

External Clinical Validation of a Synovial Fluid Machine Learning Model* for Diagnosing Periprosthetic Joint Infection at Two Large Arthroplasty Centers

Paul Edwards ³, Krista Toler ², Erik Kowalski ², Jason Cholewa ², Yale Fillingham ¹, Carl Deirmengian ¹

¹ Rothman Orthopaedic Institute, Philadelphia, Pennsylvania

² Zimmer Biomet, Warsaw, Indiana

³ Bowen Hefley Orthopaedics, North Little Rock, Arkansas

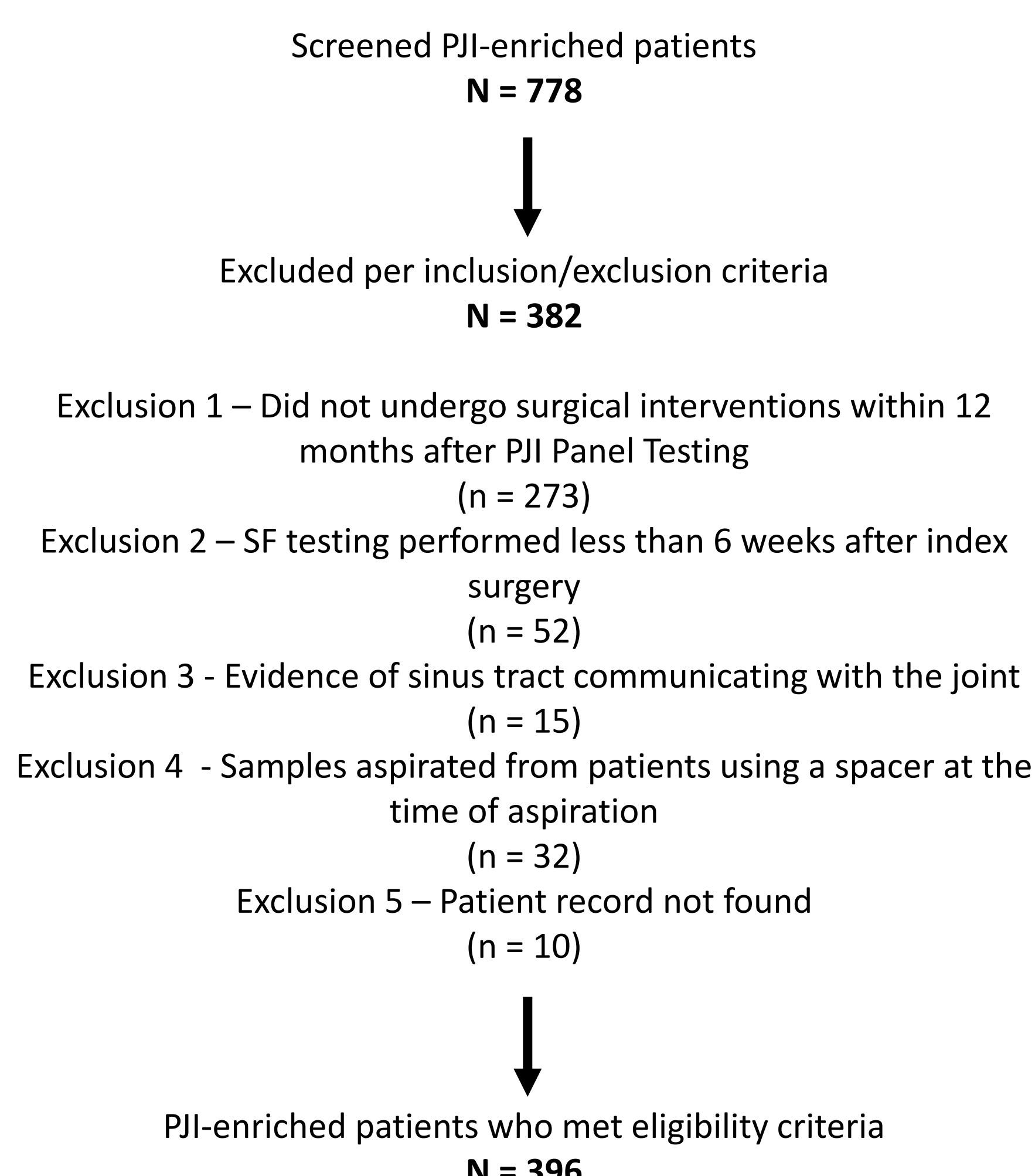
Aims

- Criteria-based definitions for periprosthetic joint infection (PJI) are complex and inconsistently applied.
- The purpose of this study was to externally validate a synovial fluid (SF)-only, laboratory-implemented machine learning (ML)-based PJI Probability Score in a multicenter clinical cohort.

Methods

- Patients from two high-volume arthroplasty centers that underwent comprehensive SF testing for hip or knee PJI at a centralized, specialized clinical laboratory (CD Laboratories, Towson, MD) between 2021-2024 were reviewed for eligibility.
- All SF-positive and SF-inconclusive patients were included, and Not Infected patients were randomly sampled at a 2:1 ratio to enrich for infected cases.
- A contract research organization performed chart review, capturing all available clinical, serologic, and intraoperative results ordered as standard of care.
- Of 778 screened PJI-enriched patients, 396 met eligibility criteria (**Figure 1**).
- The ML score was compared with the 2013 International Consensus Meeting (ICM), 2018 ICM, and 2021 European Bone and Joint Infection Society (EBJIS) definitions.
- For discordant or inconclusive cases, the treating-team clinical conclusion (operative-note diagnosis and definitive procedure) served as a pragmatic, not gold-standard, reference.

Figure 1: Patient screening and eligibility flowchart.



Results

- PJI classification rose from 34.3% (2013 ICM) to 40.7% (2018 ICM) to 49.0% (2021 EBJIS).

	2013 ICM	2018 ICM	2021 EBJIS	ML
Positive Samples	136	161	194	161
Negative Samples	260	192	199	204
Inconclusive	N/A	43	3	31
PJI Classification (Positive / 396)	34.3%	40.7%	49.0%	40.7%

- All four classifiers agreed on 324 of the 396 patients (81.8%), with 135 infected (34.1%) and 189 not infected (47.7%).

Treating-Team Clinical Conclusion for Discordant Cases

- Of the 72 discordant or inconclusive (18.2%) cases, the ML model matched the expert treating-team decision-making in 66 (91.7%), EBJIS matched in 55 (76.4%), and 2018 ICM and 2013 ICM matched in 26 (36.1%).

	Treating-Team Clinical Conclusion	2013 ICM	2018 ICM	2021 EBJIS	ML
Positive Samples	45	1	26	59	44
Negative Samples	27	71	3	10	26
Inconclusive Samples	0	0	43	3	2
Agreement with Treating-Team	-	36.1%	36.1%	76.4%	91.7%

Inconclusive Samples in 2018 ICM

- The 2018 ICM was inconclusive for 43 patients (10.9%), whereas the ML model was indeterminate for only 2 patients (0.5%).
- In 43 2018-ICM-inconclusive cases, the ML model matched 41 (95.3%) treating-team conclusions (21 of 22 treated as septic, 20 of 21 treated as aseptic).

2021 EBJIS Potential False Positives

- 33 patients were positive by 2021 EBJIS but not by 2013 or 2018 ICM. Treating teams confirmed 22 as infected and 11 as not infected, and the ML model agreed on 31 (93.9%), identifying 10 potential EBJIS false positives.
- The potential false positives were primarily driven by a single isolated positive test, most often a borderline synovial WBC elevation, under the EBJIS rule that one standalone positive can establish infection.

Conclusions

- The SF-only ML model offered clinically meaningful advantages over rule-based definitions:
 - First, it gave a definitive result in 99.5% of cases vs 89.1% for the 2018 ICM.
 - Second, it correctly classified 10 EBJIS false positives driven by isolated standalone tests.
 - Third, the ML model matched the expert treating-team decision-making in 41 of 43 (95.3%) 2018-ICM-inconclusive cases.
 - Finally, the ML model matched expert treating-team decision-making in 91.7% of difficult cases, compared to 76.4% for the next-closest-agreeing definition (2021 EBJIS); this reflects concordance with expert practice, not proven accuracy.
- By delivering a standardized, reproducible result in nearly every case, including those where criteria-based definitions disagree or return no answer, the SF-only ML model reduces the complexity and potential errors of rule-based definitions and extends expert-level arthroplasty diagnostic reasoning to any clinician.

*Synovasure® Comprehensive PJI Panel with SynTuition®; This work has been made possible by a research grant from Zimmer Biomet