

eTable 2. Methodological quality of selected studies

Reference	Study design	Sample size (patients with sepsis)	Domain definition	Determinant parameterization	Risk factor parameterization	Outcome parameterization	Overall quality score*
Arora 2007	Low-moderate (2.5) – Observational (-2) – Prospective (+0.5)	Very low (1) – n=18 (-3)	Moderate (3) – No distinction between sepsis, severe sepsis and septic shock (-0.5) – No explicit exclusion of patients with a prior history of AF (-0.5)	Moderate-high (3.5) – Duration of AF not reported (-0.5)	(n/a)	(n/a)	Moderate (2.5)
Christian 2008	Low (2) – Observational (-2)	Low-moderate (2.5) – n=274 (-1.5)	Moderate (3) – No description of sepsis diagnosis (-1)	Low-moderate (2.5) – Retrospective diagnosis of AF (-1) – Duration of AF not reported (-0.5)	Moderate (3) – Univariable analysis (-1)	Moderate (3) – Univariable analysis (-1)	Moderate (2.7)
Gomez 2012	Low-moderate (2.5) – Observational (-2) – Prospective (+0.5)	Low (2) – n= 100 (-2)	Very low (1) – No description of sepsis diagnosis (-1) – No exclusion of postcardiotomy patients (-1) – No exclusion of patients with a prior history of AF (-1)	Moderate-high (3.5) – Duration of AF not reported (-0.5)	(n/a)	(n/a)	Low (2.3)
Goodman 2007	Low-moderate (2.5) – Observational (-2) – Prospective (+0.5)	Low (2) – n=149 (-2)	Low-moderate (2.5) – No description of sepsis diagnosis (-1) – No distinction between SIRS and sepsis (-0.5)	High (4)	(n/a)	(n/a)	Moderate (2.8)
Meierhenrich 2010	Low-moderate (2.5) – Observational (-2) – Prospective (+0.5)	Very low-low (1.5) – n=50 (-2.5)	High (4)	Moderate-high (3.5) – Duration of AF not reported (-0.5)	Moderate (3) – Univariable analysis (-1)	Moderate (3) – Univariable analysis (-1)	Moderate (2.9)
Salman 2008	Low (2) – Observational (-2)	Low (2) – n=81 (-2)	Moderate (3) – No exclusion of patients with a prior history of AF (-1)	Low-moderate (2.5) – Retrospective diagnosis of AF (-1) – Duration of AF not reported (-0.5)	Moderate (3) – Univariable analysis (-1)	Moderate (3) – Univariable analysis (-1)	Moderate (2.6)
Seguin 2006	Low-moderate (2.5) – Observational (-2) – Prospective (+0.5)	Very low (1) – n=36 (-3)	Very low-low (1.5) – Only trauma patients (-1) – No exclusion of postcardiotomy patients (-1) – No explicit exclusion of patients with a prior history of AF (-0.5)	High (4)	(n/a)	(n/a)	Low (2.3)
Seguin 2004	Low-moderate (2.5) – Observational (-2) – Prospective (+0.5)	Low (2) – n= 107 (-2)	Low-moderate (2.5) – No exclusion of postcardiotomy patients (-1) – No explicit exclusion of patients with a prior history of AF (-0.5)	Moderate-high (3.5) – Duration of AF not reported (-0.5)	(n/a)	(n/a)	Moderate (2.6)
Walkey 2011	Low (2) – Observational (-2)	High (4) – n=49082 (-0)	Very low-low (1.5) – Use of ICD-9-codes to diagnose sepsis (-1) – No exclusion of postcardiotomy patients (-1) – No explicit exclusion of patients with a prior history of AF (-0.5)	Low-moderate (2.5) – Use of ICD-9-codes to diagnose AF (-1) – Duration of AF not reported (-0.5)	High (4)	High (4)	Moderate (3)
Walkey 2013	Low (2) – Observational (-2)	High (4) – n=60209 (-0)	Low (2) – Use of ICD-9-codes to diagnose sepsis (-1) – Only exclusion of postcardiotomy patients in sensitivity analysis (-1)	Low-moderate (2.5) – Use of ICD-9-codes to diagnose AF (-1) – Duration of AF not reported (-0.5)	High (4)	(n/a)	Moderate (2.9)
Wells 2011	Low (2) – Observational (-2)	Moderate (3) – n=465 (-1)	Very low (1) – Use of ICD-9-codes to diagnose sepsis (-1) – No exclusion of postcardiotomy patients (-1) – No exclusion of patients with a prior history of AF (-1)	Low-moderate (2.5) – Retrospective diagnosis of AF (-1) – Duration of AF not reported (-0.5)	Moderate (3) – Univariable analysis (-1)	Moderate (3) – Univariable analysis (-1)	Low (2.4)

AF: atrial fibrillation; ICD-9-CM: international classification of diseases-ninth revision-clinical modification; n/a: not available.

*Studies were evaluated for quality using the GRADE guidelines (Balslem, 2011). The following items were appraised: study design, sample size, domain definition and determinant parameterization. These items were scored on a scale from a score from our (high) to one point (low). We used a maximal score of 4 points, for every weak point of the study we subtracted points and for every strong point of the study we gave bonus points. With the sum of the scored items we calculated a mean overall quality score of the concerning study.