

# Do prevalence expectations affect patterns of visual search and decision-making in interpreting CT colonography endoluminal videos?

Fanshawe, Thomas R; Plumb, Andrew; Phillips, Peter; Helbren, Emma; Halligan, Steve; Taylor, Stuart A; Gale, Alastair; Mallett, Susan

DOI:

[10.1259/bjr.20150842](https://doi.org/10.1259/bjr.20150842)

License:

None: All rights reserved

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Fanshawe, TR, Plumb, A, Phillips, P, Helbren, E, Halligan, S, Taylor, SA, Gale, A & Mallett, S 2016, 'Do prevalence expectations affect patterns of visual search and decision-making in interpreting CT colonography endoluminal videos?', *British Journal of Radiology*, vol. 89, no. 1060, 20150842. <https://doi.org/10.1259/bjr.20150842>

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**

Final Version of Record, published as above

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

Cite this article as:

Fanshawe TR, Phillips P, Plumb A, Helbren E, Halligan S, Taylor SA, et al. Do prevalence expectations affect patterns of visual search and decision-making in interpreting CT colonography endoluminal videos? *Br J Radiol* 2016; **89**: 20150842.

## FULL PAPER

# Do prevalence expectations affect patterns of visual search and decision-making in interpreting CT colonography endoluminal videos?

<sup>1</sup>THOMAS R FANSHAWE, PhD, <sup>2</sup>PETER PHILLIPS, PhD, <sup>3</sup>ANDREW PLUMB, MRCP, FRCR, <sup>3</sup>EMMA HELBREN, FRCR, <sup>3</sup>STEVE HALLIGAN, PhD, FRCR, <sup>3</sup>STUART A TAYLOR, MD, FRCR, <sup>4</sup>ALASTAIR GALE, PhD and <sup>5</sup>SUSAN MALLETT, DPhil

<sup>1</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

<sup>2</sup>Health and Medical Sciences Group, University of Cumbria, Lancaster, UK

<sup>3</sup>Centre for Medical Imaging, University College London, London, UK

<sup>4</sup>Applied Vision Research Centre, Loughborough University, Loughborough, UK

<sup>5</sup>Public Health, Epidemiology and Biostatistics, Institute of Applied Health Sciences, University of Birmingham, Birmingham, UK

Address correspondence to: Dr Thomas R Fanshawe

E-mail: [thomas.fanshawe@phc.ox.ac.uk](mailto:thomas.fanshawe@phc.ox.ac.uk)

**Objective:** To assess the effect of expected abnormality prevalence on visual search and decision-making in CT colonography (CTC).

**Methods:** 13 radiologists interpreted endoluminal CTC fly-throughs of the same group of 10 patient cases, 3 times each. Abnormality prevalence was fixed (50%), but readers were told, before viewing each group, that prevalence was either 20%, 50% or 80% in the population from which cases were drawn. Infrared visual search recording was used. Readers indicated seeing a polyp by clicking a mouse. Multilevel modelling quantified the effect of expected prevalence on outcomes.

**Results:** Differences between expected prevalence were not statistically significant for time to first pursuit of the polyp (median 0.5 s, each prevalence), pursuit rate when no polyp was on screen (median  $2.7\text{ s}^{-1}$ , each

prevalence) or number of mouse clicks [mean 0.75/video (20% prevalence), 0.93 (50%), 0.97 (80%)]. There was weak evidence of increased tendency to look outside the central screen area at 80% prevalence and reduction in positive polyp identifications at 20% prevalence.

**Conclusion:** This study did not find a large effect of prevalence information on most visual search metrics or polyp identification in CTC. Further research is required to quantify effects at lower prevalence and in relation to secondary outcome measures.

**Advances in knowledge:** Prevalence effects in evaluating CTC have not previously been assessed. In this study, providing expected prevalence information did not have a large effect on diagnostic decisions or patterns of visual search.

## INTRODUCTION

If we are expecting an event, we are more alert to it and more likely to react when it occurs.<sup>1</sup> We might expect that radiologists are more alert to the presence of an abnormality when given an indication that prevalence is particularly high and, conversely, be less alert when the chance of encounter is believed to be low, as in screening.

Interpretation of medical imaging occurs in three environments: the symptomatic population, the asymptomatic/screening population and the research setting. Expected levels of abnormality vary considerably between these settings and between different medical specialities.<sup>2</sup> It follows that the effect of varying prevalence of abnormality on image interpretation is crucial to our understanding of how

diagnostic accuracy and interpretative performance might change across reporting environments.

In 2011, a systematic review<sup>3</sup> found only three medical imaging studies<sup>4-6</sup> that assessed the impact of experimentally modified prevalence on reader diagnosis. Subsequent studies have been published,<sup>7-10</sup> but the relationship between prevalence and interpretation accuracy remains unclear. Some studies report increased false negatives or reduced diagnostic confidence at lower prevalence levels, for example, for interpretation of pulmonary arteriograms,<sup>4</sup> mammograms<sup>8,11</sup> or ankle trauma radiographs.<sup>7</sup> This “rare target” effect has also been reported in non-clinical scenarios, such as baggage scanning<sup>12,13</sup> and artificial target search experiments.<sup>14</sup> By contrast, in chest radiography, the evidence for a prevalence effect on diagnostic accuracy is weaker,<sup>5,9</sup> although

two studies that used eye tracking to monitor visual search of experienced readers suggested a possible association between increased prevalence and the duration and pattern of image scrutiny.<sup>10,15</sup>

Despite increasing use of CT colonography (CTC) in routine practice, there is little research describing the effect of abnormality prevalence on diagnostic performance.<sup>3</sup> This is surprising because CTC is commonly applied across a wide range of expected prevalence, from asymptomatic individuals undergoing screening<sup>16–18</sup> to symptomatic and high-risk patients.<sup>19–21</sup> Establishing the presence or absence of a prevalence effect on reader attention, visual search and diagnostic performance is important both in understanding how CTC should be used in clinical practice and for designing future research studies.

The purpose of this study was to assess the effect of expected abnormality prevalence on visual search and decision-making in CTC.

## METHODS AND MATERIALS

Research ethics committee approval was obtained to record eye-tracking data from consenting observers in this prospective study. Institutional review board and research ethics committee approval was granted to use anonymous CTC data collated in previous studies.<sup>22,23</sup>

### Participants and cases

13 radiologists (readers) were recruited from a UK training hospital over 2 days in July 2012. All provided written, informed consent. Readers (6/13 males; mean age 32 years, range 27–36 years) were trainees with 1–7 years' experience as a radiologist and at most 50 cases CTC experience.

10 CTC endoluminal fly-through videos lasting 30 s each were generated (EH, PP) with dedicated CTC software on a medical imaging workstation (Vitrea®; Vital Images, Inc., MN) and exported for viewing. Navigation speed was fixed at approximately  $1.5 \text{ cm s}^{-1}$ . Five videos depicted a single colorectal polyp (true positive, 5–8 mm maximal transverse dimension), verified by three radiologists with >200 cases' experience.<sup>23</sup> To counteract recall, cases were excluded if they contained polyps within 5 s navigation of the caecal pole, rectal ampulla or insufflation catheter, or contained other distinctive characteristics, assessed by a radiologist with 6 years' experience (EH). Polyps were on screen for between 2.4 and 11.1 s. The remaining five videos (true negative) were selected from different sections of the colon, containing no polyps, in the same patient group.

The sample size was based on practical considerations: the number of readers available and the number of cases that could be assessed comfortably in one sitting. As the primary outcome measures have not been used before in this context, no power calculation was performed.

### Data collection

The group of 10 videos was presented to each reader three times in one sitting, with an optional break between the groups. The

order of cases was randomized for each reader. Before viewing each group, readers were told that the videos in that group came from a population with known prevalence of abnormality—20%, 50% or 80%. The ordering of the three prevalence scenarios was varied between readers using block randomization. Readers were not told that the three groups actually contained the same 10 videos repeated three times and were therefore unaware that the true prevalence was identical (50%) and the declared 20% and 80% prevalence levels were incorrect. Information given to readers was worded as:

“We are going to show you 3 groups of 10 videos in a random order.

Each group is taken from a different population, each with a different prevalence of abnormality.

Before each group we will tell you the population prevalence, either 80%, 50% or 20%.”

Readers were asked to hold a computer mouse throughout and indicate with a click (polyp identification) when they saw a lesion that they considered highly likely to represent a real polyp or cancer. Readers were not required to specify polyp location and could not pause, rewind or review videos. They were not told which videos contained polyps and were given no feedback about their performance. Data collection took 20–30 min per reader.

### Viewing conditions

Reading was conducted in a quiet room with constant, ambient light. A liquid-crystal display monitor,  $1280 \times 1024$  pixel resolution, was used (SyncMaster 971P; Samsung, Suwon, Republic of Korea and Fujitsu E19-5; Fujitsu, Tokyo, Japan; 1 pixel = 0.29 mm). The screen was positioned 60 cm in front of the reader. Videos measured  $512 \times 512$  pixels ( $14.8 \times 14.8$  cm), representing a visual angle of  $14.1^\circ$ . The eye position of readers was recorded using a Tobii X50 or X120 eye tracker (Tobii Technology AB, Danderyd, Sweden), sampling at 50 or 60 Hz, respectively, positioned beneath the screen. No headrest was used. Readers wore glasses or contact lenses as normal. They performed a nine-point calibration procedure prior to data collection and were excluded if this could not be completed. They then viewed a supplemental warm-up video prior to data collection. They were not asked to fixate a particular point before each video.

### Data preparation

The eye position data were prepared for analysis as described elsewhere;<sup>24</sup> a summary follows. True-positive polyps were approximated using a circular region of interest (ROI), manually overlaid onto each video frame-by-frame by a medical image perception scientist (PP). The centre and radius of this ROI were adjusted manually to match the polyp's transition across the screen. Within each frame, the perpendicular distance between the recorded eye position and the edge of the ROI was calculated and used in outcome measures described below. Eye gaze falling within a 50-pixel acceptance radius from the edge of the ROI was considered to be within high visual acuity. For periods when

no polyp was visible, the  $(x, y)$  eye position co-ordinates were retained for analysis. Co-ordinates located  $>100$  pixels outside the screen area were excluded as recording errors.

### Outcome measures

Eye co-ordinate data were used to derive three primary and six secondary pre-specified outcomes (metrics) (Table 1). Figure 1 shows an example eye-tracking trace (distance between eye position and ROI over time) to illustrate metric definitions. Detailed information about metric derivations has been reported previously.<sup>25</sup> Metrics reflected three aspects of reader behaviour: eye position when a polyp was on screen; eye position when no polyp was on screen; and frequency and accuracy of polyp identifications. Primary outcomes were time to first pursuit of the ROI; pursuit rate in the absence of an ROI; total number of polyp identifications. The “screen coverage” measure was defined by the proportion of eye gaze falling into three regions: within, above or below a  $256 \times 256$ -pixel square at the centre of the screen. “Any correct identification” and the “polyp on screen” metrics are defined only for true-positive videos. “Any incorrect identification” is defined only for the period before any polyp appeared, to prevent readers who delayed their decision after seeing a polyp being misclassified as making a false-positive identification.

### Statistical analysis

Metrics were analyzed using multilevel modelling, incorporating independent random intercepts for reader and video, including prevalence level as a factor. Effects of prevalence expectation

were expressed relative to the true 50% prevalence category. In a planned sensitivity analysis, to test whether results were altered by the order (first, second or third viewing) in which the prevalence categories were presented, this order was included as an additional factor variable.

Within this multilevel framework, proportional hazards, logistic and Poisson models were used, as appropriate for the data type. As most viewings had at least one missing eye position data point, short missing data runs were imputed, based on the recorded eye co-ordinates immediately before and after, and adding random measurement error. Estimates were combined using multiple imputation methods with 10 imputations.<sup>26</sup> Cases with  $>50\%$  missing values or  $>50$  consecutive missing values were examined individually by two authors (TF, AP) and removed if deemed likely to make the metric calculation highly unreliable. The electronic [Supplementary material](#) contains more details.

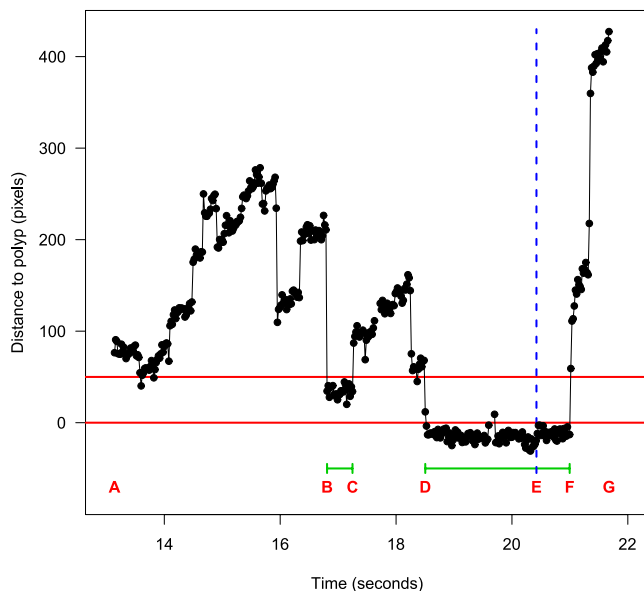
A different approach was adopted only for pursuit rate, which has no generally agreed definition.<sup>27</sup> We used the number of pursuits calculated by Tobii Studio v. 1.7.2 (50-pixel dispersion, 100-ms minimum time threshold) throughout the period when no polyp was on screen, divided by the duration of this period. Time points when the Tobii software failed to identify whether a co-ordinate belonged to any particular pursuit were excluded, and the time denominator adjusted accordingly. Cases with  $>50\%$  missing values of the pursuit classifier were excluded from analysis.

Table 1. Metric definitions. The identifying letters A, B etc. refer to time points indicated in Figure 1

Group	Name	Definition
Polyp on screen	Time to first pursuit <sup>a</sup>	Time between appearance of polyp (A) and start of first pursuit of polyp (B)
	Total assessment time span	Time between start of first pursuit of polyp (B) and polyp identification (E)
	Assessment pursuit time	Cumulative time in pursuit of polyp before polyp identification (B–C and D–E), expressed as a proportion of the total time when the polyp was visible (A–G)
	Assessment pursuit rate	Number of separate pursuits of polyp before polyp identification, divided by the total time when the polyp was visible before polyp identification (A–E)
Polyp off screen	Pursuit rate <sup>a</sup>	Number of distinct eye pursuits, divided by the total time when the polyp was off screen
	Screen coverage	Proportion of eye co-ordinates falling in to each of three regions of the screen display, “upper”, “central” and “lower” (Figure 2)
Polyp identification	Total number of identifications <sup>a</sup>	Number of identifications recorded over whole video
	Any correct identification	Binary indicator of whether an identification occurred while the polyp was visible (a reaction time of 0.5 s after the polyp left the screen was allowed)
	Any incorrect identification	Binary indicator of whether an identification occurred before the polyp was visible (or at any time, for true-negative videos)

<sup>a</sup>Primary outcome.

Figure 1. Illustration of distance between eye position and polyp [edge of region of interest (ROI)] over time for a single video viewing. Letters used in explanation of metric definitions, A: polyp becomes visible, B-C: first eye pursuit of ROI, D-F: second eye pursuit of ROI, E: polyp identification (indicated by dotted line), G: polyp disappears from view. Note short periods of missing data at 17.7 and 19.7 s. The horizontal line at distance 0 represents the edge of the ROI, and the horizontal line at distance 50 pixels represents the high visual acuity region within which eye pursuits of the ROI may occur.



Results are presented as point estimates with 95% confidence intervals (95% CIs) and  $p$ -values. A 5% significance level was used, unadjusted for multiple testing.

Statistical analysis used STATA® v. 12.1 for Windows (StataCorp, College Station, TX) and R version 3.1.1.<sup>28</sup>

## RESULTS

Eye tracking was successful and 389 of the intended 390 viewings were completed. Seven (1.8%) of these were omitted from the

analysis of one or more metrics (with the exception of pursuit rate) because patterns of missing data made calculation unreliable. For pursuit rate, 37 (9.5%) of the viewings were excluded.

Table 2 summarizes metrics across all readers within each prevalence scenario. Of the videos that contained a polyp, readers made at least one pursuit of the polyp for 185 of the 190 (97%) viewings with reliable data.

There were no statistically significant differences between expected prevalence levels in any metric relating to visual search while the polyp was visible (Table 3). In each prevalence scenario, readers took approximately half a second on average to direct their gaze to the ROI after the polyp appeared [hazard ratio 1.32 (95% CI 0.95 to 1.93,  $p = 0.14$ ) for 20% vs 50% prevalence; hazard ratio 0.95 (95% CI 0.64 to 1.40,  $p = 0.79$ ) for 80% vs 50% expected prevalence; Tables 2 and 3, Figure 3]. Average total assessment time span, assessment pursuit time and assessment pursuit rate were also similar in the three prevalence scenarios (Tables 2 and 3).

During the period when the polyp was not on screen, the average pursuit rate was approximately 2.7 pursuits per second at each of the three prevalence levels (Table 2), with no statistically significant differences (Table 3). There was a tendency for readers' gaze to fall inside the central region of the screen less often at the 80% prevalence level than at the 50% prevalence level [odds ratio 0.82 (95% CI 0.72 to 0.95,  $p = 0.008$ ), Table 3], with a concomitant increase in the upper region. This effect, however, was small, with on average 82% of gaze points falling in the central region at 80% prevalence compared with 84% at 50% prevalence (Table 2).

There were no statistically significant differences with respect to expected prevalence regarding the total number of identifications (Table 3). As expected, the average number of identifications was higher for videos that contained polyps than for those that did not (1.3 vs 0.4, Table 2). The sensitivity, or probability of a polyp being correctly identified, was higher at 50% prevalence (86%) than at 20% prevalence (71%). This difference was statistically significant ( $p = 0.01$ , Table 3) but the trend did not persist at the 80%

Figure 2. Illustration of the screen coverage metric, showing the division of the screen area into upper, central and lower regions (dashed lines). The central region occupies a  $256 \times 256$ -pixel square at the centre of the  $512 \times 512$ -pixel screen area (solid line). An additional 100-pixel margin (shown by the outer bounding box) was allowed for gaze points measured outside the screen area; this was incorporated into the upper or lower region, as appropriate. Superimposed is the pattern of gaze over the entire video duration for a single reader (Reader 11) viewing the same case (Case 3) under different prevalence conditions: 20% (left panel), 50% (middle panel) and 80% (right panel).

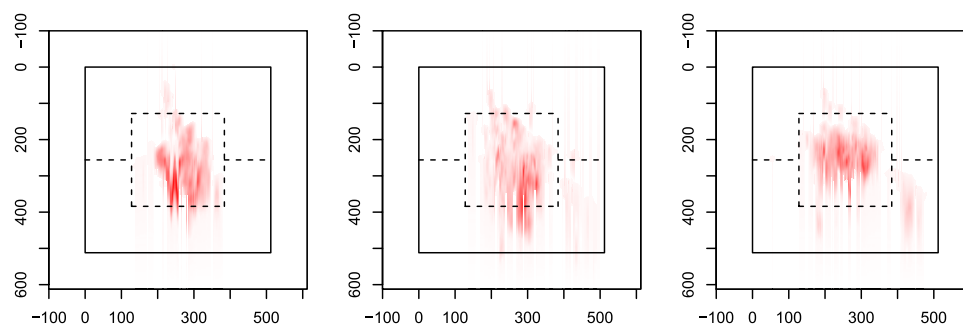


Table 2. Summary of metrics by prevalence level [number (%) or median (interquartile range), except for the total number of identifications: mean (standard deviation)]

Metric	20% prevalence	50% prevalence	80% prevalence
At least one pursuit of polyp	63/63 (100)	61/64 (95)	61/63 (97)
Immediate pursuit	5/63 (8)	4/64 (6)	10/63 (16)
Time to first pursuit (s) <sup>a</sup>	0.45 (0.26–0.65)	0.52 (0.28–0.82)	0.52 (0.37–0.95)
Total assessment time span (s) <sup>a</sup>	2.45 (1.33–5.96)	1.75 (1.00–3.49)	2.19 (1.15–5.76)
Assessment pursuit time (%)	24 (14–34)	21 (13–33)	18 (12–33)
Assessment pursuit rate (s <sup>-1</sup> )	0.59 (0.42–0.79)	0.56 (0.42–0.83)	0.69 (0.45–0.85)
Pursuit rate (s <sup>-1</sup> )	2.69 (2.19–3.09)	2.67 (2.23–3.02)	2.71 (2.26–3.11)
Screen coverage (%)			
Upper	6 (3–13)	7 (5–12)	9 (5–15)
Central	87 (77–92)	84 (77–90)	82 (73–89)
Lower	7 (4–12)	8 (5–13)	8 (6–13)
Total number of identifications	0.75 (0.82)	0.93 (0.90)	0.97 (1.07)
Videos with polyps	1.17 (0.80)	1.38 (0.90)	1.43 (1.16)
Videos without polyps	0.34 (0.59)	0.49 (0.66)	0.51 (0.73)
Any correct identification	46/65 (71)	55/64 (86)	49/65 (75)
Any incorrect identification	39/130 (30)	48/129 (37)	51/130 (39)
Videos with polyps	21/65 (32)	22/64 (34)	25/65 (38)
Videos without polyps	18/65 (28)	26/65 (40)	26/65 (40)

<sup>a</sup>Kaplan–Meier estimate, calculated without allowing for clustering, excluding viewings with immediate pursuit.

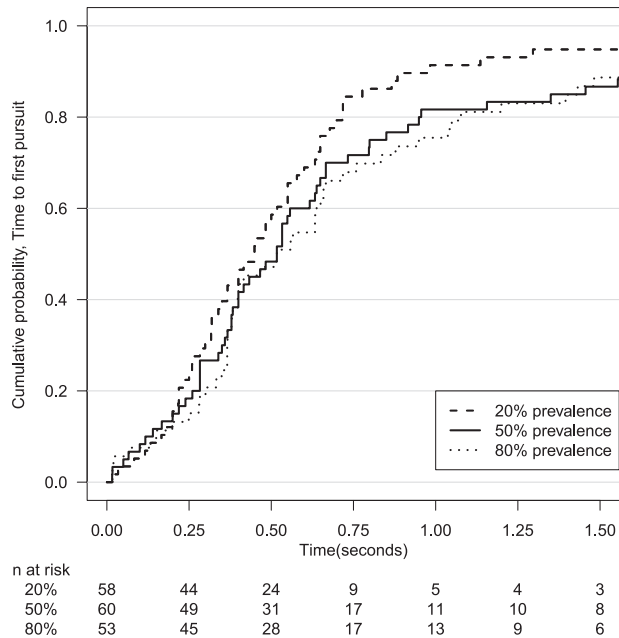
prevalence level (75%). This metric was subject to an extremely high case-specific effect (Figure 4), as in three videos 1, 2 and 4 almost every reader identified the polyp at each prevalence level; the other

two videos 3 and 5, for which the polyp was superficially more difficult to identify, are therefore likely primarily responsible for the differences in rates of correct identification.

Table 3. Comparison of metrics between prevalence levels: hazard ratio (HR), odds ratio (OR) or rate ratio (RR), as appropriate, with 95% confidence interval (CI) and *p*-value

Metric	Measure	20% vs 50% prevalence		80% vs 50% prevalence	
		Effect size (95% CI)	<i>p</i> -value	Effect size (95% CI)	<i>p</i> -value
Time to first pursuit	HR	1.32 (0.95–1.93)	0.14	0.95 (0.64–1.40)	0.79
Total assessment time span	HR	0.74 (0.50–1.12)	0.15	0.83 (0.56–1.24)	0.37
Assessment pursuit time	OR	1.27 (0.87–1.84)	0.22	0.90 (0.62–1.32)	0.60
Assessment pursuit rate	RR	0.91 (0.70–1.18)	0.47	1.07 (0.83–1.37)	0.60
Pursuit rate	RR	1.01 (0.98–1.05)	0.39	1.03 (1.00–1.07)	0.06
Screen coverage					
Upper	OR	0.93 (0.78–1.12)	0.45	1.28 (1.07–1.53)	0.007
Central	OR	1.06 (0.92–1.23)	0.39	0.82 (0.72–0.95)	0.008
Lower	OR	0.96 (0.81–1.13)	0.63	1.11 (0.94–1.31)	0.22
Total number of identifications	RR	0.81 (0.62–1.06)	0.12	1.04 (0.81–1.34)	0.75
Any correct identification	OR	0.24 (0.08–0.73)	0.01	0.37 (0.12–1.11)	0.08
Any incorrect identification	OR	0.66 (0.37–1.19)	0.17	1.11 (0.63–1.97)	0.71
Videos with polyps	OR	0.86 (0.35–2.11)	0.75	1.29 (0.54–3.10)	0.57
Videos without polyps	OR	0.53 (0.24–1.17)	0.11	1.00 (0.47–2.13)	1.00

Figure 3. Kaplan-Meier curves showing time to first pursuit in the three prevalence conditions. The vertical axis shows the proportion of viewings for which a pursuit has occurred prior to the times shown on the horizontal axis. Below the plot, the number of viewings per group for which a pursuit has not yet occurred is shown.



The probability of an incorrect identification (false positive) ranged from 30% at 20% prevalence to 39% at 80% prevalence; this difference was also not statistically significant (Table 3). On average, incorrect identifications occurred with similar frequency for videos that contained no polyps and for videos that contained polyps during periods when the polyp was not visible, although there was considerable variability between cases (Figure 4). Some false-positive features were identified with a mouse click by several readers (e.g. Case 3 at 5 s, Figures 4 and 5).

In sensitivity analysis, including as an extra factor variable, the order in which the prevalence scenarios were presented did not affect the prevalence effect sizes shown in Table 3.

## DISCUSSION

This study investigated the effect on visual search and decision-making for CTC of providing readers with substantially different expectations of the likely prevalence of abnormality in the population from which cases were drawn. We did not demonstrate a strong link between prevalence expectation and the pattern of search or decision-making.

Our conclusion differs from those of several studies<sup>8,12-14</sup> using scenarios other than CTC that found increased false-negative rate at lower prevalence levels. Our study showed a statistically significant increase in the proportion of polyp identifications between 20% and 50% expected prevalence, but for three reasons this finding should be treated cautiously. First, it did not extend to the highest prevalence level, for which the proportion was similar to that at 20%, and a non-monotonic relationship seems implausible.

Second, the effect was driven by an increased true-positive rate in just two of the five cases with polyps: a consistent increase across all cases, which would have provided more convincing evidence, was not observed. Third, this was just one of several secondary analyses performed, and so it may be a chance result.

The existence of a prevalence effect is not a universal finding in image interpretation studies. For example, Gur et al<sup>5</sup> found that varying prevalence levels between 2% and 21% did not affect the diagnostic accuracy of chest radiograph assessment. Likewise, we did not find a prevalence effect for our three primary outcomes, which were chosen to represent visual search and decision-making. Modality may therefore be an important determinant of prevalence effects.

We have shown previously that time to first pursuit of the polyp changes with reader experience and the presence of a computer-aided detection marker;<sup>29,30</sup> in the present study, this metric was unchanged across prevalence scenarios. When no polyp was visible, readers tended to spend more time, proportionally, looking at peripheral screen regions in the 80% prevalence condition, but this effect is small and is not supported by changes in other visual search metrics. However, the finding requires further investigation as our measure is based on a simple square at the centre of the screen area, which may not adequately capture gaze narrowing effects.

We used a common set of cases for each of the prevalence conditions to directly observe the effect of disclosing different prevalence information, as opposed to the effect of the true case mix. Lau et al<sup>31</sup> claim that the latter may have a larger effect on decision-making, but testing this was not our

Figure 4. Time points within each video at which polyp identifications occurred. Prevalence conditions are indicated by different colours. Cases that contain a polyp are labelled 1-5, and the bar indicates the period during which the polyp was visible on the screen. Cases with no polyps are labelled 6-10.

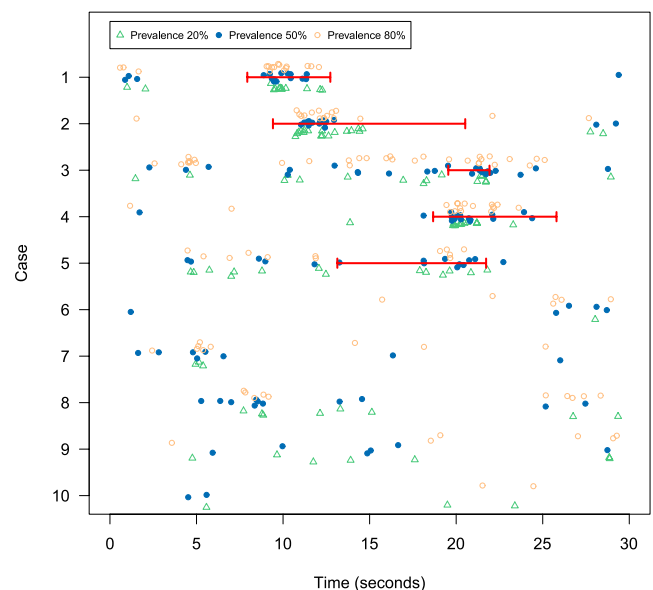
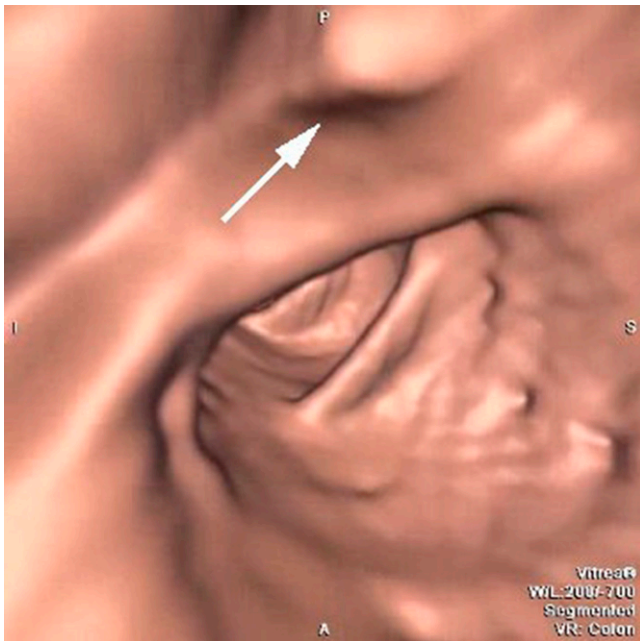


Figure 5. Screen capture from one of the displayed videos (Case 3, at around 5s) showing a feature provoking a false positive, in this case a mildly bulbous but normal fold (arrow).



objective. Indeed, it would have been infeasible for readers to make an assessment of the true underlying prevalence within a realistic time frame. It is possible that some readers realized that they had viewed videos more than once, but this is unlikely to have a major effect on our findings; the order in which the prevalence conditions were presented was determined randomly and this order was not strongly associated with outcomes. Enabling all cases to be viewed with comfort in a single sitting was an important practical consideration in our choice of the number of cases used. Despite the number of cases being moderately small, repeated viewings of the same case under different prevalence conditions enabled quantities of interest to be estimated with acceptable precision.

Future studies should assess further the possibility of a threshold effect in CTC. It is possible that the expected prevalence level needs to be  $<20\%$  for an effect to be visible, as is usually the case in everyday clinical practice, except in very high-risk patient groups such as those examined following a positive faecal occult blood test.<sup>21</sup> Evans et al<sup>8</sup> found a marked reduction in sensitivity

for breast cancer diagnosis using mammography during screening when the prevalence was extremely low (0.3%). Whether a similar effect applies to CTC remains unknown. Additionally, prevalence effects may vary according to the ease of visualization and identification of the cases chosen.

This study has limitations. This study was exploratory in nature, and therefore we may not have used enough cases for subtler prevalence effects to be detected. Endoluminal fly-through view was presented in automatic mode only, so readers could not adjust navigation speed as in usual practice. We were therefore unable to assess the effect of prevalence on the time the reader would spend scrutinizing each video; from laboratory experiments and some clinical studies, there is evidence that assessment time is affected by prevalence in static viewing modes.<sup>15,32</sup> Mouse clicks are not synonymous with definitive decisions about the presence of polyps and thus can only be regarded as proxy measures of diagnostic accuracy. Readers were not asked to identify polyp locations and so, even with eye-tracking data, it is impossible to state with certainty the cause of any particular click. Readers were inexperienced in CTC, and so our findings are not directly generalizable to experienced radiologists using CTC in day-to-day clinical practice. Finally, we did not assess the effect of providing information about the spectrum of disease severity, since readers received prevalence information alone.

In summary, CTC readers were provided with different estimates of the prevalence of abnormalities from which cases were drawn, and study results did not demonstrate a strong link between prevalence information and the pattern of visual search or decision-making. Further research should investigate effects at lower prevalence levels, such as might be present in asymptomatic populations.

## FUNDING

This work was supported by the UK National Institute for Health Research (NIHR) under its Program Grants for Applied Research funding scheme (RP-PG-0407-10338). A proportion of this work was undertaken at the University College London and University College London Hospital, which receive a proportion of funding from the NIHR Biomedical Research Centre funding scheme. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

## REFERENCES

1. Deese J. Some problems in the theory of vigilance. *Psychol Rev* 1955; **62**: 359–68. doi: <http://dx.doi.org/10.1037/h0042393>
2. Kundel HL. Disease prevalence and radiological decision making. *Invest Radiol* 1982; **17**: 107–9. doi: <http://dx.doi.org/10.1097/00004424-198201000-00020>
3. Boone D, Halligan S, Mallett S, Taylor SA, Altman DG. Systematic review: bias in imaging studies—the effect of manipulating clinical context, recall bias and reporting intensity. *Eur Radiol* 2012; **22**: 495–505. doi: <http://dx.doi.org/10.1007/s00330-011-2294-0>
4. Eggin TK, Feinstein AR. Context bias: a problem in diagnostic radiology. *JAMA* 1996; **276**: 1752–5. doi: <http://dx.doi.org/10.1001/jama.1996.03540210060035>
5. Gur D, Rockette HE, Armfield DR, Blachar A, Bogan JK, Brancatelli G, et al. Prevalence effect in a laboratory environment. *Radiology*



- 2003; **228**: 10–4. doi: <http://dx.doi.org/10.1148/radiol.2281020709>
6. Gur D, Bandos AI, Fuhrman CR, Klym AH, King JL, Rockette HE. The prevalence effect in a laboratory environment: changing the confidence ratings. *Acad Radiol* 2007; **14**: 49–53. doi: <http://dx.doi.org/10.1016/j.acra.2006.10.003>
  7. Pusic MV, Andrews JS, Kessler DO, Teng DC, Pecaric MR, Ruzal-Shapiro C, et al. Prevalence of abnormal cases in an image bank affects the learning of radiograph interpretation. *Med Educ* 2012; **46**: 289–98. doi: <http://dx.doi.org/10.1111/j.1365-2923.2011.04165.x>
  8. Evans KK, Birdwell RL, Wolfe JM. If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One* 2013; **8**: e64366. doi: <http://dx.doi.org/10.1371/journal.pone.0064366>
  9. Nocum DJ, Brennan PC, Huang RT, Reed WM. The effect of abnormality-prevalence expectation on naïve observer performance and visual search. *Radiography* 2013; **19**: 196–9. doi: <http://dx.doi.org/10.1016/j.radi.2013.04.004>
  10. Reed WM, Chow SLC, Chew LE, Brennan PC. Can prevalence expectations drive radiologists' behavior? *Acad Radiol* 2014; **21**: 450–6. doi: <http://dx.doi.org/10.1016/j.acra.2013.12.002>
  11. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The “laboratory” effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 2008; **249**: 47–53. doi: <http://dx.doi.org/10.1148/radiol.2491072025>
  12. Wolfe JM, Horowitz TS, Kenner NM. Rare items often missed in visual searches. *Nature* 2005; **435**: 439–40. doi: <http://dx.doi.org/10.1038/435439a>
  13. Wolfe JM, Horowitz TS, Van Wert MJ, Kenner NM, Place SS, Kibbi N. Low target prevalence is a stubborn source of errors in visual search tasks. *J Exp Psychol Gen* 2007; **136**: 623–38. doi: <http://dx.doi.org/10.1037/0096-3445.136.4.623>
  14. Rich AN, Kunar MA, Van Wert MJ, Hidalgo-Sotelo B, Horowitz TS, Wolfe JM. Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *J Vis* 2008; **8**: 1–17. doi: <http://dx.doi.org/10.1167/8.15.15>
  15. Reed WM, Ryan JT, McEntee MF, Evanoff MG, Brennan PC. The effect of abnormality-prevalence expectation on expert observer performance and visual search. *Radiology* 2011; **258**: 938–43. doi: <http://dx.doi.org/10.1148/radiol.10101090>
  16. Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 2003; **349**: 2191–200. doi: <http://dx.doi.org/10.1056/NEJMoa031618>
  17. Johnson CD, Chen MH, Toledano AY, Heiken JP, Dachman A, Kuo MD, et al. Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med* 2008; **359**: 1207–17. doi: <http://dx.doi.org/10.1056/NEJMoa0800996>
  18. Stoop EM, de Haan MC, de Wijkerslooth TR, Bossuyt PM, van Ballegooijen M, Nio CY, et al. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. *Lancet Oncol* 2011; **13**: 55–64. doi: [http://dx.doi.org/10.1016/S1470-2045\(11\)70283-2](http://dx.doi.org/10.1016/S1470-2045(11)70283-2)
  19. Atkin W, Dadswell E, Wooldrage K, Kralj-Hans I, von Wagner C, Edwards R, et al. Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. *Lancet* 2013; **381**: 1194–202. doi: [http://dx.doi.org/10.1016/S0140-6736\(12\)62186-2](http://dx.doi.org/10.1016/S0140-6736(12)62186-2)
  20. Regge D, Laudi C, Galatola G, Della Monica P, Bonelli L, Angelelli G, et al. Diagnostic accuracy of computed tomographic colonography for the detection of advanced neoplasia in individuals at increased risk of colorectal cancer. *JAMA* 2009; **301**: 2453–61. doi: <http://dx.doi.org/10.1001/jama.2009.832>
  21. Liedenbaum MH, van Rijn AF, de Vries AH, Dekker HM, Thomeer M, van Marrewijk CJ, et al. Using CT colonography as a triage technique after a positive faecal occult blood test in colorectal cancer screening. *Gut* 2009; **58**: 1242–9. doi: <http://dx.doi.org/10.1136/gut.2009.176867>
  22. Halligan S, Altman DG, Mallett S, Taylor SA, Burling D, Roddie M, Honeyfield L, McQuillan J, Amin H, Dehmshki J. Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. *Gastroenterology* 2006; **131**: 1690–9. doi: <http://dx.doi.org/10.1053/j.gastro.2006.09.051>
  23. Halligan S, Mallett S, Altman DG, McQuillan J, Proud M, Beddoe G, Honeyfield L, Taylor SA. Incremental benefit of computer-aided detection when used as a second and concurrent reader of CT colonographic data: multiobserver study. *Radiology* 2011; **258**: 469–76. doi: <http://dx.doi.org/10.1148/radiol.10100354>
  24. Phillips P, Boone D, Mallett S, Taylor SA, Altman DG, Manning D, Gale A, Halligan S. Method for tracking eye gaze during interpretation of endoluminal 3D CT colonography: technical description and proposed metrics for analysis. *Radiology* 2013; **267**: 924–31. doi: <http://dx.doi.org/10.1148/radiol.12120062>
  25. Helbren E, Halligan S, Phillips P, Boone D, Fanshawe TR, Taylor SA, Manning D, Gale A, Altman DG, Mallett S. Towards a framework for analysis of eye-tracking studies in the three dimensional environment: a study of visual search by experienced readers of endoluminal CT colonography. *Br J Radiol* 2014; **87**: 20130614. doi: <http://dx.doi.org/10.1259/bjr.20130614>
  26. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika* 1999; **86**: 948–55. doi: <http://dx.doi.org/10.1093/biomet/86.4.948>
  27. Salvucci D, Goldberg J. Identifying fixations and saccades in eye-tracking protocols. In: *Proceedings of the eye tracking research and applications symposium*. New York, NY: ACM Press; 2000. pp. 71–8.
  28. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012.
  29. Mallett S, Phillips P, Fanshawe TR, Helbren E, Boone D, Gale A, Taylor SA, Manning D, Altman DG, Halligan S. Tracking eye gaze during interpretation of endoluminal three-dimensional CT colonography: visual perception of experienced and inexperienced readers. *Radiology* 2014; **273**: 783–92. doi: <http://dx.doi.org/10.1148/radiol.14132896>
  30. Helbren E, Fanshawe TR, Phillips P, Mallett S, Boone D, Gale A, Altman DG, Taylor SA, Manning D, Halligan S. The effect of computer-aided detection markers on visual search and reader performance during concurrent reading of CT colonography. *Eur Radiol* 2015; **25**: 1570–8. doi: <http://dx.doi.org/10.1007/s00330-014-3569-z>
  31. Lau JS, Huang L. The prevalence effect is determined by past experience, not future prospects. *Vis Res* 2010; **50**: 1469–74. doi: <http://dx.doi.org/10.1016/j.visres.2010.04.020>
  32. Wolfe JM, Van Wert MJ. Varying target prevalence reveals two dissociable decision criteria in visual search. *Curr Biol* 2010; **20**: 121–4. doi: <http://dx.doi.org/10.1016/j.cub.2009.11.066>