

# Veterinary Clinical Pathology

An International Journal of Laboratory Medicine

## Evaluation of the analytical variability of dipstick protein pads on canine urine

|                               |  |
|-------------------------------|--|
| Journal:                      | <i>Veterinary Clinical Pathology</i>   |
| Manuscript ID                 | Draft  |
| Manuscript Type:              | Original Article   |
| Date Submitted by the Author: | n/a  |
| Complete List of Authors:     | Giraldi, Marco; University of Milan, Department of Veterinary Medicine; University of Milan, Veterinary Teaching Hospital Paltrinieri, Saverio; University of Milan, Department of Veterinary Medicine; University of Milan, Veterinary Teaching Hospital Zatelli, Andrea; Medical Consultancy Services, |
| Key Words:                    | Dogs, Proteinuria, Reagent strips, Urinalysis  |
|                               |  |

SCHOLARONE™  
Manuscripts

View

1 **Evaluation of the analytical variability of dipstick protein pads on canine urine**

2

3 Dipstick analysis of canine proteinuria

4

5 Marco Giraldi,<sup>a,b</sup> DVM

6 Saverio Paltrinieri,<sup>a,b\*</sup> DVM PhD Dipl ECVCP

7 Andrea Zatelli<sup>c</sup>, DVM PhD

8

9 <sup>a</sup>*Department of Veterinary Medicine – University of Milan, Via Celoria, 10 - 20133*

10 *Milan, Italy*

11 <sup>b</sup>*Veterinary teaching hospital – University of Milan, Via dell'Università 6 – 26900 Lodi,*

12 *Italy*

13 <sup>c</sup>*Medical Consultancy Services, Ta'Xbiex, Malta*

14

15

16 \* Corresponding author: Dr. Paltrinieri

17 *Department of Veterinary Medicine – University of Milan,*

18 *Via Celoria, 10 - 20133 Milan, Italy*

19 *Tel.: +39 0250318103.*

20 *E-mail address: saverio.paltrinieri@unimi.it*

21

**22 Abstract**

23 Background: The dipstick is a first line and inexpensive test to exclude the presence  
24 of proteinuria in dogs. No information is available on the analytical variability of  
25 dipstick analysis

26 Objectives: The aim of this study was to assess the analytical variability of two  
27 dipsticks and the inter-operator variability in dipstick interpretation

28 Methods: Canine urine supernatants (n=174) were analyzed with two commercially  
29 available dipstick. Two observers evaluated each result blinded to the other observer  
30 and to the results of the other dipstick. Intra and inter-assay variability were assessed  
31 on 5 samples (corresponding to the 5 different semi-quantitative results) tested for 10  
32 consecutive times and 5 consecutive days, respectively. The variability between  
33 observer and between dipsticks was evaluated with Cohen's k test.

34 Results: Intra-assay repeatability was good ( $\leq 3/10$  errors), whereas inter-assay  
35 variability was higher (from 1/5 to 4/5 discordant results). The concordance between  
36 operators (k=0.68 and 0.79 for the two dipsticks) and between dipsticks (k=0.66 and  
37 0.74 for the two operators) was good. However, one observer and one dipstick  
38 overestimated the results compared with the other observer or dipstick. In any case,  
39 discordant results accounted for a single unit of the semi-quantitative scale.

40 Conclusions: As for any other method, analytical variability may affect the semi-  
41 quantification of urinary proteins with dipstick. Subjective interpretation of the pad  
42 and, to a lesser extent, intrinsic staining properties of the pads could affect the  
43 results. Further studies are warranted in order to evaluate the effect of this variability  
44 on clinical decision.

45

46 **Keywords:** Dogs, Proteinuria, Reagent strips, Urinalysis

## 47 **Introduction**

48 In clinical practice proteinuria is defined as the increased amount of proteins in the  
49 urine. The presence of persistent proteinuria of renal origin has a diagnostic and  
50 prognostic value for chronic kidney disease (CKD) in dogs. The detection of  
51 proteinuria is therefore a milestone in the management of dogs with CKD.<sup>1</sup>  
52 The recommended methods to evaluate proteinuria in dogs is the quantitative  
53 evaluation of the urinary proteins and urinary creatinine, followed by the calculation of  
54 the urinary protein to creatinine (UPC) ratio, that allows to correct the magnitude of  
55 proteinuria by the dilution of urine.<sup>1,2</sup> However, in clinical practice, dry reagent test  
56 strips (dipsticks) are rapid and inexpensive methods that allow, along with other  
57 urinary physico-chemical or cytological parameters, a first evaluation of the presence  
58 or absence of proteins in urine in a point-of-care setting.<sup>3</sup> Proteins (mainly albumin),  
59 when present, react with the pad yielding a variable color change whose intensity is  
60 proportional to the protein concentration. Results are then expressed semi-  
61 quantitatively, usually as negative, trace or 1+ to 4+ (corresponding, for most of the  
62 commercially available dipsticks to 15 to 2000 mg/dL of proteins), comparing the pad  
63 against the chart on the side of the dipstick package or, alternatively, loading the strip  
64 in an automated spectrophotometric reader.  
65 Although the evaluation of protein excretion using the dipstick must be considered a  
66 screening test, erroneous interpretations of the pad may affect clinical decisions. In  
67 human medicine information from quality assurance programs revealed that the rates  
68 of misclassification of one and two scores above or below the expected value were  
69 9.7% and 2.3%, respectively,<sup>4</sup> and that the intra- and inter-observer agreement is  
70 moderate ( $k=0.53$ ) to very good ( $k=1$ ).<sup>5</sup> Moreover, the use of automated reader is  
71 recommended to minimize the observer-related errors<sup>6</sup> despite automated readings

72 only slightly improved the reproducibility of dipstick analysis.<sup>5</sup> In veterinary medicine,  
73 although similar results are likely, and quality control programs for urinalysis and  
74 dipstick tests are recommended,<sup>2</sup> little is known about the influence of inter-operator  
75 variability on the analytical variability of dipstick testing or about the intrinsic  
76 performances of the dipstick from different manufacturers, that may have their  
77 peculiar analytical sensitivity and range or different semi-quantitative interpretation  
78 charts.

79 Therefore, the aim of this study was to assess the analytical variability of dipstick  
80 analysis in the evaluation of the presence/absence of proteins in canine urine,  
81 through the comparison of results obtained using two commercially available  
82 dipsticks by two different independent operators.

83

#### 84 **Materials and Methods**

85 A total of 174 canine urine supernatants were included in this study.

86 Urine samples were collected over a period of 12 months from dogs of different age,  
87 sex and breed, presented for diagnostic investigation at the internal medicine unit of  
88 the Department of Veterinary Medicine (DIMEVET, University of Milan), by mean of  
89 cystocentesis (n=113), free catch (n=47) or unspecified method (n=14). Samples  
90 were collected for diagnostic purposes under informed consent of the owner and  
91 therefore, according to the institutional Ethical Committee (deliberation number:  
92 2/2016) a formal approval of the study from the Institutional Animal Care and Use  
93 Committee was not necessary.

94 Since this was a validation study not focused on the impact of the results on the  
95 clinical diagnosis, samples were included irrespective of health status of the dogs.

96 All urine samples underwent complete urinalysis (including USG, dipstick evaluation  
97 and sediment examination) within two hours from collection. In order to perform  
98 sediment analysis, 5 mL of sample were placed in sterile conical 10 mL tube and  
99 centrifuged 5 minutes at 1250 rpm (450 G). Supernatants were aliquoted in a 1.5 mL  
100 tube and stored at -20°C within 4 hours from collection for a maximum of 12 months.  
101 At the time of analysis, supernatants were gently thawed by transferring tubes at 4°C  
102 the day before analysis and then at room temperature one hour before analysis.  
103 Each sample was tested with two commercially available dipsticks (Dipstick 1: U-11  
104 Urine Strips, Mindray, Shenzhen, China; Dipstick 2: Multistix 10 SG Reagent Strips,  
105 Siemens, Siemens Healthcare Diagnostics Inc, Tarrytown, NY /Siemens Healthcare  
106 Diagnostics, Eschborn, Germany) after a preliminary assessment of intra- and  
107 interassay variability of each dipstick (see below). Moreover, each dipstick was  
108 evaluated by two operators with similar experience in urinalysis procedures.  
109 In each analysis, the dipstick was kept out from the case no more than 2 minutes  
110 before the use. Fifty microliters of urines supernatant were applied with a  
111 dispensable pipette only on the protein pad and, in order to prevent cross-  
112 contamination by dyes from close pads the contact of urine with adjacent pads was  
113 avoided; then, excess urine was gently discarded hitting the dipstick on a clean paper  
114 towel.

115 In order to avoid bias in interpretation of the second dipstick by each operator and to  
116 avoid excess time between urine application and reading, samples were analyzed in  
117 batches (8 samples per batch), thus allowing the evaluation within 60-120 seconds,  
118 as recommended by the manufacturer's instruction of both the dipstick kits, and urine  
119 samples were applied to the second dipstick in a different random order compared  
120 with that used for the first dipstick. Each operator interpreted the dipsticks results

121 blinded to the results of the other operator. Also, due to the different random order  
122 described above, the interpretation of the second dipstick was blinded to results of  
123 the first dipstick.

124 Interpretation of each dipstick was performed by comparing the color of the protein  
125 pads with the corresponding color chart provided by each manufacturer.

126 Because, as expected, some samples yielded a color reaction with a chromatic  
127 intensity not identical to those proposed on the chart, the following reading method  
128 was chosen: each pad was compared with one reference color block at a time; when  
129 an almost perfect match between the pad and the block on the chart was identified,  
130 the corresponding result was assigned (i.e negative or trace - N/T – or positive: 1+,  
131 2+, 3+ or 4+); when the color intensity was intermediate between two blocks on the  
132 chart, the results corresponding to the nearest reference color blocks (lighter or  
133 darker) was assigned. However, in these cases, the samples were also recorded as  
134 “difficult”.

135

### 136 **Intra-assay and inter-assay precision**

137 Five urine samples yielding results corresponding to the 5 different semi-quantitative  
138 scores of proteinuria (namely N/T = <30 mg/dL, 1+ = 30 mg/dL, 2+ = 100 mg/dL, 3+  
139 = 300 mg/dL, 4+ =  $\geq 2$  g/dL) were selected and used for analytical precision tests.

140 For intra-assay evaluation, the 5 samples were tested with both dipsticks 10  
141 consecutive times.

142 Then, inter-assay variability was assessed testing the same samples 5 times in 5  
143 consecutive days by both dipsticks, storing samples at 4°C overnight between the  
144 evaluations.

145 For both intra- and inter-assay evaluations, interpretation of pads was performed with  
146 the same method described for the whole set of samples.

147 For each semi-quantitative score, numbers of different results were counted and the  
148 percentage of the results lower (underestimation) or higher (overestimation) than the  
149 first reading were calculated.

150 Imprecision was expressed as the percentage of different results out of the sum of  
151 the two operators (i.e 20 reading/results for intra-assay evaluation and 10 results for  
152 inter-assay evaluation)

153

#### 154 **Statistical analysis**

155 Concordance between operators and between dipsticks was tested with Cohen's  
156 Kappa test<sup>7</sup> and was calculated either for the whole set of results (N/T, 1+, 2+, 3+,  
157 4+) or grouping results as  $\leq 1$  and  $\geq 2$ .

158 Moreover, for the evaluation of the concordance between dipsticks, the results of the  
159 two operators were combined in order to reach a consensus and, in case of  
160 discordant results, further intermediate categories were created (N/T-1+, 1+-2+,  
161 2+-3+, 3+-4+).

162 The k coefficients were used to define the concordance as follows: 0.00-0.20,  
163 0.21-0.40, 0.41-0.60, 0.61-0.80 and 0.81-1.00 represented poor, fair, moderate,  
164 good and very good concordance<sup>8</sup>, respectively.

165 In order to quantify the rate of discordant results at different level of positivity, the  
166 number of discordant sample yielded between two scores (e.g. between N/T and 1+  
167 or between 1+ and 2+ etc.) were counted and the percentage was calculated out of  
168 the total number of samples found among the two scores evaluated.



169 Descriptive statistics were performed with Excel software and the Analyze-it  
170 statistical software (Analyze-it Software Ltd, Leeds, West Yorkshire, England) was  
171 used to assess the level of concordance (Cohen's  $k$ )

172

## 173 **Results**

### 174 **Intra-assay variability**

175 Dipstick 1 always provided the same results recorded at first reading, except in two  
176 cases: operator 1 overestimated one 3+ sample (10%) and Operator 2  
177 underestimated one 2+ sample (10%) Difficult interpretations were more frequent for  
178 Operator 1 (5/10 at 1+; 2/10 at 2+ and 3+) than for Operator 2 (4/10 at 2+).

179 Using Dipstick 2, Operator 1 overestimated one 1+, one 2+ and one 3+ sample (10%  
180 each), while Operator 2 underestimated three 2+ and three 3+ samples (30% each).  
181 Only Operator 1 recorded difficult interpretations (3/10 at 1+ and 2+ and 1/10 at 3+).

182

### 183 **Inter-assay variability**

184 With Dipstick 1, Operator 1 overestimated one 1+ sample (20%) and Operator 2  
185 overestimated one 1+ sample (20%) and underestimated three 2+ samples (60%)  
186 and one 3+ sample (20%). Difficult interpretations were rare either for Operator 1  
187 (1/10 at N/T and at 2+) or for Operator 2 (1/10 at 3+).

188 Imprecision was more frequent for Dipstick 2: Operator 1 overestimated four 1+  
189 (40%) and four 3+ samples (80%), while Operator 2 overestimated one 1+ (10%),  
190 one 2+ (10%) and three 3+ samples (30%). Only Operator 1 recorded difficult  
191 interpretations (1/10 at 2+ and 3+).

192

### 193 **Analysis of samples**

194 Samples covered all possible results of the dipsticks but results with scores lower  
195 than 2+ were more frequent for both the dipsticks, as follows.

196 Using Dipstick 1, samples recorded as N/T, 1+, 2+, 3+ and 4+, were respectively 96,  
197 50, 17, 10 and 1 for Operator 1 and 112, 35, 16, 10 and 1 for Operator 2 (Table 1).  
198 Operators 1 and 2 recorded 23 (13,2%) and 22 (12.6%) difficult interpretations,  
199 respectively. These uncertain interpretations were mostly between negative and  
200 trace or between trace and 1+ either for Operator 1 (n=10, and n=10, respectively) or  
201 Operator 2 (n=12 and n=6, respectively). The total number of samples that were  
202 difficult for both operators was 39 (22.4%).

203 Using Dipstick 2, samples recorded as N/T, 1+, 2+, 3+ and 4+, were respectively 84,  
204 53, 19, 12 and 6 for Operator 1 and 105, 37, 17, 11 and 4 for Operator 2 (Table 1).  
205 Operators 1 and 2 recorded 23 (13,2%) and 32 (18.4%) difficult interpretations,  
206 respectively. These uncertain interpretations were mostly between negative or trace,  
207 between trace and 1+ and between 1+ and 2+ either for Operator 1 (n=7, n=8, and  
208 n= 6, respectively) or 2 (n=11, n=14 and n=3, respectively). The total number of  
209 samples that were difficult for both operators was 46 (26.4%).

210

### 211 **Concordance between operators**

212 The concordance between operators was “good” for both dipsticks, and slightly  
213 higher for Dipstick 1 (k=0.79) than for Dipstick 2 (k=0.68). Discordant results were  
214 found mostly at lower protein concentrations for dipstick 1 (N/T vs 1+: 11% of  
215 misclassifications, 1+ vs 2+: 6% of misclassifications; 2+ vs 3+: 8% of  
216 misclassifications) but at all the protein concentrations for dipstick 2 (N/T vs 1+: 17%  
217 of misclassifications, 1+ vs 2+: 14% of misclassifications; 2+ vs 3+: 12% of

218 misclassifications; 13% of misclassifications). Moreover, for both the dipsticks  
219 Operator 1 tended to provide higher scores compared with Operator 2.  
220 When results were grouped as  $\leq 1+$  and  $\geq 2$ , concordance improved consistently to  
221 the “very good” category (Dipstick 1  $k=0.94$ ; Dipstick 2  $k=0.87$ ) and only 3 and 7  
222 discordant results were misclassified with Dipstick 1 and Dipstick 2, respectively.  
223

#### 224 **Concordance between dipsticks**

225 The concordance between the two dipsticks was “good” for both operators (Operator  
226 1  $k=0.66$ ; Operator 2  $k=0.74$ ). Similarly to the inter-operator variability described  
227 above, discordant results were more frequent for Observer 1 (N/T vs 1+: 12% of  
228 misclassifications, 1+ vs 2+: 16% of misclassifications; 2+ vs 3+: 30% of  
229 misclassifications; 3+ vs 4+: 45% of misclassifications) than for Observer 2 (N/T vs  
230 1+: 11% of misclassifications, 1+ vs 2+: 11% of misclassifications; 2+ vs 3+: 17% of  
231 misclassifications; 3+ vs 4+: 27% of misclassifications)  
232 When results were grouped as  $\leq 1+$  and  $\geq 2+$ , concordance improved to “very good”  
233 category ( $k=0.83$  and  $=0.90$  for Operator 1 and Operator 2, respectively), showing 9  
234 and 5 samples misclassified by dipsticks ( $\geq 2$  using Dipstick 2 and  $\leq 1$  using Dipstick  
235 1) with Operator 1 and Operator 2, respectively.

236 Overall Dipstick 2 tended to provide higher scores compared with Dipstick 1.

237 Concordance between dipsticks using the consensual agreement between operators  
238 (Table 2) was defined as “moderate” ( $k=0.59$ ). Misclassifications were 22.5% in the  
239 interval N/T vs 1+, 42.8% in 1+ vs 2+, 66.6% in 2+ vs 3+ and 66.6% in 3+ vs 4+.  
240 Again, grouping results as  $< 1$ , 1-2,  $> 2$ , concordance improved to “good” ( $k=0.76$ ).

241

#### 242 **Discussion**

243 This study, using two commercially available dipsticks and two operators,  
244 demonstrated that a variable imprecision in the evaluation of the concentration of  
245 urinary proteins exist. The study design adopted to assess the analytical variability  
246 (e.g. application of urine by the same operator, disposal of urine only on the protein  
247 pad, analysis of 8 samples per batch, randomization of readings etc) prevented the  
248 effect of other possible sources of error and variability such as oxidation of dipsticks,  
249 insufficient amount of urine on the pad, excess urine with contamination from other  
250 pads, colour changes due to delay of readings, interpretative biases due to the  
251 sequential analysis of samples by the same operator, etc.<sup>4,9,10</sup>

252 Intra-assay imprecision, however, was similar in magnitude to that reported in human  
253 medicine where the reproducibility of visual reading was 68-85%,<sup>11</sup> and was variable  
254 between the operators. As for any other test,<sup>12</sup> inter-assay variability was even  
255 higher. Nevertheless, this inter-assay variability has a low clinical importance  
256 because rarely repeated readings are performed during different days.

257 The imprecision may depend on intrinsic characteristics of the pads, on the visual  
258 perception of the operators, on environmental factors (e.g. different light during days)  
259 or, as regards inter-assay variability, on changes of pH or of protein concentration  
260 induced by refrigeration, as shown in studies on the protein to creatinine ratio.<sup>13</sup>

261 However, the dipstick is analytically less sensitive than wet chemistry and no effects  
262 of storage on pH were reported in dogs<sup>14</sup> and therefore storage artifacts are unlikely.

263 Moreover, imprecision tests evidenced a high frequency of samples that were difficult  
264 to interpret, especially at low scores (i.e. N/T and 1+) and with Dipstick 2. For both  
265 the dipsticks, the two operators found difficulty with different samples. This points out  
266 that the difficult interpretation was operator-dependent instead of sample- or dipstick-  
267 dependent. However, to our impression, the two dipstick had slight differences in the

268 colour of the pad before the application of urines and in the hue after reaction with  
269 the same samples (especially at low protein concentration). This difference may  
270 complicate the interpretation of results. Whether sample-dependent factors (e.g.  
271 physical or chemical properties of urine such as urine dilution or presence of  
272 pigments) could affect the agreement between dipsticks needs further evaluation.  
273 This study showed also that the inter-operator concordance was not perfect, due to  
274 discordant results at all the levels of positivity. The degrees of concordance recorded  
275 in this study were similar to that found in people, where a k coefficient of 0.82 was  
276 found<sup>5</sup> but lower than that reported in dogs (k=0.92).<sup>15</sup>  
277 The majority of the discordant results were due to an overestimation of results by one  
278 operator compared with the other. Also in human medicine a tendency to  
279 overestimate or underestimate protein pads by single operators was demonstrated,<sup>11</sup>  
280 likely due to the different visual perception mentioned above. Moreover, although the  
281 reading method was standardized between the operators, each operator could  
282 consistently perceive as closer to the lower or the higher score the color reactions  
283 that were intermediate to those shown on the chart, as already demonstrated.<sup>10</sup>  
284 Similarly to the results of inter-operator variability, sub-optimal concordances  
285 between dipsticks were found with both operators. Interestingly, Dipstick 2 tended to  
286 provide higher scores compared with Dipstick 1, and about a quarter of samples  $\geq 2+$   
287 by Dipstick 2 were classified as 1+ by Dipstick 1. This result points out that, although  
288 the two manufactures declared the same analytical sensitivity (15 mg/dL) and  
289 reported equal protein concentrations for the 4 blocks on the chart, slight differences  
290 in biochemical reaction may exist between different commercially available dipsticks.  
291 According to a previous study in dogs,<sup>16</sup> samples with negative dipstick are likely non  
292 proteinuric, samples with 2+ or more are likely proteinuric and samples with 1+ may

293 or may not be proteinuric depending on the USG. Therefore, misclassification of  
294 samples as N or 1+ and 1+ or 2+ could be of clinical significance, while  
295 misclassification of samples with 2+, 3+, or 4+ may be less relevant on a clinical  
296 point of view, since the calculation of the UPC ratio, that is more accurate than  
297 dipstick, is recommended for any sample with results  $\geq 2+$ .

298 The increase of agreement grouping results as  $\leq 1+$  or  $\geq 2+$  showed that the inter-  
299 operator variability decreases. In other words, samples with more than 2+, likely  
300 corresponding to proteinuric dogs according to a previous study,<sup>16</sup> can be correctly  
301 identified independently on the operator, while this results confirms that  
302 discrepancies mainly occurred between N/T results and 1+, that according to the  
303 study cited above should be considered respectively as definitely on proteinuric and  
304 dubious, i.e. proteinuric or not depending on the USG. In practical terms this may  
305 indicate that the analytical and inter-operator variability does not affect the sensitivity  
306 of the dipstick but it may affect the specificity of the method. Therefore, further  
307 research on samples with known USG and UPC ratio, which were not available in all  
308 the cases included in the present study, is needed in order to evaluate the diagnostic  
309 performances of the two dipsticks as well as to understand which of the two dipsticks  
310 employed in this study is more accurate and which over- or underestimate proteinuria  
311 compared with the UPC ratio.

312 Interestingly, in any case errors were higher than one score. Studies in human  
313 medicine reported that errors higher than one score are possible in clinical practice  
314 and may account up to 2.4%<sup>4,9,11</sup> However, those studies evaluated not only the  
315 analytical variability (as in our study) but also the effect of preanalytical and  
316 postanalytical errors on variability. Evaluation of such a variability was beyond the  
317 aims of this study but it could be speculated that also in veterinary practice,

318 preanalytical errors may occur and, along with the analytical variability reported in the  
319 present study, may induce misclassifications higher than one score of positivity.

320 In conclusion, although dipstick is considered simple and intuitive test, analytical  
321 variability may affect the interpretation of results, as well as for any other diagnostic  
322 tests. Both the imprecision and the difficulty of interpretation may depend either on  
323 intrinsic factors of the pads or on different capability of the operators. The effect of  
324 these variables could be considerable in misclassification of samples between two  
325 contiguous scores at any level of positivity but misclassification of results between  
326 N/T, 1+ and 2+ could be of clinical significance and therefore should be interpreted  
327 with caution and confirmed with more sensitive methods such as the UPC ratio.  
328 Further studies are warranted in order to evaluate whether automated readers may  
329 reduce this variability, or to determine the accuracy of these dipsticks in comparison  
330 with a gold standard method such as the UPC ratio and the effect of the variability  
331 quantified in this study on clinical decisions.

332

### 333 **Conflict of interest statement**

334 None of the authors of this paper has a financial or personal relationship with other  
335 people or organizations that could inappropriately influence or bias the content of the  
336 paper.

337

### 338 **Acknowledgements:**

339 The Authors are grateful to Prof. Paola Scarpa that provided most of the samples  
340 included in this study and Dr. Tiziana Vitiello that performed routine urinalysis.

341

### 342 **References**

- 343 1. Grauer GF. IRIS: Proteinuria. Available at: <http://www.iris->  
344 [kidney.com/education/proteinuria.html](http://www.iris-kidney.com/education/proteinuria.html). Accessed January 22th, 2016.
- 345 2. Gunn-Christie RG, Flatland B, Friedrichs KR, et al. ASVCP quality assurance  
346 guidelines: control of preanalytical, analytical, and postanalytical factors for  
347 urinalysis, cytology, and clinical chemistry in veterinary laboratories. *Vet Clin*  
348 *Pathol.* 2012;41:18-26
- 349 3. Stockham SL, Scott MA. *Fundamentals of veterinary clinical pathology*. 2nd ed.  
350 Ames, IA: Wiley-Blackwell; 2008:415-494
- 351 4. Tighe P. Laboratory-based quality assurance program for near-patient urine  
352 dipstick testing, 1990-1997: development, management and results. *Br J Biomed*  
353 *Sci.* 1999;56:6-15.
- 354 5. Winkens RA, Leffers P, Degenaar CP, Houben AW. The reproducibility of  
355 urinalysis using multiple reagent test strips. *Eur J Clin Chem Clin Biochem.*  
356 1991;29:813-818.
- 357 6. European Confederation of Laboratory Medicine. European urinalysis guidelines.  
358 *Scand J Clin Lab Invest Suppl.* 2000;231:1-86.
- 359 7. Kouri T, Laippala P, Kutter D, Gant V. Quality specifications for ordinal scale  
360 measurements with multiproperty (multiple) urine test strips. *Scand J Clin Lab*  
361 *Invest.* 1999;59:523-526
- 362 8. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical  
363 Data. *Biometrics.* 1977;33:159–174.
- 364 9. Tighe P. Urine dry reagent strip “error” rates using different reading methods.  
365 *Accred Qual Assur.* 2000;5:488-490.
- 366 10. Berg B, Hellsing K, Jagenburg R, Kallner A. Guidelines for evaluation of reagent  
367 strips. Exemplified by analysis of urine albumin and glucose concentration using



- 368 visually read reagent strips. International Federation of Clinical Chemistry,  
369 Scientific Division. *Scand J Clin Lab Invest.* 1989;49:689-699.
- 370 11. Rumley A. Urine dipstick testing: comparison of results obtained by visual  
371 reading and with the Bayer CLINITEK 50. *Ann Clin Biochem.* 2000;37:220-221.
- 372 12. Westgard JO. *Basic method validation.* 2nd ed. Madison, Wis: Westgard QC Inc,  
373 2003.
- 374 13. Rossi G, Giori L, Campagnola S, Zatelli A, Zini E, Paltrinieri S. Evaluation of  
375 factors that affect analytic variability of urine protein-to-creatinine ratio  
376 determination in dogs. *Am J Vet Res.* 2012;73:779-788
- 377 14. Fromm P, Bieganiec B, Ehrenrich Z, Barak M. Stability of common analytes in  
378 urine refrigerated for 24 h before automated analysis by test strips. *Clin Chem.*  
379 2000;46:1384-6.
- 380 15. Bauer N, Rettig S, Moritz A. Evaluation the Clinitek status™ automated dipstick  
381 analysis device for semiquantitative testing of canine urine. *Res Vet Sci.*  
382 2008;85:467-472
- 383 16. Zatelli A, Paltrinieri S, Nizi F, Roura X, Zini E. Evaluation of a urine dipstick test  
384 for confirmation or exclusion of proteinuria in dogs. *Am J Vet Res.* 2010;71:235-  
385 240.
- 386
- 387

388 **Tables and Figures**

389 Table 1 Contingency table of the 174 canine urine samples assayed with dipstick 1

390 (above) and dipstick 2 (below) and interpreted by the two operators. Concordant

391 results are in bold, discordant results are in italics

392

|            |       | Operator 2 |           |           |          |          | Total      |
|------------|-------|------------|-----------|-----------|----------|----------|------------|
|            |       | N/T        | 1         | 2         | 3        | 4        |            |
| Dipstick 1 | N/T   | <b>96</b>  | 0         | 0         | 0        | 0        | 96         |
|            | 1+    | <i>16</i>  | <b>33</b> | <i>1</i>  | 0        | 0        | 50         |
|            | 2+    | 0          | 2         | <b>14</b> | <i>1</i> | 0        | 17         |
|            | 3+    | 0          | 0         | <i>1</i>  | <b>9</b> | 0        | 10         |
|            | 4+    | 0          | 0         | 0         | 0        | <b>1</b> | 1          |
|            | Total | 112        | 35        | 16        | 10       | 1        | <b>174</b> |
| Dipstick 2 | N/T   | <b>83</b>  | <i>1</i>  | 0         | 0        | 0        | 84         |
|            | 1+    | 22         | <b>30</b> | <i>1</i>  | 0        | 0        | 53         |
|            | 2+    | 0          | 6         | <b>13</b> | 0        | 0        | 19         |
|            | 3+    | 0          | 0         | 3         | <b>9</b> | 0        | 12         |
|            | 4+    | 0          | 0         | 0         | 2        | <b>4</b> | 6          |
|            | Total | 105        | 37        | 17        | 11       | 4        | <b>174</b> |

393 Table 2 Raw data of agreement between the two dipsticks in scoring protein pads using the results of the consensus between  
 394 operators. Concordant results are in bold, discordant results are in italics  
 395

|           |           | Dipstick2 |           |           |          |          |          |          |          |          | Total |
|-----------|-----------|-----------|-----------|-----------|----------|----------|----------|----------|----------|----------|-------|
|           |           | N/T       | N/T or 1+ | 1+        | 1+ or 2+ | 2+       | 2+ or 3+ | 3+       | 3+ or 4+ | 4+       |       |
| Dipstick1 | N/T       | <b>82</b> | <i>11</i> | <i>3</i>  | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | 96    |
|           | N/T or 1+ | <i>0</i>  | <b>9</b>  | <i>7</i>  | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | 16    |
|           | 1+        | <i>1</i>  | <i>3</i>  | <b>20</b> | <i>7</i> | <i>2</i> | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | 33    |
|           | 1+ or 2+  | <i>0</i>  | <i>0</i>  | <i>0</i>  | <b>0</b> | <i>3</i> | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | 3     |
|           | 2+        | <i>0</i>  | <i>0</i>  | <i>0</i>  | <i>0</i> | <b>8</b> | <i>3</i> | <i>3</i> | <i>0</i> | <i>0</i> | 14    |
|           | 2+ or 3+  | <i>0</i>  | <i>0</i>  | <i>0</i>  | <i>0</i> | <i>0</i> | <b>0</b> | <i>2</i> | <i>0</i> | <i>0</i> | 2     |
|           | 3+        | <i>0</i>  | <i>0</i>  | <i>0</i>  | <i>0</i> | <i>0</i> | <i>0</i> | <b>4</b> | <i>1</i> | <i>3</i> | 8     |
|           | 3+ or 4+  | <i>0</i>  | <i>0</i>  | <i>0</i>  | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | <b>1</b> | <i>0</i> | 1     |
|           | 4+        | <i>0</i>  | <i>0</i>  | <i>0</i>  | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | <i>0</i> | <b>1</b> | 1     |

---

|       |    |    |    |   |    |   |   |   |   |            |
|-------|----|----|----|---|----|---|---|---|---|------------|
| TOTAL | 83 | 23 | 30 | 7 | 13 | 3 | 9 | 2 | 4 | <b>174</b> |
|-------|----|----|----|---|----|---|---|---|---|------------|

---

396

For Peer Review