Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Schwarz CG, Kremers WK, Therneau TM, et al. Identification of anonymous MRI research participants with face-recognition software. N Engl J Med 2019;381:1684-6. DOI: 10.1056/NEJMc1908881

Contents

1 Supplementary Methods Details

1.1 Study Design

We considered the possible scenario where a participant is known to exist within a publicly-shared de-identified research study including brain magnetic resonance images (MRI), and someone seeks to identify them by using face recognition with (identified) face photographs from social media or similar sources. In addition to public research data sets, access to de-identified medical images could also be obtained via re-purposed clinical data shared with data brokers and analytics enterprises, or by having or gaining access to computer systems being used to store, transfer, or analyze de-identified data (e.g. cloud-based storage/analytics).

Knowledge or suspicion that an individual may exist in the data set could stem from many sources. For example, an individual may learn that a relative, friend, or coworker is enrolling in a study/trial. A participant may even wish to identify themselves in order to gain information such as non-disclosed biomarker statuses, or whether they are receiving treatment vs. placebo. De-identified data typically include age, sex, and approximate location (via study site location), which could be leveraged to dramatically reduce the pool of potential matches. Individuals' participation may also be inferred with high confidence when a study enrolls exclusively very small and specific populations, such as familial cohorts, the oldest-old, religious orders, mutation carriers, professional athletes, patients with very rare conditions, etc. Big data analytics firms could also infer participation by cross-referencing de-identified medical data with other databases, such as cellular location tracking data combined with study visit dates[1], and many such firms already maintain large social media photo databases they could use with face recognition.

Motives for identifying a participant could be as innocent as a family member who wants to know more about their loved one's prognosis, as invasive as corporations mining medical records to sell targeted advertising, or as malicious as agents seeking information to discredit or blackmail political or corporate opponents. Identifying a participant in this way would result in a privacy breach of all their study-associated health information that could include not only the participant's name but also diagnoses, cognitive testing, genetic data, biomarkers, results of other imaging, and participation in studies and trials.

We evaluate these scenarios by testing the rate at which someone with access to five photographs of a target individual could successfully match them to their MRI using automated face recognition.

1.2 Data Set

Participants: We recruited 84 Mayo Clinic volunteers, age 34-89 (mean=62) with recruitment stratified by sex and age-decade (see Figure S1), to allow photography of their face. All participants were clinically unimpaired and had previous head MRI within 3 months as part of their existing enrollment in either the Mayo Clinic Study of Aging[2,3] or the Alzheimer's Disease Research Center studies. All participants provided informed consent for this specific study, which was approved by the Mayo Clinic Internal Review Board.

MRI Acquisition: Sagittal 3D Fluid-Attenuated Inversion Recovery (FLAIR) head MRI were acquired using Siemens Prisma scanners with resolution 1.0x1.0x1.2mm, repetition time=4800ms, echo time=441ms, and inversion time=1650ms. We did not design this imaging sequence/protocol for face imaging; it was designed for the general assessment of brain aging and dementia pathologies, and it is identical to the protocol used in the Alzheimer's Disease Neuroimaging Initiative (ADNI) 3[4,5], a large publicly available imaging data set (whose protocols are often used as a template when designing other multi-site imaging studies). During earlier iterations of this study, we also obtained comparable face matching performance when using older GE T1-weighted scans of a subset of the same individuals using protocols identical to those in ADNI2 (data not presented).

Photographs Acquisition: We photographed each participant under indoor lighting conditions using a standard iPad (Apple Inc., Cupertino, CA; models Air 2 and 6th generation) with their face looking in each of five directions: straight ahead, slightly left, slightly right, slightly up, slightly down (total 420 photos of 84 participants). We designed this protocol in order to obtain multiple photographs of each participant that are not redundant (e.g. are not without perceptible changes), but are within the limits of a single session with minimum time expenditure and burden to participants. For the left/right/up/down directions, participants were instructed to look approximately 10 degrees in each direction, such as would be typically available in group photos or candid photos in social situations. Photos were manually cropped loosely around the head and converted to grayscale to better match MRI, which does not capture color. Our image cropping retained the entire head/hair/ears/etc. and removed only distant background and torso, in order to reduce unnecessary image size to speed up repeated image uploading during testing.

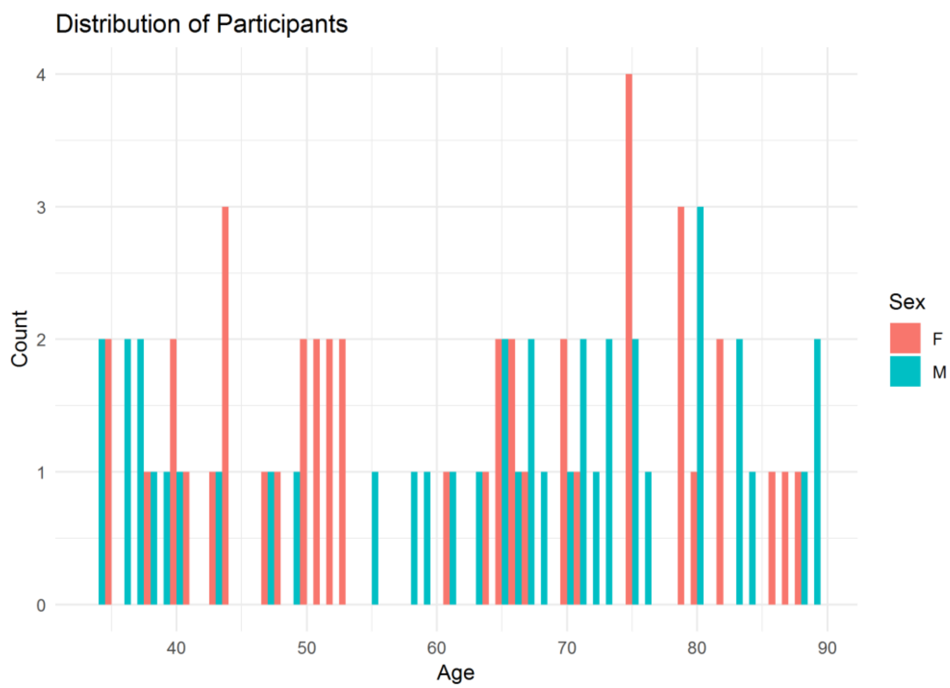

Figure S1: Distribution of age and sex of participants. All participants were clinically unimpaired, and recruitment was stratified by sex and age-decade.

1.3 Methods for creating MRI- derived face reconstruction images

As a pre-processing step before creating 3D surface reconstructions, we first applied an in-house fully-automated algorithm to detect and correct for aliasing artifacts (parts of noses/chins that appeared to be disconnected and floating behind the head), which we developed and tested using an independent MRI data set. Our algorithm used Otsu's method[6] to calculate a threshold separating tissue from air, then used a connected-components algorithm to identify contiguous regions. The largest contiguous region was assumed to be the head. If any disconnected regions were located behind the head, the largest was assumed to be the nose. Each row (in the superior/inferior direction) of the detected nose was then translated to the anterior edge of the detected head region (which varied for each slice in the left/right direction), and the intensity in its previous location was set to zero. Most of the images in this data set did not require this correction, but it was automatically applied when needed. We also examined the subset of MRI that were not correctly matched, and the images that required this correction were not qualitatively over-represented.

We then converted the .nii (volume) MRI files to .gii (surface reconstruction) files using the *nii_nii2gii.m* script from the surf_ice project repository using a mesh reduction fraction of 0.5 and a Gaussian smoothing filter with a 2 voxel diameter. For an isosurface threshold, we used Otsu's method[6] and multiplied this result by 0.3.

After the above preprocessing, we created surface render images (synthetic photograph-like 2D images of the 3D model) of the face from each MRI using Surf Ice[7]. We present visual examples of these MRI facial render images and corresponding photos in Figure 1B in the primary document. To simulate a variety of natural conditions, we rendered each face as a matte surface and saved a total of 81 2D render image files under varying directions and lighting positions. More specifically, we used the "Phong_Matte" shader with parameters: Edge=0.3, Ambient Occlusion=0, Diffuse=1, DiffuseRough=1, MeshColor=(200,200,200), BackColor=(0,0,0). All render images were saved as .png image files at 300x300 pixels, after scaling the rendered face by 95% in the left/right direction. The ranges of rotation parameters were as follows: yaw [170,190] in increments of 10, pitch [-5,5] in increments of 5. The ranges of lighting positions were as follows: yaw [-30,30] in increments of 30, pitch [30,60] in increments of 15. These .png files were the only images used to generate the PersonGroup classifier described in the next section. Original MRI storage files never used with the face recognition system, and these render images were not used again after training.

We chose the above parameters primarily through visual assessment of which generated realistic faces. Several of the more subtle parameters (Edge, Ambient Occlusion, Diffuse, DiffuseRough, and the left/right direction scaling) were additionally optimized to maximize the match-confidence score for a single participant between a photo of their face and the generated render image using the Microsoft Azure algorithm. Although these are the best rendering parameters that we found, we obtained similar photo-MRI matching rates at earlier iterations of this study using mostly default parameters (data not presented), and ultimately these parameters were not crucial to obtaining good matching performance.

1.4 Face recognition methods

All face recognition testing was performed using the Microsoft Azure Cognitive Face API[8], which attempts to match an input face photo to a user-defined set of possible faces. The methodological details underlying this cloud-based algorithm are unpublished, proprietary Microsoft technology. We used a secured, private account as part of an existing agreement between Mayo Clinic and Microsoft, but Microsoft also makes this service publicly available as a free demonstration for up to 30,000 face matches per month[9]. The software is designed for photographs, not MRI, but we hypothesized that MRI-based face render images (.png image files) would work sufficiently.

Our usage followed Microsoft's example for face identification (recognition) using the Azure "PersonGroup" classifier[10], with the exception of using MRI-generated face render images where described. The PersonGroup classifier algorithm is designed to match an input face photo ("test image") to the correct individual from a panel (set) of faces that were defined by a "training set" of images. It uses Microsoft's pre-trained models for face detection (locating faces in images) and representation (extraction of features relevant for identifying or differentiating between individuals). The set of individuals to be recognized is defined by the user by uploading a training set comprised of any number of face-containing photographs and associating each photograph with one enumerated individual. Azure then trains a classifier to distinguish between the enumerated individuals, based on the internal feature vectors extracted from their face photos. The user then inputs novel photographs of these individuals (i.e. a "testing" data set), and Azure returns a ranked list (top-50 maximum) of the best matching candidates in the data set, along with its match confidence score for each match[10]. This match confidence score ranges from 0-1, where 1 is a perfect match and 0 is a complete mismatch.

Instead of using photos of faces (as in the software's intended design), we generated a PersonGroup classifier using the MRI-derived face reconstruction images for each participant to define the appearance of their face. For each participant, we used a randomly selected set of 10 MRI face render images from the 81 we generated as described above (during preliminary testing we found that including additional images did not improve matching performance). MRI face render images were used only to define the individuals and the appearances of their faces in the training images; they not used subsequently in our experiments (nor were any other MRI-derived data or images). We then queried this classifier using the "Face-Identify" function for each of the five photographs of each subject, which returns a top-50 list of match candidates with the match confidence score for each. Although the "training" (MRI) and "test" (photograph) data sets for this task contain the same individuals, this is required for a face recognition (i.e. matching) task of differentiating between individuals, since it is impossible to test the ability to distinguish between persons A and B with an independent dataset containing neither. Here, we use only MRI-derived images as the "training" set, and only standard photographs as the "test" set, with Azure's pre-trained detection/representation models for face identification.

Typically, more than one photo of a target individual would be available to someone seeking to identify them. Therefore, we measured the software's ability to match sets of 5 photos, together, to the correct MRI. Although the Azure software only allows matching (identifying) one face at a time, we used the ranked list of candidates (with match confidence scores) for each of the five photos of each participant. Across an individual's five photos (omitting any photos where no face was detected: only 3% of our images), we summed the match confidence score for all potential candidates and ranked the candidates according to their sums. We used these sum-across-photos rankings as ranked matches for each set of five photos for each individual.

We also measured the performance of the classifier when using each photograph individually (i.e. not averaging match confidence scores across the five photographs of each participant). Matching in this scenario (i.e. only one photograph of an individual is available) was correct for the 317/420 (75%) of the photos, which is still highly significant but lower than the per-participant match rate of 70/84 (83%).

1.5 Analysis methods

Face recognition testing used Microsoft's Python SDK for the Microsoft Face API[11], running in Python 2.7.5. Statistical analyses of matching results used R version 3.5.3[12] with *tidyverse*[13] packages.

2 Bibliography

1. Wakabayashi D. Google and the University of Chicago Are Sued Over Data Sharing [Internet]. New York Times. 2019;Available from: https://www.nytimes.com/2019/06/26/technology/google-university-chicago-data-sharing-lawsuit.html

2. Roberts RO, Geda YE, Knopman DS, et al. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology [Internet] 2008 [cited 2013 May 24];30(1):58–69. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2821441&tool=pmcentrez&rendertype=abstract

3. Petersen RC, Roberts RO, Knopman DS, et al. Prevalence of mild cognitive impairment is higher in men. The Mayo Clinic Study of Aging. Neurology [Internet] 2010 [cited 2013 Jun 24];75(10):889–97. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2938972&tool=pmcentrez&rendertype=abstract

4. The ADNI MRI Core Team. ADNI MRI SCANNER PROTOCOLS [Internet]. [cited 2019 Jul 3];Available from: http://adni.loni.usc.edu/methods/documents/mri-protocols/

5. Jack CRJ, Bernstein MA, Borowski BJ, et al. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. Alzheimer's Dement [Internet] 2010 [cited 2012 Nov 7];6(3):212–20. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2886577&tool=pmcentrez&rendertype=abstract

6. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans Syst Man Cybern [Internet] 1979;9(1):62–6. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4310076

7. Rorden C. Surf Ice [Internet]. [cited 2018 Dec 11];Available from: https://www.nitrc.org/projects/surfice/

8. Microsoft Corporation. Microsoft Azure Face API Documentation [Internet]. 2019 [cited 2018 Dec 10];Available from: https://docs.microsoft.com/en-us/azure/cognitive-services/face/

9. Microsoft Corporation. Cognitive Services pricing - Face API [Internet]. [cited 2019 Apr 26];Available from: https://azure.microsoft.com/en-us/pricing/details/cognitive-services/face-api/

10. Microsoft Corporation. Microsoft Azure: Example: Identify faces in images [Internet]. Face API Doc. 2019 [cited 2019 Apr 25];Available from: https://docs.microsoft.com/en-us/azure/cognitive-services/face/face-api-how-to-topics/howtoidentifyfacesinimage

11. Microsoft Corporation. GitHub - Microsoft/Cognitive-Face-Python: Python SDK for the Microsoft Face API, part of Cognitive Services [Internet]. 2019 [cited 2019 Apr 29];Available from: https://github.com/Microsoft/Cognitive-Face-Python

12. R Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2008;Available from: http://www.r-project.org

13. Wickham H. tidyverse: Easily Install and Load the "Tidyverse" [Internet]. 2017;Available from: https://cran.r-project.org/package=tidyverse

3 Acknowledgements

The authors give their thanks to all the volunteers, participants, and coordinators who contributed to this research, with special thanks to Steven M. Smith, Josie M. Williams, Paul D. Lewis, Steven J. Demuth, Soudabeh Kargar, and Zuzana Nedelska.