The data cleaning pipeline began by addressing inconsistencies in the therapy treatment records. First, outlier values—specifically, any treatment minutes exceeding 600—were removed, and for patients with a length of stay of 7 days or less, any week‐2 therapy records were excluded; during this phase, new variables such as total therapy dose (combining week‑1 and week‑2 amounts) and a per-day dose (computed as the sum of week‑1 and week‑2 divided by the length of stay, capped at 14 days) were also created to better capture treatment intensity (​). Next, records missing both selfcare and mobility outcome measures were deemed uninformative and dropped, resulting in a final cleaned dataset of 42,938 patients (​). This refined dataset was then partitioned into training, evaluation, and test sets containing 28,625, 7,156, and 7,157 samples, respectively, as defined by an externally maintained split dictionary (​).