



Exploring the potential of generative artificial intelligence in medical image synthesis: opportunities, challenges, and future directions

Bardia Khosravi, Saptarshi Purkayastha, Bradley J Erickson, Hari M Trivedi, Judy W Gichoya

Generative artificial intelligence has emerged as a transformative force in medical imaging since 2022, enabling the creation of derivative synthetic datasets that closely resemble real-world data. This Viewpoint examines key aspects of synthetic data, focusing on its advancements, applications, and challenges in medical imaging. Various generative artificial intelligence image generation paradigms, such as physics-informed and statistical models, and their potential to augment and diversify medical research resources are explored. The promises of synthetic datasets, including increased diversity, privacy preservation, and multifunctionality, are also discussed, along with their ability to model complex biological phenomena. Next, specific applications using synthetic data such as enhancing medical education, augmenting rare disease datasets, improving radiology workflows, and enabling privacy-preserving multicentre collaborations are highlighted. The challenges and ethical considerations surrounding generative artificial intelligence, including patient privacy, data copying, and potential biases that could impede clinical translation, are also addressed. Finally, future directions for research and development in this rapidly evolving field are outlined, emphasising the need for robust evaluation frameworks and responsible utilisation of generative artificial intelligence in medical imaging.

Introduction

Generative artificial intelligence is a class of deep learning models capable of creating content that diverges from traditional discriminative models focused on interpretation or decision making. Generative artificial intelligence has seen rapid advancements over the past 3 years, with large language models gaining substantial public attention after the introduction of ChatGPT, a model trained on an extensive corpus of text to create coherent and realistic responses to user queries.¹ Large language models have shown noteworthy capabilities in the understanding and generation of natural language, paving the way for more advanced multimodal models that combine textual, visual, and contextual understanding. These large multimodal models have the potential to aid various domains, including health care, by integrating data from different input streams. Notable examples of large language models in medicine are Med-PaLM and Med-Gemini, which have shown promising results in tasks such as answering medical questions, summarising medical documents, and suggesting potential differential diagnoses on the basis of patient symptoms and test results. In addition, MedImageInsight and MedImageInsight are models trained on different types of medical images including radiology images (eg, chest x-rays, mammograms, CT), as well as dermatology and ophthalmology images, which allow end users to interact with the model using both language and images (and are thus known as multimodal foundation models). These multimodal models provide unconventional visual question answering ability and are able to learn from a few examples to perform downstream classification tasks.^{2,3}

Preliminary evidence suggests that generative artificial intelligence in the realm of visual content has made

remarkable advancements with models such as DALL-E, Stable Diffusion, Sora, and Veo, which excel in generating realistic images and videos based on textual prompts.⁴⁻⁶ Although these models primarily process text as input, with some using images for conditioning purposes, their primary focus is on generating high-quality images. Seminal works published since 2022 in medical imaging have shown the potential of generative artificial intelligence in creating realistic medical images (synthetic data), suggesting new approaches for research and clinical applications.⁷⁻¹⁰

This Viewpoint provides a comprehensive overview of synthetic data in medical imaging and critically analyses the advancements, applications, and challenges of this field. To this end, various image generation paradigms are examined, with the intention to assess how these generative technologies are changing the landscape of medical imaging research. The potential of these models and their derivative synthetic datasets, particularly their ability to augment and diversify medical research resources, are explored, in addition to their benefits in terms of data augmentation, anonymisation, and modelling biological phenomena. Finally, the challenges of using synthetic data are discussed, including the need for rigorous evaluation metrics and ethical considerations, and potential research directions are proposed that could substantially benefit the field of medical imaging.

Synthetic datasets

Generative models

The field of synthetic data is still in its nascent stages, with no consensus on a single, universally accepted definition as yet. This absence of a clear definition has led to

Lancet Digit Health 2025;

7:100890

Published Online August 14, 2025

<https://doi.org/10.1016/j.landig.2025.100890>

Department of Radiology (B Khosravi MD MPH, Prof B J Erickson MD PhD) and Department of Orthopedic Surgery (B Khosravi), Mayo Clinic, Rochester, MN, USA; Luddy School of Informatics and Computing, Indiana University, Indianapolis, IN, USA (S Purkayastha PhD); Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA, USA (H M Trivedi MD, J W Gichoya MD MS)

Correspondence to: Dr Judy W Gichoya, Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA
judywawira@emory.edu

For Med-PaLM, see <https://sites.research.google/med-palm/>

For Med-Gemini, see <https://research.google/blog/advancing-medical-ai-with-med-gemini/>

For Med-Gemina, see <https://deepmind.google/models/gemina/medgemina/>

For MedImageInsight, see <https://learn.microsoft.com/en-us/azure/ai-foundation/how-to/healthcare-ai/deploy-medimageinsight>

For DALL-E, see <https://openai.com/index/dall-e-3/>

For Stable Diffusion, see <https://stability.ai/stable-image>

For Sora, see <https://openai.com/sora/>

For more on Veo, see <https://deepmind.google/technologies/veo/veo-2/>

Generative models

inconsistencies in how the term is used and interpreted across various contexts, which in turn can affect the reproducibility and transparency of research involving synthetic data.¹¹ The Royal Society and The Alan Turing Institute put forth a working definition of synthetic data in 2022, as data that have been generated using a purpose-built mathematical model or algorithm, with the aim of solving a (set of) data science task(s).¹² This proposed definition emphasises the functional and intentional aspects of synthetic data, focusing on its strategic application in tackling complex scientific challenges rather than simply mimicking the statistical properties of the original data.

The advancement of generative artificial intelligence introduces a new concept in data sharing, which we refer to as a model as a dataset. In this concept, generative models learn and store patterns and characteristics of the original data in their internal parameters (weights).¹³ These trained weights contain a compressed version of the key features and relationships of the training data. Unlike traditional dataset sharing, which involves transferring actual images, sharing model weights provides an efficient alternative that allows others to generate new synthetic images with properties similar to the original data. These synthetic datasets have been shown to closely resemble the source data and capture their distribution, including the relationship of different anatomical features and their correlation with different pathological processes.^{8,9}

Two broad categories of generative models provide the ability to generate synthetic datasets: physics-informed and statistical models.

Physics-informed models are primarily rule-based approaches that incorporate domain-specific knowledge and physics principles through mathematical equations and explicit constraints to generate realistic and physically plausible data. Rather than learning the patterns directly from data, these models encode expert knowledge and known physics laws (eg. fluid dynamics, tissue biomechanics, or radiation physics) to simulate biological phenomena. These models have been applied successfully in medical imaging to simulate anatomical structures (such as a shape model of the femoral bone), physiological processes (such as blood flow dynamics in vascular structures), and medical interventions (such as simulating the distribution of the radiation dose in radiotherapy planning).¹⁴ Physics-informed models offer high fidelity and interpretability but might require extensive domain expertise and computational resources.

In contrast to physics-informed models, statistical models learn from data patterns and distributions (figure 1). Among them, variational autoencoders (VAEs) function by compressing data into a lower-dimensional representation, also known as latent space, and then reconstructing the data, thereby capturing the data distribution effectively.¹⁵ Generative adversarial networks (GANs) operate through a dual-network system, in which a generator creates data samples and a discriminator evaluates these data samples and provides feedback to the generator.¹⁶ This synergy

continually enhances the quality and realism of the data generated. Denoising diffusion probabilistic models (DDPMs) introduce noise into an image and learn to reverse this process, producing high-quality samples.¹⁷

Statistical models encounter the generative artificial intelligence trilemma, which involves balancing high intelligence, comprehensive mode coverage, and rapid sample quality, comprehensive mode coverage, and rapid sampling rates (figure 2).¹⁸ VAEs are notable for their quick sampling capabilities, sometimes resulting in lower sample quality. GANs excel at generating high-quality samples but might not always capture all data variations, leading to low mode coverage, known as mode collapse. DDPMs stand out for their ability to generate samples of exceptional quality and extensive mode coverage, albeit at a slower sampling rate. End users select the generative model that matches their application of interest, balancing the desired image quality and speed. For dataset generation purposes, the priority typically shifts towards ensuring high image quality and comprehensive mode coverage, often outweighing concerns of sampling speed.

Use cases in medical imaging

Generative models and their synthetic datasets have numerous applications in medical imaging (panel 1). One well studied use case involves supplementing or replacing real data to train deep learning models for downstream tasks such as classification or segmentation. Generated images can be conditioned on class labels (eg. presence or absence of pneumonia) or descriptive text (eg. right middle lobe consolidation). Research has shown that images generated by GANs and DDPMs can improve the performance of downstream pathology classifiers substantially.^{7,19,20} Notably, the classifier performance improves as more synthetic data are added to the real dataset. In some cases, a sufficiently large pool of generated images can match the performance benefit of real data, potentially opening new avenues for data sharing whereby synthetic data acts as a replacement of the original data.⁸ However, when training and evaluating generative models, caution is required to avoid distribution leakage (in which a patient is represented in both training and test data), which could overestimate performance improvements.⁸ Of note, repeatedly training image generation models on the output of other generative models (usually more than three iterations) risks mode collapse, which degrades the quality of the final model.²¹

Generative models also excel at image transformations. VAEs and GANs have long enabled low-dose CT image denoising, eventually reducing radiation exposure for patients.^{22,23} Of late, accelerated MRI techniques have been used to reduce the scan time by 30%.²⁴ Another image-to-image transformation use case generates missing MRI sequences, enabling training of downstream algorithms requiring all four sequences: T1, T2, post-contrast T1, and FLAIR.^{25,26} DDPMs have enabled inpainting, which involves selectively adding or removing specific image parts on the basis of criteria, without altering the context. For instance, trained diffusion models can introduce brain tumour

DDPMs

Trilemma
Comprehensiveness

VAEs

GANs

DDPMs

model as a dataset

DATA
AugmentationDistribute
LearnGANs
mode collapseCT
DenoisingImage
transform

Inpainting

Image Quality metrics (cristina gamin)

lesions in healthy brain MRIs or remove tumoural regions by drawing on an image.³⁷ Such edits can enrich under-represented datasets and introduce rare conditions, such as adding brain tumours to individuals with Alzheimer's disease. A more advanced version of the inpainting technique was developed to edit specific regions of a chest radiograph using text prompts.³⁸ The resulting edited images were used to stress-test existing models—for example, removing chest tubes from pneumothorax images to evaluate classifier performance without this known confounder.³⁹

Evaluating image quality

Evaluating the quality of generated images, which determines how these synthetic images are used, is crucial. Various metrics have been proposed to quantify the quality of generated images, both in the presence and absence of ground truth references. These metrics can be broadly categorised into two groups: image metrics and text-image metrics (panel 2).

Image metrics

When ground truth images are available—for example, in tasks such as super resolution and denoising of medical images—traditional metrics such as structural similarity index and the peak signal-to-noise ratio can be used to measure the similarity between the generated and reference images.^{40,41} However, in the absence of ground truth—for example, in class-conditioned image generation—alternative metrics are required. For instance, classification accuracy score trains a classification model on derived medical data and evaluates its performance on real images, providing insights into the domain adaptation capabilities of the generation models.³⁸

Another widely adopted metric is the inception score, which uses an inception network pretrained on ImageNet to evaluate class predictions for a set of generated samples.³⁹ Fréchet inception distance (FID) compares the means and covariances of features extracted by an ImageNet-pretrained inception network between the generated and real samples.⁴⁰ By accounting for the target distribution, FID provides a better estimate of image diversity than inception score. Several variants and improvements of FID have been proposed—eg, the kernel inception distance is a variant of FID that enables metric calculation using a small number of samples, unlike FID calculation, which requires generation of a large number of samples and is resource intensive.⁴¹ One limitation of these metrics is that they depend on pre-trained networks, and unlike natural images, no universally accepted model for feature extraction exists in medical imaging.

Human evaluation

In addition to computational metrics, human evaluation remains a gold standard for assessing the quality of generated medical images. The human Turing test involves domain experts who are asked to discern between real and

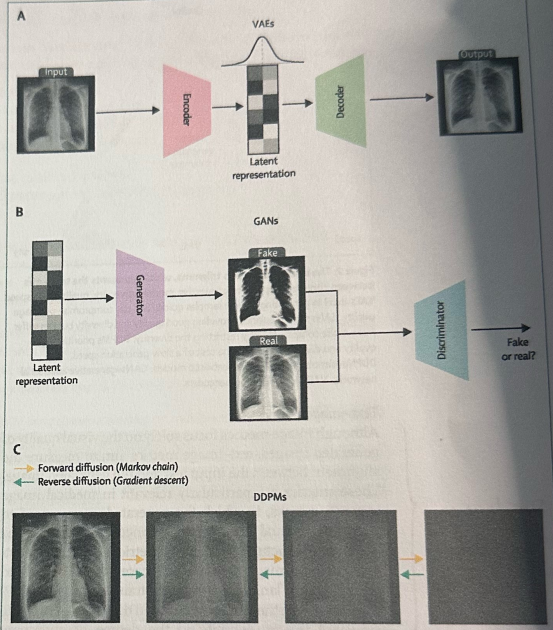


Figure 1: The architectures and key components of three popular statistical models used in image generation (A) VAEs consist of an encoder that compresses the input data into a lower-dimensional latent representation and a decoder that reconstructs the original data from the latent space. The model is trained to minimise the reconstruction error while also regularising the latent space to follow a previous distribution, typically a standard normal distribution. This training enables the generation of new samples by sampling from the learned latent distribution and decoding them. (B) GANs use a two-network architecture, with a generator that creates synthetic data samples and a discriminator that distinguishes between real and generated samples. The generator aims to produce samples that are indistinguishable from real data, whereas the discriminator provides feedback to guide the generator's improvement. Through an adversarial training process, the generator learns to capture the underlying data distribution, enabling the creation of realistic samples. (C) DDPMs generate data by learning to capture the underlying data distribution, enabling the creation of realistic samples. The model starts with a sample from a simple distribution (eg, Gaussian noise) and iteratively denoises the sample using a learned Markov chain. At each step, the model estimates the gradient of the data distribution and refines the sample accordingly. By repeatedly applying this process, DDPMs can produce high-quality samples that closely resemble the training data. The figure depicts the forward diffusion process that gradually adds noise to the data and the reverse diffusion process that progressively denoises the sample to generate a clean output. DDPMs=denoising diffusion probabilistic models. GANs=generative adversarial networks. VAEs=variational autoencoders.

derived medical images.⁴² This assessment provides insights into the perceptual quality and realism of generated images, which is crucial for medical imaging, in which accuracy and fidelity are paramount. However, as perceptual quality and realism are subjective measures, a wide range of participants with different experience levels should be involved in the image evaluation process.⁴³

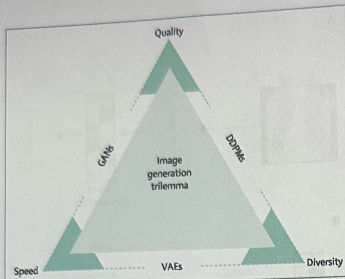


Figure 2: The image generation trilemma, which represents the trade-offs between three key aspects of generative models: diversity, quality, and speed. VAEs excel in generating diverse samples quickly but can compromise on image quality. GANs strike a balance, providing good quality and diversity but can suffer from mode collapse, thereby restricting the diversity. DDPMs prioritise high-quality and diverse samples at the cost of a slow generation speed. DDPMs=denoising diffusion probabilistic models. GANs=generative adversarial networks. VAEs=variational autoencoders.

Text-image metrics

Although image metrics focus solely on the visual quality of generated images, text-image metrics aim to measure the alignment between the input text and the generated image. These metrics are particularly relevant in medical image generation tasks, in which the generated images need to reflect the textual descriptions of medical conditions or anatomical structures accurately. Metrics such as contrastive language-image pretraining score (CLIPScore) and bootstrapping language-image pretraining score (BLIP-Score) measure the similarity between the input text and the generated image, quantifying the degree of alignment between the two modalities.^{44,45}

Image-text matching is another crucial group of metrics for evaluating the alignment between generated medical images and their corresponding textual descriptions. Compositional quality metrics assess this alignment by decomposing the text and image into individual components and measuring their correspondence, often using object detection techniques.^{46,47} These metrics go beyond overall visual similarity and focus on accurate representation of specific anatomical structures, pathologies, or medical conditions mentioned in the text. By ensuring that the generated images convey the intended medical information accurately, compositional quality metrics can play a key role in medical education and research.

Health-care-specific metrics

Evaluating synthetic medical images requires metrics tailored to health-care needs, beyond general purpose tools such as the structural similarity index or FID. Efforts are under way to adapt existing metrics for medical contexts. For instance, researchers have begun replacing ImageNet-pretrained models in FID with networks trained on medical

datasets such as RadImageNet to create a medical FID, which captures the statistical properties of radiology images better.⁴⁸ However, health-care-specific metrics remain an active research area as disease classifiers might rely more on local features than global features.⁴⁹ Similarly, anatomical accuracy is being prioritised by developing measures that use segmentation tools to ensure that crucial structures (such as organs or lesions) are preserved in synthetic images.⁵⁰ These adaptations aim to address the limitations of standard metrics, which often fail to reflect clinical relevance or diagnostic utility.

A suggested next step is to integrate clinical validation with these computational approaches. Human evaluations such as the human Turing test already involve experts distinguishing real images from synthetic ones, offering insights into the perceptual quality that is important for medical use. For text-guided image generation, metrics such as CLIP-Score are being refined by using medical foundation models such as BioMedClip.⁵¹ Testing synthetic images in practical, clinical tasks such as training classifiers for disease detection can further highlight their utility. Combining these efforts could provide a robust, health-care-specific evaluation, thereby ensuring that synthetic images meet both technical and clinical standards for advancing medical imaging research and practice.

Potentials and promises

Synthetic data generation and image generation models hold immense promise for the future of medical imaging research. By leveraging the power of generative models, researchers can unlock unprecedented levels of data diversity, privacy preservation, and multifunctionality, changing the way dataset creation, utilisation, and disease modelling are approached.

Increased dataset size and diversity

One of the key advantages of generating data via statistical models lies in their ability to increase dataset size and diversity. Preliminary evidence suggests that generative models can be trained to disentangle specific associations within data, allowing for the creation of novel combinations that might not be readily available in real-world datasets.^{32,33} For instance, a model trained on brain MRI scans can generate images with varying degrees of atrophy or lesion load, independent of factors such as age or sex. Such disentanglement enables training models to detect specific pathologies without confounding the effects of other variables. As mentioned earlier, supplementing increased dataset size with generated images could lead to enhanced downstream model performance.³ Moreover, targeted oversampling of minoritised sociodemographic groups or patients diagnosed with rare diseases through synthetic data generation has been shown to close the fairness gap by 40%.³² Synthetic data generation closes this fairness gap by facilitating an increase in dataset sizes that represent the original dataset distribution for various subgroups.

Panel 1 (continued from previous page)

Kyung et al (2024):²⁸ Forecasting chest radiograph morphology on the basis of electronic health record data

- Achieved a weighted macro AUROC of 0.72 in predicting future chest x-ray pathologies, outperforming tabular-only classifiers and previous label baselines
- Maintained sex (AUROC 0.96) and age (0.45) correlations in synthetic images

Liu et al (2025):²⁹ Forecasting tumour growth on the basis of baseline tumour characteristics and treatment plan

- SSIM of 0.92 and PSNR of 29.0 for multiparametric MRI generation, outperforming baseline models without treatment-aware conditioning
- Generated MRI quality remained high across different treatment-day ranges, with SSIM ranging from 0.88 to 0.94 depending on the treatment phase
- Tumour growth predictions were most reliable within a 4-month window, with the Dice similarity coefficient dropping from 0.85 to 0.46 as the time interval extended from 0.5 months to greater than 24 months

3D=three dimensional, AUROC=area under the receiver operating characteristic curve, DDPMs=denoising diffusion probabilistic models, PSNR=peak signal-to-noise ratio, SSIM=structural similarity index.

Panel 2: Summary of image quality metrics based on use case for medical image generation

Image super resolution, denoising, and inpainting

- SSIM: assesses structural similarity between generated and reference images by considering luminance, contrast, and structure
- PSNR: measures the ratio between the maximum possible power of a signal and the power of corrupting noise between generated and reference images

Class-conditioned and unconditional image generation

- IS: compares class predictions of generated and reference samples using an inception network pretrained on ImageNet
- FID: compares means and covariances of features extracted from generated and target distributions using an inception network pretrained on ImageNet
- KID: computes squared MMD between inception representations of generated and target distributions using an inception network pretrained on ImageNet

Domain adaptation, and class-conditioned image generation

- CAS: uses a classifier trained on derived medical images and evaluates performance on real images

Perceptual quality assessment, and realism evaluation

- Human Turing test: medical experts discern between real and derived images

Image generation from textual descriptions

- Segmentation-based metrics: volumetric analysis of different organs on generated images and comparing them with the input condition
- CLIPScore: computes cosine similarity between CLIP embeddings of text descriptions and generated images
- BLIPScore: computes cosine similarity between BLIP embeddings of text descriptions and generated images
- LLMscore: leverages an LLM for creating a detailed caption at the level of an image and different objects and compares the generated caption with the input text descriptions

BLIP=bootstrapping language-image pretraining, CAS=classification accuracy score, CLIP=contrastive language-image pretraining, FID=Fréchet inception distance, IS=inception score, KID=kernel inception distance, LLM=large language model, MMD=maximum mean discrepancy, PSNR=peak signal-to-noise ratio, SSIM=structural similarity index.

Privacy preservation

Synthetic datasets offer a privacy-preserving solution to the challenges of sharing and utilisation of data in medical research.³⁴ Generative artificial intelligence anonymises sensitive patient information by generating realistic images that mimic biological characteristics of real patient data (both visually and in the model feature space) without direct replication of original data.³⁵ Such anonymisation enables the creation of datasets that can be shared and analysed without compromising patient privacy, which further opens up new avenues for collaborative research and facilitates the development of robust, privacy-compliant artificial intelligence models in medical imaging.

Versatility across tasks

Another key potential of image generation models, especially DDPMs, lies in their multifunctional nature,

Generative models trained on medical images can be

adapted and repurposed for various tasks beyond supplementing data; for example, features learned from an unsupervised image generation model can be leveraged for few-shot image segmentation, enabling accurate delineation of anatomical structures or pathologies with only 20 expert-annotated examples.³⁷ The same model without any further training can also be used for inpainting to create diverse training samples.³⁶ Similarly, generative models without any fine-tuning after initial training can be used for anomaly detection in medical images.^{37,38} This versatility extends the value of the synthetic datasets and their generator models, as a single model can be used for multiple downstream applications, streamlining research workflows, and reducing the need for task-specific data collection and model development.

Privacy preservation
Synthetic data uses
real patient data to
train

35 30423000 0 00000000

Panel 3: Summary of challenges, considerations, and future research directions for generative artificial intelligence and synthetic datasets in medical imaging

Data copying

Generative models can inadvertently reveal sensitive patient information when they reproduce images that closely resemble the original data.

Future research directions:

- Creating metrics to quantify the privacy risk of generated images
- Developing post-hoc data anonymisation methods
- Investigating the trade-off between image quality and privacy preservation

Identification of source dataset

Identifying specific datasets used to train generative models can be challenging, hindering the assessment of potential biases or limitations in the generated data.

Future research directions:

- Creating standardised reporting guidelines for synthetic medical imaging datasets
- Developing techniques for dataset fingerprinting in generative models
- Creating trusted third-party validation services for synthetic medical datasets
- Exploring methods for reverse-engineering model-training data

Interpretability and explainability

The complex nature of generative models makes understanding how these models learn and generate data a challenge. This understanding is necessary to build trust in the model outputs.

Future research directions:

- Implementing uncertainty quantification methods for stochastic prediction models
- Creating clinically relevant interpretability metrics
- Developing interactive visualisation tools for clinicians to explore model decisions
- Investigating the integration of domain knowledge into model explanations

Potential biases

Biases in the source datasets could be propagated or amplified in the generated data, leading to skewed research findings or discriminatory applications.

Future research directions:

- Creating benchmarks for evaluating fairness in medical imaging generative models
- Establishing multi-institution collaboratives to create demographically balanced training data
- Investigating the effect of data augmentation on bias reduction

Modelling complex biological phenomena

Advanced generative models can internalise complex biological phenomena through their training procedures, enabling the intricate physiological processes to be modelled and simulated.⁵⁹ This internalised world model can be leveraged for novel applications that extend beyond the downstream tasks discussed in the previous section. One striking example of this capability is the prediction of post-operative imaging appearances. When trained on a large corpus of paired prearthroplasty and postarthroplasty pelvic radiographs, these models generated highly realistic post-operative radiographs, simulating a well executed surgery.²⁸ Remarkably, domain-expert surgeons evaluated the generated postoperative images as more robust and anatomically accurate than their real counterparts, highlighting the potential of these models in serving as virtual surgical planning tools and educational resources.²⁸

Another compelling application of this internalised world model is the prediction of disease progression.³⁰ For instance, when given an initial brain MRI scan and information about the patient's treatment regimen, advanced DDPMs can generate a series of images that depict the potential progression of a brain tumour over time.³¹ By

learning the complex interplay between disease characteristics, treatment effects, and biological processes, these models can provide valuable insights into patient prognosis and aid clinical decision making.³¹

Challenges and considerations

Although derivative synthetic datasets and image generation models hold immense promise for medical imaging research, several challenges and ethical considerations need to be addressed to ensure their responsible and effective utilisation. Panel 3 summarises these challenges and proposes some future research directions to mitigate them.

Patient privacy and data copying

Although synthetic datasets can help to preserve patient privacy by generating anonymised data, concerns regarding potential data copying still exist.⁶⁰ If a generative model is trained on a specific dataset and can replicate images that closely resemble the original data, then the model might inadvertently reveal sensitive patient information. Copying happens when multiple copies of the image or captions are present in the dataset, which not only necessitates careful data curation,⁶¹ but also raises concern about the degree of

anonymisation achieved in the training data and the potential for reidentification. Unlike tabular data, medical images contain patient-identifying information embedded within the pixel values, thus posing unique challenges for anonymisation. For instance, facial features in brain MRIs or distinctive anatomical markers in radiographs might enable reidentification even when explicit patient identifiers are removed.^{63,63}

Researchers need to carefully assess the risk of data copying and implement measures to mitigate this concern, such as using differential privacy techniques or post-hoc data anonymisation.^{64,65} Advances made over the past 4 years in privacy evaluation metrics for synthetic data, such as membership inference attacks and similarity scores between real and generated samples, can help to quantify privacy risks. Additionally, emerging standards for synthetic content provenance, including the Coalition for Content Provenance and Authenticity (C2PA) and Google's SynthID, have been developed to label artificial intelligence-generated content, addressing both transparency and intellectual property concerns.⁶⁶

For more on C2PA, see <https://c2pa.org/>
For SynthID, see <https://deepmind.google/science/synthid/>

Identification of source dataset and disclosure

Transparency regarding the source datasets used to train generative models is crucial in ensuring the integrity and reproducibility of research findings. However, identifying the specific training datasets can be challenging, especially when models are trained on multiple proprietary sources or when researchers use pretrained models without full knowledge of their training data.⁶⁹ This insufficient transparency can hinder the ability to assess potential biases or limitations in the generated data. To address this gap, researchers should strive to document and disclose all source datasets used in the training process, enabling better understanding and validation of the derived data. Additionally, specific hyperparameters used for inference, specific class or prompt conditions, and every post-processing step involved in creating the synthetic dataset should be released along with the model or dataset release, to ensure reproducibility and applicability of the downstream work.⁶⁶ Dataset documentation guidelines, such as the STANDING Together guidelines published in 2024, should be adopted for synthetic data generation models.⁶⁹

چشمه‌های داده‌ها را مشخص کنید

منابع داده‌ها را مشخص کنید

Interpretability and explainability

As generative models become increasingly complex, their interpretability and explainability will become more challenging. Understanding how these models learn and generate data is crucial for building trust in their outputs and ensuring their safe and reliable use in medical imaging research. Although some specific explainability methods devised for generative models exist to ensure proper understanding of input text or to add uncertainty measures to the datasets, adaptation and evaluation of these methods in medical imaging remains restricted.^{70,71}

Potential biases

The use of synthetic datasets and generative models raises important bias considerations. The potential for biases in the source datasets getting propagated or amplified in the generated data is a key concern.⁷² If the training data are biased towards some demographics, pathologies, or imaging protocols, then the resulting generated data could perpetuate these biases, leading to skewed research findings or discriminatory applications.²⁵ For instance, historically, many medical imaging datasets have under-represented minoritised populations, resulting in artificial intelligence systems with likely differential performance levels across demographic groups.⁷³ When representation is low, generative models could struggle to capture a true distribution of these under-represented groups. However, a 2023 study suggests that newer generative models can arrive at meaningful representations from as few as 20 samples when the overall dataset is sufficiently large to capture high-level features.⁶¹ Mitigation strategies in this case include diversity-aware sampling during training, adversarial debiasing techniques, explicit fairness constraints in model objectives, and leveraging the few-shot fine-tuning capabilities of newer generative models.⁷⁴ Researchers need to actively assess and mitigate potential biases in the source data and regularly audit the generated data for fairness and representativeness.

منابع داده‌ها را با دقت بررسی کنید

Future directions

The field of generative artificial intelligence in medical imaging is evolving rapidly, and several key areas of research and development hold promise for advancing the capabilities and applications of synthetic datasets and image generation models. One crucial direction is the development of more robust and standardised evaluation frameworks that consider the unique challenges and requirements of medical imaging, including establishment of clinically relevant metrics, benchmark datasets, and challenges concerning comparative analysis and validation of different generative models.⁷⁵

Another important avenue is the exploration of novel architectures and training strategies, such as hybrid models combining physics-informed and statistical approaches with incorporation of domain-specific knowledge and constraints. Integration of generative models with other artificial intelligence techniques such as reinforcement and active learning could enable the creation of personalised and patient-specific datasets for precision medicine and targeted treatment planning.⁷⁶

Addressing the ethical and regulatory challenges surrounding the use of synthetic datasets and image generation models is essential to realise their full potential, and requires collaboration among researchers, clinicians, ethicists, and policy makers to develop guidelines and best practices for responsible use, data privacy, consent, and accountability. Regulatory bodies, including the US Food and Drug Administration (FDA) and the European Medicines Agency, will play a crucial role in establishing

STATE-OF-THE-ART REVIEW

Atherosclerotic Cardiovascular Disease Risk Prediction Models in China, Japan, and Korea

Implications for East Asians?



Patricia K. Nguyen, MD, MBA,^a Dong Zhao, MD, PhD,^b Tomonori Okamura, MD, PhD,^c Hyeon Chang Kim, MD, PhD,^d Nathan D. Wong, PhD,^e Eugene Yang, MD, MS^f

ABSTRACT

ပြည်ထောင်စုအတွင်းရှိ အာရှတိုက်ဒေသများတွင် အာရုံကြောစနစ်နှင့် အာရုံကြောစနစ်ဆိုင်ရာ ရောဂါများကို ကာကွယ်ရန်အတွက် အာရှတိုက်ဒေသများတွင် အာရုံကြောစနစ်နှင့် အာရုံကြောစနစ်ဆိုင်ရာ ရောဂါများကို ကာကွယ်ရန်အတွက်

The management of atherosclerotic cardiovascular disease (ASCVD) in the United States is currently based upon large epidemiological studies in primarily non-Hispanic White subjects. Although this strategy provides a uniform approach that is simpler to implement, it may result in inappropriately targeting certain Asian populations for treatment based on inaccurate ASCVD risk estimation. In this state-of-the-art review, we detail the similarities and differences in the prevalence of ASCVD and its risk factors among Chinese, Japanese, and Korean people living in the United States and in their native countries. We highlight the limitations of current risk calculators when applied to East Asian immigrants and summarize risk stratification approaches in China, Japan, and Korea. Our review underscores the need to disaggregate registry, cohort, and clinical trial data by East Asian subgroups, to actively engage these populations in research, and to initiate studies to better define ASCVD risk in East Asian people living in the United States. (JACC Asia. 2025;5:333-349) © 2025 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

East Asian people make up 20.7% of the world's population based on the latest estimates by the United Nations.¹ As the fastest growing immigrant population in the United States, Asians comprise approximately 7% of the U.S. population, with East Asians making up the largest Asian subgroup (eg, ~40%). In 2019, cardiovascular disease (CVD) claimed 5.2 million lives in East Asian countries.² Atherosclerotic cardiovascular disease (ASCVD), including ischemic heart disease, ischemic stroke, and peripheral arterial disease, was the leading cause of cardiovascular morbidity and mortality

among East Asian persons.² Declining fertility rates and increasing life expectancies will ensure that ASCVD continues to be one of the most common non-communicable, chronic diseases affecting East Asian persons for future decades.³

Notably, epidemiological characteristics of ASCVD vary among East Asian persons compared with Southern, Western, and Central Asian persons.² These differences exist even within East Asian subpopulations. The proportional mortality rate of CVD is as low as 25% in the Japanese and South Korean populations but as high as 40% in Chinese people,

From the ^aDepartment of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, California, USA; ^bDepartment of Epidemiology, Beijing Anzhen Hospital, Capital Medical University, Beijing Institute of Heart, Lung, and Blood Vessel Diseases, National Clinical Research Center for Cardiovascular Diseases, Beijing, China; ^cDepartment of Preventive Medicine and Public Health, Kelo University School of Medicine, Tokyo, Japan; ^dDepartment of Preventive Medicine, Integrative Research Center for Cerebrovascular and Cardiovascular Diseases, Yonsei University College of Medicine, Seoul, Korea; ^eHeart Disease Prevention Program, Division of Cardiology, University of California, Irvine, Irvine, California, USA; and the ^fDivision of Cardiology, Department of Medicine, University of Washington School of Medicine, Seattle, Washington, USA. The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the [Author Center](#).

Manuscript received September 24, 2024; revised manuscript received January 2, 2025, accepted January 13, 2025.

10. 2023. 10. 2023. 10. 2023. 10.

ABBREVIATIONS AND ACRONYMS

- ACC = American College of Cardiology
- ASCVD = atherosclerotic cardiovascular disease
- AHA = American Heart Association
- BMI = body mass index
- CAC = coronary artery calcium
- CHD = coronary heart disease
- CVD = cardiovascular disease
- HDL-C = high-density lipoprotein cholesterol
- KSH = Korean Society of Hypertension
- LDL-C = low-density lipoprotein cholesterol
- PCE = Pooled Cohort Equation
- PR = prevalence ratio
- TC = total cholesterol

1. 2023. 10. 2023. 10. 2023. 10. 2023. 10.
 2. 2023. 10. 2023. 10. 2023. 10. 2023. 10.
 3. 2023. 10. 2023. 10. 2023. 10. 2023. 10.

highlighting the need for both targeted and personalized, therapeutic strategies for East Asian subgroups.⁷

Whether the observed differences in the epidemiological characteristics in East Asian subgroups living in Asia extend to East Asian Americans is unknown. Because the number of East Asian immigrants living in the United States is growing rapidly, this question demands further investigation. In 2019 alone, 4.2 million East Asians immigrated to the United States, a sharp rise from the 250,000 who immigrated in 1960.⁴ Similar to those living in Asia, ASCVD is a major cause of morbidity and mortality. ASCVD risk varies widely between Asian American subgroups,⁵⁻⁷ underscoring the need to disaggregate data to refine risk assessment and optimize CVD management in East Asian Americans.

Cardiovascular clinical trials inclusive of East Asian Americans are limited. Like other

Asian subgroups,⁸ East Asian U.S. immigrants and their offspring are not adequately represented among major U.S. prospective CVD cohort studies, which form the basis of the American College of Cardiology (ACC)/American Heart Association (AHA) ASCVD Pooled Cohort Equation (PCE) as well as other risk calculators. Although the recently developed AHA PREVENT (American Heart Association Predicting Risk of cardiovascular disease EVENTS) risk calculator, derived from a meta-analysis of 25 data sets, has removed race/ethnicity altogether arguing their effects may be already reflected in socioeconomic data,⁹ the accuracy of this calculator has not been specifically validated in East Asian persons. As a result, current risk stratification models are not adequate to predict the development of ASCVD in East Asian Americans. Whether ASCVD risk assessment tools based on long-term cohort studies conducted in East Asian countries can accurately predict risk in East Asian Americans is unknown. Although recalibration can create a useful ASCVD prediction model and has been proposed to improve accuracy, it is preferable to develop and validate a model using local data if available. Even among countries classified with similar risk levels, there are considerable differences in incidence rates of CVD. For example, in Korea, SCORE2-AP recalibrated SCORE2 to the Asia-Pacific region and categorized risk into 4 groups (low, moderate, high, and very-high risk). Korea is included in the low-risk group, but this low-risk regional model also overestimates risk in Korea.¹⁰ Additionally, the distribution of CVD subtypes varies by region and period, so the coefficients may differ as

well. For example, although the coefficients for each predictor are somewhat similar, regional differences between models still exist. Thus, risk calculators that between models still exist. Thus, risk calculators that include the most recent local data would be the most appropriate. These knowledge gaps highlight the need for a more thorough understanding of variations in ASCVD disease burden and cardiovascular risk factor prevalence among East Asian people residing in the United States to improve risk stratification and management strategies.

In this review, we highlight the similarities and differences in the epidemiology, diagnosis, and treatment of ASCVD for individuals of East Asian origin who immigrated to the United States and their offspring ("East Asian Americans") compared with those living in East Asia ("East Asian natives"). We identify major knowledge gaps in our understanding of ASCVD risk and explore opportunities and strategies to close these gaps through future clinical and research initiatives. Because of the lack of data and smaller size of other East Asian countries, this review focuses on studies performed in East Asian persons living or originating from China, Japan, and the Republic of Korea (South Korea).

ASCVD PREVALENCE AND RISK FACTORS IN EAST ASIAN POPULATIONS

DEFINITION OF EAST ASIAN POPULATIONS.

Based on definitions provided by the United Nations, the geographic boundaries of the Asia-Pacific region include Turkey in the west, the Pacific Island of Kiribati in the east, the Russian Federation in the north, to New Zealand in the south.¹¹ Located in the Asia-Pacific region, major East Asian countries include the People's Republic of China (China, including the special administrative regions of Hong Kong, Macao, and Taiwan), Democratic People's Republic of Korea (North Korea), Republic of Korea (South Korea), Mongolia, and Japan. The major ethnic groups in East Asia include Han, Korean, and Yamato. Minor groups include Bai, Hui, Tibetan, Turkic, Manchus, Ryukyuan, Ainu, Zhuang, Mongols, and many other groups. In 2023, approximately 1.66 billion people live in East Asia, comprising 22% of the world's population.¹

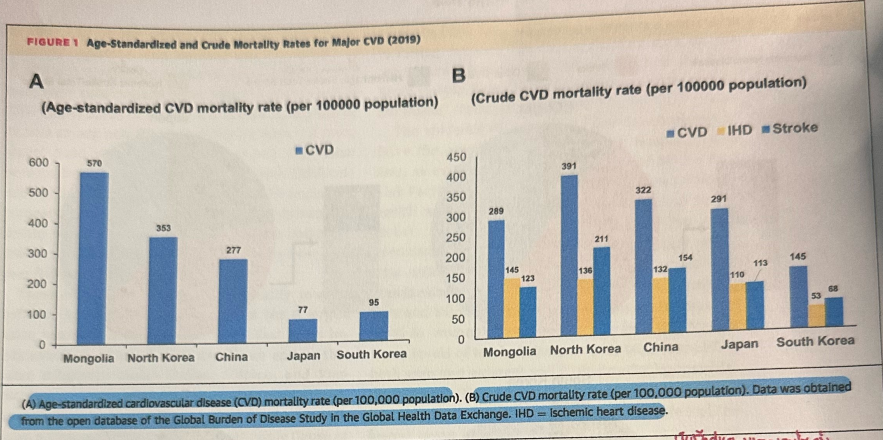
Additionally, millions of East Asian immigrants and their offspring reside outside of the geographical boundaries of East Asia. Since the mid-1960's, the number of immigrants from East Asia to the United States has risen sharply. According to calculations by the Migration Policy Institute of U.S. Census Data,⁴ between 2000 and 2019, the number of East Asian persons living in the United States has increased by

Score2-AP

2023. 10. 2023. 10. 2023. 10. 2023. 10.
 2023. 10. 2023. 10. 2023. 10. 2023. 10.
 2023. 10. 2023. 10. 2023. 10. 2023. 10.

2023. 10. 2023. 10. 2023. 10. 2023. 10.
 2023. 10. 2023. 10. 2023. 10. 2023. 10.
 2023. 10. 2023. 10. 2023. 10. 2023. 10.

Major and minor CVD mortality



(A) Age-standardized cardiovascular disease (CVD) mortality rate (per 100,000 population). (B) Crude CVD mortality rate (per 100,000 population). Data was obtained from the open database of the Global Burden of Disease Study in the Global Health Data Exchange. IHD = ischemic heart disease.

81% in the last 2 decades, outpacing Hispanic and non-Hispanic Black persons, whose numbers grew by 70% and 20%, respectively, during the same time frame. Immigrants from China and Korea accounted for 19% and 7% of the total Asian immigrant population, respectively.⁴

Korea (47%) had comparable rates. Although hemorrhagic strokes appeared to be slightly more frequent in East Asia overall, the proportion of hemorrhagic strokes varied significantly across the region. On average, 52% of stroke deaths were attributed to hemorrhage stroke, ranging from as low as 36% in Japan to as high as 50% in China.² Although peripheral vascular disease continues to be an important contributor to CVD, the availability of peripheral vascular disease mortality data in East Asian countries remains limited.

Stroke is 100,000 cases
52% of stroke deaths
Min: Japan max: China

EPIDEMIOLOGY OF ASCVD IN EAST ASIAN POPULATIONS LIVING IN ASIA AND IN THE UNITED STATES. An aging population together with the exponential rise in obesity and diabetes will continue to fuel the higher incidence of ASCVD among East Asian Americans and East Asian natives. Below, we summarize and compare the existing data on the epidemiology of ASCVD in East Asian natives and East Asian Americans.

National-level epidemiological data on ASCVD in East Asian Americans is limited, forcing us to rely on aggregated data and single center studies to estimate ASCVD prevalence and mortality rates among East Asian Americans. Based on the 2018 NHIS (National Health Interview Survey) conducted by the Centers for Disease Control and Prevention, the age-adjusted prevalence of stroke among persons 18 years of age or older was similar between non-Hispanic Asian (2.7%) and non-Hispanic White (2.7%) persons, whereas the prevalence of coronary heart disease (CHD) was lower in Non-Hispanic Asian persons at 4.4% compared with 5.7% in Non-Hispanic White persons.¹² Based on the 2019 National Vital Statistics Report conducted by the Centers for Disease Control and Prevention, the age-adjusted death rate for cerebrovascular disease was lower in non-Hispanic Asian vs non-Hispanic White persons (cerebrovascular disease: 29.3 of 100,000 vs 35.7 of 100,000).¹³ Greater differences were observed in heart disease

→ Stroke %
Asian non-Hispanic
Asian (2.7%) vs White (2.7%)
coronary heart disease (CHD)
Asian (4.4%) vs White (5.7%)
cerebrovascular (CVD)
Asian (29.3) vs White (35.7)
Heart (35.7) of 100,000
great difference

CVD rate: Min: North Korea, Max: South Korea | Ischemic non-hemorrhagic rate 87% vs stroke 13%

الدراسات التي أجريت في
الولايات المتحدة

ASCVD Risk Models in East Asia
USU BMI ≥ 25.0 kg/m²

of CHD and stroke among East Asian persons (ie, Chinese, Japanese, and Korean) was 2.07% (274 of 13,181) and 0.75% (100 of 13,181), respectively. These rates were substantially lower than the prevalence of CHD (3.9%, 2,805 of 72,701) and stroke (1.2%, 846 of 72,701) among non-Hispanic White persons.¹⁴ A more recent 2020 study that included 51,006 East Asian persons from Kaiser Permanente Northern California with CVD risk factors, the incidence of CHD (including coronary revascularization, MI, and cardiovascular death) was 4.46% in East Asian persons compared with 6.29% in non-Hispanic White persons over a 10-year follow-up period.¹⁵

Three studies have examined mortality rates of ischemic heart disease and stroke for the 6 largest Asian American subgroups, including East Asian immigrants from China, Japan, and Korea and 3 other Asian subgroups (Asian Indian, Filipino, and Vietnamese).¹⁶⁻¹⁸ Two of the earlier studies used mortality data from 34 states that reported deaths for each Asian subgroup in the 2003 to 2010 Multiple Cause of Death Mortality Database from the National Center for Health Statistics.^{16,17} A more recent study by Shah et al¹⁸ incorporated 2017 death rates and found that age-adjusted mortality rates from ischemic heart disease are lower for non-Hispanic Asian persons compared with Hispanic and non-Hispanic White persons; however, age-adjusted mortality rates from cerebrovascular disease are comparable between Asian and non-Asian persons. Analysis of Asian subgroups reveals significant differences in mortality rates with Korean persons having the highest mortality rates among the East Asian population but much lower than South Asian persons. More contemporary data on the incidence, prevalence, and mortality associated with ASCVD for Chinese, Japanese, and Korean Americans as well as other East Asian subgroups are critical to refinement of risk assessment tools and treatment strategies.

PREVALENCE OF ASCVD RISK FACTORS IN EAST ASIAN POPULATIONS LIVING IN ASIA AND IN THE UNITED STATES. Cardiovascular risk factors, including hypertension, diabetes, dyslipidemia, tobacco use, overweight/obesity, and lifestyle factors such as unhealthy nutritional practices and physical inactivity, contribute to the development of ASCVD. Understanding the prevalence of major CVD risk factors among East Asian immigrants is critical for predicting the risk of developing ASCVD. Accurate and current information on the prevalence of cardiovascular risk factors in East Asian immigrants is not only the basis for evaluation of which risk prediction models are most accurate for East Asians, but also

form the basis for recalibration of these models. The following sections summarize existing data on the prevalence of cardiovascular risk factors for East Asian people living in their native countries and in the United States, respectively.

The epidemic of unhealthy lifestyles continues to drive the prevalence of ASCVD risk factors in East Asia, as evidenced by the data collected by the NCD Risk Factor Collaboration (NCD-RisC),¹⁶ a network of health scientists that provides rigorous and timely global data on major risk factors for non-communicable diseases. According to the most recent data available, South Korea had the lowest age-standardized prevalence of hypertension (13.8% in men and 8% in women) and obesity (4.5% in men and 5.0% in women). Japanese people had the highest mean levels of total cholesterol (TC) (4.98 mmol/L in both men and women) and high-density lipoprotein-cholesterol (HDL-C) (1.46 mmol/L in men and 1.77 mmol/L in women), while Chinese people had the lowest mean levels of TC (4.56 mmol/L in men and 4.66 mmol/L in women) and HDL-C (1.20 mmol/L in men and 1.35 mmol/L in women). The U.S. population had much higher prevalence of obesity (36.0% in men and 38.1% in women) compared with the risk factor profiles of East Asian countries.^{14,19} Unfortunately, the numbers reported in the NCD-RisC reflect percentages in the United States as an entire population, which limit our ability to compare East Asian persons and their U.S. counterparts and further highlight the need for a global effort to disaggregate data by race and ethnicity. Moreover, the aforementioned percentages of obesity defined in the United States reflect body mass index (BMI) cutoffs for obesity defined in non-Asian persons. Asian American people are more likely to have central obesity, which remains a major risk factor for the development of type 2 diabetes, metabolic syndrome, and CVD at lower BMIs than non-Asian people. Evidence supporting racial differences in fat distribution between Asian and non-Asian people has prompted the lowering of the cutoff for obesity for Asian persons to ≥ 27.5 and ≥ 25.0 kg/m² rather than the cutoff 30 kg/m² for non-Asian persons by the World Health Organization and other regulatory bodies (eg, Japanese Society for the Study of Obesity, Korean Society for the Study of Obesity),²⁰ respectively. Thus, the estimates of obesity by the NCD-RisC may underestimate the true incidence of obesity in East Asian Americans and non-East Asian Americans.

The most recent estimates of cardiovascular risk factor prevalence in Asian Americans were published by Commodore-Mensah et al²⁰ and Koirala et al,⁸ who independently evaluated the heterogeneity of CVD

Handwritten notes in Hebrew script at the top of the page, including the word 'מחקר' (research) and 'אסיה' (Asia).

risk factors among U.S. Asian immigrants using a cross-sectional analysis collected from NHIS between 2010 to 2018. They compared the prevalence of CV risk factors between 474,968 White adults and 33,973 Asian immigrants including an estimated 9,753 individuals from the Asia region (including Asia Minor, China, Democratic People's Republic of Korea [North Korea], Japan, Mongolia, and Republic of Korea [South Korea]). The remaining Asian immigrants included in the analysis were from South Asia (Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan, Sri Lanka) and Southeast Asia (Burma, Cambodia, Indonesia, Laos, Malaysia, Philippines, Singapore, Thailand, and Vietnam) regions. Notably, the geoscheme for Asia used by the authors differs from the geoscheme defined by the United Nations and adopted by the writers of this review.

Compared with non-Hispanic White persons, East Asian immigrants were less likely to have hypertension (20.8% vs 34.3%), diabetes (6.3% vs 9.6%), dyslipidemia (21.9% vs 31.7%), or overweight/obesity (47.1% vs 64.1%). Rates of smoking were not significantly different between non-Hispanic white and East Asian Americans. Compared with South Asian immigrants, East Asian immigrants were more likely to have hypertension (20.8% vs 16.6%), but less likely to report having diabetes (6.3% vs 8.6%) or overweight/obese (47.1% vs 71.54%). The rate of dyslipidemia between East Asian immigrants and South Asian immigrants was comparable (21.9% vs 20.8%). Compared with Southeast Asian immigrants, East Asian immigrants were less likely to have hypertension (20.8% vs 31.2%), diabetes (6.30% vs 11.3%), dyslipidemia (21.9% vs 33.41%), or overweight/obesity (47.1% vs 59.8%). Physical inactivity was equally present in all Asian subgroups with over 50% of patients not meeting physical activity guidelines. Compared with non-Hispanic White persons, East Asian persons were more likely to be physically inactive, defined by the proportion of the population not meeting ACC/AHA guidelines for exercise and physical activity (50.8% vs 48.3%). Even after adjustment for age, sex, and socioeconomic factors, East Asian persons were more likely to report physical inactivity (prevalence ratio [PR] 1.14, 95% CI: 1.09-1.19) but less likely to report hypertension (PR 0.72, 95% CI: 0.67-0.77), overweight/obesity (PR 0.83, 95% CI: 0.80-0.87), diabetes mellitus (PR 0.80, 95% CI: 0.70-0.91), high cholesterol (PR 0.83; 95% CI: 0.7700,90), and current smoking (PR 0.53, 95% CI: 0.46-0.60) than their non-Hispanic White counterparts. Unfortunately, this analysis was unable to disaggregate subjects by their country of origin—a critical variable to estimate the incidence

and prevalence of CVD risk factors among East Asian American subgroups and essential for creating a foundation for accurate risk estimation.

THE IMPACT OF ACCULTURATION AND ENVIRONMENTAL EFFECTS OF ASCVD RISK PROFILES. It is fundamental to understand the potential differences in CVD risk to understand the potential differences in CVD risk factors and ASCVD mortality profiles of East Asian natives and East Asian Americans before we can develop an accurate risk assessment strategy. To better account for acculturation and environmental effects, risk factor tools for East Asian immigrants may need to consider the impact of immigration history and generational status on risk factor profiles. Few studies have directly compared the profiles of CVD risk factors between Asian Americans and Asian immigrants. In the REACH (Reduction of Atherothrombosis for Continued Health) registry that enrolled close to 3,000 Chinese patients with atherothrombotic disease living outside and inside mainland China, rates of obesity, hypertension, hypercholesterolemia, and diabetes were lower in mainland Chinese people than those in China Hong Kong, China Taiwan, Singapore, Western Europe, and North America.²¹ In a recent cross-sectional analysis of the Japanese population, Hirooka et al²² showed that Japanese Americans reported less sedentary behavior but were also less likely to engage in physical activity and formal exercise. Although the findings are based on studies conducted a decade ago, they highlight the need to account for environmental and cultural factors to create an accurate risk stratification tool for East Asian Americans. A recent study found that acculturation was associated with a heterogeneous pattern of CVD risk factors among Asian American subgroups, highlighting the need for more research to better understand these differences and to guide culturally concordant interventions.²³

CURRENT STATE OF ASCVD RISK CALCULATORS FOR EAST ASIAN POPULATIONS

In the 2018 ACC/AHA Guideline on the Management of Blood Cholesterol,²⁴ the PCE was used for risk assessment for primary prevention to guide the eligibility for statin therapy. It was developed by combining 4 U.S. community cohort studies, including the ARIC (Atherosclerosis Risk in Communities), CHS (Cardiovascular Health Study), CARDIA (Coronary Artery Risk Development in Young Adults), and Framingham Study. It predicts the 10-year risk of ASCVD (including nonfatal and fatal CHD and stroke). The risk factors used in this model include sex, race (non-Hispanic White, African American, or other),

CENTRAL ILLUSTRATION East Asian Cardiovascular Risk Calculators

Risk Factors	China	Japan	Korea*	ACC/AHA PCE	AHA PREVENT	ESC SCORE2
Alcohol use	○	◐	◐	○	○	○
Blood pressure	●	●	●	●	●	●
Body mass index (BMI)	◐	◐	◐	○	●	○
Cholesterol	●	●	●	●	●	●
Chronic kidney disease	◐	◐	◐	●	●	●
Diabetes mellitus	●	●	●	●	●	●
Family history of CAD	◐	○	◐	●	○	●
Physical activity	○	◐	◐	●	○	●
Tobacco use	●	●	●	●	●	●

● = Included in all risk calculators; ◐ = Included in some risk calculators; ○ = Not included in risk calculators;
 ● = Risk Enhancer (ACC/AHA) or Risk Modifier (ESC)
 * = Korea denotes the Republic of South Korea

Nguyen PK, et al. JACC Asia. 2025;5(3):333-349.

Comparing cardiovascular risk calculators: China, Japan, Korea, United States, and Europe. Filled circles: included in all risk calculators; half-filled circles: included in some risk calculators; open circles: not included in risk calculators; circles with bold outlines: risk enhancer (American College of Cardiology [ACC]/American Heart Association [AHA]) or risk modifier (European Society of Cardiology [ESC]). *Korea denotes the Republic of South Korea. CAD = coronary artery disease.

age, systolic blood pressure, treatment for hypertension, TC, HDL-C, smoking, and diabetes. Notably, the cohorts used in the PCE included very few Asian subjects. Moreover, these cohorts were initiated in the 1980s or earlier when CVD risks were higher or have tended to overestimate ASCVD risks caused by this secular effect. Also, as detailed in the following text, estimated 10-year ASCVD risk is generally lower in Asian American persons compared with non-Hispanic White persons. Accordingly, Asian-specific risk stratification calculators linked with clinical guidelines are needed to bridge these gaps (Central Illustration).¹⁸

ASCVD RISK PREDICTION IN CHINA. Several risk scores have been evaluated for ASCVD risk prediction in China. In 2003, the CMCS (Chinese Multi-provincial Cohort Study) of 30,121 adults (ages 35 to 64 years) evaluated the performance of the Framingham CHD risk score with equations derived from CMCS.²⁵ The investigators found that when compared directly and after recalibration, the original Framingham equation

significantly overestimated absolute CHD risk in the CMCS cohort, which was mainly driven by differences in the mean CHD risk and the levels of major risk factors between the 2 cohorts. Specifically, the 10-year CHD event rates were 8.0% and 2.8% in Framingham men and women, respectively, compared with 1.5% and 0.6% in the CMCS men and women. With respect to cardiac risk factors, men and women in the Framingham cohort had higher TC and lower HDL-C, respectively, than those in the CMCS cohort. However, men in the Framingham cohort had lower and women had higher smoking rates in comparison with men and women in the CMCS cohort. Importantly, a recalibration of the Framingham Risk Score (FRS) equation (ie, the mean risk of CHD and mean levels of risk factors derived from Framingham cohort), performed by replacing the corresponding estimations from CMCS cohort, substantially improved the accuracy of risk prediction.²⁵ Even after recalibrating Framingham risk calculators, the China-MUCA (Multicenter Collaborative Study of Cardiovascular Epidemiology)

CMCS: 30,121 adults (ages 35 to 64 years) evaluated the performance of the Framingham CHD risk score with equations derived from CMCS.²⁵

improved to 0.93 by adding coronary computed tomography angiographic (CTA) plaque information (both $P < 0.001$). In addition, the U.S. Multiethnic Study of Atherosclerosis showed CAC strongly predicted future ASCVD events for Chinese American participants (as well as other racial/ethnic groups) over 10 years, with those having CAC scores >100 identified to have an ASCVD risk $\geq 7.5\%$ making them eligible for initiation of statin therapy.³⁸ Recently, a new risk classification system for secondary prevention was issued by the Chinese Society of Cardiology in an expert consensus statement. Instead of considering all patients with ASCVD as a very high-risk group in previous guidelines, this statement further identified a subset at extreme ASCVD risk for more intensive lipid lowering treatment.³⁹

Although country-specific risk assessment tools are readily available in China, significant challenges remain. Most of the cohort studies collected baseline information in the 1990s, but there have been changes in risk factor prevalence among more contemporary target populations. Risk prediction models need to be recalibrated, and some studies have started to evaluate these older equations with data collected from large contemporary cohorts. For instance, Liu et al⁴⁰ recently applied the PCE and China-PAR algorithms to an electronic health record cohort of 226,406 participants and showed that although both models had good discrimination, the China-PAR model substantially underpredicted risk, and the PCE overpredicted risk in men and underpredicted risk in women, although recalibration improved this. Additionally, although most of the risk enhancing factors increased ASCVD risk in observational studies, few studies in China showed that these risk-enhancing factors, except for CAC score, could significantly improve performance of risk calculators. Several new risk factors have been considered as potential contributors to residual risk in Chinese people, including small density LDL, cholesterol-overloaded HDL particles,⁴¹ and higher circulating levels of PCSK9.⁴⁰ However, the practical use of these new risk factors in individual based ASCVD risk prediction in Chinese is uncertain. Although guidelines generally recommend risk assessment to drive prevention and treatment strategies, adherence to these recommendations are low in clinical practice.^{42,43} especially in primary care settings.⁴³ Furthermore, cost-effectiveness and feasibility of more complex and accurate risk assessment tools for clinical application are unknown. Moreover, no studies to date

have evaluated if ASCVD risk prediction calculators developed from Chinese cohorts can be applied to people living in other East Asian countries or Chinese persons living abroad, including Europe and the United States.

ASCVD RISK PREDICTION IN JAPAN. Since 1960, when the first epidemiologic studies of CVD were performed in Japan, high stroke mortality rates but low CHD mortality rates were observed.⁴⁶ Although stroke mortality has significantly improved in subsequent decades, CHD mortality remains low compared with Western populations. The 2012 Japan Atherosclerosis Society (JAS) Guidelines for Prevention of Atherosclerotic Cardiovascular Diseases was the first guideline to introduce absolute risk assessment of ASCVD and used the NIPPON DATA80 risk chart to predict 10-year CHD mortality.⁴⁴ Similarly, the 2019

Japanese Society of Hypertension incorporated absolute risk assessment as a reference to inform blood pressure treatment thresholds and management.⁴⁵

In the 2017 JAS guideline,⁴⁶ the Suita score was able to accurately estimate the absolute incidence of CHD by incorporating demographics and risk factors including age, sex, smoking, blood pressure level, HDL-C, LDL-C, impaired glucose tolerance, and family history of premature CHD. The Suita score was chosen from 10 different published risk prediction scores in Japan where internal validation was carefully performed. Although attempts were made to validate the predictive model for the guidelines in an external population, even within Japan, cardiovascular disease risk differs between urban and nonurban areas. Thus, it was extremely difficult to set a standard population for external validation.⁴⁷

Based on the algorithm for ASCVD risk assessment, individuals are first screened to determine if they should be categorized as primary or secondary prevention candidates. They are also evaluated for high-risk comorbidities including diabetes, chronic kidney disease, noncardiogenic stroke, and peripheral artery disease. If none of these conditions are present, the Suita score is calculated and individuals are stratified into low-, moderate-, and high-risk categories. The LDL-C targets in primary prevention are set at <160 mg/dL for low risk, <140 mg/dL for moderate risk, and <120 mg/dL for high risk, respectively. The LDL-C target for patients with established CHD is <100 mg/dL, but if they also have a history of familial hypercholesterolemia or acute coronary syndrome, then a lower target LDL-C level of <70 mg/dL should be considered. Patients with both diabetes

NIPPON DATA80

2019 JAS
15 predict 10 year CHD mortality

Suita score

TABLE 2 ASCVD Risk Assessment Tools in Japan

First Author	Study Population	Predictors (Other Than Age and Sex)	Outcomes	Duration of Risk Prediction, y	Validation	Risk Tools
Iso et al ⁵⁴	10,334 men and 19,542 women aged 40-69 y	hsCRP, SBP, DBP, DM, HDL, LDL, TC, alcohol, BMI, smoking habits, and medications	ASCVD	10	Internal	Odds ratio
Nippon Data80 et al ⁴⁴	4,098 men and 5,255 women age 30 y or higher	SBP, smoking, TC and glucose levels	ASCVD	19	Internal	Risk charts, Cox proportional hazard models
Hirai et al ⁴⁷	14,072 men and 21,307 women Women aged 40-79 y	SBP, DM, HDL, LDL, CKD stage, smoking habits	CHD	3	Internal	Suita risk score
Honda et al ⁵⁶	1,026 men and 1,428 women aged 40-84 y	SBP, DM, HDL, LDL, proteinuria, smoking habits, and regular exercise.	ASCVD	24	Internal	Cox proportional hazard model

CKD = chronic kidney disease; hsCRP = high sensitivity C-reactive protein; other abbreviations as in Table 1.

and ASCVD should also be treated to an aggressive LDL-C goal of <70 mg/dL.

It is important to note that the absolute ASCVD risk estimated by the Suita score only includes CHD and not stroke, unlike the PCE and SCORE2 risk calculators that include both. In Japan, however, cerebral hemorrhage accounts for a high proportion of strokes, whereas the percentages of lacunar, cardioembolic, and atherothrombotic strokes are almost equivalent,^{48,49} with the first 2 not being associated with hypercholesterolemia.⁴⁸ Although Japanese people have a high incidence of stroke, they have a relatively low proportion of stroke phenotypes associated with dyslipidemia, and therefore it has been difficult to use "stroke" when setting lipid management targets. In contrast, hypercholesterolemia is a strong risk factor for atherothrombotic infarction. Accordingly, the new 2022 JAS guideline, uses a recently published risk score from the Hisayama study that predicts incidence of the combined outcome of CHD and atherothrombotic cerebral infarction for individual risk assessment.⁵⁰ A summary of ASCVD calculators used in Japan is shown in Table 2.

The new JAS guideline also reviews the application of imaging and physiological measurements for sub-clinical atherosclerosis detection and their additive value to risk prediction models. These markers include the following: 1) periventricular hyperintensity or deep and subcortical white matter hyperintensity by brain magnetic resonance imaging; 2) carotid intima-media thickness and plaque characteristics; 3) CAC; 4) pulse wave velocity; 5) cardio-ankle vascular index; and 6) ankle-brachial index (ABI). The authors reviewed whether these potential risk enhancers improve the predictive ability of

traditional risk factors for the development of ASCVD in a primary care setting. A meta-analysis of Japanese studies showed that brachial-ankle pulse wave velocity (baPWV) improves risk prediction,⁵¹ but the model was not generated from a Japanese population. Therefore, the applicability to set baPWV cutoff values for Japanese people remains elusive. Other markers, including magnetic resonance imaging, carotid intima-media thickness, and plaque characteristics, and CAC score still lack evidence for risk prediction, and ABI did not enhance risk prediction models.

For high-risk individuals, such as stable patients with suspected CHD, the addition of CAC score to a risk prediction model that employs traditional risk factors significantly improved CHD risk classification.⁵² Given the high availability of computed tomography (CT) scanners in Japan (highest per capital in the world, 115.7 per million people in 2020),⁵³ many studies have been conducted to probe the additive value of coronary CT angiography and CAC score for symptomatic patients. However, use of these imaging modalities for comprehensive risk assessment have not been performed. This is also true for other imaging and physiological measurements, such as magnetic resonance imaging, carotid intima-media thickness, plaque characteristics, baPWV, and cardio-ankle vascular index. Biomarkers such as C-reactive protein,⁵⁴ small dense LDL,⁵⁵ and modified lipoproteins (LOX-1 ligands containing apolipoprotein B,⁵⁶ malondialdehyde-modified LDL⁵⁷) have also been validated in epidemiological and clinical studies in Japan and have been found to refine ASCVD risk when added to classical risk factors. However, they have only been assessed at the individual level, and

Hisayama
Study Risk
was for CHD and
atherothrombotic cerebral
infarction via transcranial
DUS

TABLE 3. ASCVD Risk Assessment Tools in Korea

First Author	Study Population	Predictors (Other Than Sex and Age)	Outcomes (Follow-Up)	Duration of Risk Prediction, y	Validation	Risk Tools Provided
Jee et al ⁶²	777,502 men and 446,238 women aged 30-84 y	SBP, TC, diabetes, smoking, activity, BMI, alcohol	Stroke	13	Internal	Equation, risk score table
Jee et al ⁶⁸	164,500 men and 104,310 women aged 30-74 y	BP, TC, diabetes, smoking, (HDL-C, LDL-C, TG; optional)	CHD	12	Internal	Equation, HR
Park et al ⁶⁵	3,135 men and 2,425 women aged 43-63 y	Hypertension, diabetes, SBP, DBP, TC, HDL-C, LDL-C, smoking, atrial fibrillation, white blood cells, creatinine, hemoglobin A1c	ASCVD	3	Internal	HR
Jung et al ⁶¹	119,715 men and 80,295 women aged 40-79 y	SBP, TC, HDL-C, DM, smoking, antihypertensive	ASCVD	10	Internal	Equation
Cho et al ⁶³	152,076 men and 144,354 women aged 40-79 y	BMI, SBP, DBP, TC, FBG, smoking, activity	ASCVD	10	Internal and External	Machine learning algorithm
Cho et al ⁶⁵	222,998 individuals aged 40-70 y	SBP, DBP, TC, HDL-C, LDL-C, TG, FBG, GGT, GFR, proteinuria, BMI, WC, smoking, alcohol, activity, antidiabetic, antihypertensive, statins, family history of CHD, family history of stroke	ACS + stroke	8	Internal	HR, nomogram
Choi et al ⁶⁶	10,412,947 men and 11,168,849 women Aged 40-79 y	SBP, TC, HDL-C, diabetes, smoking, antihypertensive, (+8 optional markers)	ASCVD	5	Internal	Machine learning algorithm
(Unpublished)	43,798 individuals aged 40-79 y	SBP, TC, HDL-C, diabetes, smoking, antihypertensive	ASCVD	10	Internal and external	Equation, risk chart

ACS = acute coronary syndrome; DBP = diastolic blood pressure; FBG = fasting blood glucose; GFR = glomerular filtration rate; GGT = gamma-glutamyl transferase; TG = triglycerides; other abbreviations as in Table 1.

without validation in population-based cohorts, are difficult to incorporate into risk prediction models. They have limited clinical application since they are not routinely measured at clinic visits.

Since the establishment of universal health coverage for all Japanese people in 1961, there are limited incentives to use imaging modalities such as CT scans to enhance systematic risk prediction or to evaluate the cost-effectiveness of these tests, especially for high-risk individuals. Future directions should focus on determining whether biomarkers or subclinical atherosclerosis detection can improve risk assessment for low- and intermediate-risk people.

ASCVD RISK PREDICTION IN THE REPUBLIC OF KOREA

(SOUTH KOREA). Early Korean studies attempted to recalibrate risk assessment tools developed in Western populations because they overestimated the absolute risk of ASCVD in the Korean population.⁵⁸⁻⁶¹ Later studies used Korean data to develop risk prediction models for stroke, coronary artery disease, or ASCVD.⁶¹⁻⁶⁸ The National Health Insurance Service, which operates a free health screening program for all Korean adults, provides risk

assessment for CVD and diabetes based on individualized data. Private health screening centers and research institutes are also using disease prediction on common chronic diseases, including ASCVD, diabetes, cancer, and Alzheimer's disease. However, ASCVD risk prediction assessment has not yet been adopted as the primary tool for risk assessment in clinical practice guidelines.⁶⁹ A summary of available ASCVD risk assessment tools in Korea is provided in Table 3.

The Korean Society of Lipid and Atherosclerosis recently released the fifth edition of the Korean Guidelines for the Management of Dyslipidemia. Compared with the fourth edition, the revised guidelines lowered the LDL-C target level for patients with CHD from <70 to <55 mg/dL.^{70,71} The risk classification of people with diabetes was also subdivided based on recent Korean data,⁷² with the LDL-C target value lowered in selected high-risk diabetes groups.⁷³ Lower LDL-C targets are also recommended for patients with significant carotid artery stenosis or abdominal aortic aneurysm. Although risk classification based on the presence of specific risk factors remained, it is noteworthy that the fifth KSOLA

guidelines introduced, for the first time, an optional recommendation to use a specific ASCVD prediction model. The guidelines noted that "ASCVD risk scores based on the Korean Genome and Epidemiology Study (KOGES) cohort can be used as a risk enhancer in low- or moderate-risk groups."⁷¹ The reason why it is recommended to be used only in low- to moderate-risk groups is that the absolute incidence of ASCVD is very low in Koreans, so if they rely only on the ASCVD prediction model, some patients high-risk groups with major modifiable risk factors may not be identified for treatment.⁷¹ The Korean Society of Hypertension (KSH) released the 2022 KSH Guidelines for the Management of Hypertension without significant changes in risk classification compared with the previous 2018 KSH guideline. Assessment of CVD risk and blood pressure targets are still based on the presence or absence of clinical ASCVD or the number of established cardiovascular risk factors. The KSH guidelines acknowledge the importance of accurate individual risk assessment for selecting the best ASCVD prevention strategy and recommend the use of risk assessment tools but does not endorse a specific risk prediction model.^{71,74}

There are several reasons why ASCVD risk prediction is mainly of academic interest and is not widely accepted in clinical practice guidelines in Korea. Because most of the ASCVD risk prediction models developed in Korea are based on data from limited health screening centers, there is criticism about the lack of adequate representation. Additionally, the ASCVD risk prediction models were mostly validated internally without external validation in a diverse cohort population. Furthermore, there is insufficient evidence demonstrating the clinical efficacy and cost-effectiveness of preventive pharmacological interventions. These deficiencies make it difficult to recommend an ASCVD risk prediction model in daily clinical practice.⁶⁹ Risk factors in the Korean ASCVD risk prediction models are similar to those used in other countries. The major risk predictors include sex, age, blood pressure, smoking, diabetes, and TC. It is notable that many of the Korean prediction models only use TC instead of LDL-C and HDL-C because Korea's general national health screening program often only measures TC levels.^{61-63,68} In addition, the incidence of CHD is low, and LDL-C and TC levels are strongly correlated to risk in the Korean population while the predictive power of TC is not inferior to LDL-C plus HDL-C levels.^{56,66} Efforts to use calculators, including the SCORE2 (Systemic Coronary Risk Evaluation 2) and SCORE2-OP (SCORE2-Older persons), resulted in an underestimation risk in young Korean men and women (aged 40-59 years)

and overestimation in older individuals, highlighting the need for a region-specific calculator.⁷³

The most recently developed Korean ASCVD prediction model is based on data from over 150,000 participants in the Korean Genome and Epidemiology Study. The risk predictors are similar to previous models, but this model includes large-scale prospective cohort data and was validated both internally and externally. In addition, the new model may have better clinical utility because the incidence of CVD in Koreans is rapidly increasing. Although ASCVD risk scores are not officially adopted in clinical practice guidelines, many CVD risk prediction studies have been published in Korea utilizing new biomarkers and imaging modalities.⁷⁵⁻⁸¹ Several of them have demonstrated improvements in ASCVD predictive power using CAC or CTA and employing machine learning methods. Earlier studies cross-sectionally compared CAC and Western-derived ASCVD risk scores among asymptomatic individuals and reported that high CAC levels are also observed among some Korean persons with low to moderate ASCVD risk.^{75,77} These findings indirectly support the usefulness of CAC-based risk classification. Later studies added more direct evidence of CAC score in the risk classification of asymptomatic Korean adults. The CONFIRM (COronary CT Angiography Evaluation For Clinical Outcomes International Multicenter) study evaluated the clinical utility of CAC and coronary CTA in Korean as well as European/American populations.⁷⁴ The CONFIRM study reported that CAC improved risk stratification and provided incremental value beyond FRS for predicting major adverse cardiac events.⁷⁴ Another Korean study reported that the addition of CAC to FRS improved risk prediction for CVD mortality in young adults but not in older individuals.⁸² The predictive value of coronary CTA was also evaluated in combination with CAC. In a 2-year follow-up study of Koreans participating in health screens, both CAC and coronary artery stenosis on coronary CTA are independent predictors of CVD outcomes. However, when CAC and coronary CTA stenosis were evaluated simultaneously, CVD risk was associated only with stenosis on coronary CTA but not with CACS.⁷⁶ In the CONFIRM study, coronary CTA provides incremental predictive value for asymptomatic individuals with moderate CAC scores (100-400), but not for lower or higher CAC scores.⁷⁸ There are also studies that evaluated the usefulness of funduscopy.^{80,81} A Korean study found that virtual assessment of CAC estimated from deep learning analysis of retinal photographs is comparable to CT-measured CAC in predicting CVD events and improves current risk stratification for cardiovascular

KSH updates risk
management
ASCVD tool
management

Why is ASCVD
risk prediction
of academic interest
to clinicians?

High
Study
management
of the
information