エビデンス総体の確実性評価

公益財団法人日本医療機能評価機構

Minds

Copyright (9C) 2017 JCQHC

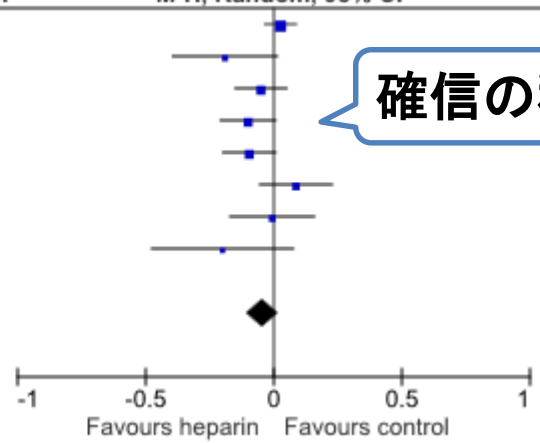
エビデンス（総体）の確実性とは

- 推奨作成における定義

効果推定値がある推奨を支持するのに適切であるという確信の程度

Study or Subgroup	Heparin Events	Heparin Total	Control Events	Control Total	Weight	Risk Difference M-H, Random, 95% CI
Agnelli 2009	333	769	155	381	22.5%	0.03 [-0.03, 0.09]
Altinbas 2004	22	42	30	42	6.6%	-0.19 [-0.39, 0.01]
Kakkar 2004	103	190	109	184	16.0%	-0.05 [-0.15, 0.05]
Klerk 2005	88	148	107	154	15.0%	-0.10 [-0.21, 0.01]
Lebeau 1994	96	138	110	139	15.7%	-0.10 [-0.20, 0.01]
Perry 2010	45	99	32	87	11.1%	0.09 [-0.05, 0.23]
Sideras 2006	40	68	41	69	9.0%	-0.01 [-0.17, 0.16]
Weber 2008	8	10	10	10	4.0%	-0.20 [-0.48, 0.08]
Total (95% CI)		**1464**		**1066**	**100.0%**	**-0.04 [-0.10, 0.02]**
Total events	735		594			

Heterogeneity: Tau² = 0.00; Chi² = 13.31, df = 7 (P = 0.06); I² = 47%
Test for overall effect: Z = 1.44 (P = 0.15)



確信の程度は？

GRADE groupでは，certainty of evidence, quality of evidence, quality of body of evidence, confidence in evidence, confidence in effect estimates, certainty in the body of evidence, strength of evidenceをほぼ同義で扱っている。本資料ではエビデンス（総体）の確実性と記載する。

Guyatt et al. J Clin Epidemiol. 2011 Apr;64(4):401-6. (No.3)
Guyatt et al. J Clin Epidemiol. 2013 Feb;66(2):151-7. (No.11)
GRADE handbook (Updated October 2013)
Schünemann HJ. Clin Epidemiol. 2016 Jul;75:6-15.

4段階のエビデンスの確実性

エビデンスの確実性	推奨を作成する際の効果推定値について
高 （A）	真の効果が効果推定値に近いことに大きな確信がある
中 （B）	効果推定値に対し中等度の確信がある。つまり、真の効果は効果推定値に近いと考えられるが、大きく異なる可能性も否めない。
低 （C）	効果推定値に対する確信性には限界がある。真の効果は効果推定値とは大きく異なるかもしれない。
非常に低 （D）	効果推定値に対し、ほとんど確信が持てない。真の効果は、効果推定値とは大きく異なるものと考えられる

http://gdt.guidelinedevelopment.org/app/

Guyatt et al. J Clin Epidemiol. 2011 Apr;64(4):401-6. (No.3)

エビデンスの確実性評価

1. 研究は適切にデザインされ、実施されたか?バイアスリスク
2. 結果は一貫しているか？非一貫性
3. 研究は設定した疑問に直接答えているか？非直接性
4. 結果は精確か？不精確さ
5. 全ての研究を含んでいるか？出版バイアス

- 観察研究では、用量反応勾配、効果の大きさ、交絡も評価する

（本講義では扱わないため、GRADE Handbook等を参照）

アウトカム毎のエビデンスの確実性を評価

(最初の) エビデンスの 確実性	グレードを下げる 5ドメイン	グレードを上げる 3ドメイン(観察研究)	(最終的な) エビデンスの確実性
RCT → 高 観察研究 → 低	・バイアスリスク -1 深刻 -2 非常に深刻 ・非一貫性 -1 深刻 -2 非常に深刻 ・非直接性 -1 深刻 -2 非常に深刻 ・不精確さ -1 深刻 -2 非常に深刻 ・出版バイアス -1 可能性あり -2 可能性高い	・効果が大きい ＋1 大きな効果 ＋2 非常に大きな効果 ・用量反応勾配 ＋1 あり ・全ての交絡を考慮した際 ＋1 効果を減弱させている ＋1 効果なしの結果であった場合に、効果を増加させる方向に働いている	**高（A）** ⊕⊕⊕⊕ **中（B）** ⊕⊕⊕ **低（C）** ⊕⊕ **非常に低(D)** ⊕

診療ガイドラインパネリストがエビデンスの確実性の評価の最終判断をする

Guyatt et al. J Clin Epidemiol. 2011 Apr;64(4):401-6. (No.3)
GRADE handbook (Updated October 2013)

アウトカムの重要性の分類

患者にとって重要なアウトカムについて

- critical（重大）
- important but not critical（重要だが重大でない）
- of limited importance（重要性は低い）.

Critical and important outcomes will bear on guideline recommendations
重大なアウトカムおよび重要だが重大でないアウトカムは推奨作成に用いられる。

rating scale:								
1	2	3	4	5	6	7	8	9
of **least** importance								of **most** importance
of limited importance for making a decision (not included in evidence profile)			important, but not critical for making a decision (included in evidence profile)			Critical for making a decision (included in evidence profile)		

意思決定に
重要性は低いアウトカム

意思決定に重要だが
重大でないアウトカム

意思決定に
重大なアウトカム

GRADE handbook 3.1 (Updated October 2013)

エビデンスの確実性評価の原則

1. 個々の研究の評価を単に平均するのではなく、質の高い研究に着目しながら、各研究の重みを考慮して評価する。
2. 各研究の重みは一般にサンプル数とイベント数で決定される。
3. グレードダウンは保守的であるべきである。
4. バイアスのリスクのグレードダウンは、ほかのドメインの限界を考慮して行う。複数のドメインの限界を合わせて1レベルグレードダウンすることもある。
5. 判断が難しい場合には、その理由や最終的な判定を行った理由を明示すべきである。

GRADE handbook 5.2.1 (Updated October 2013)

ヘパリンのシナリオ

CQ:抗凝固剤による予防や治療の適応のない癌患者に対してヘパリンを投与すべきか？

P：抗凝固剤による予防や治療の適応のない癌患者

I: 未分画、低分子量ヘパリンやフォンダパリヌクスなどの抗凝固剤の非経口投与

C：プラセボもしくは非介入

O：総死亡、症候性深部静脈血栓症、健康関連QoL、大出血、小出血など

Akl EA　et al. Cochrane Database Syst Rev. 2011

ヘパリンのシナリオの参考資料

- Cochrane Review

Cochrane Database Syst Rev. 2011 Apr 13;(4):CD006652.

Cochrane Database Syst Rev. 2014 Dec 10;(12):CD006652.　update版

- 2016年のGRADE Tokyo workshop資料

https://sites.google.com/site/gradeworkshopsba2016/tokyo2016

- 米国胸部専門医学会の診療ガイドライン

Chest 2012; 141（2）（Suppl）:e195S–e226S　Section 4.2

- サウジアラビアのCPG（McMaster大と連携）

Ann Saudi Med 2015; 35(2): 95-106.　Guideline Question 3

- 相原 守夫先生による解説資料（日本語）

http://www.grade-jpn.com/how_to_use_grade_aihara_20120426.pdf

エビデンスの確実性評価
①バイアスリスク (Risk of bias)

公益財団法人日本医療機能評価機構

Minds

エビデンスの確実性評価
①バイアスリスク

研究のデザインや実施における限界

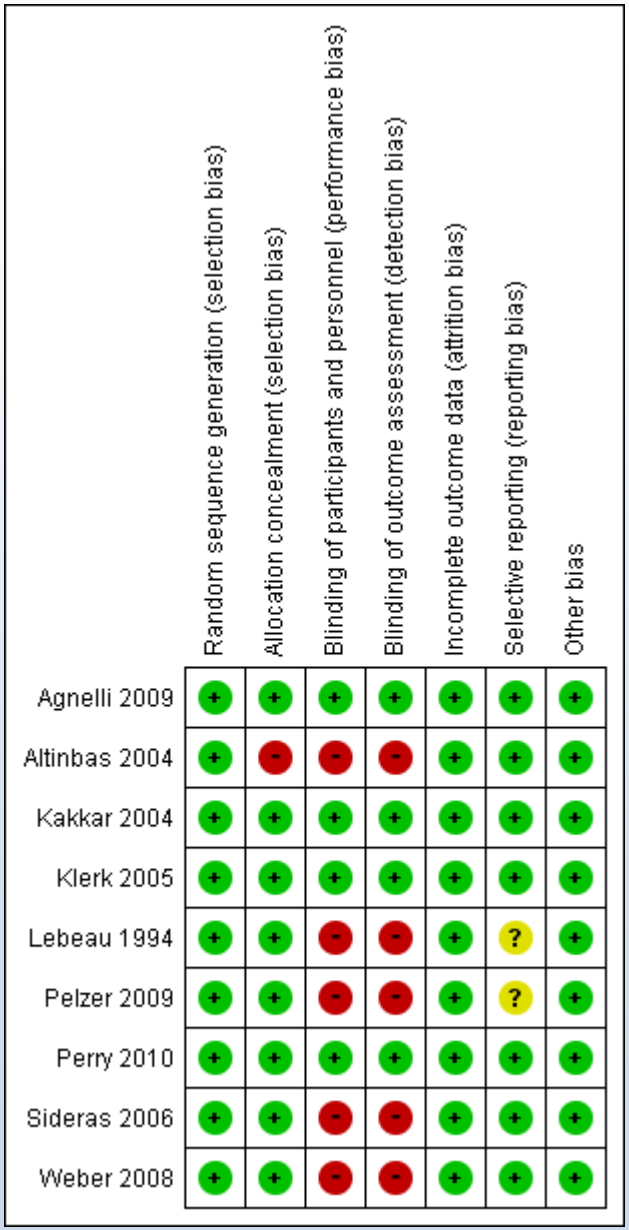
各研究の重みや結果、ドメインの重要性を考慮して判断する

Cochrane Risk of bias 評価

Risk of bias	Across studies	Interpretation	Considerations	GRADE assessment of study limitations	
Low risk of bias.	Most information is from studies at low risk of bias.	Plausible bias unlikely to seriously alter the results.	No apparent limitations.	No serious limitations, do not downgrade.	グレードダウンしない
Unclear risk of bias.	Most information is from studies at low or unclear risk of bias.	Plausible bias that raises some doubt about the results.	Potential limitations are unlikely to lower confidence in the estimate of effect.	No serious limitations, do not downgrade.	グレードダウンしない
			Potential limitations are likely to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.	グレードダウン(-1)
High risk of bias.	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results.	Plausible bias that seriously weakens confidence in the results.	Crucial limitation for one criterion, or some limitations for multiple criteria, sufficient to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.	グレードダウン(-1)
			Crucial limitation for one or more criteria sufficient to substantially lower confidence in the estimate of effect.	Very serious limitations, downgrade two levels.	グレードダウン(-2)

http://handbook.cochrane.org/chapter_12/

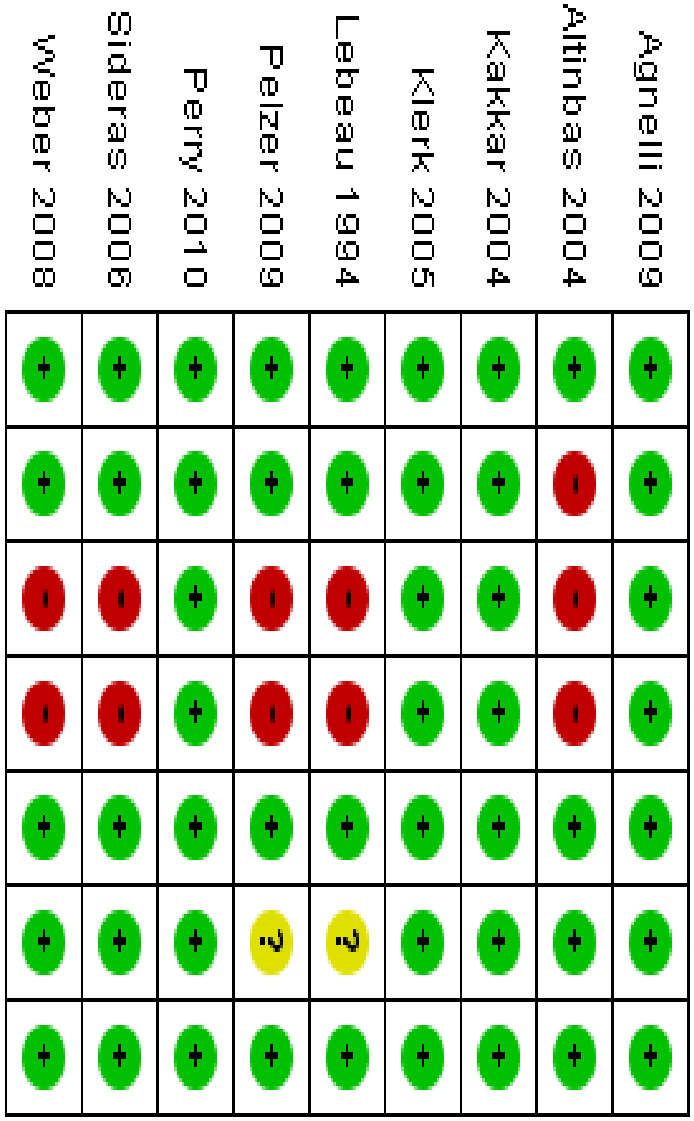
①バイアスリスク(ヘパリンのシナリオ)



アウトカム（12か月での死亡）のバイアスリスクは？
-深刻でない（グレードダウンなし）
-深刻　（1レベル　グレードダウン）
-非常に深刻（2レベル　グレードダウン）

Akl EA　et al. Cochrane Database Syst Rev. 2011

①バイアスリスク(ヘパリンのシナリオ)



Random sequence generation (selection bias)
ランダム系列の生成

Allocation concealment (selection bias)
割り付けの隠蔽化

Blinding of participants and personnel (performance bias)
参加者と研究スタッフのブラインド化

Blinding of outcome assessment (detection bias)
アウトカム評価者のブラインド化

Incomplete outcome data (attrition bias)
不完全アウトカムデータ

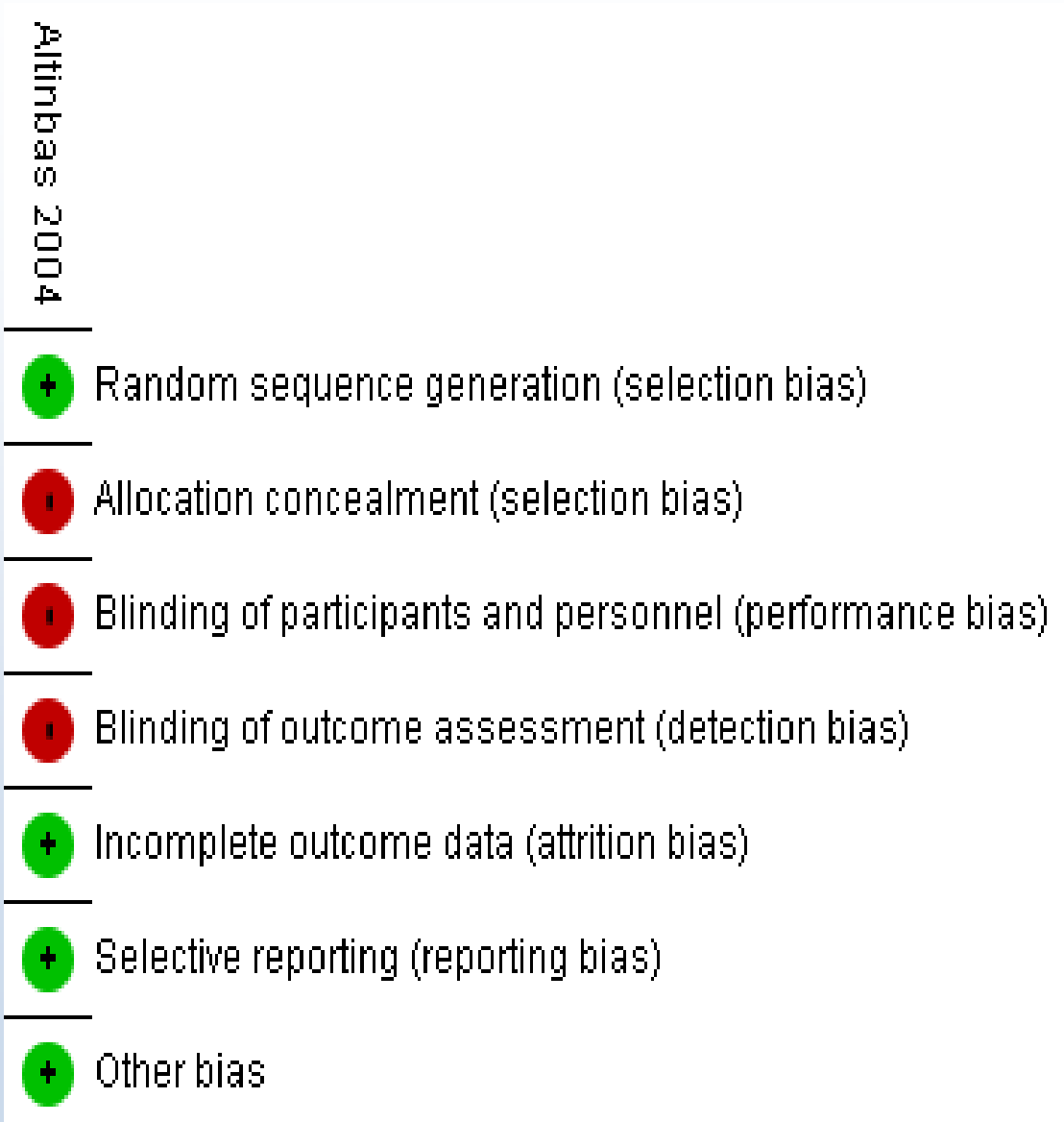
Selective reporting (reporting bias)
選択的アウトカム報告

Other bias
その他のバイアス

総死亡がアウトカムの場合、ブラインド化はさほど問題とならないが、割り付けの隠蔽化は重要

Akl EA et al. Cochrane Database Syst Rev. 2011

①バイアスリスク(ヘパリンのシナリオ)



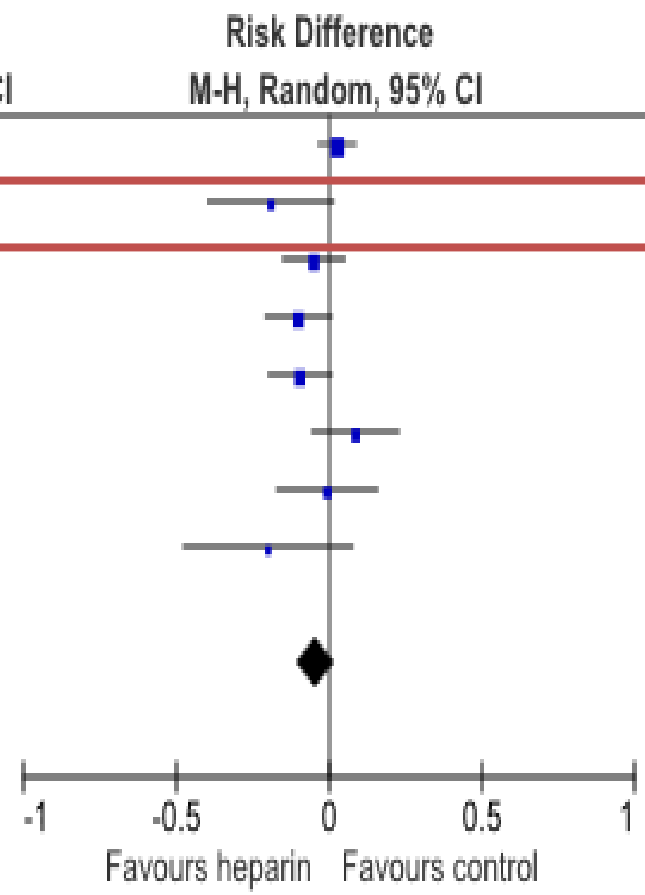
Akl EA et al. Cochrane Database Syst Rev. 2011

①バイアスリスク(ヘパリンのシナリオ)

Study or Subgroup	Heparin Events	Heparin Total	Control Events	Control Total	Weight	Risk Difference M-H, Random, 95% CI
Agnelli 2009	333	769	155	381	22.5%	0.03 [-0.03, 0.09]
Altinbas 2004	22	42	30	42	6.6%	-0.19 [-0.39, 0.01]
Kakkar 2004	103	190	109	184	16.0%	-0.05 [-0.15, 0.05]
Klerk 2005	88	148	107	154	15.0%	-0.10 [-0.21, 0.01]
Lebeau 1994	96	138	110	139	15.7%	-0.10 [-0.20, 0.01]
Perry 2010	45	99	32	87	11.1%	0.09 [-0.05, 0.23]
Sideras 2006	40	68	41	69	9.0%	-0.01 [-0.17, 0.16]
Weber 2008	8	10	10	10	4.0%	-0.20 [-0.48, 0.08]
Total (95% CI)		**1464**		**1066**	**100.0%**	**-0.04 [-0.10, 0.02]**
Total events	735		594			

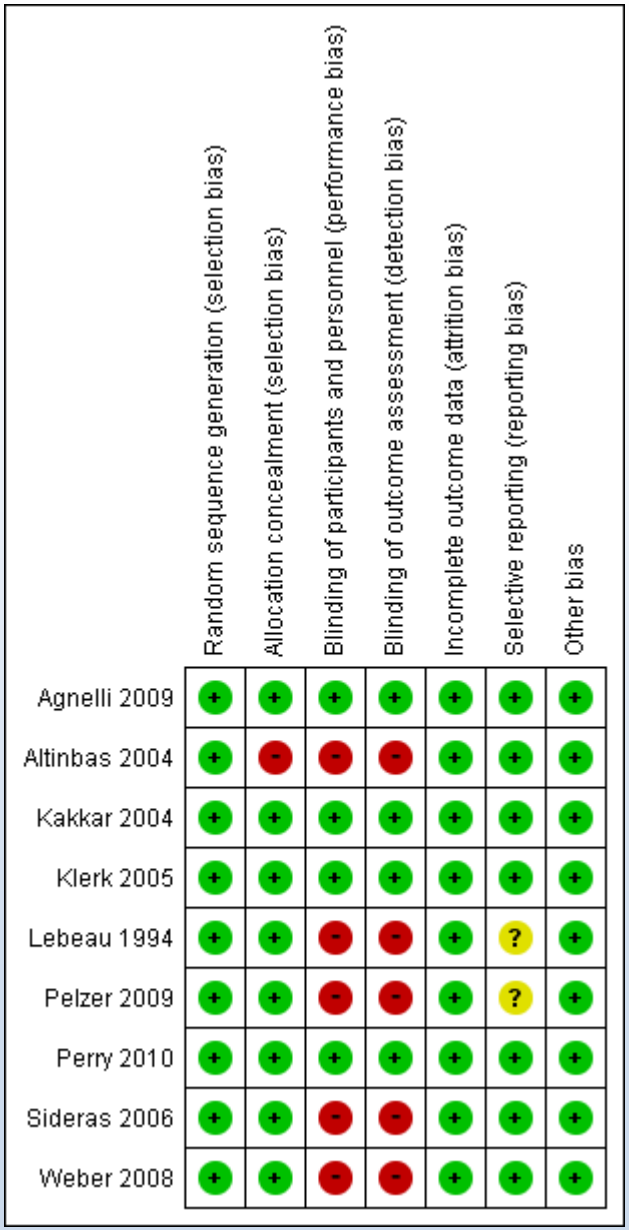
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 13.31$, df = 7 (P = 0.06); $I^2 = 47\%$

Test for overall effect: Z = 1.44 (P = 0.15)



Akl EA et al. Cochrane Database Syst Rev. 2011

①バイアスリスク(ヘパリンのシナリオ)



アウトカム（12か月での死亡）のバイアスリスクは？
-深刻でない
-深刻
-非常に深刻

ほとんどの研究が割り付けの隠蔽化やブラインド化したアウトカム判定を行っていた。いくつかの研究で非ブラインド化の懸念があったが、全体のバイアスリスクは低いと判断された。

Akl EA　et al. Cochrane Database Syst Rev. 2011

①バイアスリスク
1レベルグレードダウン例-1

P:脊髄損傷患者
I:介入A
C:非介入
O:全死亡、運動機能（詳細な身体診察に基づく）
研究デザイン：RCT

アウトカム評価者のブラインド化は行われなかった。
アウトカム評価者のブラインド化は全死亡ではあまり重要でないが、運動機能では極めて重要と考えられた。

アウトカム（全死亡）のバイアスリスクは？
-深刻でない

アウトカム（運動機能）のバイアスリスクは？
-深刻（1レベルグレードダウンに相当）
と判断された。

GRADE handbook 5.2.1 Example 3 (Updated October 2013)

①バイアスリスク
1レベルグレードダウン例-2

P:小児における慢性喘息
I:家族療法＋標準治療
C:標準治療
O:日中の喘鳴
研究デザイン:RCT

全2件のRCTのうち、1件で割り付けの隠蔽化が行われず、ほかの1件で明確でなかった。

アウトカム（日中の喘鳴）のバイアスリスクは？
-深刻（1レベルグレードダウンに相当）
と判断された。

GRADE handbook 5.2.1 Example 4 (Updated October 2013)
Allergy 2009: 64: 669–677

①バイアスリスク
2レベルグレードダウン例-1

P:腰椎椎間板ヘルニア患者
I:手術療法
C:手術なし
O:1年以降の症状
研究デザイン:RCT

全3件のRCTで、割り付けの隠蔽化が不適切で、バイアスが生じうる評価者(外科医)によりブラインド化されずに妥当性が評価されていない方法でアウトカムが評価されていた。

アウトカム(1年以降の症状)のバイアスリスクは?
-非常に深刻(2レベルグレードダウンに相当)
と判断された。

GRADE handbook 5.2.1 Example 6 (Updated October 2013)
Allergy 2009: 64: 669–677

①バイアスリスク
2レベルグレードダウン例-2

P:アレルギー性鼻炎を有する小児
I:舌下免疫療法
C:免疫療法なし
O:喘息の発症
研究デザイン:RCT

全1件のRCTで、ランダム化の方法、割り付けの隠蔽化、解析方法の記載がなく、ブラインド化されたアウトカム評価がされておらず、21%の患者が追跡不能となっていた。

アウトカム(喘息の発症)のバイアスリスクは?
-非常に深刻(2レベルグレードダウンに相当)
と判断された。

GRADE handbook 5.2.1 Example 7 (Updated October 2013)
Allergy 2009: 64: 669–677

エビデンスの確実性評価
②非一貫性 (Inconsistency)

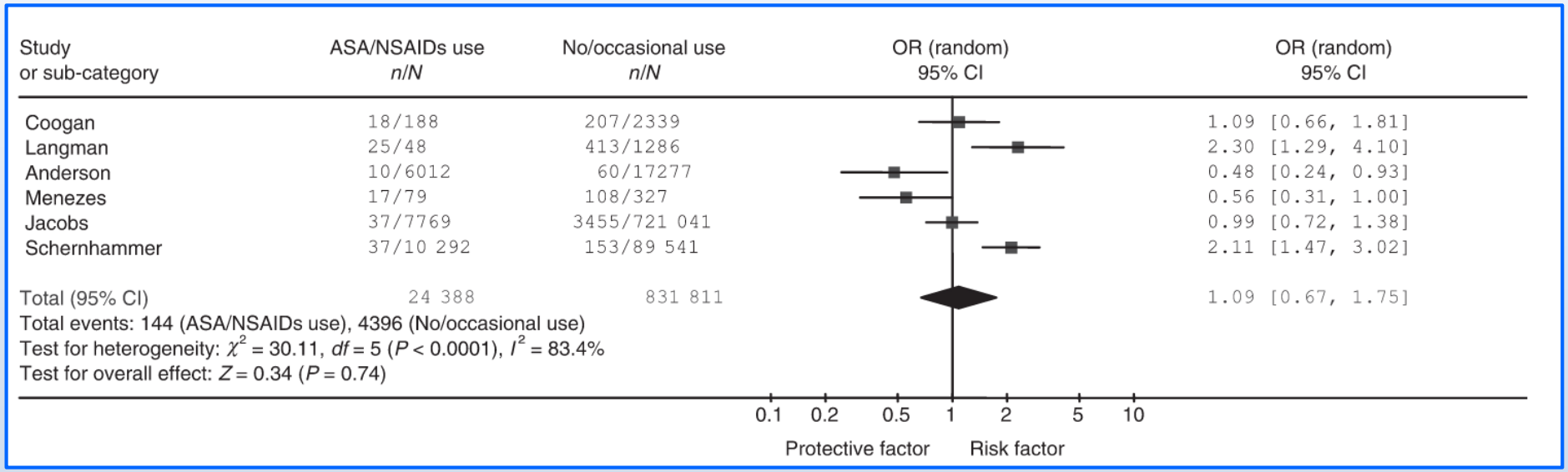
公益財団法人日本医療機能評価機構

Minds

エビデンスの確実性評価
②非一貫性

結果の説明できない 異質性のこと

リスク差などの絶対リスクを扱う指標はベースラインリスクの違いによりばらつきが生じやすいことから、一般に非一貫性はRR・HRなどの相対リスクの指標において評価される



事前に定められた、集団・介入・アウトカム・研究デザイン・解析方法の違いによるサブグループ解析が行われたが、結果の違いは説明できず、深刻な非一貫性ありと判断された。

Aliment Pharmacol Ther 26, 1089–1099

https://sites.google.com/site/gradeworkshopsba2016/tokyo2016

エビデンスの確実性評価
②非一貫性

結果の説明できない 異質性のこと

評価基準

1. 点推定値が研究間でかなり異なる。
2. 信頼区間に重なりがほとんどまたは全くない。
3. 異質性検定のp値が小さい。
4. I^2 統計量が大きい。

(τ^2 統計量も使用可)

I^2統計量による評価の目安

0% - 40%: is low（小さい）

30% - 60%: may be moderate（中程度）

50% - 90%: may be substantial（かなり大きい）

75% - 100%: is considerable（非常に大きい）

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1294-302 (No.7).

②非一貫性

- 異質性を認めた場合、集団・介入・アウトカム等の違いにより結果の違いがあるかを探索する。
- 異質性が説明できない場合、非一貫性が意思決定をするための効果推定値の確信の程度を低下させるか（グレードダウンを行うべきか）について判断する。

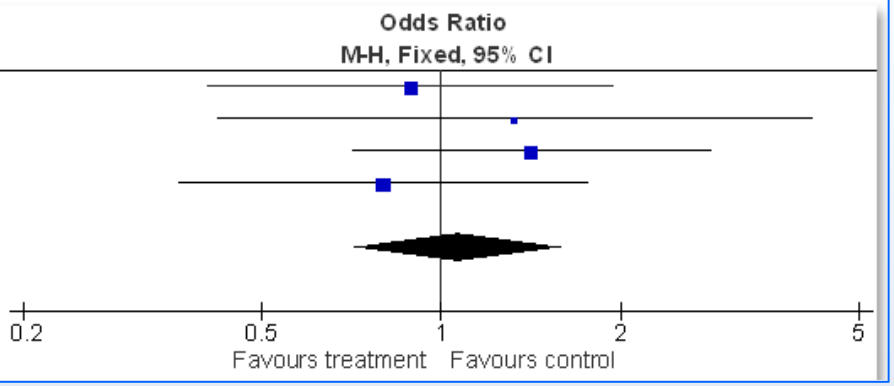
Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1294-302 (No.7).

②非一貫性

1．点推定値の向きは異なるが、深刻な非一貫性はないと判断される

Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Blake 1990	24	50	26	51	29.0%	0.89 [0.41, 1.94]
Gallefoss 1999a	18	26	17	27	11.1%	1.32 [0.42, 4.15]
Littlejohns 1991	33	68	26	65	29.6%	1.41 [0.71, 2.81]
Rea 2004	28	63	21	42	30.3%	0.80 [0.37, 1.75]
Total (95% CI)		**207**		**185**	**100.0%**	**1.07 [0.71, 1.59]**
Total events	103		90			

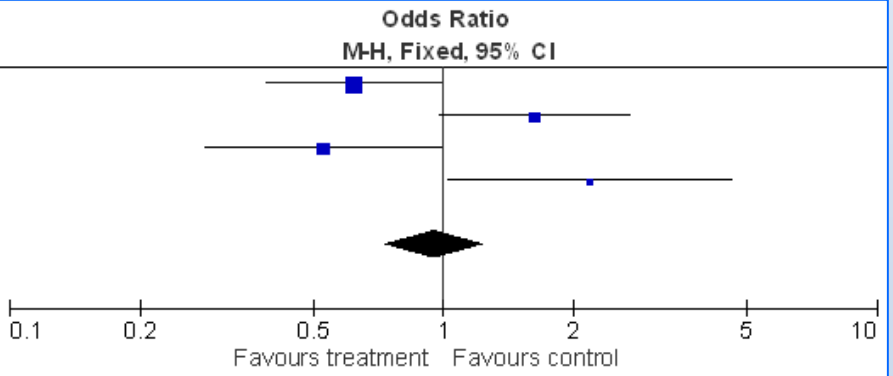
Heterogeneity: Chi² = 1.52, df = 3 (P = 0.68); I² = 0%
Test for overall effect: Z = 0.31 (P = 0.76)



2．信頼区間の重なりがなく、深刻な非一貫性が示唆される

Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Blake 1990	36	234	58	256	43.7%	0.62 [0.39, 0.98]
Coultas 2005a	60	149	37	126	22.3%	1.62 [0.98, 2.68]
Coultas 2005b	19	151	30	140	25.4%	0.53 [0.28, 0.99]
Rea 2004	28	52	22	63	8.6%	2.17 [1.02, 4.61]
Total (95% CI)		**586**		**585**	**100.0%**	**0.95 [0.73, 1.25]**
Total events	143		147			

Heterogeneity: Chi² = 15.63, df = 3 (P = 0.001); I² = 81%
Test for overall effect: Z = 0.34 (P = 0.73)

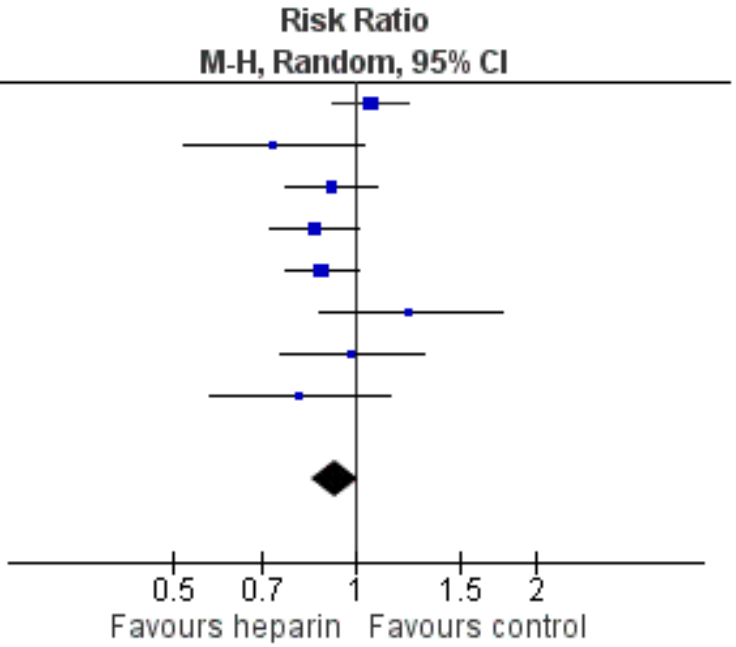


GRADE Handbook(ver 3.2):
http://www.who.int/hiv/topics/mtct/grade_handbook.pdf

②非一貫性 (ヘパリンのシナリオ)

Study or Subgroup	Weight	Risk Ratio M-H, Random, 95% CI
Agnelli 2009	19.7%	1.06 [0.92, 1.23]
Altinbas 2004	6.1%	0.73 [0.52, 1.04]
Kakkar 2004	16.0%	0.92 [0.77, 1.09]
Klerk 2005	16.9%	0.86 [0.72, 1.01]
Lebeau 1994	20.6%	0.88 [0.76, 1.01]
Perry 2010	6.0%	1.24 [0.87, 1.75]
Sideras 2006	8.6%	0.99 [0.75, 1.31]
Weber 2008	6.1%	0.81 [0.57, 1.14]
Total (95% CI)	**100.0%**	**0.93 [0.85, 1.02]**
Total events		

Heterogeneity: Tau² = 0.01; Chi² = 10.75, df = 7 (P = 0.15); I² = 35%
Test for overall effect: Z = 1.57 (P = 0.12)

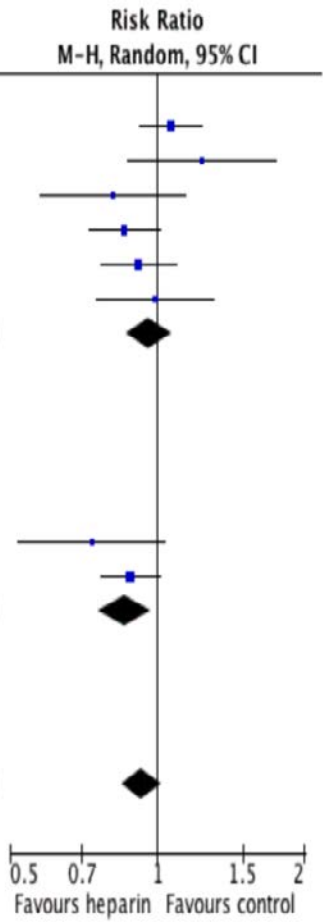


アウトカム（12か月での死亡）の
非一貫性は？
-深刻でない
-深刻
-非常に深刻

Akl EA et al. Cochrane Database Syst Rev. 2011

②非一貫性 (ヘパリンのシナリオ)

Study or Subgroup	Heparin Events	Heparin Total	Control Events	Control Total	Weight	Risk Ratio M-H, Random, 95% CI
1.1.1 Cancers other than SCLC						
Agnelli 2009	333	769	155	381	19.7%	1.06 [0.92, 1.23]
Perry 2010	45	99	32	87	6.0%	1.24 [0.87, 1.75]
Weber 2008	8	10	10	10	6.1%	0.81 [0.57, 1.14]
Klerk 2005	88	148	107	154	16.9%	0.86 [0.72, 1.01]
Kakkar 2004	103	190	109	184	16.0%	0.92 [0.77, 1.09]
Sideras 2006	40	68	41	69	8.6%	0.99 [0.75, 1.31]
Subtotal (95% CI)		**1284**		**885**	**73.3%**	**0.96 [0.86, 1.07]**
Total events	617		454			
Heterogeneity: $Tau^2 = 0.01$; $Chi^2 = 7.37$, df = 5 (P = 0.19); $I^2 = 32\%$						
Test for overall effect: Z = 0.73 (P = 0.47)						
1.1.2 SCLC						
Altinbas 2004	22	42	30	42	6.1%	0.73 [0.52, 1.04]
Lebeau 1994	96	138	110	139	20.6%	0.88 [0.76, 1.01]
Subtotal (95% CI)		**180**		**181**	**26.7%**	**0.86 [0.75, 0.98]**
Total events	118		140			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.93$, df = 1 (P = 0.33); $I^2 = 0\%$						
Test for overall effect: Z = 2.34 (P = 0.02)						
Total (95% CI)		**1464**		**1066**	**100.0%**	**0.93 [0.85, 1.02]**
Total events	735		594			
Heterogeneity: $Tau^2 = 0.01$; $Chi^2 = 10.75$, df = 7 (P = 0.15); $I^2 = 35\%$						
Test for overall effect: Z = 1.57 (P = 0.12)						

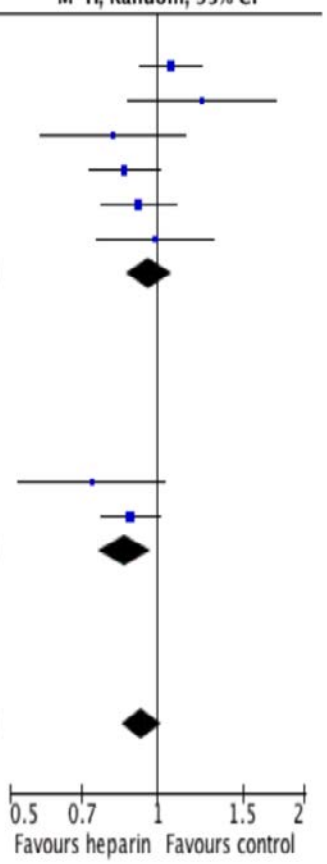


アウトカム（12か月での死亡）の非一貫性は？
-深刻でない
-深刻
-非常に深刻

Akl EA et al. Cochrane Database Syst Rev. 2011

②非一貫性 (ヘパリンのシナリオ)

Study or Subgroup	Heparin Events	Heparin Total	Control Events	Control Total	Weight	Risk Ratio M-H, Random, 95% CI
1.1.1 Cancers other than SCLC						
Agnelli 2009	333	769	155	381	19.7%	1.06 [0.92, 1.23]
Perry 2010	45	99	32	87	6.0%	1.24 [0.87, 1.75]
Weber 2008	8	10	10	10	6.1%	0.81 [0.57, 1.14]
Klerk 2005	88	148	107	154	16.9%	0.86 [0.72, 1.01]
Kakkar 2004	103	190	109	184	16.0%	0.92 [0.77, 1.09]
Sideras 2006	40	68	41	69	8.6%	0.99 [0.75, 1.31]
Subtotal (95% CI)		**1284**		**885**	**73.3%**	**0.96 [0.86, 1.07]**
Total events	617		454			
Heterogeneity: $Tau^2 = 0.01$; $Chi^2 = 7.37$, df = 5 (P = 0.19); $I^2 = 32\%$						
Test for overall effect: Z = 0.73 (P = 0.47)						
1.1.2 SCLC						
Altinbas 2004	22	42	30	42	6.1%	0.73 [0.52, 1.04]
Lebeau 1994	96	138	110	139	20.6%	0.88 [0.76, 1.01]
Subtotal (95% CI)		**180**		**181**	**26.7%**	**0.86 [0.75, 0.98]**
Total events	118		140			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.93$, df = 1 (P = 0.33); $I^2 = 0\%$						
Test for overall effect: Z = 2.34 (P = 0.02)						
Total (95% CI)		**1464**		**1066**	**100.0%**	**0.93 [0.85, 1.02]**
Total events	735		594			
Heterogeneity: $Tau^2 = 0.01$; $Chi^2 = 10.75$, df = 7 (P = 0.15); $I^2 = 35\%$						
Test for overall effect: Z = 1.57 (P = 0.12)						



アウトカム（12か月での死亡）の非一貫性は？

-深刻でない

-深刻（※1レベル下げるほどではない）

-非常に深刻

中等度の非一貫性（I^2=35%)を認め、サブグループ解析では小細胞肺がん（SCLC)患者においては生存が延長することが示唆されたが、ほかのがん患者では示唆されなかった。不精確さと併せて、エビデンスの確実性が1レベルグレードダウンされた。

Akl EA　et al. Cochrane Database Syst Rev. 2011

エビデンスの確実性評価
③非直接性 (Indirectness)

公益財団法人日本医療機能評価機構

Minds

エビデンスの確実性評価
③非直接性

CQで設定したPICOとシステマティックレビューで選択された研究のPICOとの差異のこと

4種類の非直接性を評価する

1. 対象集団の違い (Patients or populations)
2. 介入(や対照）の違い (Intervention or comparison)
3. アウトカムの違い (Outcome：代用アウトカム等)
4. 間接比較

GRADE handbook (Updated October 2013)

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1303-10 (No.8).

③非直接性の例

１．対象集団の違い

鳥インフルエンザ患者に対するオセルタミビルの効果の評価に季節性インフルエンザ患者における研究結果が用いられた場合。

２．介入の違い

大腸内視鏡検査による介入効果の評価に、便潜血検査による死亡リスク減少効果を示した結果を用いた場合。

３．アウトカムの違い

カルシウムやリンの排泄率を血管石灰化や血管イベントのサロゲートアウトカムとして使用した場合。

４．間接比較

RCTで直接比較されていない２つの抗うつ剤の効果を間接比較した場合。

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1303-10.(No.8).

③非直接性 (ヘパリンのシナリオ)

Altinbas 2004

Methods	Randomized clinical trial
Participants	84 patients with histologically confirmed small cell lung cancer (both limited and extensive) undergoing combination chemotherapy, ECOG performance status < 3; median age 58; 81% males; country: Turkey
Interventions	LMWH (Dalteparin; prophylactic dose) versus no intervention for 18 weeks or less if disease progressed
Outcomes	Outcomes: all-cause mortality (at 12 and 24 months), symptomatic DVT (no PE events; personal communication with author), and bleeding. Screening and diagnostic testing for DVT: not reported
Notes	Funding: not reported Follow up: median of 10 months; ra HR not adjusted (analyses were uni

アウトカム（12か月での死亡）の非一貫性は？
-深刻でない
-深刻
-非常に深刻
深刻な非直接性はないと判断された。

Akl EA　et al. Cochrane Database Syst Rev. 2011

エビデンスの確実性評価
④不精確さ(Imprecision)

公益財団法人日本医療機能評価機構

Minds

エビデンスの確実性評価
④不精確さ

- 不精確さ:ランダム誤差（random error）のこと
- 不精確さは、診療ガイドラインにおける推奨作成とシステマティックレビューで評価方法が異なる。

GRADE handbook (Updated October 2013)

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1283-93. (No.6).

④不精確さ:推奨作成

- 推奨作成においては、

１．（95%）信頼区間の上限と下限の値によって臨床的な決断が異なると判断される場合にグレードダウンする。

２．サンプル数とイベント数が少ない場合（*）には、効果推定値が大きい結果であってもグレードダウンを考慮する。

* Optimal information size（OIS 最適情報量）を満たさない場合

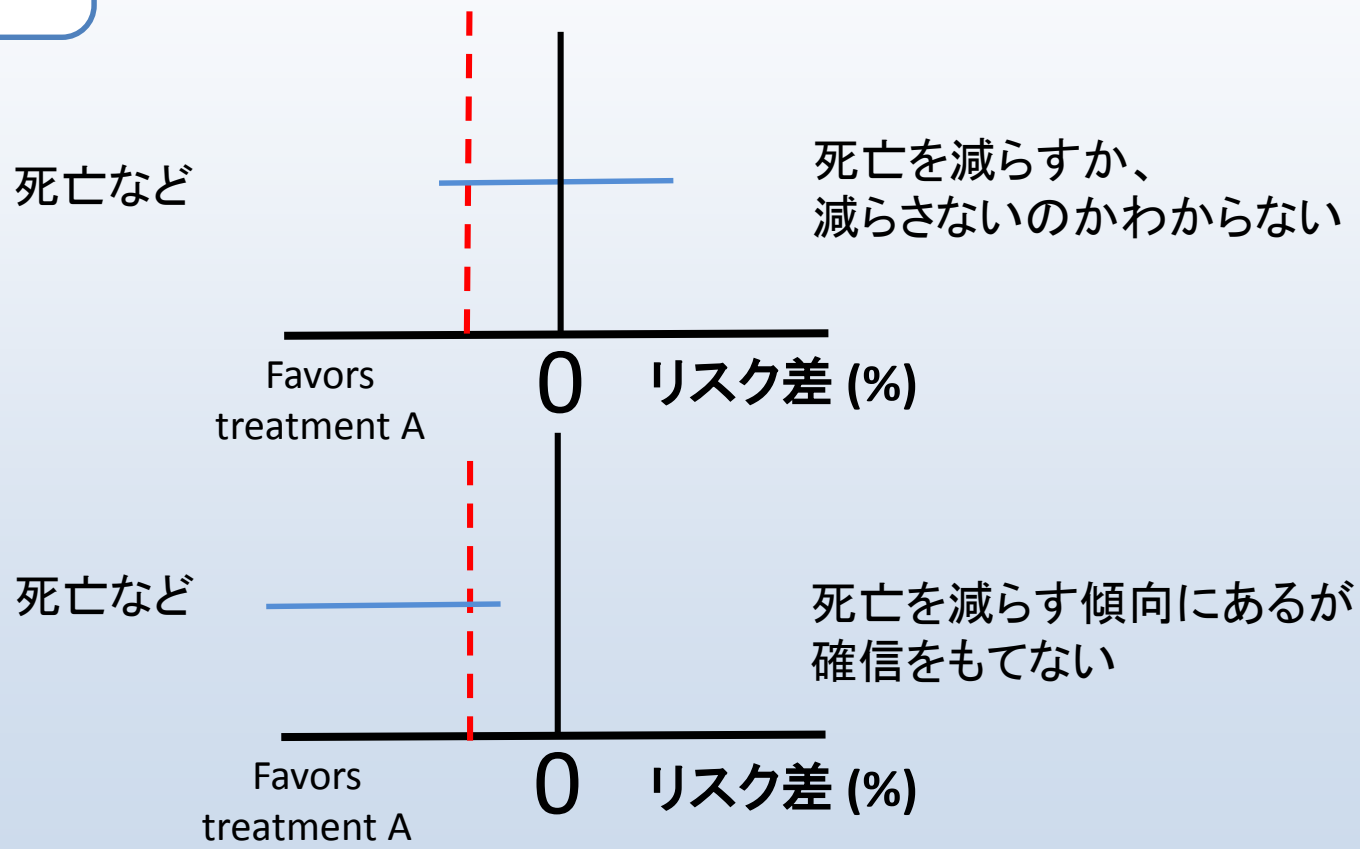
GRADE handbook (Updated October 2013)

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1283-93. (No.6).

臨床的決断の閾値

ほかのアウトカム
(コスト含む)を考慮
して決定する

治療を推奨するための治療効果の閾値
(効果がこの値より大きければ治療を推奨できる)

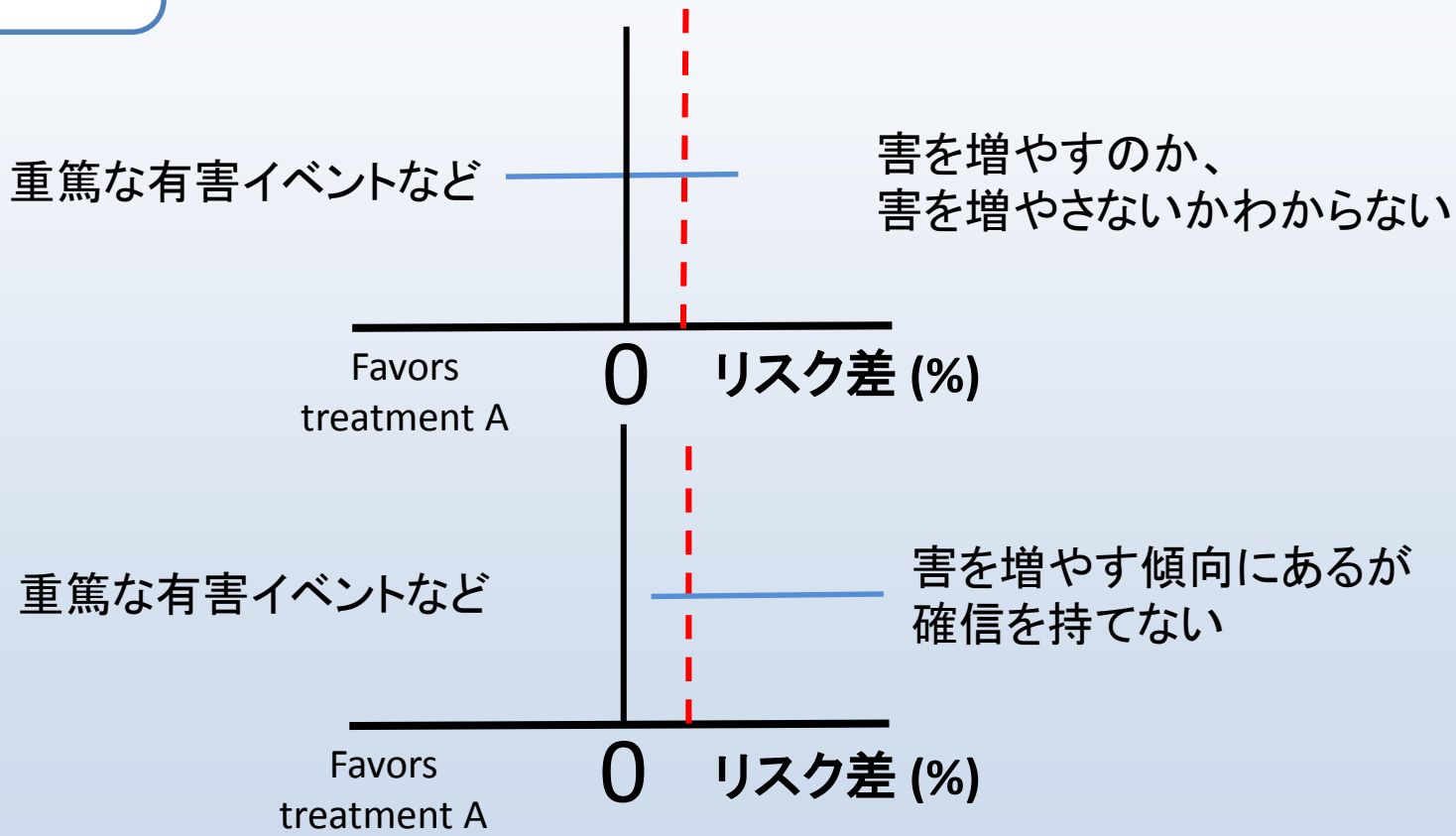


GRADE Handbook(ver 3.2):
http://www.who.int/hiv/topics/mtct/grade_handbook.pdf

臨床的決断の閾値

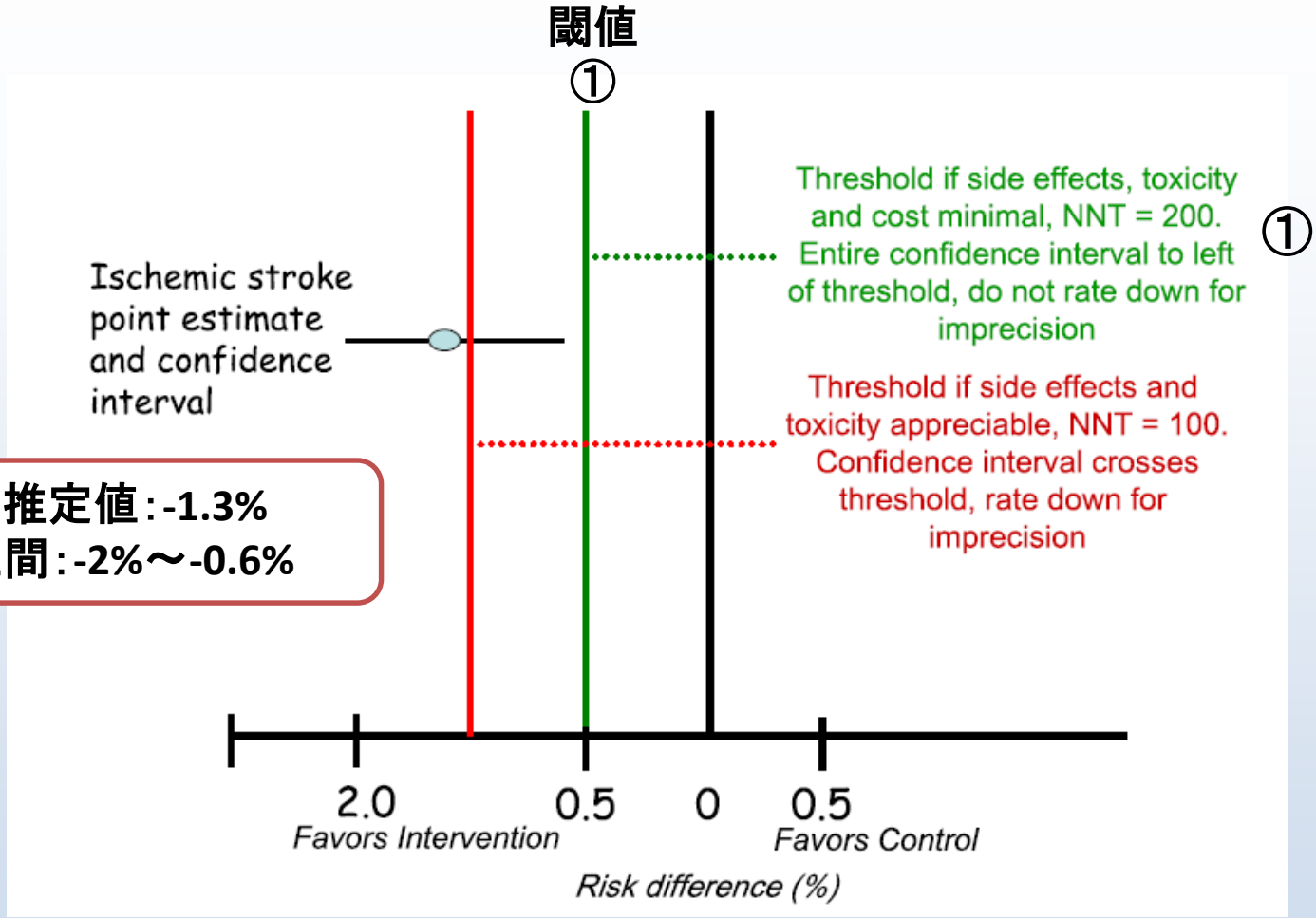
ほかのアウトカム
(コスト含む)を考慮
して決定する

治療を推奨するために許容できる害の閾値
(この値より小さければ治療を推奨できる)



GRADE Handbook(ver 3.2):
http://www.who.int/hiv/topics/mtct/grade_handbook.pdf

臨床的決断の閾値

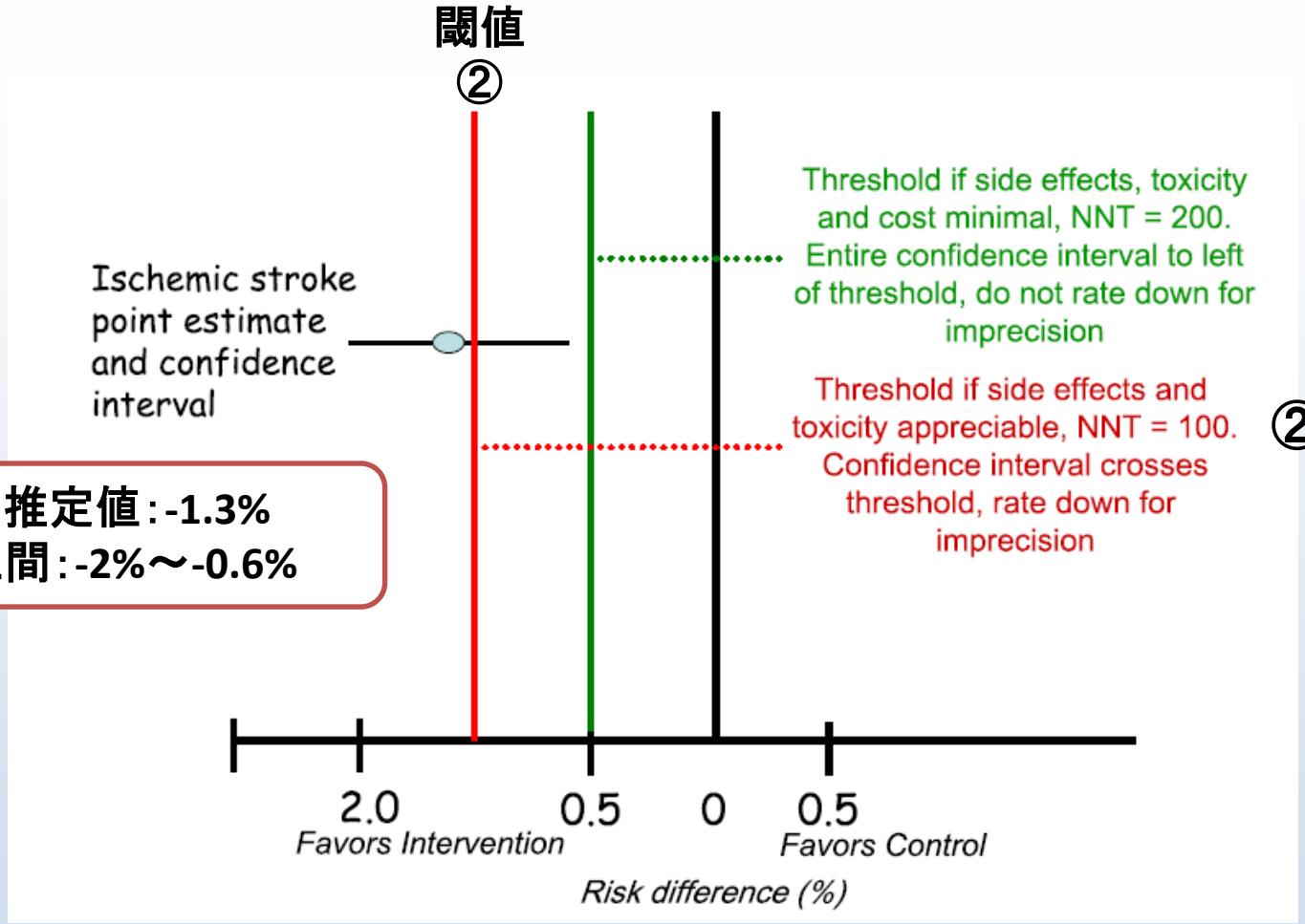


リスク比 点推定値：-1.3%
95%信頼区間：-2%～-0.6%

①治療による害やコストが小さく、治療を推奨する閾値はリスク差=0.5%と考えられた場合、95%信頼区間が閾値をまたいでいない→必要サンプル数を満たせばグレードダウンしない。

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1283-93. (No.6).

臨床的決断の閾値



リスク比 点推定値：-1.3%
95%信頼区間：-2%～-0.6%

②治療による害やコストが大きく、治療を推奨する閾値はリスク差=1.0%と考えられた場合、95%信頼区間が閾値をまたいでいる→グレードダウンする

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1283-93. (No.6).

臨床的決断の閾値-Rule of thumb

- レビューワーが閾値の判断ができない場合

リスク比(RR) =0.75, 1.25

を相当な益もしくは害の閾値の目安として使用可。

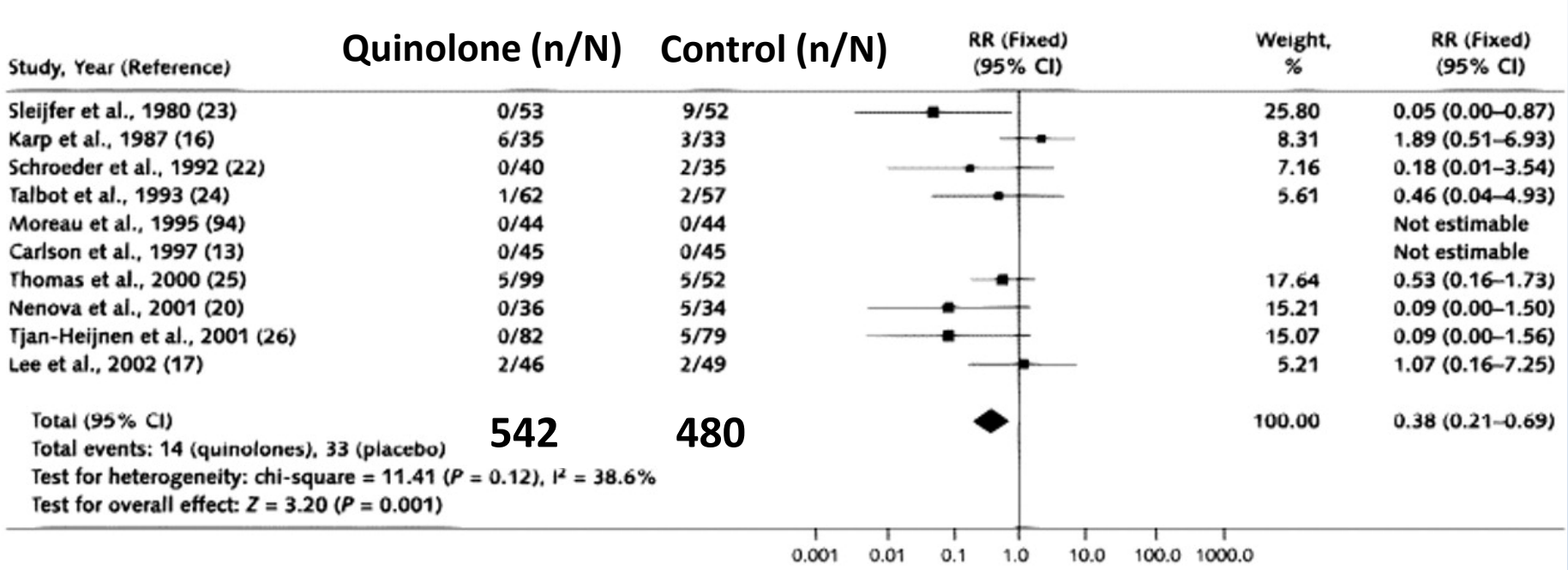
ただし、推奨決定の際には、この閾値が適切かどうかを判断する必要がある

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1283-93. (No.6).

信頼区間の限界

P:好中球減少患者
I:Fluoroquinoloneによる予防
C:Fluoroquinoloneなし
O:感染症による死亡
デザイン：RCT

対照群のリスク: 0.069, Δ=0.25 (RRR)
α=0.05, β=0.20
OIS(必要総サンプル数）: N=6400

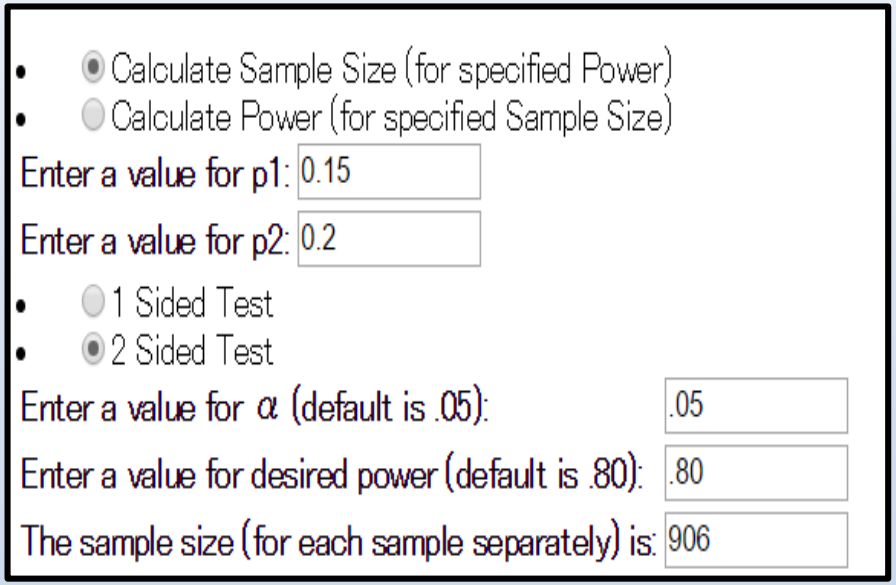


95%信頼区間は大きな治療効果を示唆したが、サンプル数(N=1022)がOISを満たしていないのでグレードダウンが必要と考えられた。

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):Box4 (No.6).
Ann Intern Med. 2005;142:979-995. Figure 3

Optimal information size
（OIS 最適情報量）

- 十分な検出力のある試験を設計するために計算されたサンプル数のこと
- サンプル数計算ソフトを用いて計算できる
  http://stat.ubc.ca/~rollin/stats/ssize/b2.html



対照群のリスク(p2): 0.20
介入群のリスク(p1):0.15
Δ=0.25 (relative risk reduction : RRR)
α=0.05, β=0.20
と設定

対照群のリスクの目安：
対照群発生割合の平均値や中央値

Δ(RRR)の目安：0.2～0.3

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1283-93. (No.6).

グレードダウンを考慮する
イベント数、サンプル数の目安

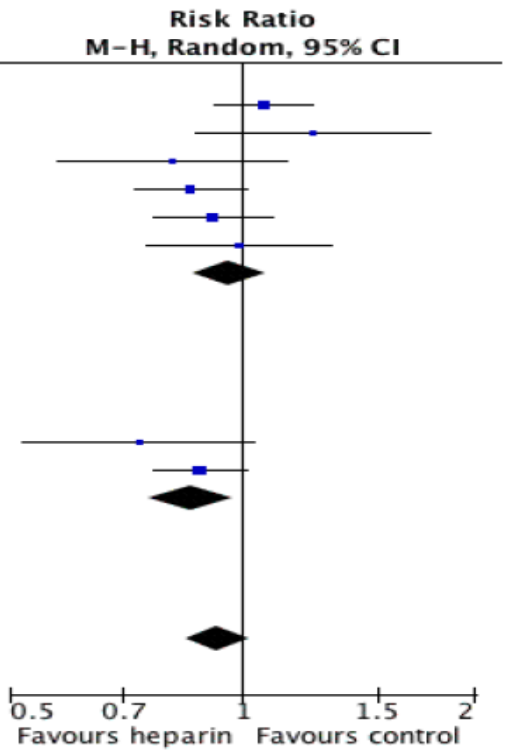
アウトカムが2値変数：総イベント数< 300

アウトカムが連続変数：総サンプル数< 400

GRADE Handbook(ver 3.2):
http://www.who.int/hiv/topics/mtct/grade_handbook.pdf

④不精確さ:ヘパリンのシナリオ

Study or Subgroup	Heparin Events	Heparin Total	Control Events	Control Total	Weight	Risk Ratio M-H, Random, 95% CI
1.1.1 Cancers other than SCLC						
Agnelli 2009	333	769	155	381	19.7%	1.06 [0.92, 1.23]
Perry 2010	45	99	32	87	6.0%	1.24 [0.87, 1.75]
Weber 2008	8	10	10	10	6.1%	0.81 [0.57, 1.14]
Klerk 2005	88	148	107	154	16.9%	0.86 [0.72, 1.01]
Kakkar 2004	103	190	109	184	16.0%	0.92 [0.77, 1.09]
Sideras 2006	40	68	41	69	8.6%	0.99 [0.75, 1.31]
Subtotal (95% CI)		**1284**		**885**	**73.3%**	**0.96 [0.86, 1.07]**
Total events	617		454			
Heterogeneity: Tau² = 0.01; Chi² = 7.37, df = 5 (P = 0.19); I² = 32%						
Test for overall effect: Z = 0.73 (P = 0.47)						
1.1.2 SCLC						
Altinbas 2004	22	42	30	42	6.1%	0.73 [0.52, 1.04]
Lebeau 1994	96	138	110	139	20.6%	0.88 [0.76, 1.01]
Subtotal (95% CI)		**180**		**181**	**26.7%**	**0.86 [0.75, 0.98]**
Total events	118		140			
Heterogeneity: Tau² = 0.00; Chi² = 0.93, df = 1 (P = 0.33); I² = 0%						
Test for overall effect: Z = 2.34 (P = 0.02)						
Total (95% CI)		**1464**		**1066**	**100.0%**	**0.93 [0.85, 1.02]**
Total events	735		594			
Heterogeneity: Tau² = 0.01; Chi² = 10.75, df = 7 (P = 0.15); I² = 35%						
Test for overall effect: Z = 1.57 (P = 0.12)						



アウトカム（12か月での死亡）の不精確さは？
-深刻でない
-深刻
-非常に深刻

Akl EA et al. Cochrane Database Syst Rev. 2011

④不精確さ:ヘパリンのシナリオ

アウトカム:12か月での死亡

総サンプル数	2530人
総死亡数	1329件
リスク比	0.93（95%CI 0.85～1.02）
リスク差	1000人当たり45件減少 (92減少～ 13増加）

アウトカム（12か月での死亡）の不精確さは？
-深刻でない
-深刻
-非常に深刻

Akl EA　et al. Cochrane Database Syst Rev. 2011

④不精確さ:ヘパリンのシナリオ

OIS計算

- (●) Calculate Sample Size (for specified Power)
- ( ) Calculate Power (for specified Sample Size)

Enter a value for p1: 0.65

対照群発生割合の中央値

Enter a value for p2: 0.4875

対照群発生割合の中央値×0.75
Δ(RRR)=0.25に相当

- ( ) 1 Sided Test
- (●) 2 Sided Test

Enter a value for α (default is .05): .05

Enter a value for desired power (default is .80): .80

The sample size (for each sample separately) is: 145

総サンプル数(=2530)は、OIS (=290)より多い

https://www.stat.ubc.ca/~rollin/stats/ssize/b2.html

④不精確さ:ヘパリンのシナリオ

アウトカム:12か月での死亡
総サンプル数　2530人
総死亡数　1329件
リスク比　0.93（95%CI 0.85～1.02）
リスク差　4.5%減少 (9.2%減少～ 1.3%増加）

システマティック・レビューの場合

アウトカム（12か月での死亡）の不精確さは？
-深刻でない
-深刻（※1レベル下げるほどではない）
-非常に深刻
(OISは満たしていたが)
95%信頼区間が益と効果なしを含んでおり、非一貫性と併せて、エビデンスの確実性を1レベルグレードダウンされた。

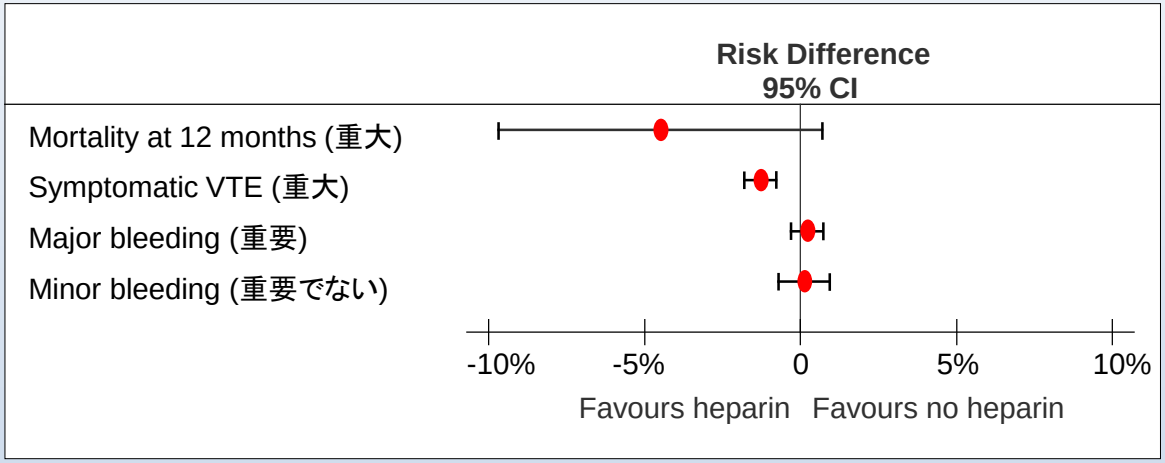
Akl EA　et al. Cochrane Database Syst Rev. 2011

④不精確さ:ヘパリンのシナリオ

アウトカム:12か月での死亡

総サンプル数	2530人
総死亡数	1329件
リスク比	0.93（95%CI 0.85～1.02）
リスク差	4.5%減少 (9.2%減少～ 1.3%増加）

推奨作成の場合
他のアウトカムやコストを考慮して、閾値を判断する。

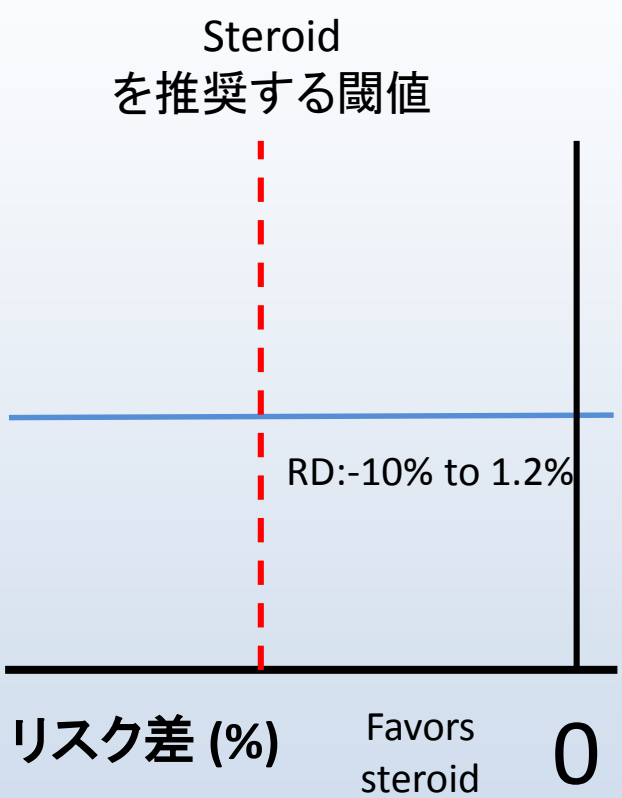


コスト:ヘパリンナトリウム注　5千単位:約200円

アウトカム（12か月での死亡）の不精確さは？
-深刻でない
-深刻
-非常に深刻
他のアウトカムやコストを考慮して、治療を推奨するための閾値を-0.5%と判断すると、1レベルグレードダウン相当。

Akl EA　et al. Cochrane Database Syst Rev. 2011

④不精確さ
1レベルグレードダウン 例



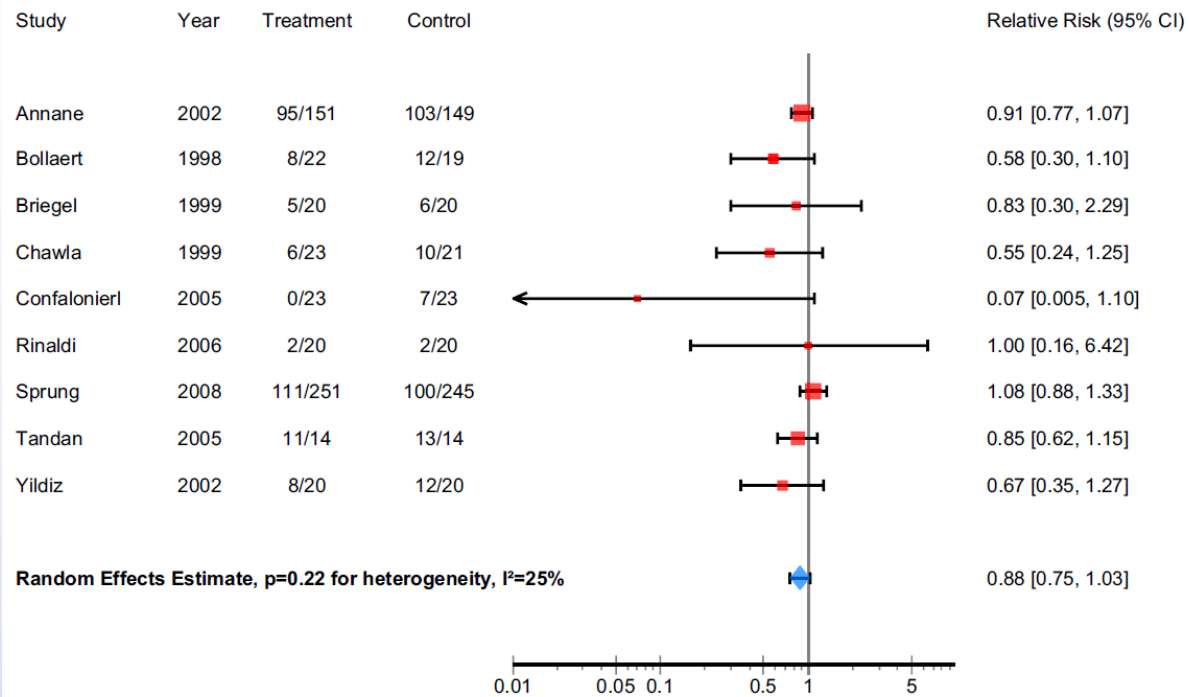


Fig. 2. Corticosteroids to reduce hospital mortality in septic shock.

アウトカム（死亡）の不精確さは？
-深刻（1レベルグレードダウンに相当）
総イベント数は511でOISを満たしているが、リスク差の信頼区間は臨床的決断の閾値をまたいでいると判断された。

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12): No.6. Box 2

④不精確さ
2レベルグレードダウン 例

P:クローン病患者
I:プロバイオティクス
C:プロバイオティクスなし
O:寛解
研究デザイン:RCT

全1件のRCTで、介入群で5人中4人で、対照群で6人中5人で寛解
リスク比: 0.96 (95%CI 0.56~1.69)
リスク差: -3% (95%CI -49%~43%)

アウトカム(寛解)の不精確さは?
-非常に深刻(2 レベルグレードダウン相当)
寛解の総数は9件で非常に少なく、リスク比・リスク差の両方の信頼区間が相当な益と害の両方をまたいでいると判断された。

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12): No.6. p.1290

エビデンスの確実性評価
⑤出版バイアス(Publication bias)

公益財団法人日本医療機能評価機構

Minds

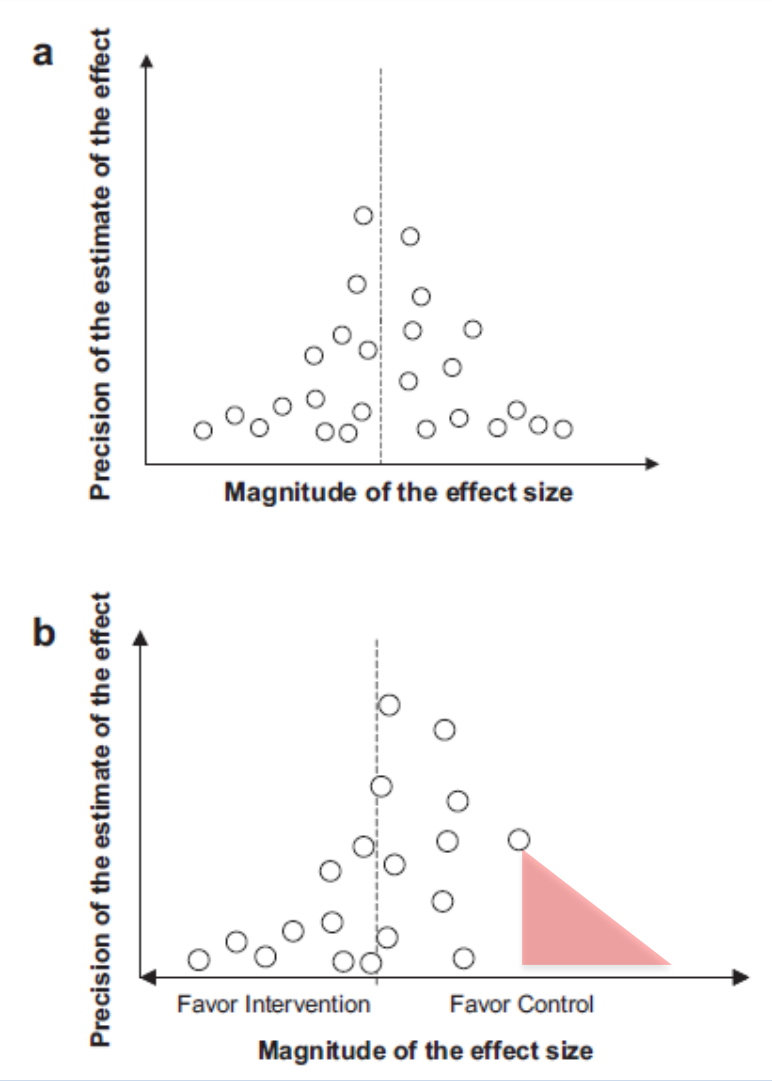
エビデンスの確実性評価
⑤出版バイアス

研究が選択的に出版されることにより、本来ある有益もしくは有害な効果を系統的に過小もしくは過大評価してしまうこと。

GRADE handbook (Updated October 2013)

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1277-82. (No.5).

⑤出版バイアス

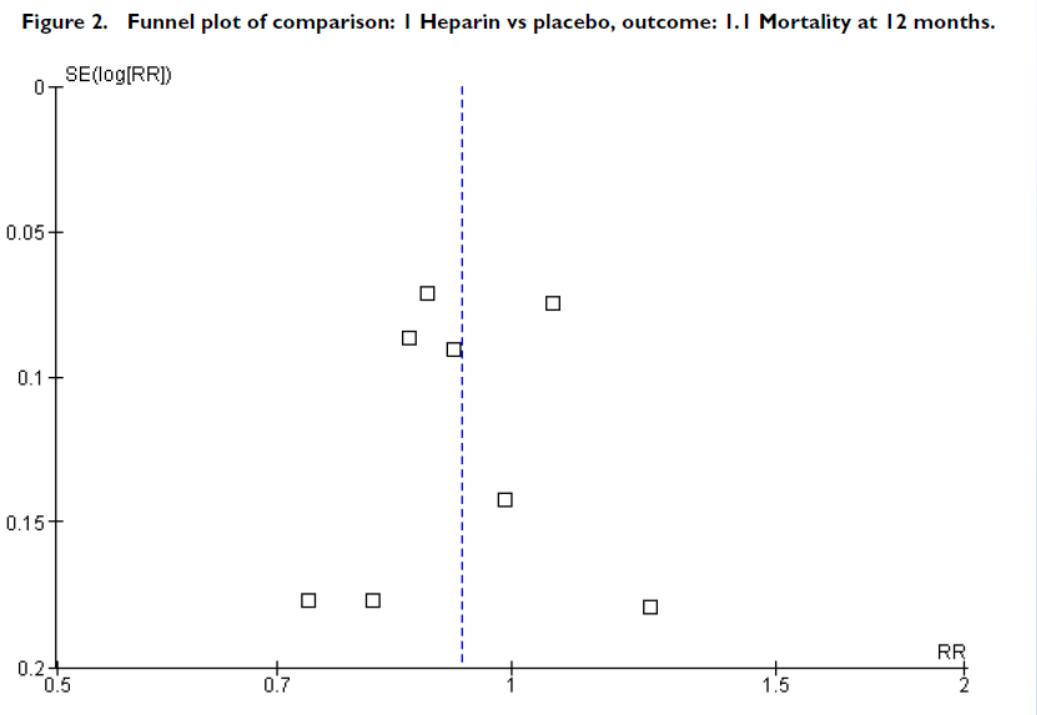


出版バイアスを評価する方法
1. ファンネルプロットの左右対称性の目視確認
2. ファンネルプロットの左右対称性の統計学的検定（EggerやBeggの検定）
3. Trim-and-Fillの方法による評価
ただし、研究数が10件以上（5件以上という専門家もいる）のときに限り、ファンネルプロットや検定を用いるべきとされている。

GRADE handbook (Updated October 2013)

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1277-82. (No.5).

⑤出版バイアス (ヘパリンのシナリオ)



アウトカム（12か月での死亡）の出版バイアスは？

-深刻でない

-深刻

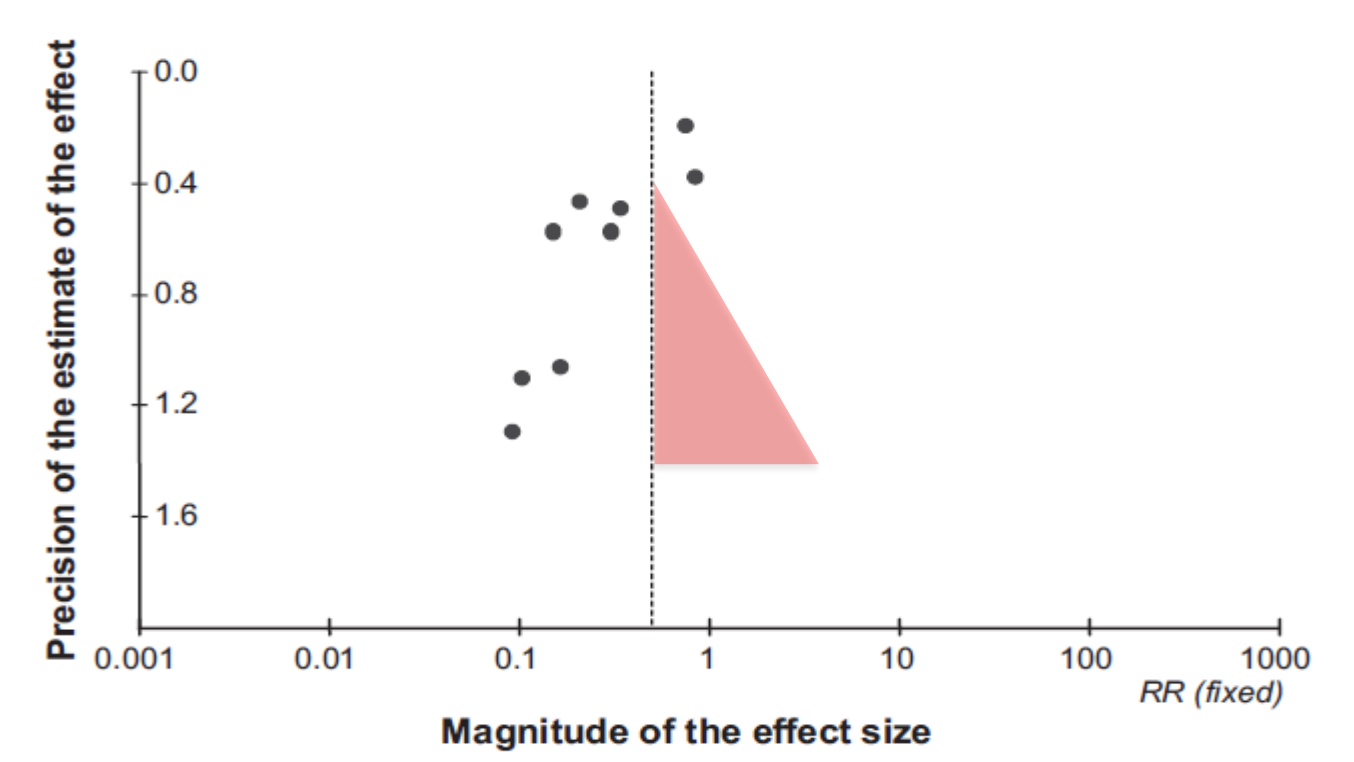
-非常に深刻

深刻な出版バイアスはないものと判断された。

Akl EA　et al. Cochrane Database Syst Rev. 2011

⑤出版バイアス
(1レベルグレードダウン例)

P:痔核
I:フラボノイド
C:フラボノイドなし
O:症状改善
研究デザイン:RCT



アウトカム(症状改善)の出版バイアスは？
-深刻(1レベルグレードダウン相当)
全てが企業が出資した研究であり、ファンネルプロットが非対称であることから、1レベルグレードダウンするのが適切だろう

Guyatt et al. J Clin Epidemiol. 2011 Dec;64(12):1277-82. (No.5)　Fig 2.

エビデンスの確実性評価
まとめ

公益財団法人日本医療機能評価機構

Minds

アウトカム毎のエビデンスの確実性を評価

(最初の) エビデンスの 確実性	グレードを下げる 5ドメイン	グレードを上げる 3ドメイン(観察研究)	(最終的な) エビデンスの確実性
RCT → 高 観察研究 → 低	・バイアスリスク -1 深刻 -2 非常に深刻 ・非一貫性 -1 深刻 -2 非常に深刻 ・非直接性 -1 深刻 -2 非常に深刻 ・不精確さ -1 深刻 -2 非常に深刻 ・出版バイアス -1 可能性あり -2 可能性高い	・効果が大きい ＋1 大きな効果 ＋2 非常に大きな効果 ・用量反応勾配 ＋1 あり ・全ての交絡を考慮した際 ＋1 効果を減弱させている ＋1 効果なしの結果であった場合に、効果を増加させる方向に働いている	**高（A）** ⊕⊕⊕⊕ **中（B）** ⊕⊕⊕ **低（C）** ⊕⊕ **非常に低(D)** ⊕

診療ガイドラインパネリストがエビデンスの確実性の評価の最終判断をする

Guyatt et al. J Clin Epidemiol. 2011 Apr;64(4):401-6. (No.3)
GRADE handbook (Updated October 2013)

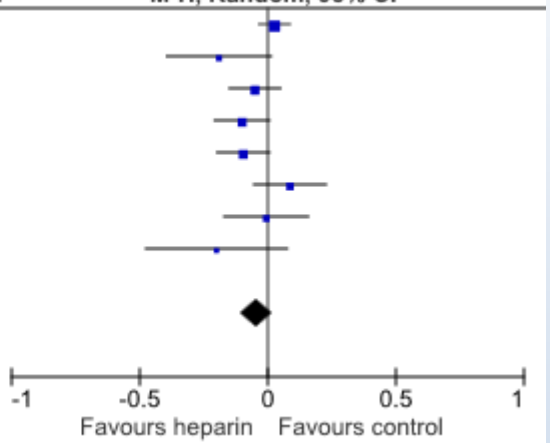
エビデンス（総体）の確実性とは

- 推奨作成における定義

効果推定値がある推奨を支持するのに適切であるという確信の程度

Study or Subgroup	Heparin Events	Heparin Total	Control Events	Control Total	Weight	Risk Difference M-H, Random, 95% CI
Agnelli 2009	333	769	155	381	22.5%	0.03 [-0.03, 0.09]
Altinbas 2004	22	42	30	42	6.6%	-0.19 [-0.39, 0.01]
Kakkar 2004	103	190	109	184	16.0%	-0.05 [-0.15, 0.05]
Klerk 2005	88	148	107	154	15.0%	-0.10 [-0.21, 0.01]
Lebeau 1994	96	138	110	139	15.7%	-0.10 [-0.20, 0.01]
Perry 2010	45	99	32	87	11.1%	0.09 [-0.05, 0.23]
Sideras 2006	40	68	41	69	9.0%	-0.01 [-0.17, 0.16]
Weber 2008	8	10	10	10	4.0%	-0.20 [-0.48, 0.08]
Total (95% CI)		**1464**		**1066**	**100.0%**	**-0.04 [-0.10, 0.02]**
Total events	735		594			

Heterogeneity: Tau² = 0.00; Chi² = 13.31, df = 7 (P = 0.06); I² = 47%
Test for overall effect: Z = 1.44 (P = 0.15)



アウトカム（12か月での死亡）のエビデンスの確実性は？

中 ⊕⊕⊕

バイアスリスク、非直接性、出版バイアスの深刻な問題はなく、若干の非一貫性および不精確さがあったため、併せて1レベルグレードダウンし、最終的なエビデンスの確実性は中と判断された。

GRADE handbook (Updated October 2013)

Evidence Profile (シナリオ例)

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	heparin	no heparin	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: 12 months)												
8	randomised trials	not serious [a]	serious [b]	not serious	not serious [c]	none	735/1464 (50.2%)	64.9%	**RR 0.93** (0.85 to 1.02)	**45 fewer per 1,000** (from 13 more to 97 fewer)	⊕⊕⊕◯ MODERATE	CRITICAL
Symptomatic VTE (follow up: 12 months)												
7	randomised trials	not serious [a]	not serious	not serious	not serious	none	38/1338 (2.8%)	2.9%	**RR 0.55** (0.37 to 0.82)	**13 fewer per 1,000** (from 5 fewer to 18 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Major bleeding (follow up: 12 months)												
9	randomised trials	not serious [a]	not serious	not serious	serious [d]	none	30/1624 (1.8%)	0.7%	**RR 1.30** (0.59 to 2.88)	**2 more per 1,000** (from 3 fewer to 13 more)	⊕⊕⊕◯ MODERATE	IMPORTANT

a. Vast majority of studies had allocation concealment , and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.
b. There is moderate heterogeneity among studies included in the analysis of death at 12 months (I2=41%). The subgroup analysis for mortality at 12 months was statistically significant and suggested survival benefit in patients with SCLC but not in patients with advanced cancer. Overall we decided to downgrade by one level when considering these issues along with imprecision.
c. CI interval includes effects suggesting benefit as well as no benefit.
d. CI includes possibility of both harms or benefits.

https://sites.google.com/site/gradeworkshopsba2016/tokyo2016

Akl EA et al. Cochrane Database Syst Rev. 2011

Summary of Findings Table (ヘパリンのシナリオ)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no heparin	Risk with heparin				
Mortality follow up: 12 months	**Moderate**		**RR 0.93** (0.85 to 1.02)	2531 (8 RCTs)	⊕⊕⊕◯ MODERATE [a,b,c]	
	649 per 1,000	**604 per 1,000** (552 to 662)				
Symptomatic VTE follow up: 12 months	**Moderate**		**RR 0.55** (0.37 to 0.82)	2264 (7 RCTs)	⊕⊕⊕⊕ HIGH [a]	
	29 per 1,000	**16 per 1,000** (11 to 24)				
Major bleeding follow up: 12 months	**Moderate**		**RR 1.30** (0.59 to 2.88)	2843 (9 RCTs)	⊕⊕⊕◯ MODERATE [a,d]	
	7 per 1,000	**9 per 1,000** (4 to 20)				
Minor bleeding follow up: 12 weeks	**Moderate**		**RR 1.05** (0.75 to 1.46)	2345 (7 RCTs)	⊕⊕⊕◯ MODERATE [a,d]	
	27 per 1,000	**28 per 1,000** (20 to 39)				
Health related quality of life assessed with: the Uniscale and the Symptom Distress Scale (SDS); Better indicated by lower values follow up: 12 months	0 per 1,000	**0 per 1,000** (0 to 0)	not estimable	(1 RCT)	⊕⊕◯◯ LOW [e]	

***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

https://sites.google.com/site/gradeworkshopsba2016/tokyo2016

Akl EA et al. Cochrane Database Syst Rev. 2011