Supplementary Data

Impact of sample size on amygdala subdivision contrast

A key question when constructing group midspace templates is how many individual images are required to generate sufficient contrast-to-noise ratio between tissue regions for manual labeling. For this work, we approached this question empirically and compared averages of 21, 42, 84, 128 and 168 from the 168 warp registered individual images used to construct the final templates (Supplementary Figure 1). The intensity contrast within the amygdala was estimated from the interquartile range of intensities within the entire amygdala (summation of all individual labels including AMY (Other)) in the T1w and T2w templates. The standard deviation of the noise was estimated from the residual obtained by subtracting the T1w or T2w template (without Laplacian sharpening and intensity normalization) from a representative single subject T1w or T2w image. Lower and upper quartile intensities within the amygdala were 626 and 678 (IQR = 52) for the T1w template and 342 and 375 (IQR = 33) with respective residual noise standard deviations of 24 and 15, resulting in an estimated CNR of 2.2 in both T1w and T2w individual amygdala data.

Impact of contrast-to-noise ratio on registration accuracy

One concern with the use of individual images with low CNR between subdivisions of the amydgala, is whether the diffeomorphic registration has sufficient information to generate a meaningful warp field within the boundaries of the amygdala. We explored this issue using a 3D numerical phantom consisting of two ellipsoidal regions, one within the other, with normally distributed axis lengths and centers. One hundred T1w and T2w phantom pairs were generated with 48 x 48 x 48 voxels with dimensions 0.7 mm x 0.7 mm x 0.7 mm. The phantoms contained filled ellipsoids axis lengths drawn from the normal distribution N(μ, σ) = N(16,1) and center displacements in each axis drawn from N(0, 1). The background, outer and inner ellipsoid intensities were (3.0, 2.0, 1.0) for the T1w phantom and (1.0, 2.0, 3.0) for the T2w phantom respectively. White noise drawn from N(0, σ) with σ = 0.0, 0.2, 0.5, 1.0, 2.0 was added to each phantom and the absolute intensity calculated, resulting in Rician distributed noise. Unbiased templates were constructed from the 100 phantom pairs at each noise levels using the same joint cost function diffeomorphic parameters used to generate the CIT168 template (Supplementary Figure 2a). CNR was calculated between the inner and outer ellipsoids and was therefore equal to $1/\sigma$.

Correlations between ground-truth volume estimates for the inner ellipsoid and those estimated by

inversion of the diffeomorphic warp from individual to template space as summarized in Supplementary Figure 2b-d. Inner ellipsoid volumes become increasingly overestimated as CNR declines and become unstable for CNRs less than 1. Similarity measures between ground truth and estimated inner volume labels are stable for CNRs of 5 and 2 and only begin to decline substantially for CNRs less than 1 (Supplementary Figures 2c and 2d).

Accuracy of Reflection Warping of the Left Amygdala onto the Right Amygdala

The bilateral probabilistic amygdala atlas is generated from repeated labeling of the left amygdala alone in a set of validation templates. This approach is consistent with many other MRI-based segmentations where only unilateral labels are constructed, from example the right hippocampus in ten subjects by two raters (Van Leemput et al., 2009) and the left hippocampus of one subject by 21 research groups (Yushkevich et al., 2015). The validity of left-right comparisons of amygdala and subdivision volumes hinges on the accuracy of the mapping of homotopic structures in the left and right hemispheres. In this work, we define a reflection warp as a reflection about the mid-sagittal plane, followed by a diffeomorphic transformation or warp of the reflected image to the original unreflected image. This transformation initially maps the left amygdala to the approximate location of the right amygdala, then refines the mapping using the diffeomorphic warp. The warp cost function was computed using both the T1w and T2w images, and so makes use of all available tissue contrast features in both images. We can assess the accuracy of the reflection warp by computing the residual intensity difference between the reflected or reflection warped left amygdala and the unwarped right amygdala in the original template (Supplementary Figure 3). It is apparent that the structured residuals within the bounds of the amygdala following reflection alone are substantially reduced by the additional diffeomorphic warp (Supplementary Figure 3e). The contrast-to-noise ratio between major amygdala subdivisions is far higher in the validation templates than in individual images, and it appears very likely that the diffeomorphic warp is adequately constrained by a subset of internuclear boundaries. We therefore have reasonable confidence that the reflection warp is sufficiently accurate to allow interhemispheric comparison of subnuclear volumes without introducing substantial bias.

References

Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B (2009) Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* 19:549–557.

Yushkevich PA, Amaral RSC, Augustinack JC, Bender AR, Bernstein JD, Boccardi M, Bocchetta M, Burggren AC, Carr VA, Chakravarty MM, Chételat G, Daugherty AM, Davachi L, Ding SL, Ekstrom A, Geerlings MI, Hassan A, Huang Y, Iglesias JE, La Joie R, Kerchner GA, LaRocque KF, Libby LA, Malykhin N, Mueller SG, Olsen RK, Palombo DJ, Parekh MB, Pluta JB, Preston AR, Pruessner JC, Ranganath C, Raz N, Schlichting ML, Schoemaker D, Singh S, Stark CEL, Suthana N, Tompary A, Turowski MM, Van Leemput K, Wagner AD, Wang L, Winterburn JL, Wisse LEM, Yassa MA, Zeineh MM, Hippocampal Subfields Group (HSG) (2015) Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. *NeuroImage* 111:526–541.

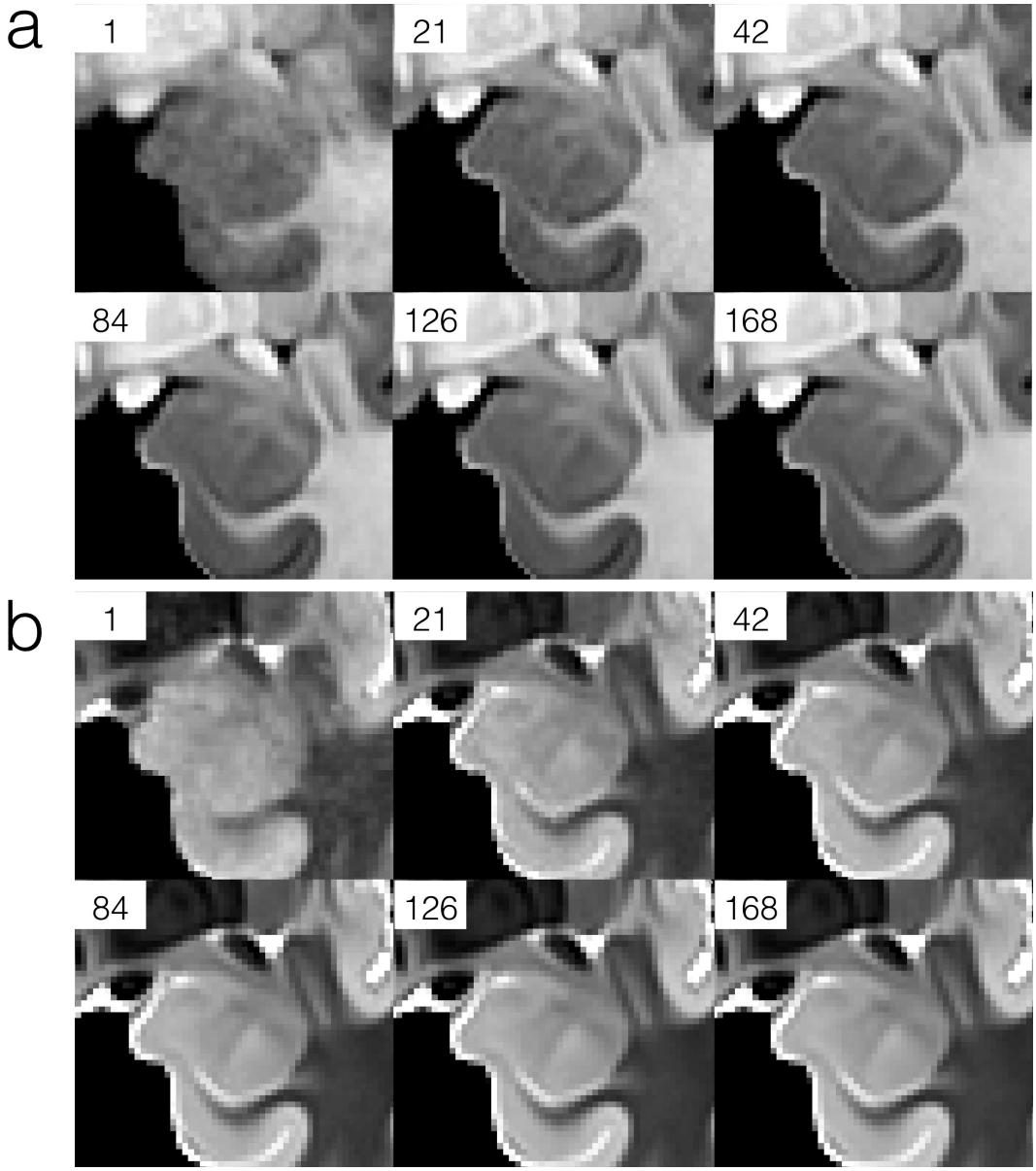

Figure 1: Impact of the number of individual images contributing to an average template on the visibility of amygdala subdivision boundaries in (a) T1w and (b) T2w. Boundary visibility stabilizes between 80 and the full 168 individuals in both imaging contrasts. Cumulative noise statistics are non-Gaussian due to the use of magnitude only, parallel imaging reconstruction in the original data. Note that a Laplacian edge sharpening filter has been applied to all templates except the individual images (n = 1). This is a default behavior of the ANTs template construction and has been retained in this figure to highlight residual noise in the averages

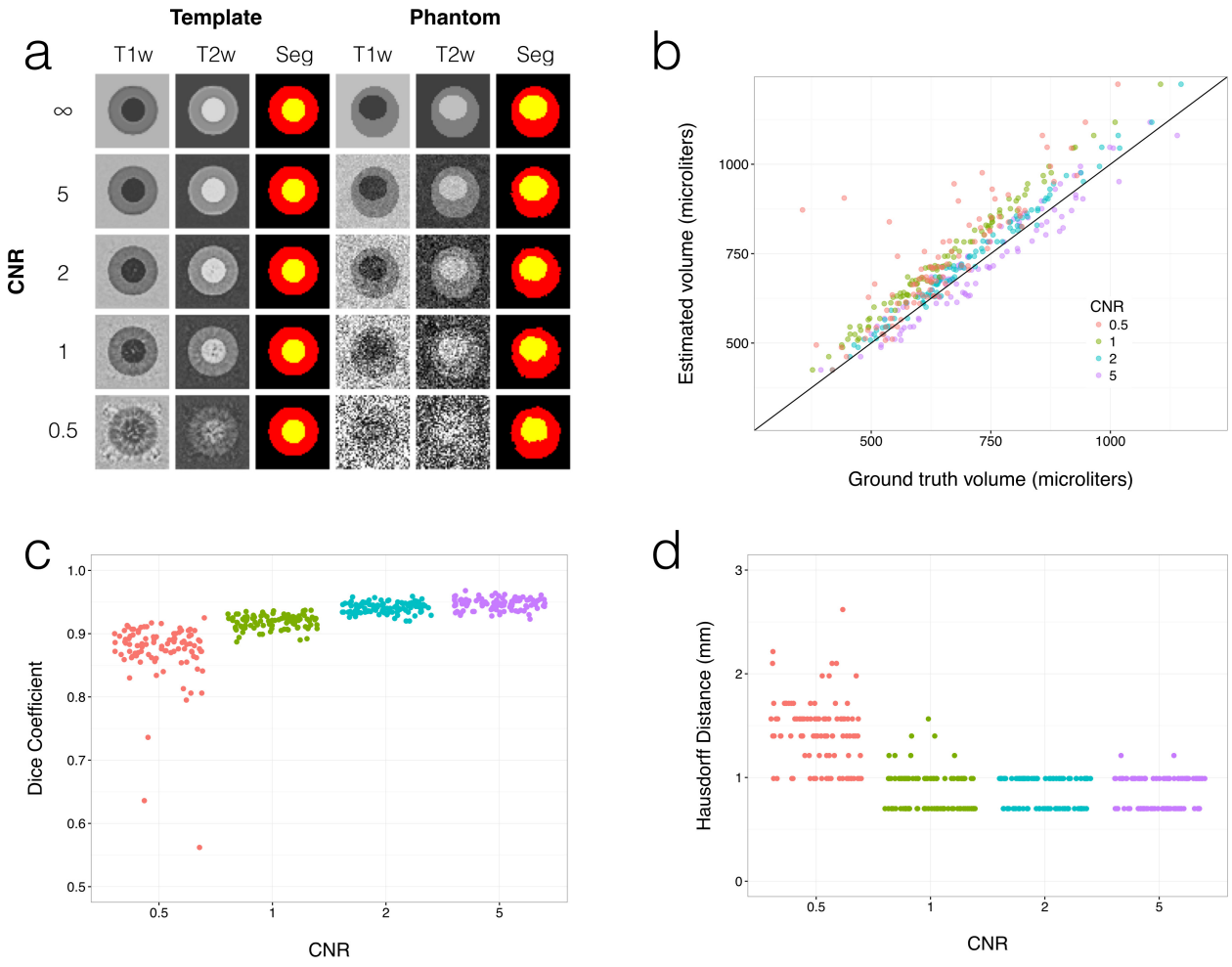

Figure 2: Impact of contrast-to-noise ratio on the robustness of volume estimations in numerical phantoms. One hundred simulated T1w and T2w phantom images consisting of two approximately concentric ellipsoids where generated with randomly jittered axis lengths and center positions. (a) Bivariate templates were constructed from the set of phantoms with various levels of Rician distributed noise. Note that the templates are normalized to the mean signal of the entire volume and are Laplacian sharpened as part of the template construction process (b) Inner ellipsoid volumes were estimated by inversion of the individual to template diffeomorphic warps. Estimate volumes were consistently overestimated for CNRs less than 1.0. (c) Dice coefficients and (d) Hausdorff distances between the estimated and ground truth inner ellipsoids reveal the destabilization of the inverse warp as the CNR falls below 1.0.

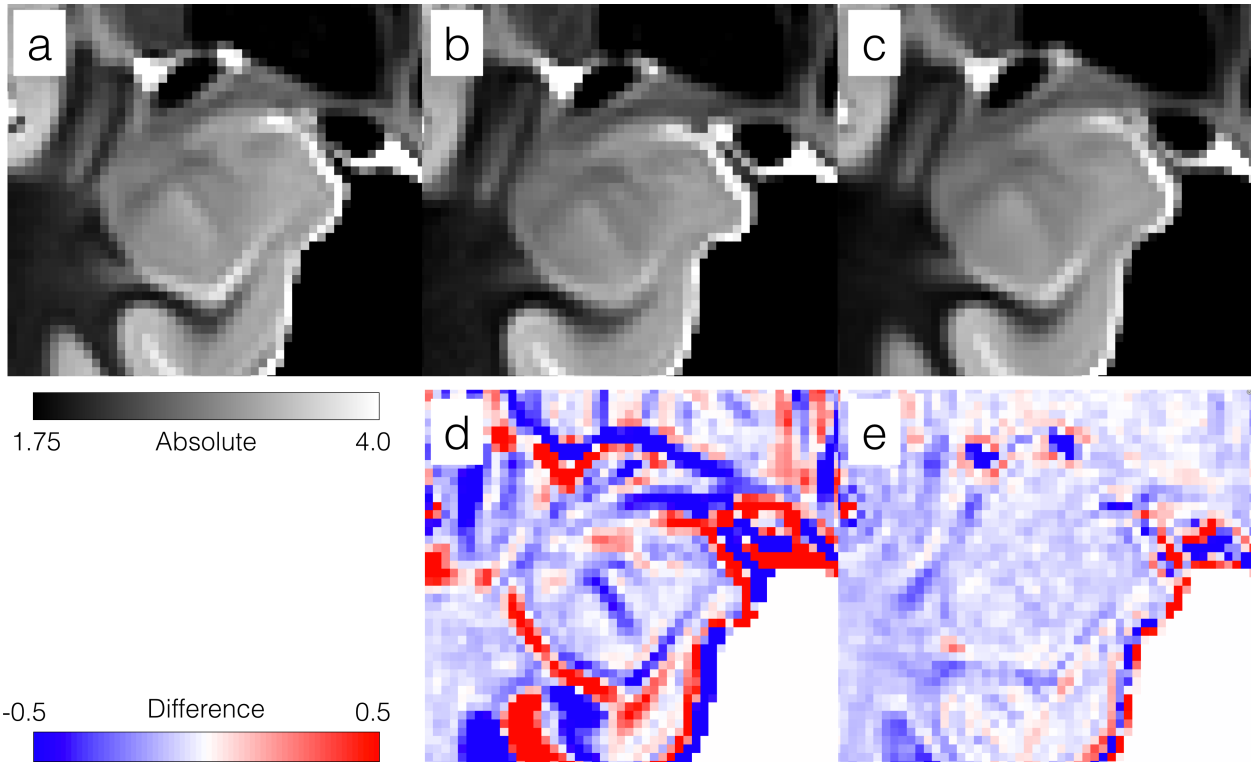

Figure 3: Accuracy of reflection warping of the left amygdala onto the right amygdala. (a) Original right amygdala in one of the eight validation templates used for amygdala labeling. (b) The reflection of the left amygdala across the midsagittal plane. (c) The left amygdala following both reflection about the midsagittal plane and diffeomorphic warping to the original, unreflected template. (d) Intensity difference between the reflected left amygdala (b) and original right amygdala (a). (e) Intensity difference between the reflection warped left amygdala (c) and the original right amygdala (a). The additional diffeomorphic warping step greatly reduces residual intensity differences within the bounds of the amygdala.