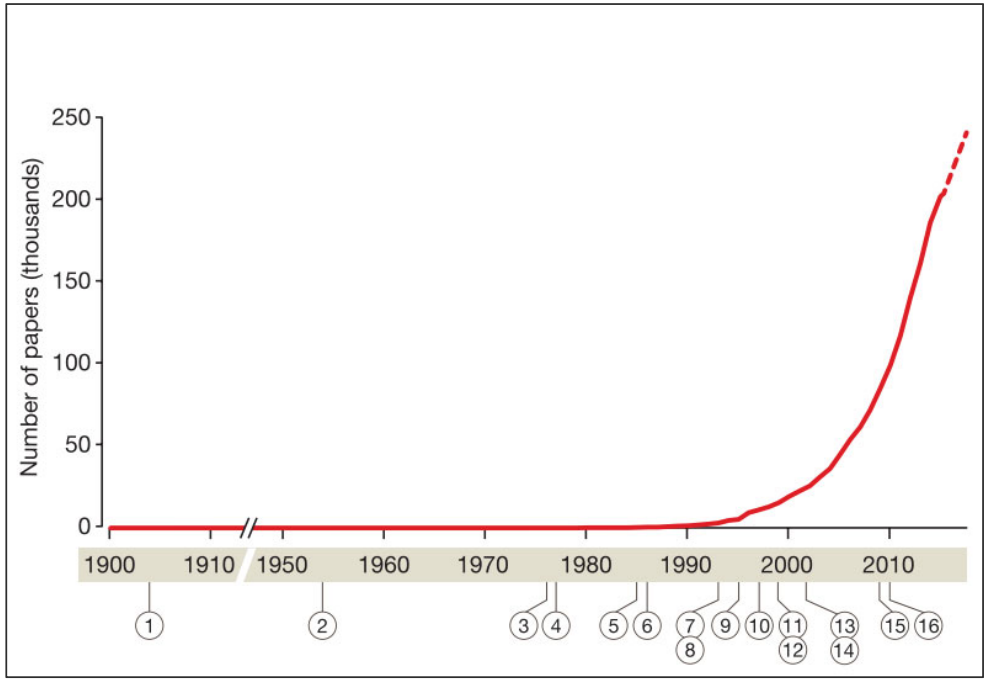
最近のNMAのhot topics

東京大学大学院　情報学環
（兼）医学系研究科　生物統計学分野
准教授　大庭　幸治
oba@epistat.m.u-tokyo.ac.jp

メタアナリシスの歴史

- 医学分野では90年代以降、メタアナリシスを実施した論文が急増
  - EBM(Evidence Based Medicine)の台頭



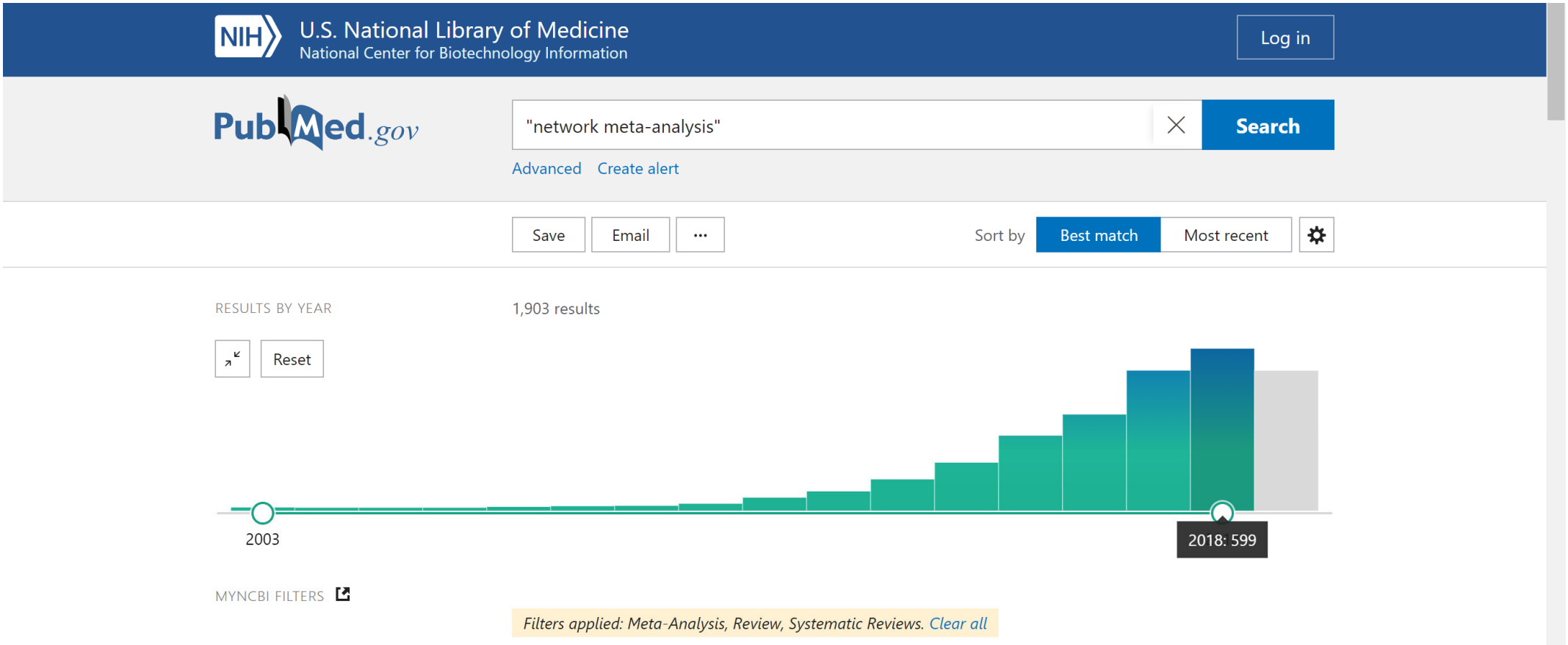
① **1904** First (medical) meta-analysis published (effect of innoculation against typhoid) (ref. 83)
② **1954** First meta-analytic methods formalized (fixed- and random-effects models) (ref. 86)
③ **1976** Term 'meta-analysis' coined (ref. 95)
④ **1977** First social science meta-analysis published (efficacy of psychotherapy) (ref. 87)
⑤ **1985** Statistics textbook dedicated to meta-analytic methods released (ref. 16)
⑥ **1986** Method for calculating between-study variance developed (ref. 96)
⑦ **1993** Review of 302 social science meta-analyses on treatment efficacy published (ref. 97)
⑧ **1993** Cochrane Collaboration established
⑨ **1995** Term 'systematic review' introduced (ref. 98)
⑩ **1997** Methods for assessing publication bias introduced (funnel plot and Egger's test) (ref. 19)
⑪ **1999** QUOROM (Quality of Reporting of Meta-analyses) standards developed (ref. 99)

Network meta-analysisという言葉の出現

⑭ **2002** Term 'network meta-analysis' coined (ref. 74)
⑮ **2009** PRISMA guidelines established (ref. 12)
⑯ **2010** *metafor* (free and comprehensive R package for meta-analysis) released (ref. 17)

Gurevitch et al, Nature 2018

NMAの報告も年々増加



NMAという名前でなくても、その概念は昔から

つい先日教えてもらったので個人的にHotな論文

Bridging Different Eras in Sports

Scott M. BERRY, C. Shane REESE, and Patrick D. LARKEY

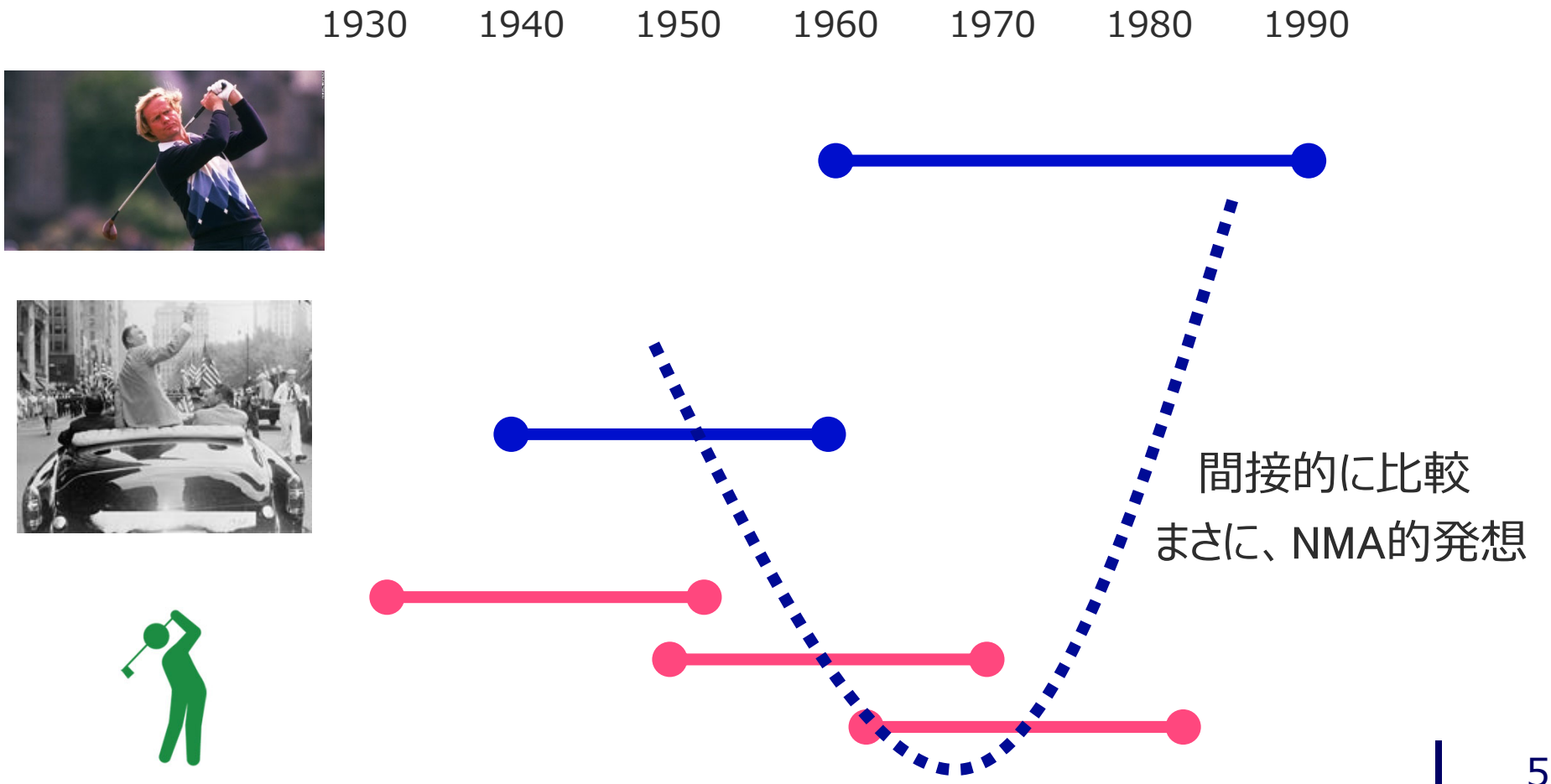
This article addresses the problem of comparing abilities of players from different eras in professional sports. We study National Hockey League players, professional golfers, and Major League Baseball players from the perspectives of home run hitting and hitting for average. Within each sport, the careers of the players overlap to some extent. This network of overlaps, or bridges, is used to compare players whose careers took place in different eras. The goal is not to judge players relative to their contemporaries, but rather to compare all players directly. Hence the model that we use is a statistical time machine. We use additive models to estimate the innate ability of players, the effects of aging on performance, and the relative difficulty of each year within a sport. We measure each of these effects separated from the others. We use hiera ers and specify separate distributions for each decade, thus allowing the "taler the changing talent pool in each sport and address Gould's conjecture about the way in which populations change. Nonparametric aging functions allow us to estimate the league-wide average aging function. Hierarchical random curves allow for individuals to age differently from the average of athletes in that sport. We characterize players by their career profile rather than a one-number summary of their career.

Statistical time machine

KEY WORDS: Aging function; Bridge model; Hierarchical model; Population dynamics; Random curve.

Berry SM, et al. JASA 1999

過去の選手の中で最も優れた選手は？



統計モデル

- 4大トーナメントのスコアに対してベイズ流階層モデルを利用
- 選手の各ラウンドのスコアを単純比較はできない
  - 各選手のピーク時に1997年の平均難易度におけるスコアを比較
- 考慮すべき要因を踏まえて各選手のスコアをモデル化
  - 年齢の影響（ピーク時の評価）
  - 年ごとのコースの難易度
  - 年の中でのコースの難易度のバラツキ
- 更に、スコアを表すパラメータに対して、時代（道具などの影響）ごとに平均の異なる事前分布を考慮

Table 8. The Top 25 Peak Players in the Golf Study

Rank	Name	Born	θ	ψ_1	ψ_2
1	J. Nicklaus	1940	70.42 (.29)	1.03	.99
2	T. Watson	1949	70.82 (.23)	.92	1.19
3	B. Hogan	1912	71.12 (.29)	1.13	.27
4	N. Faldo	1957	71.19 (.21)	1.19	1.21
5	A. Palmer	1929	71.33 (.28)	1.19	.95
6	G. Norman	1955	71.39 (.19)	1.21	.64
7	J. Leonard	1972	71.40 (.45)	.68	1
8	E. Els	1969	71.45 (.34)	.78	1
9	G. Player	1935	71.45 (.23)	.87	.62
10	F. Couples	1959	71.50 (.21)	1.00	.97
11	H. Irwin	1945	71.56 (.26)	1.02	.68
12	C. Peete	1943	71.56 (.36)	1	.80
13	J. Boros	1920	71.62 (.37)	1	.61
14	R. Floyd	1942	71.63 (.24)	1.22	.38
15	L. Trevino	1939	71.63 (.29)	1.00	.72
16	S. Snead	1912	71.64 (.27)	1.10	.21
17	J. Olazabal	1966	71.69 (.39)	.74	1
18	T. Kite	1949	71.71 (.23)	.98	.70
19	B. Crenshaw	1952	71.74 (.22)	.43	1.22
20	T. Woods	1975	71.77 (.64)	.52	1
21	B. Casper	1931	71.77 (.26)	1.00	1.09
22	B. Nelson	1912	71.78 (.31)	1.00	1.11
23	P. Mickelson	1970	71.79 (.44)	.79	1
24	L. Wadkins	1949	71.79 (.22)	1.13	.78
25	T. Lehman	1959	71.82 (.30)	1.05	.79

NOTE: The standard deviations are in parentheses. The means of ψ_1 and ψ_2 are also presented.

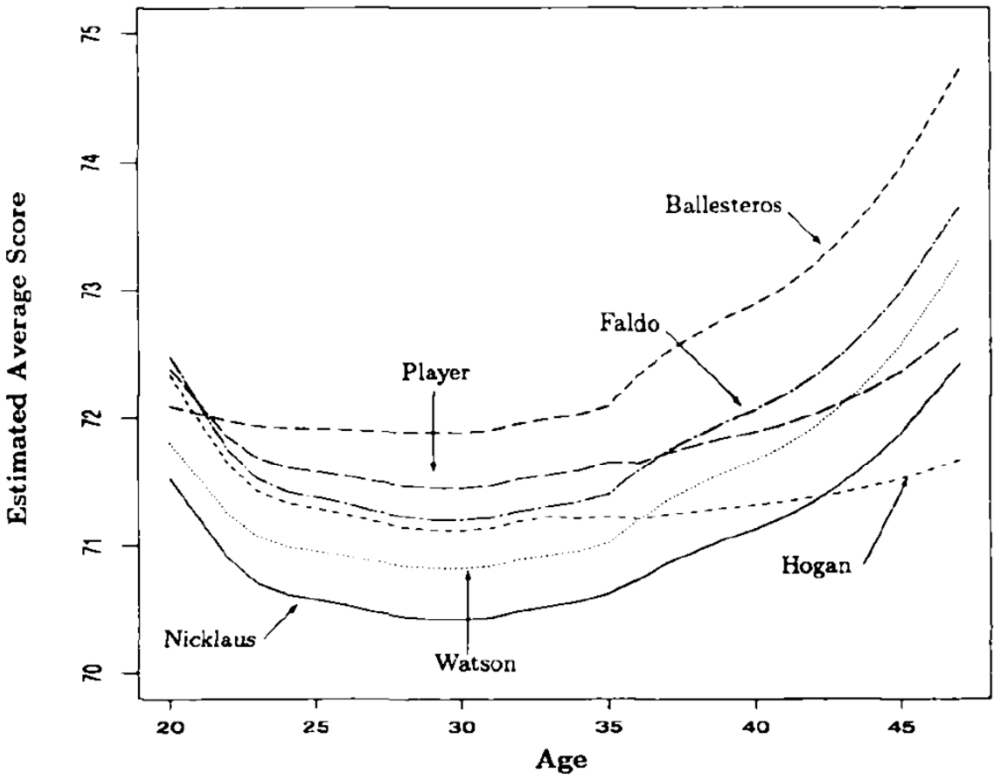


Figure 9. A Profile of Some of the Best Players in the Golf Study. The estimated mean score for each age of the player, if that round were an average 1997 round.

Berry SM, et al. JASA 1999

この結果、あなたは信用できますか？

- 信用できる？
- 信用できない？

- 情報が足りないと思った方も多いと思います
- どのような情報があれば信用できるでしょうか？

自己紹介も兼ねて…

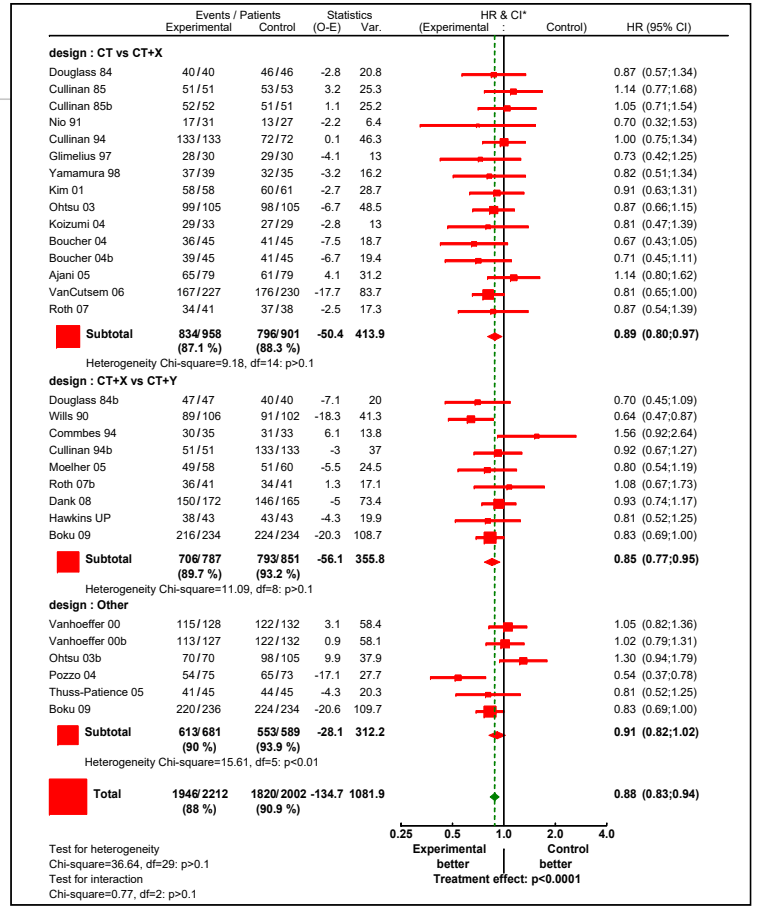
The GASTRIC

Global Advanced/Adjuvant Stomach Tumor Research through International Collaboration

Japanese

Welcome

The Global Advanced/Adjuvant Stomach Tumor Research through International Collaboration (GASTRIC) is an academic, worldwide project conducting individual patient data based meta-analyses of randomized trials testing post-operative adjuvant chemotherapy for resectable gastric cancer, or chemotherapy for advanced/recurrent gastric cancer. The GASTRIC 1st round started in 2006 and published 2 results about efficacy of chemotherapy for curatively resected gastric cancer and advanced/recurrent gastric cancer and 2 results about the evaluation of surrogate endpoints (disease free survival or progression free survival) against overall survival. Now, the GASTRIC 2nd round has started.



	Events / Patients Experimental	Control	Statistics (O-E)	Var.	HR & CI* (Experimental : Control)	HR (95% CI)
design : CT vs CT+X						
Douglass 84	40/40	46/46	-2.8	20.8		0.87 (0.57;1.34)
Cullinan 85	51/51	53/53	3.2	25.3		1.14 (0.77;1.68)
Cullinan 85b	52/52	51/51	1.1	25.2		1.05 (0.71;1.54)
Nio 91	17/31	13/27	-2.2	6.4		0.70 (0.32;1.53)
Cullinan 94	133/133	72/72	0.1	46.3		1.00 (0.75;1.34)
Glimelius 97	28/30	29/30	-4.1	13		0.73 (0.42;1.25)
Yamamura 98	37/39	32/35	-3.2	16.2		0.82 (0.51;1.34)
Kim 01	58/58	60/61	-2.7	28.7		0.91 (0.63;1.31)
Ohtsu 03	99/105	98/105	-6.7	48.5		0.87 (0.66;1.15)
Koizumi 04	29/33	27/29	-2.8	13		0.81 (0.47;1.39)
Boucher 04	36/45	41/45	-7.5	18.7		0.67 (0.43;1.05)
Boucher 04b	39/45	41/45	-6.7	19.4		0.71 (0.45;1.11)
Ajani 05	65/79	61/79	4.1	31.2		1.14 (0.80;1.62)
VanCutsem 06	167/227	176/230	-17.7	83.7		0.81 (0.65;1.00)
Roth 07	34/41	37/38	-2.5	17.3		0.87 (0.54;1.39)
Subtotal	**834/958 (87.1 %)**	**796/901 (88.3 %)**	**-50.4**	**413.9**		**0.89 (0.80;0.97)**
Heterogeneity Chi-square=9.18, df=14: p>0.1						
design : CT+X vs CT+Y						
Douglass 84b	47/47	40/40	-7.1	20		0.70 (0.45;1.09)
Wills 90	89/106	91/102	-18.3	41.3		0.64 (0.47;0.87)
Commbes 94	30/35	31/33	6.1	13.8		1.56 (0.92;2.64)
Cullinan 94b	51/51	133/133	-3	37		0.92 (0.67;1.27)
Moelher 05	49/58	51/60	-5.5	24.5		0.80 (0.54;1.19)
Roth 07b	36/41	34/41	1.3	17.1		1.08 (0.67;1.73)
Dank 08	150/172	146/165	-5	73.4		0.93 (0.74;1.17)
Hawkins UP	38/43	43/43	-4.3	19.9		0.81 (0.52;1.25)
Boku 09	216/234	224/234	-20.3	108.7		0.83 (0.69;1.00)
Subtotal	**706/787 (89.7 %)**	**793/851 (93.2 %)**	**-56.1**	**355.8**		**0.85 (0.77;0.95)**
Heterogeneity Chi-square=11.09, df=8: p>0.1						
design : Other						
Vanhoeffer 00	115/128	122/132	3.1	58.4		1.05 (0.82;1.36)
Vanhoeffer 00b	113/127	122/132	0.9	58.1		1.02 (0.79;1.31)
Ohtsu 03b	70/70	98/105	9.9	37.9		1.30 (0.94;1.79)
Pozzo 04	54/75	65/73	-17.1	27.7		0.54 (0.37;0.78)
Thuss-Patience 05	41/45	44/45	-4.3	20.3		0.81 (0.52;1.25)
Boku 09	220/236	224/234	-20.6	109.7		0.83 (0.69;1.00)
Subtotal	**613/681 (90 %)**	**553/589 (93.9 %)**	**-28.1**	**312.2**		**0.91 (0.82;1.02)**
Heterogeneity Chi-square=15.61, df=5: p<0.01						
Total	**1946/2212 (88 %)**	**1820/2002 (90.9 %)**	**-134.7**	**1081.9**		**0.88 (0.83;0.94)**

0.25 0.5 1.0 2.0 4.0

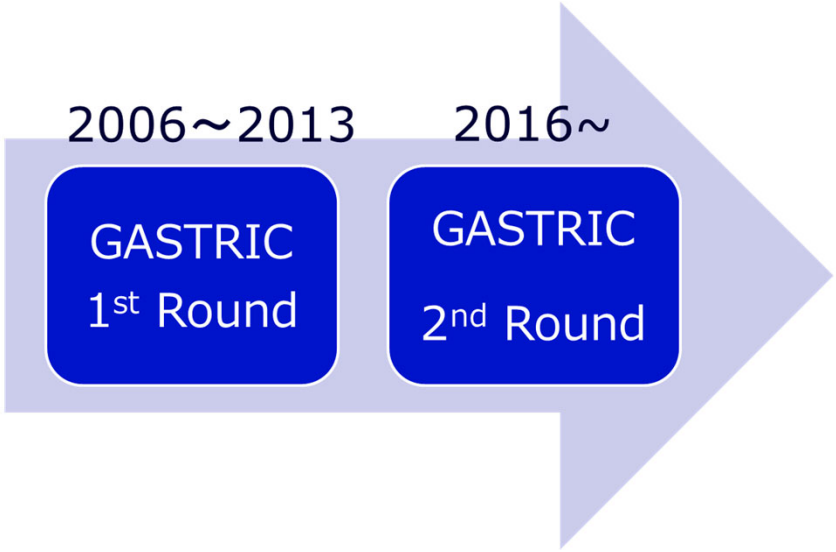
Experimental better | Control better

Treatment effect: p<0.0001

Test for heterogeneity
Chi-square=36.64, df=29: p>0.1
Test for interaction
Chi-square=0.77, df=2: p>0.1

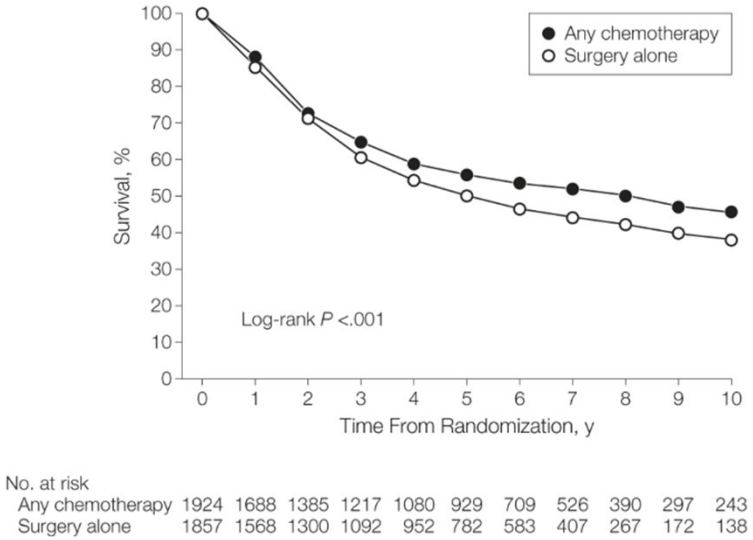
The GASTRIC project

- **G**lobal **A**dvanced/Adjuvant **S**tomach **T**umor **R**esearch through **I**nternational Collaboration (GASTRIC)
- 根治術後胃がんを対象とした術後化学療法を評価したランダム化比較試験、進行／再発胃がんを対象とした化学療法を評価したランダム化比較試験の個人データに基づくメタアナリシスを行うアカデミア主導の国際プロジェクト

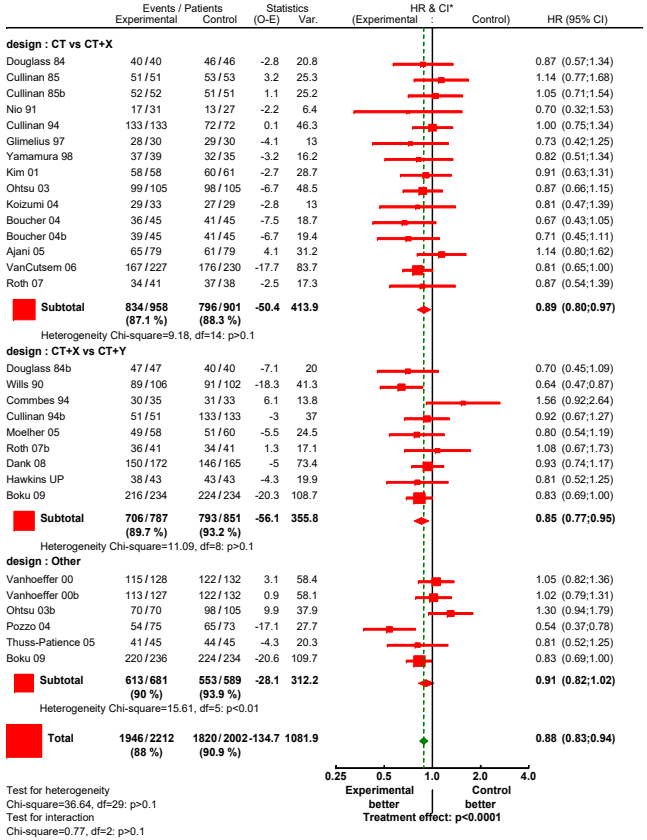


1st Roundの成果

- 化学療法の有効性に関するメタアナリシス
  - 術後胃がんに対する術後化学療法
  - 進行・再発胃がんに対する化学療法



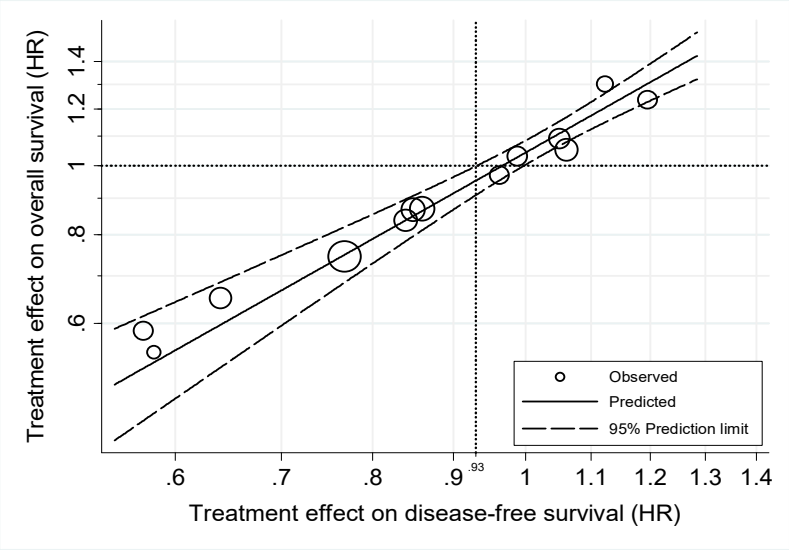
JAMA 2010



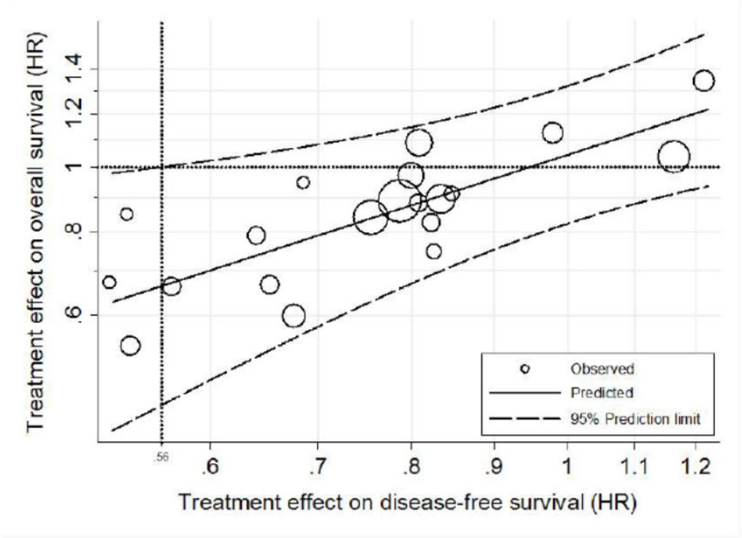
Eur J Cancer 2012

1st Roundの成果

- 真のエンドポイント(全生存期間)に対する代替エンドポイントの評価
  - 術後胃がんでの無病生存期間
  - 進行・再発胃がんでの無増悪生存期間



JNCI 2013



JNCI 2013

2nd Roundでの検討予定の事項

- 進行・再発胃がんに限定
  - 術後胃がんについては、研究テーマが曖昧だったため
- 2nd Roundでの研究テーマ
  - 化学療法の有効性・安全性に関するメタアナリシス
  - RECIST基準で評価されたランダム化比較試験における無増悪生存期間、腫瘍縮小効果（奏効率）の全生存に対する代替性評価
  - 分子標的薬の有効性に関するメタアナリシスと効果予測因子の検討
  - NMAを用いたS-1、Capecitabine、5FUの非劣性の検討
  - …

2nd Roundでの検討予定の事項

- 進行・再発胃がんに限定
  - 術後胃がんについては、研究テーマが曖昧だったため
- 2nd Roundでの研究テーマ
  - 化学療法の有効性・安全性に関するメタアナリシス
  - RECIST基準で評価されたランダム化比較試験における無増悪生存期間、腫瘍縮小効果（奏効率）の全生存に対する代替性評価
  - 分子標的薬の有効性に関するメタアナリシスと効果予測因子の検討
  - NMAを用いたS-1、Capecitabine、5FUの非劣性の検討
  - …

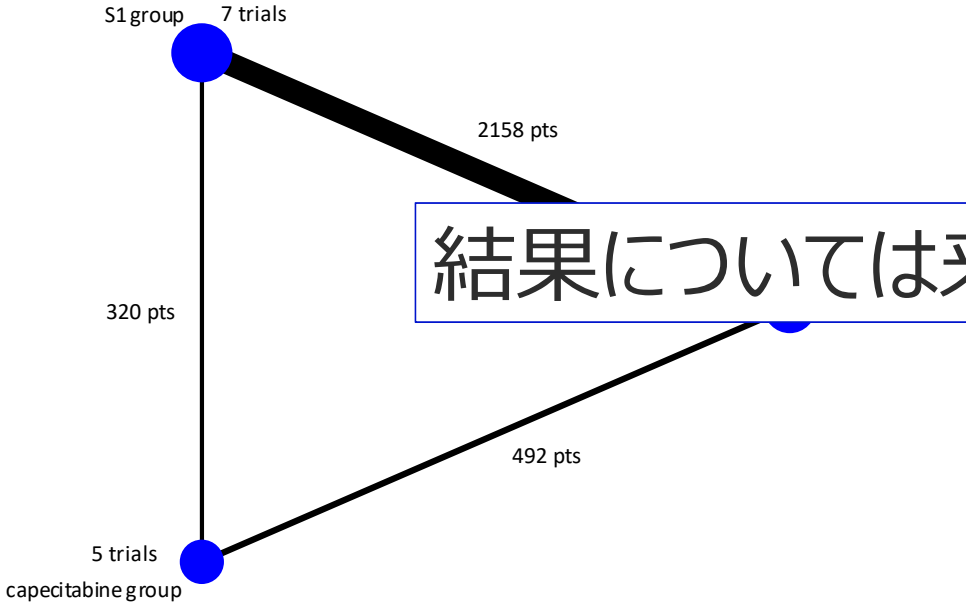
進行・再発胃癌治療の歴史

- 5FU系薬剤がキードラッグだった時代が長く続いた
  - 特にアジア
- 日本でS1が開発され標準治療になった一方、同じくフッ化ピリミジン系の薬剤であるCapecitabineが開発され、こちらも使用されるように
- S1とCapecitabineの直接比較があまり行われないまま、それぞれをベースにした併用療法が開発され、今に至る
  - FOLFOX療法、CAPOX療法、SOX療法…

これからわざわざ前向き試験をやる人・企業はおそらくないのでは…

5FU、S1、Capecitabineに関連する比較試験

- 単剤、上乗せ効果を検討した試験のネットワークプロット



結果については来年発表予定

No	Author	year	N	Ph	Arm 1	Arm 2
1	Lee JL	2008	96	2	Capecitabine	S-1
2	Kang YK	2009	316	3	Capecitabine /cisplatin	5-FU /cisplatin
3	Boku N	2009	468	3	Fluorouracil	S-1
				2	CapeOX	SOX
				2	5FU/PTX	S1/PTX
6	Tsuburaya A	2012	100	2	Capecitabine /cisplatin	S-1 /cisplatin
7	Huang D	2013	229	2	Paclitaxel /S-1	Paclitaxel /5-FU
8	Ajani JA	2013	1053	3	S-1 /Cisplatin	5-FU /Cisplatin
9	VanCutsem E	2015	175	2	docetaxel/ oxaliplatin/ 5-FU	docetaxel/ oxaliplatin/ capecitabine
	Total		2794			

この検討、あなたは信用できますか？

- 一応、非劣性マージンも事前に規定（後づけではない形で検討）
  - ハザード比で1.25

- 信用できる？
- 信用できない？

- どのような情報があれば信用できるでしょうか？

NMAの活用・解釈に当たって

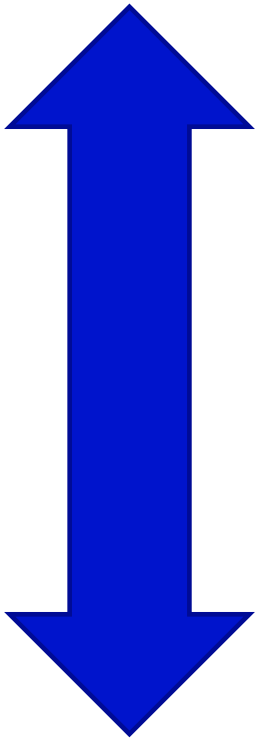
- 統計的な方法論の整備は進み、様々な場面で活用する段階に

- もし、信用できないと感じるとしたら、どんな点が気になるのか
  - その感覚を定量化して、どのように解析で考慮されているか
- 仮定されていることがあれば、その仮定が崩れた時にどのような影響があるか
- 様々な専門の目で確認することが重要

- 今日のセミナーでは、NMAにおける重要な仮定や特性を理解できる
  - 実際どうやっているのか（デモ）も経験できる

様々なレベルでの利用例

- 利用可能な全ての薬剤もしくは関連する薬剤について、有効性・安全性を系統的に比較、医療経済評価
  - 薬物治療ガイドライン作成
- 不足したエビデンスを補うための探索的意味合いの強いNMA
  - 実施する試験の対照群の検討



エビデンスの充実度

PRISMA NMA, Ann Int Med 2015

2型糖尿病患者に対する糖尿病治療薬の事例

Research

JAMA | Original Investigation

Association Between Use of Sodium-Glucose Cotransporter 2 Inhibitors, Glucagon-like Peptide 1 Agonists, and Dipeptidyl Peptidase 4 Inhibitors With All-Cause Mortality in Patients With Type 2 Diabetes

A Systematic Review and Meta-analysis

Sean L. Zheng, BM BCh, MA, MRCP; Allstair J. Roddick, BSc; Rochan Aghar-Jaffar, BMedSci, BMBS, MRCP; Matthew J. Shun-Shin, BM BCh, MRCP; Darrel Francis, MB BChir, FRCP, MD; Nick Oliver, MBBS, FRCP; Karim Meeran, MBBS, MD, FRCP, FRCPath

IMPORTANCE The comparative clinical efficacy of sodium-glucose cotransporter 2 (SGLT-2) inhibitors, glucagon-like peptide 1 (GLP-1) agonists, and dipeptidyl peptidase 4 (DPP-4) inhibitors for treatment of type 2 diabetes is unknown.

OBJECTIVE To compare the efficacies of SGLT-2 inhibitors, GLP-1 agonists, and DPP-4 inhibitors on mortality and cardiovascular end points using network meta-analysis.

DATA SOURCES MEDLINE, Embase, Cochrane Library Central Register of Controlled Trials, and published meta-analyses from inception through October 11, 2017.

STUDY SELECTION Randomized clinical trials enrolling participants with type 2 diabetes and a follow-up of at least 12 weeks were included, for which SGLT-2 inhibitors, GLP-1 agonists, and DPP-4 inhibitors were compared with either each other or placebo or no treatment.

DATA EXTRACTION AND SYNTHESIS Data were screened by 1 investigator and extracted in duplicate by 2 investigators. A Bayesian hierarchical network meta-analysis was performed.

MAIN OUTCOMES AND MEASURES The primary outcome: all-cause mortality; secondary outcomes: cardiovascular (CV) mortality, heart failure (HF) events, myocardial infarction (MI), unstable angina, and stroke; safety end points: adverse events and hypoglycemia.

RESULTS This network meta-analysis of 236 trials randomizing 176 310 participants found SGLT-2 inhibitors (absolute risk difference [RD], −1.0%; hazard ratio [HR], 0.80 [95% credible interval {CrI}, 0.71 to 0.89]) and GLP-1 agonists (absolute RD, −0.6%; HR, 0.88 [95% CrI, 0.81 to 0.94]) were associated with significantly lower all-cause mortality than the control groups. SGLT-2 inhibitors (absolute RD, −0.9%; HR, 0.78 [95% CrI, 0.68 to 0.90]) and GLP-1 agonists (absolute RD, −0.5%; HR, 0.86 [95% CrI, 0.77 to 0.96]) were associated with lower mortality than were DPP-4 inhibitors. DPP-4 inhibitors were not significantly associated with lower all-cause mortality (absolute RD, 0.1%; HR, 1.02 [95% CrI, 0.94 to 1.11]) than were the control groups. SGLT-2 inhibitors (absolute RD, −0.8%; HR, 0.79 [95% CrI, 0.69 to 0.91]) and GLP-1 agonists (absolute RD, −0.5%; HR, 0.85 [95% CrI, 0.77 to 0.94]) were significantly associated with lower CV mortality than were the control groups. SGLT-2 inhibitors were significantly associated with lower rates of HF events (absolute RD, −1.1%; HR, 0.62 [95% CrI, 0.54 to 0.72]) and MI (absolute RD, −0.6%; HR, 0.86 [95% CrI, 0.77 to 0.97]) than were the control groups. GLP-1 agonists were associated with a higher risk of adverse events leading to trial withdrawal than were SGLT-2 inhibitors (absolute RD, 5.8%; HR, 1.80 [95% CrI, 1.44 to 2.25]) and DPP-4 inhibitors (absolute RD, 3.1%; HR, 1.93 [95% CrI, 1.59 to 2.35]).

CONCLUSIONS AND RELEVANCE In this network meta-analysis, the use of SGLT-2 inhibitors or GLP-1 agonists was associated with lower mortality than DPP-4 inhibitors or placebo or no treatment. Use of DPP-4 inhibitors was not associated with lower mortality than placebo or no treatment.

JAMA. 2018;319(15):1580-1591. doi:10.1001/jama.2018.3024

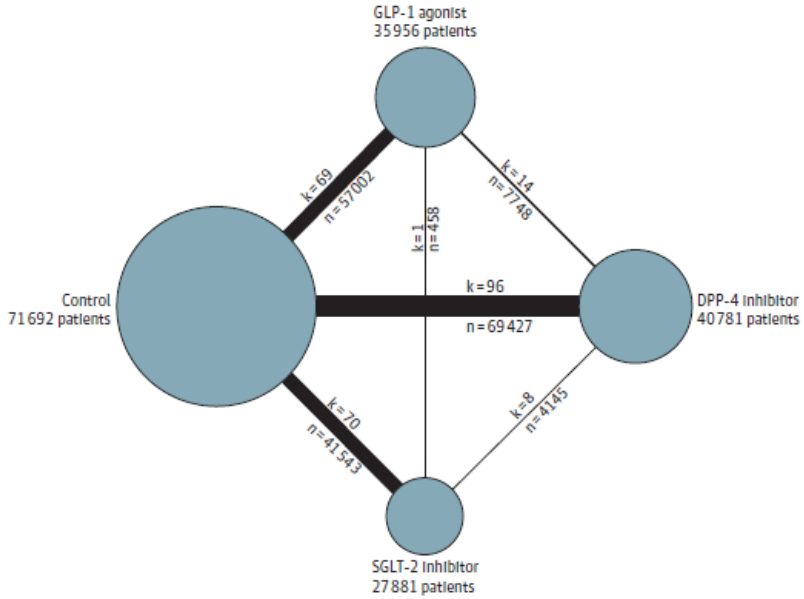
Animated Summary Video

Supplemental content

Author Affiliations: Department of Endocrinology, Imperial College Healthcare NHS Foundation Trust, London, United Kingdom (Zheng, Aghar-Jaffar, Oliver, Meeran); Department of Cardiology, Royal Brompton and Harefield NHS Foundation Trust, London, United Kingdom (Zheng); Imperial College London, United Kingdom (Zheng, Aghar-Jaffar, Shun-Shin, Francis, Oliver, Meeran); Faculty of Life Sciences and Medicine, King's College London, United Kingdom (Roddick); Division of Diabetes, Endocrinology and Metabolism, Imperial College London, United Kingdom (Aghar-Jaffar, Oliver, Meeran).

Corresponding Author: Sean L. Zheng, BM BCh, MA, MRCP, Department of Cardiology, Royal Brompton Hospital, Sydney Street, London, UK SW3 6NP (sean.zheng@nhs.net).

Figure 2. Network Plot for All Studies



Graphical representation of network for all included trials. Connecting lines represent head-to-head comparisons between drugs, indicated by nodes. Multigroup trials contribute multiple comparisons, resulting in 258 comparisons from 236 trials. The thickness of lines between nodes is proportional to the number of trials comparing the treatments. The sizes of the nodes are proportional to the number of patients in each treatment. Patients may be included in multiple comparisons: for example, in a study of 3 groups consisting of control and 2 different drug classes, the control group is compared with each drug class. This is accounted for within the network model and does not constitute duplication of participants.

DPP-4 indicates dipeptidyl peptidase 4 GLP-1, glucagon-like peptide 1; k, the number of comparisons; n, the number of patients per comparison; and SGLT-2, sodium-glucose cotransporter 2.

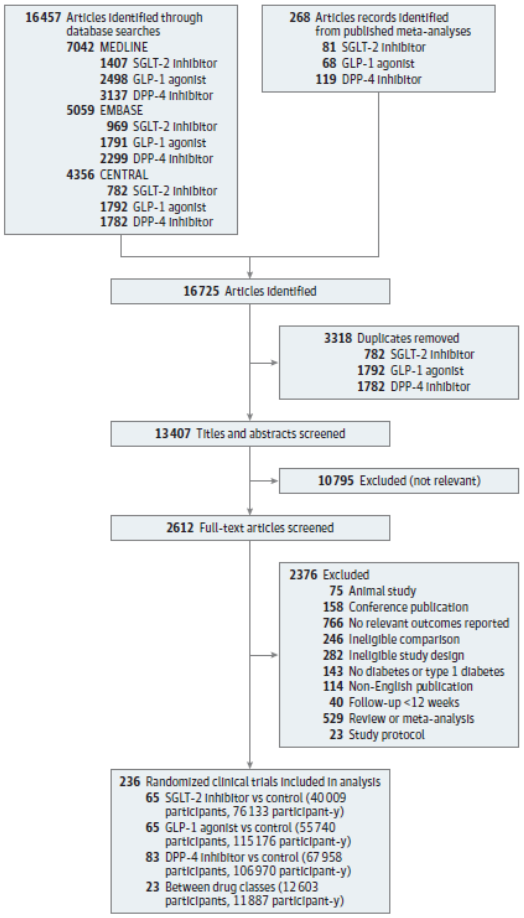
Zheng SL et al, JAMA 2018

JAMA論文における研究目的

- 2型糖尿病患者を対象に
- SGLT-2阻害薬、GLP-1受容体作動薬、DPP-4阻害薬、コントロール治療の間で
- 主要エンドポイント：全生存期間（総死亡）
- 副次エンドポイント：循環器死亡、心不全、心筋梗塞、不安定狭心症、脳卒中
- 安全性エンドポイント：有害事象、低血糖
- ネットワークメタ解析の手法を用いて、比較を行う

対象となった試験のフローダイアグラム

Figure 1. Summary of Study Retrieval and Identification for Network Meta-analysis



DPP-4 indicates dipeptidyl peptidase 4; GLP-1, glucagon-like peptide 1; and SGLT-2, sodium-glucose cotransporter 2.

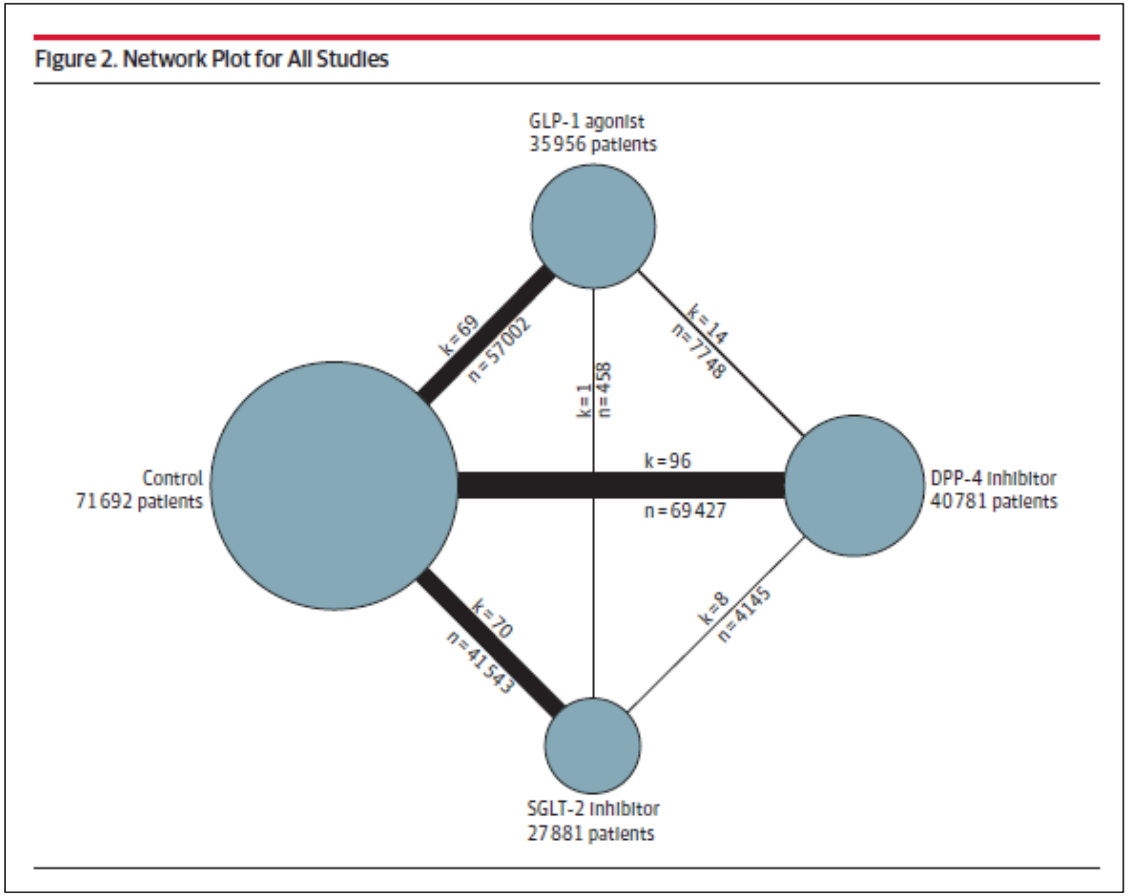
対象試験

- プラセボ・無治療とSGLT-2、GLP-1、DPP-4を比較したランダム化比較試験
- 最低12か月以上の追跡
- 目的としたエンドポイントを評価

検索の正確性・再現性の担保

最終的に236試験が解析対象

ネットワークプロット



- 対象となった試験の関係性をグラフ化
- ノード（集合点：各治療群）と
  エッジ（辺：直接比較の存在）で
  構成されるプロット

Statistical analysis（抜粋）

- Network meta-analysis comprises direct and indirect comparisons between multiple interventions, allowing comparisons to be made when direct trial evidence is scarce (乏しい).
  - This approach respects randomization but does not represent randomized evidence.
- A Bayesian hierarchical network meta-analysis was performed … Fixed- or random-effects models were selected for each outcome based on the deviance information criterion (DIC).
- Analyses were performed using Markov-chain Monte Carlo methods. Results were presented as hazard ratios (HRs) with 95% credible intervals (CrIs).

残りの演者のパートで説明されます

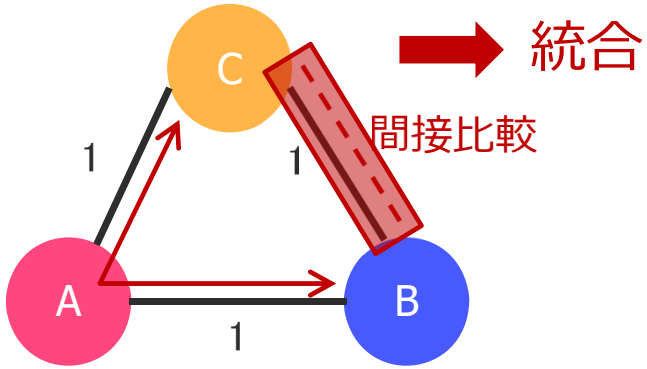
ネットワークメタ解析の1つの肝

- 直接比較(direct comparison)の統合
  - 通常のメタ解析で実施される、同じ比較を行っているRCTの統合
  - 網羅的に論文が把握され、それぞれの試験の質が良く、
    結果が類似していれば、統合結果で得られるエビデンスの説得力は強い

- 直接比較と間接比較(indirect comparison)の統合
  - ネットワークメタアナリシスにおける肝となる部分
- 統合が可能と考えられる前提とは？
  - 網羅的に論文が把握され、それぞれの試験の質が良く、
    直接比較の結果と間接比較の結果が類似していれば、統合結果で得られる
    エビデンスの説得力は強い
  - consistency assumption (一致性の仮定)

一致性の仮定

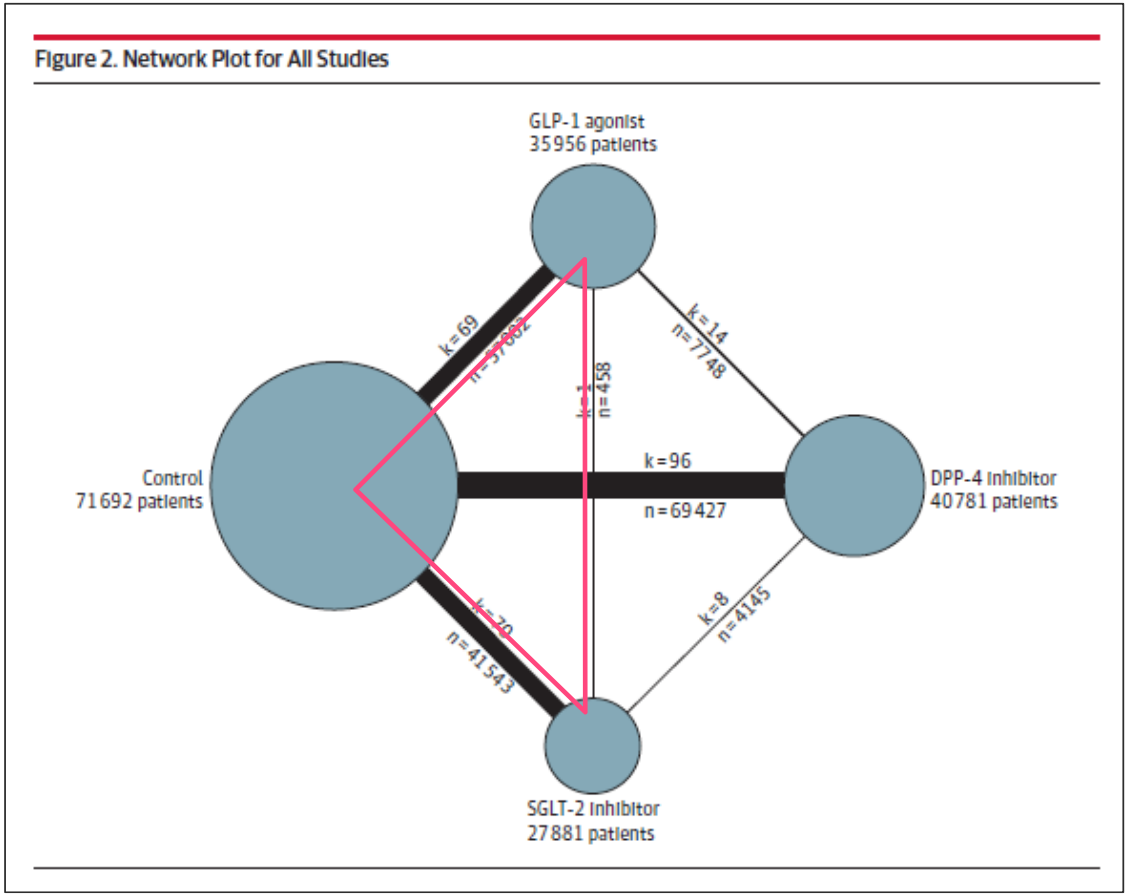
- ネットワークの全てのパスにおいて、直接比較と間接比較の効果の大きさが一致する(consistentである)こと



B vs Cの結果：ハザード比=0.80
A vs C, A vs Bからの間接的な結果: ハザード比=0.75
➡ 結果的に整合している

B vs Cの結果：ハザード比=0.80
A vs C, A vs Bからの間接的な結果: ハザード比=2.00
➡ 結果的に不一致 Inconsistency が見られる
➡ ネットワーク上の比較の妥当性が失われているのでは？

一致性の評価



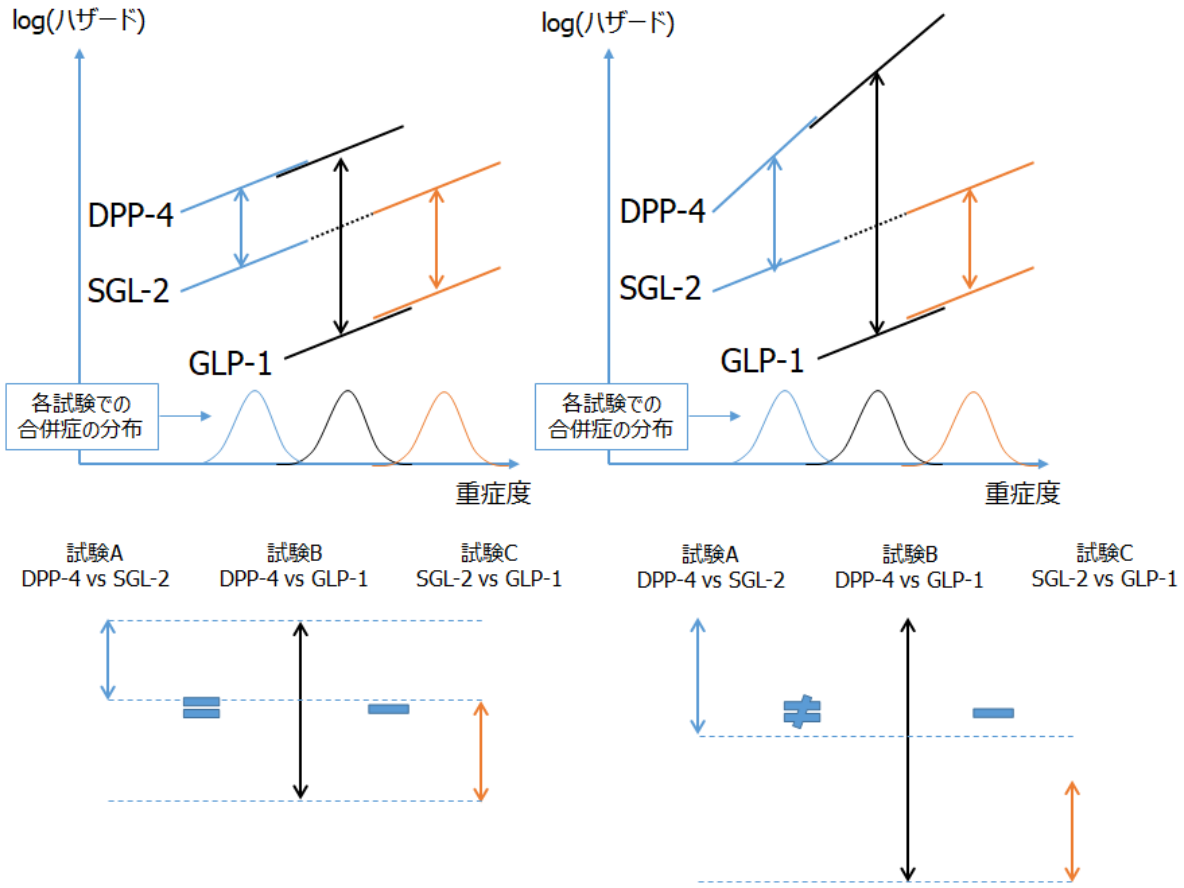
- ネットワーク上の特定のTriangleに対して直接比較と間接比較がconsistentか？
- 局所的な一致性の程度は、

$$\frac{\text{直接比較の効果} - \text{間接比較の効果}}{\sqrt{\text{直接比較の分散} + \text{間接比較の分散}}} \sim N(1,0)$$

のような検定で評価することが可能

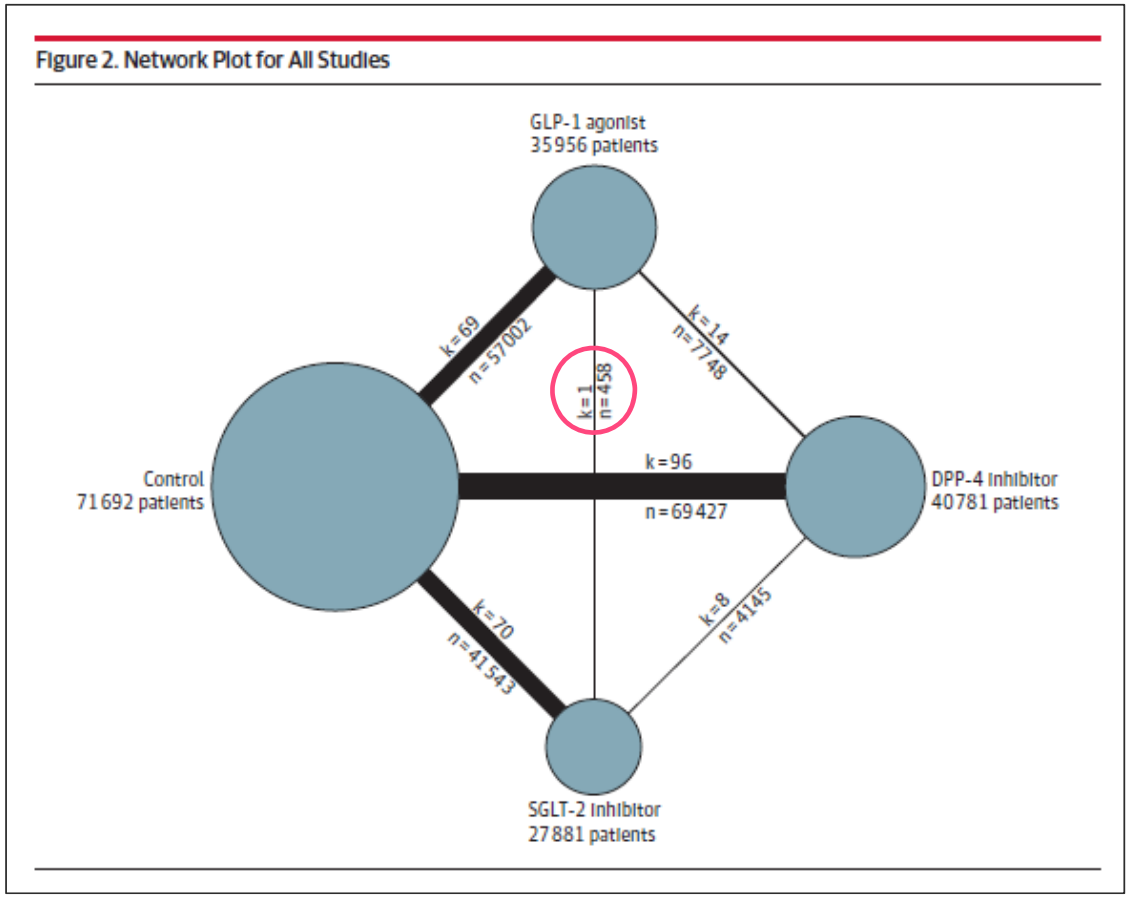
- ネットワークメタ解析は、ネットワーク上の全ての統合に一致性を仮定しているので、1つでも崩れると妥当性が成り立たない
  - ネットワーク全体の一致性も評価

どういった時にinconsistencyが生じる？

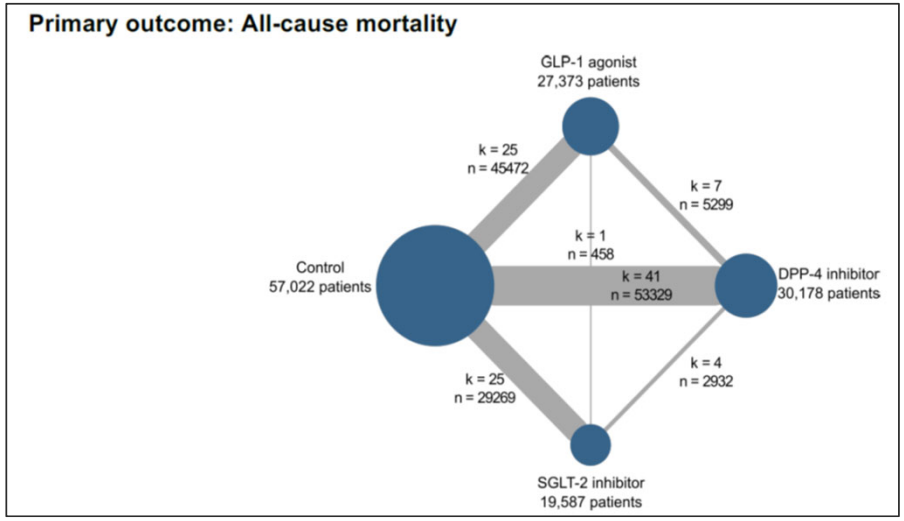


- 治療との交互作用を示すような因子が試験間で異なる場合
  - 左図の状況
- DPP-4が重症度が高くなれば、心不全リスクが、他の薬よりも、より起きやすくなる
  - 直接比較と間接比較の結果は系統的にズレる
- 交互作用を起こしうる要因
  - 個人レベルの特性: 年齢、疾患重症度、合併症など
  - 試験レベルの特性: 治療計画、併用薬、治療環境、時代効果、効果指標の選択、など

結果の確認：薬剤クラス間の比較



- クラス間比較については、十分な数の直接比較が存在
  - ただし、SGLT-2とGLP-1の直接比較は1試験のみ
- アウトカムよってプロットは変わることに注意



Zheng SL et al, JAMA 2018: Figure2, eFigure3

結果の確認：直接比較における背景

Table. Study Participant Characteristics[a]

Drug Type	No. of Trials	Total No. Randomized	Mean (SD)			
			Men, %	Age, y	BMI	HbA_{1c}, %
DPP-4 inhibitor vs control	83	67 958	54.7 (9.4)	57.9 (5.3)	29.3 (2.9)	8.16 (0.61)
GLP-1 agonist vs control	65	55 740	55.1 (11.4)	57.1 (3.8)	31.5 (3.5)	8.11 (0.36)
SGLT-2 inhibitor vs control	65	40 009	57.9 (10.4)	58.0 (3.7)	29.3 (5.0)	8.05 (0.32)
DPP-4 inhibitor vs GLP-1 agonist	14	8024	50.9 (7.5)	52.9 (4.4)	32.6 (2.3)	8.2 (0.20)
DPP-4 inhibitor vs SGLT-2 inhibitor	8	4121	56.0 (5.5)	55.5 (2.1)	30.9 (1.2)	8.0 (0.39)
GLP-1 agonist vs SGLT-2 inhibitor	1	458				

Abbreviations: BMI, body mass index, calculated as weight in kilograms divided by height in meters squared; DPP-4, dipeptidyl peptidase 4; GLP-1, glucagon-like peptide 1; HbA_{1c}, hemoglobin A_{1c}; SLGT2, sodium-glucose cotransporter 2.

[a] The Table represents data from studies stratified by the intervention and comparator. Control refers to placebo or no treatment. There was 1 study assessing GLP-1 agonist compared with a SGLT-2 inhibitor.

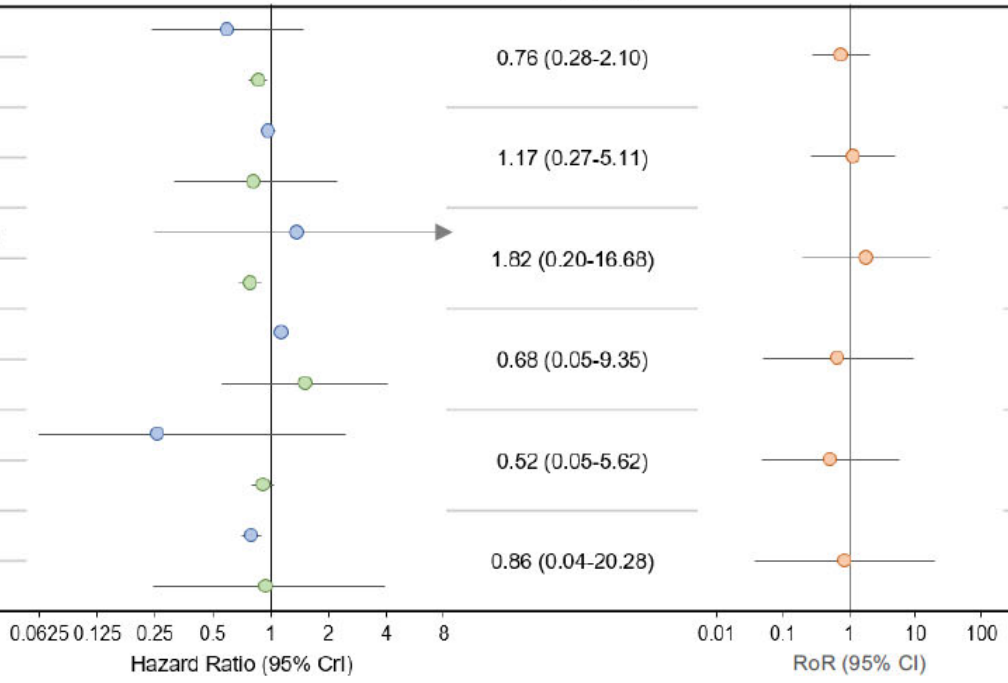
DPP-4 vs SGLT-2の比較だけ、少し集団が異なるが、概ね類似

Zheng SL et al, JAMA 2018: Table

結果の確認：全死亡について、一致性の確認

All-cause mortality

Comparison	Events	Trials	Evidence	HR (95% CrI)	RoR (95% CI)	P-value
DPP4:GLP1	20	7	Direct	0.59 (0.24-1.47)	0.76 (0.28-2.10)	0.60
			Indirect	0.86 (0.77-0.97)		
DPP4:Control	2288	41	Direct	0.98 (0.90-1.07)	1.17 (0.27-5.11)	0.83
			Indirect	0.82 (0.32-2.21)		
DPP4:SGLT2	6	4	Direct	1.38 (0.25-11.23)	1.82 (0.20-16.68)	0.60
			Indirect	0.78 (0.68-0.90)		
GLP1:Control	2520	25	Direct	1.14 (1.06-1.23)	0.68 (0.05-9.35)	0.78
			Indirect	1.52 (0.56-4.06)		
GLP1:SGLT2	4	1	Direct	0.26 (0.01-2.45)	0.52 (0.05-5.62)	0.59
			Indirect	0.92 (0.80-1.05)		
Control:SGLT2	1207	25	Direct	0.80 (0.71-0.89)	0.86 (0.04-20.28)	0.93
			Indirect	0.95 (0.24-3.99)		



Zheng SL et al, JAMA 2018: eFigure9

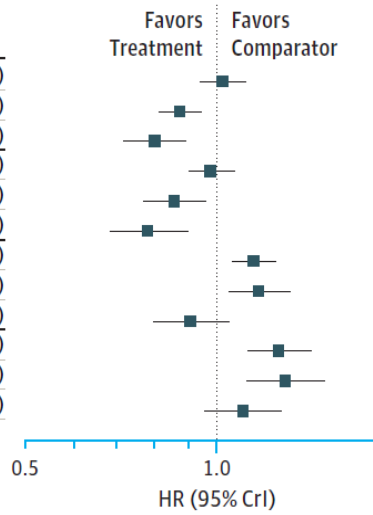
メインの結果

- 全ての組み合わせのフォレストプロット
- GLP-1とSGLT-2はコントロールよりも死亡を抑制
  - ハザード比が、0.88、080
  - リスク差が-0.6%、-1.0%なので、Number Needed to Treatは、166.7、100.0人
- GLP-1とSGLT-2はDPP-4より死亡抑制
  - ハザード比が、0.86、0.78

A Primary outcome: all-cause mortality, 97 trials; $I^2 = 12\%$

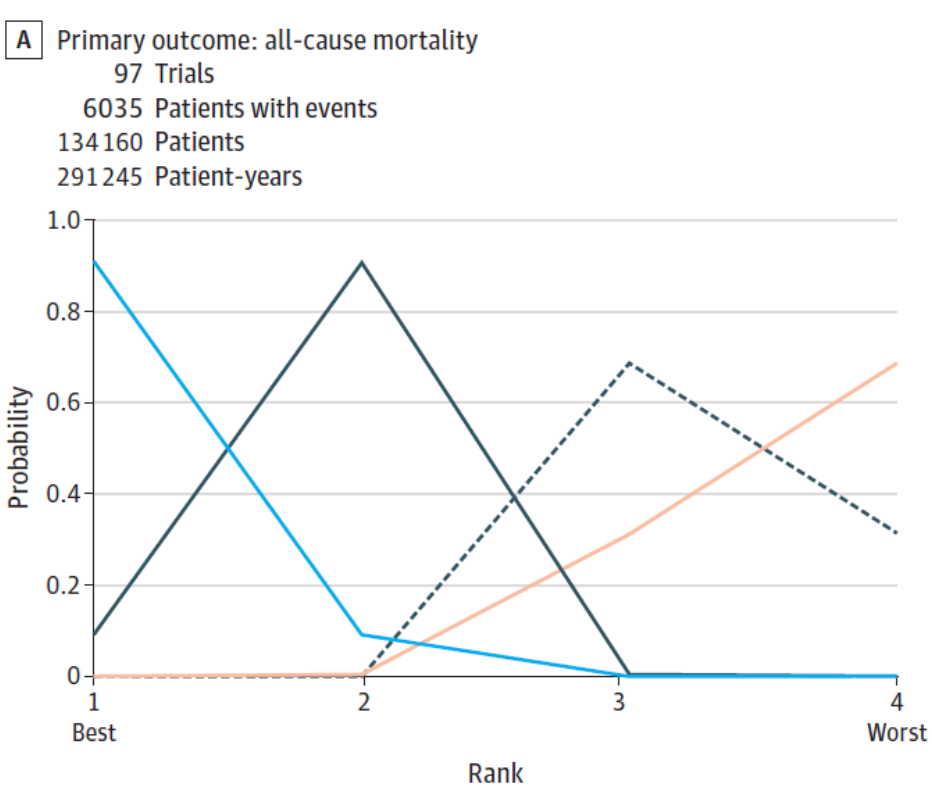
Treatment	Comparator	Absolute RD (95% CrI), %	HR (95% CrI)
DPP-4 inhibitor	vs Control	0.1 (-0.3 to 0.6)	1.02 (0.94 to 1.11)
GLP-1 agonist		-0.6 (-1.0 to -0.3)	0.88 (0.81 to 0.94)
SGLT-2 inhibitor		-1.0 (-1.5 to -0.6)	0.80 (0.71 to 0.89)
Control	vs DPP-4 inhibitor	-0.1 (-0.4 to 0.2)	0.98 (0.90 to 1.06)
GLP-1 agonist		-0.5 (-0.9 to -0.2)	0.86 (0.77 to 0.96)
SGLT-2 inhibitor		-0.9 (-1.2 to -0.4)	0.78 (0.68 to 0.90)
Control	vs GLP-1 agonist	0.6 (0.3 to 1.0)	1.14 (1.06 to 1.23)
DPP-4 inhibitor		0.7 (0.2 to 1.3)	1.17 (1.04 to 1.30)
SGLT-2 inhibitor		-0.4 (-0.9 to 0.2)	0.91 (0.79 to 1.04)
Control	vs SGLT-2 inhibitor	0.9 (0.4 to 1.5)	1.25 (1.12 to 1.40)
DPP-4 inhibitor		1.0 (0.4 to 1.7)	1.28 (1.11 to 1.47)
GLP-1 agonist		0.4 (-0.1 to 0.9)	1.10 (0.96 to 1.26)

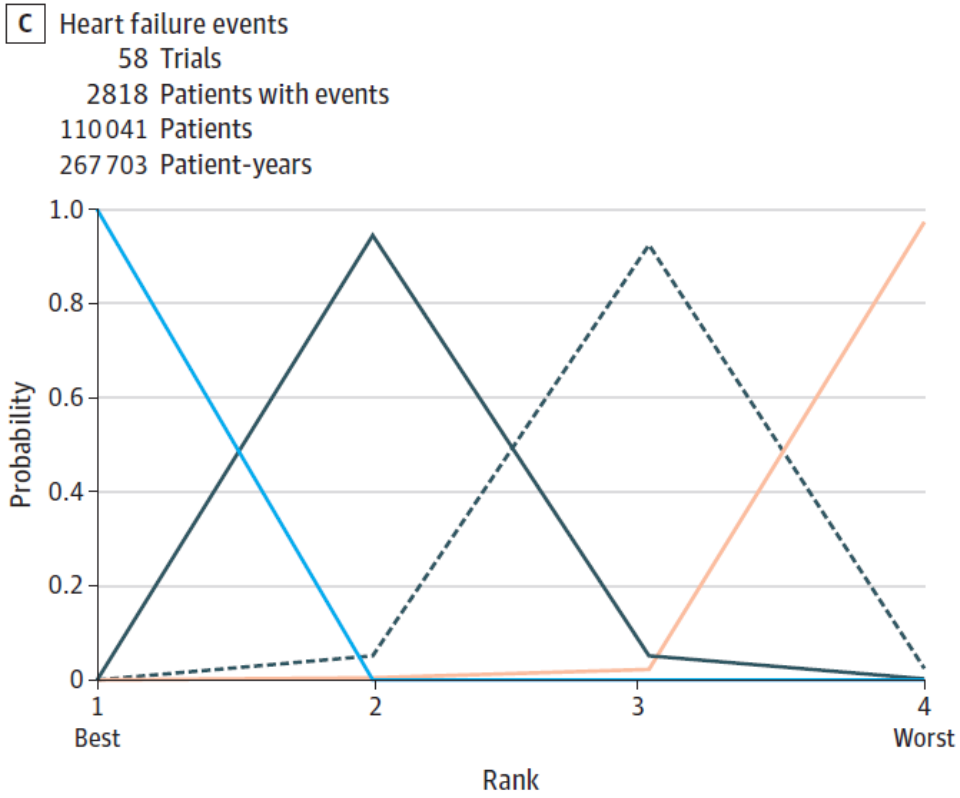
Treatment	No. of Trials	No. With Events (%)	Total No. of Patients
Control	88	2955 (5.2)	57022
DPP-4 inhibitor	49	1171 (3.9)	30178
GLP-1 agonist	32	1195 (4.4)	27373
SGLT-2 inhibitor	29	714 (3.6)	19587



Zheng SL et al, JAMA 2018: Figure3

結果提示例：事後確率に基づくランキング





Zheng SL et al, JAMA 2018: Figure5

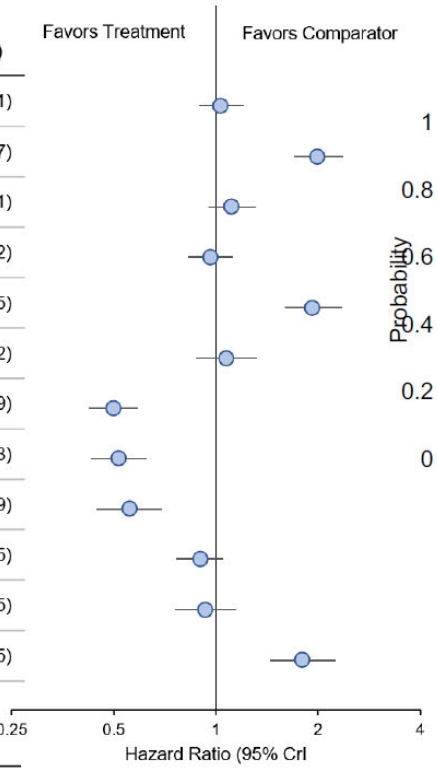
他のエンドポイント：中止に至った有害事象

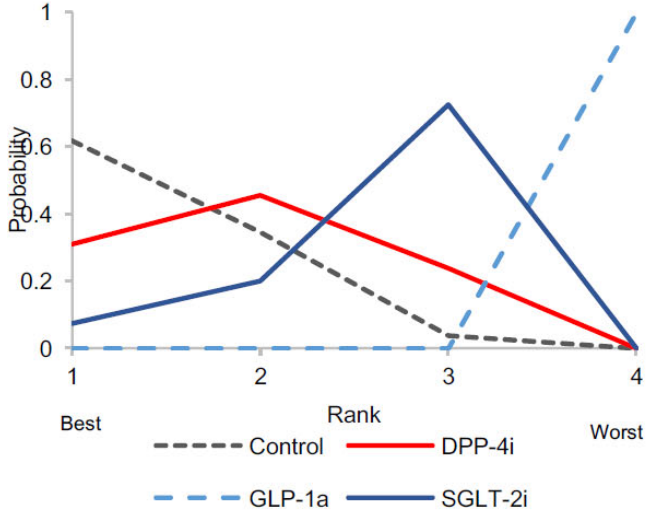
Adverse events leading to withdrawal (Random-effects)

Treatment	Comparator	ARD (95% CI)	HR (95% CrI)
DPP-4 inhibitor		+0.2% (-0.5% to +1.0%)	1.04 (0.89-1.21)
GLP-1 agonist	vs Control	+4.7% (+3.3% to +6.5%)	2.00 (1.70-2.37)
SGLT-2 inhibitor		+0.5% (-0.2% to +1.5%)	1.11 (0.95-1.31)
Control		-0.1% (-0.6% to +0.4%)	0.97 (0.83-1.12)
GLP-1 agonist	vs DPP-4 inhibitor	+3.1% (+2.0% to +4.5%)	1.93 (1.59-2.35)
SGLT-2 inhibitor		+0.2% (-0.4% to +1.1%)	1.07 (0.87-1.32)
Control		-3.5% (-4.1% to -2.9%)	0.50 (0.42-0.59)
DPP-4 inhibitor	vs GLP-1 agonist	-3.4% (-4.0% to -2.6%)	0.52 (0.43-0.63)
SGLT-2 inhibitor		-3.1% (-4.0% to -2.2%)	0.56 (0.44-0.69)
Control		-0.7% (-1.7% to +0.4%)	0.90 (0.76-1.05)
DPP-4 inhibitor	vs SGLT-2 inhibitor	-0.5% (-1.7% to +1.1%)	0.93 (0.76-1.15)
GLP-1 agonist		+5.8% (+3.2% to +9.0%)	1.80 (1.44-2.25)

186 trials; I^2 = 4%

Treatment	Number of trials	Events (%)	Total Patients
Control	172	2112 (4.7)	44,766
DPP-4 inhibitor	86	659 (3.4)	19,620
GLP-1 agonist	60	2278 (7.1)	32,094
SGLT-2 inhibitor	64	1584 (7.2)	21,921

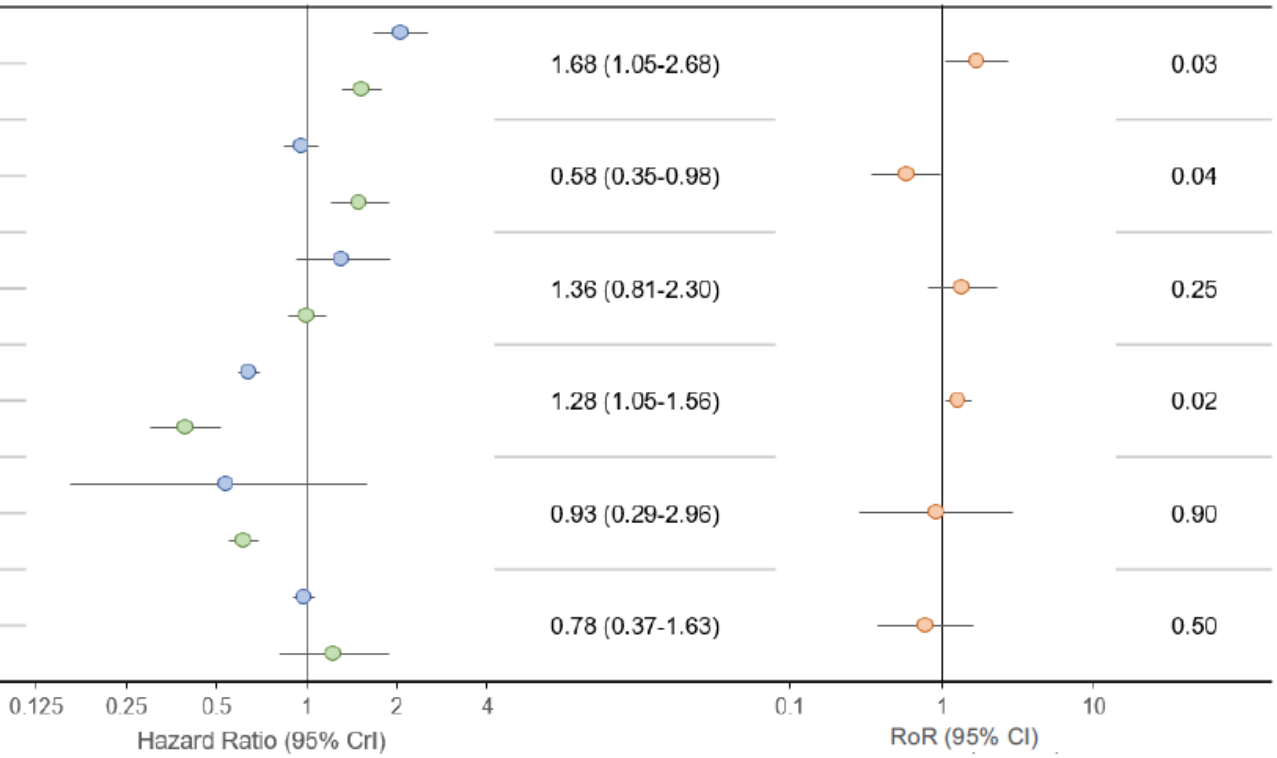




一致性の確認

Adverse events leading to withdrawal

Comparison	Events	Trials	Evidence	HR (95% CrI)	RoR (95% CI)	P-value
DPP4:GLP1	510	12	Direct	2.05 (1.68-2.53)	1.68 (1.05-2.68)	0.03
			Indirect	1.53 (1.31-1.79)		
DPP4:Control	898	71	Direct	0.96 (0.83-1.10)	0.58 (0.35-0.98)	0.04
			Indirect	1.50 (1.20-1.88)		
DPP4:SGLT2	152	8	Direct	1.31 (0.92-1.89)	1.36 (0.81-2.30)	0.25
			Indirect	1.00 (0.87-1.16)		
GLP1:Control	2902	48	Direct	0.64 (0.59-0.69)	1.28 (1.05-1.56)	0.02
			Indirect	0.39 (0.30-0.52)		
GLP1:SGLT2	14	1	Direct	0.53 (0.16-1.59)	0.93 (0.29-2.96)	0.90
			Indirect	0.61 (0.55-0.68)		
Control:SGLT2	2351	59	Direct	0.98 (0.90-1.07)	0.78 (0.37-1.63)	0.50
			Indirect	1.23 (0.81-1.88)		



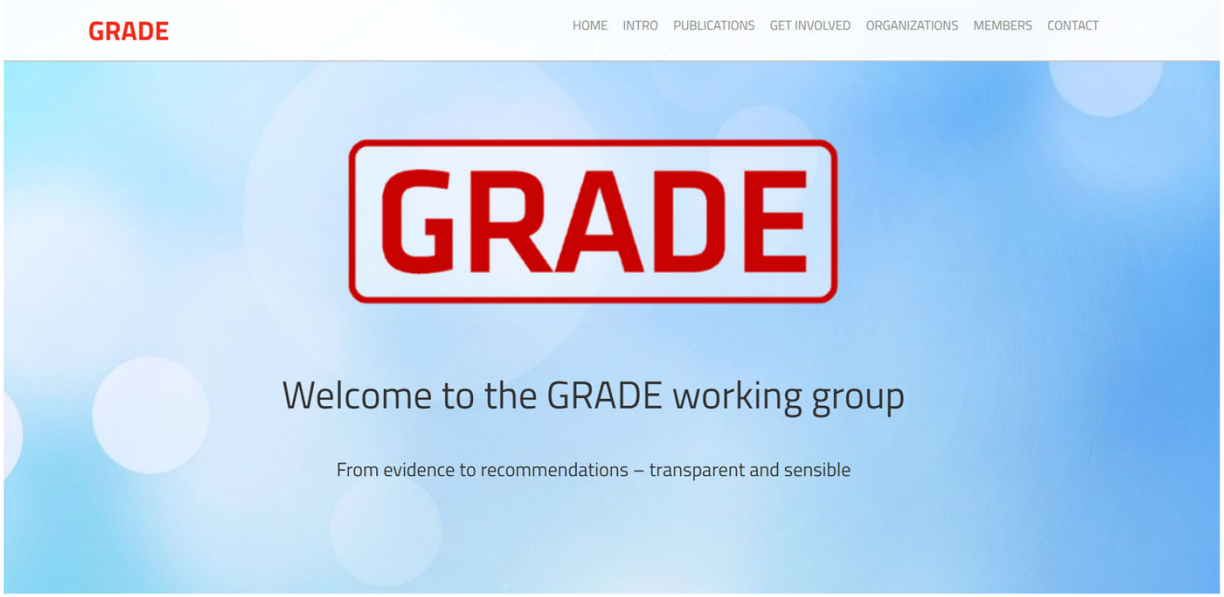
この結果を受けて、臨床のMDからは

- 薬物治療ガイドラインの変更を考えたほうが良い？
  - 直接比較と間接比較、そんなに違わないのは分かった
  - SGLT2が良さそうな気がする
  - それで、この結果、信用していいの？
- エビデンスの確かさ（Certainty of evidence）を整理するガイドライン
  - ISPOR
  - NICE Technical Support Documents

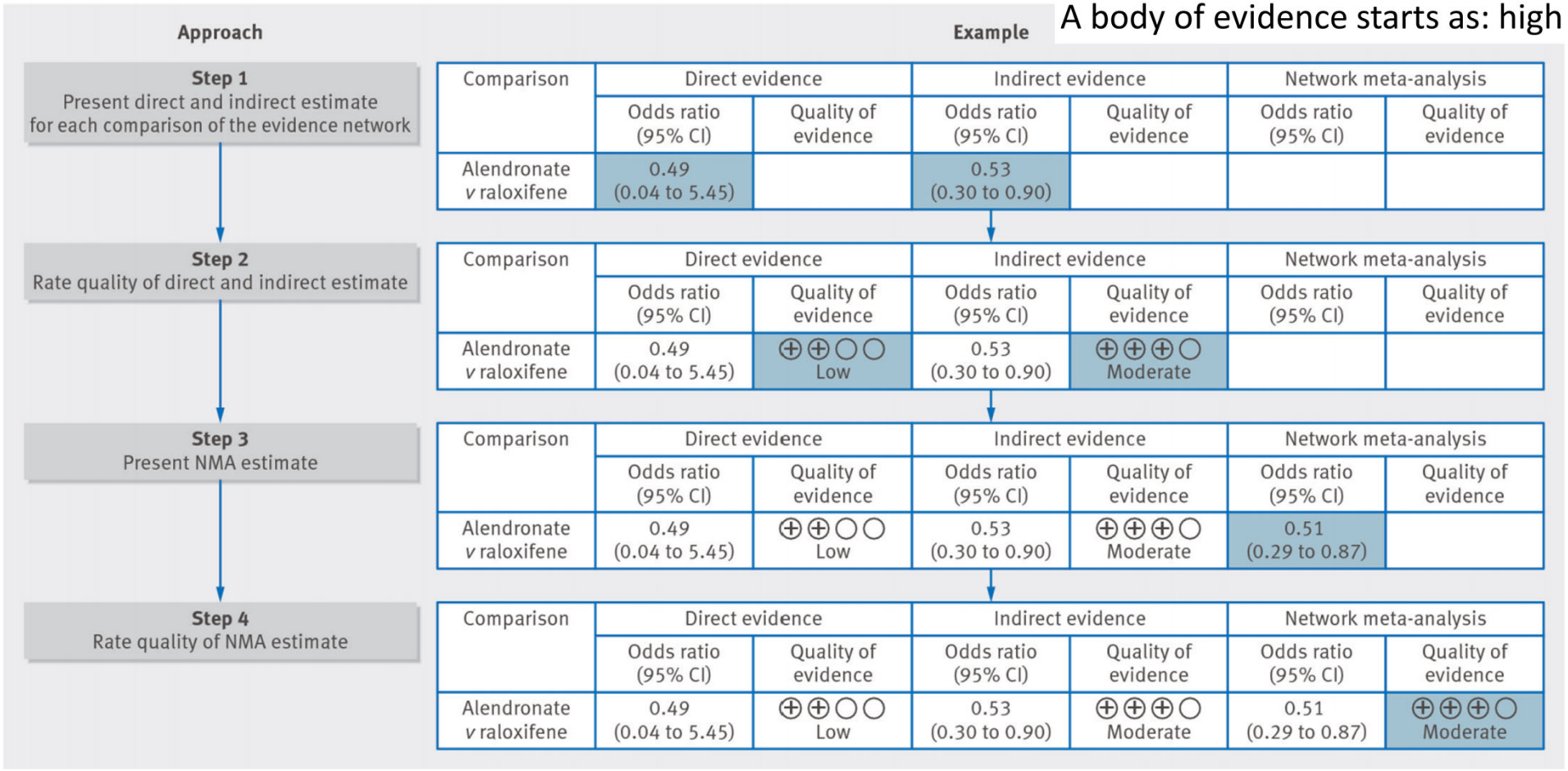
  今日のセミナーで解説あり
  - GRADE system（システマティックレビュー・臨床ガイドライン作成）
  - PRISMA NMA（論文報告ガイドライン）

GRADE System

- To assess the quality (or certainty) of evidence and strength of recommendations
- GRADEハンドブック英語版
  （https://gradepro.org/resources/#handbook）



GRADEでの判定基準

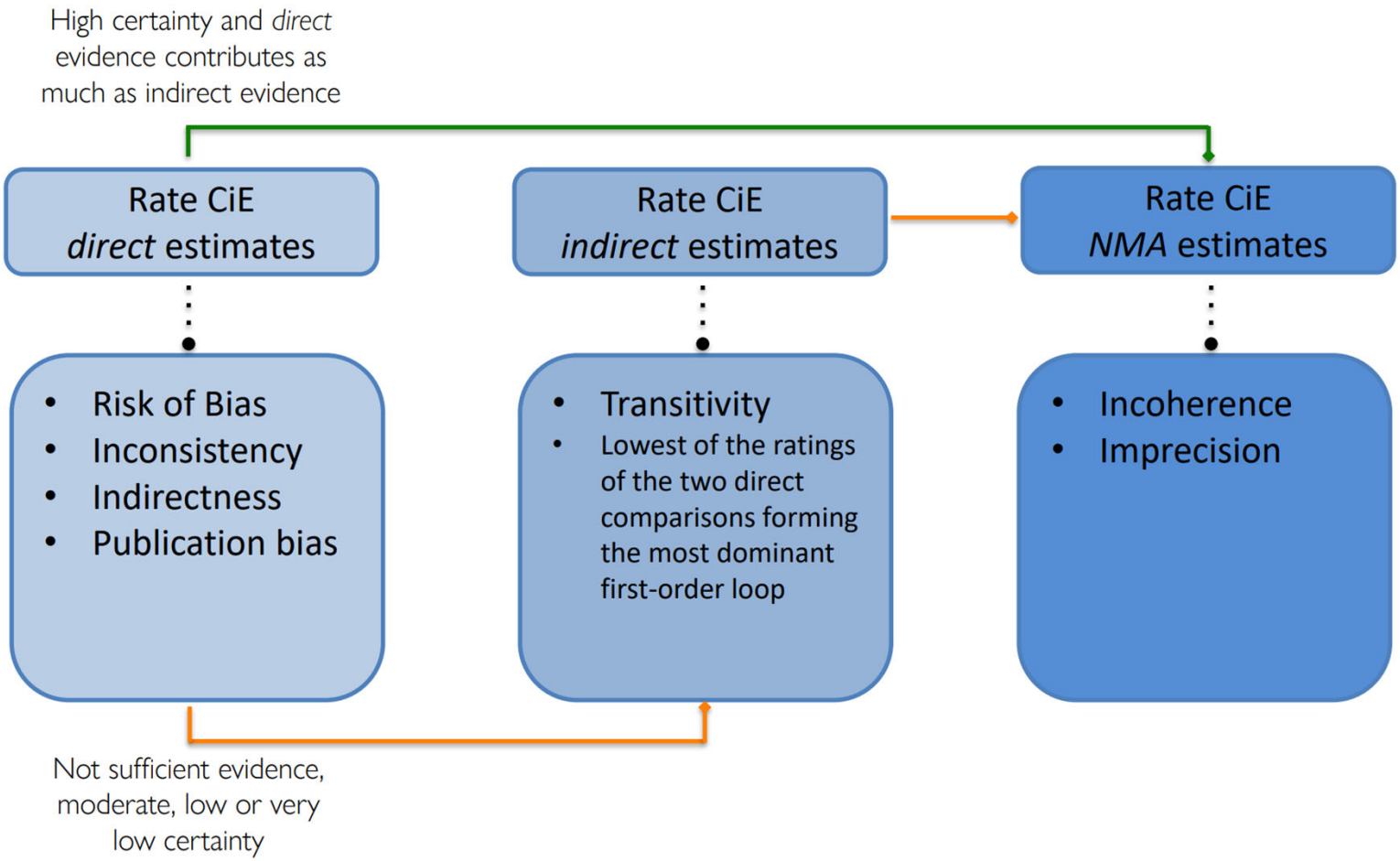


NMAに対する各比較の質評価

- 5 factors that can lower quality
  1. Risk of bias criteria 
     - Lack of randomization (non-randomized or observational studies) lowers confidence to low 
  2. Inconsistency (or heterogeneity) 
  3. Indirectness (PICO and applicability) 
  4. Imprecision 
  5. Publication bias 
- 3 factors can increase quality
  1. large magnitude of effect 
  2. opposing plausible residual bias or confounding 
  3. dose-response gradient 

https://training.cochrane.org/sites/training.cochrane.org/

NMAの推定結果の対する質評価



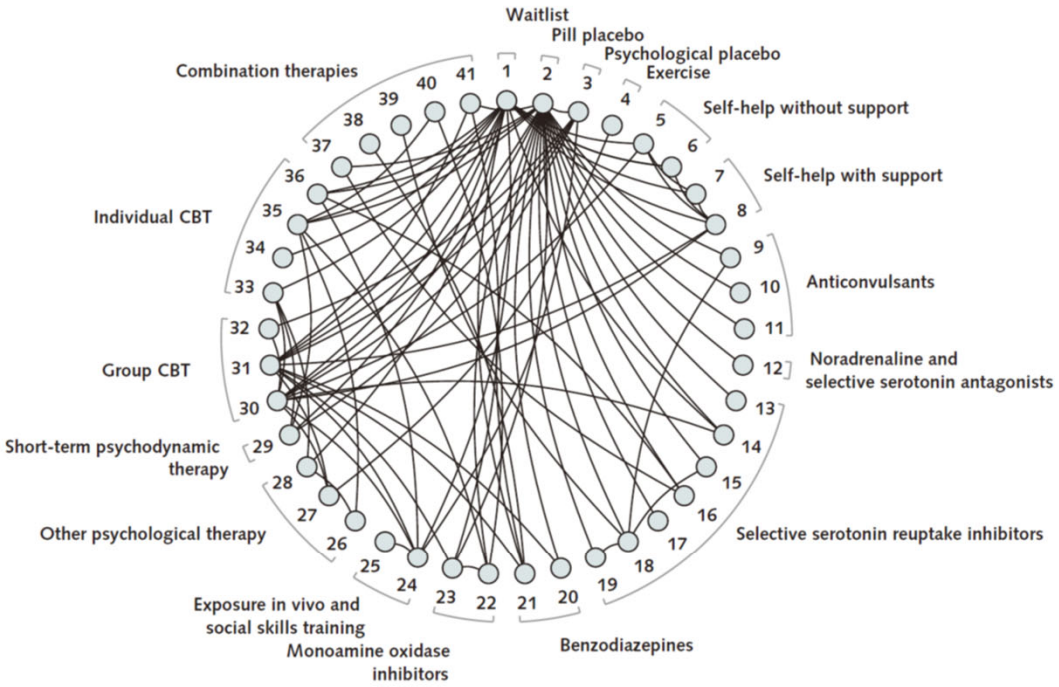
https://training.cochrane.org/sites/training.cochrane.org/

CiNEMA

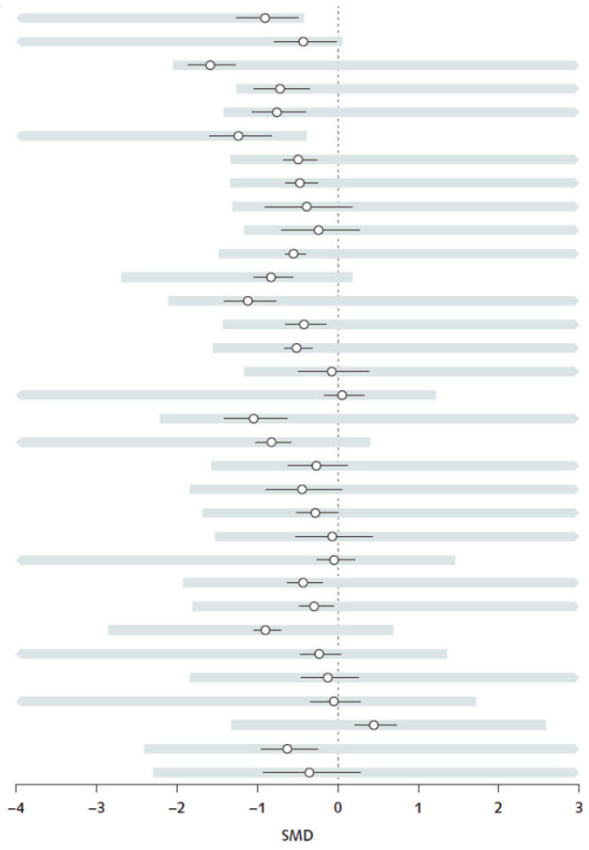
- 質評価だけでは、結果に不均衡（incoherence）が生じうる
  - A vs B、B vs C、A vs Cで三竦みが生じうる

https://cinema.ispm.unibe.ch/#

感度分析



Contrast	SMD	95% CrI		Invariant Interval	
41 vs. 31	−0.88	(−1.27 to −0.49)	–	(NT to −0.42)	36
41 vs. 23	−0.41	(−0.80 to −0.02)	–	(NT to 0.06)	36
36 vs. 1	−1.56	(−1.86 to −1.27)	36	(−2.05 to 13.61)	7
36 vs. 16	−0.69	(−1.04 to −0.35)	36	(−1.26 to 4.91)	16
36 vs. 24	−0.73	(−1.06 to −0.40)	36	(−1.42 to 5.61)	25
41 vs. 2	−1.21	(−1.60 to −0.83)	–	(NT to −0.39)	36
17 vs. 2	−0.47	(−0.68 to −0.26)	17	(−1.33 to 50.71)	12
13 vs. 2	−0.45	(−0.65 to −0.25)	13	(−1.33 to 139.66)	4
39 vs. 18	−0.36	(−0.91 to 0.18)	39	(−1.31 to 3.60)	36
38 vs. 21	−0.22	(−0.70 to 0.27)	38	(−1.16 to 3.71)	36
18 vs. 2	−0.53	(−0.65 to −0.40)	39	(−1.48 to 8.79)	36
23 vs. 2	−0.81	(−1.05 to −0.56)	23	(−2.69 to 0.17)	36
36 vs. 2	−1.09	(−1.41 to −0.77)	36	(−2.11 to 34.05)	18
16 vs. 2	−0.40	(−0.65 to −0.14)	36	(−1.43 to 17.40)	12
19 vs. 2	−0.49	(−0.67 to −0.32)	19	(−1.55 to 29.73)	23
25 vs. 24	−0.05	(−0.49 to 0.39)	25	(−1.16 to 554.74)	39
31 vs. 8	0.07	(−0.17 to 0.32)	31	(−13.67 to 1.22)	36
34 vs. 1	−1.02	(−1.42 to −0.63)	34	(−2.21 to 101.29)	5
31 vs. 1	−0.80	(−1.02 to −0.58)	31	(−16.38 to 0.40)	36
32 vs. 30	−0.25	(−0.61 to 0.12)	32	(−1.57 to 5.66)	36
11 vs. 2	−0.42	(−0.89 to 0.04)	11	(−1.84 to 856.05)	34
9 vs. 2	−0.25	(−0.51 to 0.01)	9	(−1.68 to 46.63)	36
40 vs. 35	−0.05	(−0.52 to 0.43)	40	(−1.53 to 3.12)	36
30 vs. 24	−0.02	(−0.27 to 0.22)	30	(−8.80 to 1.46)	36
15 vs. 2	−0.41	(−0.63 to −0.19)	15	(−1.92 to 49.40)	23
22 vs. 2	−0.27	(−0.48 to −0.06)	22	(−1.81 to 5.92)	18
8 vs. 1	−0.88	(−1.04 to −0.71)	8	(−2.86 to 0.68)	36
8 vs. 6	−0.21	(−0.46 to 0.04)	8	(−13.40 to 1.36)	6
37 vs. 30	−0.10	(−0.46 to 0.25)	37	(−1.84 to 7.10)	36
8 vs. 7	−0.03	(−0.34 to 0.28)	8	(−17.13 to 1.72)	7
31 vs. 23	0.47	(0.21 to 0.74)	36	(−1.32 to 2.58)	23
21 vs. 2	−0.60	(−0.95 to −0.25)	21	(−2.41 to 10.57)	36
12 vs. 2	−0.33	(−0.93 to 0.28)	12	(−2.30 to 59.55)	36

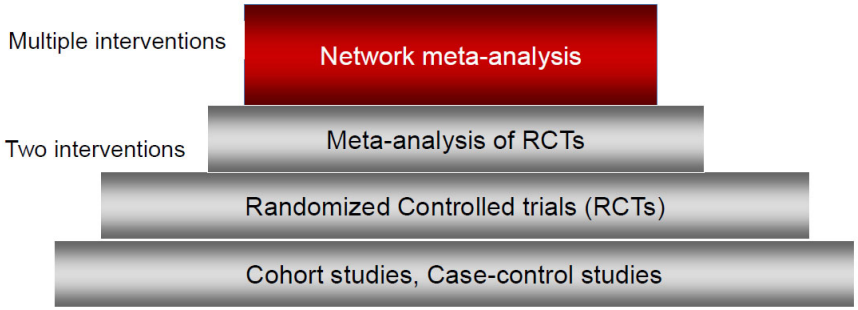


Base-case recommended treatment is 41.

Phillippo DM et al, Ann Int Med 2019

まとめ

- NMAの方法論が整理されてきて、治療・薬剤間のエビデンスをまとめることができるようになってきた
- 実施環境も整備されてきたことで、まだまだ事例は増えていく
  - 治療ガイドラインにおける利用
  - 新規試験の参考データとしての利用
  - 地域フォーミュラリ、院内フォーミュラリにおける利用
- 結果が目的と比して利用（信用）できるかを個別に判断する必要がある



従来のエビデンスヒエラルキーのように単純ではない