Technical Details of the Results

Exploratory Factor Analysis

Table 3 contains the results of the EFA and internal consistency reliability analyses for the 10-item single factor model (Full Scale), the 9-item single factor model and the 9-item four-factor imposed model (Subscales) tested by Cadarette et al. The Kaiser-Mayer-Olkin and Bartlett’s sphericity test were significant in each model indicating that EFA was appropriate. The amount of variance explained, however, was low (only 36.6% in the 9-item four-factor model). The individual factor loadings were adequate (> 0.30) for our large sample size [20] except for the reverse-coded items. While items loaded onto their respective factors from the Cadarette et al. model, the factor loadings were weak (range = 0.36-0.92, data not shown). Cronbach’s alpha for the 10-item overall instrument was 0.68 and the reliability coefficients for the domain subscales were all lower.

Confirmatory Factor Analysis

We tested each of the factor domains and their respective items using CFA as depicted in Figure 1. Several goodness-of-fit test statistics indicated that overall the model reasonably fit the data well, but the modification indices for the error terms suggested that the error terms for Age Related and No Osteoporosis Prevention (e3 and e9), Diet and Thin Bones (e5 and e8), and Diet and No Osteoporosis Prevention (e5 and e9) were correlated. The CFA model incorporating these correlated error terms showed slight improvements in fit. The χ² statistic (χ² = 6.47) for the model was significant (p < 0.0001), which signifies poor model fit, but this test is sensitive to sample size and will reject more often than it should [21] We adjusted for the sample size by using the ratio of the χ² to the degrees of freedom (CMIN/DF = 3.14), but that value was still above the desired threshold of < 3, again suggesting poor fit [22].

We also examined the comparative fit index (CFI) which compares the model to a more restrictive, nested model. CFI ranges from 0 to 1 with values close to 1 indicating better goodness-of-fit. The CFI for the model is close to 1 (CFI = 0.996) indicating that the model fits better than a more restrictive model that has no relationship between the variables. The Tucker-Lewis Index (TLI) is a variation on the CFI with an additional penalty for including freely estimated variables and can be interpreted in a similar fashion to the CFI. The TLI for our model (TFI = 0.992) was high, further suggesting that the model is reasonable. We used the parsimony correction index called the Root Mean Square Error of Approximation, which penalizes for poor parsimony and is sensitive to model
complexity but is not affected by sample size [21]. The RMSEA (0.017) was below the threshold of 0.05 further supporting that the model was a good fit.

The correlations between the factors range from a low of 0.23 between Consequences and Prevention and Treatment to a high of 0.73 between Biological and Consequences. Consistent with EFA, item correlation was low with correlations ranging from 0.02 to 0.47 (data not shown). The highest correlation occurs between items 4 and 5 (\( \rho = 0.47 \)), which make up the Lifestyle domain, although this is far below the 0.80 threshold signifying collinearity and poor discrimination. The low correlations of the other items indicate that each item is measuring a distinct component.

Each of the factor items significantly load on their respective latent variables with values ranging from 0.44 to 0.74, or poor to excellent (data not shown) [23]. The items in the Biological and Lifestyle domains generally had the largest loadings. The items for the Consequences domain were somewhat lower while the reverse-coded items in the Prevention and Treatment domains were the lowest.

**Measurement Invariance**

Measurement invariance was examined by testing for structural or configural invariance by age, sex, education, history of prior DXA, baseline study DXA result, site and pharmacotherapy. The results of the model fit indices are given in Table 4. While testing for configural invariance, the fit indices of each subgroup for the unconstrained models met the desirable respective thresholds (RMSEA < 0.05, CFI > 0.95, NFI > 0.95, TLI > 0.95) although the \( \chi^2 \) statistic was significant in each subgroup. We tested for weak invariance by constraining the factor loadings to be equal in each subgroup and obtained similar results except for the diagnosis subgroups which indicated weak invariance for all model fit indices. For all other subgroups the model fit indices indicate at least configural invariance.

**Responsiveness to Change**

We performed a responsiveness to change analysis comparing the three groups mentioned previously. Improvements in knowledge by 0.50 standard deviations were reported among 14.7% (N = 1,137) of participants at 12-weeks and 17.7% (N = 1,375) at 52-weeks. There was a statistically significant difference in the scale scores at 12- and 52-weeks among those who improved versus those who did not improve relative to baseline (p < 0.0001).

The logistic regression models for correctly identifying the results of their baseline study DXA scan are shown in Table 5. Models 1 and 3 included as the predictor variables the change score, and whether or not the score
improved at follow-up for 12- and 52-weeks, respectively. Models 2 and 4 further adjust for the baseline knowledge score for 12- and 52-weeks, respectively.

Generally, the change score, while accounting for improvement in knowledge, indicated that patients were less likely to correctly identify their baseline study DXA result at either 12- or 52-weeks. This was apparent in Group 1, newly diagnosed patients, and Group 3, pharmacotherapy- and DXA-experienced patients at 12-week follow-up, but only in Group 2, normal bone density patients, was this statistically significant (OR=0.85, 95% CI: 0.75, 0.96). Similarly, at 52-weeks, only 12% (OR=0.88, 95% CI: 0.78, 0.996) of Group 1 and 18% (OR=0.82, 95% CI: 0.72, 0.94) of Group 2 patients were significantly able to correctly recall their baseline study DXA result. However, an improved score at 12-weeks suggests that 30% of Group 3 patients were more likely to be able to correctly recall their baseline study DXA result. This significant effect was lost when baseline score was included in the model (Model 2).

Adjusting for the baseline knowledge score consistently improved the association between the change score and correctly recalling the baseline study DXA result. The indicator variable for improved knowledge, however, becomes non-significant in Models 2 and 4 at both 12- and 52-weeks, while the baseline knowledge score is a more significant predictor of correctly recalling the baseline study DXA result among all groups (p < 0.05).

In Model 2 at 12-weeks and Model 4 at 52-weeks, the baseline knowledge score is more predictive of patients correctly identifying their baseline study DXA result for Group 2 than for the others. For each additional question that a participant answered correctly, the odds of identifying their baseline study DXA result increased by 20% (OR = 1.20, 95% CI = 1.08 to 1.33). The influence of the baseline score on the outcome did not change from 12- to 52-week for Group 2; however, there was a decrease in the proportion of Group 1 and 3 patients who could correctly identify their baseline study DXA results at 52-weeks (14% and 16%, respectively) compared to 12-weeks (both 17%). Patients with any DXA experience, diagnosis or pharmacotherapy (Group 3) have higher odds of knowledge than those who are newly diagnosed (Group 1) at 52-weeks.

Except for Group 3 in Model 3, the indicator variable for improved score increased study DXA result recall at 52-week follow-up. This distinction is not marked for Group 3 until the baseline score is included in Model 4. According to Model 4, improvements in recall was greatest in normal bone density patients (Group 2; OR=1.53, 95% CI: 0.82, 2.84) followed by newly diagnosed patients (Group 1; OR=1.51, 95% CI: 0.84, 2.69) and was least notable in experienced patients (Group 3; OR=1.13, 95% CI: 0.85, 1.51).