

## Supplementary Materials

# Binned Data Provide Better Imputation of Missing Time Series Data from Wearables

Shweta Chakrabarti <sup>1</sup>, Nupur Biswas <sup>1,\*</sup>, Khushi Karnani <sup>2</sup>, Vijay Padul <sup>1</sup>, Lawrence D. Jones <sup>3</sup>, Santosh Kesari <sup>4</sup> and Shashaanka Ashili <sup>3</sup>

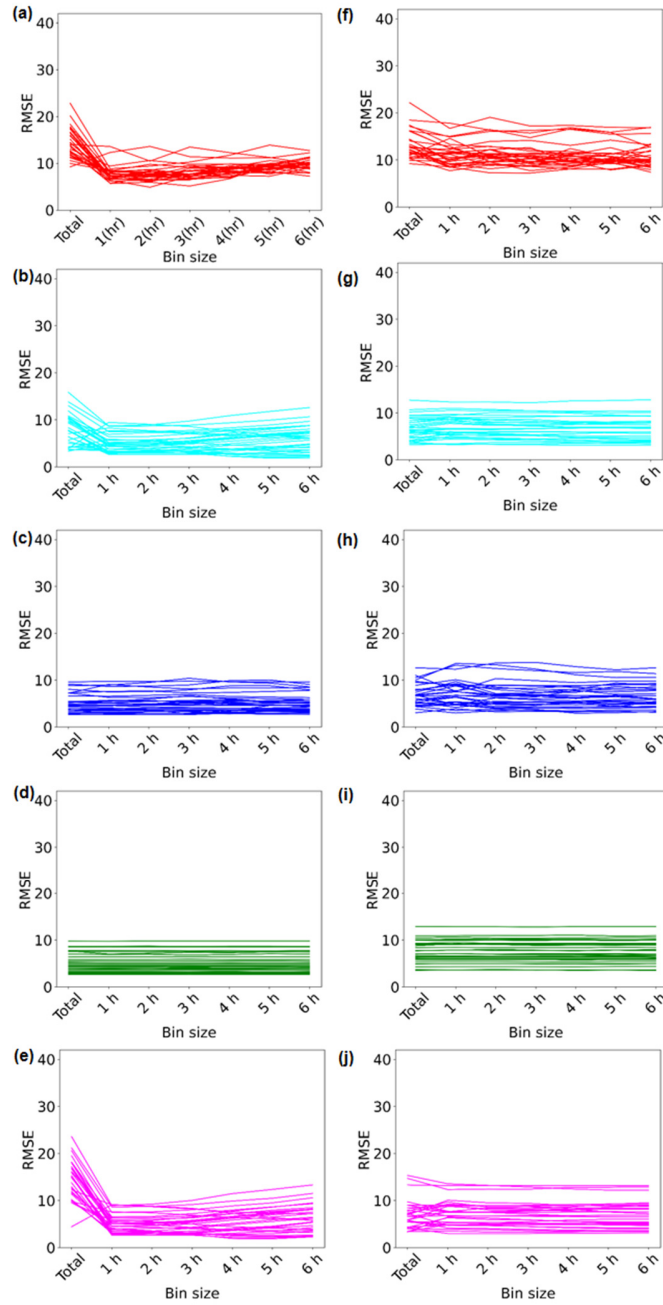
<sup>1</sup> Rhenix Lifesciences, Hyderabad 500038, India

<sup>2</sup> Department of BioSciences and BioEngineering, Indian Institute of Technology, Guwahati 781039, India

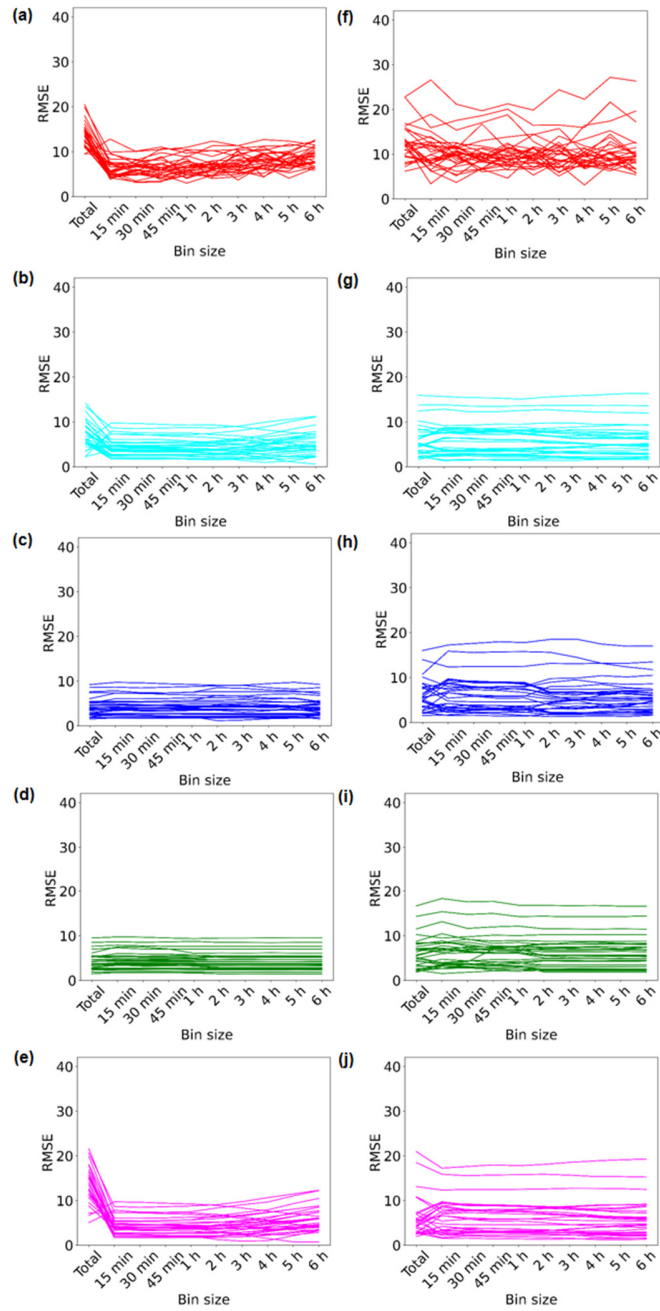
<sup>3</sup> CureScience, 5820 Oberlin Dr, 202, San Diego, CA 92121, USA

<sup>4</sup> Department of Translational Neurosciences, Pacific Neuroscience Institute and Saint John's Cancer Institute at Providence Saint John's Health Center, Santa Monica, CA 90404, USA

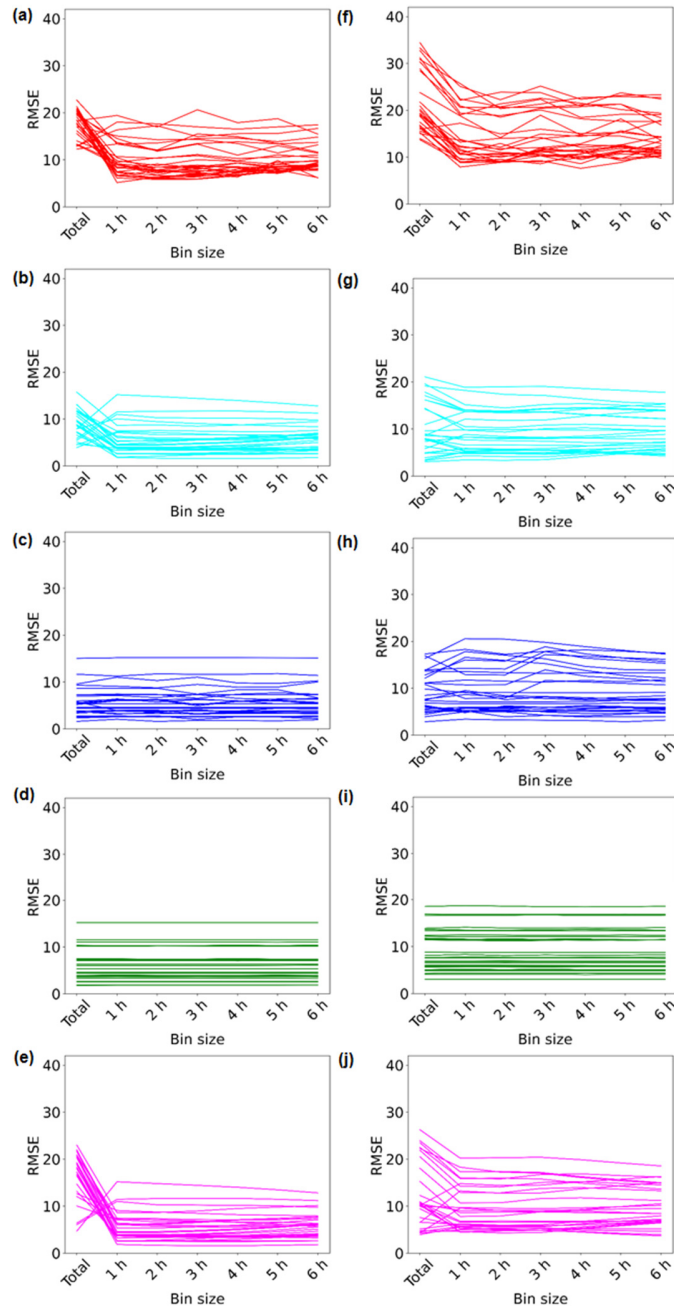
\* Correspondence: nupur@rhenix.org or nupurbiswas@gmail.com



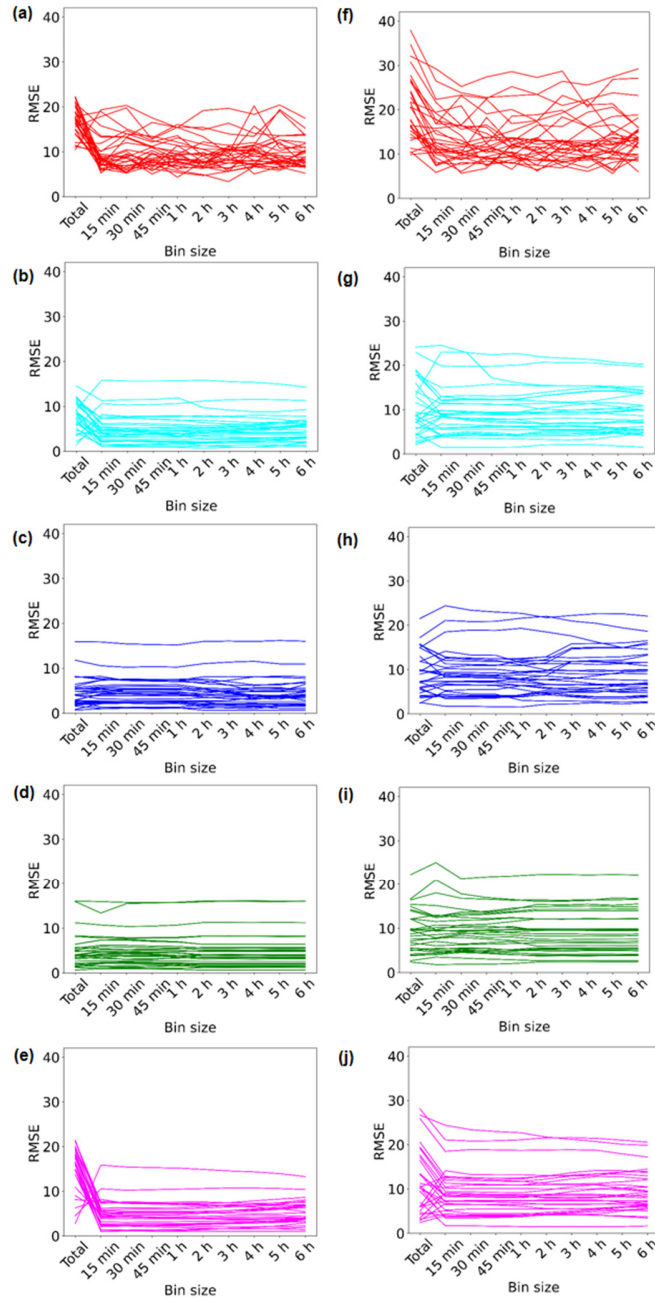
**Figure S1:** Imputation for 1 h missing data for volunteer V2. Variation in RMSE when data of different bin sizes was used for imputing missing data of ‘inactive’ period of 3–4 a.m. using (a) EM, (b) IL, (c) kNN, (d) RF, and (e) SI methods respectively. Variation in RMSE when data of different bin sizes was used for imputing missing data of ‘active’ period of 3–4 p.m. using (f) EM, (g) IL, (h) kNN, (i) RF, and (j) SI methods respectively.



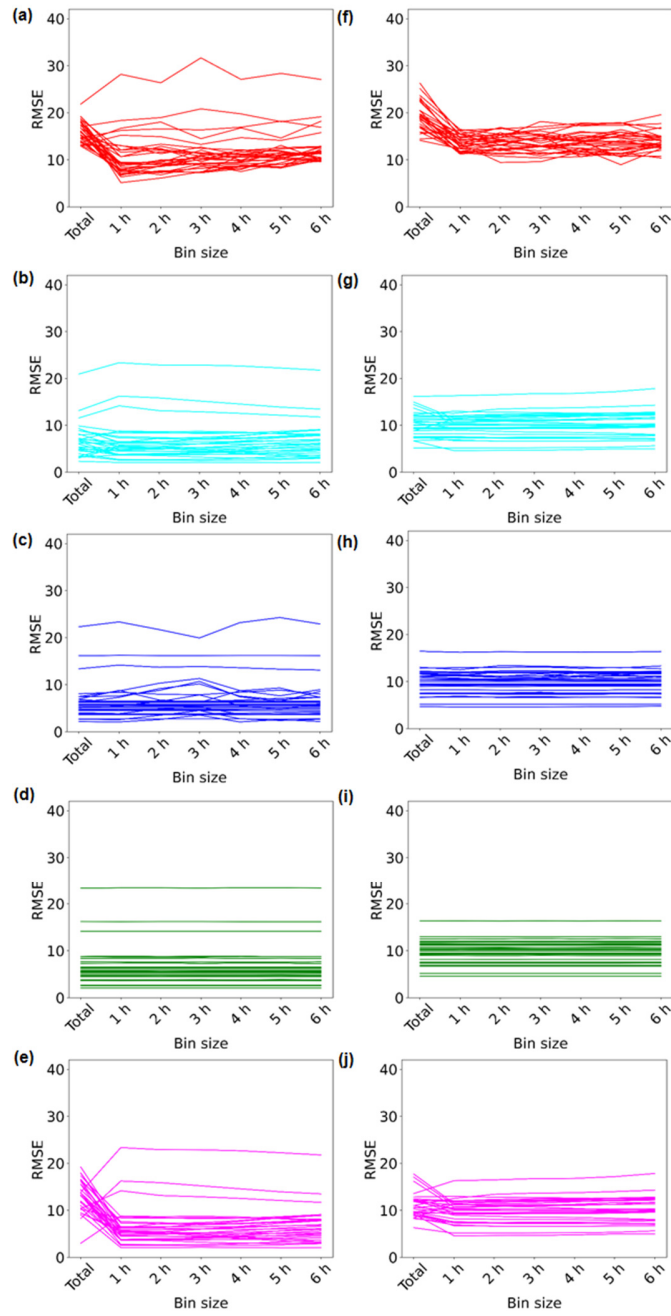
**Figure S2:** Imputation for 15 min missing data for volunteer V2. Variation in RMSE when data of different bin sizes was used for imputing missing data of 'inactive' period of 3–3:15 a.m. using (a) EM, (b) II, (c) kNN, (d) RF, and (e) SI methods respectively. Variation in RMSE when data of different bin sizes was used for imputing missing data of 'active' period of 3–3:15 p.m. using (f) EM, (g) II, (h) kNN, (i) RF, and (j) SI methods respectively.



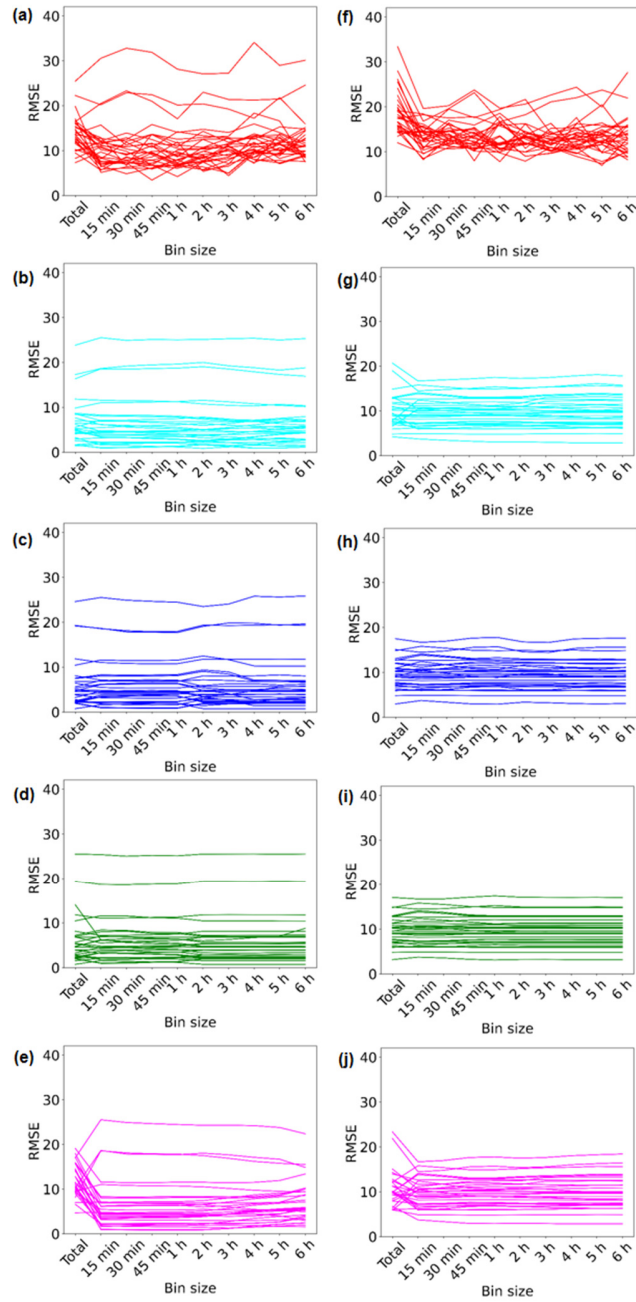
**Figure S3:** Imputation for 1 h missing data for volunteer V3. Variation in RMSE when data of different bin sizes was used for imputing missing data of ‘inactive’ period of 3–4 a.m. using (a) EM, (b) IL, (c) kNN, (d) RF, and (e) SI methods respectively. Variation in RMSE when data of different bin sizes was used for imputing missing data of ‘active’ period of 3–4 p.m. using (f) EM, (g) IL, (h) kNN, (i) RF, and (j) SI methods respectively.



**Figure S4:** Imputation for 15 min missing data for volunteer V3. Variation in RMSE when data of different bin sizes was used for imputing missing data of 'inactive' period of 3–3:15 a.m. using (a) EM, (b) IL, (c) kNN, (d) RF, and (e) SI methods respectively. Variation in RMSE when data of different bin sizes was used for imputing missing data of 'active' period of 3–3:15 p.m. using (f) EM, (g) IL, (h) kNN, (i) RF, and (j) SI methods respectively.



**Figure S5:** Imputation for 1 h missing data for volunteer V4. Variation in RMSE when data of different bin sizes was used for imputing missing data of ‘inactive’ period of 3–4 a.m. using (a) EM, (b) IL, (c) kNN, (d) RF, and (e) SI methods respectively. Variation in RMSE when data of different bin sizes was used for imputing missing data of ‘active’ period of 3–4 p.m. using (f) EM, (g) IL, (h) kNN, (i) RF, and (j) SI methods respectively.



**Figure S6:** Imputation for 15 min missing data for volunteer V4. Variation in RMSE when data of different bin sizes was used for imputing missing data of 'inactive' period of 3–3:15 a.m. using (a) EM, (b) II, (c) kNN, (d) RF, and (e) SI methods respectively. Variation in RMSE when data of different bin sizes was used for imputing missing data of 'active' period of 3–3:15 p.m. using (f) EM, (g) II, (h) kNN, (i) RF, and (j) SI methods respectively.