Table S1. Studies selected	Selection F vs M	Comparability	DD ascertainment	m-NOS (0-8)
Aggarwal 2009	3	0	2	5 medium-high
Almousa 2023	3	1	2	6 high
Aloush 2007	3	0	2	5 medium-high
Atagunduz 2010	4	1	2	7 high
Bakland 2011	4	1	2	7 high
Bandinelli 2016	4	0	2	6 high
Blasco Blasco 2017	2	1	2	5 medium-high
Bodur 2012	4	0	2	6 high
Dehodar 2016	1	0	2	3 low
Dincer 2008	4	1	2	7 high
Garrido-cumbrera 2021	4	0	2	6 high
Hajialilo 2014	4	0	2	6 high
Ibn Yacoub 2012	4	1	2	7 high
Jovani 2018	4	1	2	7 high
Landi 2016	4	0	2	6 high
Li 2019	4	1	2	7 high
Ma 2012	3	1	2	6 high
Marks 1983	1	1	2	4 medium
Neunshwander 2020	4	1	2	7 high
Ogdie 2019	1	0	2	3 low
Reed 2008	3	0	2	5 medium-high
Ringsdal 1989	1	1	2	4 medium
Roussou 2011	1	1	2	4 medium
Shahlaee A 2015	3	1	3	7 high
Slobodin 2011	4	1	1	6 high
Zink 2000	4	1	2	7 high
Selection (0-4): female (F) disease definition and representativeness, men (M) selection and disease definition; Sexes comparability (0-1); Ascertainment of diagnostic delay (DD) (records 0-1, same method between sexes 0-1, not response rate 0-1 =0-3). Abbreviation: m-NOS: modified newcastle-ottawa quality assessment scale.				

Table S1. m-NOS quality assessment of 26 studies eligible of systematic review

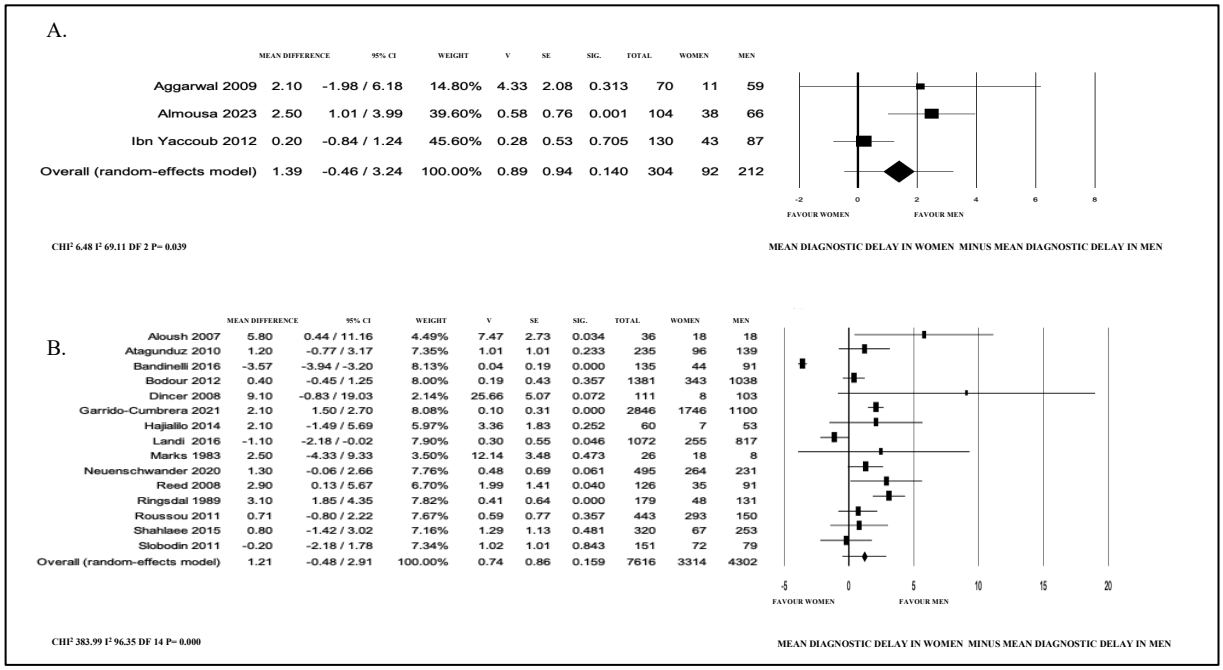

Figure S1: Forest plot of the differences between the mean diagnosis delay of Ax-SpA and AS in women versus men according to World Bank economic class in: A. papers from low and lower-middle income countries; B. papers from upper-middle- and high-income countries.

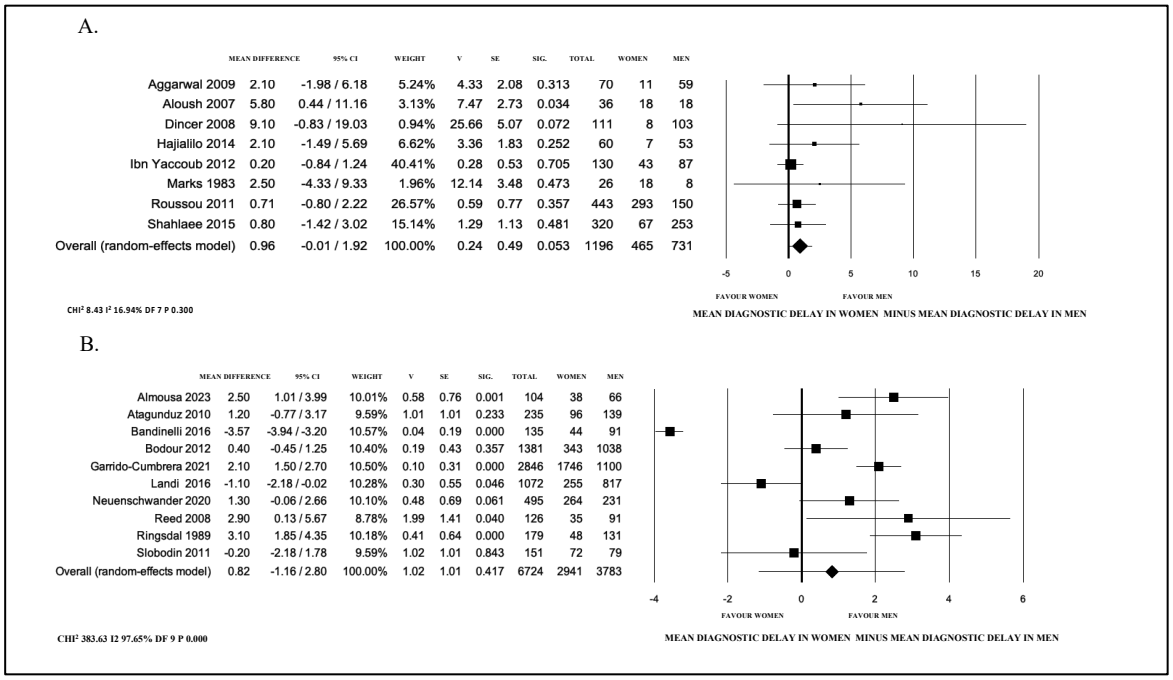

Figure S2: Forest plot of the differences between the mean diagnosis delay of of Ax-SpA and AS in women versus men according to samples of:
A. mono-centric B multi-centric papers.