FULL PAPER

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

[1]E HELBREN, MRCS, FRCR, [1]S HALLIGAN, FRCP, FRCR, [2]P PHILLIPS, PhD, [1]D BOONE, MRCS, FRCR, [3]T R FANSHAWE, PhD, [1]S A TAYLOR, MRCP, FRCR, [4]D MANNING, DSc, [5]A GALE, PhD, [6]D G ALTMAN, DSc and [3]S MALLETT, DPhil

[1]Centre for Medical Imaging, University College London, London, UK
[2]Health and Medical Sciences Group, University of Cumbria, Lancaster, UK
[3]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[4]School of Medicine, Lancaster University, Lancaster, UK
[5]Applied Vision Research Centre, Loughborough University, Loughborough, UK
[6]Centre for Statistics in Medicine, University of Oxford, Oxford, UK

Address correspondence to: Professor Steve Halligan
E-mail: s.halligan@ucl.ac.uk

Objective: Eye tracking in three dimensions is novel, but established descriptors derived from two-dimensional (2D) studies are not transferable. We aimed to develop metrics suitable for statistical comparison of eye-tracking data obtained from readers of three-dimensional (3D) "virtual" medical imaging, using CT colonography (CTC) as a typical example.
Methods: Ten experienced radiologists were eye tracked while observing eight 3D endoluminal CTC videos. Subsequently, we developed metrics that described their visual search patterns based on concepts derived from 2D gaze studies. Statistical methods were developed to allow analysis of the metrics.
Results: Eye tracking was possible for all readers. Visual dwell on the moving region of interest (ROI) was defined as pursuit of the moving object across multiple frames. Using this concept of pursuit, five categories of metrics

were defined that allowed characterization of reader gaze behaviour. These were time to first pursuit, identification and assessment time, pursuit duration, ROI size and pursuit frequency. Additional subcategories allowed us to further characterize visual search between readers in the test population.
Conclusion: We propose metrics for the characterization of visual search of 3D moving medical images. These metrics can be used to compare readers' visual search patterns and provide a reproducible framework for the analysis of gaze tracking in the 3D environment.
Advances in knowledge: This article describes a novel set of metrics that can be used to describe gaze behaviour when eye tracking readers during interpretation of 3D medical images. These metrics build on those established for 2D eye tracking and are applicable to increasingly common 3D medical image displays.

Eye tracking is widely used in both medical and commercial settings to assess patterns of visual search. In radiological settings, eye tracking has been used to investigate visual search patterns associated with medical image interpretation. The majority of such work has focused on two-dimensional (2D) medical image displays, such as plain radiographic film, mammography and stacked 2D CT series.[1–3] Metrics such as "time to first hit" and "dwell time" are well-established measurements used to compare the performance of different observers.[2,4–6] However, it is now increasingly common for radiologists to interpret three-dimensional (3D) medical images, considerably complicating the perceptive task. Multiplanar imaging, 3D reconstructions and endoluminal "fly-through" viewing all

demand patterns of visual search that are more complex than those associated with the interpretation of 2D displays. In particular, the 3D image is often moving and so gaze strategies are more akin to looking at a "video" than at a static image.

Methods to obtain gaze-tracking information from readers of moving 3D medical images have been described recently.[7] However, standard metrics for analysis of visual search that have been derived from static 2D images might not be applicable to new 3D display paradigms, especially where pathology is often both moving and changing in size during display. Using experienced readers to observe videos obtained from 3D endoluminal CT colonography (CTC)

examinations, we aimed to develop a range of measurements intended to allow investigation of visual search, recognition and decision making in the 3D environment. Our intention was to propose a comprehensive framework of metrics suitable for application in the 3D paradigm that builds on previous initial work.[7]

METHODS AND MATERIALS

Institutional review board approval was granted to use anonymized CTC data for eye tracking collected during two prior studies,[8,9] and to obtain eye-tracking data from volunteer readers. All readers gave informed written consent, and all data generated were anonymized.

Video preparation

Eight endoluminal fly-though videos, each of 20 s duration, were recorded from 3D CTC fly-through examinations viewed on a Viatronix® V3D colon imaging workstation (Viatronix Inc., Stony Brook, NY). Patient cases were selected from a bank of CTC studies used for previous research related to interpretation of CTC by both experienced and novice readers.[8,9] Studies included those from symptomatic and asymptomatic patients accrued from four US and three European centres. All studies had both prone and supine acquisitions following full bowel purgation and colonic insufflation. A reference truth as to the location and size of polyps on each patient case had been established previously by three radiologists experienced in CTC interpretation, in consensus and with the aid of the original radiological reports and of endoscopic correlation.

Cases were selected to obtain a subset of videos that were neither "too easy" nor "too difficult" to interpret. Consulting data from the previous reader studies, 20 cases were obtained in which a false-negative or -positive diagnosis of a polyp had been made previously by approximately 50% of experienced readers. Cases were then reviewed by a radiologist (DB) with experience of >500 endoscopically validated cases and excluded if the target polyp could not be demonstrated on either endoluminal projection or if it was within 5 s navigation of the rectal ampulla or caecal pole during endoluminal fly through. Where the polyp was visible in both prone and supine acquisitions, the least conspicuous view was selected. Five videos with true-positive polyps (with diameters of 6, 8, 11, 12 and 25 mm) were selected. Three videos with prior false-positive polyps (with diameters of 5, 7 and 10 mm) were also selected to provide true- and false-positive lesions in a 2:1 ratio.

Readers

We collected eye-tracking data from ten experienced readers who were the teaching faculty at a CTC "hands-on" workshop (ESGAR Amsterdam workshop, April 2010). "Experienced readers" were defined as radiologists who had previously interpreted >300 CTC studies independently. All readers were unaware of the prevalence of abnormality prior to viewing, and no feedback was given regarding their diagnostic performance.

Eye tracking

An infrared eye tracker (Tobii X50®; Tobii Technology, Danderyd, Sweden) was positioned beneath the viewing screen, and Studio™

capture software (Tobii Technology) was hosted on a laptop. Eye-tracking accuracy was 0.5° and 20 screen pixels at approximately 60 cm viewing distance. The video area was 512 × 512 pixels. Eye tracking was overseen by an image perception scientist (PP) with 8 years of experience.

Videos were viewed during the workshop, in a quiet area of a reporting room that was specifically designated for this study. Following a five-point calibration exercise, a "warm-up" video was used to assess the ability of the eye tracker to obtain sufficient data and to familiarize readers with the procedure. Readers wore glasses/contact lenses as per normal. Each reader held a computer mouse prior to commencing each video. The following instructions were then displayed onscreen: "You are about to be shown some short video clips of fly-throughs. Some will have pathology and some will not. Please click the mouse when you see a lesion which you consider highly likely to represent a real polyp or cancer". Readers were told to click once for each lesion.

The eight videos were then presented to readers in a randomized order. Readers took approximately 10 min to complete the set-up described above and to view all videos. Eye-tracking data and number/timing of mouse clicks were recorded for all ten readers viewing all eight videos. Readers were not required to target any polyp with the mouse; they simply clicked to indicate their belief that pathology might be present on the 3D fly through.

Data preparation and display

Following data collection, a circular region of interest (ROI) was applied around the polyp on each individual video frame where the polyp (true- and false-positive) was visible. This was then related to the readers' gaze for each individual frame by calculating the distance from the gaze point to the closest ROI boundary point for each point of gaze data acquired during the time the polyp was onscreen. Gaze points recorded within 50 pixels beyond the outer rim of the ROI were considered to have fixed upon the polyp to ensure that all gaze directed at the polyp was captured. This represented a 1.25° visual acceptance radius,[7] where the ROI boundary fell within very high visual acuity.[10] For each reader, the distance from the gaze point to the ROI boundary was plotted against time, for the duration that any polyp was onscreen, and included the identification time (mouse click), if any. A representative graph is displayed in Figure 1.

Statistical analysis

Eye pursuits were defined when within 50 pixels from the polyp ROI boundary and lasting for at least 100 ms. Missing data were imputed using multiple imputation methods[11] adapted for longitudinal data. A pursuit distance >50 pixels was characterized as a pursuit termination, provided that (to allow for measurement error) either the average pursuit distance of four to six near-contemporaneous gaze points was >50 pixels or the observed pursuit distance was more than two standard deviations of measurement error greater than the average pursuit distance within that pursuit. Gaze metrics were defined as in Figure 1 (time to first pursuit corresponding to A to B; overall assessment time A to E); pursuit time being total time within a 50-pixel distance from the ROI boundary. For each metric, data sets with

Figure 1. The graph shows the distance in pixels (*y*-axis) between individual gaze points (dots) and the region of interest (ROI) drawn around the polyp after viewing, plotted against time (milliseconds, *x*-axis). The horizontal lines indicate the 50-pixel margin to the ROI boundary. Thus, a gaze point falling within this boundary denoted the observer looking at the polyp. Consecutive data points lasting >100 ms within this margin were defined as gaze pursuits (highlighted by horizontal bars). The vertical dashed line represents the timing of any reader mouse click (observer identification of the polyp). The following labels denote key events: (A) point at which the ROI first becomes visible onscreen; (B) time at which pursuit is first recorded; (C) time at which the final pursuit immediately preceding polyp identification begins; (D) time of mouse click; and (E) time at which the ROI leaves the screen.
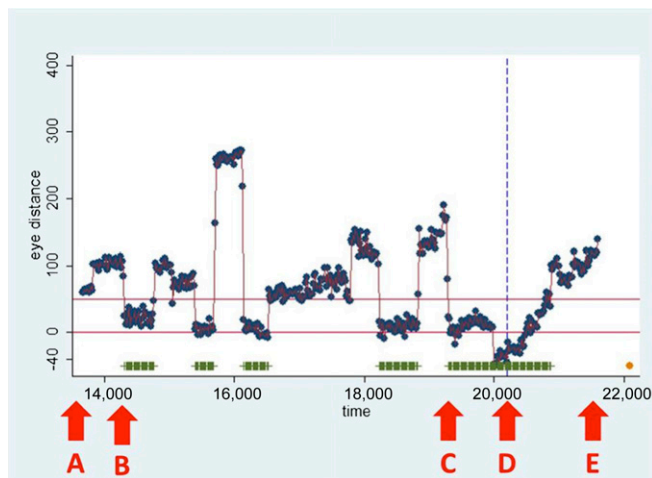
either >50% missing data or at least 1 block of 50 consecutive missing observations were examined to identify the values that would be unreliable and should be excluded from the analysis.

For metrics that measured time to an event, *e.g.* time to first pursuit, the event was censored if the event did not occur and was truncated at the time point when the event was no longer possible. For eye pursuits, the censor time was defined as occurring when the ROI was no longer visible (*i.e.* when the polyp left the screen), and at 500 ms after this for events involving time of ROI identification. Data were analysed using STATA® 12 (StataCorp, College Station, TX). Median and interquartile ranges (IQRs) were calculated to summarize percentage assessment pursuit time across readers and cases.

RESULTS
For clarity, metrics are defined by referring to one reader's eye tracking of a visible ROI in a single video (Figure 1). Additionally, Table 1 details two readers viewing the same video with corresponding gaze graphs in Figure 2; this demonstrates how search varies between readers and how the metrics reflect this.

Pursuit
We defined "pursuit" as a consecutive contiguous gaze point related to the ROI (*i.e.* within a boundary distance of 50 pixels) and lasting for 100 ms or longer.

Time to first pursuit
Since "time to first fixation" is a commonly used outcome for 2D, we defined "time to first pursuit" as the time elapsing between the first onscreen appearance of the ROI and commencement of first pursuit, if any (Figure 1). Both readers pursued the polyp soon after it became visible (at 0.42 and 0.57 s), both within 10% of the total onscreen time.

Identification and assessment time
To distinguish different components of identification and decision time, we defined the following three features and extracted them from the gaze data, in seconds and milliseconds (Table 1, Figure 1):
- *Identification time span*: time elapsing between first onscreen appearance of the ROI and the time of polyp identification (if any) by the reader (represented by the mouse click); time A to D on Figure 1.
- *Total assessment time span*: time from first fixation on the ROI (if any) to identification by the reader; time B to D on Figure 1.
- *Last assessment time span*: time from commencement of the ROI fixation immediately preceding identification to the time of identification; time C to D on Figure 1.

A reaction time of 500 ms was included to capture mouse clicks occurring very soon after the polyp left the screen.

Pursuit time
We identified two different components of pursuit time. We expressed both as a percentage of the total onscreen time during the periods described:
- *Assessment pursuit time*: the aggregated time for individual pursuits (if any) occurring prior to polyp identification; the summed length of the horizontal bars on Figure 1 prior to the mouse click.
- *Total pursuit time*: the aggregated time for individual pursuits (if any) occurring for the total onscreen time of the ROI (*i.e.* including pursuits occurring after identification); the summed length of the horizontal bars on Figure 1.

Region of interest size
We expressed the size of the ROI as a percentage of visible video area at crucial points during reader gaze as follows:
- *size at first pursuit*: B on Figure 1
- *size at longest pursuit*: C on Figure 1.

Readers pursued polyps at relatively small sizes, as a percentage screen area; 0.26% and 0.31% at the first pursuit and 0.69% and 2.38% at the longest pursuit. In our example video, the largest polyp size is 8.54%, indicating that both the first and longest pursuits were at relatively small polyp sizes. The polyp size when readers clicked was much larger, at 2.90% and 7.10%.

Pursuit frequency
We identified two different components of pursuit frequency expressed as the rate of pursuits per second. This facilitated comparison across videos where onscreen ROI time will vary.

Table 1. Metrics applicable to eye tracking of three dimensional moving studies

Metric category	Metric descriptor	Reader 1 (s) (% of total time ROI onscreen)	Reader 2 (s) (% of total time ROI onscreen)
Time to first pursuit	Time to first pursuit	0.42 (5.2%)	0.57 (7.1%)
Identification and assessment time	Identification time span	6.87 (87%)	5.99 (76%)
	Total assessment time span	6.46 (81%)	5.43 (68%)
	Last assessment time span	2.28 (29%)	1.26 (16%)
Pursuit time	Assessment pursuit time	69%	30%
	Total pursuit time	71%	43%
ROI size	Size at first pursuit	0.26%	0.31%
	Size at longest pursuit	0.69%	2.38%
Number of pursuits	Assessment pursuit rate	0.44 s^{-1}	0.67 s^{-1}
	Total pursuit rate	0.38 s^{-1}	0.5 s^{-1}

ROI, region of interest.
Metric values are presented for two readers reading the same video, corresponding to the gaze graphs in Figure 2.

- *Assessment pursuit rate*: the rate of individual pursuits (if any) occurring prior to the point of polyp identification; the number of individual horizontal bars per second on Figure 1 prior to the mouse click.
- *total pursuit rate*: the rate of individual pursuits (if any) occurring for the total onscreen time of the ROI (*i.e.* including pursuits occurring after identification); the number of individual horizontal bars per second on Figure 1.

Analysis of metrics across readers

The total time the ROI was visible varied from 2.47 to 8.87 s, with a median of 5.73 s (IQR, 3.05–7.93 s). Some metrics were highly dependent on the time the polyp was onscreen. Accordingly, some metrics were expressed as a percentage of total onscreen time. To illustrate the power of these metrics to summarize visual search patterns across readers and cases, we present results that assess pursuit time across all ten readers and eight videos: the median assessment pursuit time was 43% (IQR, 23–53%).

DISCUSSION

When viewing an image, features of interest are brought into the centre of field of view via "foveal fixation", providing the sharpest visual detail in the region of conscious attention. Multiple fixations ("spatial clustering") imply a feature of particular interest. It is relatively straightforward to record the location and duration of foveal fixations for 2D medical images.[5,12] By contrast, when images are moving, we fix and follow objects using rotational eye movements to stabilize the fovea on the target. Because both the image and the location of

Figure 2. Gaze graphs for two readers demonstrating the differing visual search characteristics that can be seen when viewing the same video. Metric values for the same readers are summarized in Table 1.
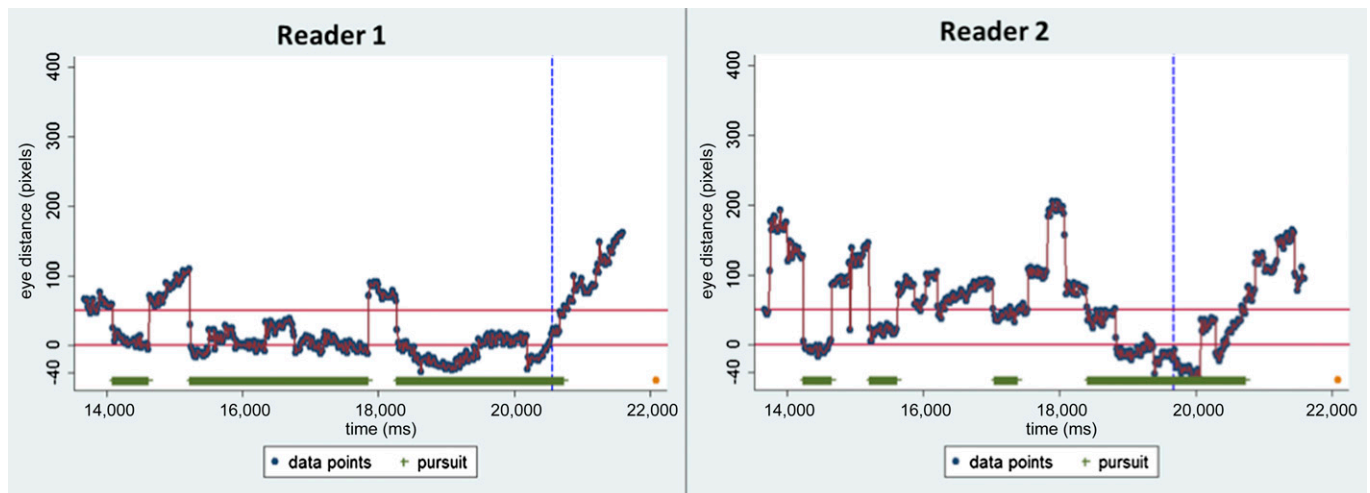
any fixed feature change frame by frame, and the nature of eye movements involved is different, the simple *x*, *y* co-ordinates and "heat maps" (for fixation duration) used to represent visual search in 2D images can no longer be applied. Using CTC as an example of the 3D medical imaging paradigm, we sought to develop a comprehensive range of metrics intended to facilitate investigation of visual search, recognition and decision making in the 3D environment, building on recent preliminary descriptions by our group.[7] By relating gaze to an ROI in terms of proximity and time, we derived a set of metrics applicable to a wide range of 3D imaging paradigms, basing these on parameters already established for 2D studies. We considered the unique qualities of the 3D environment, *e.g.* feature variation between individual frames (expressed via metrics describing the ROI size), and the time-dependent nature of the viewing task. Time pressure is irrelevant for static images, and the location and nature of background features are also constant. Care was taken so that our metrics were potentially applicable to readers of all experience and would extend to studies using different software and eye-tracking systems.

In 2D gaze tracking, a "hit" occurs when readers gaze at lesions directly for a specified minimum time period. Spatial clustering of fixation points over a static ROI is known as "dwell time" and their individual summation as "cumulative dwell". In 3D gaze tracking, readers' eyes must follow the ROI across the screen. Assessment is then reflected by time spent "pursuing" the ROI, which we propose as the 3D surrogate of 2D dwell time. We defined pursuit as when readers' uninterrupted gaze was within a moving ROI boundary for 100 ms or more. In 2D, the number of fixation clusters associated with an ROI has been shown to correlate with identification of true-positive lesions.[4] We were able to identify the number of individual pursuits in 3D and measure their individual and summated duration, with the expectation that this could examine any relationship between repeat pursuits and lesion identification in future studies. It is possible that the time-limited nature of lesion identification in 3D will enhance the importance of such metrics.

Many 2D eye-tracking studies allow readers to control the total time an image is displayed and viewed. In our study, readers could not control display time, since polyps appeared on the screen and disappeared subsequently at pre-determined points. In fly-through 3D, the observer does not change case once a lesion has been detected nor can he/she eliminate a lesion from view once it has been characterized. It is therefore desirable to separate pursuit frequency and times into those that occur before and after lesion identification. We achieved this using a mouse click. Pursuits prior to any click are probably related to lesion recognition and decision. Not all viewers will identify an abnormality as such, so we believe a metric that reflects total pursuit time when they occur, while the ROI is on-screen, is important. Alternatives to a click, such as verbal response, are possible.

Time to identify an onscreen lesion was considered to reflect both "viewing time", as for 2D studies, as well as providing insight regarding visual search. Unlike 2D, where viewing time essentially terminates interrogation, noting "decision time" in 3D allows the observer to indicate that a lesion has been identified and to then continue searching for further abnormalities present in the remaining video. If no abnormality is apparent elsewhere, the observer may return to pursue an abnormality already identified. The mouse click allowed us to separate pursuits relating to search and detection. A marker of lesion identification distinct from pursuit was also necessary, because it is possible that experienced readers may perceive abnormalities via peripheral vision, without formal pursuit. This is particularly important where moving images are concerned.

Whereas lesions remain unchanged in size and location in 2D, 3D necessitates a complex ROI that changes in size and position frame by frame. We therefore hypothesized that the ROI size (in pixels) at time of first pursuit, and immediately prior to identification, will aid understanding of 3D perception. Experienced readers might pursue and identify smaller ROIs than would novices or, alternatively, they might appreciate that resolution is maximal when a potential abnormality reaches the image foreground (*i.e.* just prior to it leaving the screen). This might delay identification time to extract maximal visual information. We were mindful that our methods should be transferable; an ROI can be created for different and complex lesion morphologies.

Our study has limitations. We investigated endoluminal fly through in automatic mode. In clinical practice, readers can adjust navigation speed and also stop to inspect potential abnormalities. A circular ROI was convenient; we adjusted the diameter to represent change in the polyp size over time of approach. However, irregular polyps and those seen in the profile are more difficult to characterize in this way. Boundary accuracy could be improved via more representative descriptions but this will increase complexity. The 50-pixel threshold was constant for all polyps, which may have entailed distant polyps being defined as "seen" too early. Possible perceptual errors would then be classified as recognition errors. A threshold based on a fixed proportion of the ROI would have the opposite effect. Future thresholds should account for all polyp sizes. We investigated only experienced readers, since our aim was simply to derive a set of metrics applicable to 3D gaze tracking; ultimately, we believe that these metrics will facilitate comparisons between experienced and inexperienced readers that might reveal factors associated with correct lesion detection that can be used to inform training schedules. For example, we have found that time to first pursuit is significantly reduced in experienced readers. We used both true- and false-positive polyps. False-positive polyps were visible abnormalities (*e.g.* residues) that had consistently been labelled as polyps by experienced readers previously. In the future, we intend to determine if there are any gaze characteristics that differentiate these from true-positive polyps. While we believe this work represents important steps towards 3D gaze tracking, further work should investigate gaze when the ROI is offscreen and for multiple simultaneous ROIs. We have also used these metrics to examine how observers' gaze is affected by the presence of computer-assisted-detection marks on the screen, a technology

used to increase polyp sensitivity for CTC. We present only one metric summarized across all readers and cases. Future work will use multilevel analyses accounting for clustering of data within readers and cases, hence allowing use of time-to-event survival analysis and count data.

In summary, we propose a comprehensive range of metrics applicable to studies of eye tracking in the 3D paradigm, where potential lesions are both moving and changing in size. We believe these metrics provide a reproducible framework to investigate 3D visual search and are potentially applicable to a wide range of research studies performed in this new, exciting environment. These metrics should facilitate identification of factors related to expertise in interpretation of 3D medical images.

REFERENCES

1. Manning DJ, Ethell S, Donovan T, Crawford T. How do radiologists do it? the influence of experience and training on searching for chest nodules. *Radiography* 2006; **12**: 134–42.

2. Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology* 2007; **242**: 396–402. doi: 10.1148/radiol.2422051997

3. Ellis SM, Hu X, Dempere-Marco L, Yang GZ, Wells AU, Hansell DM. Thin-section CT of the lungs: eye-tracking analysis of the visual approach to reading tiled and stacked display formats. *Eur J Radiol* 2006; **59**: 257–64.

4. Krupinski EA. Visual search of mammographic images: influence of lesion subtlety. *Acad Radiol* 2005; **12**: 965–9. doi: 10.1016/j.acra.2005.03.071

5. Matsumoto H, Terao Y, Yugeta A, Fukuda H, Emoto M, Furubayashi T, et al. Where do neurologists look when viewing brain CT images? an eye-tracking study involving stroke cases. *PLoS One* 2011; **6**: e28928. doi: 10.1371/journal.pone.0028928

6. Leong JJ, Nicolaou M, Emery RJ, Darzi AW, Yang GZ. Visual search behaviour in skeletal radiographs: a cross-specialty study. *Clin Radiol* 2007; **62**: 1069–77. doi: 10.1016/j.crad.2007.05.008

7. Phillips P, Boone D, Mallett S, Taylor SA, Altman DG, Manning D, et al. 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: 10.1148/radiol.12120062

8. Halligan S, Mallett S, Altman DG, McQuillan J, Proud M, Beddoe G, et al. 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: 10.1148/radiol.10100354

9. Halligan S, Altman DG, Mallett S, Taylor SA, Burling D, Roddie M, et al. Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. *Gastroenterology* 2006; **131**: 1690–9. doi: 10.1053/j.gastro.2006.09.051

10. Chakraborty D, Yoon HJ, Mello-Thoms C. Spatial localization accuracy of radiologists in free-response studies: inferring perceptual FROC curves from mark-rating data. *Acad Radiol* 2007; **14**: 4–18. doi: 10.1016/j.acra.2006.10.015

11. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011; **30**: 377–99. doi: 10.1002/sim.4067

12. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. *Acad Radiol* 1996; **3**: 137–44.