

P-276 Tuesday, October 9, 2018 6:30 AM

## EMBRYOLOGISTS TEAM VS. AUTOMATED ANNOTATION SOFTWARE OUTCOMES.

L. Alegre,<sup>a</sup> R. Del Gallego, <sup>a</sup> T. Peura,<sup>b</sup> L. Bori,<sup>a</sup> B. Aparicio Ruiz,<sup>a</sup> A. Adam,<sup>c</sup> A. Coello,<sup>a</sup> D. Castello,<sup>a</sup> M. Meseguer.<sup>a</sup> aIVIRMA Valencia, Valencia, Spain; bGenea Biomedx, Sydney, Australia; cINCLIVA, Valencia, Spain. FERTILITY & STERILITY\_ e217

**OBJECTIVE:** To compare outcomes of embryo morphokinetic event annotations performed in a routine clinical practice to those performed by an automated embryo assessment software.

**DESIGN:** Key development events of embryos cultured in Geri time-lapse incubator were annotated manually by embryologists at IVIRMA clinics. Embryo videos were separately analysed with an automated annotations software, and event detection and annotation timing accuracies between the two compared.

**MATERIALS AND METHODS:** A busy embryologist team annotated nine developmental events of 311 embryos as per normal clinical practice using Geri Assess (GA) 1.3 software. MP4 videos were uploaded to a server and analysed with a Beta version of GA 2.0 automated annotations software (bGA 2.0), and the outcome data analysed.

**RESULTS:** Detection: From 2,799 putative developmental events, IVI detected 89%, bGA 2.0 94%, and both concurrently 86%. Mismatch rate of bGA 2.0 annotation without IVI (IVI-no annotation), was 8%, and of IVI annotation without bGA 2.0 (bGA 2.0-no annotation) was 2%. Accuracy: Annotation timings varied between the methods, with largest differences in late M and EB events. Mean difference of PNd to 6-cell events varied between 12-46 (median 5-13), and of M and EB between 89-110 (median 39-63) frames (1 frame¼5 min). Non-annotated events accumulated largely on same embryos, with only 152 embryos (49%) having all events annotated by both methods. Lack of annotations reflects interpretation difficulties associated with poor embryo quality. Such embryos also had the highest differences in annotation timings. The expected accuracy may vary between clinics, but e.g. an arbitrarily assigned limit of \_10 frames, when applied to ALL embryos, resulted in 45% (PNa, )88% (PNd), 91% (2-cell), 74% (3-cell), 71% (4-cell), 67% (5-cell), 62% (6-cell), 16% (M) and 34% (EB) of events within the limits.

**CONCLUSIONS:** Despite near 90% detection rate accordance, a high discordancy was found in annotation timings. Possible reasons may be: 1) Different event definitions by IVI vs. bGA 2.0; 2) Variability between annotators; 3) Inadequate time for accurate manual annotations due to stresses of clinical practice; 4) Chaotic embryos with aberrant divisions making exact annotations difficult for both IVI and bGA 2.0. Supported by: Spanish Ministry of Economy and Competitiveness (PI14/ 00523) through Instituto de Salud Carlos III program. The Grant from Fertility Innovation from Merck Serono.