

総説 (2016年度横浜市立大学医学研究奨励賞受賞研究)

進行／局所進行／再発非小細胞肺癌の一次化学療法の
第Ⅲ相ランダム化試験において
無増悪生存期間／奏効率／病勢制御率は全生存期間の代用となるか

堀 田 信 之

横浜市立大学大学院医学研究科 呼吸器病学
National Human Genome Research Institute, National Institutes of Health

要 旨：近年、化学療法未施行の進行/局所進行/再発非小細胞肺癌患者を対象としたRCTにおいて全生存期間（overall survival, OS）の代用として無増悪生存期間（progression-free survival, PFS）、奏効率（response rate, RR）、病勢制御率（disease control ratio, DCR）がしばしば使われている。しかし、これらの指標がOSと相關しているかどうかは十分に検証されていない。私たちは2005年1月1日以降に発表された22709人のNSCLC患者を対象とした44のrandomized controlled trialのデータを用いてsystematic reviewを行い、これを検証した。二群間のhazard ratio (HR), 中央値の差 (Δ), odds ratio (OR) 等の相関係数・決定係数を評価した。HRpfs, Δ PFS, ORrr, およびORdcrはすべて, HRosとの中等度のtrial-levelの関係を有していた。私たちのデータは、ORrrが最善のHRosの代用評価項目であり、HRpfsがそれに次ぐと判断された。PFSとRRが以前の研究から予想されたよりも正確にOSを反映していることを示唆されたが、同時にPFSおよびRRは十分な精度にてOSを予測できないと判断された。今後OSの延長に伴い、PFSはさらにOSを反映しなくなるだろう。

Key words: 無増悪生存期間 (Postprogression survival), 代理エンドポイント (Surrogate marker), 殺細胞性抗癌剤 (Cytotoxic agent), 分子標的治療 (Molecular targeted therapy)

緒 言

原発性肺癌は、現在、世界で最も罹患者の多い悪性新生物のひとつであり、世界の癌死の主要な要因である¹⁾。既報のランダム化臨床試験 (randomized controlled trial, RCT) により、現在標準治療とされている化学療法が、非小細胞肺癌 (non-small cell lung cancer, NSCLC) 患者の全生存期間 (overall survival, OS) を延長することが明らかにされている。OSは、肺癌のRCTにおいて最も広く受け入れられている評価項目である²⁻⁴⁾。これは、OSは評価者によるバイアスを受けず、測定が簡単で、解釈が容易で、患者にとって明らかな利点を示すためである。米国食品医薬品局 (Food and Drug Administration, FDA)

および欧州医薬品庁 (European Medicines Agency, EMA) は、OSはNSCLC化学療法の評価において最も信頼できる評価項目であると考えている^{2, 5)}。しかし、OSの評価には長期間のフォローアップと多数の患者を必要とするここと、後治療や非癌死亡の影響を受けやすいこと、などの欠点がある⁶⁻⁸⁾。OSを、無増悪生存期間 (progression-free survival, PFS)、奏効率 (response rate, RR)、病勢制御率 (disease control ratio, DCR) などの代用評価項目に置き換えることにより、これらの欠点を克服できる可能性がある⁴⁾。

2000年から2012年までのデータによると、腫瘍評価項目としてRRは第II相進行NSCLC試験の21%に使用され、PFSは第II相試験の32%と第III相試験の27%に使用

されている²⁾。出版された複数の論文で、PFS および RR が進行 NSCLC に対する化学療法の効果を評価するための良い代用評価項目となりうるかどうかを調べられている⁹⁻¹⁴⁾。しかしながら、これらの研究のいずれも、第 III 相 RCT における第一選択化学療法を中心に評価したものではない⁹⁻¹¹⁾。近年の化学療法薬は副作用が低下し、第二次および第三次以降の化学療法は、多くの進行/局所進行/再発 NSCLC 患者に投与可能となっている。その結果、増悪後生存期間 (post-progression survival, PPS) が OS に占める割合は増大しつつある^{6, 8, 11)}。したがって、PFS, RR, および DCR が OS を正確に反映しているかどうか再評価する必要がある。仮に PFS, RR, および DCR が OS を反映していないとすると、これらの代用評価項目を用いた RCT が誤った結論を招く可能性がある⁴⁾。

本 systematic review は化学療法未施行の進行/局所進行/再発 NSCLC 患者を対象とした RCT において PFS, RR, および DCR が OS と相関しているか否かを評価するために行われた。

事前登録

私たちの行った解析のプロトコールは the international prospective register of systematic reviews (PROSPERO) に登録され (CRD42015029449)¹⁵⁾, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) に従って行われた。

評価研究組み込み基準

化学療法未施行の進行/局所進行/再発 NSCLC 患者に対する第一次化学療法を投与した 2 つの群を比較する第 III 相 RCT を対象とした。主要評価項目は問わないが、full article のみを対象とし、学会抄録等は除外した。2005 年 1 月 1 日以降に発表された研究のみを対象とした。これは、標準レジメンが変更になっていること、PET-CTを中心とする画像評価機器の進歩により病勢進行の検出が容易になったこと、OS および PPS が延長していること等により 2005 年以前の研究で現在の治療評価技術を議論することが困難なためである⁹⁾。

治療内容は、細胞傷害性薬剤および分子標的療法を含む第一次化学療法とした。免疫療法および免疫調節薬は対象から外した。術前化学療法、術後化学療法、放射線療法、放射線化学療法は除外した。酵素療法、非ステロイド抗炎症薬および抗生物質による治療等の非抗癌内科的治療も対象としなかった。組み込まれた研究の多くは 2 つの異なる薬剤あるいは 2 つの異なる薬剤の組み合わせの比較であったが、同一薬剤の異なる投与方法すなわち固定容量レジメン対漸増容量レジメン等の形式の比較

も可能とした。

化学療法の数サイクル後または維持療法の直前にランダム化が行われた場合は除外とした。

評価研究の検索方法

2015 年 11 月 16 日に、Medline, PubMed, Cochrane Database, EMBASE, Web of Science を検索した。さらに組み込まれた論文の参考文献を検索した。2人の研究者がタイトルと抄録を独立してスクリーニングし、その後全文を精査した。同じ研究の重複使用は適宜除外された。研究者間の意見の相違は話し合いにて解決された。

評価研究のバイアスリスクの評価

個々の RCT におけるバイアスリスクは、Cochrane Risk of Bias にて評価された¹⁶⁾。

評価項目の評価

ハザード比 (Hazard ratio, HR) は、イベント発生までの時間データを含む OS や PFS を評価するための最良の方法であると考え、OS のハザード比 (HRos) を真のアウトカムと仮定した¹⁶⁾。PFS の HR (HRpfs), 試験治療群の PFS 中央値から対照群の PFS の中央値を引いた差 (Δ PFS), RR のオッズ比 (ORrr), DCR のオッズ比 (ORdcr) の 4 つの代用評価項目のうち少なくとも 1 つを評価している論文を組み込み対象とした。RR, DCR, および PFS 評価のための病勢評価は、Response Evaluation Criteria in Solid Tumors (RECIST) 2000 ガイドラインおよび RECIST 2009 改訂ガイドラインの Response Evaluation Criteria (反応評価基準) に従った¹⁷⁾。無増悪期間 (Time to progression, TTP) は PFS とみなさなかった。

データ抽出

データは 2 人の研究者によって独立して抽出され、ダブルチェックされた。

ひとつの評価対象から 2 つ以上タイミングの OS/PFS が得られる場合、プロトコールにて事前に宣言された方法に基づいて得られた生存データを選択した。観察期間が短く生存データが初回報告の論文に記載されていない場合、二次報告等のデータで代用した。治験が中間解析により早期終了した場合は、終了時に入手可能な生存データを使用した。多変量解析により調整された評価項目ではなく、調整されていない評価項目を用いた。サブグループ分析と探索的分析のデータは使用しなかった。

試験治療群と対照群は、HR 評価時の比較に基づいて決

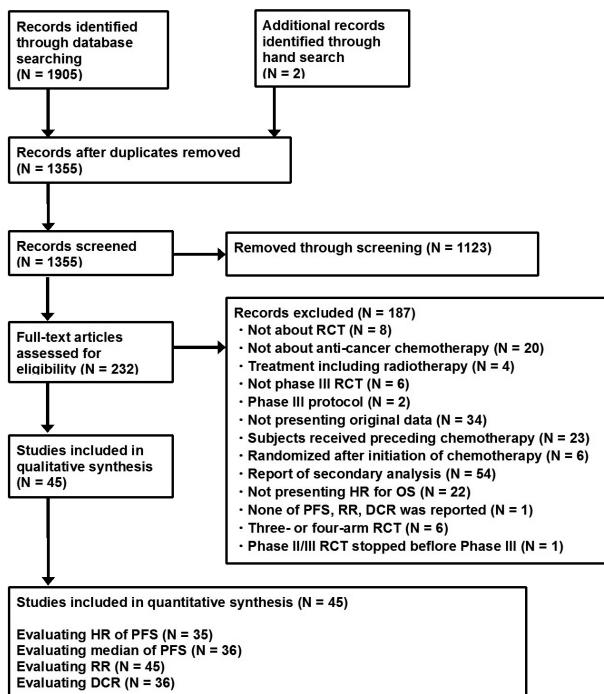


図1 評価研究の検索経過

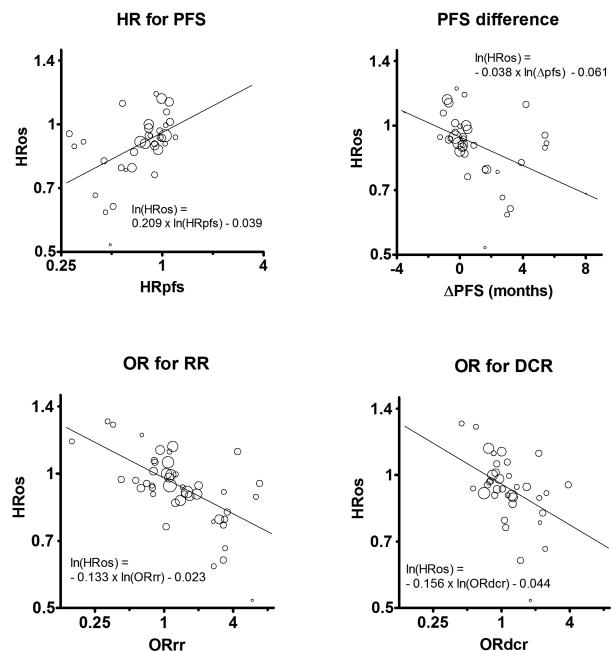


図2 全生存期間と代用評価項目の散布図

OS: overall survival. PFS: progression-free survival. RR: response rate. DCR: disease control rate. HR: hazard ratio. OR: odds ratio. ln: natural logarithm.

定された。すなわち、 $HR_{ros} < 1$ 、 $HR_{pfs} < 1$ 、 $\Delta PFS > 0$ 、 $OR_{rr} > 1$ 、および $OR_{dcr} > 1$ は、試験治療群で治療成績が良いことを示している。

Intention-to-treat, full-analysis-set および per-protocol-population analyses はいずれも許容し、各研究が主要評価項目と定義した解析法に基づくデータを優先した。

データ統合と解釈

真的アウトカムと仮定された HR_{ros} を代用評価項目が正確に評価できているかどうかを下記の3項目で評価した。(i)重み付きスピアマン順位相関係数、(ii)重み付け最小二乗線形回帰にて求めた標準化重み付き回帰係数、および(iii)調整重み付き決定係数^{4, 9, 18, 19)}。各試験の重みは、OSについて評価された患者の数に基づいて決定された。標準化重み付き回帰係数と、調整重み付き決定係数の計算にはHRとORの自然対数(ln)を使用した⁹⁾。

重み付きスピアマン順位相関係数は、ソフトウェア「R」の「boot」パッケージ内の「corr」コマンドを使用して推定された²⁰⁾。ソフトウェア「R」の「lm」コマンドを使用して、標準化された重み付き回帰係数および調節された重み付け係数の係数を推定した²¹⁾。

サブグループ感度解析は、サンプルサイズ500人を基準に各研究をグループ分けした⁹⁾。

出版バイアスは、funnel plotを用いて評価した¹⁶⁾。

評価研究の検索方法と研究の特徴

基準を満たした1907件の報告のうち、重複削除にて552件、タイトル/抄録のスクリーニングにて1123件、全文評価にて188本が除外された。私たちは最終的に44の論文を評価対象とした(表1、図1)。このうち8本が非扁平上皮癌、4本が腺癌、5本が上皮増殖因子受容体(epidermal growth factor receptor, EGFR)変異陽性患者、および1本がechinoderm microtubule-associated protein-like 4-anaplastic lymphoma kinase (EML 4-ALK)転座患者に特化した研究であった。他の研究はNSCLC病理型亜分類に関わらず患者集積していた。OSおよびPFSは、それぞれ30および11のRCTにおける主要評価項目とされていた。OSを評価された患者の数は56から1725の範囲で、中央値は387であり、合計患者数は22709人であった。 HR_{pfs} 、 ΔPFS 、 OR_{rr} 、および OR_{dcr} は、それぞれ34、35、44、および35のRCTで評価された。個々のRCTのバイアス評価の結果、36本のRCTは非盲検化治療デザインのために高いリスクを有し、34は画像評価時に盲検化がされていないために検出バイアスのリスクが高いと判断された。

HR_{ros} と代用評価項目の関連

HR_{pfs} 、 ΔPFS 、 OR_{rr} 、および OR_{dcr} は、 HR_{ros} に対する重み付きスピアマン順位相関係数値は順に、0.496、0.477、0.570および0.470であり、中程度の相関と

表 1 評価対象論文一覧

Study name	Year	Pathological type	EGFR/ALK status	Experimental arm	Control arm	Primary endpoint	Independent radiologic review	Premature stop	n for OS evaluation	Analysis method (ITT/FAS)	Outcomes			
											HRos	HRpfs	ΔPFS	ORrr
BEYOND ¹	2015	Non-sq	NS	Car + Pac + Bev	Car + Pac + Pla	PFS	No	276	ITT	0.68	0.40	2.70	3.40	2.43
BMS099 ²	2010	NSCLC	NS	Car + (Dtx/Ptx) + Cetuximab	Car + (Dtx/Ptx)	PFS	Yes	676	ITT	0.89	0.90	0.16	1.67	1.27
CAPPA ²³	2013	NSCLC	NS	Cis + Gem	Gem	OS	No	56	ITT	0.52	0.49	1.60	5.87	1.80
CATS trial ⁴	2015	NSCLC	NS	Cis + S-1	Cis + Doc	OS	No	596	FAS	1.01	1.11	-0.30	0.81	0.89
Cobo ⁵	2007	NSCLC	NS	Cis + Doc/Gem after genotyping	Cis + Doc	ORR	No	366	ITT	0.90	0.90	0.90	1.51	0.87
ECOG4599 ⁶	2006	Non-sq	NS	Car + Ptx + Bev	Car + Ptx	OS	No	850	ITT	0.79	0.66	1.70	3.03	NA
ENROUTE study ⁷	2015	NSCLC	EGFR+	Erl	Cis + Gem	PFS	Yes	217	ITT	0.91	0.34	5.50	3.32	2.49
FASTACT-2 ⁸	2013	NSCLC	NS	(Cis/Car) + Gem + Erl	(Cis/Car) + Gem + Pla	PFS	Yes	451	ITT	0.79	0.57	1.60	3.37	1.06
First-SIGNAL ⁹	2015	Ad	NS	Gef	Cis + Gem	OS	Yes	309	ITT	0.93	1.20	-0.60	1.45	0.56
FLEX ¹⁰	2009	NSCLC	EGFR+	Cis + Vin + Cetuximab	Cis + Vin	OS	No	1125	ITT	0.87	0.94	0.00	1.39	NA
G.O.I.M. ¹¹	2008	NSCLC	NS	Cis + weekly-Vin	Cis + day 1 / 8-Vin	OS, toxicity	No	278	ITT	1.12	NA	1.11	0.85	NA
Gatzemeier ¹²	2007	NSCLC	NS	Cis + Gem + Erl	Cis + Gem + Pla	OS	No	1159	ITT	1.06	NA	1.08	NA	NA
GEICO ¹³	2007	NSCLC	NS	Cis + Gem (120 mins)	Cis + Gem (30 mins)	OS	No	400	ITT	0.93	0.97	0.23	0.79	1.02
Georgoulias ¹⁴	2008	NSCLC	NS	Dtx + Gem	Dtx + Gem	OS	Yes	312	ITT	1.29	NA	NA	0.36	0.60
IFCT-0501 ¹⁵	2011	NSCLC	NS	Car + Ptx	Car + Ptx	OS	No	451	ITT	0.64	0.51	3.20	3.28	1.47
JCOG0803 ¹⁶	2015	NSCLC	NS	Cis + Dtx	Dtx	OS	Yes	272	ITT	1.18	0.92	0.30	0.16	NA
JML ¹⁷	2014	Non-sq	NS	Cis + Pen	Cis + Gem	OS	No	256	ITT	1.00	1.05	0.03	1.26	1.16
Komidis ¹⁸	2008	NSCLC	NS	Car + Gem	Gen+Ptx	OS	No	452	ITT	1.06	NA	NA	0.83	0.91
Kubota ¹⁹	2008	NSCLC	NS	Vin + Gem f/b Dtx	Car + Ptx	OS	No	393	ITT	0.97	0.97	-0.30	0.57	0.79
Langer ²⁰	2008	NSCLC	NS	Car + Ppx	Car + Ppx	OS	No	400	ITT	0.97	NA	NA	0.42	0.80
Lee ²¹	2009	NSCLC	NS	Car + Gem + thalidomide	Car + Gem + Pla	OS	No	722	ITT	1.13	1.10	-0.70	0.92	1.01
LETS ²²	2010	NSCLC	NS	Car + S-1	Car + Ptx	OS	No	563	ITT	0.93	1.00	-0.70	0.63	0.91
LUX-Lung ²³	2013	Ad	EGFR+	Afa	PFS	Yes	No	345	ITT	1.12	0.58	4.20	4.37	2.13
LUX-Lung6 ²⁴	2014	Ad	EGFR+	Afa	PFS	Yes	No	364	ITT	0.95	0.28	5.40	6.80	3.88
Mok ²⁵	2009	Ad	NS	Gef	Car + Ptx	PFS	No	1217	ITT	0.91	0.74	-0.10	1.59	0.70
NEJ002 ²⁶	2010	NSCLC	EGFR+	Gef	Car + Ptx	PFS	Yes	224	ITT	0.89	0.30	5.40	6.32	2.15
NEJUS ²⁷	2012	Non-sq	NS	Cis + Gem + Sorafenib	Cis + Gem + Pla	OS	Yes	772	FAS	0.98	0.83	0.50	1.10	0.96
NVALT-3 ²⁸	2011	NSCLC	NS	Car + Ptx	QOL	No	No	181	ITT	1.22	NA	-0.20	0.64	NA
O'Brien ²⁹	2008	NSCLC	NS	Ppx	Gem/Vin	OS	No	381	ITT	0.95	NA	NA	0.75	NA
Paccagnella ³⁰	2006	NSCLC	NS	Car + Ptx	OS, ORR	OS	No	319	ITT	1.31	NA	NA	0.32	0.44
PointBreak ³¹	2013	Non-sq	NS	Car + Pem + Bev f/b Pem + Bev	Car + Ptx + Gem	OS	No	939	ITT	1.00	0.83	0.40	1.05	0.84
PROFILE ³²	2014	Non-sq	ALK+	Carz	Car + Gem	PFS	Yes	343	ITT	0.82	0.45	3.90	3.55	2.22
PRONOUNCE ³³	2015	Non-sq	NS	Car + Pem f/b Pem	Car + Ptx + Bev f/b Bev	G4PFS	No	361	ITT	1.07	1.06	-1.05	0.82	1.13
Pujol ³⁴	2005	NSCLC	NS	Gem + Dtx	Cis + Vin	PFS	No	311	ITT	0.90	1.04	0.20	0.80	1.14
Ruid ³⁵	2005	NSCLC	NS	Car + Gem	Cis + Ifo + Mtc	OS	No	422	ITT	0.76	0.90	0.50	1.04	1.09
Scagliotti_2008 ³⁶	2008	NSCLC	NS	Cis + Pen	Cis + Gen	OS	No	1725	ITT	0.94	1.04	-0.30	1.13	NA
Scagliotti_2010 ³⁷	2010	NSCLC	NS	Car + Ptx + Sor	Car + Ptx + Pla	OS	No	926	ITT	1.15	0.99	-0.80	1.19	0.77
Sederholm ³⁸	2005	NSCLC	NS	Gem + Car	Gem	OS	No	334	ITT	0.77	NA	NA	3.29	NA
TOPICAL ³⁹	2012	NSCLC	NS	Erl	Pla	OS	No	670	ITT	0.94	0.83	0.20	2.00	1.67
TRIBUTE ⁴⁰	2005	NSCLC	NS	Car + Ptx + Erl	Car + Ptx + Pla	OS	No	1079	ITT	1.00	NA	NA	1.15	NA
WTOGO203 ⁴¹	2010	NSCLC	NS	3 cycles of Pt-D f/b Gef	Pt-D	FAS	0.86	0.68	0.30	1.26	1.26	NA	NA	
WTOGO9904 ⁴²	2006	NSCLC	NS	Dtx	Vin	OS	No	180	ITT	0.78	0.61	2.40	2.68	2.19
Yang ⁴³	2014	Non-sq	NS	Cis + Pem f/b Gef	Gef	PFS	No	236	ITT	0.94	0.85	-1.25	0.79	1.37
Zuk ⁴⁴	2013	NSCLC	NS	Car + Pen	Pen	OS	No	205	ITT	0.62	0.46	3.00	2.71	NA

表2 結果要約

	Studies	Patients	R	src	R^2
ln HRpfs	34	17814	0.496	0.439	0.224
Δ PFS	35	17995	0.477	-0.376	0.161
ln ORrr	44	22709	0.570	-0.605	0.350
ln ORdcr	35	15779	0.470	-0.381	0.176

R: the weighted Spearman's rank correlation coefficient

src: the standard weighted correlation coefficient

R^2 : the adjusted weighted coefficient of determinant

判断された（表2）。4つの代用評価項目は、HRosに対して標準化重み付き回帰係数が順に0.439, -0.376, -0.605, -0.381であった（表2、図2）。同様に調整重み付き決定係数はそれぞれ0.224, 0.161, 0.350, および0.176であった（表2）。これらの結果から、4項目間の比較ではORrrならびにHRpfsがHRosの良い代用評価項目であると判断された。

サンプルサイズに基づくサブグループ分析では、これらの傾向が再び確認された。サンプルサイズが500未満の研究のサブグループでは、HRpfs, Δ PFS, ORrr, およびORdcrの重み付きスピアマン順位相関係数は、それぞれ0.461, 0.423, 0.589および0.495であった。サンプルサイズ \geq 500の研究のサブグループでは、重み付きスピアマン順位相関係数の値は、HRpfs, Δ PFS, ORrr, およびORdcrにおいて0.472, 0.450, 0.592および0.361であった。これは、サンプルサイズに関係なく、私たちの分析が頑強であることを示している。

プロトコールに計画されていない解析であるが、主要評価項目としてPFSを使用した11の試験において、サブグループ分析を実施した。11のRCTのうち9つがEGFRチロシンキナーゼ阻害剤（EGFR-TKI）またはBevacizumabを評価していた。HRosとHRpfsの関連は、重み付きスピアマン順位相関係数が-0.032、標準化重み付き回帰係数が0.073、調整重み付き決定係数が-0.105であった（図3）。これらの値は、主要評価項目としてのPFSがこのサブグループ内のOSをまったく予測できないことを意味している。

Funnel plot

Funnel plotからはHRpfsおよび Δ PFSの出版バイアスが否定できなかった（図4）。

本解析の意義

本解析は、PFS/RR/DCRとOSと間の関連性を、第III相試験における化学的未治療の進行/局所進行/再発NSCLC患者の第一次化学療法に焦点を当てて調べた初め

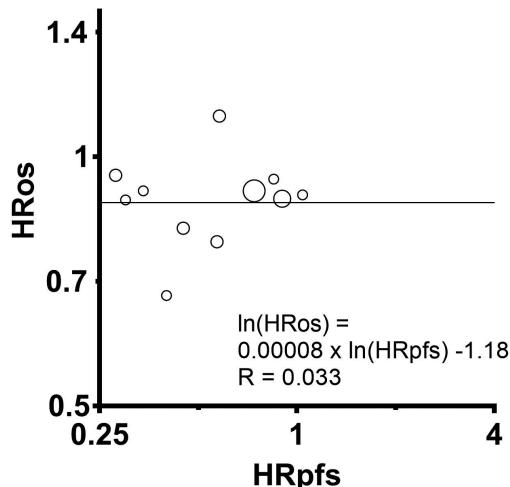


図3 無増悪生存を主要評価項目とするtrialによるサブ解析

てのsystematic reviewである。今回の結果は、PFSとRRが以前の研究から予想されたよりも正確にOSを反映していることを示唆した。RRは第III相試験の設定で主要評価項目として使用されることはあるが、RRはPFSと同様OSの容認できる代用評価項目となり得る（表2）。一方、PFSが主要評価項目として使用された場合、HRpfsとHRosとの間に関連は見られなかった。研究者が「OSで有意差なしPFS有意差あり」の結果が予見されるBevacizumab等の薬剤の研究にて、PFSを主要評価項目として選択したためと思われる。PPSの延長に伴いPFSはOSを反映しにくくなるが、OSを主要評価項目とした研究デザインの実施が困難になるためPFSが主要評価項目として使われる傾向になる。同様の問題はすでに乳がんで顕著である²²⁾。PFSとOSの関連は中等度であり、PFSにおいて $P < 0.05$ を用いた統計的有意性を確認することで、新レジメンの有用性を証明したとは判断することは困難である²³⁾。

無増悪生存期間（PFS）

PFSは、治療開始から癌進行または全死亡までの経過時間として定義される。2003年に米国FDAのOncologic Drugs Advisory Committeeが、進行NSCLC臨床試験の評価項目としてPFSの使用を推奨した後、PFSを主要評価項目として使用する治験が増えている^{2, 11, 24)}。しかし、PFSがOSを正確に予測するかどうかは十分に評価されていない。

2015年に、Blumenthalらは、第二次あるいはそれ以後次の治療を除外せず、FDAに登録されている14の進行NSCLCを対象とした試験のデータを用いて、PFS/RRとOSと関連を探るために、trial-levelおよびindividual-levelの分析を行った⁹⁾。HRosに対するHRpfsおよびORrrの係

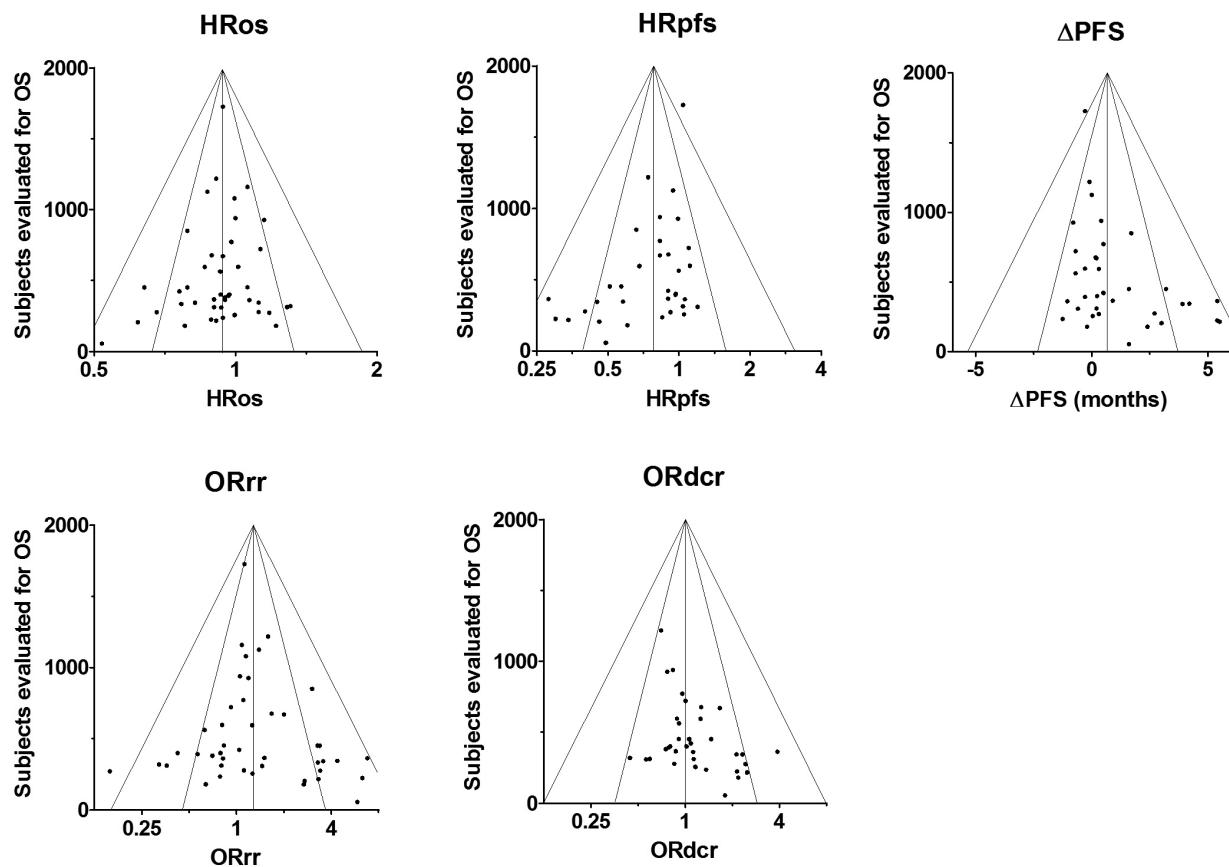


図4 Funnel plot

OS: overall survival

HRos: hazard ratio for overall survival

HRpfs: hazard ratio for progression-free survival

△ PFS: median PFS in experimental arm minus median PFS in control arm

ORrr: odds ratio for response rate

ORdcr: odds ratio for disease control rate

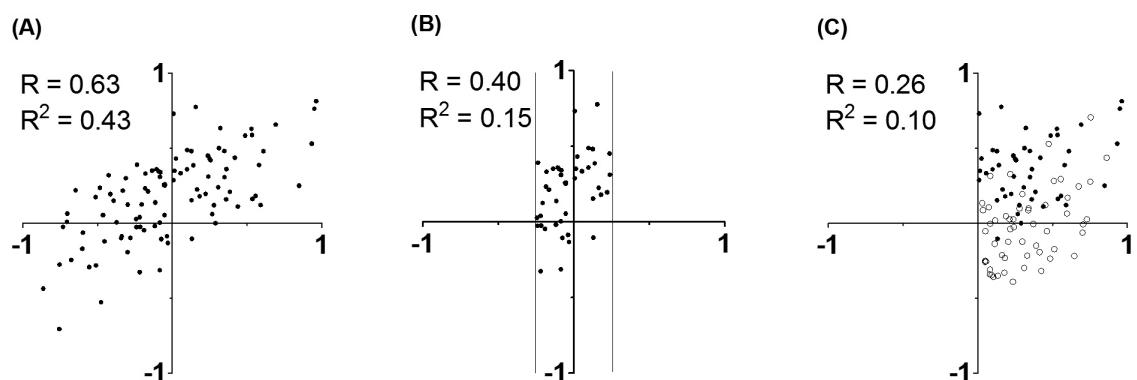


図5 Spearmanの順位相関係数に関する考察

R: The weighted Spearman's rank correlation coefficient.

R^2 : the adjusted weighted coefficient of determinant.

(A) : Dots are plotted from the left-lower area to the right-upper area.

(B) : When $|X|$ is limited to < 0.25 , R and R^2 decrease.

(C) : When the left half dots are 180 degree rotated around the origin, R and R^2 decrease. Open circles indicate rotated data.

数は、それぞれ0.08および0.09であり、私たちの解析で算出された調整重み付き決定係数よりもはるかに低かった（表2）。変数が広く分布している場合（図5-A）、一般的に強い相関係数と強い決定係数が算出される。Blumenthalらの研究はFDAに登録された「無難な」研究のみを対象としていたため、観察されたHRosは0.80～1.04の狭い範囲にあった。変数の範囲が限られているので、相関係数と決定係数は0に近くなった（図5-B）²⁵⁾。私たちはFDA登録以外の「挑戦的な」研究も対象としたため結果としてHRosが0.5-1.4の範囲で広く分布した。私たちの結果とBlumenthalらの結果の乖離は治療効果アウトカムの分布に起因している。

Hottaらは二次治療以降の化学療法を除外せずにEGFR-TKIで治療されたNSCLC患者のPFSとOSとの関連性を調べるために、35本のRCTのデータを用いてtrial-levelの分析を行った¹⁰⁾。HRosに対するHRpfsによる決定係数は0.14であった。Hottaらは、クロスオーバーが存在する場合、PFSはOSとほとんど関連しないと結論付した¹⁰⁾。私たちの研究の結果とHottaらの研究の結果との間の相違は、組織型、EGFR変異の有無、およびEML4-ALK変異の有無の相違によって説明することができる。EGFR-TKIとEML4-ALK阻害剤とのクロスオーバーが許容されている場合、PFSはOSの代用評価項目とはなり得ないというHottaらの見解に賛成である¹⁰⁾。

OSと比較して、PFSはRCTのデザイン上で魅力的な選択肢である。なぜなら、OSよりも早く入手できるPFSは、時間的経済的費用が少ないのみでなく、後治療ならばに非癌死の影響を受けにくく、薬剤の生物学的活動をより正確に評価できると考えられるからである^{6, 7)}。そのため、化学療法の有効性を発見するのにPFSがより適していると考える者もいる^{6, 7)}。現実的に、倫理的に許容せざるを得ない分子標的療法等の有効な治療のクロスオーバー、および近年のOSの延長は、腺癌患者のRCTにおけるOS差異検出することを非常に困難にしている⁸⁾。FDAは、薬剤の認可のためにPFSを主要評価項目として利用することを推奨している^{10, 11)}。

しかし、私たちはPFSには多くの欠点があることを忘れてはならない。試験中の真の進行と画像評価との間には常にタイムラグがあるため、測定された疾患の進行は、常に真の進行と誤差がある。OSおよびQOL（quality of life）とは対照的に、PFSは患者のメリットを直接反映していない。分子標的剤、すなわちベバシズマブまたはエルロチニブを細胞傷害性抗癌剤に添加することは、OSの大きな改善なしにPFSを延長することが知られている^{26, 27)}。しかし、これらの設定では、分子標的薬の上乗せは副作用のリスクを増加させQOLを低下させやすいと想定される。薬価は当然上昇する。OSが伸びず、副作用のリスクが上がり、QOLが下がり、薬価が上がる場合、

PFSの延長が患者にとってどれだけのメリットになるのだろうか？OSの改善を伴わないPFSの改善は慎重に評価すべきである。さらに、画像評価を必要とするPFS評価は、画像評価が盲検化されていない限りバイアスの影響を受けやすい。しかし、主治医と独立した盲検化画像評価を臨床試験で利用することはかなり困難である²⁸⁾。私たちの解析において、PFSのfunnel plotは明らかに非対称である（図4）。通常非対称なfunnel plotは出版バイアスの存在を示唆するが、ここでは別の解釈も可能である。つまり、より小さなサンプルサイズのRCTは質が悪く、盲検化のされていない画像評価で新治療が有利になるようなバイアスの影響を強く受けている可能性が示唆される。

PFSが主要評価項目として使用される場合、PFSはOS（図3）とは相関がみられない。この驚くべき事実からは2つの相反する解釈が可能である。一つ目の解釈はPFSはOSの代用にならないため、PFSを主要評価項目として使用すべきでないという考え方である。二つ目の解釈は、PFSは、クロスオーバーや長期フォローの困難のためにOSが検出できない場合でさえ生物学的活性を検出できる可能性があり、PFSを積極的に使用すべきという考え方である。

HRは、イベント発生までの時間データを含むOSやPFSを評価するための最良の方法であることが知られている¹⁶⁾。実際、私たちの結果からは、△PFSと比較して、HRpfsがHRosをより正確に予測することが確認された（表2、図2）。

奏効率（RR）

客観的奏効率は、1981年に世界保健機関（WHO）によって最初に提案された。現在使用されているRRは2009年のRECISTver1.1に記載されている¹⁷⁾。RRはOSよりもPFSに近い性質を持っている。RRは通常、治療開始後に6～8週目に評価され、非癌死亡および後治療の影響を受けにくい¹⁷⁾。私たちの解析によれば、RRはPFSより優れたOSの代用評価項目となる可能性がある。

RRの長い歴史にもかかわらず、最近の研究者はNSCLC患者を対象とする第III相RCTの主要評価項目としてRRよりもPFSを使用することが多い（表1）^{2, 11, 24)}。現在、RRは主に第II相試験で十分な治療効果と生物学的活性をも薬剤を確認するために使用されており、第III相で評価される薬剤のスクリーニングに使われる。しかし、第II相試験での良好なRRは、しばしばPFSやOSを主要評価項目とした第III相試験で再現されない^{29, 30)}。これはRRの性質に起因するものではなく、「中心への回帰」と呼ばれる統計学的現象である。新たな治療法は、第II相試験で一定の基準を満たした後にのみ、第III相試験で評価さ

れる機会に恵まれる。第II相試験の良好な結果試行は本当に良い治療効果かランダムエラーの影響の双方が考えられる^{29, 30)}。初回の六面サイコロ試行を第II相とし、4以上の目が出た場合のみ第III相に相当する2回目の施行を行うとする。この筋書きでは、第III相に相当する2回目の施行でのサイコロの目の平均値(=3.5)が、第II相に相当する1回目の施行でのサイコロの目の平均値(=5)より小さくなるのは当然である。

RRがOSを予測するかどうかを評価したsystematic reviewは少ない。Tsujinoらの分析によれば、エルロチニブまたはゲフィチニブに関する28の研究のデータを用いた解析を行い、arm-levelのRRはarm-levelのOSに有意に関連していることを示した¹⁴⁾。Johnsonらは相を問わずにsystematic reviewを行い、trial-levelのRR差(%)とOS差との間の決定係数が0.16であることを明らかにした¹²⁾。この決定係数は、私たちの分析において、ln(ORrr)とln(HRos)との間の調節重み付け決定係数0.350よりもはるかに低かった。Johnsonらは「素のRR差」すなわち実験群のRRから対照群のRRを差し引いた値の代わりに「RR差の絶対値」を用いたため決定係数を過小評価している。素のRR差の代わりに絶対値を使用すると、相関係数と決定係数は0に近づく(図5-C)。Blumenthalらによる別のレビューではORrrとHRosとの間の決定係数はtrial-levelで0.09であり、私たちの解析で観察された調整重み付け決定係数よりもはるかに低かった²⁵⁾。これは、PFS(図5-B)の項で説明したように、変数の幅が狭いことにより起因している²⁵⁾。

最近承認された免疫調整薬は、RRの解釈を困難にするとと思われる。

病勢制御率(DCR)

いくつかの研究は、DCRがOSの優れた代用評価項目であることを示している^{31, 32)}。しかし、Tsujinoらのsystematic reviewはOSを予測する際にDCRよりRRが良い代用評価項目であることを示している¹⁴⁾。私たちのデータは、DCRと比較してRRがOSの良い代用評価項目であることを再確認した(表2、図2)。OSの代用評価項目としてRRの代わりにDCRを選択するする積極的理由はないと思われる。

本解析の限界

私たちは本研究の限界についてコメントしなければならない。代用評価項目を検証するためには、individual-levelとtrial-levelの両方の解析がなされることが理想であるとされていたが私たちはtrial-levelの解析しか行っていない。もっとも、最近の理解では、trial-levelの検証は

individual-levelの検証よりも重要であるとされている^{4, 6, 33)}。

結 語

私たちは化学療法未施行の進行/局所進行/転移性非小細胞肺癌患者に対して抗癌剤を投与した第III相RCTにおけるPFS/RR/DCRとOSとの間のtrial-levelの関連性を評価するため22709人のNSCLC患者を含む44のRCTのデータを用いてsystematic reviewを実施した。HRpfs、△PFS、ORrr、およびORdcrはすべて、HRosとの中等度のtrial-levelの関係を有していた。私たちのデータは、ORrrが最善のHRosの代用評価項目であり、HRpfsがそれに次ぐと判断された。しかしながら、これらの値からは、PFSおよびRRは十分な精度にてOSを予測できないと判断される。今後PPSが長くなるにつれて、PFSはさらにOSを反映しなくなるだろう³⁴⁾。

謝 辞

長年にわたるご指導を賜った呼吸器病学教室金子猛主任教授、ならびに、下記の論文の一部転載許諾を頂いたELSEVIERに感謝する。

Nakashima K, Horita N, Nagai K, Manabe S, Murakami S, Ota E, Kaneko T. Progression-Free Survival, Response Rate, and Disease Control Rate as Predictors of Overall Survival in Phase III Randomized Controlled Trials Evaluating the First-Line Chemotherapy for Advanced, Locally Advanced, and Recurrent Non-Small Cell Lung Carcinoma. *J Thorac Oncol.* 2016;11:1574-85.

文 献

- 1) Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: Global cancer statistics. *CA Cancer J Clin*, 2011; **61** (2) : 69 – 90.
- 2) Ghimire S, Kyung E, Kim E: Reporting trends of outcome measures in phase II and phase III trials conducted in advanced-stage non-small-cell lung cancer. *Lung*, 2013; **191** (4) : 313 – 319.
- 3) Fleming TR: Objective response rate as a surrogate end point: a commentary. *J Clin Oncol*, 2005; **23** (22) : 4845 – 4846.
- 4) Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *Int J Clin Oncol*, 2009; **14** (2) : 102 – 111.
- 5) George SL: Response rate as an endpoint in clinical trials. *J Natl Cancer Inst*, 2007; **99** (2) : 98 – 99.

- 6) Saad ED, Katz A, Hoff PM, Buyse M: Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. *Ann Oncol*, 2010; **21** (1) : 7 – 12.
- 7) Amit O, Bushnell W, Dodd L, Roach N, Sargent D: Blinded independent central review of the progression-free survival endpoint. *Oncologist*, 2010; **15** (5) : 492 – 495.
- 8) Broglio KR, Berry DA: Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst*, 2009; **101** (23) : 1642 – 1649.
- 9) Blumenthal GM, Karuri SW, Zhang H, et al: Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: US Food and Drug Administration trial-level and patient-level analyses. *J Clin Oncol*, 2015; **33** (9) : 1008 – 1014.
- 10) Hotta K, Suzuki E, Di Maio M, et al: Progression-free survival and overall survival in phase III trials of molecular-targeted agents in advanced non-small-cell lung cancer. *Lung Cancer*, 2013; **79** (1) : 20 – 26.
- 11) Hotta K, Kiura K, Fujiwara Y, et al: Role of survival post-progression in phase III trials of systemic chemotherapy in advanced non-small-cell lung cancer: a systematic review. *PLoS One*, 2011; **6** (11) : e26646.
- 12) Johnson KR, Ringland C, Stokes BJ, et al: Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol*, 2006; **7** (9) : 741 – 746.
- 13) Yamamoto N, Nambu Y, Fujimoto T, Koshiji M: A landmark point analysis with cytotoxic agents for advanced NSCLC. *J Thorac Oncol*, 2009; **4** (6) : 697 – 701.
- 14) Tsujino K, Kawaguchi T, Kubo A, et al: Response rate is associated with prolonged survival in patients with advanced non-small cell lung cancer treated with gefitinib or erlotinib. *J Thorac Oncol*. 2009; **4** (8) : 994 – 1001.
- 15) Booth A, Clarke M, Dooley G, et al: The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst Rev*. 2012; **1**: 2 .
- 16) Higgins PJ, Green S: Cochrane Handbook for Systematic Reviews of Interventions (Version5.1.0) . (2011) . Available from: http://handbook.cochrane.org/front_page.htm. (Accessed on 25th July 2015) .
- 17) Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1) . *Eur J Cancer*, 2009; **45** (2) : 228 – 247.
- 18) Sargent DJ, Patiyil S, Yothers G, et al: End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. *J Clin Oncol*, 2007; **25** (29) :4569 – 4574.
- 19) Sargent DJ, Wieand HS, Haller DG, et al: Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol*, 2005; **23** (34) : 8664 – 8670.
- 20) Ripley B. Package 'boot'.2015. Available from: <https://cran.r-project.org/web/packages/boot/boot.pdf#search='corr+boot+r'>. Accessed on Oct. 15th, 2015.
- 21) R documentation: Fitting Linear Models. Available from: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/lm.html>. Accessed on Oct. 1 st, 2015.
- 22) Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al: Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol*, 2008; **26** (12) : 1987 – 1992.
- 23) Fleming TR, Rothmann MD, Lu HL: Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol*, 2009; **27** (17) : 2874 – 2880.
- 24) Garon EB: Issues surrounding clinical trial endpoints in solid malignancies with a focus on metastatic non-small cell lung cancer. *Lung Cancer*, 2012; **77** (3) : 475 – 481.
- 25) Renfro LA, Shi Q, Xue Y, Li J, Shang H, Sargent DJ: Center-Within-Trial Versus Trial-Level Evaluation of Surrogate Endpoints. *Comput Stat Data Anal*, 2014; **78**: 1 – 20.
- 26) Reck M, von Pawel J, Zatloukal P, et al: Overall survival with cisplatin-gemcitabine and bevacizumab or placebo as first-line therapy for nonsquamous non-small-cell lung cancer: results from a randomised phase III trial (AVAiL) . *Ann Oncol*, 2010; **21** (9) : 1804 – 1809.
- 27) Xu JL, Jin B, Ren ZH, et al: Chemotherapy plus Erlotinib versus Chemotherapy Alone for Treating Advanced Non-Small Cell Lung Cancer: A Meta-Analysis. *PLoS One*, 2015; **10** (7) : e0131278.
- 28) Bhattacharya S, Fyfe G, Gray RJ, Sargent DJ: Role of sensitivity analyses in assessing progression-free survival in late-stage oncology trials. *J Clin Oncol*, 2009; **27** (35) : 5958 – 5964.
- 29) Chen TT, Chute JP, Feigal E, Johnson BE, Simon R: A model to select chemotherapy regimens for phase III

- trials for extensive-stage small-cell lung cancer. *J Natl Cancer Inst*, 2000; **92** (19) : 1601 – 1607.
- 30) Zia MI, Siu LL, Pond GR, Chen EX: Comparison of outcomes of phase II studies and subsequent randomized control studies using identical chemotherapeutic regimens. *J Clin Oncol*. 2005; **23** (28) : 6982 – 6991.
- 31) Lara PN Jr, Redman MW, Kelly K, Edelman MJ, Williamson SK, Crowley JJ, Gandara DR; Southwest Oncology Group: Disease control rate at 8 weeks predicts clinical benefit in advanced non-small-cell lung cancer: results from Southwest Oncology Group randomized trials. *J Clin Oncol*, 2008; **26**: 463 – 467.
- 32) Claret L, Gupta M, Han K, et al: Prediction of overall survival or progression free survival by disease control rate at week 8 is independent of ethnicity: Western versus Chinese patients with first-line non-small cell lung cancer treated with chemotherapy with or without bevacizumab. *J Clin Pharmacol*, 2014; **54** (3) : 253 – 257.
- 33) Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A: Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Control Clin Trials*, 2002; **23** (6) : 607 – 625.
- 34) Nakashima K, Horita N, Nagai K, et al: Progression-Free Survival, Response Rate, and Disease Control Rate as Predictors of Overall Survival in Phase III Randomized Controlled Trials Evaluating the First-Line Chemotherapy for Advanced, Locally Advanced, and Recurrent Non-Small Cell Lung Carcinoma. *J Thorac Oncol*, 2016; **11** (9) : 1574 – 1585.

Abstract

PROGRESSION-FREE SURVIVAL, RESPONSE RATE, AND DISEASE CONTROL RATE AS
PREDICTORS OF OVERALL SURVIVAL IN PHASE III RANDOMIZED CONTROLLED TRIALS
EVALUATING THE FIRST-LINE CHEMOTHERAPY FOR NON-CURABLE NON-SMALL CELL LUNG
CARCINOMA.

Nobuyuki HORITA

Yokohama City University Graduate School of Medicine, Department of Pulmonology
National Human Genome Research Institute, National Institutes of Health

An increasing number of recent randomized controlled trials for chemo naive advanced/locally-advanced/recurrent non-small cell lung cancer (NSCLC) adopts progression-free survival (PFS), response rate (RR), and disease control rate (DCR) as primary endpoint instead of overall survival (OS). However, whether PFS, RR, and DCR accurately reflect OS is still not clear. We conducted the first systematic review to evaluate the trial-level association between PFS/RR/DCR and OS in phase III RCTs for advanced/locally-advanced/metastatic NSCLC subjects treated with first-line chemotherapy using data from 44 RCTs involving 22709 NSCLC patients. Hazard ratio (HR), odds ratio (OR), and difference of median values (Δ) were used as summary estimates. All of HR_{PFS}, Δ PFS, OR_{RR}, and OR_{DCR} demonstrated moderate trial-level correlation coefficient, regression coefficient, and coefficient of determinant. Based on our analysis, OR_{RR} is the best surrogate endpoint for HR_{OS} followed by HR_{PFS}. Although our results are favorable for PFS and RR compared to previous studies, PFS and RR cannot be sufficiently reliable surrogate of OS. As OS prolongs, PFS is going to less reflect OS.

