1 PROTOCOL AND QUALITY ASSURANCE FOR CAROTID IMAGING IN 100,000

2 PARTICIPANTS OF UK BIOBANK: DEVELOPMENT AND ASSESSMENT

3

4

5 Sean Coffey[1,2], Adam J Lewandowski[1], Sarah Hudson[3], Steve Garratt[3], Rudy Meijer[4], Steven

6 Lynum[5], Ram Bedi[6], James Paterson[7], Mohammad Yaqub[8], J. Alison Noble[8], Stefan

7 Neubauer[1], Steffen E Petersen[9], Naomi Allen[3,10], Cathie Sudlow[3,11], Rory Collins[3,10], Paul M.

8 Matthews[12], Paul Leeson[1]

9

10

11 Affiliations:

12 [1] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of

13 Oxford, Oxford, United Kingdom.

14 [2] Department of Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New

15 Zealand.

16 [3] UK Biobank, Cheadle, Stockport, United Kingdom.

17 [4] Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht,

18 The Netherlands.

19 [5] Panasonic Healthcare Corporation of North America, New Jersey USA

20 [6] Department of Bioengineering, University of Washington, Seattle WA USA

21 [7] Intelligent Ultrasound Ltd, Abingdon, United Kingdom.

22 [8] Institute of Biomedical Engineering, Department of Engineering Science, University of

23 Oxford, Oxford, United Kingdom.

1 [9] William Harvey Research Institute, NIHR Cardiovascular Biomedical Research Unit at

2 Barts, Queen Mary University of London, London, United Kingdom.

3 [10] Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of

4 Population Health, University of Oxford, Oxford, United Kingdom.

5 [11] Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom.

6 [12] Division of Brain Sciences, Imperial College, London, United Kingdom

7

17

18 **Address for correspondence**

19 Professor Paul Leeson, Oxford Cardiovascular Clinical Research Facility, Division of

20 Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, John

21 Radcliffe Hospital, Oxford. OX39DU. e-mail: paul.leeson@cardiov.ox.ac.uk.

22 Tel:+44(0)1865572846, Fax:+44(0)1865572840

23

1   **Word count, including references:** 3938

2

Abstract

Background - Ultrasound imaging is able to quantify carotid arterial wall structure for assessment of cerebral and cardiovascular disease risks. We describe a protocol and quality assurance process to enable carotid imaging at large scale that has been developed for the UK Biobank Imaging Enhancement Study of 100,000 individuals.

Design - An imaging protocol was developed to allow measurement of carotid intima-media thickness (CIMT) from the far wall of both common carotid arteries. Six quality assurance criteria were defined and a web-based interface (Intelligent Ultrasound) was developed to facilitate rapid assessment of images against each criterion.

Results and Conclusions – Excellent inter- and intra-observer agreements were obtained for image quality evaluations on a test dataset from 100 individuals. The image quality criteria then were applied in the UK Biobank Imaging Enhancement Study. Data from 2560 participants was evaluated. Feedback of results to the imaging team led to improvement in quality assurance (QA), with QA failures falling from 16.2% in the first two-month period examined to 6.4% in the last. 80% had all CIMT images graded as of acceptable quality, with at least one image acceptable for 98% of participants. CIMT measures showed expected associations with increasing age and gender.

Carotid imaging can be performed consistently, with semi-automated quality assurance of all scans, in a limited time frame within a large scale multimodality imaging assessment. Routine feedback of quality control metrics to operators can improve the quality of the data collection.

Abstract word count: 239

Keywords: Carotid, UK Biobank, Ultrasound, Carotid Intima Media Thickness, Carotid IMT, Quality Assurance

1 **INTRODUCTION**

2 UK Biobank is a prospective study of 500,000 people aged between 40 and 69 years at

3 recruitment between 2006 and 2010.[1] An initial feasibility assessment (pilot phase) was

4 initiated in 2014, to evaluate the feasibility of conducting a multi-modal imaging study in

5 100,000 participants over 6 years, including MRI scans of the brain, heart and body, a dual-

6 energy X-ray absorptiometry (DEXA) scan and a carotid ultrasound.[2] The large scale and the

7 richness of the phenotyping in UK Biobank allows researchers to examine aspects of

8 cardiovascular epidemiology that previously would not have been possible.

9

10 Imaging of the carotid vessels can provide information on the thickness of the carotid vessels,

11 expressed as the carotid intima media thickness (CIMT), and the presence of carotid plaque.

12 Information concerning both has been associated with incident cardiovascular events.[3–5] Use

13 of ultrasound to image the carotids has the advantage of being relatively low-cost, fast, free

14 of ionizing radiation, and requiring relatively little training to use,[6] leading to its successful

15 use in a number of large population-based cohort studies.[4,5,7,8] Good reproducibility is seen

16 even among relatively inexperienced operators under routine clinical conditions.[6] However,

17 imaging on the scale of UK Biobank is unprecedented and requires adaptation of usual

18 approaches that allow both efficient data acquisition and exceptional attention to quality

19 assurance, to avoid bias and minimise measurement variation over time. Previous

20 publications have described the UK Biobank in general,[1] and the imaging substudies

21 involving cardiac MRI,[2,9] liver MRI,[10] and brain MRI.[11] In this article, we describe the

22 carotid ultrasound scanning protocol used within the UK Biobank Imaging Enhancement

23 study and the development of a quality assurance process. We also report an evaluation of

24 this process based on carotid ultrasound scans from the first 2,560 participants.

1

METHODS

3

Protocol Development

A protocol was developed based on reported guidelines for carotid image acquisition[12,13] that

took account of the practical aspects of space for equipment, time available and the limited

experience of the technicians. In addition, the requirement for potential future analysis of raw

ultrasound data was considered in acquisition of images. The protocol was devised to operate

alongside the other planned assessments for the Imaging Enhancement of UK Biobank

(cardiovascular, brain and abdominal magnetic resonance imaging, DEXA, as well as a

repeat of the study's baseline measures, which incorporate a comprehensive computer-based

questionnaire with interview, physical measures, and blood and urine sampling).

13

The flowchart for the protocol developed is shown in Figure 1. An abbreviated version of the

UK Biobank Standard Operating Protocol is available from

http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=511. In the pilot phase, imaging was

performed solely at the UK Biobank imaging facility at Cheadle, Stockport, UK. In brief, a

CardioHealth Station (Panasonic Healthcare Corporation of North America, Newark, NJ,

USA) was used, with a 9MHz linear array transducer. The carotid study was performed in a

darkened room and participants lay supine supported by a triangular pillow. Initially, a 2D

sweep of both carotids was performed along the short-axis (transverse plane) from low in the

neck up to the jaw, at least to the level of the carotid bifurcation, with the right carotid

scanned first. The conventionally configured CardioHealth Station user interface software

allows these sweeps to be performed but not stored. To meet the needs of the UK Biobank,

this interface was modified prior to commencement of the study to allow cine image loop

1   storage. The sweep was repeated in the long-axis (longitudinal plane) and stored. These

2   longitudinal images were also used to identify the bifurcation for measurement of CIMT.

3   This was performed at two angles for each carotid, giving a total of four CIMT

4   measurements, which when measured from the vertical axis were 150° and 120° on the right

5   carotid, and 210° and 240° on the left carotid. Automated position location identification

6   provided a real-time readout of the transducer angle to the technician. The CIMT was

7   measured at time of acquisition using the semi-automated CardioHealth Station hardware and

8   software, with 'Auto ROI' and 'Auto Freeze' enabled. The CHS software places a marker on

9   the screen to guide the user in alignment of the flow divider and a 10mm region-of-interest

10   box to facilitate CIMT measurement. The far-wall of the common carotid was then

11   automatically tracked by the region-of-interest box. When three consecutive cardiac cycles

12   met internal quality thresholds, the image auto-froze in end-diastole, giving the mean,

13   maximum and minimum of the CIMT tracking. Manual override of the automatically derived

14   CIMT boundaries was performed in cases where the detected boundaries did not align with

15   the underlying B-mode image, and manual override of the auto-freeze feature was performed

16   if auto-freeze was not invoked after prolonged period of scanning. The technician also had

17   the facility to adjust the image to ensure the CIMT measurement was not taken at sites of

18   focal plaque, defined based on the Mannheim Consensus as: "a focal structure that

19   encroaches into the arterial lumen of at least 0.5mm or 50% of the surrounding IMT value or

20   demonstrates a thickness >1.5mm as measured from the media-adventitia interface to the

21   intima-lumen interface".[12]

22

23   ***Quality Assurance***

Development of Quality Criteria - To develop quality assurance (QA) processes for carotid imaging, we reviewed available literature on protocols used in previous studies and also criteria described for carotid image quality.[5,12–15] No specific guidelines for quality assurance for a study of this size (i.e. for 100,000 participants) were identified in literature review. Thus, we defined six QA criteria based on how a technician is trained to handle the probe and identify the region of interest when acquiring a carotid image, as described below:

1. **Horizontal**. The whole carotid (not just at the location of the box) lies approximately horizontal on the image, with >75% of the long axis of the carotid < 15 degrees above or below the horizontal.

2. **Oblique (**also referred to as **arterial misalignment)**. The section obtained should be along the long-axis of the carotid, with >80% of the visualised common carotid showing parallel near and far carotid walls.

3. **Box location**. The distal edge of the box for automated CIMT measurement should be positioned 1cm proximal to the flow divider of the carotid bulb, and there should be no focal plaque within the box.

4. **White-black-white**. The CIMT should have distinct borders that define a clear "white-black-white" pattern on the monitor.

5. **CIMT tracking.** The automated CIMT measurement was recorded at the correct time in the cardiac cycle (end-diastole) and > 50% of the CIMT within the measurement box has been tracked. All areas of CIMT tracked should accurately align with the B-mode image.

6. **Angle.** The acquisition angle of the image should have been recorded within 30 degrees of the selected angle category (i.e., for images selected to represent 120 degrees, the acceptable range is 90-150 degrees).

1   Based on these six quality criteria, we developed a single Pass/Fail QA flag for images. For

2   an image to pass the overall QA, an image needed to pass *all* of the individual criteria of box

3   position and CIMT tracking, but could fail *one* of horizontal, oblique or white-black-white

4   criteria. For each participant, a carotid study consisted of four images showing the CIMT

5   measurement acquired at predefined angles of acquisition (i.e., right 150°, right 120°, left

6   210°, and left 240°). A separate QA flag was generated for each angle of acqusition.

7   ***Assessment of Quality Criteria -*** To evaluate our approach, we used a test dataset of 100

8   carotid studies to estimate both inter and intra-observer reproducibility of the quality criteria,

9   after which we evaluated the process as a whole in the Pilot Phase (2567 subjects, of whom

10   2560 had a QA assessment at each angle of acquisition). To ensure consistency in application

11   of these criteria, we developed a web-based interface to allow the rapid semi-automated

12   review of images. The images were loaded into a simple custom designed grading tool

13   (Intelligent Ultrasound Ltd, Abingdon, United Kingdom), and each image was assessed twice

14   (on separate occasions) by three clinicians experienced in grading carotid ultrasound images,

15   henceford called graders. When assessing inter- and intra-observer reliability or agreement,

16   measures that adjust the raw percentage agreement by that expected due to chance are used,

17   such as Cohen's kappa. However, Cohen's kappa and its direct extensions tend to show

18   paradoxical behaviour at the extremes of prevalence (for example, as in our case, if a large

19   majority of images passed QA assessment), with very high agreement between observers

20   sometimes even leading to negative kappas.[16,17] We therefore used Gwet's first order

21   agreement coefficient, referred to as the AC1 statistic, to assess reliability. This, like Cohen's

22   kappa, is 0 in cases of no agreement and 1 in cases of perfect agreement, but incorporates

23   observer uncertainty into the model and tends to be less affected by prevalence.[17,18]

24   ***Application of Quality Assurance and Quality Audit in Pilot Phase -*** Participant age was

25   provided and analysed according to decade of birth, with age groups of $\leq$50, 51-60, 61-70,

1 and ≥71 years at the time of carotid ultrasound. Once the QA criteria were finalised, all

2 studies acquired were reviewed for quality assurance (QA) purposes by a trained UK

3 Biobank senior radiographer and information fed back to the sonographers on a weekly basis.

4 Images from each of the four acquired angles were graded as Pass/Fail, according to the

5 developed criteria and based on an overall assessment of each image. A random selection of

6 approximately 10% of these images, as well as any borderline images, were reviewed by

7 expert staff in the Oxford Cardiovascular Clinical Research Facility. Results from the

8 external review were fed back to the imaging centre on a monthly basis to complete the audit

9 cycle, with teleconferences as required to discuss any specific issues.

10 *Statistical Analysis -* At the end of the pilot phase, we used the available data on CIMT,

11 quality assurance and the age and sex of participants to investigate the overall validity of the

12 acquired carotid measurements. Specifically, we studied associations between CIMT and age

13 and sex, reasoning that measured CIMT should be higher in older age groups and in male

14 participants. We used multivariate linear regression, with predictor variables of age (as an

15 ordered categorical variable based on decade) and sex. We then assessed the impact of the

16 quality assurance process on carotid measures by studying how the model fit varied if images

17 that failed QA were excluded. We did this based on exclusion of all studies in which one

18 image failed QA. All statistical analyses were performed using R version 3.2.2.[19]

19

20

21 **RESULTS**

22 *Assessment of Quality Criteria -* 588 images from 100 participants acquired during the pilot

23 phase of the UK Biobank Imaging Enhancement were graded in the semi-automated tool

24 (Figure 2A) by three graders on two separate occasions. More than one image from each

25 angle of acquisition was eligible for assessment during this phase of the QA process. Overall,

26 there was a very good agreement between evaluations performed by different graders

1 (Supplementary Figure 1, Supplementary Table 1). All criteria showed excellent intra-

2 observer reliability (minimum Gwet AC1 88%), with good inter-observer reliability

3 (minimum Gwet AC1 78%). Examples of images passing and failing QA are shown in Figure

4 2B-G.

5 ***Application of Quality Assurance and Quality Audit -*** Images from the first 2567

6 participants after implementation of the QA protocol were evaluated. Pass/Fail QA flags for

7 all four angles of acquisition were available for 2560 participants, and, of these, 2519 (98%)

8 had at least one image passing QA. 523 (20%) studies had at least one image failing QA but

9 the proportion of images failing QA decreased over time, with an average across views of

10 16.2% failing QA in the first two month period compared to 6.4% failing in the last two

11 month period analysed (Supplementary Figure 2).

12 ***Evaluation of validity of carotid datasets***

13 Data for evaluation of associations were available for 2558 of the 2567 participants. Based on

14 the entire dataset, mean CIMT was higher in older, compared with younger, age groups and

15 was higher in men than in women (Table 1, Supplementary Figure 3). A linear regression

16 model (Table 2) showed a 61µm (95% CI: 55-67µm) increase in CIMT per decade, and a

17 48µm (95% CI: 39-57µm) increase in male compared to female participants compared to the

18 average CIMT of 553µm (95% confidence interval (95% CI): 542-564µm) in 45-50 year old

19 females in the model (adjusted R^2 0.191, $p<2x10^{-16}$). Studies with one image that failed QA

20 tended to have on average a 46µm greater CIMT (Supplementary Figure 3) and a linear

21 regression model with a QA flag (indicating any image failing QA) had slightly higher

22 adjusted R^2 than the model without the QA flag (Table 2). However, the confidence interval

23 for reference CIMT in younger females, as well as those of older age and male sex, was very

1 similar under both models, suggesting that the impact of including 'failed' images on the

2 overall assessment is small.

3

4 **DISCUSSION**

5 We describe a protocol for carotid image acquisition and semi-automated measurement of

6 carotid IMT within a limited timeframe for application within a large cohort study.

7 Furthermore, we describe development of quality assurance criteria to allow evaluation of

8 quality of images and audit control. Application of these procedures reduced the number of

9 quality fails within scans and helped to ensure that 98% of participants had at least one CIMT

10 measure. Analysis of the dataset confirms that the CIMT measures within the UK Biobank

11 Imaging Enhancement pilot phase associate with age and sex in the expected way, implying

12 face validity of the measures.

13

14 A key feature of the UK Biobank imaging protocol is that it needs to be performed within a

15 limited time frame in participants who also undergo abdominal, cardiac and brain magnetic

16 resonance imaging, dual energy X-ray absorptiometry (DEXA), blood sampling and an

17 extensive questionnaire. During the study visit, only 10 minutes is available for carotid

18 imaging, including time for patient registration. In addition to this time constraint, we expect

19 that data acquisition will be performed by multiple technicians at three separate imaging

20 assessment centres over the course of the Imaging Enhancement Study (which will take 6-7

21 years). Our protocol was therefore developed to achieve a relatively detailed image

22 acquisition in a highly reproducible manner. A recent meta-analysis highlights the variability

23 in CIMT measurements across studies.[20] The use of semi-automated measurement techniques

24 is recommended in international consensus statements to reduce inter-observer variability,[12,13]

25 and we therefore chose an ultrasound system that implements these. In addition, the

CardioHealth Station system we used has been shown to have excellent reproducibility even in less-experienced users.[6,21,22]

A potential limitation of the protocol at this time is the lack of quantification of the plaque that may be present in the images. Focal plaque has been shown to provide additional discrimination above that provided by CIMT measurement alone.[23,24] However, the additional discrimination is modest (area under the receiver operating curve 0.64 for plaque vs 0.61 for CIMT as a predictor of myocardial infarction over an average of eight years in a large meta-analysis).[24] As the assessment of plaque is much more subjective, it is both operator dependent and time consuming. Therefore it was not considered practical to provide evaluation of plaque at the time of image acquisition. However, the image acquisition protocol provides both a short axis 2D sweep along the length of the carotid as well as longitudinal views to above the bifurcation. Retrospective evaluation of stored image datasets will therefore be possible to assess plaque characteristics and other novel carotid parameters – these images will be available to external researchers. In addition, these stored raw image datasets have the potential to be used for future novel metric evaluations including, for example, assessment of geometric variation or measurements of intima media thickness in other locations within the vessel. By linking these measures to the available imaging modalities or later clinical outcomes, potentially new carotid biomarkers of cerebral or cardiovascular health could be identified.

The within and between rater reliability of our QA criteria suggest they can be applied with little inter- or intra-observer variability. A number of these criteria, such as the "White-Black-White" criterion, are integral to CIMT measurement, while others, such as the exclusion of

1   focal plaque from the measurement box, were chosen to increase reproducibility. The

2   "Angle" criterion has relatively relaxed requirements for the angle of acquisition, as we

3   decided that the exact angle of acquisition was not as important as repeated measurements of

4   the two carotids in this single time point study, and therefore should not lead to failure of the

5   entire carotid study. With the development of a standardised interface for grading the quality

6   of the images acquired and a larger number of studies acquired it may be possible to develop

7   image analysis techniques that would automate the QA process further.

8

9   The impact of inclusion or exclusion of QA criteria on the model fit was very small and

10   therefore data released by UK Biobank includes all CIMT measures irrespective of the QA

11   flag. There was a higher rate of QA fails with older age, and the provision of QA flag may

12   therefore be of scientific interest. There is a distinct possibility that future analyses will reveal

13   that images that fail QA have predictive value in themselves. For example, it is possible that

14   highly tortuous carotid arteries would be more likely to fail QA, and that the tortuosity is

15   predictive of future cardiovascular events. The QA flag for each image is therefore provided

16   along with the CIMT measure on the UK Biobank data showcase (available at

17   biobank.ctsu.ox.ac.uk/showcase).

18

19   A major advantage of the QA process is that real-time feedback is possible during training.

20   The number of images failing QA during the course of the pilot phase more than halved

21   following such feedback based on QA measures, with a rapid fall on implementation of the

22   QA audit process and maintenance of lower levels thereafter. Interestingly, both left sided

23   angles of acquisition had a higher failure rate initially, and appeared to have a more rapid

24   improvement compared to right sided angles of acquisition. As we did not record handedness

1   of operators, we were unable to investigate this further. The audit process involved feedback

2   to the technicians of both number of image fails and reasons for image quality failures.

3   Furthermore, as each image could be identified by time and operator, the advice could be

4   personalised for individual technicians to improve their carotid imaging technique. Several

5   operators performed carotid imaging during the pilot phase and new operators started during

6   that period. The QA criteria were introduced as part of the training for new technicians and

7   each was required to show competency against the QA criteria in test cases before being

8   allowed to scan participants. Therefore the maintenance of a low failure rate provides

9   reassurance that the use of the QA criteria within training processes was also robust.

10   In summary, carotid images and semi-automated CIMT measures are now available from the

11   UK Biobank Imaging Enhancement Study. This is the first published description of the

12   protocol used for acquisition of these images and the rationale for the design. Furthermore,

13   we provide a detailed evaluation of the quality assurance process, description of a semi-

14   automated online tool for rapid image evaluation and demonstrate that the data generated fits

15   expected patterns of association with age and gender.

16

1

Conflicts of interest

SL is an employee of, and RB and RM are consultants to Panasonic Corporation. JP is an

employee of, and AN, MY and PL are consultants to Intelligent Ultrasound Ltd. SH, SG, NA,

CS, RC, PM, and PL are employees of or have received support from UK Biobank.

REFERENCES

1. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* 2015; 12: e1001779.

2. Petersen SE, Matthews PM, Bamberg F, et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *J Cardiovasc Magn Reson* 2013; 15: 46.

3. Naqvi TZ, Lee M-S. Carotid Intima-Media Thickness and Plaque in Cardiovascular Risk Assessment. *JACC Cardiovasc Imaging* 2014; 7: 1025–1038.

4. Polak JF, Pencina MJ, Pencina KM, et al. Carotid-wall intima-media thickness and cardiovascular events. *N Engl J Med* 2011; 365: 213–21.

5. Nambi V, Chambless L, Folsom AR, et al. Carotid intima-media thickness and presence or absence of plaque improves prediction of coronary heart disease risk: the ARIC (Atherosclerosis Risk In Communities) study. *J Am Coll Cardiol* 2010; 55: 1600–7.

6. Vanoli D, Wiklund U, Lindqvist P, et al. Successful novice's training in obtaining accurate assessment of carotid IMT using an automated ultrasound system. *Eur Heart J Cardiovasc Imaging* 2014; 15: 637–42.

7. Herder M, Johnsen SH, Arntzen K a., et al. Risk Factors for Progression of Carotid Intima-Media Thickness and Total Plaque Area: A 13-Year Follow-Up Study: The Tromsø Study. *Stroke* 2012; 43: 1818–1823.

8. Clarke R, Du H, Kurmi O, et al. Burden of carotid artery atherosclerosis in Chinese adults: Implications for future risk of cardiovascular diseases. *Eur J Prev Cardiol* 2017; 24: 647–656.

9. Petersen SE, Matthews PM, Francis JM, et al. UK Biobank's cardiovascular magnetic resonance protocol. *J Cardiovasc Magn Reson* 2015; 18: 8.

10. Wilman HR, Kelly M, Garratt S, et al. Characterisation of liver fat in the UK Biobank cohort. *PLoS One* 2017; 12: e0172921.

11. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 2016; 19: 1523–1536.

12. Touboul P-J, Hennerici MG, Meairs S, et al. Mannheim Carotid Intima-Media Thickness and Plaque Consensus (2004-2006-2011). *Cerebrovasc Dis* 2012; 34: 290–296.

13. Stein JH, Korcarz CE, Hurst RT, et al. Use of Carotid Ultrasound to Identify Subclinical Vascular Disease and Evaluate Cardiovascular Disease Risk: A Consensus Statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force Endorsed by the Society for Vascular. *J Am Soc Echocardiogr* 2008; 21: 93–111.

14. Bots ML, den Ruijter HM. Should We Indeed Measure Carotid Intima-Media Thickness for Improving Prediction of Cardiovascular Events After IMPROVE? *J Am Coll Cardiol* 2012; 60: 1500–1502.

17

15. Nambi V, Chambless L, He M, et al. Common carotid artery intima-media thickness is as good as carotid intima-media thickness of all carotid artery segments in improving prediction of coronary heart disease risk in the Atherosclerosis Risk in Communities (ARIC) study. *Eur Heart J* 2012; 33: 183–90.

16. Feinstein AR, Cicchetti D V. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43: 543–9.

17. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol* 2008; 61: 29–48.

18. Feng GC. Factors affecting intercoder reliability: a Monte Carlo experiment. *Qual Quant* 2013; 47: 2959–2982.

19. R Core Team. R: A Language and Environment for Statistical Computinghttps://www.r-project.org/ (2016).

20. Liao X, Norata GD, Polak JF, et al. Normative values for carotid intima media thickness and its progression: Are they transferrable outside of their cohort of origin? *Eur J Prev Cardiol* 2016; 23: 1165–1173.

21. Singh S, Nagra A, Maheshwari P, et al. Rapid Screening for Subclinical Atherosclerosis by Carotid Ultrasound Examination: The HAPPY (Heart Attack Prevention Program for You) Substudy. *Glob Heart* 2013; 8: 83–9.

22. Vanoli D, Lindqvist P, Wiklund U, et al. Fully automated on-screen carotid intima-media thickness measurement: a screening tool for subclinical atherosclerosis. *J Clin Ultrasound* 2013; 41: 333–9.

23. Baldassarre D, Hamsten A, Veglia F, et al. Measurements of Carotid Intima-Media Thickness and of Interadventitia Common Carotid Diameter Improve Prediction of Cardiovascular Events. *J Am Coll Cardiol* 2012; 60: 1489–1499.

24. Inaba Y, Chen JA, Bergmann SR. Carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: A meta-analysis. *Atherosclerosis* 2012; 220: 128–133.

1 **Table 1. Carotid intima-media thickness by age.**

2

3

Age (years)	n	Overall mean (sd)	Female n (%)	Female mean (sd)	Male mean (sd)	Poor QA n (%)
≤ 50	170	588 (84)	95 (56%)	578 (76)	600 (93)	23 (14%)
51-60	787	628 (107)	451 (57%)	607 (83)	657 (126)	119 (15%)
61-70	1180	701 (123)	596 (51%)	675 (104)	728 (134)	238 (20%)
≥ 71	421	763 (145)	175 (42%)	739 (126)	781 (155)	143 (34%)

4

5

6 "Poor QA" refers to studies where at least one of the four angles of acquisition was deemed to

7 not pass quality assurance criteria. CIMT is measured in micrometres. Abbreviations: CIMT,

8 carotid intima-media thickness; sd, standard deviation.

9

Table 2. Linear regression, with carotid intima media thickness as outcome variable.

	Basic model	Incorporating QA
n	2558	2558
Adjusted R^2	0.19	0.21
Average CIMT for ≤ 50 year old females	553 (542 - 564)	551 (539 - 562)
Increase per decade	61 (55 - 67)	58 (52 - 63)
Increase in males	48 (39 - 57)	45 (36 - 54)
Increase if study had any image failing QA		46 (35 - 57)

CIMT is measured in µm, with the 95% confidence interval shown in brackets. The average

CIMT for ≤ 50 year old females is based on the linear regression model intercept. Abbreviations:

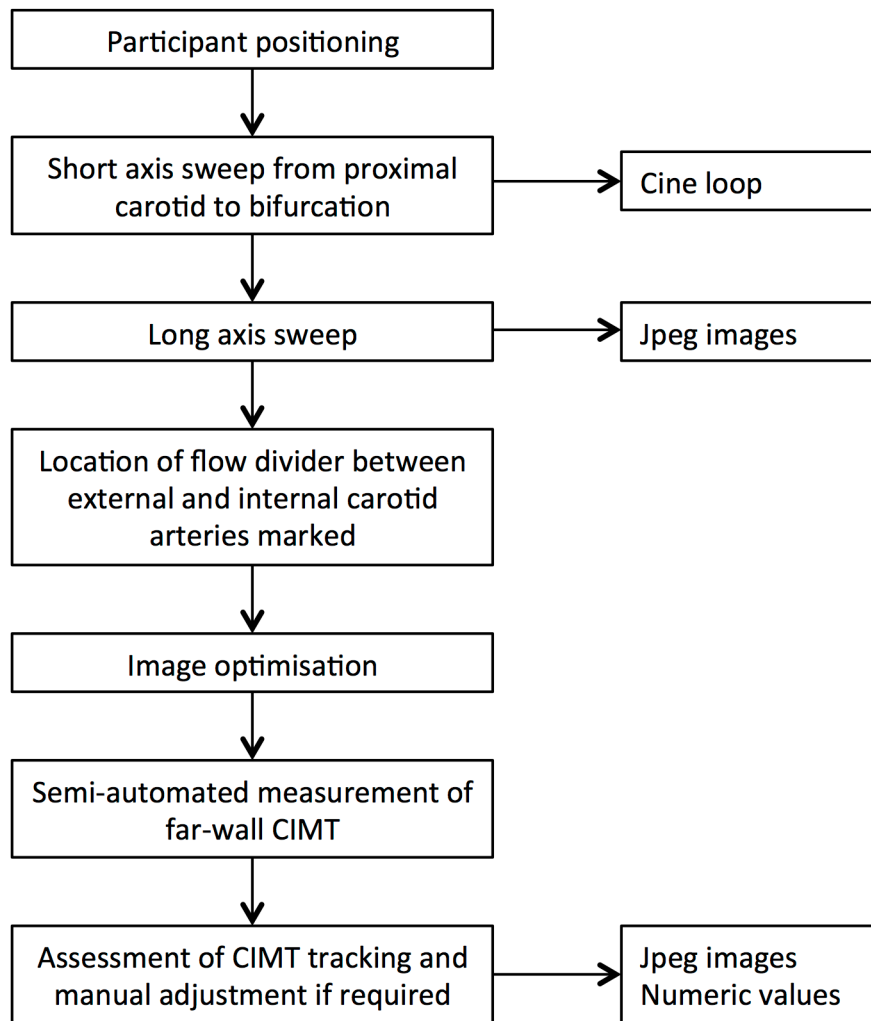CIMT, carotid intima media thickness; QA, quality assurance.

**Figure 1. Overview of UK Biobank carotid ultrasound protocol.**

Outms from the protocol are shown on the right of the figure. Four separate sets of CIMT

measurements were obtained: right carotid at 150°, right carotid at 120°, left carotid at 210°, and

left carotid at 240°. Numeric values stored were maximum, mean, and minimum CIMT.

Abbreviations: CIMT, carotid intima-media thickness.

**Figure 2. Quality assurance grading process.**
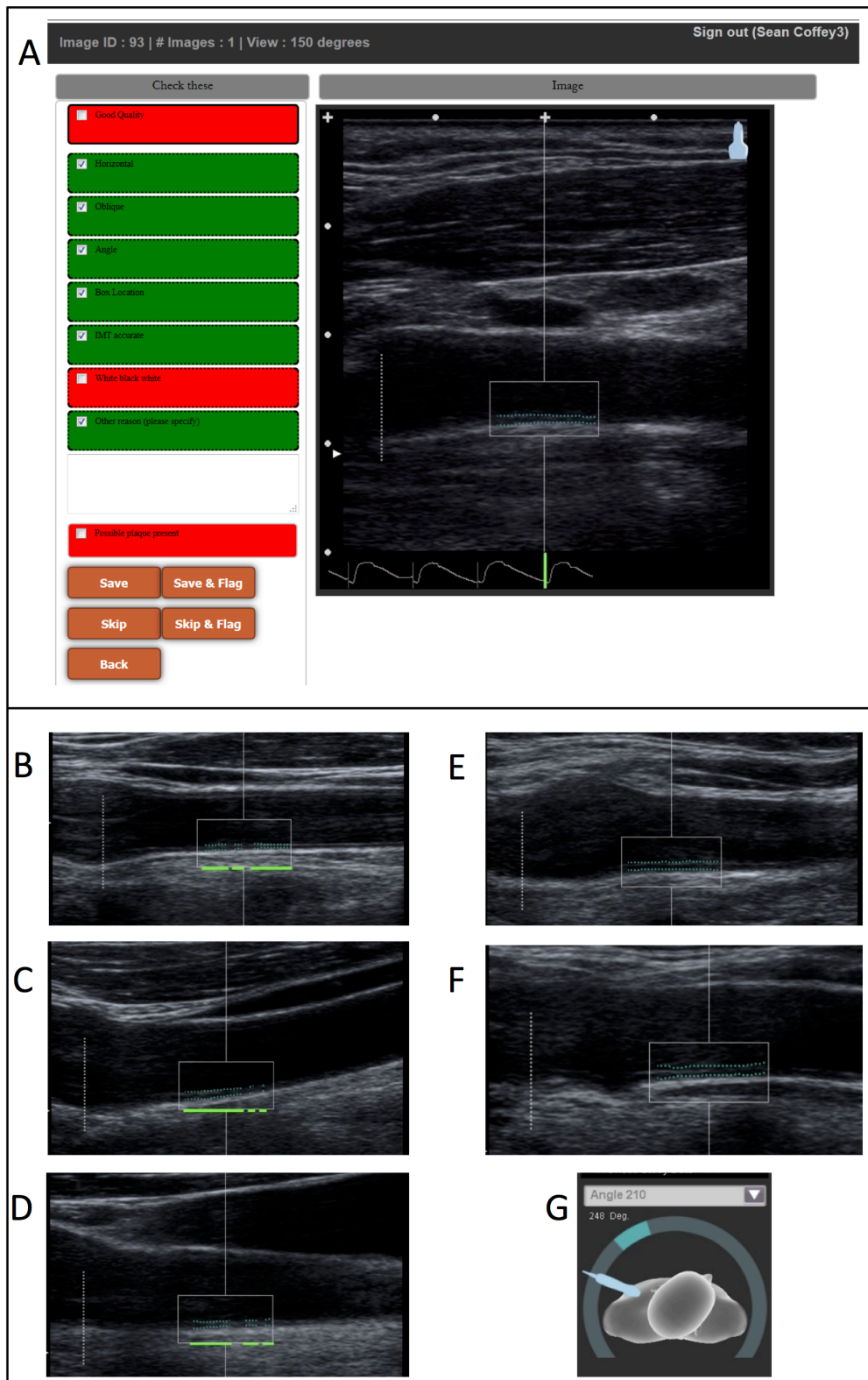
2

3   Figure showing the grading interface (A) and example images (lower panel). The grading

4   interface shows an image not meeting the "White-black-white" quality assurance (QA) criterion

5   (indicated by the lack of a tick in this box). The example images show a good quality image (B),

1   and images failing quality assurance criteria: Horizontal (C), Oblique (D), Box location (E), CIMT

2   tracking (F), and Angle (G).   Assessment of the Angle criterion was facilitated by the use of an

3   automated transducer position identifier, as shown in 2G.

Supplementary Material for Coffey et al, "Protocol and Quality Assurance for Carotid Imaging in 100,000 Participants of UK Biobank: Development and Assessment"

Supplementary Figure 1. Variability of quality assurance criteria.
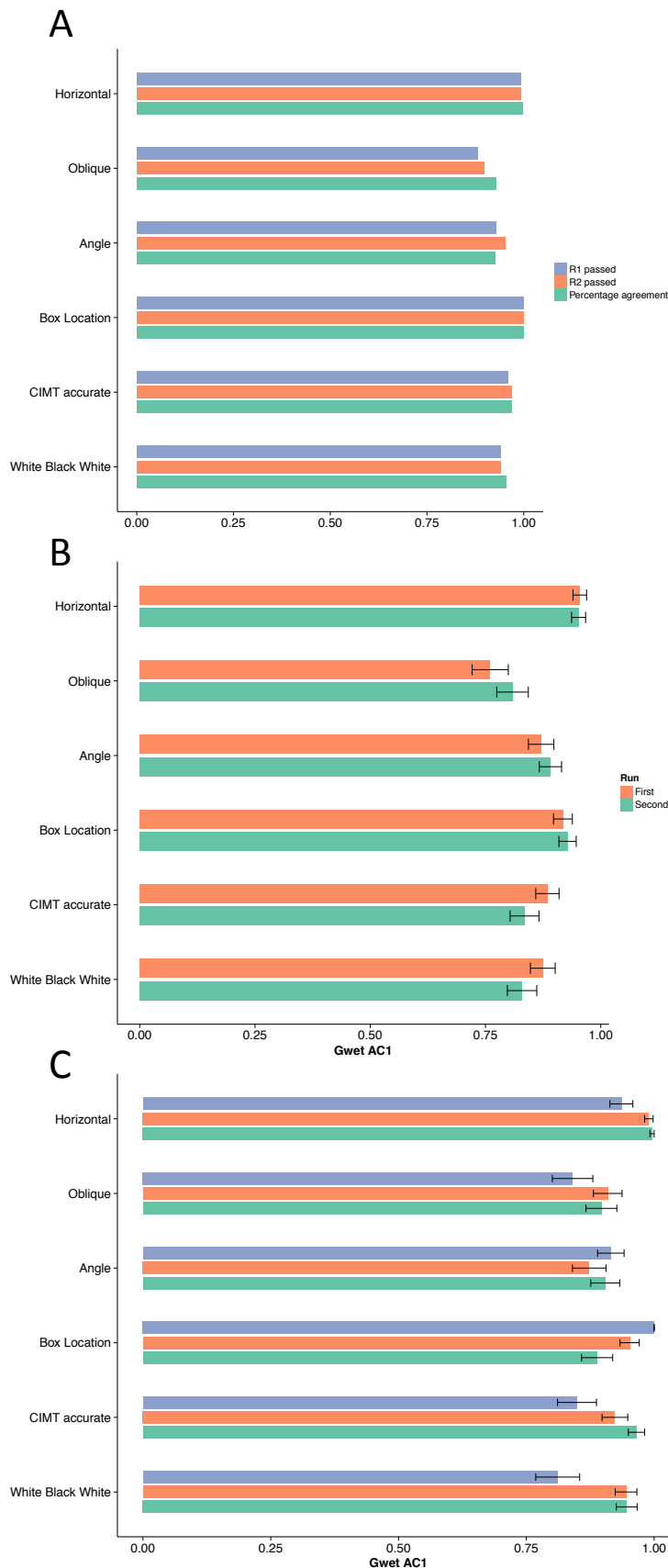

(A) Percentage of images meeting quality assurance criterion during first run (R1), second run (R2), and the percentage agreement between runs.

(B) Interobserver agreement measured by Gwet AC1 statistic for first and second runs.

(C) Intraobserver agreement measured by Gwet AC1 statistic, with each colour representing each of the three different graders. Error bars represent 95% confidence intervals.

Supplementary Figure 2. Images failing quality assurance over time.
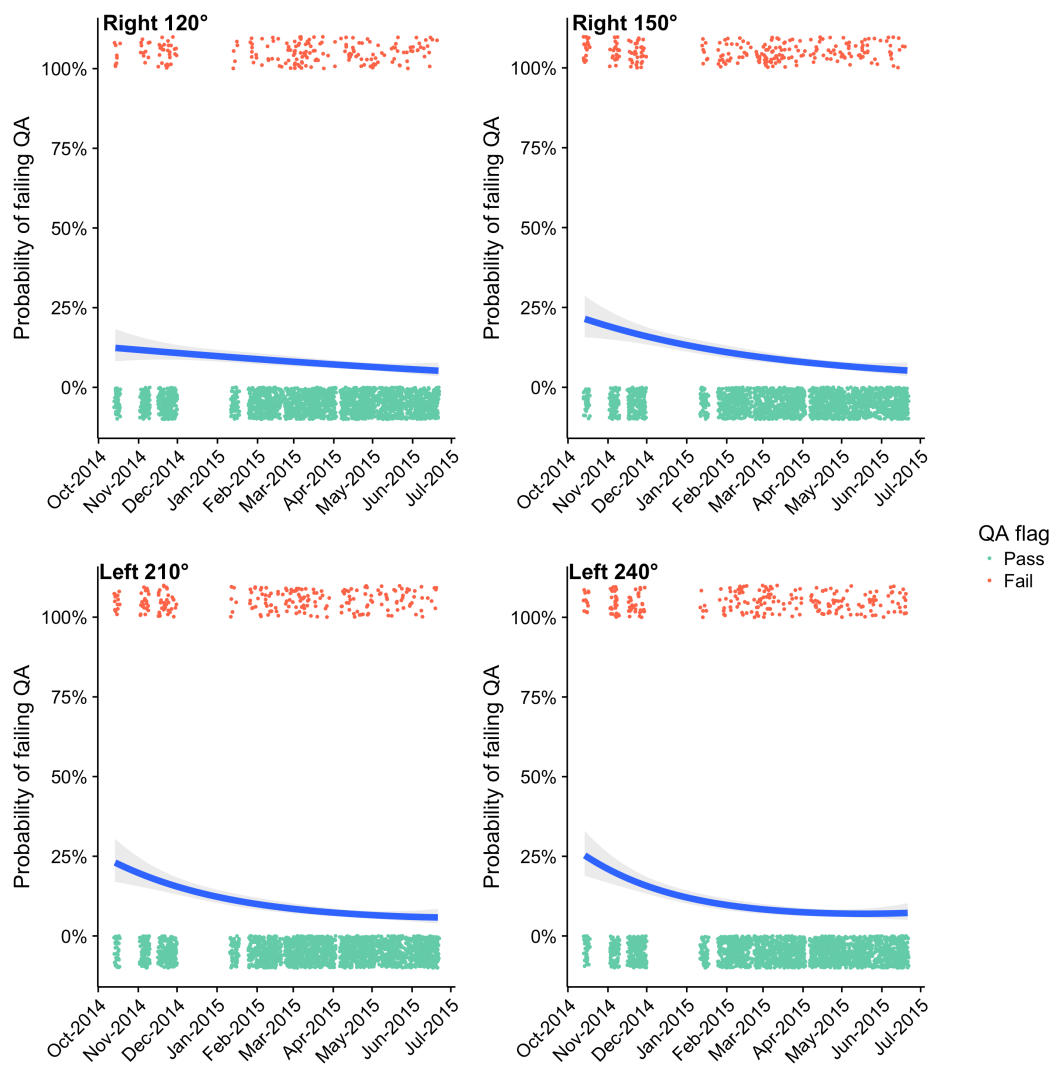
Curves were fitted with a polynomial logistic model, with shaded area representing the 95% confidence interval. Images failing QA are shown as dots at the top of each graph, while those passing are shown at the bottom. Increasing numbers of participants led to narrower confidence intervals over time. Abbreviations: QA, quality assurance.

Supplementary Figure 3. Carotid intima media thickness, shown by age, sex, and QA status.
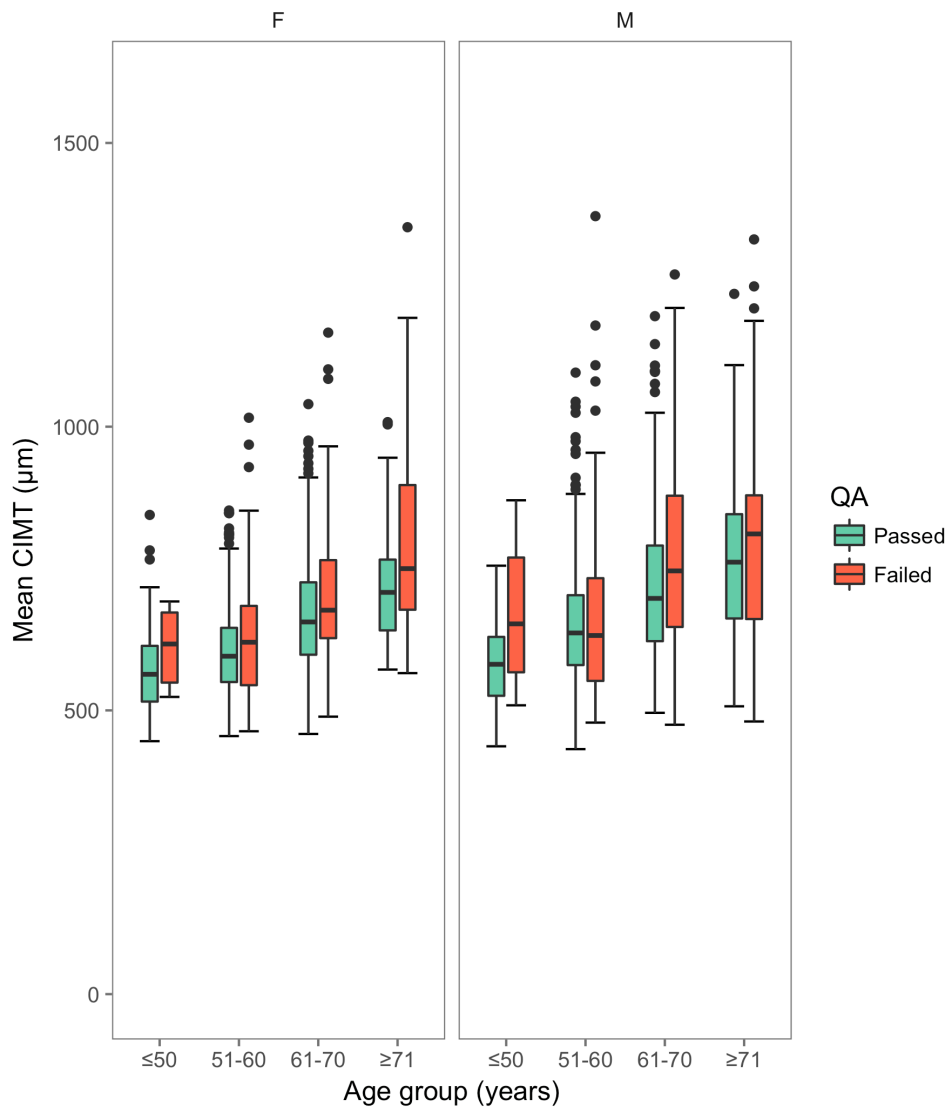
QA failed refers to studies where at least one of the four angles of acquisition was deemed to not pass quality assurance criteria. Abbreviations: CIMT, carotid intima media thickness; F, female; M, male; QA, quality assurance.

Supplementary Table 1. Summary of inter-observer agreement according to
quality metric.

Quality criteria	Percentage agreement	Inter-observer Gwet AC1 (95% CI)	Intra-observer Gwet AC1 (95% CI)
White Black White	88.0%	85.2% (82.2% - 88.1%)	90.1% (87.3% - 92.9%)
CIMT accurate	88.4%	85.9% (83.1% - 88.8%)	91.3% (88.6% - 93.9%)
Box Location	92.9%	92.3% (90.3% - 94.3%)	94.7% (93% - 96.3%)
Angle	89.8%	88.1% (85.5% - 90.7%)	89.8% (86.8% - 92.7%)
Oblique	83.5%	78.4% (74.8% - 82.1%)	88.2% (85% - 91.5%)
Horizontal	95.6%	95.3% (93.8% - 96.8%)	97.4% (96.2% - 98.5%)

The inter-observer and intra-observer Gwet AC1 statistic is the mean value
across quality criteria assessment runs and observers, respectively.
Abbreviation: CI, confidence interval.