

Table S1: Risk of bias assessment for included articles

✓ = One point scored, ● = Half point scored, ✕ = Zero points scored

^a = full article, ^b = conference abstract

	Validity of study methods					Results interpretation		Applicability of results	
First Author	Population sampling	Location sampling	Sample size	Standard Measures	Unbiased assessor	Confidence intervals	Complete reporting	Study subjects	Quality Score
Aydemir ¹²	✕	✕	✕	✓	✓	✓	✓	✕	4 ^a
Beser ¹³	✓	✕	✓	✓	✓	✕	✕	✕	4 ^b
Chanis ¹⁵	✓	✕	✕	✓	✓	✕	✓	✓	5 ^b
Chew ¹⁶	✓	●	✓	✓	✓	✓	✓	✓	7.5 ^a
Chikunov ¹⁷	✓	✕	✓	✓	✕	✕	✓	✕	4 ^b
Ekinici ¹⁸	✕	✕	✕	✓	✓	✕	✓	✓	4 ^a
Ekinici ¹⁹	✕	✕	✓	✓	✕	✕	✓	✓	4 ^a
Jativa-Marino ²⁰	✓	✓	✓	●	✓	✕	✕	✓	5.5 ^a
Llanos-Chea ²¹	✓	●	✓	✓	✓	✕	✕	✕	4.5 ^b
Ozdemir ²²	✓	✕	✓	✓	✕	✕	✕	✕	3 ^b
Robin ²³	●	✓	✓	✓	✓	✕	✓	✓	6.5 ^a
Saps ²⁴	✓	✓	✓	✓	✓	✕	✓	✓	7 ^a
Saps ²⁵	✓	✕	✓	✓	✓	✕	✕	✓	5 ^a
Steutel ²⁶	✓	●	✓	✓	✓	✓	✓	✓	7.5 ^a
Velasco-Benitez ²⁷	✓	✕	✓	✓	✓	✕	✕	✓	5 ^b
Velasco-Benitez ²⁸	✓	✕	✓	✓	✓	✕	✕	✓	5 ^b
Velasco-Benitez ²⁹	✓	✓	✓	✓	✓	✕	✓	✓	7 ^a
Velasco-Benítez ³⁰	✓	✓	✓	✓	✓	✕	✓	✓	7 ^a
Vlad ³¹	✓	●	✓	●	✕	✕	✓	✕	4 ^b
Zeevenhooven ³²	✕	✕	✕	✓	✓	✓	✓	✓	5 ^b

Population sampling - Is the population sampling method appropriate for the research question?

Scores: Random or whole population sampling = 1 point, selecting respondents within these methods = 0.5 point

Location sampling - Is the sampling location appropriate for the general population?

Scores: Regional sampling = 1 point, Sampling from more than one general health clinic = 0.5 point

Sample size adequate - Is the sample size adequate? Determined by previous work among children showing FGID prevalence at approximately 30%^{8,9}, therefore, a minimal sample size of 150 could get adequate precision of FGID prevalence.

Scores: Sample size greater than 150 = 1 point

Standard measures - Are objective criteria used for measuring the health outcome, with appropriate validation?

Scores: English Rome IV criteria, or translation carried out with appropriate methodology = 1 point, translated Rome IV if methodology not stated or validation not carried out = 0.5 point.

Un-biased assessor - Is the health outcome measured in a non-biased fashion?

Scores: Study reported who completed Rome IV = 1 point

Confidence intervals - Are the results presented with confidence intervals for their estimates?

Scores: Confidence intervals reported = 1 point

Complete reporting - Is there complete reporting of all results?

Scores: Complete reporting of data (including 'zero' if no participants) = 1 point

Study subjects - Are study subjects sufficiently described?

Scores: Reporting age and gender distribution for study population = 1 point