 Comparison of analytical performance and biological variability of three bone resorption assays

Hsin-Shan J. Ju, Sunny Leung, Bradley Brown, Matthew A. Stringer, Scott Leigh, Christy Scherrer, Karen Shepard, Dean Jenkins, Jane Knudsen, and Robert Cannon

We have compared the analytical performance and biological variability of three commercially available bone resorption assays: Pyrilinks®-D, Osteomark®, and CrossLaps®, for the measurement of urinary free deoxypyridinoline (Dpd), cross-linked N-telopeptides of type I collagen (NTx), and linear C-telopeptides of type I collagen (CTx), respectively. The intraassay and interassay CVs for precision of the Dpd and NTx assays were <10% for analyte concentrations greater than the second calibrator (i.e., 3 nmol/L Dpd or 30 nmol bone collagen equivalents/L NTx). The CTx assay demonstrated poor precision for analyte concentration lower than the third calibrator (i.e., 200 μg/L). The NTx assay exhibited nonlinear recovery for sample dilutions prepared in buffer; however, this nonlinear recovery could be corrected for sample dilutions made in urine at a low analyte concentration. Supplement recoveries of each of the three assays were within 100% ± 10% on average. All three analytes showed stability through five freeze–thaw cycles. The mean day-to-day variations were 16% for Dpd, and 23% for both NTx and CTx. Similar diurnal rhythm was observed for all three assays on average, with the peak in the early morning and the nadir in the afternoon. Mean amplitude of the diurnal variation was 37% for Dpd and NTx, and 57% for CTx. Variations within the reference intervals for a healthy premenopausal population were 28% for Dpd, 57% for NTx, and 56% for CTx. Pyrilinks-D has demonstrated analytical precision and accuracy equal or superior to Osteomark and CrossLaps in all areas. Dpd exhibits the least biological variability day-to-day, within individuals across the diurnal cycle, and within a healthy premenopausal population.

The organic matrix of bone consists of ~90% type I collagen [1]. Trifunctional pyridinium cross-links—pyridinoline (Pyd) or deoxypyridinoline (Dpd)—form between hydroxylysine or lysine residues at the C- and N-telopeptide ends of one collagen molecule and the helical portion of a neighboring molecule during collagen maturation [2]. The integrity and rigidity of bone collagen derive from the cross-linking between the adjacent collagen fibrils. Pyridinium cross-links have been found in bone, cartilage, tendon, and most other connective tissues, but not skin [2–4]. Although Pyd has a wider tissue distribution, Dpd is present almost exclusively in bone. The molar ratio of Pyd and Dpd in urine is very similar to that in bone [5, 6]. This indicates that both Pyd and Dpd are derived mainly from bone. However, Dpd would be a more appropriate bone marker for patients with cartilage disease, e.g., rheumatoid arthritis, who have significant increases in Pyd from nonbone tissue [7]. After bone degradation, cross-links are released into the circulation and are not reused in new collagen synthesis or metabolized in the liver, and are eventually excreted in urine. The molecular distribution of urinary collagen cross-links is 40–50% of the free form and 50–60% of peptide-linked forms [8–10]. The peptide-linked forms are mixtures generated from N- and C-telopeptides of collagen.

Pyridinium cross-links have proven to be specific and sensitive bone resorption markers for evaluation of met-
abolic bone diseases such as osteoporosis, hyperthyroid-
ism, hyperparathyroidism, Paget disease, and malignancy
involving bone [11–17]. A HPLC technique with a pre-
treatment step was the first method devised to measure
both cross-links in urine [18, 19]. Recently, immunoassays
for the measurement of urinary free Dpd [20], cross-
linked N-telopeptides of type I collagen (NTx) [21], and
linear C-telopeptides of type I collagen (CTx) [22] were
developed. The new methods offer alternatives to the
laborious and complicated HPLC method for measurement
of the cross-links.

In addition to biochemical marker evaluations, bone
imaging techniques and bone biopsies are available for the
diagnosis of metabolic bone diseases. Bone markers,
bone biopsies, and radioisotope labeling can evaluate
bone turnover status. Dpd has been shown to correlate
with resorption histomorphometry of bone biopsies and
85Sr kinetic analyses for estimation of bone resorption
[23–25]. Bone biopsies are invasive methods and they are
impractical for frequent and routine use. Consensus has
been reached that measurement of bone density is impor-
tant in the diagnosis of osteoporosis (i.e., 2.5 SD below the
young adult mean) [26]. However, detecting high bone
turnover at onset stage or monitoring the acute changes in
bone is difficult with bone density. Therefore, measure-
ment of biochemical markers of bone resorption can serve
as a routine and noninvasive way to detect and assess the
progress of metabolic bone diseases. The markers are also
sufficiently sensitive to effectively monitor the acute
changes in bone turnover [27], such as after initiation of
antiresorptive therapy or evaluating patients’ compliance.

The appropriate interpretation of biochemical marker
results should consider all sources of variability that
include the analytical performance characteristics of the
method and the biological variability of the marker itself
[28]. Several factors, such as the availability of highly
specific and sensitive antibodies, pure and discrete cali-
briors, and assay design determine the analytical capa-
bility of the assay. The tissue specificity, specificity for the
resorption process, rate of bone metabolism, and other
factors may contribute to biological variability. We have
investigated the analytical performance, and day-to-day,
diurnal, and population variability of healthy subjects for
three commercially available assays for urinary markers
of bone resorption: Pyrilinks®-D, Osteomark®, and Cross-
Laps®. The aim of the study was to assess and compare
the analytical reliability and biological attributes of the
three assays and the markers they measure.

**Materials and Methods**

**Urine samples for biological variability evaluation**

*Day-to-day.* Seven women and 11 men participated in the
day-to-day variability study. First morning void (FMV)
urine was collected on 8 nonconsecutive days over a
5-month period.

*Diurnal.* Seventeen women and 21 men were included in
the diurnal variation study. Urine was collected over a
24-h period. The first sample consisted of all urine ex-
creted in the 3-h interval between 0700 and 1000. Subse-
quently, urine was collected in 3-h intervals for the rest of
the day. A total of eight samples per subject was collected.

*Premenopausal reference population.* FMV urine was col-
clected from 216 women who participated in a reference
interval study.

Urine samples for analytical evaluations were collected
from in-house volunteers. Urine samples for the biological
variability studies were obtained from healthy subjects
between the ages of 25 and 44 years. The mean age was 34
years for both day-to-day and diurnal studies and 35
years for the reference interval study. No subjects had a
past medical history of endocrine, renal, metabolic, bone/
articular disease, or any type of malignancy. None of
subjects had used drugs that could affect bone resorption
for at least 2 years. Women were premenopausal, not
pregnant, lactating, or using oral contraceptives. All the
urine samples for the biological variability studies were
stored at ≅20 °C until testing.

**Assays**

Dpd was measured with Pyrilinks-D (Metra Biosystems)
[20]. NTx were measured with Osteomark (Oste Pas
International) [21]. CTx were measured with CrossLaps (Oste-
ometer Biotech) [22]. All three assays were run according
to manufacturers’ directions. Resorption marker results in
the biological variability studies were corrected for differ-
ences in urine concentrations by expression relative to the
urinary creatinine concentration.

All urinary creatinine concentrations were measured
by Corning Nichols Institute reference laboratories by
using a standard colorimetric method.

**Analytical evaluation**

The analytical performance was evaluated by determin-
ing intraassay and interassay precisions, and accuracy
was assessed by determining the linearity of dilution and
the recovery of supplemented analyte. Three operators
assessed intraassay precision by measuring six urine
samples with 26 or 28 replicates. Five operators assessed
interassay precision with three urine samples and the
manufacturer’s kit controls over a 2-week period. For
linearity testing, urine samples and the kits’ high calibra-
tors were serially diluted two-, four-, eight-, and 16-fold
with the kits’ assay buffer (Dpd and NTx assays) or zero
(buffer) calibrator (CTx assay). These are the diluents
recommended by the manufacturers of the Dpd and CTx
assays for retesting specimens exceeding the value of
the highest kit calibrator. Linearity was also evaluated with
serial dilution into a urine with low analyte concentration
for all three assays. This is the diluent recommended by
the manufacturer of the NTx assay for retesting specimens
exceeding the value of the highest kit calibrator. When
using urine as the diluent, the analyte concentration was incorporated into the calculation of recovery. Supplement recovery was assessed by supplementing 1/10th volume of the two highest kit calibrators into nine urine samples. For freeze–thaw stability evaluation, urine samples were analyzed over five freeze–thaw cycles.

**STATISTICAL ANALYSIS**

Statistical calculations were performed with StatView® software (Abacus Concepts). For the analysis of diurnal variability, a repeated-measures linear model was estimated by using a third-order polynomial equation to model the effect of time by Pacific Research Associates. Factors were added to the model if they significantly ($P < 0.05$) contributed to the underlying variability as measured by an $F$-test. Adequacy of the final model was evaluated by an $F$-test and by a review of the diagnostic information (e.g., residual plots).

**Results**

**ANALYTICAL EVALUATION**

**Precision.** Intraassay precision profiles of the three assays are plotted as CV vs analyte concentration (Fig. 1). Table 1 shows the data for interassay precision determined with three urine samples and the manufacturer-supplied kit controls. The interassay precision was analyzed in 20 different runs by five operators over a 2-week period. The data for interassay precision were comparable with in-

<table>
<thead>
<tr>
<th>Sample</th>
<th>Dpd Mean ± SD, nmol/L</th>
<th>Dpd CV, %</th>
<th>NTx Mean ± SD, nmol BCE/L</th>
<th>NTx CV, %</th>
<th>CTx Mean ± SD, µg/L</th>
<th>CTx CV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine 1</td>
<td>7.4 ± 0.6</td>
<td>8.1</td>
<td>23.3 ± 6.2</td>
<td>26.6</td>
<td>564.0 ± 71.1</td>
<td>12.6</td>
</tr>
<tr>
<td>Urine 2</td>
<td>31.3 ± 2.6</td>
<td>8.3</td>
<td>294.0 ± 23.0</td>
<td>7.8</td>
<td>1845.2 ± 179.3</td>
<td>9.7</td>
</tr>
<tr>
<td>Urine 3</td>
<td>128.6 ± 11.5</td>
<td>8.9</td>
<td>2473.4 ± 152.1</td>
<td>6.1</td>
<td>842.1 ± 84.2</td>
<td>10.0</td>
</tr>
<tr>
<td>Kit control 1</td>
<td>14.5 ± 1.1</td>
<td>7.6</td>
<td>397.5 ± 22.6</td>
<td>5.7</td>
<td>842.1 ± 84.2</td>
<td>10.0</td>
</tr>
<tr>
<td>Kit control 2</td>
<td>104.5 ± 7.0</td>
<td>6.7</td>
<td>1335.7 ± 56.8</td>
<td>4.3</td>
<td>842.1 ± 84.2</td>
<td>10.0</td>
</tr>
</tbody>
</table>

* $n = 2$ replicates/run, 20 runs, 5 operators, 6–8 days.
* Undetectable (signal higher than zero calibrator).
* CTx assay provides only one kit control.

**Fig. 1.** Intraassay precision profiles of the Dpd (A), NTx (B), and CTx (C) 10pssays determined with six urine samples (n = 26 or 28).

**Fig. 2.** Frequency distributions of the Dpd (A), NTx (B), and CTx (C) analytical values for the 440 first or second morning void urine specimens used in the biological variation studies.

CTx values for six specimens were undetectable; CTx values for 18 specimens greater than the assay’s dynamic range not shown.
tra assay precision. Both the Dpd and NTx assays had intra- and interassay CVs <10% for concentrations greater than the second calibrator [i.e., 3 nmol/L Dpd or 30 nmol bone collagen equivalents (BCE)/L NTx]. The CTx assay exhibited poor precision at lower concentrations, but precision improved for analyte concentrations greater than the third calibrator (200 μg/L).

The frequency distribution of the 440 first and second morning void urines collected for the biological variability studies is depicted in Fig. 2. Dpd values ranged from 8.37 to 207.74 nmol/L. NTx values ranged from 31.6 to 2793.0 nmol BCE/L. CTx values ranged from undetectable (six specimens) to 14 853 μg/L. CTx values for 22 specimens (6.4%) were below the concentration at which CTx intraassay imprecision exceeded 20%.

**Linearity and supplement recovery.** Fig. 3 shows the recovery of the kits’ highest calibrators and four urine samples diluted in assay buffer (Dpd and NTx) or buffer (zero) calibrator (CTx). The Dpd assay demonstrated acceptable recovery with either urine (Table 2) or buffer (Fig. 3) as diluent. When buffer was used as diluent, the NTx assay exhibited linear recovery for the kit calibrator but not for urine specimens (Fig. 3). This nonlinear recovery was corrected with dilutions made in a urine of low analyte concentration (Table 2). Nonlinear recovery of the CTx assay observed at lower concentrations when using the kit’s zero calibrator may be due to imprecision of the assay (Fig. 3). The average recoveries of the three assays were within 100% ± 12% when urine of low analyte concentration was used as diluent. All three assays demonstrated supplement recovery within 100% ± 20% (Table 3).

### Table 2. Dilution recovery of eight urine samples diluted in urine of low analyte concentration.

<table>
<thead>
<tr>
<th>Dilution</th>
<th>Dpd</th>
<th>NTx</th>
<th>CTx</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:2</td>
<td>100 (10)</td>
<td>88 (8)</td>
<td>97 (12)</td>
</tr>
<tr>
<td>1:4</td>
<td>94 (2)</td>
<td>93 (5)</td>
<td>94 (7)</td>
</tr>
<tr>
<td>1:8</td>
<td>92 (2)</td>
<td>94 (12)</td>
<td>102 (1)</td>
</tr>
<tr>
<td>1:16</td>
<td>96 (5)</td>
<td>93 (10)</td>
<td>110 (5)</td>
</tr>
</tbody>
</table>

### Table 3. Supplement recovery (mean) of the two highest kit calibrators in nine urine samples.

<table>
<thead>
<tr>
<th>Added, nmol/L</th>
<th>Recovery (SD), %</th>
<th>Added, nmol BCE/L</th>
<th>Recovery (SD), %</th>
<th>Added, μg/L</th>
<th>Recovery (SD), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.0</td>
<td>97 (18)</td>
<td>105</td>
<td>102 (10)</td>
<td>176</td>
<td>104 (20)</td>
</tr>
<tr>
<td>30.9</td>
<td>95 (8)</td>
<td>298</td>
<td>107 (6)</td>
<td>509</td>
<td>103 (10)</td>
</tr>
</tbody>
</table>

### Table 4. Freeze–thaw stability: mean recovery (SD) of eight urine samples thawed at two temperatures.

<table>
<thead>
<tr>
<th>Freeze/thaw cycle number</th>
<th>37 °C</th>
<th>25 °C</th>
<th>37 °C</th>
<th>25 °C</th>
<th>37 °C</th>
<th>25 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>106 (4)</td>
<td>104 (8)</td>
<td>88 (6)</td>
<td>102 (10)</td>
<td>112 (19)</td>
<td>97 (6)</td>
</tr>
<tr>
<td>3</td>
<td>97 (5)</td>
<td>106 (9)</td>
<td>94 (3)</td>
<td>97 (6)</td>
<td>104 (21)</td>
<td>99 (20)</td>
</tr>
<tr>
<td>4</td>
<td>103 (5)</td>
<td>102 (10)</td>
<td>95 (4)</td>
<td>97 (7)</td>
<td>121 (36)</td>
<td>113 (56)</td>
</tr>
<tr>
<td>5</td>
<td>102 (6)</td>
<td>97 (8)</td>
<td>98 (3)</td>
<td>97 (5)</td>
<td>104 (30)</td>
<td>92 (15)</td>
</tr>
</tbody>
</table>
**Freeze–thaw stability.** None of the analytes exhibited any significant change after the freeze–thaw treatment up to five cycles (Table 4).

**Biological variability evaluation**

**Day-to-day variation.** Day-to-day variability (CV) in 18 individuals was 16.3% for creatinine-corrected Dpd, 23.1% for NTx, and 22.8% for CTx. Mean daily variation of creatinine-corrected Dpd, NTx, and CTx values is shown in Table 5.

**Diurnal variation.** Fig. 4 shows the mean diurnal variation in creatinine-corrected Dpd, NTx, and CTx for 38 healthy subjects. All three assays exhibited similar diurnal rhythm, with peak excretion between 0400 to 0700 and a nadir between 1300 to 1600. The peak was 20–28% greater than the 24-h mean, and the nadir was 15–29% below the 24-h mean for the three bone resorption assays. The rhythm was statistically significant ($P < 0.0001$) for all three assays. The mean amplitude (peak to nadir) was 37% for both Dpd and NTx, and 57% for CTx. CVs within individual subjects were 10.4% for Dpd, 26.7% for NTx, and 28.9%, for CTx.

**Population variation.** Mean, standard deviation, and CV of 216 healthy premenopausal women are presented in Table 6. Six women with undetectable CTx values were not included in the calculations.

**Discussion**

Considerable efforts have been made to demonstrate the clinical utility of the pyridinium cross-links as markers of bone resorption [2, 29, 30]. The new immunoassays are much simpler and easier than the complicated and cumbersome HPLC methods [18, 19]. However, analytical, biological, and clinical differences between the immunoassays have been reported [9, 31–35]. Proper interpretation of assay results determined for these bone resorption markers requires an understanding of their analytical performance and biological variability. We have assessed these parameters in a comparison of three commercially available bone resorption assays.

Ideally, bone resorption markers and the assays used to measure them should possess minimal and predictable biological variability and acceptable analytical performance to provide useful clinical information. Biological

---

**Table 5. Day-to-day variation.**

<table>
<thead>
<tr>
<th></th>
<th><strong>Dpd</strong></th>
<th></th>
<th></th>
<th><strong>NTx</strong></th>
<th></th>
<th></th>
<th><strong>CTx</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean, mmol/mmol Cr</td>
<td>Mean CV, %</td>
<td>CV range, %</td>
<td>Mean, mmol BCE/mmol Cr</td>
<td>Mean CV, %</td>
<td>CV range, %</td>
<td>Mean, µg/mmol Cr</td>
</tr>
<tr>
<td>All (18)*</td>
<td>4.56</td>
<td>16.3</td>
<td>7–25</td>
<td>47.7</td>
<td>23.1</td>
<td>13–35</td>
<td>240.0</td>
</tr>
<tr>
<td>Men (11)</td>
<td>4.06</td>
<td>16.0</td>
<td>7–25</td>
<td>48.9</td>
<td>24.0</td>
<td>13–33</td>
<td>257.3</td>
</tr>
<tr>
<td>Women (7)</td>
<td>5.33</td>
<td>17.4</td>
<td>12–24</td>
<td>45.7</td>
<td>22.9</td>
<td>15–35</td>
<td>212.8</td>
</tr>
</tbody>
</table>

*a Number of subjects.
Cr, creatinine.
characteristics are inherent in each marker but are affected by the analytical determination of marker values. Analytical performance characteristics such as precision, linearity, and recovery support the reliability of the tests. Highly specific and sensitive antibodies, pure and discrete calibrators, and assay design determine the analytical capability of the assay method.

On the basis of these criteria and results of this study, only the Dpd assay meets all the analytical and biological requirements. The CTx assay lacked precision at lower analyte concentrations at which >6% of clinical specimens were found. The NTx assay demonstrated nonlinear recovery with buffer as a diluent. The manufacturer recommends that specimens with values exceeding the highest kit calibrator be diluted in urine of low analyte concentration and not assay buffer. Use of urine as diluent did obviate the nonlinearity of dilution within the assay’s dynamic range. However, the results with buffer as diluent suggest that the kit’s calibrators may be different from immunoreactive N-telopeptides detected in urine, or the composition of the standard matrix contributes to quantification differences compared with urine. All three assays demonstrated good supplement recovery and good analyte stability up to five freeze–thaw cycles.

In 18 healthy subjects, Dpd demonstrated 16% average day-to-day CV; NTx and CTx each demonstrated 23%. Free Dpd determined by immunoassay in this study exhibited less day-to-day variation than has been reported for total Dpd measured by HPLC [36]. The difference may be due to methodology, subject differences, or the contribution of high peptide-bound Dpd variability to the total measurement as suggested by the present study. Others have also reported that Dpd exhibits lower day-to-day variability than NTx and CTx [31–33]. Lower day-to-day variation is likely to provide more reliable results for a patient’s single visit to the physician’s office.

The mean amplitude of the diurnal variation was 37% for creatinine-corrected Dpd and NTx, and 57% for CTx. All three bone resorption markers showed similar rhythm, with high values during early morning and low values during the afternoon. The amplitude of pyridinium cross-links diurnal variation has been reported greater by HPLC methods [37–39]. Diurnal variations in NTx, both average amplitude and within individuals, observed in the present study confirm previous reports [39, 40]. CTx diurnal variability results may be impaired by the assay’s analytical limitations in addition to the rhythm. Since bone resorption assays will reflect the circadian rhythm of bone resorption, a spot urine collected without regard for timing may yield inaccurate results. Only urine collected at a consistent time of day will provide clinically relevant information for patient follow-up and interpretation relative to an appropriate reference interval.

Low biological variability within a reference interval for healthy individuals is necessary to accurately distinguish normal from decreased or increased bone resorption. Population variability in a large group of premenopausal women was 28% for Dpd and >55% for NTx and CTx. The relative differences in population variation between these assays in our study are consistent with previous reports [32, 35].

In summary, our data suggest that the three commercial assays for measuring type I collagen degradation products differ slightly in analytical performance. The CTx assay requires additional sensitivity for detection of lower concentrations of analyte. Acceptable dilution recovery of the NTx assay requires the use of a urine of low analyte concentration as diluent. The biological variability studies also indicate differences in the three assays. All exhibit comparable diurnal variation on average with peak excretion in the early morning and the nadir in the afternoon. Within individuals, however, variability of repeated measures across a 24-h period for Dpd is half the variability observed for NTx and CTx. The assays also do not appear comparable in terms of day-to-day variation within individuals or between individuals in a reference population. These findings suggest that Pyrilinks-D provides the most consistent and reliable laboratory results among these three bone resorption assays. The impact these differences in assay variability have on their utility in identifying individuals with high rates of bone resorption or monitoring antiresorptive agents needs to be investigated.

References

7. Seibel MJ, Duncan A, Robins SP. Urinary hydroxypyridinium


