# REPRODUCIBILITY OF MODIFIED WALDENSTRÖM CLASSIFICATION IN PERTHES DISEASE

## REPRODUTIBILIDADE DA CLASSIFICAÇÃO DE WALDENSTRÖM MODIFICADA NA DOENÇA DE PERTHES

FELIPPI GUIZARDI CORDEIRO[1] (iD), PATRICIA MORENO GRANGEIRO[1] (iD), BRUNO SÉRGIO FERREIRA MASSA[1] (iD),
NEI BOTTER MONTENEGRO[1] (iD), ROBERTO GUARNIERO[1] (iD)

1. Universidade de São Paulo, Medical School, Hospital das Clínicas, Orthopedics and Traumatology Institute, São Paulo, SP, Brazil.

## ABSTRACT

Objective: The purpose of our study is to evaluate intraobserver and interobserver reliability of modified Waldenström classification system for Legg-Calvé-Perthes disease and assess the influence of the professional's area of expertise in the assessment. Methods: Twelve evaluators assessed 40 pairs of pelvic radiographs of patients with Legg-Calvé-Perthes disease. After two weeks, a new evaluation was performed by the same evaluators. Kappa and Kendall's W indexes were used to evaluate both intraobserver and interobserver reliability and determine the influence of the evaluators' experience and area of expertise. Results: The average intraobserver kappa value was 0.394, with a reasonable agreement level. The interobserver Kappa value was 0.243 in the first evaluation (95% CI, 0.227-0.259 and p < 0.0001) and 0.245 in the second evaluation (95% CI, 0.229-0.260 and p < 0.0001). The Kendall's W values obtained for pediatric orthopedists, radiologists and resident physicians were 0.686, 0.630 and 0.529 (p < 0.0001), respectively. Conclusion: The modified Waldenström classification presented both moderate and reasonable levels of intraobserver agreement, and reasonable level of interobserver agreement. The evaluators' degree of experience and area of expertise influenced the concordance level found. ***Level of Evidence II, Diagnostic Studies – Investigating a Diagnostic Test.***

**Keywords:** Legg-Calve-Perthes Disease. Radiography. Osteonecrosis. Child. Hip Joint.

## RESUMO

*Objetivo: O objetivo deste estudo é avaliar a confiabilidade intraobservador e interobservador do sistema de classificação de Waldenström modificado para a doença de Legg-Calvé-Perthes e avaliar a influência da experiência do profissional na avaliação. Métodos: Doze avaliadores analisaram 40 pares de radiografias pélvicas de pacientes com doença de Legg-Calvé-Perthes. Após duas semanas, nova avaliação foi realizada pelos mesmos avaliadores. Os índices Kappa e Kendall's W foram usados para analisar a confiabilidade intraobservador e interobservador e para determinar a influência da experiência e perícia dos avaliadores. Resultados: O valor médio do kappa intraobservador foi de 0,394, com razoável concordância. O valor de Kappa interobservador foi de 0,243 na primeira avaliação (IC 95%, 0,227-0,259 e p < 0,0001) e 0,245 na segunda (IC 95%, 0,229-0,260 e p < 0,0001). Os valores W de Kendall obtidos para ortopedistas pediátricos, radiologistas e médicos residentes foram 0,686, 0,630 e 0,529 (p < 0,0001), respectivamente. Conclusão: A classificação de Waldenström modificada apresentou níveis moderados e razoáveis de concordância intraobservador e razoável concordância interobservador. O grau de experiência e especialização dos avaliadores influenciou o nível de concordância encontrado. **Nível de Evidência II, Estudos diagnósticos – Investigação de um exame para diagnóstico.***

***Descritores:** Doença de Legg-Calvé-Perthes. Radiografia. Osteonecrose. Criança. Articulação do Quadril.*

## INTRODUCTION

Legg-Calvé-Perthes disease (LCPD) is defined as idiopathic a femoral head osteonecrosis that occurs in childhood. The annual prevalence ranges from 5.1 to 16.9 per 100,000 children between ages 2 and 14 in several regions of the world[1-3] and affects more the boys. The etiology is still unknown, but many theories have been proposed an origin, including trauma, inflammatory process, vascular occlusion, thrombophilia, maternal smoking and Type II collagen mutations. Despite the different theories, most authors agree that it is a multifactorial disease with genetic and environmental factors leading to a common clinical and pathological presentation.[4] Although the etiology remains unconfirmed, the disease key manifestation is the vascular occlusion of the femoral head, causing an ischemic necrosis followed by structural changes that lead to

sphericity loss and abnormal hip growth. Since the vascular occlusion is temporary, the disease is self-limiting. Nevertheless, once revascularization and reossification occurs, its acknowledgment and management are essential to prevent deformity, which may occur due to the femoral head susceptibility to mechanical load during the necrosis phase, and may evolve to early arthrosis, pain and functional loss of the affected hip.[4]

One of LCPD challenges is to predict the disease course based on symptoms' manifestation. Clinical and radiographic events have been used to establish prognostic factors and indicate therapeutic interventions in a timely manner, aiming at reducing deformation risks.

Although magnetic resonance imaging is currently more sensitive and accurate in assessing the extension degree of existent necrosis,[5] radiography remains as the most affordable and used examination. There is still no consensus about the use of radiography to predict the natural course of the disease, and several radiographic rating systems have been used and modified to establish prognostic factors and guide interventions.

In 1922, Waldenström[6] proposed a classification system based on four LCPD radiographic stages: I – initial or avascular stage; II – fragmentation stage; III – reossification or healing stage; and IV – healed stage.

The three most important subsequent classifications were intended to indicate a prognosis and were described in the fragmentation phase: Salter and Thompson,[7] Catterall[8] and Herring et al.[9,10]

However, none of these classifications ensure a prognostic value for a specific patient. Moreover, they present conflicting results in both inter and intraobserver evaluations.[11-17] In 2003, a study by Joseph et al.[13] proposed a new classification, which was meant to be a modified and updated version of the classification proposed by Canale et al.,[14] which became known as "modified Waldenström classification" (1972). This classification seems to be very reproducible,[13,15] according to the authors, and it should be routinely applied to study the evolution of Legg-Calvé-Perthes disease.

In the modified Waldenström classification, the stages are: IA, IB, IIA, IIB, IIIA and IIIB and IV. The stage IA is the sclerotic phase with no loss of epiphyseal height, while IB presents both sclerosis and epiphyseal height loss. In stage IIA, the epiphysis begins to fragment, and one or two vertical fissures can be seen. In stage IIB, this fragmentation is at advanced stage, but without lateral bone formation at the fragmented epiphysis. In stage IIIA, there is evidence of new bone formation at the peripheral necrotic region, but still without a normal texture and it covers less than a third of the epiphyseal region. In stage IIIB, the forming bone texture is normal and covers over a third of the epiphyseal region. In stage IV, there is no more avascular bone.[13]

The new classification system includes the disease evolution and helps to identify the onset of epiphyseal, metaphyseal and acetabular events related to the prognosis worsening. The authors argue that the understanding of these aspects may help determine how any treatment could alter the illness natural course.[13,15]

Therefore, the purpose of our study is to evaluate intraobserver and interobserver reliability of the modified Waldenström classification system for Legg-Calvé-Perth disease. Secondly, it aims at assessing if different results are obtained when the evaluation is performed by professionals of two different areas of expertise (Pediatric Orthopedics and Radiology) with diverse work experience (assistants, interns, and resident physician) from our institution.

## MATERIALS AND METHODS

Institutional Review Board approval from our institution were obtained (number of the decision: 1.967.830).

In total, 12 evaluators participated in our study, four pediatric orthopedists with over five years of work experience, four skeletal muscle radiologists and four orthopedics and trauma residents. The evaluators attended an introductory lecture of approximately 10 minutes on the aspects of the disease and its classification stages. Each evaluator received a table (Appendix 1) to be filled during the evaluation and an appendix containing the classification to be used. After the delivery of the material, 40 pairs of anteroposterior and frog-leg pelvic radiographs were presented to the evaluators, (Figure 1) distributed randomly according to the different disease stages.



**Figure 1.** Slide presentation model of radiographic examinations in anteroposterior and frog-leg views.

After 2 weeks, a new evaluation was performed in the same way and following the same criteria of the first one. The same radiography pairs were randomly assigned; however, they followed a different order from the first evaluation.

The statistical analysis was performed by using Kappa index (Table 1).[18]

**Table 1.** Correlation between K values and concordance level.

| Kappa index | Concordance level |
|---|---|
| < 0.00 | No concordance |
| 0.00-0.20 | Minimal concordance |
| 0.21-0.40 | Reasonable concordance |
| 0.41-0.60 | Moderate concordance |
| 0.61-0.80 | Substantial concordance |
| 0.81-1.00 | Perfect concordance |

To check agreement level regarding the different areas of expertise and work experience, a new statistical analysis was performed using Kendall's W test.[19]

## RESULTS

The intraobserver evaluation obtained an average Kappa value of 0.394, with 95% confidence interval and with a reasonable agreement level. All pediatric orthopedists, one radiologist, and one resident physician presented a moderate agreement level. The other participants, that is, three radiologists and three resident physicians, presented Kappa values between 0.21 and 0.40, with a reasonable agreement level. Table 2 shows the results of Kappa indexes and confidence intervals regarding the twelve participants.

**Table 2.** Kappa values and the correlation with the concordance level in intraobserver evaluation of twelve evaluators.

| Evaluators | 95% CI | Kappa | Concordance Level |
|---|---|---|---|
| Pediatric orthopedics | | | |
| 1 | 0.274 to 0.607 | 0.441 | Moderate |
| 2 | 0.338 to 0.700 | 0.518 | Moderate |
| 3 | 0.260 to 0.615 | 0.437 | Moderate |
| 4 | 0.343 to 0.700 | 0.521 | Moderate |
| Radiologist | | | |
| 1 | 0.236 to 0.591 | 0.413 | Moderate |
| 2 | 0.197 to 0.555 | 0.376 | Reasonable |
| 3 | 0.133 to 0.475 | 0.304 | Reasonable |
| 4 | 0.195 to 0.565 | 0.380 | Reasonable |
| Residents | | | |
| 1 | 0.127 to 0.497 | 0.312 | Reasonable |
| 2 | 0.293 to 0.661 | 0.477 | Moderate |
| 3 | 0.098 to 0.458 | 0.278 | Reasonable |
| 4 | 0.102 to 0.444 | 0.273 | Reasonable |

The interobserver evaluation obtained a Kappa value of 0.243 among all twelve evaluators in the first evaluation (95% CI, 0.227-0.259 and $p < 0.0001$) and Kappa value = 0.245 among the twelve evaluators in the second evaluation (95% CI, 0.229-0.260 and $p < 0.0001$). The two evaluation levels of inter-evaluative concordance were considered reasonable.

A new statistical analysis was performed using Kendall's W test to compare the influence of the area of expertise and work experience on the interobserver evaluation. Figure 2 shows the results referring to each of the three groups studied.
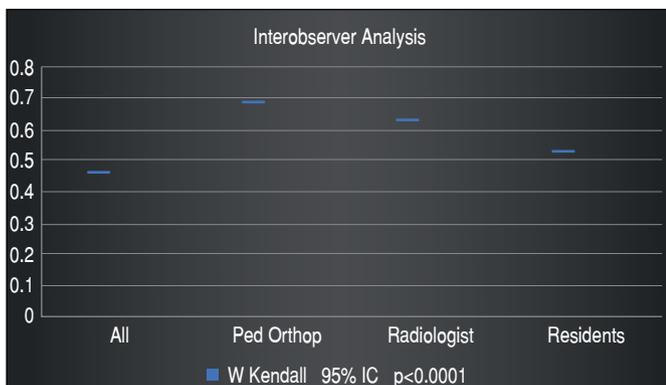


**Figure 2.** W Kendall values for interobserver agreement evaluation.

## DISCUSSION

Legg-Calvé-Perthes disease can be defined as a juvenile form of idiopathic osteonecrosis of the femoral head, since it affects children aged between 2 and 14. It is considered the most common form of femoral head osteonecrosis, with prevalence ranging from 5.1 to 16.9 per 100,000 children in several regions of the world.[1-3]

Since the first reports of this unique condition, made approximately 100 years ago by Legg, Calvé, and Perthes, several studies have been published on its etiology, epidemiology, natural history, radiographic classifications and treatment. Despite the increase in research related to Legg-Calvé-Perthes disease, it remains one of the most controversial conditions in pediatric orthopedics.

Current treatment for Legg-Calvé-Perthes disease is focused on obtaining and maintaining a good hip movement amplitude by restricting femoral head movement in the acetabulum. It is believed that, by following this treatment, the femoral head will be congruent with the acetabulum, reducing residual problems.

However, the difficulty of correlating the degree of femoral head lateral subluxation with the final residual deformity raises controversy on what would be the most effective treatment for patients with the disease. The available procedures range from non-surgical interventions, including physical therapy, orthoses and plaster cast to surgical interventions, such as: proximal femoral varus osteotomy, pelvic osteotomy,[14] and arthrodiastasis. Studies suggested that surgical containment may benefit a certain age group of Legg-Calvé-Perthes patients. These studies, however, did not address the relevant question on the best time to intervene with surgical treatment. Understanding the natural history of the disease and identifying the pathology stage affect both treatment decisions and outcome.

Monitoring the disease evolution by radiographic examinations is essential for therapeutic decision-making. Studies with other imaging methods, such as arthrography and MRI[5] were also proposed; however, the difficulty of conducting these serial examinations interferes with the evolutionary monitoring of the pathology.

Therefore, some radiographic classifications were proposed to determine the disease severity and relate it to the final prognosis. Three main classification systems were proposed by Catterall,[8] Salter and Thompson[7] and Herring.[9,10]

Despite the relevance of these classification systems to estimate the disease severity, it also requires a reliable classification system to determine the disease stage, since the intervention time may have a material effect on the treatment results.

The Waldenström classification is the only classification for staging Legg-Calvé-Perthes disease. Thorough its course, the disease produces characteristic radiographic changes in the femoral epiphysis that were subdivided into four temporal stages [6]: necrosis, fragmentation, reossification and residual stage.

Joseph et al.[13] modified Waldenström classification system and subdivided the first three stages into early (A) and late (B). Joseph et al.[13] proved that children surgically treated until the stage of early fragmentation IIA have a greater chance of presenting spherical and congruent femoral head and acetabulum in the residual stage when compared with those surgically treated in later stages of the disease. Therefore, they defended the use of modified Waldenström classification system to guide treatment decision-making. In this study, the modified classification showed solid interobserver and intraobserver reliability (Kappa = 0.72 and 0.71, respectively) [13]. Hyman et al.[15] also obtained a substantial and almost perfect interobserver and intraobserver agreement in several stages of their study. Twenty evaluators, all highly experienced pediatric orthopedists, evaluated 40 pairs of AP and frog-leg pelvic radiographs. Together with the images, the evaluators were provided data, such as patients' age, gender and time of symptom onset.

In both inter and intraobserver evaluation of our study, a slightly lower agreement level was obtained when compared with previous studies. The intraobserver evaluation obtained moderate and reasonable agreement levels, with Kappa value ranging from 0.273 to 0.521, whereas the interobserver analysis obtained a reasonable agreement level in the two evaluations, with Kappa values of 0.243, in the first evaluation, and 0.245, in the second one. Considering that our study involved the same number of radiographs and participants of other studies,[13,15] we attributed the agreement difference to two main factors.

The first factor is related to the evaluators' varied work experience and area of expertise in our study, which was greater than in Hyman et al.[15] and Joseph et al.[13] studies. In the latter two studies, only pediatric orthopedists with prior classification knowledge and highly experienced were assigned as evaluators. We divided the

evaluators in three groups to assess the influence of this factor on the results: experienced pediatric orthopedists, skeletal muscle radiologists, and orthopedics and trauma residents. The intraobserver analysis showed that in, the group of experienced pediatric orthopedists, four observers showed moderate concordance, with kappa ranging from 0.441 to 0.521, whereas the muscle skeletal radiologists' group presented Kappa ranging from 0.304 to 0.413 and only one observer showed moderate concordance, and the resident physicians' group, in which only one observer presented moderate concordance level, with Kappa values of 0.273 and 0.447 among observers. We used Kendall's W test in the interobserver evaluation to estimate the agreement difference among the three groups. The Kendall's W concordance coefficient indicates the degree of ordinal association assessments made by several evaluators when assessing the same samples. Kendall's W values can range from 0 to 1, and the highest Kendall value refers to the strongest association. In our study, Kendall's W coefficient results were higher for experienced pediatric orthopedists (W = 0.686) when compared with the skeletal muscle radiologists (W = 0.630) and the orthopedics and trauma residents (W = 0.529). Therefore, we can say the evaluators' degree of experience and area of expertise influences the agreement level found for the modified Waldenström classification.

The second factor is related to the fact that, unlike the other studies, our study does not provide extra data, such as age, gender and time of symptom onset. We believe that only acetabular and femoral changes presented in AP and frog-leg pelvic radiographs should be enough for the assertive classification of the disease. However, considering that this is an evolutionary classification, with progressive radiographic changes, and the Legg-Calvé-Perthes disease natural history already described by some studies, the time of the disease onset should be considered and may have influenced the increased level of intra and interobserver concordance. The effect of the presence of such data on the evaluation, however, has not been studied and its inclusion should be considered in future studies on classification description.

Although intraobserver and interobserver agreement levels of our study are lower than those presented in previous literature, we believe that our findings show that Waldenström classification is reproducible and it is the best current option for the evolutionary monitoring of Legg-Calvé-Perthes disease. Prior knowledge of the classification, experience in treating patients diagnosed with Legg-Calvé-Perth and additional data, such as the time of disease onset, gender and age, can increase the level of intraobserver and interobserver concordance.

Our study has one limitation: the absence of a clear gold standard for staging each of the cases, which could not be obtained due to the difficulty of determining the stage of each case, despite the large experience of the pediatric orthopedists' group. We decided not to rank the answers, as it was done in Hyman et al.,[15] as early stages (up to stage IIA) and late stages (after stage IIB), because we believe that it is important to distinguish all stages of the disease evolution.

## CONCLUSION

The modified Waldenström classification presented a moderate and reasonable level of intraobserver agreement. The interobserver agreement level was considered reasonable in the two evaluations. The evaluators' degree of experience and area of expertise influenced the agreement level among the different groups studied. The concordance level among pediatric orthopedists with over five years of experience was higher than the one between muscle skeletal radiologists and orthopedics and trauma residents.

Prior knowledge of the classification, experience in the disease treatment, and additional data, such as gender, age, and time of the disease onset, probably play an important role on the classification agreement level, adding the need for further studies to confirm this hypothesis.

## REFERENCES

1. Gray IM, Lowry RB, Renwick DH. Incidence and genetics of Legg-Perthes disease (osteochondritis deformans) in British Columbia: evidence of polygenic determination. J Med Genet. 1972;9(2):197-202.
2. Molloy MK, MacMahon B. Incidence of Legg-Perthes disease (osteochondritis deformans). N Engl J Med. 1966;275(18):988-90.
3. Margetts BM, Perry CA, Taylor JF, Dangerfield PH. The incidence and distribution of Legg-Calvé-Perthes' disease in Liverpool, 1982-95. Arch Dis Child. 2001;84(4):351-4.
4. Kim HKW, Herring JA. Pathophysiology, classifications, and natural history of Perthes disease. Orthop Clin North Am. 2011;42(3):285-95.
5. Du J, Lu A, Dempsey M, Herring JA, Kim HK. MR perfusion index as a quantitative method of evaluating epiphyseal perfusion in Legg-Calve-Perthes disease and correlation with short-term radiographic outcome: a preliminary study. J Pediatr Orthop. 2013;33(7):707-13.
6. Waldenstrom H. The definite form of the coxa plana. Acta Radiol. 2016;57(7):e79-94.
7. Salter RB, Thompson GH. Legg-Calvé-Perthes disease: the prognostic significance of the subchondral fracture and a two-group classification of the femoral head involvement. J Bone Joint Surg Am. 1984;66(4):479-89.
8. Catterall A. The natural history of Perthes' disease. J Bone Joint Surg Br. 1971;53(1):37-53.
9. Herring JA, Neustadt JB, Williams JJ, Early JS, Browne RH. The lateral pillar classification of Legg-Calvé-Perthes disease. J Pediatr Orthop. 1992;12(2):143-50.
10. Herring JA, Kim HT, Browne R. Legg-Calve-Perthes disease. Part II: prospective multicenter study of the effect of treatment on outcome. J Bone Joint Surg Am. 2004;86(10):2121-34.
11. Ritterbusch JF, Shantharam SS, Gelinas C. Comparison of lateral pillar classification and Catterall classification of Legg-Calvé-Perthes' disease. J Pediatr Orthop. 1993;13(2):200-2.
12. Wiig O, Terjesen T, Svenningsen S. Inter-observer reliability of radiographic classifications and measurements in the assessment of Perthes' disease. Acta Orthop Scand. 2002;73(5):523-30.
13. Joseph B, Varghese G, Mulpuri K, Narasimha Rao KL, Nair NS. Natural evolution of Perthes disease: a study of 610 children under 12 years of age at disease onset. J Pediatr Orthop. 2003;23(5):590-600.
14. Canale ST, D'Anca AF, Cotler JM, Snedden HE. Innominate osteotomy in Legg-Calvé-Perthes disease. J Bone Joint Surg Am. 1972;54(1):25-40.
15. Hyman JE, Trupia EP, Wright ML, Matsumoto H, Jo CH, Mulpuri K, et al. Interobserver and intraobserver reliability of the modified Waldenström classification system for staging of Legg-Calvé-Perthes disease. J Bone Joint Surg Am. 2015;97(8):643-50.
16. Hardcastle PH, Ross R, Hamalainen M, Mata A. Catterall grouping of Perthes' disease. An assessment of observer error and prognosis using the Catterall classification. J Bone Joint Surg Br. 1980;62-B(4):428-31.
17. Mahadeva D, Chong M, Langton DJ, Turner AM. Reliability and reproducibility of classification systems for Legg-Calvé-Perthes disease: a systematic review of the literature. Acta Orthop Belg. 2010;76(1):48-57.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
19. Kendall MG. A new measure of rank correlation. Biometrika. 1938;30(1-2):81-93.

Modified Waldenström classification – Date: ___/___/___
Did you receive any orientation about the classification? ___Y ___N

1. What is your degree of experience?
a) Orthopedic Assistant ( ) How long? ____
b) Orthopedic Resident
c) Radiologist
2. Classify the radiographs below:

| | |
|---|---|
| 1 Ia Ib IIa IIb IIIa IIIb IV | 21 Ia Ib IIa IIb IIIa IIIb IV |
| 2 Ia Ib IIa IIb IIIa IIIb IV | 22 Ia Ib IIa IIb IIIa IIIb IV |
| 3 Ia Ib IIa IIb IIIa IIIb IV | 23 Ia Ib IIa IIb IIIa IIIb IV |
| 4 Ia Ib IIa IIb IIIa IIIb IV | 24 Ia Ib IIa IIb IIIa IIIb IV |
| 5 Ia Ib IIa IIb IIIa IIIb IV | 25 Ia Ib IIa IIb IIIa IIIb IV |
| 6 Ia Ib IIa IIb IIIa IIIb IV | 26 Ia I1b IIa IIb IIIa IIIb IV |
| 7 Ia Ib IIa IIb IIIa IIIb IV | 27 Ia Ib IIa IIb IIIa IIIb IV |
| 8 Ia Ib IIa IIb IIIa IIIb IV | 28 Ia Ib IIa IIb IIIa IIIb IV |
| 9 Ia Ib IIa IIb IIIa IIIb IV | 29 Ia Ib IIa IIb IIIa IIIb IV |
| 10 Ia Ib IIa IIb IIIa IIIb IV | 30 Ia Ib IIa IIb IIIa IIIb IV |
| 11 Ia Ib IIa IIb IIIa IIIb IV | 31 Ia Ib IIa IIb IIIa IIIb IV |
| 12 Ia Ib IIa IIb IIIa IIIb IV | 32 Ia Ib IIa IIb IIIa IIIb IV |
| 13 Ia Ib IIa IIb IIIa IIIb IV | 33 Ia Ib IIa IIb IIIa IIIb IV |
| 14 Ia Ib IIa IIb IIIa IIIb IV | 34 Ia Ib IIa IIb IIIa IIIb IV |
| 15 Ia Ib IIa IIb IIIa IIIb IV | 35 Ia Ib IIa IIb IIIa IIIb IV |
| 16 Ia Ib IIa IIb IIIa IIIb IV | 36 Ia Ib IIa IIb IIIa IIIb IV |
| 17 Ia Ib IIa IIb IIIa IIIb IV | 37 Ia Ib IIa IIb IIIa IIIb IV |
| 18 Ia Ib IIa IIb IIIa IIIb IV | 38 Ia Ib IIa IIb IIIa IIIb IV |
| 19 Ia Ib IIa IIb IIIa IIIb IV | 39 Ia Ib IIa IIb IIIa IIIb IV |
| 20 Ia Ib IIa IIb IIIa IIIb IV | 40 Ia Ib IIa IIb IIIa IIIb IV |