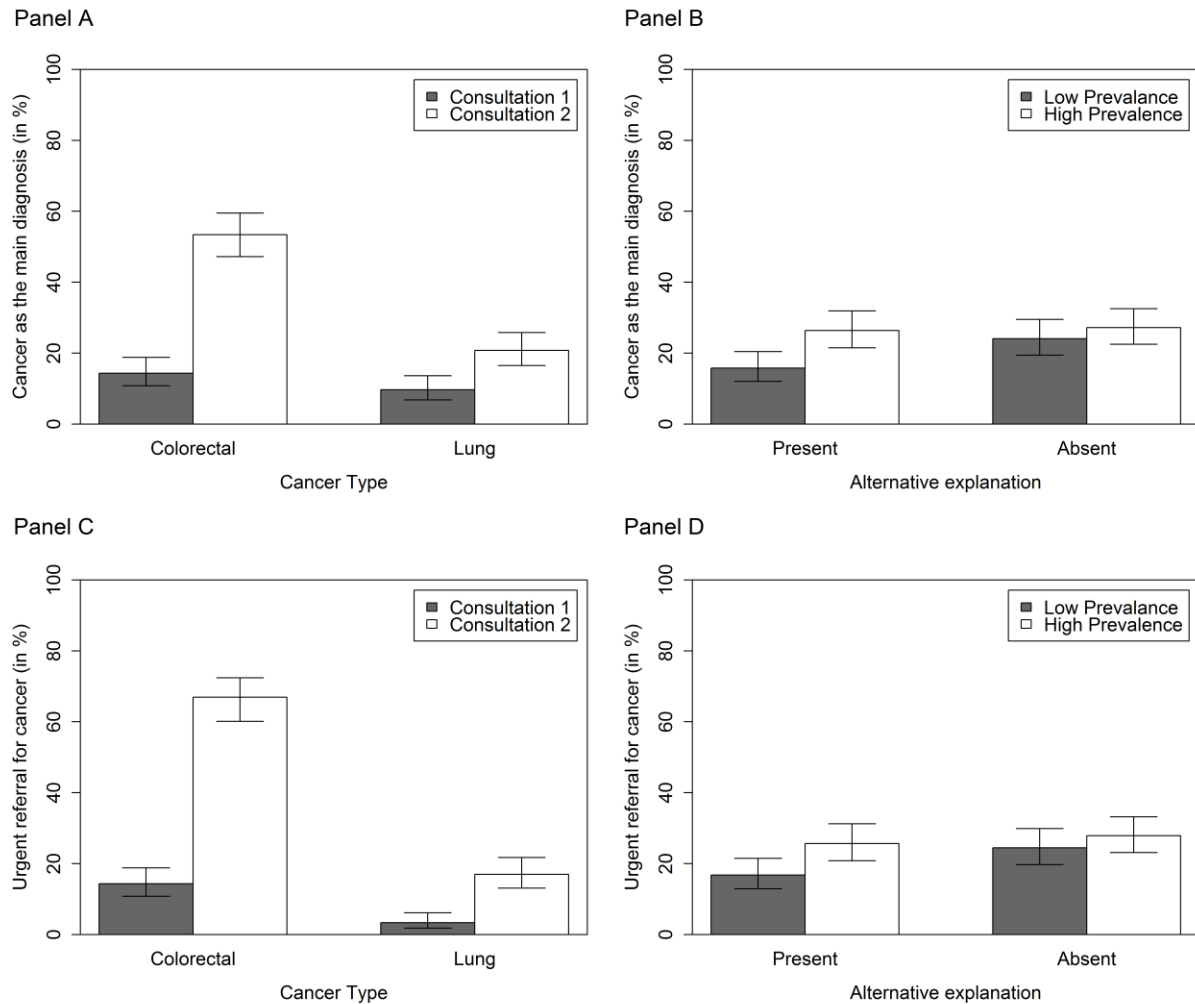
Supplemental Materials (A)

Vignettes

Colorectal cancer vignettes

1st consultation

Richard Drake is **42** [**70**] years old. He has been registered with your practice for the past 5 years. According to his medical record, he does not smoke and drinks alcohol rarely. *He had several episodes of constipation 3 years ago with bleeding PR. Sigmoidoscopy confirmed haemorrhoids. There are no other significant illnesses.* [*He has not had any significant illnesses in the past.*] You last saw him 5 months ago for an ankle sprain. He was treated with Ibuprofen and physiotherapy.

Today, he comes in complaining of constipation. He reports hard and infrequent stools for the past month. He normally has softer stools and opens his bowels once a day or once every two days. Now, however, he opens his bowels once every 4 to 5 days and complains of mild discomfort whilst passing stools. [*He has not experienced long bouts of constipation before.*] He has increased his fruits, vegetables and fluid intake, but it has not made much difference to his motions. He also complains of pain in his left lower abdomen. It comes and goes and does not relate to any particular food or time of day. Bowel motions seem to relieve the pain. He has not noticed any blood or mucus with his stools, his appetite is fine and he has not lost any weight.

Upon examination, he is not pale. Abdominal examination is normal and there is no discomfort. Digital rectal examination is also normal.

Note: The vignette presents a case with 'low prevalence' and a 'prior alternative explanation' at the first consultation. The manipulation of the prevalence is in **bold** *and the wording for high prevalence is in the squared brackets []; the manipulation of the prior alternative explanation is in* italics *and the wording for no alternative explanation available is in the squared brackets []. Please note that we have not used italics or bold fonts in the vignettes presented to participants. The wording in the vignettes was grammatically consistent with the gender of the presenting patient.*

2nd consultation

Richard Drake is **42** [**70**] years old. He has been registered with your practice for the past 5 years. According to his medical record, he does not smoke, drinks alcohol rarely. *He had several episodes of constipation 3 years ago with bleeding PR. Sigmoidoscopy confirmed haemorrhoids. There are no other significant illnesses.* [*He has not had any significant illnesses in the past.*]

Two weeks ago, he came in complaining of constipation and abdominal pain for one month. Further history and examination revealed no abnormality.

Today, he comes back to see you again. His motions have been a little better with laxatives but he has noticed blood with his stools, not with every stool but about 3-4 times during last week. The blood is red in colour and mixed with the stools. He has not noticed any mucus with it. He has not had pain while passing stools but he continues to have abdominal pain from time to time.

The digital rectal examination does not reveal any masses, no blood on the finger.

Results of the Full Blood Count:

Hb	12.6 g/dL	(13.5-18)
MCV	79fL	(76-100)

Supplemental Materials

WCC	8.6×10^9	(4.0-11.0)
Neutrophils	6×10^9	(2.0-7.5)
Lymphocytes	2.1×10^9	(1.3-3.5)
Monocytes	0.3×10^9	(0.2-0.8)
Eosinophils	0.2×10^9	(0-0.4)
Platelets	280×10^9	(150-450)

Note: The vignette presents a case with 'low prevalence' and a 'prior alternative explanation' at the second consultation. The manipulation of the prevalence is in **bold** *and the wording for high prevalence is in the squared brackets []; the manipulation of the prior alternative explanation is in* italics *and the wording for no alternative explanation available is in the squared brackets []. Please note that we have not used italics or bold fonts in the vignettes presented to participants. The wording in the vignettes was grammatically consistent with the gender of the presenting patient.*

Lung cancer vignettes

1st consultation
Christine Evans [**Michael Thompson**] is 60 years old. She has been registered with your practice for 15 years. She was diagnosed with essential hypertension 5 years ago and is currently on Amlodipine 10 mg. She has been smoking 10 cigarettes a day for 30 years and is a light alcohol drinker. *You saw her last month and diagnosed her with a viral upper respiratory tract infection.*[*the last sentence was omitted*]
She presents to you today with a cough productive of white sputum, *which has not resolved since her viral URTI last month* [*for one month*]. She always had a smoker's cough but it was intermittent and not too bothersome. This cough is a bit more noticeable and occasionally wakes her up at night. She has no orthopnoea and no paroxysmal nocturnal dyspnoea. Upon questioning, she does not report any other symptoms.
On examination she is apyrexial. ENT, cardiovascular and respiratory examinations are normal.

Note: The vignette presents a case with 'low prevalence' and a 'prior alternative explanation' at the first consultation. The manipulation of the prevalence is in **bold** *and the wording for high prevalence is in the squared brackets []; the manipulation of the prior alternative explanation is in* italics *and the wording for no alternative explanation available is in the squared brackets []. Please note that we have not used italics or bold fonts in the vignettes presented to participants. The wording in the vignettes was grammatically consistent with the gender of the presenting patient.*

2nd consultation
Christine Evans [**Michael Thompson**] is 60 years old. She has been registered with your practice for 15 years. She was diagnosed with essential hypertension 5 years ago and is currently on Amlodipine 10 mg. She has been smoking 10 cigarettes a day for 30 years and is a light alcohol drinker.
You saw her 4 months ago and diagnosed her with a viral URTI. She consulted again a month later with a cough productive of white sputum, which was not resolving following her URTI. [*You saw her 3 months ago with a cough productive of white sputum*] The cough was a bit more persistent and bothersome than her usual smoker's cough. The tests done at that time (chest X-ray, spirometry, FBC, U&E, LFTs, fasting lipids and glucose) were all normal.

She comes to see you again today. Her previous episode of cough resolved but the cough came back a few weeks ago *following another bout of cold*. [*the last four words were omitted*]. She also reports feeling a bit out of breath when she runs for the bus, more tired and not having her usual appetite. She has no chest pain, orthopnoea or paroxysmal nocturnal dyspnoea. She has not lost any weight. Upon questioning, she does not report any other symptoms.
On examination she is apyrexial. ENT, cardiovascular and respiratory examinations are normal. There is no ankle oedema.

*Note: The vignette presents a case with 'low prevalence' and a 'prior alternative explanation' at the second consultation. The manipulation of the prevalence is in **bold** and the wording for high prevalence is in the squared brackets []; the manipulation of the prior alternative explanation is in* italics *and the wording for no alternative explanation available is in the squared brackets []. Please note that we have not used italics or bold fonts in the vignettes presented to participants. The wording in the vignettes was grammatically consistent with the gender of the presenting patient.*

Supplemental Materials (B)

Example of a response page

Now, please enter the diagnosis/es you are considering and a management plan.

Michael Thompson is 60 years old. He has been registered with your practice for 15 years. He was diagnosed with essential hypertension 5 years ago and is currently on Amlodipine 10 mg. He has been smoking 10 cigarettes a day for 30 years and is a light alcohol drinker.

He presents to you today with a cough productive of white sputum for one month. He always had a smoker's cough but it was intermittent and not too bothersome. This cough is a bit more noticeable and occasionally wakes him up at night. He has no orthopnoea and no paroxysmal nocturnal dyspnoea. Upon questioning, he does not report any other symptoms.

On examination he is apyrexial. ENT, cardiovascular and respiratory examinations are normal.

Please enter:

1) What is your working diagnosis? (Please enter only one) * Required Field

☐

2) If you are you considering any other diagnoses, enter them below.

☐

3) How would you manage this patient?

(Please tick as many as appropriate)

Select at least one option

☐ Arrange follow up

☐ Ask patient to come back if symptoms persist

☐ Order investigations

☐ Prescribe

☐ Refer to secondary care

Next

Note. Selecting the option