Supplement 6. GRADE Summary of Findings & Evidence Profiles tables & forest plots

Evidence Set 1. Table 1.1 GRADE Summary of Findings – Benefits and harms of screening compared to no screening

Screening compared to no screening for asymptomatic bacteriuria in pregnant women

Patient or population: asymptomatic bacteriuria in pregnant women

Setting: Any primary or clinical care setting providing care to pregnant women

Intervention: screening

Comparison: no screening

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no screening	**Risk with screening**				
Maternal mortality	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on maternal mortality.
Maternal sepsis	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on maternal sepsis.
Pyelonephritis	**Median**		**RR 0.28** (0.15 to 0.54)	5659 (3 observational studies)	⨁◯◯◯ VERY LOW [1,a]	We are very uncertain about the effects of screening on pyelonephritis.
	18 per 1,000	**13 fewer per 1,000** (from 8 fewer to 16 fewer)				
Perinatal mortality	**Median**		**RR 1.21** (0.01 to 102.93)	724 (2 observational studies)	⨁◯◯◯ VERY LOW [1,b]	We are very uncertain about the effects of screening on perinatal mortality.
	19 per 1,000	**4 more per 1,000** (from 19 fewer to 1,000 more)				
Spontaneous abortion	55 per 1,000	**2 fewer per 1,000** (from 32 fewer to 70 more)	**RR 0.96** (0.41 to 2.27)	370 (1 observational study)	⨁◯◯◯ VERY LOW [1,c]	We are very uncertain about the effects of screening on spontaneous abortion.

Screening compared to no screening for asymptomatic bacteriuria in pregnant women

Patient or population: asymptomatic bacteriuria in pregnant women

Setting: Any primary or clinical care setting providing care to pregnant women

Intervention: screening

Comparison: no screening

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no screening	**Risk with screening**				
Neonatal sepsis	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on neonatal sepsis.
Preterm delivery	**Median**		**RR 8.70** (0.32 to 240.07)	722 (2 observational studies)	⨁◯◯◯ VERY LOW [1, d]	We are very uncertain about the effects of screening on preterm delivery.
	13 per 1,000	**102 more per 1,000** (from 9 fewer to 1,000 more)				
Low birthweight	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on low birthweight.
Maternal serious harm(s)	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on maternal serious harms.
Neonatal serious harm: fetal abnormalities	11 per 1,000	**5 more per 1,000** (from 8 fewer to 85 more)	**RR 1.50** (0.25 to 8.87)	372 (1 observational study)	⨁◯◯◯ VERY LOW [1, e]	We are very uncertain about the effects of screening on fetal abnormalities.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

Screening compared to no screening for asymptomatic bacteriuria in pregnant women

Patient or population: asymptomatic bacteriuria in pregnant women

Setting: Any primary or clinical care setting providing care to pregnant women

Intervention: screening

Comparison: no screening

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no screening	**Risk with screening**				

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] **The imprecision domain is assessed using GRADE guidance[42] relevant for systematic reviews as follows: when optimal information size (OIS) criterion is met, and the 95% confidence interval overlaps no effect, consideration of important benefit or important harm will be assessed using a relative risk of 1.0 (0.75 to 1.25).**

Pyelonephritis [a] → Very Low Quality Evidence: Three non-concurrent cohort studies (Gérard 1983, Gratacós 1994, Uncu 2001) reported this outcome (n=5,659). Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics. The optimal information size is met (sample size >5600), therefore downgrading for **imprecision** is not warranted. There were no serious concerns to warrant downgrading for **inconsistency**, **indirectness**, or **other considerations**.

Perinatal mortality [b] → Very Low Quality Evidence: Two non-concurrent cohort studies (n=724; Gérard 1983, Uncu 2001) reported this outcome. Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics, and suspected reporting bias as two studies did not report on perinatal mortality. Further downgrading is warranted for **imprecision** due to optimal information size not being met with a small sample size. There were no serious concerns to warrant downgrading for **inconsistency, indirectness** or **other considerations**.

Spontaneous abortion [c] → Very Low Quality Evidence: One non-concurrent cohort study reported this outcome (n=370; Gérard 1983). Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk

factors or other patient characteristics, and suspected reporting bias as two studies did not report on spontaneous abortion. Only one study provided data on spontaneous abortion, so this warrants downgrading for **inconsistency**. Further downgrading for **imprecision** is warranted due to low event rates (total of 20) without optimal information size. There were no serious concerns to warrant downgrading for **indirectness** or **other considerations**.

Preterm delivery [d] → Very Low Quality Evidence: Two non-concurrent cohort studies (n=722; Gérard 1983, Uncu 2001) reported this outcome. Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics, and suspected reporting bias as two studies did not report on preterm delivery. Further downgrading is warranted for **imprecision** for inadequate sample size and optimal information size not being met (total of 38 events). There were no serious concerns to warrant downgrading for **inconsistency, indirectness** or **other considerations**.

Neonatal serious harm: fetal abnormalities (harm) [e] → Very Low Quality Evidence: One non-concurrent cohort study reported this outcome (n=370; Uncu 2001). Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics, and suspected reporting bias as three studies did not report on fetal abnormalities. Only one study provided data on this outcome so this warrants downgrading for **inconsistency**. Further downgrading for **imprecision** is warranted due to the optimal information size not being met for rare events. There were no serious concerns to warrant downgrading for **indirectness** or **other considerations**.

Evidence Set 1. Table 1.2 GRADE Evidence Profile – Benefits and harms of screening compared to no screening

Question: Screening compared to no screening for asymptomatic bacteriuria in pregnant women

Setting: Any primary or clinical care setting providing care to pregnant women

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision** [1]	**Other considerations**	**screening**	**no screening**	**Relative (95% CI)**	**Absolute (95% CI)**		
Maternal mortality												
0									not estimable		-	CRITICAL
Maternal sepsis												
0									not estimable		-	CRITICAL
Pyelonephritis												
3	observational studies	serious	not serious	not serious	serious	none	10/2008 (0.5%)	1.8%	**RR 0.28** (0.15 to 0.54)	**13 fewer per 1,000** (from 8 fewer to 16 fewer)	⨁◯◯◯ VERY LOW [1, a]	CRITICAL
Perinatal mortality												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision** [1]	**Other considerations**	**screening**	**no screening**	**Relative (95% CI)**	**Absolute (95% CI)**		
2	observational studies	serious	not serious	not serious	serious	none	6/349 (1.7%)	1.9%	**RR 1.21** (0.01 to 102.93)	**4 more per 1,000** (from 19 fewer to 1,000 more)	⨁◯◯◯ VERY LOW [1, b]	CRITICAL
Spontaneous abortion												
1	observational studies	serious	serious	not serious	serious	none	9/170 (5.3%)	11/200 (5.5%)	**RR 0.96** (0.41 to 2.27)	**2 fewer per 1,000** (from 32 fewer to 70 more)	⨁◯◯◯ VERY LOW [1, c]	CRITICAL
Neonatal sepsis												
0									not estimable		-	CRITICAL
Preterm delivery												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision** [1]	**Other considerations**	**screening**	**no screening**	**Relative (95% CI)**	**Absolute (95% CI)**		
2	observational studies	serious	not serious	not serious	serious	none	33/347 (9.5%)	1.3%	**RR 8.70** (0.32 to 240.07)	**102 more per 1,000** (from 9 fewer to 1,000 more)	⨁◯◯◯ VERY LOW [1, d]	CRITICAL
Low birthweight												
0									not estimable		-	IMPORTANT
Maternal serious harm(s)												
0									not estimable		-	CRITICAL
Neonatal serious harm: fetal abnormalities												
1	observational studies	serious	serious	not serious	serious	none	3/186 (1.6%)	2/186 (1.1%)	**RR 1.50** (0.25 to 8.87)	**5 more per 1,000** (from 8 fewer to 85 more)	⨁◯◯◯ VERY LOW [1, e]	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

[1] The imprecision domain is assessed using GRADE guidance[42] relevant for systematic reviews as follows: when optimal information size (OIS) criterion is met, and the 95% confidence interval overlaps no effect, consideration of important benefit or important harm will be assessed using a relative risk of 1.0 (0.75 to 1.25).

Pyelonephritis [a] → Very Low Quality Evidence: Three non-concurrent cohort studies (Gérard 1983, Gratacós 1994, Uncu 2001) reported this outcome (n=5,659). Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics. The optimal information size is met (sample size >5600), therefore downgrading for **imprecision** is not warranted. There were no serious concerns to warrant downgrading for **inconsistency**, **indirectness**, or **other considerations**.

Perinatal mortality [b] → Very Low Quality Evidence: Two non-concurrent cohort studies (n=724; Gérard 1983, Uncu 2001) reported this outcome. Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics, and suspected reporting bias as two studies did not report on perinatal mortality. Further downgrading is warranted for **imprecision** due to optimal information size not being met with a small sample size. There were no serious concerns to warrant downgrading for **inconsistency, indirectness** or **other considerations**.

Spontaneous abortion [c] → Very Low Quality Evidence: One non-concurrent cohort study reported this outcome (n=370; Gérard 1983). Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics, and suspected reporting bias as two studies did not report on spontaneous abortion. Only one study provided data on spontaneous abortion, so this warrants downgrading for **inconsistency**. Further downgrading for **imprecision** is warranted due to low event rates (total of 20) without optimal information size. There were no serious concerns to warrant downgrading for **indirectness** or **other considerations**.

Preterm delivery [d] → Very Low Quality Evidence: Two non-concurrent cohort studies (n=722; Gérard 1983, Uncu 2001) reported this outcome. Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics, and suspected reporting bias as two studies did not report on preterm delivery. Further downgrading is warranted for **imprecision** for inadequate sample size and optimal information size not being met (total of 38 events). There were no serious concerns to warrant downgrading for **inconsistency, indirectness** or **other considerations**.

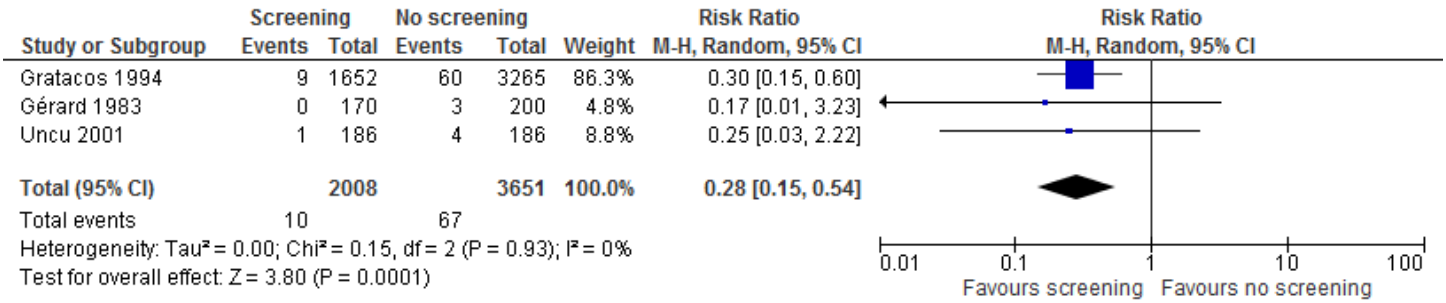
Neonatal serious harm: fetal abnormalities [e] → Very Low Quality Evidence: One non-concurrent cohort study reported this outcome (n=370; Uncu 2001). Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** across studies associated with: 1) no demonstration of comparability between screening and no screening groups, and 2) no adjustment to analyses for risk factors or other patient characteristics, and suspected reporting bias as three studies did not report on fetal abnormalities. Only one study provided data on this outcome so this warrants downgrading for **inconsistency**. Further downgrading for **imprecision** is warranted due to the optimal information size not being met for rare events. There were no serious concerns to warrant downgrading for **indirectness** or **other considerations**.

Evidence Set 1. Forest Plots 1.1-1.5 – Benefits and harms of screening compared to no screening

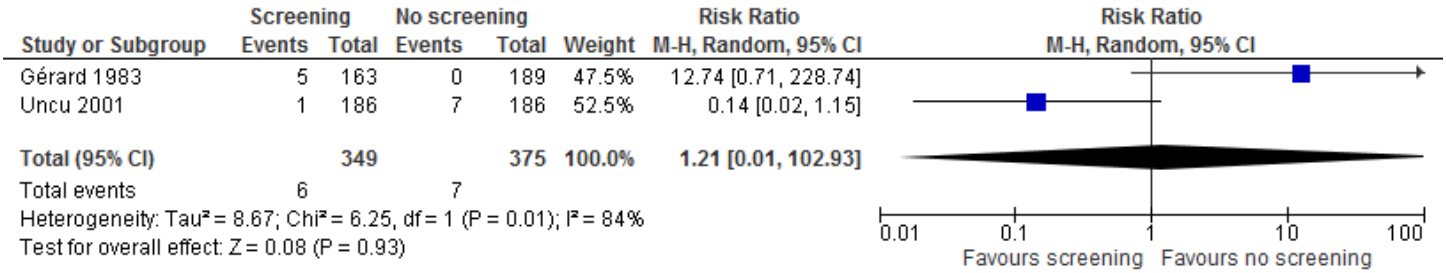
Outcome	No. of studies	No. of participants	Effect size (Risk Ratio; M-H, Random, 95%CI)
1.1 Pyelonephritis	3	5659	0.28 [0.15, 0.54]
1.2 Perinatal mortality >=20 wks GA note: Gérard >=31 wks; Uncu >20 wks	2	724	1.21 [0.01, 102.93]
1.3 Spontaneous abortion <20 wks GA note: 1 study <=28 wks (all occurred 7-21 wks)	1	370	0.96 [0.41, 2.27]
1.4 Preterm delivery <37 wks GA	2	722	8.70 [0.32, 240.07]
1.5 Neonatal serious harm: fetal abnormalities	1	372	1.50 [0.25, 8.87]

CI: confidence interval; GA: gestational age; M-H: Mantel-Haenszel; No.: number; wks: weeks

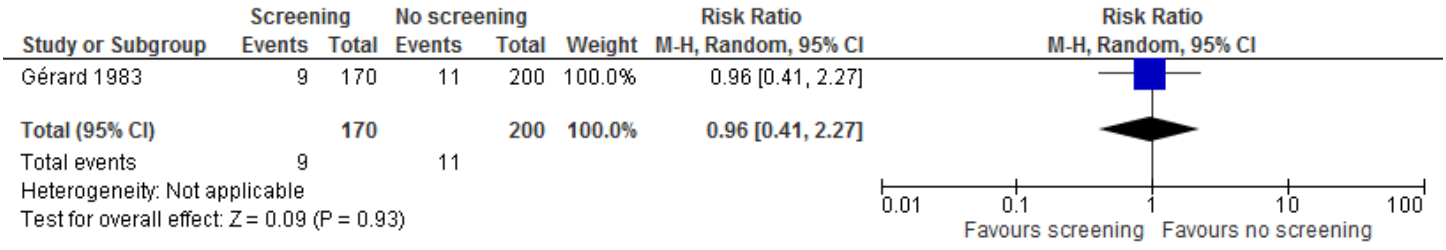
1.1 Pyelonephritis



1.2 Perinatal mortality (>=20 wks GA)



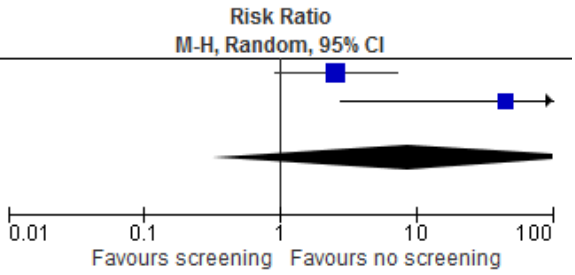
1.3 Spontaneous abortion (<20 wks GA)



1.4 Preterm delivery (<37 wks GA)

Study or Subgroup	Screening Events	Screening Total	No screening Events	No screening Total	Weight	Risk Ratio M-H, Random, 95% CI
Gérard 1983	11	161	5	189	57.5%	2.58 [0.92, 7.28]
Uncu 2001	22	186	0	186	42.5%	45.00 [2.75, 736.39]
Total (95% CI)		**347**		**375**	**100.0%**	**8.70 [0.32, 240.07]**
Total events	33		5			

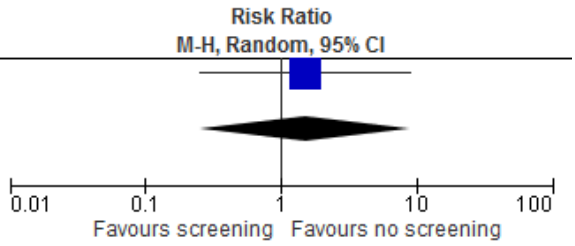
Heterogeneity: Tau² = 4.70; Chi² = 5.07, df = 1 (P = 0.02); I² = 80%
Test for overall effect: Z = 1.28 (P = 0.20)



1.5 Neonatal serious harm: fetal abnormalities

Study or Subgroup	Screening Events	Screening Total	No screening Events	No screening Total	Weight	Risk Ratio M-H, Random, 95% CI
Uncu 2001	3	186	2	186	100.0%	1.50 [0.25, 8.87]
Total (95% CI)		**186**		**186**	**100.0%**	**1.50 [0.25, 8.87]**
Total events	3		2			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.45 (P = 0.65)



Evidence Set 2. Table 2.1 GRADE Summary of Findings - Benefits and harms of frequent screening compared to one-time screening

Frequent screening compared to one-time screening for asymptomatic bacteriuria

Patient or population: asymptomatic bacteriuria

Setting: Any primary clinical care setting providing care to pregnant women

Intervention: frequent screening

Comparison: one-time screening

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with one-time screening	**Risk difference with frequent screening**				
Pyelonephritis	4 per 1,000	**0 fewer per 1,000** (from 3 fewer to 13 more)	**RR 1.09** (0.27 to 4.35)	1952 (1 observational study)	⨁◯◯◯ VERY LOW [1, a]	We are very uncertain about the effects of frequent screening compared to one-time screening on pyelonephritis.
Preterm delivery	49 per 1,000	**28 more per 1,000** (from 5 more to 60 more)	**RR 1.57** (1.11 to 2.23)	1952 (1 observational study)	⨁◯◯◯ VERY LOW [1, b]	We are very uncertain about the effects of frequent screening compared to one-time screening on preterm delivery.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] The imprecision domain is assessed using GRADE guidance[42] relevant for systematic reviews as follows: when

optimal information size (OIS) criterion is met, and the 95% confidence interval overlaps no effect, consideration of important benefit or important harm will be assessed using a relative risk of 1.0 (0.75 to 1.25).

CI: Confidence interval; **RR:** Risk ratio

Pyelonephritis [a] → Very Low Quality Evidence: One non-concurrent cohort study (n=1952; Rhode 2007) reported this outcome. Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to serious **risk of bias** associated with: 1) no demonstration that pyelonephritis was not present at start of study, 2) no demonstration of comparability between frequent and one-time screening groups, and 3) no adjustment to analyses to account for risk factors or other patient characteristics. Only one study provided data for this outcome so downgrading is warranted for **inconsistency**. Further downgrading is warranted for **indirectness** as the women are predominantly medically underserved, Hispanic and receiving care from a midwifery clinic, with a high rate of gestational diabetes (9%). The optimal information size is not met (8 events) with sample size (n=1952), therefore this warrants downgrading for **imprecision**. There were no serious concerns to warrant downgrading for **other considerations**.

Preterm delivery [b] → Very Low Quality Evidence: One non-concurrent cohort study (n=1952; Rhode 2007) reported this outcome. Quality of evidence is downgraded from high to low due to observational level data. Further downgrading from low to very low is warranted due to very serious **risk of bias** associated with: 1) no demonstration of comparability between frequent and one-time screening groups, 2) no adjustment to analyses to account for risk factors or other patient characteristics, and 3) suspected reporting bias among outcomes reported by studies (did not report on spontaneous abortion, perinatal mortality or fetal abnormalities). Only one study provided data for this outcome so downgrading is warranted for **inconsistency**. Further downgrading is warranted for **indirectness** as the women are predominantly medically underserved, Hispanic and receiving care from a midwifery clinic, with a high rate of gestational diabetes (9%). The event rate is low (122 events) without meeting optimal information size, so this is downgraded for **imprecision**. There were no serious concerns to warrant downgrading for **other considerations**.

Evidence Set 3. Table 3.1 GRADE Summary of Findings – Benefits and harms of treatment compared to no treatment

Treatment compared to no treatment for asymptomatic bacteriuria

Patient or population: asymptomatic bacteriuria

Setting: Any primary or clinical care setting providing care to pregnant women

Intervention: treatment

Comparison: no treatment

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no treatment	**Risk with treatment**				
Maternal mortality	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on maternal mortality.
Maternal sepsis	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on maternal sepsis.
Pyelonephritis	**Median**		**RR 0.24** (0.13 to 0.41)	2017 (12 RCTs)	⨁⨁◯◯ LOW [1, a]	There may be a reduction in pyelonephritis from treatment.
	232 per 1,000	**176 fewer per 1,000** (from 137 fewer to 202 fewer)				
Perinatal mortality	**Median**		**RR 0.96** (0.27 to 3.39)	1104 (6 RCTs)	⨁◯◯◯ VERY LOW [1, b]	We are very uncertain about the effects of treatment on perinatal mortality.
	40 per 1,000	**2 fewer per 1,000** (from 29 fewer to 97 more)				
Spontaneous abortion	**Median**		**RR 0.60** (0.11 to 3.10)	379 (2 RCTs)	⨁◯◯◯ VERY LOW [1, c]	We are very uncertain about the effects of treatment on spontaneous abortion.
	33 per 1,000	**13 fewer per 1,000** (from 30 fewer to 70 more)				

Treatment compared to no treatment for asymptomatic bacteriuria

Patient or population: asymptomatic bacteriuria

Setting: Any primary or clinical care setting providing care to pregnant women

Intervention: treatment

Comparison: no treatment

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no treatment	**Risk with treatment**				
Neonatal sepsis	**Median**		**RR 0.22** (0.01 to 4.54)	154 (2 RCTs)	⨁◯◯◯ VERY LOW [1, d]	We are very uncertain about the effects of treatment on neonatal sepsis.
	22 per 1,000	**17 fewer per 1,000** (from 22 fewer to 79 more)				
Preterm delivery	**Median**		**RR 0.57** (0.21 to 1.56)	533 (4 RCTs)	⨁◯◯◯ VERY LOW [1, e]	We are very uncertain about the effects of treatment on preterm delivery.
	158 per 1,000	**68 fewer per 1,000** (from 125 fewer to 88 more)				
Low birth weight	**Median**		**RR 0.63** (0.45 to 0.90)	1522 (7 RCTs)	⨁◯◯◯ LOW [1, f]	There may be a reduction in low birth weight from treatment.
	118 per 1,000	**44 fewer per 1,000** (from 12 fewer to 65 fewer)				
Maternal serious harm(s)	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(0 studies)	-	No study reported on maternal serious harms.
	Median					

Treatment compared to no treatment for asymptomatic bacteriuria

Patient or population: asymptomatic bacteriuria

Setting: Any primary or clinical care setting providing care to pregnant women

Intervention: treatment

Comparison: no treatment

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no treatment	**Risk with treatment**				
Neonatal serious harm: fetal abnormalities	19 per 1,000	**9 fewer per 1,000** (from 15 fewer to 8 more)	**RR 0.49** (0.17 to 1.43)	821 (4 RCTs)	⨁◯◯◯ VERY LOW [1, g]	We are very uncertain about the effects of treatment on harms (fetal abnormalities).
Neonatal serious harm: hemolytic anemia	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	265 (1 RCT)	⨁◯◯◯ VERY LOW [1, h]	We are very uncertain about the effects of treatment on harms (hemolytic anemia).

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] The imprecision domain is assessed using GRADE guidance[42] relevant for systematic reviews as follows: when optimal information size (OIS) criterion is met, and the 95% confidence interval overlaps no effect, consideration of important benefit or important harm will be assessed using a relative risk of 1.0 (0.75 to 1.25).

CI: Confidence interval; **RR:** Risk ratio

Pyelonephritis, overall [a] → Low Quality Evidence: Twelve trials (Brumfitt 1975, Elder 1971, Foley 1987, Furness 1975, Gold 1966, Kass 1960, Kazemier 2015, Kincaid-Smith 1965, Little 1966, Mulla 1960, Pathak 1969, Williams 1969) reported this outcome (n=2,017). Quality of evidence is downgraded from high to moderate due to serious **risk of bias**

associated with use of alternation for sequence generation (Elder 1971, Gold 1966, Kass 1960), inadequate allocation concealment (Elder 1971, Gold 1966, Kass 1960), and incomplete reporting (Brumfitt 1975, Furness 1975). This body of evidence on treatment effectiveness is downgraded from moderate to low for **indirectness** due to studies that did not explicitly include asymptomatic women (only 3 studies included exclusively asymptomatic women; Kazemier 2015, Mulla 1960, and Williams 1969), and studies that included high-risk women (Elder 1971, Kincaid-Smith 1965, Little 1966, and Pathak 1969). The optimal information size criterion is met (control group event rate=20%; total number of events=253) with an adequate sample size (n=2,017), and the confidence interval (0.13 to 0.41) indicates there may be important benefit; therefore, downgrading is not warranted for **imprecision**. There were no concerns with **inconsistency** or **other considerations** to warrant further downgrading.

Perinatal mortality [b] → Very Low Quality Evidence: Six trials (n=1,104; Elder 1971, Kass 1960, Kazemier 2015, Kincaid-Smith 1965, Little 1966, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate due to serious **risk of bias** associated with use of alternation for sequence generation (Elder 1971, Kass 1960, Wren 1969), and inadequate allocation concealment (Elder 1971, Kass 1960). This body of evidence on treatment effectiveness is downgraded for **indirectness** due to studies that did not explicitly include asymptomatic women as well as studies that included high-risk women. Further downgrading is warranted for **imprecision** due to the samples size not being met for optimal information size criterion (37 events). There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Spontaneous abortion [c]→ Very Low Quality Evidence: Two trials (n=379; Furness 1975, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate due to serious **risk of bias** associated with use of alternation for sequence generation (Wren 1969), inadequate allocation concealment (Wren 1969) and incomplete reporting (Furness 1975). Further downgrading from moderate to low is warranted for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women. The sample size is inadequate with optimal information size not met (10 events) to warrant downgrading twice from low to very low for **imprecision**. There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Neonatal sepsis [d] → Very Low Quality Evidence: Two trials (n=154; Kazemier 2015, Thomsen 1987) reported this outcome. Quality of evidence is downgraded for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women. The sample size (<2000) is not met with only 2 events to warrant downgrading twice for **imprecision**. There were no concerns to warrant downgrading for **risk of bias, inconsistency** or **other considerations**.

Preterm delivery [e] → Very Low Quality Evidence: Four trials (n=533; Furness 1975, Kazemier 2015, Thomsen 1987, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate for **risk of bias** associated with use of alternation for sequence generation (Wren 1969), inadequate allocation concealment (Wren 1969), and incomplete reporting (Furness 1975). There is substantial heterogeneity (I^2=70%) with point estimates on both sides of the line of no effect to warrant downgrading for **inconsistency**. Downgrading from moderate to low for **indirectness** is warranted due to studies that did not explicitly include exclusively asymptomatic women. There were no concerns to warrant downgrading for **imprecision or other considerations**.

Low birth weight [f] → Low Quality Evidence: Seven trials (n=1,522; Brumfitt 1975, Elder 1971, Kass 1960, Kazemier 2015, Kincaid-Smith 1965, Little 1966, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate for serious **risk of bias** associated with use of alternation for sequence generation (Elder 1971, Kass 1960, Wren 1969), inadequate allocation concealment (Elder 1971, Kass 1960, Wren 1969), and incomplete reporting (Brumfitt 1975). Further downgrading from moderate to low is warranted for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women as well as studies that included high-risk women. The optimal information size was not quite met (<2000 patients and <200 events), but we did not think the concerns were serious enough to downgrade for this outcome for **imprecision.** There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Neonatal serious harm: fetal abnormalities [g] → Very Low Quality Evidence: Four trials (n=821; Elder 1971, Furness 1975, Kazemier 2015, Little 1966) reported this outcome. Quality of evidence is downgraded from high to moderate for

serious **risk of bias** associated with use of alternation for sequence generation (Elder 1971), inadequate allocation concealment (Elder 1971), and incomplete reporting (Furness 1975). Downgrading from moderate to low is warranted for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women as well as studies that included high-risk women. Further downgrading from low to very low for **imprecision** is warranted due to optimal information size (sample size of 821) not being met for rare events. There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Neonatal serious harm: hemolytic anemia [h] → Very Low Quality Evidence: One trial (n=265; Elder 1971) reported this outcome. Quality of evidence is downgraded from high to moderate for **risk of bias** associated with use of alternation for sequence generation and inadequate allocation concealment. Only one study provided data for this outcome so downgrading from moderate to low for **inconsistency** is warranted. Further downgrading from low to very low is warranted for **indirectness** due the study not explicitly including exclusively asymptomatic women as well as studies that included high-risk women. Due to optimal information size (sample size of 265) not being met for rare events, downgrading twice is warranted for **imprecision**. There were no concerns to warrant downgrading for **other considerations**.

Evidence Set 3. Table 3.1 GRADE Evidence Profile – Benefits and harms of treatment compared to no treatment

Question: Treatment compared to no treatment for asymptomatic bacteriuria

Setting: Any primary or clinical care setting providing care to pregnant women

Bibliography:

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision** [1]	**Other considerations**	**treatment**	**no treatment**	**Relative (95% CI)**	**Absolute (95% CI)**		
Maternal mortality												
0									not estimable		-	CRITICAL
Maternal sepsis												
0									not estimable		-	CRITICAL
Pyelonephritis												
12	randomised trials	serious	not serious	serious	not serious	none	55/1023 (5.4%)	23.2%	**RR 0.24** (0.13 to 0.41)	**176 fewer per 1,000** (from 137 fewer to 202 fewer)	⨁⨁◯◯ LOW [1, a]	CRITICAL
Perinatal mortality												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision** [1]	**Other considerations**	**treatment**	**no treatment**	**Relative (95% CI)**	**Absolute (95% CI)**		
6	randomised trials	serious	not serious	serious	serious	none	16/529 (3.0%)	4.0%	**RR 0.96** (0.27 to 3.39)	**2 fewer per 1,000** (from 29 fewer to 97 more)	⨁◯◯◯ VERY LOW [1, b]	CRITICAL
Spontaneous abortion												
2	randomised trials	serious	not serious	serious	very serious	none	4/222 (1.8%)	3.3%	**RR 0.60** (0.11 to 3.10)	**13 fewer per 1,000** (from 30 fewer to 70 more)	⨁◯◯◯ VERY LOW [1, c]	CRITICAL
Neonatal sepsis												
2	randomised trials	not serious	not serious	serious	very serious	none	0/77 (0.0%)	2.2%	**RR 0.22** (0.01 to 4.54)	**17 fewer per 1,000** (from 22 fewer to 79 more)	⨁◯◯◯ VERY LOW [1, d]	CRITICAL
Preterm delivery												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision** [1]	**Other considerations**	**treatment**	**no treatment**	**Relative (95% CI)**	**Absolute (95% CI)**		
4	randomised trials	serious	serious	not serious	very serious	none	34/299 (11.4%)	15.8%	**RR 0.57** (0.21 to 1.56)	**68 fewer per 1,000** (from 125 fewer to 88 more)	⨁◯◯◯ VERY LOW [1, e]	CRITICAL
Low birth weight												
7	randomised trials	serious	not serious	serious	not serious	none	64/769 (8.3%)	11.8%	**RR 0.63** (0.45 to 0.90)	**44 fewer per 1,000** (from 12 fewer to 65 fewer)	⨁⨁◯◯ LOW [1, f]	IMPORTANT
Maternal serious harm(s)												
0									not estimable		-	CRITICAL
Neonatal serious harm: fetal abnormalities												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision** [1]	**Other considerations**	**treatment**	**no treatment**	**Relative (95% CI)**	**Absolute (95% CI)**		
4	randomised trials	serious	not serious	serious	very serious	none	4/425 (0.9%)	1.9%	**RR 0.49** (0.17 to 1.43)	**9 fewer per 1,000** (from 15 fewer to 8 more)	⨁◯◯◯ VERY LOW [1, g]	CRITICAL
Neonatal serious harm: hemolytic anemia												
1	randomised trials	serious	serious	serious	very serious	none	0/122 (0.0%)	0/143 (0.0%)	not estimable		⨁◯◯◯ VERY LOW [1, h]	CRITICAL

[1] The imprecision domain is assessed using GRADE guidance[42] relevant for systematic reviews as follows: when optimal information size (OIS) criterion is met, and the 95% confidence interval overlaps no effect, consideration of important benefit or important harm will be assessed using a relative risk of 1.0 (0.75 to 1.25).

CI: Confidence interval; **RR:** Risk ratio

Pyelonephritis, overall [a] → Low Quality Evidence: Twelve trials (Brumfitt 1975, Elder 1971, Foley 1987, Furness 1975, Gold 1966, Kass 1960, Kazemier 2015, Kincaid-Smith 1965, Little 1966, Mulla 1960, Pathak 1969, Williams 1969) reported this outcome (n=2,017). Quality of evidence is downgraded from high to moderate due to serious **risk of bias** associated with use of alternation for sequence generation (Elder 1971, Gold 1966, Kass 1960), inadequate allocation concealment (Elder 1971, Gold 1966, Kass 1960), and incomplete reporting (Brumfitt 1975, Furness 1975). This body of evidence on treatment effectiveness is downgraded from moderate to low for **indirectness** due to studies that did not explicitly include asymptomatic women (only 3 studies included exclusively asymptomatic women; Kazemier 2015, Mulla 1960, and Williams 1969), and studies that included high-risk women (Elder 1971, Kincaid-Smith 1965, Little 1966, and Pathak 1969). The optimal information size criterion is met (control group event rate=20%; total number of events=253) with an adequate sample size (n=2,017), and the confidence interval (0.13 to 0.41) indicates there may be important benefit; therefore, downgrading is not warranted for **imprecision**. There were no concerns with **inconsistency** or **other considerations** to warrant further downgrading.

Perinatal mortality [b] → Very Low Quality Evidence: Six trials (n=1,104; Elder 1971, Kass 1960, Kazemier 2015, Kincaid-Smith 1965, Little 1966, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate due to serious **risk of bias** associated with use of alternation for sequence

generation (Elder 1971, Kass 1960, Wren 1969), and inadequate allocation concealment (Elder 1971, Kass 1960). This body of evidence on treatment effectiveness is downgraded for **indirectness** due to studies that did not explicitly include asymptomatic women as well as studies that included high-risk women. Further downgrading is warranted for **imprecision** due to the samples size not being met for optimal information size criterion (37 events). There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Spontaneous abortion [c]→ Very Low Quality Evidence: Two trials (n=379; Furness 1975, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate due to serious **risk of bias** associated with use of alternation for sequence generation (Wren 1969), inadequate allocation concealment (Wren 1969) and incomplete reporting (Furness 1975). Further downgrading from moderate to low is warranted for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women. The sample size is inadequate with optimal information size not met (10 events) to warrant downgrading twice from low to very low for **imprecision**. There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Neonatal sepsis [d] → Very Low Quality Evidence: Two trials (n=154; Kazemier 2015, Thomsen 1987) reported this outcome. Quality of evidence is downgraded for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women. The sample size (<2000) is not met with only 2 events to warrant downgrading twice for **imprecision**. There were no concerns to warrant downgrading for **risk of bias, inconsistency** or **other considerations**.

Preterm delivery [e] → Very Low Quality Evidence: Four trials (n=533; Furness 1975, Kazemier 2015, Thomsen 1987, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate for **risk of bias** associated with use of alternation for sequence generation (Wren 1969), inadequate allocation concealment (Wren 1969), and incomplete reporting (Furness 1975). There is substantial heterogeneity (I^2=70%) with point estimates on both sides of the line of no effect to warrant downgrading for **inconsistency**. Downgrading from moderate to low for **indirectness** is warranted due to studies that did not explicitly include exclusively asymptomatic women. There were no concerns to warrant downgrading for **imprecision or other considerations**.

Low birth weight [f] → Low Quality Evidence: Seven trials (n=1,522; Brumfitt 1975, Elder 1971, Kass 1960, Kazemier 2015, Kincaid-Smith 1965, Little 1966, Wren 1969) reported this outcome. Quality of evidence is downgraded from high to moderate for serious **risk of bias** associated with use of alternation for sequence generation (Elder 1971, Kass 1960, Wren 1969), inadequate allocation concealment (Elder 1971, Kass 1960, Wren 1969), and incomplete reporting (Brumfitt 1975). Further downgrading from moderate to low is warranted for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women as well as studies that included high-risk women. The optimal information size was not quite met (<2000 patients and <200 events), but we did not think the concerns were serious enough to downgrade for this outcome for **imprecision.** There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Neonatal serious harm: fetal abnormalities [g] → Very Low Quality Evidence: Four trials (n=821; Elder 1971, Furness 1975, Kazemier 2015, Little 1966) reported this outcome. Quality of evidence is downgraded from high to moderate for serious **risk of bias** associated with use of alternation for sequence generation (Elder 1971), inadequate allocation concealment (Elder 1971), and incomplete reporting (Furness 1975). Downgrading from moderate to low is warranted for **indirectness** due to studies that did not explicitly include exclusively asymptomatic women as well as studies that included high-risk women. Further downgrading from low to very low for **imprecision** is warranted due to optimal information size (sample size of 821) not being met for rare events. There were no concerns to warrant downgrading for **inconsistency** or **other considerations**.

Neonatal serious harm: hemolytic anemia [h] → Very Low Quality Evidence: One trial (n=265; Elder 1971) reported this outcome. Quality of evidence is downgraded from high to moderate for **risk of bias** associated with use of alternation for sequence generation and inadequate allocation concealment. Only one study provided data for this outcome so downgrading from moderate to low for **inconsistency** is warranted. Further downgrading from low to very low is warranted for **indirectness** due the study not explicitly including exclusively asymptomatic women as well as studies that included high-risk women. Due to

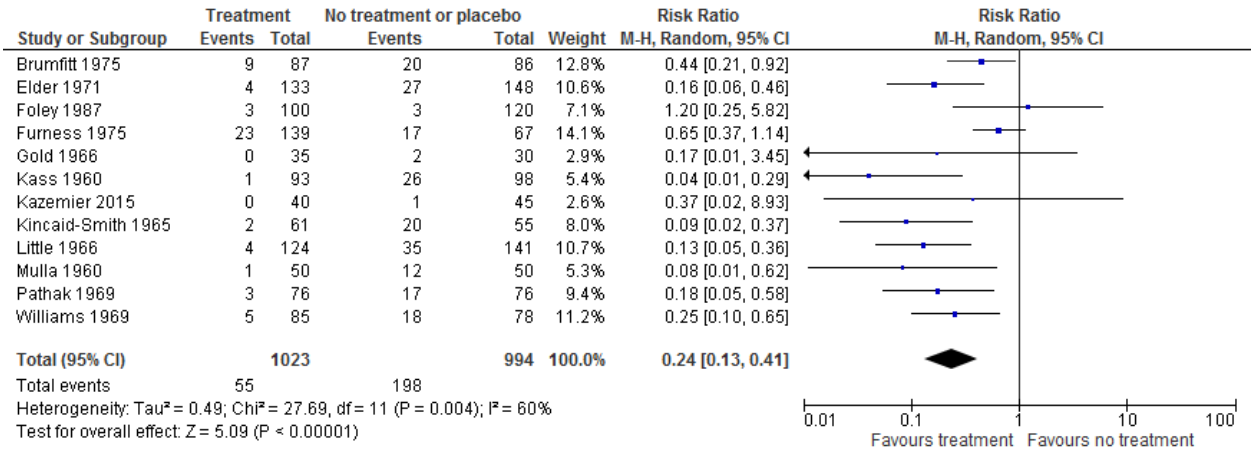
optimal information size (sample size of 265) not being met for rare events, downgrading twice is warranted for **imprecision**. There were no concerns to warrant downgrading for **other considerations**.

Evidence Set 3: Forest Plots 3.1-3.8 - KQ4: Benefits and harms of treatment compared to no treatment

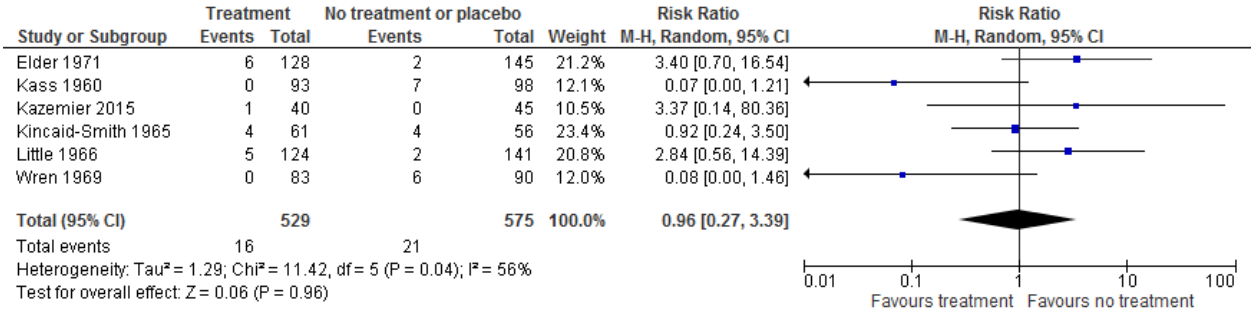
Outcome	No. of studies	No. of participants	Effect size (Risk Ratio; M-H, Random, 95%CI)
3.1 Pyelonephritis	12	2017	0.24 [0.13, 0.41]
3.2 Perinatal mortality (≥20 wks, including intrauterine demise, stillbirth, early neonatal death)	6	1104	0.96 [0.27, 3.39]
3.3 Spontaneous abortion (<20 wks)	2	379	0.60 [0.11, 3.10]
3.4 Neonatal sepsis	2	154	0.22 [0.01, 4.54]
3.5 Preterm delivery (<38 wks)	4	533	0.57 [0.21, 1.56]
3.6 Low birth weight (≤2500g; SGA <10th percentile & <5th percentile)	7	1522	0.63 [0.45, 0.90]
3.7 Neonatal serious harm: fetal abnormalities	4	821	0.49 [0.17, 1.43]
3.8 Neonatal serious harm: hemolytic anemia	1	265	Not estimable

CI: confidence interval; g: grams; M-H: Mantel-Haenszel; No.: number; SGA: small for gestational age; wks: weeks

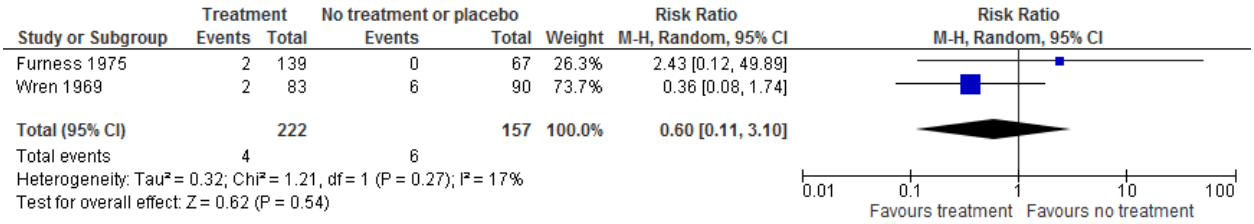
3.1 Pyelonephritis



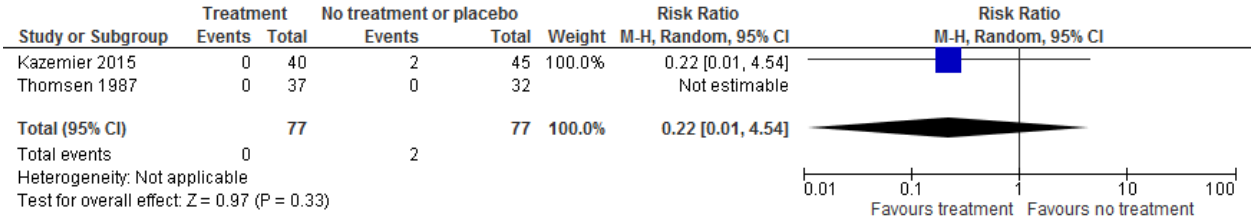
3.2 Perinatal mortality



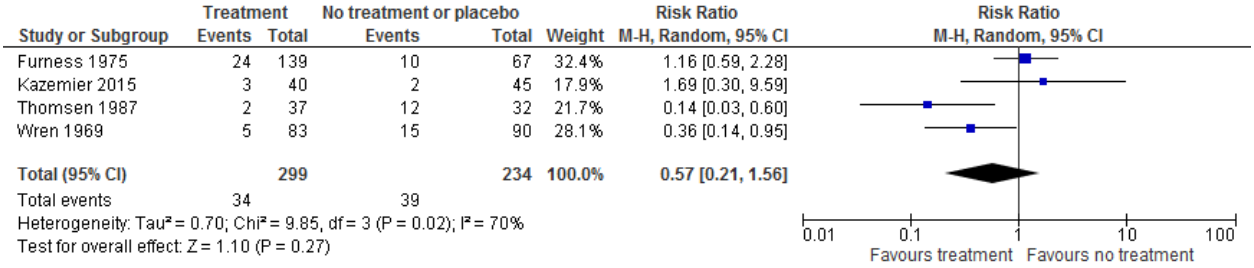
3.3 Spontaneous abortion



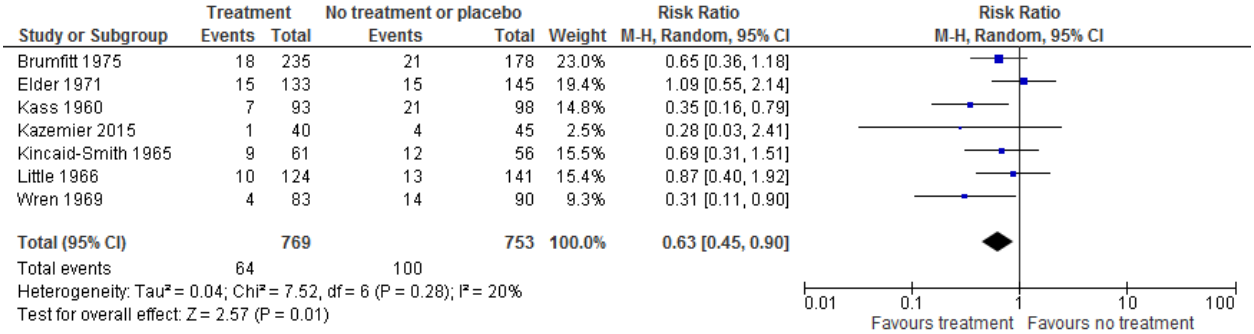
3.4 Neonatal sepsis



3.5 Preterm delivery



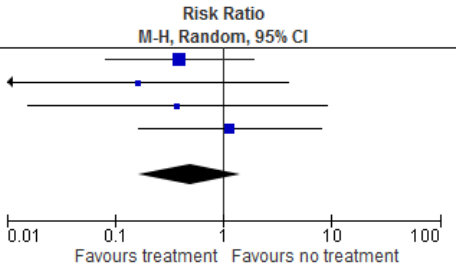
3.6 Low birthweight



3.7 Neonatal serious harm: fetal abnormalities

Study or Subgroup	Treatment Events	Total	No treatment or placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Elder 1971	2	122	6	143	46.4%	0.39 [0.08, 1.90]
Furness 1975	0	139	1	67	11.4%	0.16 [0.01, 3.92]
Kazemier 2015	0	40	1	45	11.5%	0.37 [0.02, 8.93]
Little 1966	2	124	2	141	30.7%	1.14 [0.16, 7.95]
Total (95% CI)		**425**		**396**	**100.0%**	**0.49 [0.17, 1.43]**
Total events	4		10			

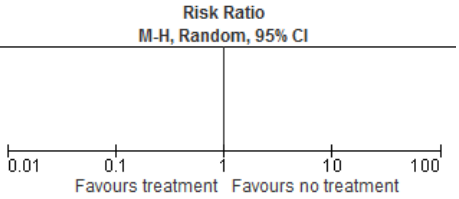
Heterogeneity: Tau² = 0.00; Chi² = 1.29, df = 3 (P = 0.73); I² = 0%
Test for overall effect: Z = 1.31 (P = 0.19)



3.8 Neonatal serious harm: hemolytic anemia

Study or Subgroup	Treatment Events	Total	No treatment or placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Elder 1971	0	122	0	143		Not estimable
Total (95% CI)		**122**		**143**		**Not estimable**
Total events	0		0			

Heterogeneity: Not applicable
Test for overall effect: Not applicable

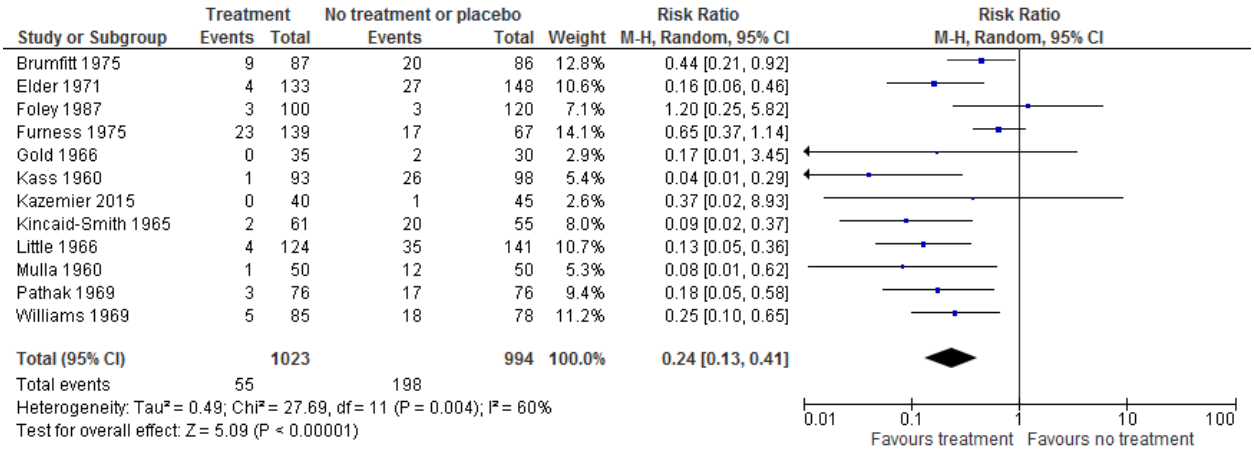


Evidence Set 3. Forest Plots for Subgroup Analyses 3.1.1-3.1.4 – KQ4: Benefits and harms of treatment compared to no treatment

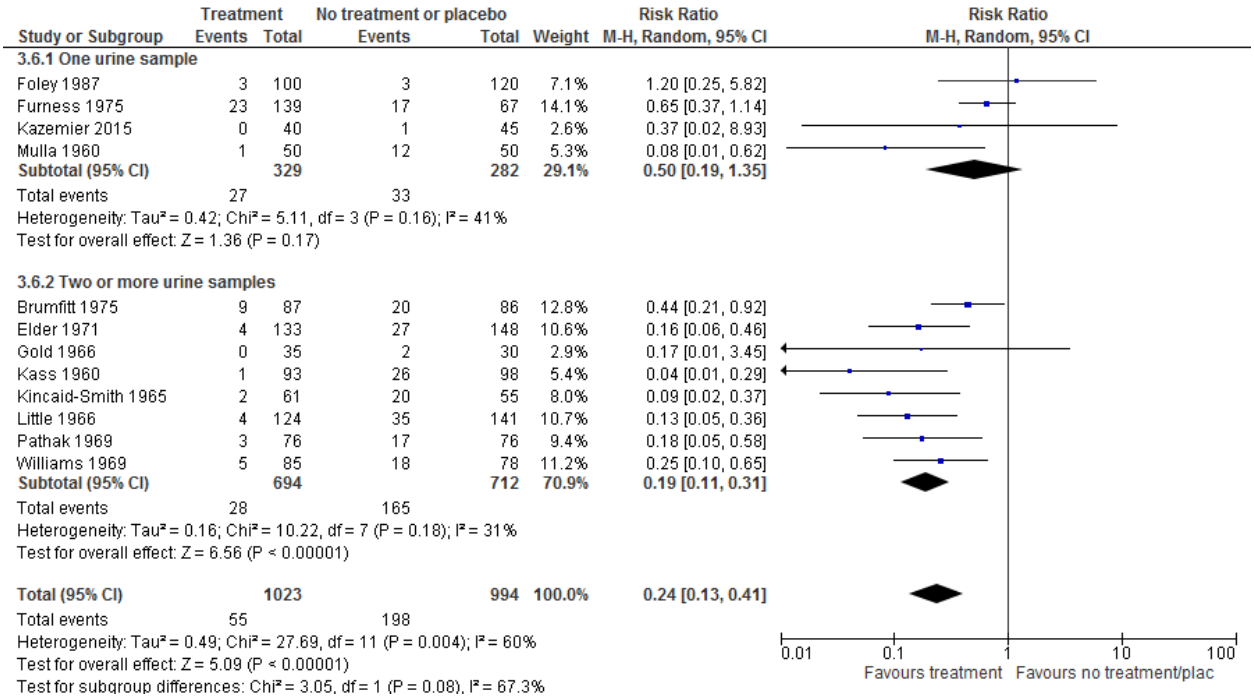
Outcome	No. of studies	No. of participants	Effect size (Risk Ratio; M-H, Random, 95%CI)
3.1 Pyelonephritis (overall)	**12**	**2017**	**0.24 [0.13, 0.41]**
3.1.1 Subgroup analysis: no. of urine samples before confirming bacteriuria and giving treatment			
One urine sample	4	611	0.50 [0.19, 1.35]
Two or more urine samples	8	1406	0.19 [0.11, 0.31]
3.1.2 Subgroup analysis: testing for persistent bacteriuria			
Tested for persistent bacteriuria during pregnancy	8	1352	0.26 [0.15, 0.45]
Testing for persistent bacteriuria post-delivery only	1	206	0.65 [0.37, 1.14]
Testing for persistent bacteriuria during pregnancy and post-delivery	3	459	0.11 [0.05, 0.25]
3.1.3 Subgroup analysis: follow-up			
Follow-up until delivery or puerperium (≤6 wks post-delivery)	9	1558	0.31 [0.18, 0.54]
Follow-up until >6 wks post-delivery	3	459	0.11 [0.05, 0.25]

CI: confidence interval; M-H: Mantel-Haenszel; No.: number; wks: weeks

3.1 Pyelonephritis (overall)

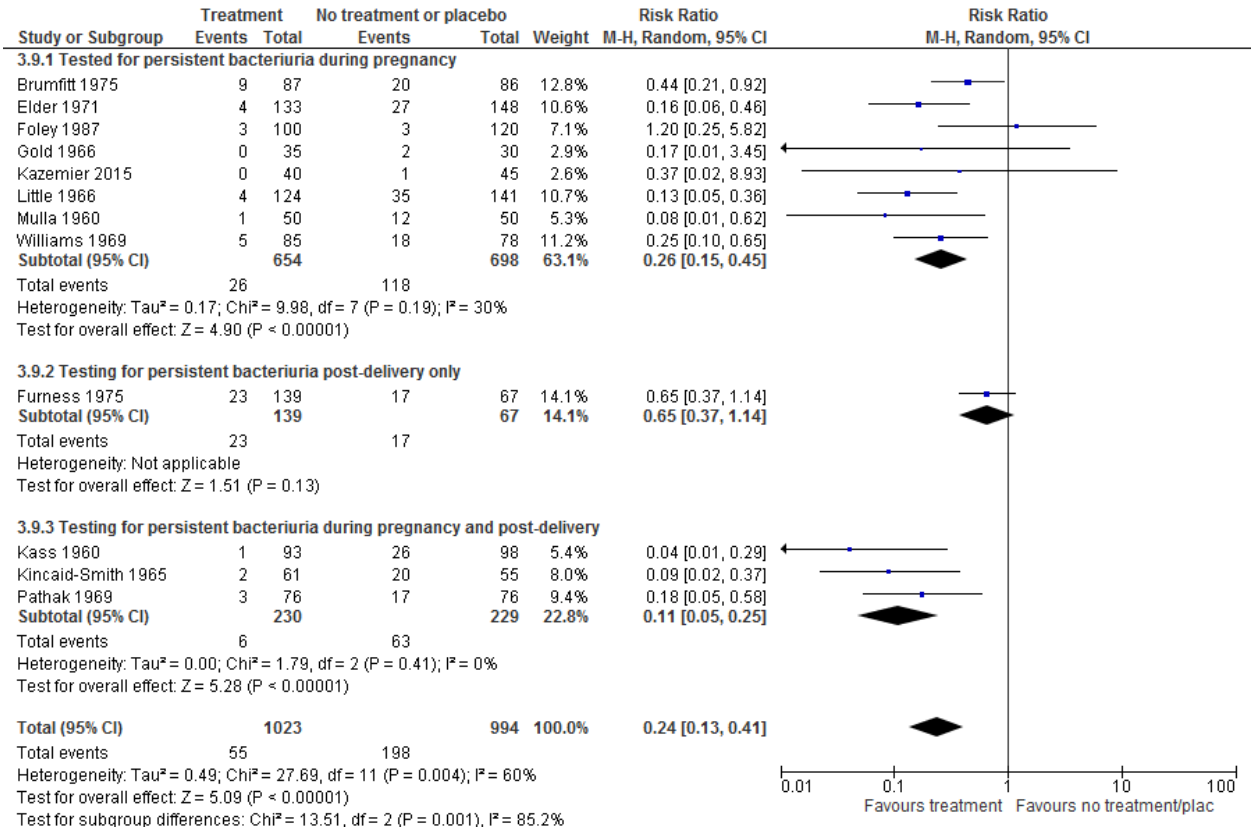


3.1.1 Pyelonephritis subgroup: number of urine samples at each screening visit*



*The additional culture(s) was used to confirm levels of bacteriuria.

3.1.2 Pyelonephritis subgroup: timing of testing for persistent bacteriuria



3.1.3 Pyelonephritis subgroup: duration of follow-up