

Supplemental Document 1

Study Procedures

Bone microarchitecture, density and strength

We used HR-pQCT to assess bone microarchitecture and strength at the left and right distal radii and tibias (four sites) before and after spaceflight with the standard *in vivo* scanning protocol^{1,2}. Limb dominance was determined based on throwing arm and kicking leg. Fracture history was obtained using a health history questionnaire. We acquired a 10.2 mm slice at 9.5 mm and 22.5 mm proximal to the distal radius and tibia reference lines, respectively. Reference lines at follow-up were placed as close as possible to pre-flight reference lines. Each scan acquired 168 slices at a 60.7 μm nominal isotropic resolution using the manufacturer's standard protocol of 68.0 kVp, 1470 μA , matrix size of 2304 X 2304, and 43 ms integration time. The effective dose equivalent for the scan is approximately 5 μSv per scan, with a measurement time of 2.0 min per scan. We manually scored motion artifact on a scale from 1 (no motion) to 5 (discontinuities in the cortical shell and significant blurring of the periosteal surface).³ Scans with motion >3 were excluded from analyses. The scanner was calibrated prior to each measurement session using a phantom provided by the manufacturer.

We performed image analysis according to the manufacturer's standard patient protocol.^{1,2} The periosteal and endocortical contours were automatically generated using a dual-threshold technique.^{4,5} Contours were visually inspected and manually corrected, as previously described.⁶ We performed 3D image registration to ensure the same bone volume was assessed at each time point (Figure 1).^{7,8} Briefly, intensity-based, rigid body registration was used, such that the postflight scan was registered to the baseline scan and a binary mask was created to represent the

common scan region for both time points. The binary mask was transformed based on results of the 3D image registration into the image space of the follow-up scan. All morphological analyses were conducted on the region defined by the common mask. Percent overlap was calculated as the ratio of the total volume in the common region to the total volume in the baseline scan.

Morphological standard measures included total volumetric bone mineral density (Tt.vBMD; mg HA/cm³) and trabecular vBMD (Tb.vBMD; mg HA/cm³), bone volume fraction (BV/TV; %), number (Tb.N; 1/mm), thickness (Tb.Th; mm), and separation (Tb.Sp; mm); cortical vBMD (Ct.vBMD; mg HA/cm³), thickness (Ct.Th, mm) and porosity (Ct.Po; %).^{4,5} Reproducibility in our laboratory range from <3% for density, trabecular and cortical microarchitecture to <14% for Ct.Po.⁹

Failure load (F.Load; N) was estimated on the unregistered segmented HR-pQCT images using custom finite element analysis (μ FE) software (FAIM, version 8.0, Numerics88 Solutions Ltd, Canada).¹⁰ In brief, images were filtered and segmented, and a mesh was generated using the voxel-by-voxel approach, where bone was modeled as non-linear homogeneous material with Young's modulus of 8748 MPa and Poisson's ratio of 0.3.¹¹ An axial compression test was simulated to 1% strain, and failure load estimated using the yield criterion of 2% critical volume and 0.7% critical strain.¹²

Densitometry

Areal bone mineral density (aBMD, g/cm²) of the femoral neck (FN), total hip (TH) and lumbar spine (LS) were acquired pre- and post-flight using dual-energy X-ray absorptiometry (DXA; Hologic QDR Discovery). Scans of both hips and lumbar spine were conducted at NASA

Johnson Space Center in the Bone and Mineral Laboratory and for one crew member at ESA's DLR Institute of Aerospace Medicine. Scans were analyzed using Hologic's automated software. The global hip box was manually determined as previously described.¹³ Weekly and day-of test calibration was conducted using a calibration phantom. Precision values at JSC are <4% for FN, TH and LS.¹³

Bone biochemistry

Biochemical data were obtained through data sharing with NASA's Spaceflight Standard Measures and Biochemical Profile studies. Blood and urine samples were collected before flight at approximately launch -180 and -45 days, in-flight at flight day (FD)15, FD30, FD60, FD120, and FD180 and upon return (R+0). Pre-flight and R+0 urine were collected as two consecutive 24-hour urine pools while in-flight urine was one 24-hour urine pool. Blood samples were collected following an overnight fast. Since flight duration varied, not all crews had five in-flight sample collections. Biochemical assays were performed as previously described.¹⁴ For this study, pre-flight results (180 and 45 days before launch) were averaged.

Urine and blood specimens were analyzed by the NASA Nutritional Biochemistry Lab at JSC. Serum and urine creatinine were assessed using a colorimetric method on an Ace Alera clinical chemistry analyzer (inter assay %CV: 4.5). Serum and urine calcium were analyzed by atomic absorption spectroscopy (Perkin Elmer Flame 3.1% CV for blood and 3.3% CV for urine). Osteocalcin was analyzed by radioimmunoassay (Alpco, Salem, NH, 5.9% CV). Bone-specific alkaline phosphatase and sclerostin were analyzed by commercially available enzyme-linked immunosorbent assays (ELISA) (Quidel, 7.8% CV and 3.7% CV, respectively). 1,25-

dihydroxyvitamin D and 25-hydroxyvitamin D were analyzed by commercially available radioimmunoassays (Diasorin, Inc, Stillwater, MN) 16.8% and 13.3%, respectively). Urine N-telopeptide (NTx) was analyzed using a commercially available ELISA (9.6% CV). Parathyroid hormone was measured using a commercially available immunoradiometric assay (Scantibodies, Inc., Santee, CA) 4.7% CV), and osteoprotegerin (OPG) and RANKL were assessed by enzyme immunoassay (10.6% CV and 15.6% CV, respectively). Serum phosphorus was assessed using a colorimetric method on a clinical chemistry analyzer (Alera, 2.3% CV). C-telopeptide (CTx) was measured using a commercially available ELISA (6.9 % CV) and procollagen 1 intact N-terminal propeptide (PINP) was analyzed using a radioimmunoassay (4.1 % CV).

Exercise

Crewmembers completed a pre-flight health history questionnaire inquiring about exercise history during the past year for a typical week. Questions included the number of sessions and hours per week spent running, cycling and weight training. Weight training questions further probed about the number of sets, repetitions and average load per week for various exercises including squats, deadlifts and heel raises. During spaceflight, crewmembers followed an individualized exercise program prescribed by the space agency exercise physiologist. On-orbit exercise was prescribed six days per week and consisted of aerobic exercise (using the cycle ergometer or treadmill) and resistance training using the ARED (Advanced Resistive Exercise Device). In brief, the ARED allows 17 exercise configurations, including squats, deadlifts, heel raises, bench press, bicep curl, triceps extension, rows, shrugs, and shoulder press. In this study, we examined data from the in-flight treadmill and cycle ergometer: frequency (number of sessions per week), duration (average exercise time per session) and weekly volume (average

time per week); and three lower body ARED exercises and their variations: squats (back, single leg, and sumo), deadlifts (stiff leg, Romanian and sumo) and heel raises (both and single leg): number of resistance training sessions per week, average number of sets, repetitions, and volume (repetitions/week). We also calculated change in aerobic and resistance training volume from pre-flight to in-flight.

Statistical analysis

All analyses were conducted using Stata (version 16, StataCorp, College Station, USA). Change in bone variables from pre- to post-flight were assessed using Wilcoxon signed rank test. Spearman correlations examined relationships between: 1) change in bone variables and potential continuous covariates (e.g., age, mission duration, total number of days spent in microgravity including prior missions); 2) pre-flight bone variables with change in bone variables; 3) pre-flight and change in biochemical markers with bone variables; and pre-flight and in-flight exercise with bone variables. Mixed effects models with Kenward-Roger small sample size adjustment¹⁵ examined changes in bone variables pre- to post-flight, while adjusting for mission duration. A random intercept allowed individuals their own intercept for the effect of time. Bone models included the fixed effects of time (pre/post), mission duration, and the interaction between time and mission duration. We examined linear, quadratic and piecewise models for the effect of mission duration. Wald test p-values were used to determine significance of individual fixed effects and the best fitting models were determined by the largest reduction in the deviance test ($-2LL$) and parsimony of the model (Akaike and Bayesian information criterion (AIC and BIC) values). The margins command was used to predict the mission duration at which a significant change from pre-flight occurred, using a Bonferroni correction to account for

multiple comparisons. All bone analyses were conducted on raw pre- and post-flight values; however, in Figure 2 we dichotomized mission duration into shorter or longer than a 6-month mission and present data as percent change from pre-flight values, normalized to a 6-month mission.

Mixed effects models also examined changes in biochemical markers of bone turnover across time. Biochemistry data included before flight, FD15, FD30, FD60, FD120, FD180 and return (R+0). Bonferroni correction accounted for multiple comparisons. We log-transformed 25-hydroxy vitamin D, OC and RANKL for analyses, but present raw data in tables. We were unable to normalize 25-hydroxy vitamin D, but present results of the log-transformed analyses. Finally, we examined the relationship between biochemical markers and exercise with change in bone variables by individually adding biochemical markers and exercise variables as fixed effects (along with their interactions with time) to the mixed effects bone models. Model assumptions were assessed graphically using plots of residuals and significance was set at $p < 0.05$. Study data were managed using REDCap electronic data capture tools hosted at The University of Calgary.

References

1. Manske SL, Zhu Y, Sandino C, et al. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone* 2015;79:213-21. doi: 10.1016/j.bone.2015.06.006
2. Whittier DE, Boyd SK, Burghardt AJ, et al. Guidelines for the assessment of bone density and microarchitecture in vivo using high-resolution peripheral quantitative computed tomography. *Osteoporos Int* 2020 doi: 10.1007/s00198-020-05438-5 [published Online First: 2020/05/28]
3. Pauchard Y, Liphardt A-M, Macdonald HM, et al. Quality control for bone quality parameters affected by subject motion in high-resolution peripheral quantitative computed tomography. *Bone* 2012;50(6):1304-10. doi: 10.1016/j.bone.2012.03.003
4. Burghardt AJ, Buie HR, Laib A, et al. Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT. *Bone* 2010;47(3):519-28. doi: 10.1016/j.bone.2010.05.034
5. Buie HR, Campbell GM, Klinck RJ, et al. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone* 2007;41(4):505-15. doi: 10.1016/j.bone.2007.07.007
6. Whittier DE, Mudryk AN, Vandergaag ID, et al. Optimizing HR-pQCT workflow: a comparison of bias and precision error for quantitative bone analysis. *Osteoporos Int* 2020;31(3):567-76. doi: 10.1007/s00198-019-05214-0 [published Online First: 2019/12/01]
7. Kemp T, de Bakker CMJ, Gabel L, et al. Longitudinal bone microarchitectural changes are best detected using image registration. *Osteoporos Int* 2020 doi: doi:10.1007/s00198-020-05449-2
8. Ellouz R, Chapurlat R, Van Rietbergen B, et al. Challenges in longitudinal measurements with HR-pQCT: evaluation of a 3D registration method to improve bone microarchitecture and strength measurement reproducibility. *Bone* 2014;63:147-57. doi: 10.1016/j.bone.2014.03.001
9. MacNeil JA, Boyd SK. Improved reproducibility of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med Eng Phys* 2008;30(6):792-99. doi: 10.1016/j.medengphy.2007.11.003
10. MacNeil JA, Boyd SK. Bone strength at the distal radius can be estimated from high-resolution peripheral quantitative computed tomography and the finite element method. *Bone* 2008;42(6):1203-13. doi: 10.1016/j.bone.2008.01.017
11. Whittier DE, Manske SL, Kiel DP, et al. Harmonizing finite element modelling for non-invasive strength estimation by high-resolution peripheral quantitative computed tomography. *J Biomech* 2018;80:63-71. doi: 10.1016/j.jbiomech.2018.08.030
12. Pistoia W, van Rietbergen B, Lochmüller E-M, et al. Estimation of distal radius failure load with micro-finite element analysis models based on three-dimensional peripheral quantitative computed tomography images. *Bone* 2002;30(6):842-48.
13. Sibonga J, Matsumoto T, Jones J, et al. Resistive exercise in astronauts on prolonged spaceflights provides partial protection against spaceflight-induced bone loss. *Bone* 2019;128:112037. doi: 10.1016/j.bone.2019.07.013

14. Smith SM, Heer MA, Shackelford LC, et al. Benefits for bone from resistance exercise and nutrition in long-duration spaceflight: Evidence from biochemistry and densitometry. *J Bone Miner Res* 2012;27(9):1896-906. doi: 10.1002/jbmr.1647
15. McNeish D. Small Sample Methods for Multilevel Modeling: A Colloquial Elucidation of REML and the Kenward-Roger Correction. *Multivariate Behav Res* 2017;52(5):661-70. doi: 10.1080/00273171.2017.1344538