Automated CT biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study

Perry J Pickhardt, Peter M Graffy, Ryan Zea, Scott J Lee, Jiamin Liu, Veit Sandfort, Ronald M Summers

Summary

Background Body CT scans are frequently done for a wide range of clinical indications, but potentially valuable biometric information typically goes unused. We aimed to compare the prognostic ability of automated CT-based body composition biomarkers derived from previously developed deep-learning and feature-based algorithms with that of clinical parameters (Framingham risk score [FRS] and body-mass index [BMI]) for predicting major cardiovascular events and overall survival in an adult screening cohort.

Methods In this retrospective cohort study, mature and fully automated CT-based algorithms with predefined metrics for quantifying aortic calcification, muscle density, ratio of visceral to subcutaneous fat, liver fat, and bone mineral density were applied to a generally healthy asymptomatic outpatient cohort of adults aged 18 years or older undergoing abdominal CT for routine colorectal cancer screening. To assess the association between the predictive measures (CT-based *vs* FRS and BMI) and downstream adverse events (death or myocardial infarction, cerebrovascular accident, or congestive heart failure subsequent to CT scanning), we used both an event-free survival analysis and logistic regression to compute receiver operating characteristic curves (ROCs) .

Findings 9223 people (mean age 57·1 years [SD 7·8]; 5152 [56%] women and 4071 [44%] men) who underwent CT scans between April, 2004, and December, 2016, were included in this analysis. In the longitudinal clinical follow-up (median 8·8 years [IQR 5·1–11·6]), subsequent major cardiovascular events or death occurred in 1831 (20%) patients. Significant differences were observed for all five automated CT-based body composition measures according to adverse events (p<0·001). Univariate 5-year area under the ROC (AUROC) values for predicting death were 0·743 (95% CI 0·705–0·780) for aortic calcification, 0·721 (0·683–0·759) for muscle density, 0·661 (0·625–0·697) for ratio of visceral to subcutaneous fat, 0·619 (0·582–0·656) for liver density, and 0·646 (0·603–0·688) for vertebral density, compared with 0·499 (0·454–0·544) for BMI and 0·688 (0·650–0·727) for FRS. Univariate hazard ratios for highest-risk quartile versus others for these same CT measures were 4·53 (95% CI 3·82–5·37) for aortic calcification, 3·58 (3·02–4·23) for muscle density, 2·28 (1·92–2·71) for the ratio of visceral to subcutaneous fat, 1·82 (1·52–2·17) for liver density, and 2·73 (2·31–3·23) for vertebral density, compared with 1·36 (1·13–1·64) for BMI and 2·82 (2·36–3·37) for FRS. Multivariate combinations of CT biomarkers further improved prediction over clinical parameters (p<0·05 for AUROCs). For example, the 2-year AUROC from combining aortic calcification, muscle density, and liver density for predicting death was 0·811 (95% CI 0·761–0·860).

Interpretation Fully automated quantitative tissue biomarkers derived from CT scans can outperform established clinical parameters for presymptomatic risk stratification for future serious adverse events and add opportunistic value to CT scans performed for other indications.

Funding Intramural Research Program of the National Institutes of Health Clinical Center.

Lancet Digital Health 2020

Published Online
March 2, 2020
https://doi.org/10.1016/S2589-7500(20)30025-X

See Online/Comment
https://doi.org/10.1016/S2589-7500(20)30061-3

Department of Radiology, The University of Wisconsin School of Medicine & Public Health, Madison, WI, USA (Prof P J Pickhardt MD, P M Graffy MPH, R Zea MS, S J Lee MD)**; and Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD, USA** (J Liu PhD, V Sandfort MD, Prof R M Summers MD)

Correspondence to:
Prof Perry J Pickhardt, Department of Radiology, University of Wisconsin School of Medicine & Public Health, Madison, WI 53792-3252, USA
ppickhardt2@uwhealth.org

Introduction

There has been substantial and growing interest in applying artificial intelligence (AI) to medicine by use of various machine-learning and deep-learning algorithms.[1] Along with other big data challenges, diagnostic imaging has been identified as a logical early target.[2–4] In particular, body CT represents an ideal modality with vast potential because these scans are used widely and contain additional robust, objective volumetric data that are highly reproducible and consistent across patients. In fact, opportunistic use of CT data beyond the clinical indication has already shown value from a range of manual and semi-automated approaches, most notably with incidental osteoporosis screening.[5–7] Beyond bone mineral density (BMD) information, every abdominal CT scan contains additional rich body composition data that can be objectively measured, including vascular calcification, muscle mass and density, visceral and

Research in context

Evidence before this study

We searched MEDLINE for research articles published in English between Jan 1, 1950, and Aug 31, 2019, using search terms including "CT", "prediction", "outcomes", "opportunistic screening", "body composition", and "artificial intelligence". There is a robust literature on how certain objective measures derived from abdominal CT scans can provide useful health information beyond the specific clinical indication for scanning. We and others have previously shown that manual opportunistic measures of aortic calcification, abdominal musculature, visceral fat, liver fat, and bone mineral density can help to stratify patient risk in terms of future adverse cardiometabolic events, including death. In some cases, these manual CT-based measures outperformed established clinical predictive tools. We have also recently demonstrated that these CT-based biometric measures can all be fully automated using artificial intelligence techniques to allow for objective, large-scale investigation of large patient cohorts.

Added value of this study

This is the first study, to our knowledge, to apply a battery of validated, fully automated CT biomarkers to a large screening cohort of asymptomatic adults with long-term clinical follow-up to assess their ability to predict future adverse clinical events, such as myocardial infarction, stroke, and death. Predictive ability of these CT biomarkers was compared with the well established Framingham risk score (FRS) and body-mass index (BMI). We found that the automated CT-based prediction was overall superior to the FRS and BMI. Some univariate CT measures outperformed the multivariate FRS, with further improvement in CT-based prediction when combining biomarkers. These CT biomarkers are typically ignored in current clinical practice, but this tissue-based information resides in all CT scans, regardless of clinical indication for imaging.

Implications of all the available evidence

Our study shows the rich prognostic value that can be automatically derived from abdominal CT scans, incidental to the indication for imaging. Given the many millions of CT scans performed each year in many countries, harnessing these valuable data could identify many presymptomatic patients who are at high risk of future serious adverse events, potentially allowing for earlier intervention and prevention.

subcutaneous fat, and liver fat content.[8–13] If properly leveraged, these additional opportunistic data could further augment the value of CT scans for the benefit of patients by potentially providing risk stratification for future adverse events and overall mortality. Of note, a recent report has emphasised the scarcity of such prevention research among studies supported by the US National Institutes of Health (NIH).[14] Importantly, these body composition data are freely available on essentially any abdominal CT scan, regardless of the initial clinical indication for imaging.

We have previously developed, trained, tested, and validated fully automated algorithms for measuring body composition at abdominal CT, including quantification of aortic calcification, muscle density, visceral and subcutaneous fat, liver fat, and BMD.[15–19] These CT-based biomarkers might hold potential for identifying people at an increased risk of a range of adverse clinical outcomes. With all of the AI learning steps complete and the specific automated tool outputs already preselected, our next logical step was to apply these mature predefined static tools to an external cohort. To this end, we had access to a unique external screening cohort of generally healthy asymptomatic adults who underwent abdominal CT for the purpose of colorectal cancer prevention and screening, using the CT colonography (CTC) technique.[20] Through longitudinal follow-up, we have identified subsequent defined adverse events in this patient cohort, including heart attack, stroke, and death. The main purpose of this study was to investigate the prognostic ability of a fully automated, predefined panel of CT-based body composition biomarkers compared with that of clinical parameters for presymptomatic prediction of future cardiovascular events and overall survival in a healthy adult screening cohort.

Methods

Patient cohort and CT protocol

In this retrospective cohort study, we used the deep-learning and image-processing algorithms that were previously developed, trained, tested, and validated at the NIH Clinical Center in separate CT cohorts.[21] Previous validation focused mainly on technical performance and optimal measurement parameters. These CT-based algorithms include automatically segmenting and quantifying the spine, aortic calcium, abdominal musculature, visceral and subcutaneous fat, and liver. Adults aged 18 years or older who were generally healthy consecutive asymptomatic outpatients, undergoing low-dose unenhanced abdominal CT for colorectal cancer screening (as part of routine health maintenance) at University of Wisconsin Hospital and Clinics (Madison, WI, USA) were included. Individuals with inadequate follow-up (<1 year in the absence of an adverse event) were excluded. The low-dose, non-contrast supine multidetector CT scans used for this investigation were all performed at 120 kV_p using a single vendor (GE Healthcare, Waukesha, WI, USA), with modulated mA to achieve a noise index of 50, typically resulting in an effective dose of 2–3 mSv. The specific additional CTC-related techniques for bowel preparation and colonic distention have been previously described[20,22] and are beyond the scope of this investigation.

This investigation was Health Insurance Portability and Accountability Act compliant and approved by the Institutional Review Board at the University of Wisconsin and the Office of Human Subjects Research Protection at the NIH Clinical Center (Bethesda, MD, USA). The requirement for signed informed consent was waived.

Automated CT biomarkers

Both the preliminary work and this culminating predictive trial made use of the high-performance computing capabilities of the Biowulf system at the NIH. The specific AI methodology for these automated CT-based anatomical tissue segmentation and quantification tools have been previously described elsewhere[15–19,23–29] (see appendix pp 6–7 for additional methodology details). Briefly, these tools fall into two main categories: a deep-learning group and a feature-based, image-processing group. Deep-learning algorithms were used to segment and analyse the entire liver, the abdominal wall musculature, and calcified atherosclerotic aortic plaque. These models consisted of a modified 3D U-Net for segmentation of liver and muscle, and the Mask RCNN algorithm for segmentation of aortic calcium. For bone and fat quantification, we used feature-based, image-processing algorithms, starting with fully automated spine segmentation and labelling software to identify each vertebral level from T12 to L5. This step was followed by isolation of the anterior trabecular space of each vertebra for BMD assessment, as well as the visceral and subcutaneous fat compartments at each level. Because these validated CT-based tools were used herein in a static manner whereby no additional learning was used, the need for additional training, testing, or cross-validation was obviated.

Preliminary work using the CT screening cohort in this study was done for each automated CT tool to establish normative values and success and failure rates, and to narrow down each tool to a single, stable quantitative measure for each tissue composition, without additional learning or adjustment.[15–19] Each tissue measure can be reported in various ways. For example, CT attenuation numbers measured in Hounsfield units (HU) reflect mean tissue density. Tissue bulk can be expressed according to cross-sectional area at specified levels or by volume. The final selected static measure for each of the five body composition areas (figure 1) was chosen according to our preliminary investigations to optimise overall success, and included: the visceral-to-subcutaneous fat ratio at the L1 level, mean muscle density (in HU) at the L3 level, volumetric liver density (in HU), aortic calcification between the L1 and L4 vertebral levels (quantified by an Agatston score), and trabecular BMD at the L1 level (in HU). The technical failure rates for these tools were all on the order of 1% or less. Figure 1 depicts visual correlates of the quantitative output for the automated CT tools. This final panel of biomarkers was derived from CT scans in this study cohort in a fully automated manner.

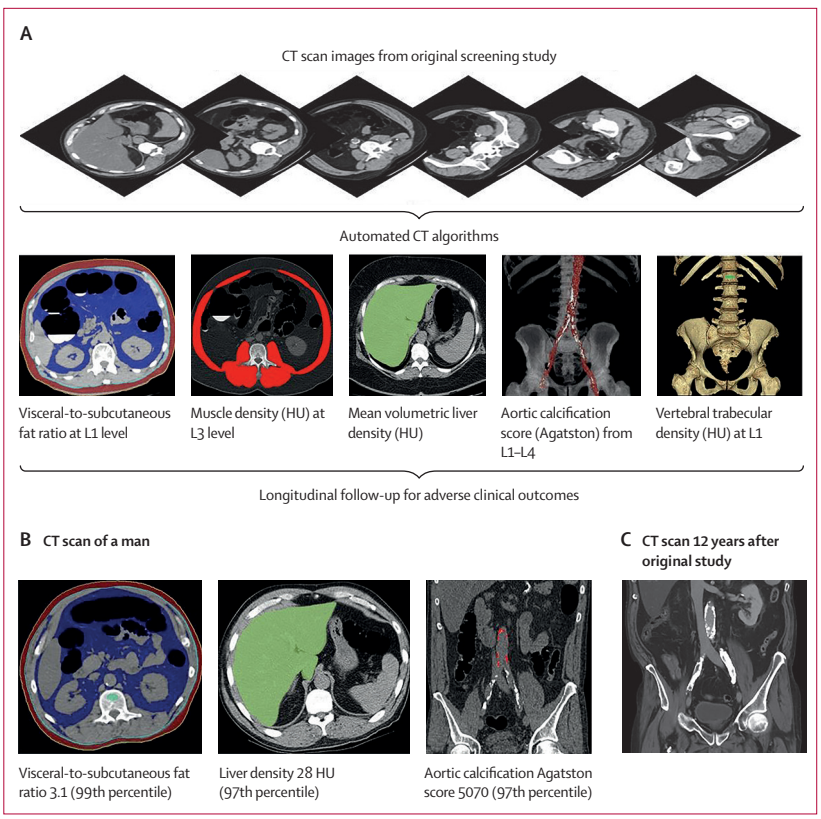

Figure 1: **Depiction of the fully automated CT biomarkers tools used in this study**
(A) Schematic depiction of the automated process for assessing fat, muscle, liver, aortic calcification, and bone from original abdominal CT scan data. (B) Case example in an asymptomatic 52-year-old man undergoing CT for colorectal cancer screening. At the time of CT screening, he had a body-mass index of 27·3 and Framingham risk score of 5% (low risk). However, several CT-based metabolic markers were indicative of underlying disease. Multivariate Cox model prediction based on these three CT-based results put the risk of cardiovascular event at 19% within 2 years, at 40% within 5 years, and at 67% within 10 years, and the risk of death at 4% within 2 years, 11% within 5 years, and 27% within 10 years. At longitudinal clinical follow-up, the patient suffered an acute myocardial infarction 3 years after this initial CT and died 12 years after CT at the age of 64 years. (C) Contrast-enhanced CT performed 7 months before death for minor trauma was interpreted as negative but does show significant progression of vascular calcification, visceral fat, and hepatic steatosis. HU=Hounsfield units.

Clinical parameters and adverse outcomes

See **Online** for appendix

Beyond patient age and sex, the main clinical parameters that we considered were body-mass index (BMI), defined as weight (kg) divided by the square of height (m^2), and the data inputs necessary for the Framingham risk score (FRS). The FRS for assessing risk for cardiovascular disease is a well established, validated multivariate algorithm combining the factors of age, sex, blood pressure, cholesterol, lipids, diabetes status, and smoking.[30] Data points closest to the timing of the CT scan were included.

Adverse clinical outcomes were defined by either patient death or major cardiovascular events subsequent to CT scanning—myocardial infarction, cerebrovascular accident, or development of congestive heart failure—to reflect the endpoints considered by the FRS for cardiovascular disease. We constructed a broad algorithmic electronic health record search for the relevant clinical data points and the defined clinical events.

	Total cohort (n=9223)	Cardiovascular event* or death		Death	
		Yes (n=1831)	No (n=7392)	Yes (n=549)	No (n=8674)
Aortic calcification, Agatston score					
Mean (SD)	699 (1748)	1628 (2802)	469 (1265)	2471 (3572)	587 (1494)
Median (IQR)	59 (0–493)	449 (38–1945)	31 (0–313)	873 (131–3541)	48 (0–428)
Muscle density, HU					
Mean (SD)	28·9 (12·1)	25·4 (13·7)	29·8 (11·5)	20·8 (15·3)	29·4 (11·7)
Median (IQR)	31 (22–38)	27 (17–35)	31 (23–38)	22 (12–31)	31 (23–38)
Visceral-to-subcutaneous fat ratio					
Mean (SD)	0·91 (0·71)	1·13 (0·85)	0·58 (0·66)	1·22 (0·93)	0·89 (0·69)
Median (IQR)	0·7 (0·5–1·2)	0·9 (0·6–1·4)	0·7 (0·5–1·1)	1·0 (0·7–1·5)	0·7 (0·5–1·1)
Liver density, HU					
Mean (SD)	55·4 (10·4)	54·1 (10·8)	55·7 (10·2)	53·6 (10·8)	55·5 (10·3)
Median (IQR)	58 (52–62)	56 (50–61)	58 (52–62)	56 (50–60)	58 (52–62)
Vertebral density, HU					
Mean (SD)	171·2 (42·2)	159·0 (44·4)	174·2 (41·1)	150·9 (47·5)	172·4 (41·6)
Median (IQR)	168 (142–197)	156 (128–186)	171 (146–200)	146 (113–180)	169 (144–198)

All comparisons of CT biomarkers in patients with versus those without events were significant (p<0·001). HU=Hounsfield units. *Defined as acute myocardial infarction, cerebrovascular accident, or congestive heart failure.

Table 1: **Summary data of CT biomarkers according to clinical outcomes**

Statistical analysis

The analysis and modelling were developed specifically for this study and have not been applied previously to other cohorts or scenarios. We compiled and compared summary statistics for patients with and for those without subsequent adverse events. To assess the association between the predictive measures and downstream adverse events, we used both an event-free survival analysis and logistic regression to compute receiver operating characteristic curves (ROCs). Relevant p values were derived with two-sided Student's *t* tests for normally distributed variables, and the Wilcoxon rank-sum test when the normality assumption did not hold. We made area under the ROC (AUROC) comparisons using DeLong's method; p values of less than 0·05 indicate statistical significance. For the time-to-event survival analysis, we generated Kaplan-Meier curves by splitting predictor variables into quartiles. We used Cox proportional hazards models to derive concordance values and individual risk predictions. For ROC curve analysis, datasets were restricted to defined time intervals since time-to-event was not considered. Three arbitrary cutoffs included only patients having at least 2-year, 5-year, or 10-year follow-up, respectively, if they did not have an event within those timeframes. We calculated AUROCs with 95% CIs. We did univariate and multivariate analyses of CT biomarkers, and age and sex were considered as potential confounders. Hazard ratios (HRs) with 95% CIs were computed for each CT biomarker, comparing the highest-risk quartile with the other three quartiles. We used R (version 3.6) for statistical analyses.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The final study cohort consisted of 9223 generally healthy asymptomatic adults (mean age 57·1 years [SD 7·8]; 5152 [56%] women and 4071 [44%] men) who underwent low-dose, unenhanced, abdominal CT between April, 2004, and December, 2016. After final longitudinal clinical follow-up subsequent to CT scanning (median time interval 8·8 years [IQR 5·1–11·6]), adverse clinical outcomes of interest, including major cardiovascular events (myocardial infarction, cerebrovascular accident, or congestive heart failure) or death, were confirmed in 1831 (20%) patients. Of the 549 (6%) patients who died during the surveillance interval, the median time from CT scan to death was 6·1 years (IQR 3·2–9·2; mean 6·2 years [SD 3·6]). Median time to cardiovascular event was 4·4 years (IQR 2·0–7·8; mean 5·0 years [SD 3·6]). Significant differences (p<0·001) were observed in all five automated CT-based measures (aortic calcification, muscle density, ratio of visceral to subcutaneous fat, liver density, and vertebral density) between patients with and those without an adverse event (table 1). Summary data for clinical parameters according to adverse events are shown in the appendix (p 1).

For all data points (ie, 2-year, 5-year, and 10-year AUROC; and Cox model concordance), the automated CT-based univariate results for aortic calcification and muscle density were higher than the FRS, without including any demographic input data (table 2). All five CT-based measures outperformed BMI for all data points. For example, the univariate 5-year AUROC value was

	2-year AUROC (95% CI)	5-year AUROC (95% CI)	10-year AUROC (95% CI)	Cox proportional hazards model concordance (95% CI)
Number of patients	7849	6891	4029	..
Clinical parameters				
FRS	0·700 (0·638–0·762)	0·688 (0·650–0·727)	0·693 (0·665–0·720)	0·681 (0·654–0·707)
BMI	0·546 (0·475–0·617)	0·499 (0·454–0·544)	0·533 (0·502–0·565)	0·520 (0·493–0·546)
Automated CT biomarkers				
Univariate analysis				
Aortic calcification, Agatston score	0·746 (0·688–0·803)	0·743* (0·705–0·780)	0·746* (0·720–0·773)	0·735 (0·709–0·760)
Muscle density, HU	0·736 (0·676–0·796)	0·721 (0·683–0·759)	0·717 (0·691–0·743)	0·700 (0·674–0·725)
Visceral-to-subcutaneous fat ratio	0·685 (0·631–0·739)	0·661 (0·625–0·697)	0·656 (0·629–0·682)	0·648 (0·622–0·674)
Liver density, HU	0·644 (0·589–0·699)	0·619 (0·582–0·656)	0·628 (0·601–0·655)	0·602 (0·576–0·628)
Vertebral density, HU	0·627 (0·554–0·700)	0·646 (0·603–0·688)	0·640 (0·610–0·671)	0·637 (0·612–0·663)
Multivariate analysis†				
Aortic calcification + muscle density	0·780 (0·723–0·836)	0·768 (0·732–0·805)	0·768 (0·743–0·793)	0·772 (0·729–0·781)
Aortic calcification + muscle density + liver density	0·811 (0·761–0·860)	0·782 (0·747–0·817)	0·777 (0·752–0·802)	0·778 (0·739–0·791)
Aortic calcification + muscle density + liver density + visceral-to-subcutaneous fat ratio	0·817 (0·768–0·866)	0·789 (0·755–0·824)	0·780 (0·763–0·813)	0·780 (0·745–0·798)
Aortic calcification + FRS‡	0·774 (0·714–0·833)	0·744 (0·707–0·781)	0·746 (0·716–0·768)	0·733 (0·706–0·760)
Aortic calcification + muscle density + liver density + visceral-to-subcutaneous fat ratio + FRS‡	0·847 (0·796–0·897)	0·796 (0·759–0·834)	0·792 (0·766–0·818)	0·778 (0·750–0·806)

AUROC=area under the receiver operating characteristic curve. BMI=body-mass index. FRS=Framingham risk score. HU=Hounsfield units. *p<0·05 compared with FRS. †p<0·05 for all multivariate comparisons with FRS. ‡No significant improvement compared with CT-based performance without FRS (p values ranged from 0·509 to 0·965 for aortic calcification comparison and p values ranged from 0·406 to 0·806 for aortic calcification, muscle density, liver density, and visceral-to-subcutaneous fat ratio comparison).

Table 2: **Diagnostic performance for predicting overall survival**

0·743 (95% CI 0·705–0·780) for CT-based aortic calcification and 0·721 (0·683–0·759) for muscle density compared with 0·688 (0·650–0·727) for FRS (p=0·047 for aortic calcification *vs* FRS; figure 2A). Automated CT-based fat (5-year AUROC 0·661 [95% CI 0·625–0·697]), liver (0·619 [0·582–0·656]), and bone (0·646 [0·603–0·688]) measures also performed fairly well as univariate measures, whereas BMI was a poor predictor, with a 5-year AUROC value of 0·499 (95% CI 0·454–0·544; p<0·001 for all CT-based parameters *vs* BMI; figure 2A). Similar performance was observed for prediction of downstream major cardiovascular events (appendix p 2). For example, all AUROC values for aortic calcification, whether alone or in combination with other CT-based automated measures, were significantly greater than for FRS (p<0·05). In general, multivariate combinations of CT biomarkers further improved prediction over clinical parameters (p<0·05 for AUROCs; table 2; appendix p 2). For example, combining the three CT-based quantitative biomarkers of aortic calcification, muscle density, and liver density resulted in a 2-year AUROC of 0·811 (95% CI 0·761–0·860; figure 2B) for overall survival.

When FRS was added to the CT-based aortic calcification score, there was no significant improvement of this automated CT measure alone for either cardiovascular events or overall survival, with p values all 0·509 or greater for all AUROC comparisons (table 2; appendix p 2). Similarly, adding FRS to CT-based multivariate combinations did not significantly improve performance (table 2; appendix p 2). Also of note, adding potential confounders of patient age and sex to the multivariate analysis provided only minor incremental benefit to the automated CT data alone (appendix p 5).

Separation between the highest-risk quartile versus the other three was greater for CT-based aortic calcification score, muscle density, and vertebral density compared with FRS (figure 3). Although quartile separation was less pronounced for the CT-based fat and liver measures, each is noticeably better than BMI. Univariate HRs comparing the highest-risk quartile with the other three quartiles were 4·53 (95% CI 3·82–5·37) for aortic calcification, 3·58 (3·02–4·23) for muscle density, 2·28 (1·92–2·71) for visceral-to-subcutaneous fat ratio, 1·82 (1·52–2·17) for liver density, and 2·73 (2·31–3·23) for vertebral density. Corresponding HRs were 1·36 (1·13–1·64) for BMI and 2·82 (2·36–3·37) for FRS (figure 3). Similar time-to-event results were observed when cardiovascular events are included (appendix p 4); univariate HRs ranged from 1·62 (95% CI 1·48–1·79) to 3·53 (3·22–3·87) for the five metabolic CT markers, and were 1·34 (1·21–1·49) for BMI and 2·59 (2·35–2·85) for FRS (appendix p 3). When combining CT-based parameters in a multivariate manner, further improvement in highest-risk quartile separation was observed (appendix p 5).

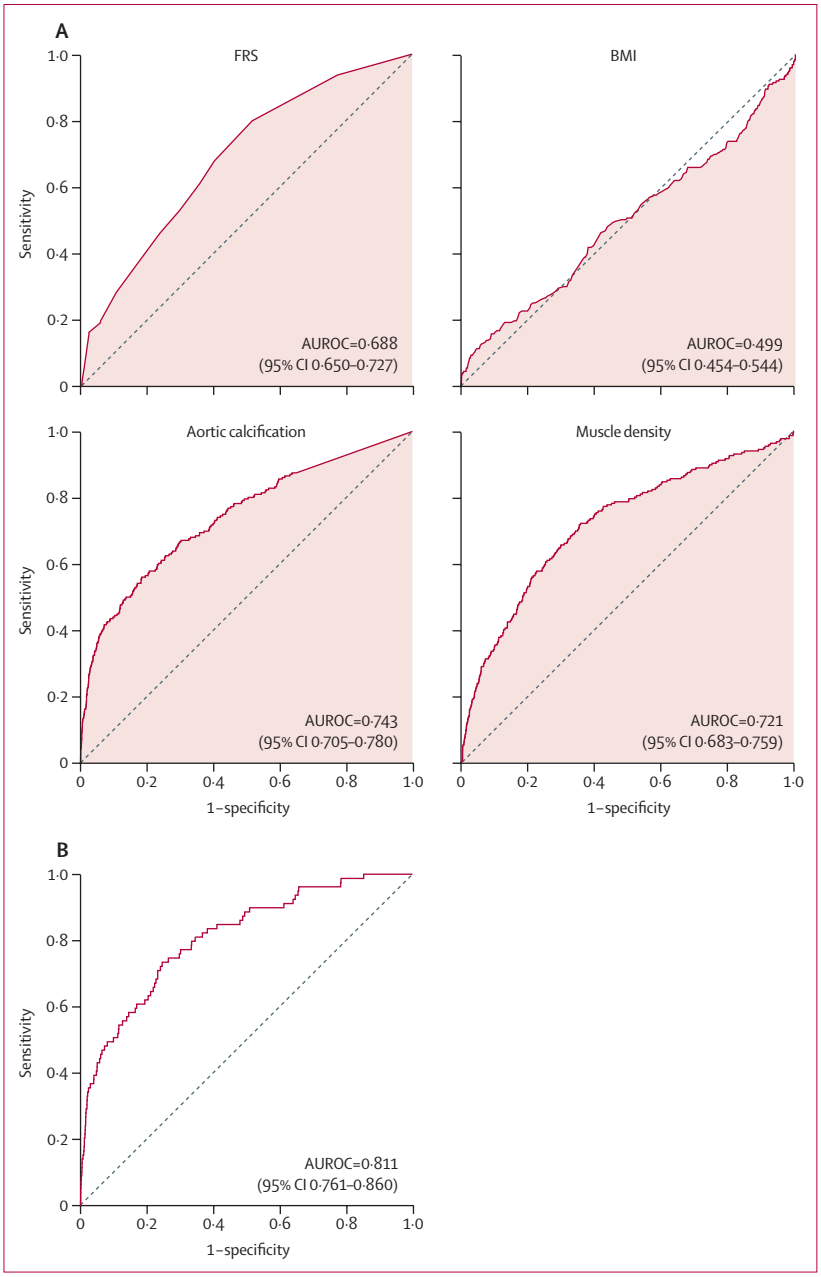

***Figure 2*: AUROCs for predicting overall survival**

(A) AUROCs for the clinical parameters of FRS and BMI, as well as univariate CT measures of aortic calcification and muscle density for predicting overall survival over a 5-year time period. (B) Multivariate CT-based aortic calcification, muscle density, and liver density for predicting overall survival over a 2-year time period. AUROC=area under the receiver operating characteristic curve. BMI=body-mass index. FRS=Framingham risk score.

Figure 1B demonstrates a case example that shows how predictive modelling derived from the quantitative CT data can be applied to an individual patient, similar to the multivariate FRS approach.

Discussion

We found that the AI panel of automated CT-based tissue biomarkers used in this study compared favourably with the FRS and BMI for presymptomatic prediction of future cardiovascular events and death. In fact, in terms of AUROCs and HRs, the univariate CT-based measures of aortic calcification alone significantly outperformed the multivariate FRS for major cardiovascular events and overall survival. On the basis of previous preliminary work that required manual case-by-case interaction for abdominal aortic calcium scoring in a smaller cohort,[9] we expected the automated calcium tool to be valuable for cardiovascular risk profiling. BMI, which does not account for the relative anatomical distribution of fat,[10] was a poor predictor of cardiovascular events and overall survival, whereas the CT-based visceral-to-subcutaneous fat ratio performed significantly better than BMI. In fact, all five automated CT-based measures clearly outperformed BMI for adverse event prediction. Although BMI quartile separation was minimal, the slightly greater risk of death observed for the first and fourth quartiles (figure 3) probably reflects the previously described U-shaped risk curve for this parameter.[31] Liver density at non-contrast CT directly correlates with fat content[11,19] and reflects the high prevalence of hepatic steatosis, which has relevance for metabolic syndrome. Although its univariate performance was not remarkable, liver density appears to have complementary value in terms of AUROC when combined with other CT biomarkers, such as aortic calcification and muscle density. In general, a multivariate combination of these CT-based biomarkers is probably the best way forward for optimised risk stratification. Furthermore, these CT biomarkers appear to be stronger predictors of future cardiovascular events than a panel of previously studied blood-based and urine-based biomarkers.[32]

This study demonstrates the potential value of harnessing the rich biometric tissue data embedded within all body CT scans that typically go unused in routine practice. Although such an opportunistic approach can be applied using manual or semi-automated measures, the maturation of robust, fully automated AI algorithms provides for a more efficient and objective means for high-volume, population-based opportunistic screening. With more than 80 million body CT scans performed each year in the USA,[33] much of the focus has been placed on negative concerns about so-called incidentalomas and radiation exposure.[34,35] However, since most scans are performed on older adults (eg, age 50 years or older), the opportunistic screening potential also becomes apparent. We applied these CT-based tools for assessing body composition to a generally healthy outpatient screening cohort to start, but this approach can also be applied to other cohorts, including those with symptoms or increased risk factors. We envision a (not-too-distant) future in which this valuable prognostic CT information might be routinely captured and reported for the benefit of the patient, regardless of the clinical indication for imaging. The added value from these CT-based metabolic biomarkers requires no

additional patient time or radiation exposure, and has the potential for improved individualised risk profiling.

A recent study[14] has emphasised the relative lack of prevention research that measures leading risk factors for death or disability as outcomes among studies supported by the NIH. Our study was intended in part to help address this research gap. Although the current study focused only on the clinical outcomes of subsequent cardiovascular events and overall survival, these automated CT biomarkers have prognostic value for other cardiometabolic endpoints, such as osteoporotic fragility fractures and metabolic syndrome. We are only advocating use of these additional CT-based body composition data in an opportunistic manner, and not as the sole reason for scanning. However, when coupled with an established indication such as CTC for colorectal cancer prevention, the concept of stand-alone population-based CT screening of asymptomatic adults could potentially be considered. In this scenario, the cumulative value of the screening CT data would need to clearly outweigh the potential harms, including cost and radiation exposure, and provide benefit beyond the more typical clinical parameters. Nonetheless, in current practice, these additional CT data are largely going unused in the many patients being scanned for a wide range of established clinical indications. Automated CT measures of muscle, fat, and bone might also be valuable for opportunistic frailty monitoring in patients with cancer, who often undergo repeated CT scanning for treatment response and surveillance.

The ever-increasing attention focused on the potential of AI in medicine is nearly ubiquitous, both in the medical literature and the lay press.[1] The application of countless algorithms ranging from classic machine learning to more complex deep learning with convolutional neural networks is omnipresent. Along with a few other specific areas in medicine, medical image analysis represents a logical target for AI application.[3,4] Despite predictions by some that disruptive AI technology is destined to soon displace the radiologist,[36] the complexity of creating, training, and modifying the vast number of necessary algorithms argues instead for active engagement over replacement.[3,4,37,38] Furthermore, use of AI is not new in radiology, and co-authors of the current work have been involved in CT-based computer-aided detection for many years.[39] Although we believe that AI-based advances will ultimately enhance the practice of radiology, the recent hype has greatly outpaced true progress so far. The validated CT-based AI tools that we demonstrate herein represent the culmination of years of development, training, and testing. Although some of the processes are rooted in deep-learning algorithms, the output of these quantitative tools is straightforward and can be visually confirmed for quality assurance (ie, explainable AI), as opposed to the more black box feel of many other deep-learning AI solutions.

We acknowledge limitations to our investigation. All CT scans were performed with non-contrast technique; we

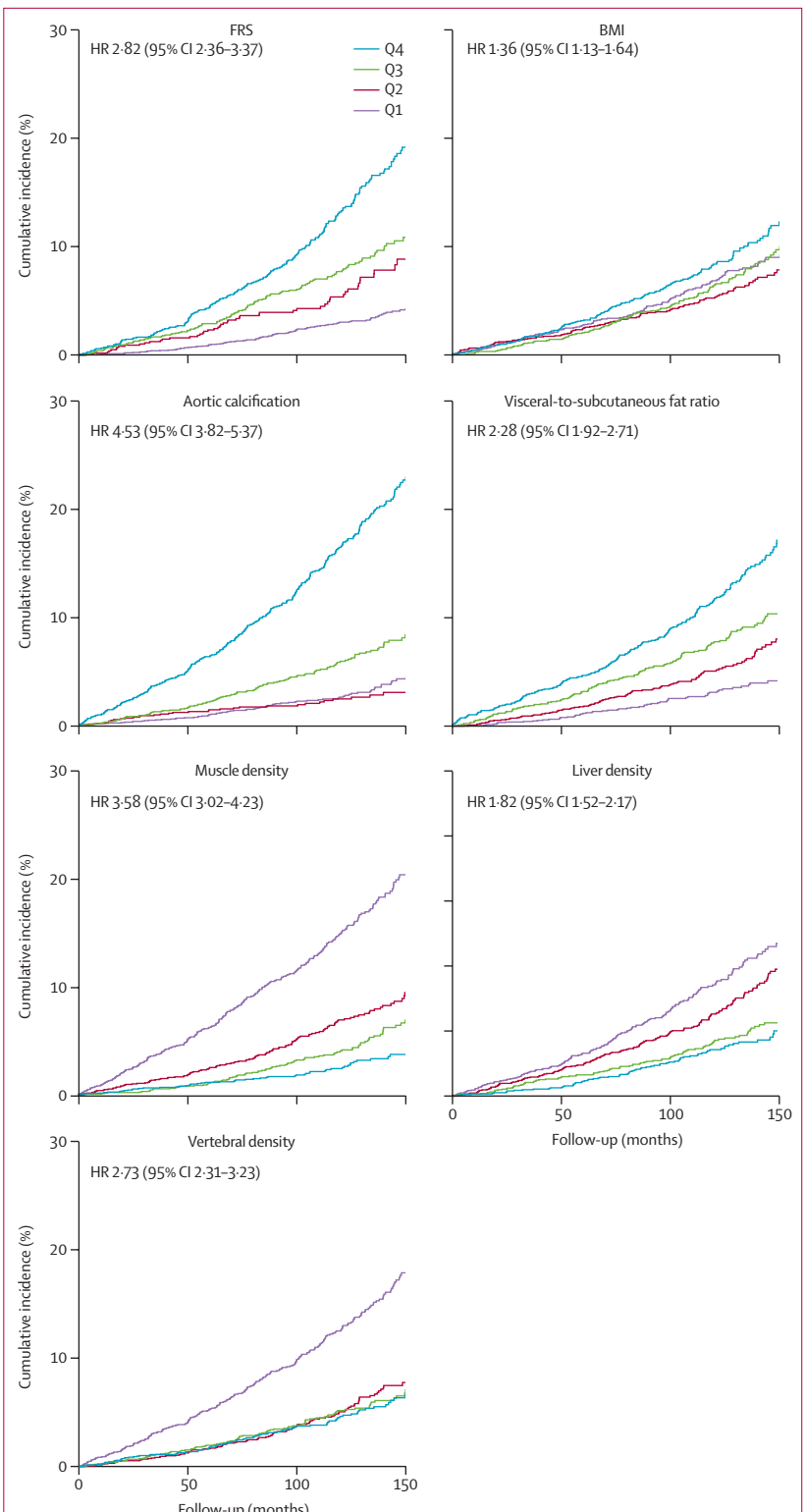

***Figure 3*: Kaplan-Meier time-to-death plots by quartile for clinical parameter and univariate CT biomarkers**
BMI=body-mass index. FRS=Framingham risk score. HR=hazard ratio. Q=quartile.

are currently validating the use of these automated tools in a separate asymptomatic healthy cohort who underwent CT both without and with intravenous contrast. Risk stratification was based on analysis of the initial CTC examination in this screening cohort. A subset of more than 2000 patients underwent subsequent CT screening 5–10 years later, for which we plan to assess for interval changes in these automated measures that might offer additive value. Although our relatively unique CT screening cohort comprising generally healthy outpatient adults was ideal for initial investigation, external validation in other screening populations with broader racial diversity is warranted, as our cohort was about 90% white. Application to new cohorts, including symptomatic patients at other centres, would also allow for further testing of the predictive models. This testing could be done with a federated approach.[40] One could argue that the FRS is outdated as a clinical comparator and used less often in the clinic. However, the FRS has served well for several previous trials, providing greater context as a common reference standard. Furthermore, the recent 2019 American College of Cardiology/American Heart Association guidelines state that the FRS might still be appropriate for use as an alternative risk prediction tool.[41] It is conceivable that unforeseen confounders between the earlier testing or training cohorts and the current study group could exist with regard to measuring body composition by CT. Given the nature of the electronic health record search for adverse outcomes, it is possible that some definable events were not captured. However, our population was geographically stable. The quartile separation approach we have chosen probably does not reflect the optimal division of the data, but instead represents a starting point for further investigation, potentially with even larger cohorts. Finally, it is also important to consider the potential for possible unintended harm if subsequent intervention or inaction resulted from an incorrect classification of cardiovascular risk based on the CT-based body composition data.

In conclusion, we have shown that fully automated quantitative tissue biomarkers derived from abdominal CT scans can outperform established clinical parameters for presymptomatic prediction of future cardiovascular events and overall survival. This approach leverages robust biometric data embedded in all such scans and can add opportunistic value to abdominal CT scans performed for a wide range of other indications.

Contributors

PJP and RMS conceived study design. PJP, PMG, and RMS did the literature search. PJP, PMG, SJL, JL, and RMS collected data. All authors did data analysis or interpretation. PJP, PMG, RZ, and RMS created the figures. PJP wrote the report, which was edited and approved by all authors.

Declaration of interests

RMS receives royalties from iCAD, Philips, PingAn, and ScanMed and research support from PingAn and NVIDIA. PJP is an adviser or consultant for Zebra Medical Vision and Bracco Diagnostics, and shareholder in Cellectar, Elucent, and SHINE.

Data sharing

The summary numerical output of the automated CT-based tools will be shared upon request, subject to an internal review by PJP and RMS to ensure that participant privacy is protected and subject to completion of a data sharing agreement, as well as approval from both the University of Wisconsin School of Medicine & Public Health (Madison, WI, USA) and the US National Institutes of Health Clinical Center (Bethesda, MD, USA). Pending the aforementioned approvals, data sharing will be made in a secure setting, on a per-case-specific manner. Please submit such requests to PJP.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH) Clinical Center. This study used the high-performance computing capabilities of the NIH Biowulf cluster.

References

1 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019; **380:** 1347–58.

2 Wang S, Summers RM. Machine learning and radiology. *Med Image Anal* 2012; **16:** 933–51.

3 Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. *Radiology* 2018; **288:** 318–28.

4 Dreyer KJ, Geis JR. When machines think: radiology's next frontier. *Radiology* 2017; **285:** 713–18.

5 Jang S, Graffy PM, Ziemlewicz TJ, Lee SJ, Summers RM, Pickhardt PJ. Opportunistic osteoporosis screening at routine abdominal and thoracic CT: normative L1 trabecular attenuation values in more than 20 000 adults. *Radiology* 2019; **291:** 360–67.

6 Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med* 2013; **158:** 588–95.

7 Lee SJ, Graffy PM, Zea RD, Ziemlewicz TJ, Pickhardt PJ. Future osteoporotic fracture risk related to lumbar vertebral trabecular attenuation measured at routine body CT. *J Bone Miner Res* 2018; **33:** 860–67.

8 Shah RV, Yeri AS, Murthy VL, et al. Association of multiorgan computed tomographic phenomap with adverse cardiovascular health outcomes: the Framingham Heart Study. *JAMA Cardiol* 2017; **2:** 1236–46.

9 O'Connor SD, Graffy PM, Zea R, Pickhardt PJ. Does nonenhanced CT-based quantification of abdominal aortic calcification outperform the Framingham risk score in predicting cardiovascular events in asymptomatic adults? *Radiology* 2019; **290:** 108–15.

10 Abraham TM, Pedley A, Massaro JM, Hoffmann U, Fox CS. Association between visceral and subcutaneous adipose depots and incident cardiovascular disease risk factors. *Circulation* 2015; **132:** 1639–47.

11 Pickhardt PJ, Graffy PM, Reeder SB, Hernando D, Li K. Quantification of liver fat content with unenhanced MDCT: phantom and clinical correlation with MRI proton density fat fraction. *AJR Am J Roentgenol* 2018; **211:** W151–57.

12 Brown JC, Caan BJ, Prado CM, et al. Body composition and cardiovascular events in patients with colorectal cancer: a population-based retrospective cohort study. *JAMA Oncol* 2019; **5:** 967–72.

13 Derstine BA, Holcombe SA, Ross BE, Wang NC, Su GL, Wang SC. Skeletal muscle cutoff values for sarcopenia diagnosis using T10 to L5 measurements in a healthy US population. *Sci Rep* 2018; **8:** 11369.

14 Vargas AJ, Schully SD, Villani J, Ganoza Caballero L, Murray DM. Assessment of prevention research measuring leading risk factors and causes of mortality and disability supported by the US National Institutes of Health. *JAMA Netw Open* 2019; **2:** e1914718.

15 Lee SJ, Liu J, Yao J, Kanarek A, Summers RM, Pickhardt PJ. Fully automated segmentation and quantification of visceral and subcutaneous fat at abdominal CT: application to a longitudinal adult screening cohort. *Br J Radiol* 2018; **91:** 20170968.

16 Graffy PM, Liu J, O'Connor S, Summers RM, Pickhardt PJ. Automated segmentation and quantification of aortic calcification at abdominal CT: application of a deep learning-based algorithm to a longitudinal screening cohort. *Abdom Radiol* 2019; **44:** 2921–28.

17 Pickhardt PJ, Lee SJ, Liu J, et al. Population-based opportunistic osteoporosis screening: validation of a fully automated CT tool for assessing longitudinal BMD changes. *Br J Radiol* 2019; **92:** 20180726.

18 Graffy PM, Liu J, Pickhardt PJ, Burns JE, Yao J, Summers RM. Deep learning-based muscle segmentation and quantification at abdominal CT: application to a longitudinal adult screening cohort for sarcopenia assessment. *Br J Radiol* 2019; **92:** 20190327.

19 Graffy PM, Sandfort V, Summers RM, Pickhardt PJ. Automated liver fat quantification at nonenhanced abdominal CT for population-based steatosis assessment. *Radiology* 2019; **293:** 334–42.

20 Kim DH, Pickhardt PJ, Taylor AJ, et al. CT colonography versus colonoscopy for the detection of advanced neoplasia. *N Engl J Med* 2007; **357:** 1403–12.

21 Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016; **18:** e323.

22 Pickhardt PJ, Choi JR, Hwang I, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 2003; **349:** 2191–200.

23 Burns JE, Yao J, Chalhoub D, Chen JJ, Summers RM. A machine learning algorithm to estimate sarcopenia on abdominal CT. *Acad Radiol* 2019; published online May 22. DOI:10.1016/j.acra.2019.03.011.

24 Chellamuthu K, Liu J, Yao J, et al. Atherosclerotic vascular calcification detection and segmentation on low dose computed tomography scans using convolutional neural networks. 2017 IEEE 14th International Symposium on Biomedical Imaging; Melbourne, VIC; 2017: 388–91.

25 Liu J, Lu L, Yao J, Bagheri M, Summers RM. Pelvic artery calcification detection on CT scans using convolutional neural networks. SPIE Medical Imaging 2017; Orlando, FL; Feb 11–16, 2017 (abstr 101341A).

26 Summers RM, Baecher N, Yao J, et al. Feasibility of simultaneous computed tomographic colonography and fully automated bone mineral densitometry in a single examination. *J Comput Assist Tomogr* 2011; **35:** 212–16.

27 Yao JH, O'Connor SD, Summers RM. Automated spinal column extraction and partitioning. 2006 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006; Arlington, VA: 2006: 390–93.

28 Liu JM, Yao JH, Bagheri M, Sandfort V, Summers RM. A semi-supervised CNN learning method with pseudo-class labels for atherosclerotic vascular calcification detection. 2019 IEEE 16th International Symposium on Biomedical Imaging; Venice, Italy; 2019: 780–83.

29 Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* 2019; **9:** 16884.

30 D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; **117:** 743–53.

31 Zheng W, McLerran DF, Rolland B, et al. Association between body-mass index and risk of death in more than 1 million Asians. *N Engl J Med* 2011; **364:** 719–29.

32 Wang TJ, Gona P, Larson MG, et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med* 2006; **355:** 2631–39.

33 IMV. 2018 CT Market Outlook Report. Des Plains, IL: IMV Medical Information Division, 2018.

34 Berland LL, Silverman SG, Gore RM, et al. Managing incidental findings on abdominal CT: white paper of the ACR incidental findings committee. *J Am Coll Radiol* 2010; **7:** 754–73.

35 Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. *N Engl J Med* 2007; **357:** 2277–84.

36 Obermeyer Z, Emanuel EJ. Predicting the Future - Big data, machine learning, and clinical medicine. *N Engl J Med* 2016; **375:** 1216–19.

37 Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 2017; **208:** 754–60.

38 Langlotz CP, Allen B, Erickson BJ, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology* 2019; **291:** 781–91.

39 Summers RM, Yao J, Pickhardt PJ, et al. Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology* 2005; **129:** 1832–44.

40 Yang Q, Liu Y, Chen TJ, Tong YX. Federated machine learning: concept and applications. *ACM Trans Intell Syst Technol* 2019; **10:** 19.

41 Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* 2019; **74:** e177–232.