独立行政法人国立病院機構
水戸医療センター

研 究 業 績 集

令和5年度

巻頭言

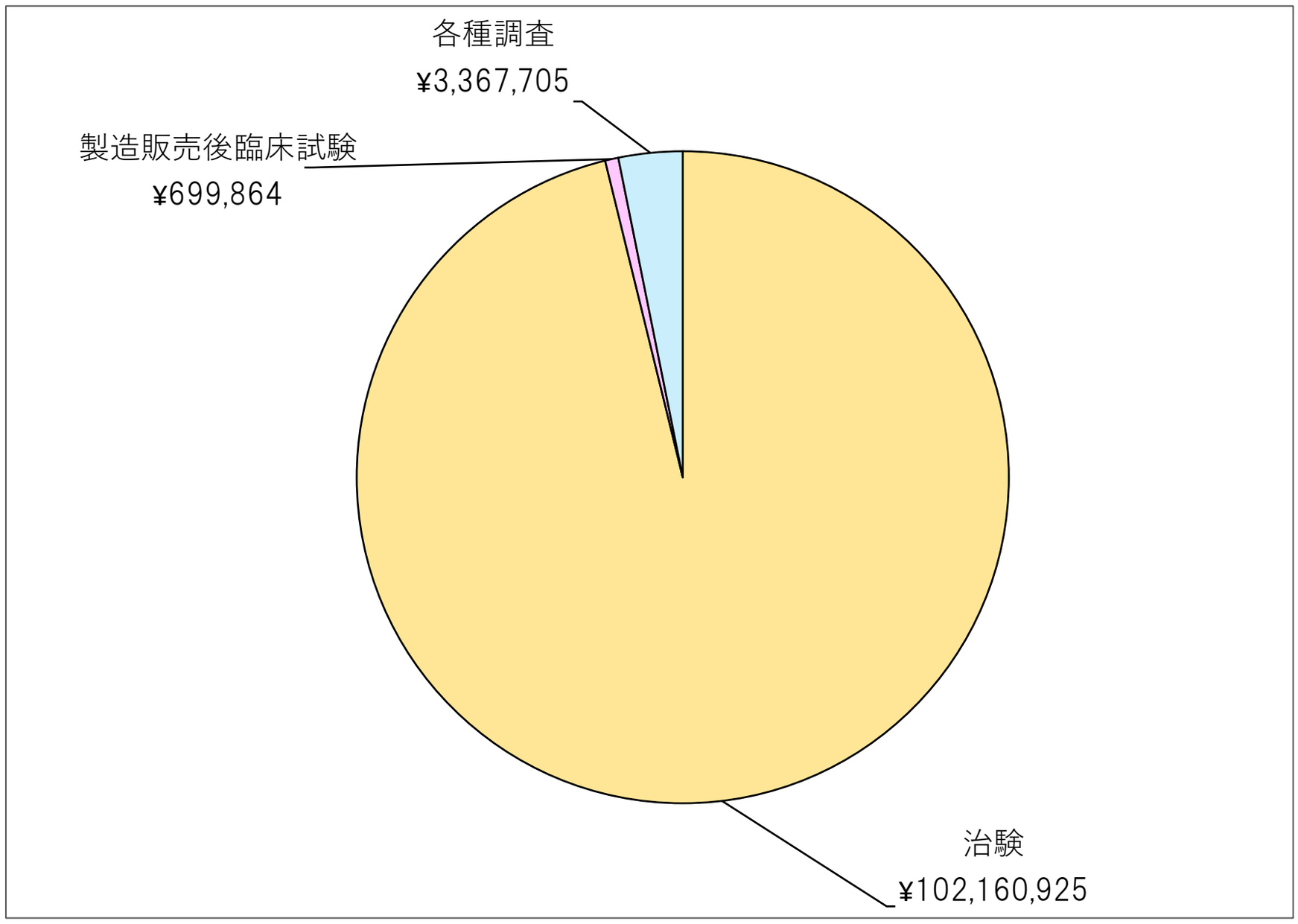
病院長　米野琢哉

水戸医療センターの基本方針として、「臨床研究を積極的に推進します」を掲げています。働き方改革で研究に対する時間を捻出するのが難しくなっておりますが、日常診療を黙々とこなし、臨床研究にも真摯に取り組んでいただきました。職員の皆様のご努力に敬意をはらいたいと思います。

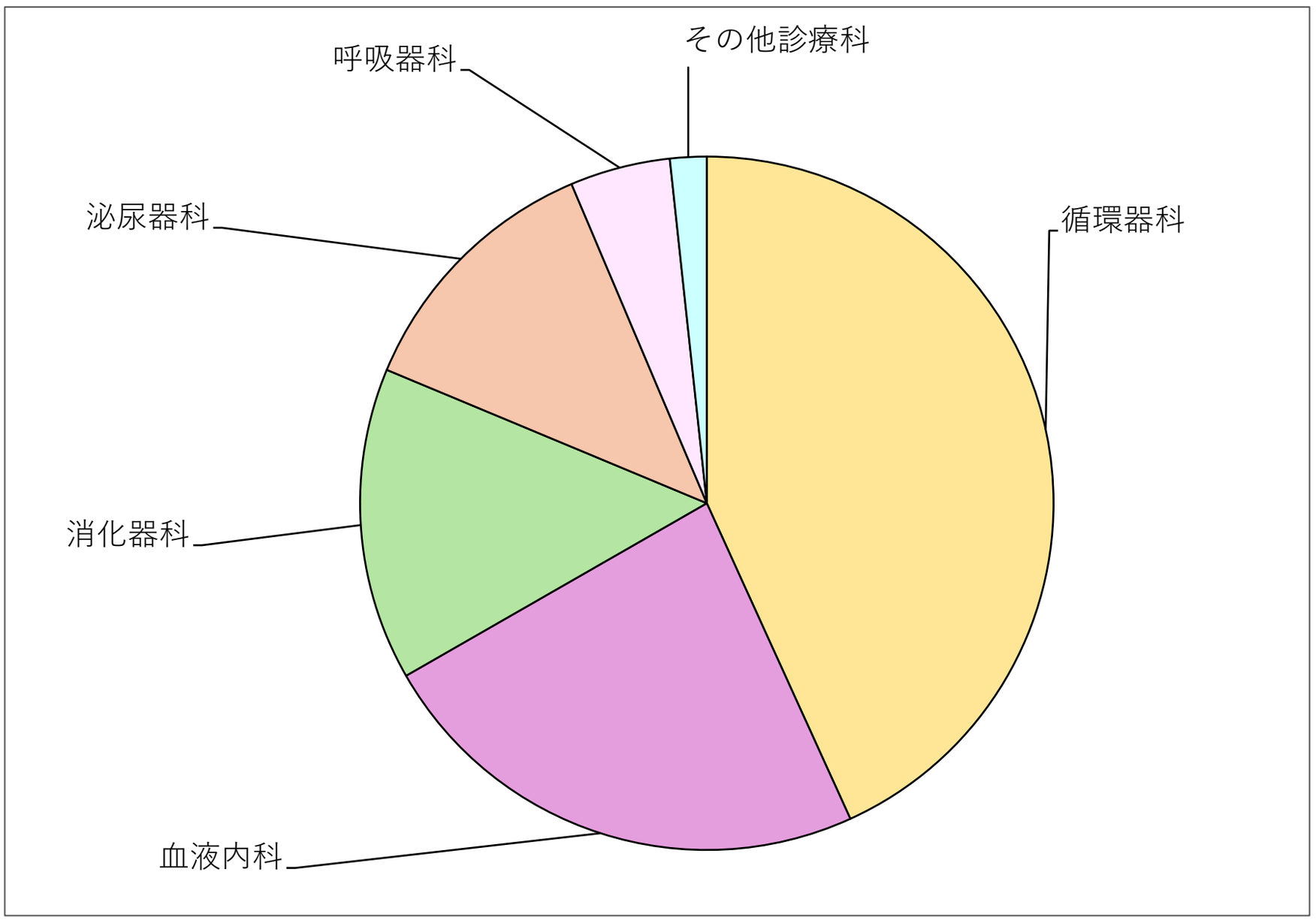
臨床研究の推進は、診療の質向上のためにも必須です。様々な職種の方々が研究に取り組むことによって、診療の活性化にもつながると期待しております。病院としても、臨床研究部を中心に資金供給・研究に必要な文書作成のアドバイス等、研究実施のサポートを継続していきます。今後も是非臨床研究にチャレンジしてみてください。

2023年度　　受託研究実績金額　10,622万円

契約種類別グラフ

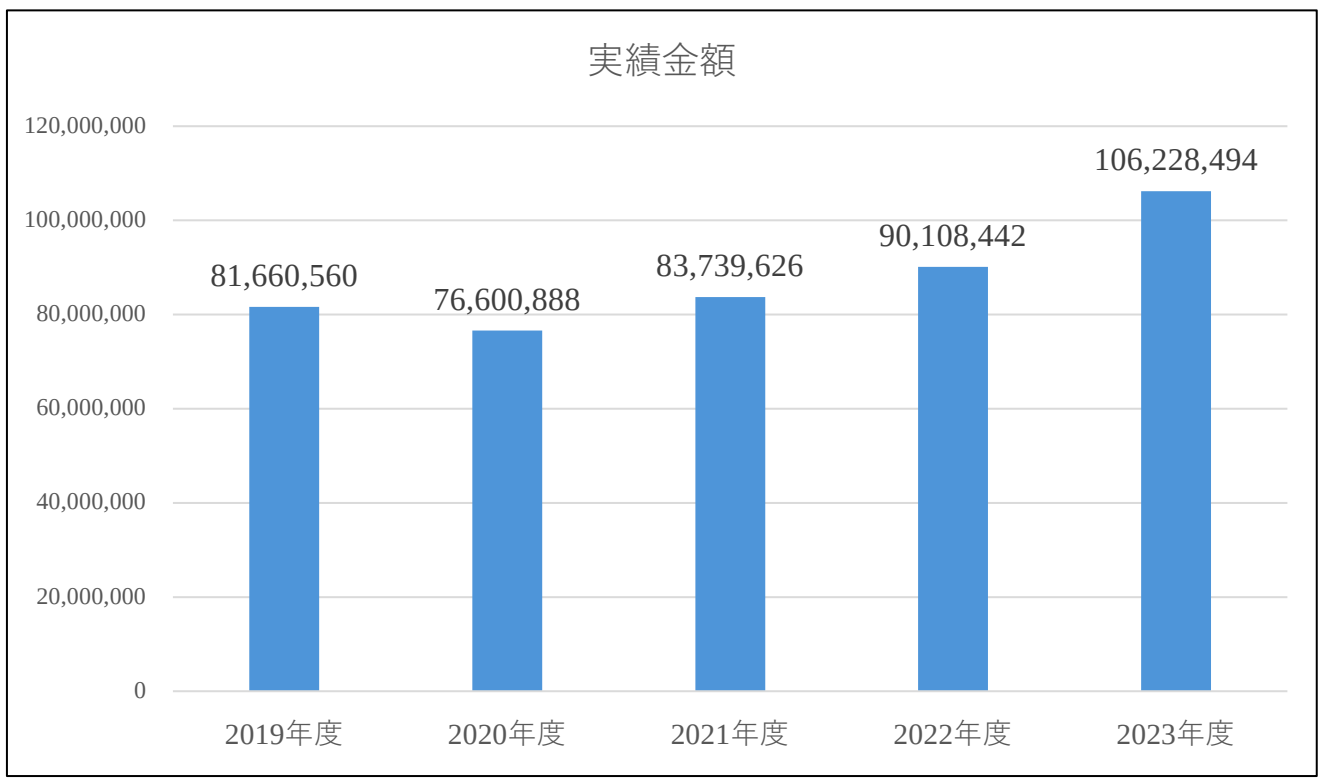


診療科別グラフ

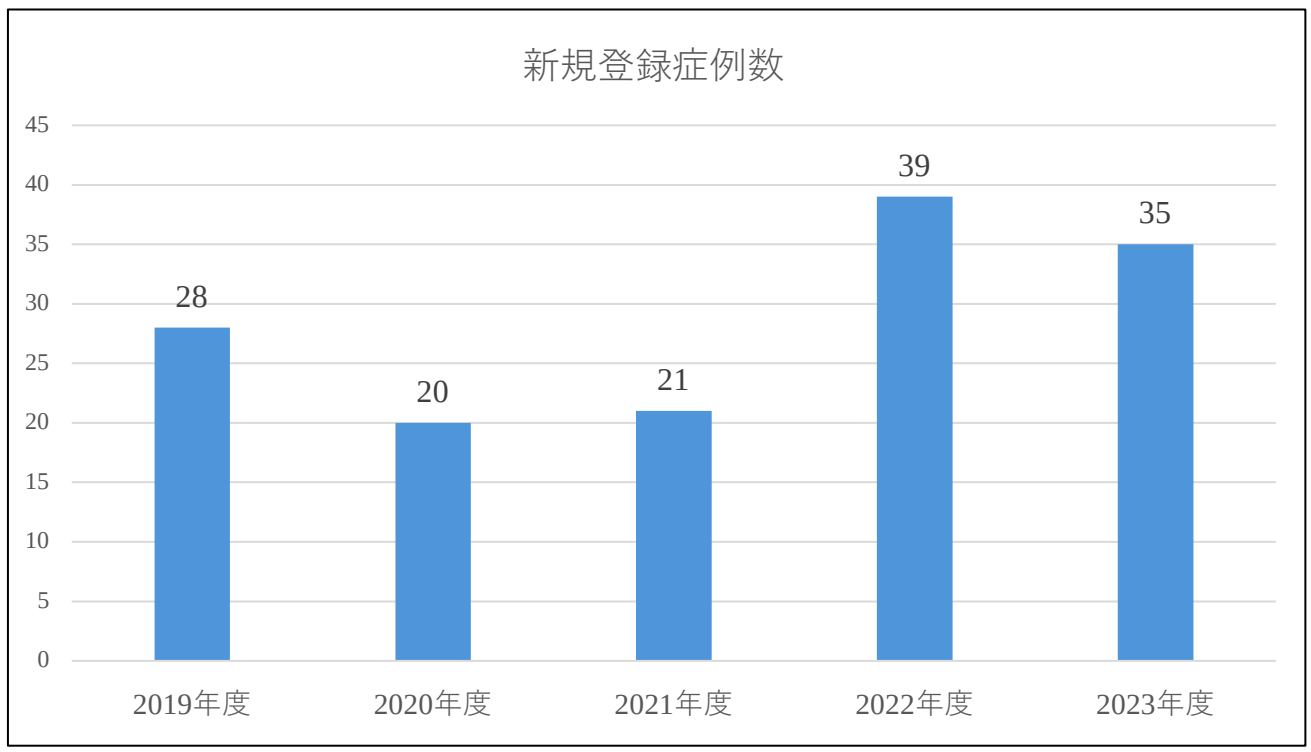


受託研究実績報告

①　受託研究実績金額 （治験･製造販売後臨床試験･製造販売後調査）



②　治験･製造販売後臨床試験　新規登録症例数



ＮＨＯネットワーク共同研究　新規症例登録数

研究責任者	研究課題名(採択番号)	研究代表者(施設名)	文書同意 有・無	当該施設 新規症例 登録数
武藤　亮	膵癌における腹腔洗浄細胞診を補完する新規バイオマーカーの確立に関する研究 （採択番号：R3-NHO(消化)-01）	末永 雅也 （名古屋医療センター）	有	5
福本　英樹	DOAC服用患者における抜歯の安全性の確立に関する研究:ガイドライン確立のための多施設共同前向き研究 （採択番号：R3-NHO(他研)-01）	吉川 博政 （九州医療センター）	有	5
加藤　徳之	急性期BAD型脳梗塞に対する抗血栓療法の種類と神経学的予後に関する前向き探索研究 （採択番号：R4-NHO(心脳)-01）	土井 健人 （京都医療センター）	有	2

競争的研究費

項目	研究課題名	研究者名	研究事業者名 (依頼者名)	主任 分担	研究費 受領日	研究費 単位：円
科学研究費助成事業 (学術研究助成基金助成金)	医師の病院前診察における網羅的文献データベース構築とエビデンス診療ギャップの解明 (21K10386)	堤　悠介	水戸医療センター (堤　悠介)	主任	R5.4.10	1,365,000
科学研究費助成事業 (学術研究助成基金助成金)	包括的外傷長期予後データベースを用いたテーラーメイド型社会復帰支援システムの確立 (22K10476)	堤　悠介	東海大学 (土谷飛鳥)	分担	R5.8.10	65,000
科学研究費助成事業 (学術研究助成基金助成金)	病院前輸血療法における全国悉皆的疫学調査と最適病院前輸血療法戦略の構築 (23K09584)	堤　悠介	東海大学 (三浦直也)	分担	R5.8.10	39,000
厚生労働科学研究費	HAMならびに類縁疾患の患者レジストリによる相談機能の強化と診療ガイドラインの改訂 (22FC1013)	湯沢賢治	聖マリアンナ医科大学医学研究科 (山野嘉久)	分担	R5.8.18	200,000
日本医療研究開発機構研究費	HAM・HTLV-1陽性難治性疾患の患者レジストリ活用によるエビデンス創出 (23ek0109529s0303)	湯沢賢治	聖マリアンナ医科大学医学研究科 (山野嘉久)	分担	R5.8.30	325,000
科学研究費助成事業 (学術研究助成基金助成金)	文献レジストリ構築とリアルワールドデータによる膠原病予後因子の網羅的負荷推計 (22K10423)	堤　悠介	昭和大学 (辻本　康)	分担	R5.9.7	65,000

英文論文

No.	タイトル	著者	ポイント
1	Epidemiology of post-suboccipital craniotomy headache: A multicentre retrospective study	山崎友郷	4.200
2	Venetoclax plus low-dose cytarabine in patients with newly diagnosed acute myeloid leukemia ineligible for intensive chemotherapy: an expanded access study in Japan	吉田近思	4.900
3	Correlations between 3D preoperative planning and postoperative reduction in the osteosynthesis of distal humeral fractures	小川　健	5.800
4	Clinical benefit of platinum doublet combination therapy in older adults with advanced non-small cell lung cancer: A prospective multicenter study by the National Hospital Organization in Japan	遠藤健夫	5.300
5	Publication hyper-inflation in the field of intensive care	堤　悠介	29.100
6	Prehospital shock index predicts 24-h mortality in trauma patients with a normal shock index upon emergency department arrival	堤　悠介	5.700
7	Survival Impact of Second-Line Immune Checkpoint Inhibitors in Older Patients With Advanced Squamous-Cell NSCLC: Post Hoc Analysis of the CAPITAL Study	遠藤健夫	6.000
8	Twenty-year follow-up of promising clinical studies reported in highly circulated newspapers: a meta-epidemiological study	堤　悠介	7.100
9	Prospective exosome-focused translational research for afatinib (EXTRA) study of patients with nonsmall cell lung cancer harboring EGFR mutation: an observational clinical study	遠藤健夫	7.300
10	PREFACE	湯沢賢治	1.400
11	Follow-up focused on psychological intervention initiated after intensive care unit in adult patients and informal caregivers: a systematic review and meta-analysis	堤　悠介	2.150
12	An open competition involving thousands of competitors failed to construct useful abstract classifiers for new diagnostic test accuracy systematic reviews	堤　悠介	8.000

No.	タイトル	著者	ポイント
13	Variations in S100A8/A12 Gene Expression Are Associated with the Efficacy of Nintedanib and Acute Exacerbation Development in Idiopathic Pulmonary Fibrosis Patients	箭内英俊	3.950
14	Ultra-early rt-PA administration should improve patient outcome on mechanical thrombectomy: Post hoc analysis of SKIP	加藤徳之	6.600
15	Pathophysiology of sex difference in refractoriness in lateral epicondylitis: Biomechanical study of wrist torque	小川　健	5.100
16	Atezolizumab Monotherapy for Non-small Cell Lung Cancer Patients: An Observational Study in Ibaraki Group (ATTENTION-IBARAKI)	沼田岳士	4.800
17	The impact of SAH finding on CT to the clinical outcome after mechanical thrombectomy for large vessel occlusion	加藤徳之	6.600
18	Atezolizumab for EGFR-mutated Non-small Cell Lung Cancer Patients: An Observation Study in Ibaraki Group (ATTENTION-IBARAKI)	沼田岳士	4.600
19	Japanese Clinical Practice Guidelines for Rehabilitation in Critically Ill Patients 2023 (J-ReCIP 2023)	堤　悠介	2.900
20	Development and evaluation of a rapid one-step high sensitivity real-time quantitative PCR system for minor BCR-ABL (e1a2) test in Philadelphia-positive acute lymphoblastic leukemia (Ph plus ALL)	吉田近思	4.900
21	Female and preserved platelet count subgroups of myelodysplastic syndrome patients benefit from standard-dose azacitidine	吉田近思	4.500
22	Small Intestinal Adenocarcinoma Arising at the Anastomotic Site after Kasai Operation for Biliary Atresia: A Case Report and Literature Review	小林仁存	1.850
23	Detection of Factors Related to the Development of Osteochondritis Dissecans in Youth Baseball Players Screening	小川　健	6.000
24	The Effect of Axial Traction MRI on the Articular Cartilage Visibility in Thumb Carpometacarpal Arthritis	小川　健	4.000

No.	タイトル	著者	ポイント
25	Effect of Preoperative Oral Antibiotics and Mechanical Bowel Preparations on the Intestinal Flora of Patients Undergoing Laparoscopic Colorectal Cancer Surgery: A Single-Center Prospective Pilot Study	伊瀬谷和輝	4.000
26	Prognostic Impact of Preoperative Assessment of Muscle Mass and Strength in Surgically Resected Lung Cancer	栗原秀輔	11.200
27	Optimal Limb Position for the Stress Ultrasound Evaluation of Elbow Valgus Laxity in Baseball Players	小川　健	5.400
28	Proton Pump Inhibitors and Cyclin-Dependent Kinase 4/6 Inhibitors in Patients With Breast Cancer	小坂真吉	7.800
29	Exploring the relationship between plasma substance P and glottal incompetence in the elderly	瀬成田雅光	6.400
30	Is a Novel Fluoroscopic Intraoperative Reference System Superior to Conventional Management for Distal Radius Fracture Reduction? A Propensity-matched Comparative Study	小川　健	7.200
31	Intoxication with massive doses of amlodipine and candesartan requiring venoarterial extracorporeal membrane oxygenation	堤　悠介	1.750
32	Excess mortality in COVID-19-affected solid organ transplant recipients across the pandemic	湯沢賢治	11.900

和文論文

No.	論文名	著者	ポイント
1	2015年9月関東・東北豪雨での茨城県常総水害について	安田 貢	1.000
2	周術期トリプルネガティブ乳癌に対する免疫チェックポイント阻害剤の使用経験	小坂真吉	1.500
3	令和5年度　認定HLA　検査技術者認定制度試験問題に関する報告	湯沢賢治	1.000
4	Nephron mass定量化による生体腎移植後グラフト機能予測	湯沢賢治	1.000
5	【上肢の骨壊死疾患治療～最新の知見～】2) Kienböck病の病態について	小川 健	1.500
6	【上肢の骨壊死疾患治療-最新の知見-】Kienboeck病に対する自己骨髄血移植治療	小川 健	1.500
7	【局所麻酔で行える手外科手術のコツとピットフォール】母指手根中手(CM)関節症に対する関節鏡視下滑膜切除術	小川 健	1.500
8	上腕骨遠位端骨折に対するA.L.P.S.Elbow Plating SystemTMの治療成績	小川 健	1.500
9	頸部固定用補助具の背景抑制広範囲拡散強調MRIでの有効性評価	金居啓介	1.000

国際学会

学会名	塩題名	演者名	発表年月日
The 13th JSH International Symposium 2023 in Tsukuba	Asciminib for Chronic Myeloid Leukemia Patients Who Are Intolerance to TKI Treatment	橋川　諒	2023/6/21
American Heart Association Scientific Session 2023, Philadelphia, USA	Differences in Erythrocyte Morphology in Thrombus From Infarct-Related Artery: Cardioembolic Thrombosis vs. Atherothrombosis, Pathological Analyses From MITO Study	小泉智三	2023/11/11
The 88th Annual Scientific Meeting of the Japanese Circulation Society	A Case of Acute Coronary Syndrome Due to Organic Stenosis and Coronary Spasm	茂木奈穂	2024/3/10
European Congress of Radiology 2024, Vienna, Austria	The Utility of Maximum Intensity Projection Images in Non-Enhanced CT for Detecting the Hyperdense Cerebral Artery Sign in Acute Thromboembolic Ischemic Stroke	金居　啓介	2024/3/1

国 内 学 会

学会名	演題名	演者名	発表年月日
第85回日本血液学会学術集会	Asciminib in chronic phase chronic myeloid leukemia: A single center experience	橋川　諒	2023/10/15
第85回日本血液学会学術集会	FLT3/TKD mutations in patients with acute myeloid leukemia: HM-SCREEN-Japan 02 study	堤　育代	2023/10/15
第693回日本内科学会関東地方会	梅毒関連の溶血性貧血を生じた1例	仲野谷　純	2024/2/10
The 88th Annual Scientific Meeting of the Japanese Circulation Society	A Case of Bevacizumab as a Suspected Cause of Angina Pectoris Due to Cronary Microvascular Dysfunction	黒田 裕和	2024/3/9
The 88th Annual Scientific Meeting of the Japanese Circulation Society	A Case of Developed a Systemic Embolism while Taking Warfarin for Atrial Fibrillation and Performed Thoracoscopic Left Atrial Appendage Exclusion	鈴木健太	2024/3/10
The 88th Annual Scientific Meeting of the Japanese Circulation Society	Autopsy Case of Acute Myocardial Infarction with Cardiac Rupture	鮎澤 祥吾	2024/3/9
The 88th Annual Scientific Meeting of the Japanese Circulation Society	A Case of the Patient who Underwent Complete Revascularization for Severe Multi-vessel Coronary Disease after Vf Resuscitation Triggered by STE-ACS	丸田　俊介	2024/3/8
第63回日本呼吸器学会学術講演会	肺小細胞癌に対するデュルバルマブ投与後にirAEによる自己免疫性脳炎を発症した一例	山崎健斗	2023/5/29
第254回日本呼吸器学会関東地方会	尿中抗原検査が陰性であったが、臨床的にレジオネラ肺炎を疑って治療し、救命し得た一例	櫻井優樹	2023/5/13
第257回日本呼吸器学会関東地方会	ECMO と CHDF を用いて救命しえた重症レジオネラ肺炎の一例	宮坂直樹	2023/11/11
第258回日本呼吸器学会関東地方会	気管支肺胞洗浄液中で好酸球増多を認めた過敏性肺炎様の病像を呈した 1 例	松下祐真	2024/2/17
第223回茨城内科学会	尿中抗原検査が陰性であったが、臨床的にレジオネラ肺炎を疑って治療し、救命し得た一例	藤田弘輝	2023/6/17
第224回茨城内科学会	気管支肺胞洗浄液（BAL）で好中球の有意な増多を認めたが，急性好酸球性肺炎の病像を呈した一例	横瀬直希	2023/10/15
第225回茨城内科学会	Nivolumab 投与後に免疫関連細気管支炎を発症した 1 例	小島原史大	2024/3/16
第20回県南チェストカンファレンス	進展型小細胞肺癌に対する治療～イミフィンジ3年フォローアップ解析結果から考える～	箭内英俊	2023/4/4
Lung Cancer Expert Symposium in Mito	Ⅳ期日扁平上皮非小細胞肺癌における個別化治療を考える	沼田岳士	2023/5/17
第3回Lung Cancer Expert Symposium in Mito	「進行・再発非小細胞肺癌に対するアテゾリズマブの使用経験について～茨城県後ろ向き観察研究『ATTENSION-IBARAKI』を踏まえて～」	沼田岳士	2023/12/6
第8回桜の郷チェストカンファレンス	たかが咳、されど咳－慢性咳嗽診療 最近の話題－	遠藤健夫	2023/11/22

学会名	演題名	演者名	発表年月日
好酸球性重症喘息を考える会	好酸球性重症喘息における生物学的製剤	箭内英俊	2023/12/20
AZ Immuno-Oncology Online Meeting	進行期NSCLCの治療戦略～当院のPOSEIDON使用経験～	羽鳥貴士	2024/2/19
水戸予防医学セミナー2023	高齢者の肺炎予防について～肺炎球菌ワクチンを中心に～	遠藤健夫	2023/4/19
第140回ひたちなか市胸部疾患カンファレンス	COPDについて考える～最新の知見を踏まえて～	山崎健斗	2023/8/24
令和5年度アレルギー疾患医療拠点病院事業　住民向け講演会	喘息・アレルギー性鼻炎	遠藤健夫	2023/12/9
日本消化器病学会第377回関東支部例会	潰瘍性大腸炎に対してベドリズマブ導入後に生じた食道潰瘍の一例	小野田翼	2023/12/9
第20回日本消化管学会総会学術集会	潰瘍性大腸炎術後の慢性回腸嚢炎における内視鏡表現型頻度と生物学的製剤使用に関する検討	小野田翼	2024/2/9
第694回日本内科学会関東地方会	抜歯を契機に発症した多発肝膿瘍、肺膿瘍の1例	安部計雄	2024/3/16
第225回茨城内科学会	18年の時を経て再燃した自己免疫性肝炎の1例	鈴木健太	2024/3/16
第694回内科学会関東地方会	潜在的に抗AchR抗体を有し，免疫チェックポイント阻害薬投与後にir AEとして発症した重症筋無力症の1例	相澤　哲史	2024/3/16
第59回日本小循環器学会総会・学術集会	茨城県における学校BLS教育の強化～地域悉皆教育～	安田　貢、吉澤あずさ	2023/7/6
第47回茨城県救急医学会	救急医療における介護。医療連携とACP	安田　貢	2023/9/9
第40回　日本呼吸器外科学会学術集会	術前GNRIとPNIは根治的肺葉切除術を施行した非小細胞肺癌の予後予測因子となるか	栗原秀輔	2023/7/14
第76回　日本胸部外科学会定期学術集会	骨格筋減少と非小細胞肺癌切除例の予後との関係	栗原秀輔	2023/10/19
第253回　茨城外科学会	低侵襲手術で治癒が得られた魚骨による小腸穿孔・膿瘍形成の1例	成田保和	2023/10/15
第15回 日本Acute care surgery学会	外腸骨静脈損傷を伴った杙創による直腸穿孔の1例	成田保和	2023/10/6
第60回　日本腹部救急医学会	十二指腸に嵌頓した胆石イレウスの1例	成田保和	2024/3/21
第85回　日本臨床外科学会	横行結腸の双孔式人工肛門が重積嵌頓し,緊急で人工肛門を造設した1例	伊瀬谷和輝	2023/11/18
第77回　日本食道学会学術集会	胸腔鏡下食道切除後に遅発性の心嚢内出血による心タンポナーデをきたした1例	福富俊明	2023/6/30

学会名	演題名	演者名	発表年月日
第78回　日本消化器外科学会総会	腰ヘルニアに対して異なるアプローチ法で修復術を行った3例	福富俊明	2023/7/13
第85回　日本臨床外科学会	高度な混合性換気障害を有するupside down stomachを呈した食道裂孔ヘルニアに対して腹腔鏡下修復術を施行した2例	福富俊明	2023/11/17
第31回　日本乳癌学会学術総会	乳腺外科医が不足する地域において、医療の質を低下させず医療を継続するためには何が必要か	森千子	2023/7/1
第19回日本乳癌学会関東地方会	キイトルーダを投与した周術期トリプルネガティブ乳癌の検討	小坂真吉	2023/12/2
第21回日本乳癌学会東北地方会	ペンブロリズマブを投与した周術期トリプルネガティブ乳癌の検討	小坂真吉	2024/3/2
第40回　日本呼吸器外科学会学術集会	術前筋量と筋力を評価した肺癌手術症例の検討	中村亮太	2023/7/14
第76回　日本胸部外科学会定期学術集会	肺癌周術期体組成変化の検討	中村亮太	2023/10/19
第78回　日本消化器外科学会総会	幽門側胃切除術後再建方法と骨格筋減少の関連	米山智	2023/7/14
第36回　日本内視鏡外科学会総会	大腿-大腿動脈バイパス術後の鼠径ヘルニアに対してSelf-Fixating Meshを用いて腹腔鏡下手術を行った一例	米山智	2023/12/9
国立病院総合医学会	脳血栓回収療法実施医、脳血管内治療医を育てる	加藤徳之	2023/10/20
第82回日本脳神経外科総会	ビデオシンポジウム「最適な治療を提供するための脳血管造影と脳動脈瘤コイル塞栓術」	佐藤允之	2023/10/25
第39回日本脳神経血管内治療学会学術集会	脳血管内治療周術期の非血栓性遠位塞栓	加藤徳之	2023/11/23
第39回日本脳神経血管内治療学会学術集会	血栓回収療法におけるT1脂肪抑制造影3D画像を用いた血栓長と閉塞後分岐血管の評価	佐藤允之	2023/11/23
STROKE24	三叉神経痛および歩行障害で発症したテント状硬膜動静脈瘻の一例	丸山沙彩	2024/3/8
STROKE24	救急隊による病院評価スケールで失語症と診断された主幹動脈閉塞患者の機能転帰	佐藤允之	2024/3/9
第67回日本透析医学会学術集会・総会	長期透析患者における腎移植成績～我が国の腎移植統計から～	湯沢賢治	2023/6/17
第49回日本骨折治療学会	橈骨遠位端骨折における一時的創外固定の有用性についての検討	森田純一郎	2023/6/30
第38回日本整形外科学会基礎学術集会	骨盤screw挿入に関する新たな手技の提案	森田純一郎	2023/10/19
第49回日本骨折治療学会	模擬骨を用いた骨折手術時透視教育のニーズについて検討	森田純一郎	2023/6/29

学会名	演題名	演者名	発表年月日
第49回日本骨折治療学会	茨城県における上肢外傷に対する上肢外傷担当医とハンドセラピストの連携	森田純一郎	2023/6/29
第49回日本骨折治療学会	ゲームエンジンを用いたVRプレート術前計画手法の新規開発	森田純一郎	2023/7/1
第49回日本骨折治療学会	大腿骨頚部骨折に対する内固定術 - Femoral Neck SystemとHansson Pinの臨床成績の検討 -	森田純一郎	2023/7/1
第32回　日本脊椎インストゥルメンテーション学会	セメント注入型椎弓根スクリューにおける椎体内に拡がるセメントの特徴	江藤文彦	2023/11/25
第38回　日本整形外科学会基礎学術集会	肘内側側副靱帯（UCL）のMRI所見 ー野球肘内側障害の病態を考察するー	小川健	2023/10/19
第34回　日本臨床スポーツ医学会学術集会	茨城県内陸上競技大会における傷病調査　COVID-19流行前後の比較	小川健	2023/11/13
第34回　日本臨床スポーツ医学会学術集会	茨城県における陸上競技医務サポートの黎明と発展	小川健	2023/11/13
第34回　日本臨床スポーツ医学会学術集会	大学野球選手における原テストと主観的パフォーマンススコアとの関係	小川健	2023/11/12
第34回　日本臨床スポーツ医学会学術集会	大学野球投手における投球による肘関節外反弛緩性の増大と投球時最大内反トルクとの関係	小川健	2023/11/12
第38回日本整形外科学会基礎学術集会	野球選手の投球時肩痛発症予測に有用なメディカルチェック項目の検討	小川健	2023/10/19
第38回日本整形外科学会基礎学術集会	野球選手の投球時肩痛発症予防のためにメディカルチェックで介入を要する肩関節回旋角度の検討	小川健	2023/10/19
第49回日本骨折治療学会学術集会	上腕骨遠位端骨折に対する3D術前計画の臨床的意義　従来法との比較	小川健	2023/6/29
第66回日本手外科学会学術集会	母指CM関節症に対する関節鏡下滑膜切除術の有用性	小川健	2023/4/20
第66回日本手外科学会学術集会	DDA創外固定器を用いて治療したPIP関節内(中節骨基部)骨折の治療成績	小川健	2023/4/20
第38回日本整形外科学会基礎学術集会	野球肘の診療・研究の最前線　肘内側側副靱帯(UCL)のMRI所見 野球肘内側障害の病態を考察する	小川健	2023/10/19
第66回日本手外科学会学術集会	キーンベック病(Stage3A,B)に対し骨釘移植を追加した骨髄血移植・創外固定・低出力超音波併用療法の治療成績	小川健	2023/4/20
第66回日本手外科学会学術集会	キーンベック病の治療～原点と挑戦～　キーンベック病の原点「壊死骨再生への挑戦」 骨髄血移植・創外固定・低出力超音波併用治療	小川健	2023/4/20
第49回日本骨折治療学会学術集会	上腕骨遠位部骨折に対する3D術前計画の再現性評価	小川健	2023/6/29
第38回日本整形外科学会基礎学術集会	肘関節離断性骨軟骨炎に対する三次元MRI・CT合成画像を用いた術前評価と手術計画	小川健	2023/10/19

学会名	演題名	演者名	発表年月日
第38回東日本手外科研究会	尺骨神経皮下前方移動術後に再手術を要したStruthers arcadeによる肘部管症候群の一例	小川健	2024/2/3
第36回日本肘関節学会学術集会	上腕骨外側上顆炎に対し2回の手術で効果不十分であったが、最終的に除神経術が奏功した一例	小川健	2024/3/1
第48回日本足の外科学会学術集会	痙性尖足に対する鏡視下腓腹筋膜切離術の経験	大山和生	2023/10/27
第36回日本四肢再建・創外固定学会学術集会	両側対称性に発症した慢性再発性多発性骨髄炎（CRMO）による外反膝変形の 1 例	大山和生	2023/7/14
第36回日本四肢再建・創外固定学会学術集会	外傷後成長軟骨障害により生じた下肢変形の病態と治療	大山和生	2023/7/15
第21回茨城形成外科研究会	外傷性動静脈瘻からの出血により眼球圧迫を呈した神経線維腫症Ⅰ型の1例	笠井丈博	2023/6/9
第110回日本泌尿器科学会総会	Pharmacogenomic features on clinical outcomes in patients with castration-sensitive and -resistant prostate cancer who received abitraterone acetate	高橋佳子	2023/4/21
第36回日本内視鏡外科学会総会	腎尿管全摘術と同時に行ったロボット支援膀胱全摘術の2例	高橋佳子	2023/12/8
第27回秋田腎不全研究会	両側上部尿路腫瘍摘出後に透析を導入した3例	高橋佳子	2023/12/17
第127回日本泌尿器科学会茨城地方会	両側同時性に発生した尿管癌に対して尿管部分切除・Boari手術を行い、右腎機能を温存した1例	高橋祥太	2023/10/15
第128回日本泌尿器科学会茨城地方会	PSA低値、CEA・CA19-9高値を示してドセタキセルが奏功した1例	高橋祥太	2024/2/17
第110回日本泌尿器科学会総会	当院におけるMRI/US融合画像下標的生検の初期成績	飯沼昌宏	2023/4/20
第37回日本泌尿器内視鏡・ロボティクス学会	PI-RADS3病変におけるprostate health indexの有用性について	飯沼昌宏	2023/11/11
第36回日本内視鏡外科学会総会	当院における80歳以上の高齢者に対するロボット支援膀胱全摘術の5例	飯沼昌宏	2023/12/8
第88回日本泌尿器科学会東部総会	当院における上部尿路癌に対するエンホルツマブベドチンの初期治療経験	飯沼昌宏	2023/10/6
第36回日本老年泌尿器科学会	当院における80才以上の高齢者における根治的膀胱全摘所術の治療成績	飯沼昌宏	2023/5/26
第36回日本口腔診断学会、第33回日本口腔内科学会、第43回歯科薬物療法学会、第32回日本口腔感染症学会　4学会合同学術大会	顎矯正手術後に生じた上顎前歯部surgical ciliated cystの１例	福本秀樹	2023/9/23
第36回日本口腔診断学会、第33回日本口腔内科学会、第43回歯科薬物療法学会、第32回日本口腔感染症学会　4学会合同学術大会	MRONJによる腐骨除去後の下顎骨内に乳癌転移を認めた１例	高野智穂	2023/9/23
第６８回（公社）日本口腔外科学会総会・学術大会	国立病院機構水戸医療センターにおける顎矯正手術症例の臨床的検討	岩崎敬大	2023/11/11

学会名	演題名	演者名	発表年月日
第47回茨城県救急医療学会	当院救命救急センターにおけるERダイヤリーの効果について -ERダイヤリー活用後のアンケート調査から家族の思いを知る-	住谷いずみ	2023/9/9
第39回 日本診療放射線技師学術大会	放射線診断における遠隔3D画像作成運用の可能性について	金居　啓介	2023/9/30
第77回国立病院総合医学会	re-entryを認めないstanfordA型大動脈解離の一症例	國嶋雄太	2023/10/21
第51回国臨協関信支部学会	左房前後径と左房容積係数の比較検討	渡邉隼	2023/9/2
第77回 国立病院総合医学会	入院中にCOVID-19に感染したギラン・バレー症候群患者に介入した症例	井口朋重	2023/10/20
第77回 国立病院総合医学会	示指・中指・環指デグロービング損傷後、複数のスプリントを用いてADLの獲得に繋がった一症例	金田玲央	2023/10/20
第77回　国立病院総合医学会	血栓回収術を施行された脳梗塞患者の再開通時間と認知・高次脳機能予後の関連	辻健太	2023/10/21

令和5年度院内臨床研究課題

研究代表者	課題名	配分額
神経内科 相澤哲史	消化器癌および泌尿器癌患者における免疫関連有害事象（irAE）発現に影響を与える因子を探索する後ろ向き症例集積研究	360,000
救急科 堤　悠介	論文出版における地域格差の検証：メタ疫学研究	550,000
外科 中村亮太	肺癌手術症例における、周術期骨格筋減少部位の検討	600,000
外科 米山　智	ロボット支援下直腸手術の導入・手技習得・教育にむけた臓器モデルの有用性について	650,000
脳神経外科 山崎友郷	脳主幹動脈閉塞に対する急性期再開通療法におけるcombined techniqueの臨床転帰に与える影響　後ろ向き症例　集積研究	500,000
脳神経外科 佐藤允之	脳血管造影検査の手技到達度評価の確立	470,000
脳神経外科 丸山沙彩	慢性硬膜下血種の再発例に対する塞栓術の有用性と安全性、再再発に関連する因子を探索する前向き症例　観察研究	450,000
整形外科 小川　健	手指化膿性疾患治療に関する後ろ向き症例集積研究	500,000
整形外科 江藤文彦	高エネルギー外傷に伴う胸腰椎椎体骨折に対する安静臥床期間を設けない保存治療に関する前向き観察研究	800,000
整形外科 森田純一郎	透視装置と複合現実デバイスを用いた簡易ナビゲーションシステムの有効性を検討する観察研究	850,000
放射線科 金居啓介	急性期脳梗塞における最大値投影画像用いた頭部単純ＣＴの有用性	120,000
放射線科 杉原理菜	外傷に対する３ＤＣＴを用いた骨折の初期評価法	200,000
放射線科 田中善啓	気管支動脈塞栓術に対する仮想透視画像の有用性	80,000
栄養管理室 味原　稔	胃がん患者周術期及び術後の栄養評価と体組成変化について	400,000
看護部 小川直子	食道癌患者への動画による経腸栄養管理指導の検討	610,000

水戸医療センター

令和5年度　代表的論文

JJCO Japanese Journal of Clinical Oncology

Japanese Journal of Clinical Oncology, 2023, 53(7), 595–603
https://doi.org/10.1093/jjco/hyad027
Advance Access Publication Date: 4 April 2023
Original Article

OXFORD

Original Article

Venetoclax plus low-dose cytarabine in patients with newly diagnosed acute myeloid leukemia ineligible for intensive chemotherapy: an expanded access study in Japan

Noboru Asada[1,*], Jun Ando[2], Satoru Takada[3], Chikashi Yoshida[4], Kensuke Usuki[5], Atsushi Shinagawa[6], Kenichi Ishizawa[7], Toshihiro Miyamoto[8,†], Hiroatsu Iida[9], Nobuaki Dobashi[10], Sumiko Okubo[11], Hideyuki Honda[12], Tomomi Soshin[12], Yasuko Nishimura[12], Atsuko Tsutsui[12], Harumi Mukai[13,‡] and Kazuhito Yamamoto[14]

[1]Department of Hematology and Oncology, Okayama University Hospital, Okayama, Japan, [2]Department of Hematology, School of Medicine, Juntendo University, Tokyo, Japan, [3]Leukemia Research Center, Gunma Saiseikai Maebashi Hospital, Maebashi, Japan, [4]Department of Hematology, National Hospital Organization Mito Medical Center, Ibaraki-machi, Japan, [5]Department of Hematology, NTT Medical Center Tokyo, Tokyo, Japan, [6]Department of Internal Medicine, Hitachi General Hospital, Hitachi, Japan, [7]Department of Internal Medicine III, Yamagata University Faculty of Medicine, Yamagata, Japan, [8]Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan, [9]Department of Hematology, National Hospital Organization Nagoya Medical Center, Nagoya, Japan, [10]Division of Clinical Oncology/Hematology, The Jikei University Daisan Hospital, Tokyo, Japan, [11]Department of Hematology and Cell Therapy, AbbVie GK, Osaka, Japan, [12]Department of Hematology and Cell Therapy, AbbVie GK, Tokyo, Japan, [13]Department of Hematology and Cell Therapy, Abbvie Inc., Singapore and [14]Department of Hematology and Cell Therapy, Aichi Cancer Center, Nagoya, Japan

*For reprints and all correspondence: Noboru Asada, Okayama University Hospital, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan. E-mail: nasada@okayama-u.ac.jp

[†]Present address: Department of Hematology, Faculty of Medicine, Institute of Medical, Pharmaceutical and Health Sciences, Kanazawa University, Kanazawa, Japan

[‡]Present address: Moderna Japan Co., Ltd, Tokyo, Japan

Received 17 November 2022; Editorial Decision 14 March 2023; Accepted 15 March 2023

Abstract

Background: In a Phase 3 international clinical trial (VIALE-C), venetoclax plus low-dose cytarabine improved the response rate and overall survival versus placebo plus low-dose cytarabine in patients with newly diagnosed acute myeloid leukemia who were ineligible for intensive chemotherapy. After the enrollment period of VIALE-C ended, we conducted an expanded access study to provide preapproval access to venetoclax in combination with low-dose cytarabine in Japan.

Methods: Previously, untreated patients with acute myeloid leukemia who were ineligible for intensive chemotherapy were enrolled according to the VIALE-C criteria. Patients received venetoclax (600 mg, Days 1–28, 4-day ramp-up in Cycle 1) in 28-day cycles and low-dose cytarabine (20 mg/m^2,

Days 1–10). All patients took tumor lysis syndrome prophylactic agents and hydration. Safety endpoints were assessed.

Results: Fourteen patients were enrolled in this study. The median age was 77.5 years (range = 61–84), with 78.6% over 75 years old. The most common grade ≥ 3 treatment-emergent adverse event was neutropenia (57.1%). Febrile neutropenia was the most frequent serious adverse event (21.4%). One patient developed treatment-related acute kidney injury, leading to discontinuation of treatment. Two patients died because of cardiac failure and disease progression that were judged not related to study treatment. No patients developed tumor lysis syndrome.

Conclusions: The safety outcomes were similar to those in VIALE-C without new safety signals and were well managed with standard medical care. In clinical practice, more patients with severe background disease are expected, in comparison with in VIALE-C, suggesting that it is important to carefully manage and prevent adverse events.

Key words: acute myeloid leukemia, venetoclax, low-dose cytarabine, expanded access study, tumor lysis syndrome

Introduction

The standard treatment strategy for newly diagnosed acute myeloid leukemia (AML) is an intensive curative chemotherapy, and a combination of cytarabine (AraC) and anthracycline is recommended as remission induction therapy. However, many AML patients are ineligible for intensive therapy because of advanced age or co-morbidities (1–3). Treatment options are limited for these patients, especially older patients, who account for a large proportion of patients with newly diagnosed AML (1,4,5). According to the guidelines of the Japanese Society of Hematology at the time of study initiation, the only recommended treatment for older patients with AML, in whom standard therapy is unsuitable but who are treatable, is low-dose AraC (LDAC) or participation in a clinical study (6).

Venetoclax is an orally available, small-molecule selective B-cell leukemia/lymphoma-2 inhibitor (7,8). In two placebo-controlled Phase 3 trials, the safety and efficacy of venetoclax-based therapy were confirmed in treatment-naive patients with AML who were ineligible for intensive chemotherapy owing to advanced age or co-morbidities. In the VIALE-A study, the venetoclax plus azacitidine (AZA) arm demonstrated significantly better outcomes compared with the placebo plus AZA arm (9). In the VIALE-C study, the venetoclax plus LDAC arm did not meet its primary endpoint of a statistically significant improvement in overall survival compared with the placebo plus LDAC arm (10). In the 6-month follow-up analysis of the VIALE-C study, the addition of venetoclax to LDAC increased the rates of complete remission (CR) and CR with incomplete blood count recovery (CRi) compared with the control arm (48 vs. 13%; $P < 0.001$) and extended median overall survival [8.4 vs. 4.1 months (hazard ratio = 0.70; $P = 0.04$)] (10,11). In the subgroup analysis, venetoclax plus LDAC was well tolerated in Japanese patients (5).

In November 2018, the US Food and Drug Administration granted accelerated approval for venetoclax in combination with AZA, decitabine or LDAC for the treatment of newly diagnosed AML in adults aged ≥75 years or who have co-morbidities that preclude the use of intensive induction chemotherapy. The expanded access study (EAS) framework (Japanese compassionate use program) was established in January 2016. This framework can provide preapproval access to unapproved or off-label drugs for patients under the following conditions: the target disease is serious and life-threatening with no effective therapy available; the drug, either unapproved or off-label, is under development in Japan and is in the final stage of development, that is, the pivotal trial (confirmatory trial for new drug application) has ended or patient enrollment in the trial has finished (12). A supplemental new drug application for venetoclax in AML was submitted in June 2020. Owing to the limited treatment options available for patients with AML who are not candidates for intensive chemotherapy, this EAS was conducted to support the use of venetoclax until approval. The VIALE-C regimen was adopted to provide venetoclax treatment to a broader range of patients, including those who had been pretreated with hypomethylating agents, such as AZA, which is in line with the purpose of the EAS as opposed to the VIALE-A study where those patients were excluded. In March 2021, venetoclax was approved for the treatment of AML by the Ministry of Health, Labour and Welfare of Japan based on results of the VIALE-A and VIALE-C studies. Here, we aim to present the safety results of venetoclax plus LDAC in Japanese patients who were ineligible for intensive chemotherapy in the EAS.

Patients and methods

Study design

This study was a single-arm, open-label, multicenter, EAS of venetoclax in combination with LDAC in newly diagnosed patients with AML who were ineligible for intensive induction therapy in Japan. The primary objective was to provide a treatment option with venetoclax plus LDAC to eligible patients in the EAS prior to the approval of venetoclax by the Ministry of Health, Labour and Welfare in Japan. There were no efficacy endpoints and only safety was assessed, but bone marrow and disease assessment were conducted at the investigator's discretion to evaluate the disease condition based on patients' physical findings, peripheral blood counts and/or bone marrow examination during study treatment. The protocol and informed consent form were reviewed and approved by an independent ethics committee/institutional review board at each site before initiation. All patients provided written informed consent before participating. The study was conducted in accordance with the International Council for Harmonization requirements, Good Clinical Practice guidelines and the Declaration of Helsinki.

Patients

This study enrolled the following patients and had the identical eligibility criteria as the VIALE-C study. Eligible patients were adults

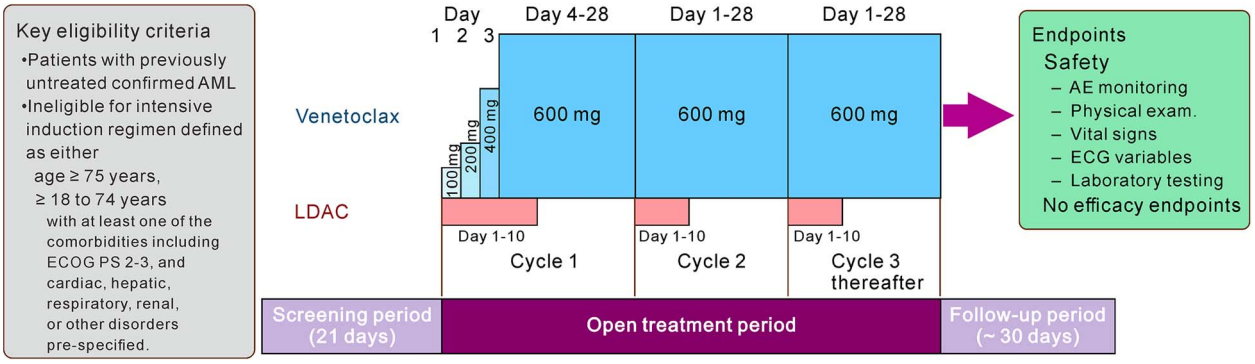


Figure 1. Design of the expanded access study. AE, adverse event; ECG, electrocardiogram; LDAC, low-dose cytarabine; AML, acute myeloid leukemia; ECOG, Eastern Cooperative Oncology Group; PS, performance status.

(≥18 years old) with newly diagnosed AML according to World Health Organization criteria (13). Patients were considered to be ineligible for standard induction therapy if they were aged ≥75 years or with the presence of at least one of the following: Eastern Cooperative Oncology Group performance status (PS) 2 or 3; cardiac history of congestive heart failure requiring treatment or ejection fraction ≤50% or chronic stable angina; diffusion capacity of the lung for carbon monoxide ≤65% or forced expiratory volume in 1 second ≤65%; creatinine clearance of ≥30 ml/min to <45 ml/min; total bilirubin >1.5 to ≤3.0 times the upper limit of normal or other co-morbidities deemed incompatible with standard intensive chemotherapy.

Treatment

All patients received venetoclax 600 mg orally once a day or daily (QD) on Days 1–28 in combination with LDAC 20 mg/m^2 subcutaneously on Days 1–10 in each 28-day treatment cycle, except for the first cycle (Cycle 1). In Cycle 1, venetoclax dosing began at 100 mg on Day 1 of the cycle and was then increased stepwise over 4 days (ramp-up period) to reach the target dose of 600 mg (100, 200, 400 and 600 mg). Treatment with the study drugs was continued until progressive disease (PD), unacceptable toxicity or other pre-established treatment discontinuation criteria were met, or until venetoclax was commercially available after its approval. All patients were followed up for 30 days after the last dose of venetoclax (Fig. 1).

Tumor lysis syndrome (TLS) prophylaxis and monitoring were implemented for all patients during the study as TLS risk mitigation measures. Specifically, all patients (i) were hospitalized prior to the initial dose of study treatment for at least 24 hours after reaching the target dose of venetoclax in Cycle 1 to monitor for TLS; (ii) received a uric acid reducing agent and hydration prior to and during the ramp-up period and (iii) underwent blood sampling for TLS chemistry tests, including calcium, inorganic phosphorus, potassium, uric acid and creatinine on Day 1 of each cycle and each day during the ramp-up period within 4 hours prior to dosing and 6–8 hours post-dosing of the study drug. Anti-infective prophylaxis for viral, fungal, bacterial or pneumocystis infections was required for patients with an absolute neutrophil count of <500/μl.

Assessments

Safety evaluations were performed in enrolled patients throughout the study, including adverse event (AE) monitoring, physical examination, vital sign measurement, variables in electrocardiogram/two-dimensional echocardiogram/multi-gated acquisition scans and clinical laboratory testing (hematology, chemistry, liver functions and urinalysis) as measures of safety and tolerability for the entire study duration.

Treatment-emergent AEs (TEAEs) were defined as those that occurred between the first dose of the study drug until 30 days after the last dose of the study drug. AEs were graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events Version 4.03.

Disease assessments were conducted at the discretion of the investigator according to patients' physical findings, peripheral blood counts and/or bone marrow assessment during the study treatment, mainly at screening, the end of Cycle 1 and every three cycles thereafter. Clinical responses were defined according to the modified International Working Group Criteria for AML (14), and PD was defined as per European Leukemia Net recommendations (15).

Transfusion independence was defined as a period of at least 56 consecutive days with no red blood cell or platelet transfusion during the evaluation period. The post-baseline transfusion evaluation period was from the first dose of the study drug to the last dose of the study drug plus 30 days, PD, confirmed morphological relapse or death, whichever occurred earlier.

Statistical methods

The sample size was not determined statistically. Safety was assessed through reported TEAEs, serious AEs (SAEs), AEs leading to discontinuation, death or changes in laboratory and vital sign parameters.

Results

Patients

Patient demographics and clinical characteristics are summarized in Table 1. This study was conducted at 11 sites in Japan between 5 October 2020 and 13 May 2021. Eighteen patients were screened, among whom 14 patients with AML were enrolled and received venetoclax in combination with LDAC; 4 patients were excluded for reasons of not meeting the eligibility criteria (n = 3) or early death before enrollment (n = 1). The median age was 77.5 years, 11 patients (78.6%) were ≥ 75 years old, 1 patient was 61 years old with moderate hepatic impairment (total bilirubin >1.5 to ≤3.0 upper limit of normal) and the other 2 patients were 70 and 72 years old with co-morbidities that the physicians judged to be incompatible with intensive chemotherapy. Most patients had Eastern Cooperative Oncology Group PS 0–1: seven patients (50.0%) were PS 0 and

Table 1. Patient demographics and baseline characteristics

Characteristics	*n* (%) or median (range) *N* = 14
Age (years)	
Median (range)	77.5 (61–84)
≥75	11 (78.6%)
Sex	
Female	3 (21.4%)
Male	11 (78.6%)
ECOG PS	
0	7 (50.0%)
1	6 (42.9%)
≥2	1 (7.1%)
AML type	
De novo AML	6 (42.9%)
Secondary AML	8 (57.1%)
Type of secondary AML (*n* = 8)	
Treatment-related	1/8 (12.5%)
Post-MDS	7/8 (87.5%)
AML with MDS-related changes	
Yes	6 (42.9%)
No	8 (57.1%)
Prior systemic therapy	
Prior AZA treatment	5 (35.7%)
Bone marrow blast count	
<30%	8 (61.5%)
≥30% to <50%	4 (30.8%)
≥50%	1 (7.7%)
Missing	1
Baseline neutrophil count ($\times 10^9$/l)	
Median (range)	0.6 (0.0–7.4)
Baseline hemoglobin value (g/l)	
Median (range)	79.5 (66.0–109.0)
Baseline platelet count ($\times 10^9$/l)	
Median (range)	42.0 (14.0–338.0)
RBC transfusion dependence at baseline[a]	
Yes	9 (64.3%)
Platelet transfusion dependence at baseline[a]	
Yes	9 (64.3%)
RBC or platelet transfusion dependence at baseline[a]	
Yes	9 (64.3%)

ECOG, Eastern Cooperative Oncology Group; PS, performance status; AML, acute myeloid leukemia; MDS, myelodysplastic syndrome; AZA, azacitidine; RBC, red blood cell.

[a]Transfusion dependence at baseline was defined as transfusion within 56 days prior to the first dose of study drug.

six patients (42.9%) were PS 1. Secondary AML was reported in 8 of 14 patients (57.1%), among whom 7 patients (87.5%) had myelodysplastic syndrome (MDS) overt AML. Five patients (35.7%) had a treatment history of AZA. Blast counts in bone marrow (<30%) were reported for 8 of 13 patients (61.5%), who were diagnosed with AML according to the World Health Organization classification.

The median treatment period was 2.0 months (range = 0.7–5.1). All patients discontinued the study treatment. Seven patients (50.0%) continued to receive the same combination treatment using commercially available venetoclax instead of the study drug after its approval. In the other seven patients, the primary reasons for discontinuation included the physician's decision in three cases (21.4%), PD in two cases (14.2%) and AEs related to and not related to aggravation of AML in one case each (7.1%).

Table 2. Common TEAEs

AEs, ≥20% of patients (VEN + LDAC)	Any grade	Grade 3 or 4
Any TEAEs	14 (100%)	14 (100%)
Hematological AEs		
Neutropenia	8 (57.1%)	8 (57.1%)
Anemia	5 (35.7%)	5 (35.7%)
Decreased WBC count	5 (35.7%)	5 (35.7%)
Lymphopenia	5 (35.7%)	5 (35.7%)
Thrombocytopenia	5 (35.7%)	5 (35.7%)
Leukopenia	4 (28.6%)	4 (28.6%)
Febrile neutropenia	4 (28.6%)	4 (28.6%)
Non-hematological AEs		
Decreased appetite	6 (42.9%)	1 (7.1%)
Nausea	6 (42.9%)	0
Constipation	4 (28.6%)	0
Hypokalemia	3 (21.4%)	1 (7.1%)

AE, adverse event; VEN, venetoclax; LDAC, low-dose cytarabine; TEAE, treatment-emergent AE; WBC, white blood cell.

Safety

TEAEs reported in ≥20% of patients, regardless of severity or relationship to the study drug, are listed in Table 2. All patients experienced at least one grade ≥ 3 TEAE. The most common TEAE (grade ≥ 3) was neutropenia in eight patients (57.1%) followed by other hematological TEAEs, including anemia, lymphopenia, thrombocytopenia and decreased white blood cell count in five patients each (35.7% each), involving patients who experienced multiple TEAEs. Additionally, febrile neutropenia was reported in four patients (28.6%). Any grade of infection was reported in five patients (35.7%), among which two experienced grade ≥ 3 infections, including nasopharyngitis, pneumonia and sepsis (*n* = 1 each) involving a patient who experienced multiple TEAEs.

SAEs were reported in six patients (42.9%) (Table 3). Febrile neutropenia was the most frequently reported SAE (*n* = 3, 21.4%), which was followed by cardiac failure, gastroenteritis, nasopharyngitis, AML (aggravation of the disease) and acute kidney injury (AKI) reported in one patient each. TLS was not reported in this study. The incidence of dose interruption, dose reduction and permanent venetoclax discontinuation owing to TEAEs was 35.7% (*n* = 5), 7.1% (*n* = 1) and 14.3% (*n* = 2), respectively (Table 4). AKI and AML were reported as TEAEs that led to the discontinuation of venetoclax, and heart failure and AML were reported as TEAEs that led to death.

A Grade 3 AKI was observed in an 80-year-old man with hypertension and diabetic nephropathy. He also used nifedipine and furosemide for coronary spastic angina and cardiac failure during study treatment. On Day 24, the antifungal prophylaxis was changed from caspofungin to fluconazole, and the dose of venetoclax was reduced to 300 mg accordingly. Subsequently, elevation of serum creatinine, hypercalcemia and decreased blood pressure were observed. Venetoclax was discontinued on Day 26 owing to persistently elevated creatinine level. Nifedipine and furosemide were also discontinued. On Day 45, renal failure improved. Drug-induced renal injury, decreased blood pressure and hypercalcemia were suspected to be the causes of acute renal failure. The causality of the study drugs could not be ruled out.

Fatal Grade 5 cardiac failure was reported in a 61-year-old male patient during post-treatment. He was previously diagnosed with

Table 3. SAEs of venetoclax and LDAC

SAEs (SOC/PT)	Patients, $N = 14$, n (%)
Any SAEs	6 (42.9%)
Blood and lymphatic system disorders	3 (21.4%)
Febrile neutropenia	3 (21.4%)
Cardiac disorder/cardiac failure	1 (7.1%)
Infections and infestations	2 (14.3%)
Gastroenteritis	1 (7.1%)
Nasopharyngitis	1 (7.1%)
Neoplasms: benign, malignant and unspecified/AML[a]	1 (7.1%)
Renal and urinary disorders/AKI	1 (7.1%)

No patients reported tumor lysis syndrome (TLS) in this study, where all patients received either TLS-prophylactic agents or hydration. SAE, serious adverse event; SOC/PT, MedDRA system organ class and preferred term; AKI, acute kidney injury.
[a]MedDRA PT for 'aggravation of AML,' which was defined as an investigator-reported AE.

Table 4. TEAEs of venetoclax leading to death and action taken

AE (PT) leading to action taken	Patients (n) $N = 14$ (100%)
Venetoclax dose interruption, n (%)[a]	
Any AEs	5 (35.7%)
Febrile neutropenia	1 (7.1%)
Leukopenia	1 (7.1%)
Neutropenia	1 (7.1%)
Thrombocytopenia	1 (7.1%)
Vomiting	1 (7.1%)
Nasopharyngitis	1 (7.1%)
Decreased WBC count	1 (7.1%)
AKI	1 (7.1%)
Venetoclax dose reduction, n (%)[a]	
Any AEs	1 (7.1%)
Leukopenia	1 (7.1%)
Neutropenia	1 (7.1%)
Venetoclax discontinuation, n (%)	
Any AEs	2 (14.3%)
AML[a]	1 (7.1%)
AKI	1 (7.1%)
AE leading to death, n (%)	
Any AEs	2 (14.3%)
Cardiac failure[b]	1 (7.1%)
AML[b,c]	1 (7.1%)

AE, adverse event; AML, acute myeloid leukemia; PT, MedDRA preferred term; TEAE, treatment-emergent AE; WBC, white blood cell.
[a]Including patients with multiple AEs. [b]Cause of death in both cases was considered not to be venetoclax or LDAC in the opinion of the investigator. [c]MedDRA PT for 'aggravation of AML,' which was defined as an investigator-reported AE.

MDS 75 days before initial dosing of the study drug and was not reported to have received AZA. After his diagnosis of AML, he consented to participate in this study and the Study Cycle 1 was started. On Day 24, bone marrow aspiration showed no therapeutic effect. The study drug was discontinued at the end of Cycle 1 on Day 28. Post-treatment was started on Day 31 with reduced-dose 7 + 3 AraC (67 mg/m^2 QD for 7 days) and idarubicin (3.4 mg/m^2 QD for 3 days). The following day, he had pyrexia and dyspnea with decreased oxygen saturation, which led to suspicion of heart failure. On Day 50, oxygen therapy was started due to low oxygen saturation between 89 and 92%. On Day 53, his blood pressure and oxygen saturation levels decreased, and despite intervention, he died. The final administrations of venetoclax and LDAC were >3 weeks and >6 weeks before the onset of the event, respectively. Therefore, this event was considered likely to be associated with the complications of AML and not related to the study treatment.

Prophylaxis

Anti-fungal agents were concomitantly used in eight patients, mostly as a prophylactic for patients with an absolute neutrophil count <500/μl; namely, fluconazole was used in four patients and voriconazole, micafungin, caspofungin was used for one patient each as antifungal prophylaxis. Dose modification for venetoclax was defined in the protocol when moderate and strong CYP3A inhibitors such as azoles were concomitantly administered. Granulocyte-colony stimulating factor (G-CSF) was used in five patients during venetoclax treatment, and two of them used it during febrile neutropenia (Fig. 2).

Disease assessment

Figure 3 shows responses at assessment time points in each patient. Best responses were CR (n = 2), CRi (n = 3), morphologic leukemia-free state (n = 2), resistant disease (n = 4), PD (n = 1) and not evaluable (n = 1) by investigator assessment. Marrow blasts were decreased from baseline in 10 out of 14 patients, and 7 patients achieved <5% during the study period (Fig. 4). Transfusion independence for red blood cells and platelets was 28.6 and 57.1%, respectively. Total transfusion independence was 28.6%.

Discussion

In this EAS, venetoclax in combination with LDAC showed a similar safety profile to that in the VIALE-C study. Regarding patient characteristics, the venetoclax plus LDAC arm of the VIALE-C study (N = 143) included 82 patients ≥75 years old (57%), 58 patients (41%) with secondary AML, 52 patients (36%) with prior hematologic disorder and 28 patients (20%) who had received prior treatment with hypomethylating agents (AZA or decitabine) for MDS (10). In comparison, the EAS (N = 14) included 11 patients ≥aged 75 years (78.6%), 8 patients (57.1%) with secondary AML, 7 patients (50.0%) with an MDS history and 5 (35.7%) who had received prior AZA treatment. These proportions were higher than those in the VIALE-C study, suggesting that patients in the EAS had more severe background disease. The incidence of neutropenia was higher in the patients of this study than in the Japanese subgroup of the VIALE-C study (57.1 vs. 16.7%), but the incidence of febrile neutropenia was lower (28.6 vs. 50.0%) (5). TLS was not reported in this EAS; it is thought to be manageable using risk mitigation measures, including appropriate prophylaxis and monitoring.

The patient with AKI developed renal failure and decreased blood pressure after changing the antifungal prophylaxis to fluconazole. Fluconazole, which is a moderate cytochrome P450 3A (CYP3A) inhibitor, may have increased the blood concentration of nifedipine, which is a substrate of CYP3A, and decreases blood pressure and could have resulted in prerenal failure. Additionally, the patient's history of diabetic nephropathy may have contributed to the development of AKI. In the VIALE-C study, 7 out of 142 patients (4.9%) in the venetoclax + LDAC group had AKI (including 1 with SAE) compared with 5 of 68 patients (7.4%) in the placebo + LDAC group,

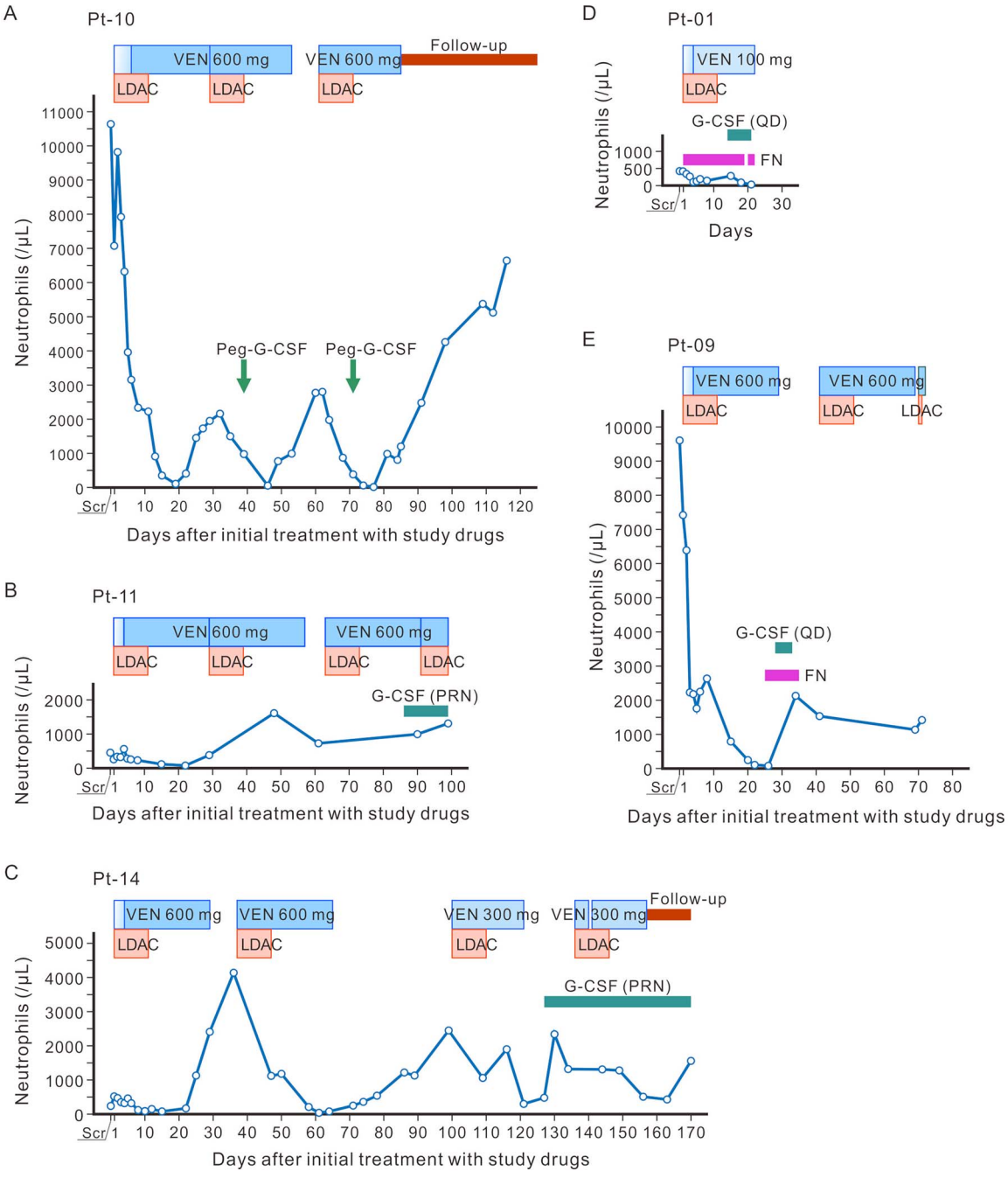


Figure 2. Concomitant use of G-CSF and neutrophil counts. (A–C) Time courses of neutrophil counts in Pt-10, -11 and -14 who received prophylactic G-CSF. (D, E) Pt-01 and -09 received G-CSF as treatment for febrile neutropenia. G-CSF, granulocyte-colony stimulating factor; Peg, pegfilgrastim; FN, febrile neutropenia; PRN, pro re nata; QD, quaque die (once a day or daily); VEN, venetoclax; Pt, patient; Scr, screening.

with no increase in venetoclax treatment, although no SAEs were reported in the latter group (16). Venetoclax is mainly metabolized by CYP3A in the liver, and <0.1% is excreted into the urine (17). Therefore, the risk of renal injury is presumed to be low. However, coadministration of venetoclax with CYP3A and/or P-glycoprotein inhibitors increases venetoclax blood concentrations and requires a dose reduction of venetoclax (18,19). Because older patients often have co-morbidities and use multiple medications, drug interactions should be carefully monitored. In addition, it should also be noted that Ca antagonists, which are frequently used as antihypertensive drugs in Japan, are also metabolized by CYP3A; thus, caution is needed when Ca antagonists are coadministered with CYP3A inhibitors, which is similar to venetoclax (20,21).

G-CSF was used in 5 out of 14 patients (35.7%), including for prophylactic purposes (n = 3). Among these, one patient received pegylated G-CSFs after LDAC administration (22). The use of G-CSF may have contributed to the lower incidence of febrile neutropenia (28.6% in this EAS vs. 50.0% in the Japanese VIALE-C population) despite the higher incidence of neutropenia (57.1 vs. 16.7%), although this cannot be confirmed owing to the coincided

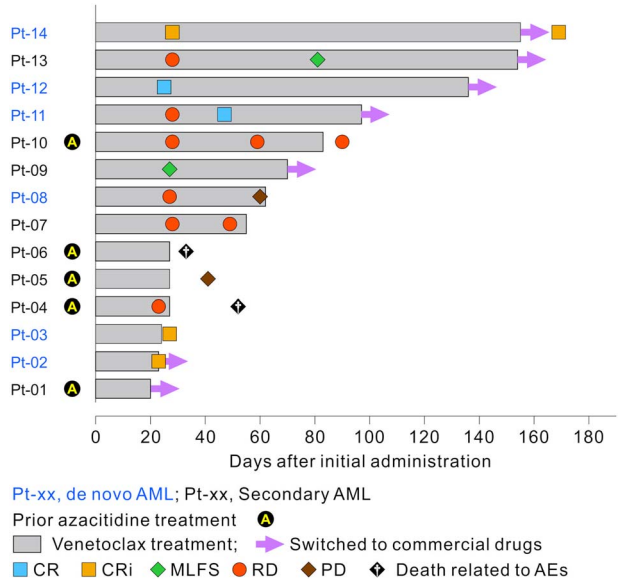


Figure 3. Disease response and treatment duration in each patient, with baseline characteristics. Commercial drugs: patients received commercially available venetoclax + LDAC after approval of venetoclax in Japan. venetoclax treatment, open treatment period; CR, complete remission; CRi, CR with incomplete blood count recovery; MLFS, morphologic leukemia-free state; RD, resistant disease; PD, progressive disease.

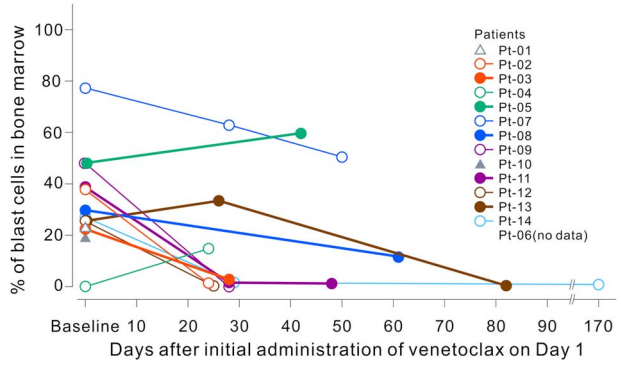


Figure 4. Bone marrow blast count change after the initial administration of venetoclax. Bone marrow aspiration or biopsy for disease assessment was conducted at screening and was done at the investigator's discretion according to patients' physical findings or peripheral blood results.

drug holiday and the small sample size in both studies. Furthermore, neither study was designed to assess the effect of G-CSF, and its use was not mandated by the protocol.

This study did not include efficacy endpoints because this was an EAS focusing on 'early access' for patients with no effective therapy; however, bone marrow assessment after Cycle 1 was performed to evaluate the disease condition at the discretion of each investigator. The rates of CR and CRi appeared to be similar to the outcomes of the VIALE-C study. Additionally, the achievement rate of transfusion independence suggested efficacy of venetoclax plus LDAC, which is comparable with that in the VIALE-C study (11).

The VIALE-C study confirmed the benefit–risk balance of venetoclax treatment in combination with LDAC among patients with untreated AML who are ineligible for intensive chemotherapy and for whom treatment options are limited (10). This study provided supportive data regarding the benefit–risk balance of venetoclax treatment in combination with LDAC, more closely applicable to the real-world clinical treatment of Japanese patients with AML than that in the VIALE-C study. This EAS study included more older patients and patients with secondary AML than the VIALE-C study, highlighting the fact that the patient background will be more serious in actual clinical practice. These results indicate that it is important to pay close attention to complications and concomitant medications during this treatment.

In conclusion, the AEs reported in this EAS were consistent with the known safety profile of venetoclax plus LDAC and were successfully managed with standard medical care. The findings of this EAS further support the benefit–risk profile of venetoclax plus LDAC shown in the VIALE-C study.

Acknowledgements

AbbVie and the authors thank all the study investigators and the patients who participated in this clinical study. Medical writing support was provided by ASCA Corporation and was funded by AbbVie.

Conflict of interest statement

Noboru Asada has received honoraria from AbbVie, Astellas Pharma, Chugai Pharmaceutical, Celgene, Eisai, Kyowa Kirin, Otsuka Pharmaceutical, Novartis Pharma and Sanofi. Chikashi Yoshida has received research funding from Bristol-Myers Squibb and honoraria from Bristol-Myers Squibb, Novartis Pharma, Pfizer Japan, Otsuka Pharmaceutical, AbbVie, Janssen Pharmaceutical, Nippon Shinyaku and Chugai Pharmaceutical, Takeda Pharmaceutical and Ono Pharmaceutical. Kensuke Usuki has received research funding from Astellas-Amgen-Biopharma, Otsuka Pharmaceutical, Apellis Pharmaceuticals, SymBio Pharmaceuticals, Takeda Pharmaceutical, Nippon Shinyaku, Novartis Pharma, AbbVie, Janssen Pharmaceutical, Bristol-Myers Squibb, Ono Pharmaceutical, Chugai Pharmaceutical, Daiichi Sankyo, MSD, Astellas Pharma, Alexion Pharmaceuticals, Kyowa-Kirin, Gilead Sciences, Pfizer, Incyte, SymBio Pharmaceuticals, Celgene, Sumitomo Dainippon Pharma, Mundipharma International, Yakult, Eisai and Celgene; honoraria from Novartis Pharma, Astellas Pharma, Alexion Pharmaceuticals, Eisai, MSD, Otsuka Pharmaceutical, Ono Pharmaceutical, Kyowa-Kirin, Celgene, Daiichi Sankyo, Takeda Pharmaceutical, Nippon-Shinyaku, PharmaEssentia, Bristol-Myers Squibb, Yakult, Sanofi, Pfizer, AbbVie and Chugai Pharmaceutical; and has served on consulting bureaus for Alexion Pharmaceuticals, SymBio Pharmaceuticals, Nippon Shinyaku, Otsuka Pharmaceutical, Chugai Pharmaceutical, Sanofi, Takeda Pharmaceutical, Kyowa-Kirin, Astellas Pharma and Alnylam Japan. Kenichi Ishizawa has received research funding from Novartis Pharma, AbbVie, Bayer, SymBio Pharmaceuticals, Pfizer, Sanofi and Otsuka Pharmaceutical; honoraria from Takeda Pharmaceutical, Celgene, Ono Pharmaceutical, Novartis Pharma, Chugai Pharmaceutical, Eisai, IQVIA Services Japan, MSD, Kyowa Kirin, Janssen Pharmaceutical and Micron. Toshihiro Miyamoto has received honoraria from AbbVie, Janssen Pharmaceutical, Bristol-Myers Squibb, Otsuka Pharmaceutical, Astellas Pharma and MSD. Hiroatsu Iida has received research funding from Chugai Pharmaceutical; honoraria from Janssen Pharmaceutical, AbbVie, Astellas Pharma, Chugai Pharmaceutical and Novartis Pharma. Nobuaki Dobashi has received research finding from Astellas Pharma, Chugai Pharmaceutical, Daiichi Sankyo, Kyowa Kirin, Otsuka Pharmaceutical, AbbVie, Takeda Pharmaceutical, Ono Pharmaceutical, Bristol-Myers Squibb, Pfizer and Japan Agency for Medical Research and Development;

honoraria from Otsuka Pharmaceutical and Nippon Shinyaku and paid expert testimony from Otsuka Pharmaceutical. Sumiko Okubo, Hideyuki Honda, Tomomi Soshin, Yasuko Nishimura and Atsuko Tsutsui are employees of AbbVie and may own stock or possess stock options. Harumi Mukai was the former employee of AbbVie (currently affiliated with Moderna Japan). Kazuhito Yamamoto reports grants and personal fees from AbbVie, grants and personal fees from AstraZeneca, grants and personal fees from Bristol-Myers Squibb/Celgene, grants and personal fees from Chugai Pharmaceutical, grants and personal fees from Eisai, grants from IQVIA/Genmab, grants and personal fees from Meiji Seika Pharma, personal fees from Nippon Shinyaku, grants and personal fees from Novartis Pharma, grants and personal fees from Ono Pharmaceutical, grants and personal fees from Otsuka Pharmaceutical, grants from Solasia Pharma, personal fees from SymBio Pharmaceuticals, grants and personal fees from Takeda Pharmaceutical, grants from Yakult, grants and personal fees from Zenyaku Kogyo, personal fees from HUYA Bioscience International/IQVIA Services Japan, personal fees from Janssen Pharmaceutical, personal fees from Kyowa Kirin, personal fees from Micron, personal fees from MSD, personal fees from Sanofi, grants from Incyte, personal fees from Pfizer and personal fees from Nippon Kayaku, outside the submitted work. Jun Ando has received honoraria from Novartis Pharma and AbbVie and research funding from AbbVie and Daiichi Sankyo. Atsushi Shinagawa and Satoru Takada have nothing to disclose.

Funding

This work was supported by AbbVie, which contributed to its design; collected, analyzed and interpreted the data and participated in the writing, review and approval of the work. All authors had access to the relevant data and participated in the drafting, review and approval of this manuscript. No honoraria or payments were made for authorship.

Abbreviations

AE, adverse event; AKI, acute kidney injury; AML, acute myeloid leukemia; AraC, cytarabine; AZA, azacitidine; CR, complete remission; CRi, complete remission with incomplete blood count recovery; CYP3A, cytochrome P450 3A; EAS, expanded access study; G-CSF, granulocyte-colony stimulating factor; LDAC, low-dose cytarabine; MDS, myelodysplastic syndrome; PD, progressive disease; PS, performance status; QD, once a day or daily; SAE, serious adverse event; TEAE, treatment-emergent adverse event; TLS, tumor lysis syndrome

Data sharing and data accessibility statement

AbbVie is committed to responsible data sharing regarding the clinical trials we sponsor in which this EAS is contained. This includes access to anonymized, individual- and trial-level data (analysis data sets) as well as other information (e.g. protocols and Clinical Study Reports or analysis plans) as long as the trials are not part of an ongoing or planned regulatory submission. This includes requests for clinical trial data for unlicensed products and indications.

These clinical trial data can be requested by any qualified researchers who engage in rigorous, independent scientific research and will be provided following review and approval of a research proposal and statistical analysis plan and execution of a data sharing agreement. Data requests can be submitted at any time after approval in the US and Europe and after acceptance of this manuscript for publication. The data will be accessible for 12 months, with possible extensions considered. For more information on the process, or to submit a request, visit the following link: https://www.abbvie.com/our-science/clinical-trials/clinical-trials-data-and-information-sharing/data-and-information-sharing-with-qualified-researchers.html.

References

1. Kiyoi H, Yamaguchi H, Maeda Y, Yamauchi T. JSH practical guidelines for hematological malignancies, 2018: I. Leukemia-1. Acute myeloid leukemia (AML). *Int J Hematol* 2020;111:595–613.
2. Miyawaki S. Clinical studies of acute myeloid leukemia in the Japan adult leukemia study group. *Int J Hematol* 2012;96:171–7.
3. Kantarjian H, Ravandi F, O'Brien S, et al. Intensive chemotherapy does not benefit most older patients (age 70 years or older) with acute myeloid leukemia. *Blood* 2010;116:4422–9.
4. Heuser M, Ofran Y, Boissel N, et al. Acute myeloid leukaemia in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2020;31:697–712.
5. Yamauchi T, Yoshida C, Usuki K, et al. Venetoclax plus low-dose cytarabine in Japanese patients with untreated acute myeloid leukaemia ineligible for intensive chemotherapy. *Jpn J Clin Oncol* 2021;51:1372–82.
6. Miyawaki S. JSH guideline for tumors of hematopoietic and lymphoid tissues: leukemia 1. Acute myeloid leukemia (AML). *Int J Hematol* 2017;106:310–25.
7. Li Q, Cheng L, Shen K, et al. Efficacy and safety of Bcl-2 inhibitor venetoclax in hematological malignancy: a systematic review and meta-analysis of clinical trials. *Front Pharmacol* 2019;10:697.
8. Souers AJ, Leverson JD, Boghaert ER, et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat Med* 2013;19:202–8.
9. DiNardo CD, Jonas BA, Pullarkat V, et al. Azacitidine and venetoclax in previously untreated acute myeloid leukemia. *N Engl J Med* 2020;383:617–29.
10. Wei AH, Montesinos P, Ivanov V, et al. Venetoclax plus LDAC for newly diagnosed AML ineligible for intensive chemotherapy: a phase 3 randomized placebo-controlled trial. *Blood* 2020;135:2137–45.
11. Wei AH, Panayiotidis P, Montesinos P, et al. 6-month follow-up of VIALE-C demonstrates improved and durable efficacy in patients with untreated AML ineligible for intensive chemotherapy. *Blood Cancer J* 2021;11:163.
12. Fujiwara Y. Evolution of frameworks for expediting access to new drugs in Japan. *Nat Rev Drug Discov* 2016;15:293–4.
13. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016;127:2391–405.
14. Cheson BD, Bennett JM, Kopecky KJ, et al. Revised recommendations of the international working group for diagnosis, standardization of response criteria, treatment outcomes, and reporting standards for therapeutic trials in acute myeloid leukemia. *J Clin Oncol* 2003;21:4642–9.
15. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017;129:424–47.
16. AbbVie Inc. *M16–043 Study in 2.7.6 Synopses of Individual Studies of Venetoclax*. (in Japanese). CTD Module 2 for Japan submission. https://www.pmda.go.jp/drugs/2021/P20210318002/index.html (6 July 2022, date last accessed).
17. Liu H, Michmerhuizen MJ, Lao Y, et al. Metabolism and disposition of a novel B-cell lymphoma-2 inhibitor venetoclax in humans and characterization of its unusual metabolites. *Drug Metab Dispos* 2017;45:294–305.
18. Agarwal SK, DiNardo CD, Potluri J, et al. Management of venetoclax-posaconazole interaction in acute myeloid leukemia patients: evaluation of dose adjustments. *Clin Ther* 2017;39:359–67.

19. Venclexta (venetoclax). *Package insert. AbbVie Inc*. Revised October 2021: https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/208573s026lbl.pdf (23 May 2022, date last accessed).
20. Blockman M, Maartens G. Drug interactions with the calcium-channel blockers: clinical pharmacology. *CME* 2006;24:274–5. https://journals.co.za/doi/pdf/10.10520/EJC63124 (12 May 2022, date last accessed).
21. Choi JS, Choi I, Choi DH. Effects of nifedipine on the pharmacokinetics of repaglinide in rats: possible role of CYP3A4 and P-glycoprotein inhibition by nifedipine. *Pharmacol Rep* 2013;65:1422–30.
22. DiNardo CD, Wei AH. How I treat acute myeloid leukemia in the era of new drugs. *Blood* 2020;135:85–96.

Journal of Orthopaedic Surgery and Research

RESEARCH ARTICLE

Open Access

Correlations between 3D preoperative planning and postoperative reduction in the osteosynthesis of distal humeral fractures

Yuichi Yoshii[1*], Sho Iwabuchi[2], Akira Ikumi[2], Sho Kohyama[3], Takeshi Ogawa[4] and Tomoo Ishii[1]

Abstract

Background Three-dimensional preoperative planning has been applied to the osteosynthesis of distal humerus fractures. The present study investigated the correlations between 3D preoperative planning and postoperative reduction for the osteosynthesis of distal humerus fractures using 3D parameters.

Methods Twenty-three elbows of 23 distal humerus fracture patients who underwent osteosynthesis with three-dimensional preoperative planning were evaluated. 3D images of the distal humerus were created after taking preoperative CT scans of the injured elbow. Fracture reduction, implant selection, and placement simulations were performed based on 3D images. Postoperative CT images were taken 1 month after surgery. Correlations were evaluated with preoperative plans and postoperative 3D images. The longitudinal axis and coordinates of the humerus were defined on the 3D images. The coronal angle (CA) was defined as the angle formed by the long axis and the line connecting the medial and lateral margins of the trochlea of the humerus on a coronal plane image. The sagittal angle (SA) was defined as the angle formed by the long axis and the line connecting the top of the lateral epicondyle and the center of the humeral capitellum on a sagittal plane image. The axial angle (AA) was defined as the angle between the sagittal plane and the line connecting the medial and lateral margins behind the trochlea of the humerus. The intraclass correlation coefficients (ICC) of each measurement value were assessed between preoperative planning and postoperative images.

Results Preoperative planning and postoperative measurement values were CA: 85.6 ± 5.9°/85.8 ± 5.9°, SA: 140.9 ± 8.5°/139.4 ± 7.9°, and AA: 84.0 ± 3.1°/82.6 ± 4.9°, respectively. ICCs were CA: 0.75 ($P < 0.01$), SA: 0.78 ($P < 0.01$), and AA: 0.34 ($P < 0.05$), respectively.

Conclusions The 3D preoperative planning of distal humeral fractures achieved the good correlations of coronal and sagittal angles, but the relatively poor correlation of the axial angle. This may be attributed to an inability to assess the rotation angle during surgery. We propose the measurement indices shown in the present study as a three-dimensional evaluation index for distal humerus fractures.

Trial registration Registered as NCT04349319 at ClinicalTrials.gov.

Keywords Distal humerus fracture, Preoperative plan, Computed tomography, Three dimensions, Osteosynthesis, Computer-assisted orthopedic surgery

*Correspondence:
Yuichi Yoshii
yyoshii@tokyo-med.ac.jp
Full list of author information is available at the end of the article

Background

A distal humerus fracture is a fracture of the distal end of the humerus, one of the three bones (the humerus, radius, and ulna) that make up the elbow joint. Fractures of the distal humerus in adults account for 2% of all fractures and approximately 30% of all humeral fractures [1–4]. Anatomically, the distal humerus has a triangular shape that comprises two columns and a tie arch [4, 5]. The medial column holds at its distal end of the non-articular medial epicondyle with the insertion of the flexor muscles and the medial part of the humeral trochlea. The lateral column holds at its distal end of the capitellum and more proximally at the lateral epicondyle with the insertion of the extensor muscles.

Open reduction and internal fixation has become the standard treatment for distal humerus fractures [6–9]. The aim of surgical treatment is to reconstruct the strong triangular structure at the distal humerus [10]. Rigid internal fixation and anatomical remodeling are essential for the recovery of elbow function, bone healing, and the avoidance of cartilage degeneration [11]. Regarding rigid fixation, biomechanical studies demonstrated the advantages of double plating over single plating in metaphysis and intraarticular fractures of the distal humerus [12–15]. However, the number of screws that may be inserted into distal humerus fragments is limited due to the interference of screws inserted from the medial and lateral plates.

The utility of a three-dimensional (3D) surgical simulation has recently been reported [16–19]. Evaluations of 3D bone morphology and preoperative planning are considered to be an effective means for increasing the accuracy of surgery and reducing complications. In a previous study, a 3D preoperative planning system was developed to manage fractures around the elbow [20]. This system allows the reduction process and implant placement/choices to be visualized in a virtual space. It has the advantage of being able to predict in advance the interference of screws inserted from the medial and lateral plates. However, a method has not yet been established to three-dimensionally evaluate the reduction shape accuracy of the 3D preoperative plan. In the present study, we developed a method to evaluate reduction shape accuracy based on the 3D coordinates of the distal humerus. Using this method, the correlations between 3D preoperative planning and postoperative reduction were evaluated in the osteosynthesis of distal humerus fractures. We hypothesized that 3D preoperative planning for osteosynthesis of distal humeral fractures would have good correlations between 3D preoperative planning and postoperative reduction with an assessment of 3D parameters.

Methods

This study protocol was approved by the Institutional Review Board (approved No. 14-21, T2022-0041). The present study was registered as NCT04349319 at ClinicalTrials.gov. This was a prospective case series (level of evidence II). Twenty-three elbows of 23 distal humerus fracture patients who underwent osteosynthesis with 3D preoperative planning (14 females, 9 males, mean age 61.3 years, age range 21–87) were evaluated. Written consent was obtained from all study participants. Patients were excluded if they had a previous history of traumatic arm injuries. All patients had CT images of the injured elbow taken before and 1 month after surgery. According to preoperative X-ray (posterior-anterior and lateral view) and CT scans, fractures were classified using the AO classification system. CT images were taken with tube settings of 120 kV and 100 mAS, a section thickness of 0.8 mm, and a pixel size of 0.3×0.3 mm (Sensation Cardiac, Siemens). Images were taken in a range of approximately 20 cm centered on the elbow joint.

3D preoperative planning

3D preoperative planning and a surgical simulation were performed prior to surgery (Fig. 1). Preoperative planning software (Zed-Trauma Distal Humerus Stage, LEXI Co., Ltd., Tokyo, Japan) was used for the reduction and implant placement simulation. A 3D image of the distal humerus was created from the DICOM data of CT scans. A fracture reduction simulation was performed by separating bone fragments along the fracture line. Each distal humerus fragment was segmented according to the fracture line. The main reduction criteria were the improvement of shortening, angular and rotational deformation, the recovery of joint surface compatibility, and connecting between bone fragments. After repositioning the fragment, we confirmed the 3D shape of the distal humerus. Fragments larger than 10 mm were considered for reduction, while smaller ones were excluded from the reduction simulation. After reduction, simulations of implantation with the locking plates and screws of various sizes were performed. Computer-aided design models of different-sized implants were installed in the software. Criteria for plate selection were as follows: (1) the proximal portion of the plate reaches the diaphysis beyond the fracture line, allowing at least 3 screws to be inserted into the diaphysis; (2) enables screw fixation to the major distal fragments; (3) the insertion direction or length of the distal screw does not perforate the articular surface. Distal screws that were long enough to support distal humeral fragments were selected. Regarding proximal screws, screws of sufficient lengths were selected to reach the contralateral bone cortex. After reduction and

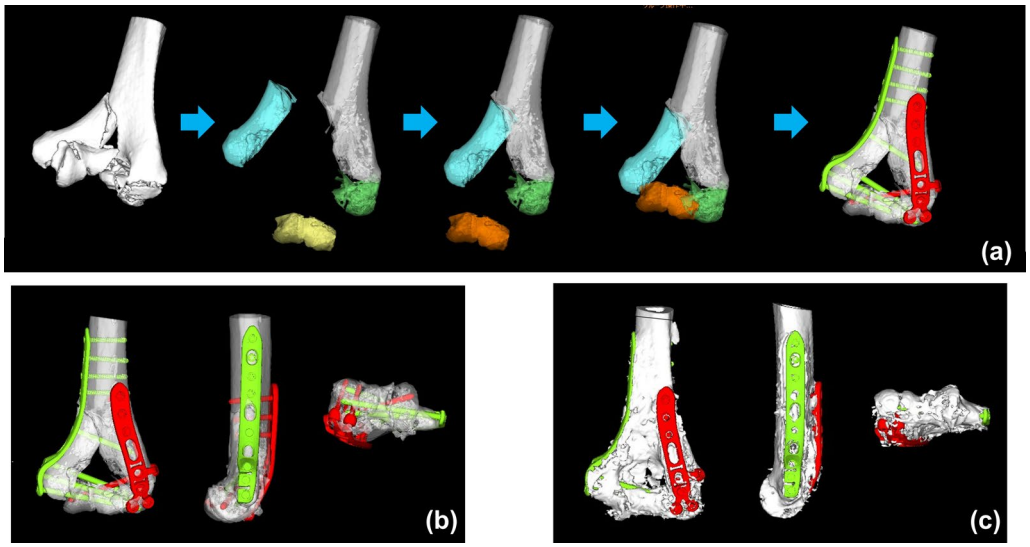


Fig. 1 An example of the preoperative planning process. **a** Reduction and implant placement simulation. **b** Completed preoperative plan. **c** Postoperative images

implantation simulations, osteosynthesis was performed under general anesthesia. Surgery was conducted in the lateral position with the injured side up, and the injured limb was fixed by placing a support table under the elbow. A posterior approach to the elbow joint was used to expand the fracture site. Osteotomy of the olecranon was added where necessary to reconstruct the articular surface. During surgery, the surgeon performed reduction and the placement of implants while comparing images between the preoperative plan and fluoroscopy images obtained during surgery. The positions of the plates were selected based on the distance from the articular surface of the distal end of the plate and fluoroscopic images. Screw lengths were selected by intraoperative depth gauge measurements with reference to preoperative measurements. Surgeries were performed by nine trainees (residents and fellows) and one hand surgeon. The hand surgeon participated in all surgeries.

Evaluations

Preoperative and postoperative 3D images of the distal humerus were analyzed with image analysis software (BoneSimulator, Orthree, Osaka, Japan). After importing image data into the software, a surface construction algorithm was used to construct a 3D surface model of the humerus (Fig. 2). The long axis of the humerus was calculated from a preoperative 3D surface model of the intact portion of the humerus. An intact portion of the

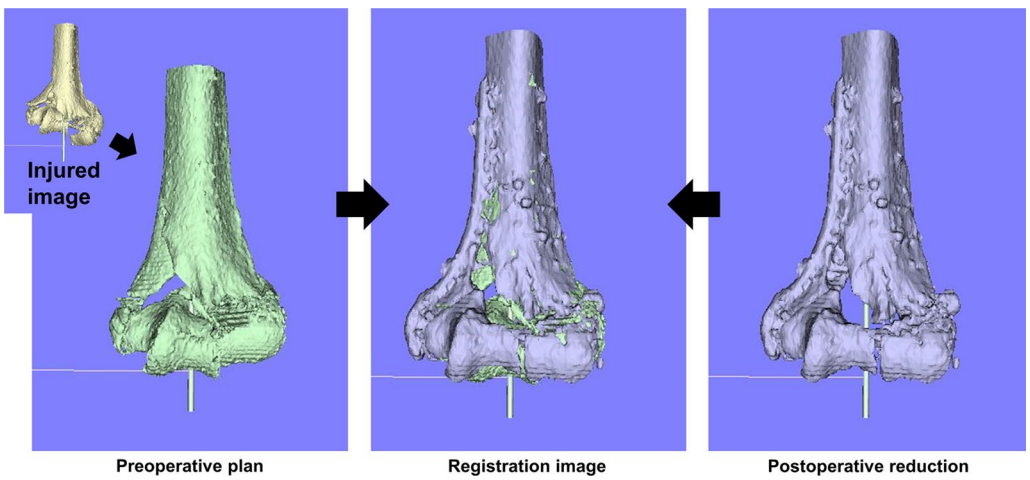


Fig. 2 An example of a registration image for the coronal view

distal humerus image was used for registration between the preoperative planning image and postoperative reduction image. The coronal plane is parallel to the long axis of the humeral shaft and includes the long axis and passes through the top of the medial epicondyle, the sagittal plane is the plane including the long axis and perpendicular to the coronal plane, and the plane perpendicular to the long axis is the axis defined as a cross-section. The origin of coordinates was defined as the intersection of the joint surface and the humerus long axis on the preoperative plan image. Preoperative planning and postoperative 3D models were evaluated in the same coordinate system.

In the correlation analysis, three angular parameters were measured according to anatomical landmarks (Fig. 3). The coronal angle (CA) was defined as the angle formed by the long axis and the line connecting the medial and lateral margins of the trochlea of the humerus on the coronal plane image. The sagittal angle (SA) was defined as the angle formed by the long axis and the line connecting the top of the lateral epicondyle and the center of the humeral capitellum on the sagittal plane image. The axial angle (AA) was defined as the angle between the sagittal plane and the line connecting the medial and lateral margins behind the trochlea of the humerus. Each parameter was measured on preoperative planning and postoperative images. Two raters independently assessed images. One rater was involved in the surgeries and the other was not involved in the surgeries. After evaluating the reliability of the two raters' measurements, the mean values for each parameter were used in further analyses.

Statistical analysis

In this study, we assessed the sample size with non-parametric binominal reliability demonstration test. For the calculation, we set the number of allowable test failures: 1, reliability: 80%, test confidence level: 95%. Subsequently, the sample size was determined as 22. Results are expressed as the mean ± standard deviation. The Shapiro–Wilk test was used to test the normality of datasets. Interrater reliability was assessed using the intraclass correlation coefficient (ICC). In addition, the ICCs for the parameters between the preoperative plan and postoperative reduction were assessed. According to the previous recommendation, ICC values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 were defined as poor, moderate, good, and excellent correlations, respectively [21]. P values < 0.05 were considered to be significant. All analyses were performed using BellCurve for Excel version 2.12 (SSRI Co., Tokyo, Japan).

Results

There were six patients with A2 fractures, four with A3 fractures, six with C2 fractures, and five with C3 fractures. There was one case each of B1 and B2 fractures. Screw fixation was performed in two cases. Single plate fixation was conducted in five cases (lateral plate: two cases, medial plate: two cases, postero-lateral plate: one case). Double plate fixation was performed in 16 cases (combination of medial and lateral plates: 12 cases, combination of medial and postero-lateral plates: four cases). The mean surgical time was 203.2 min (105–335 min).

The results of measurements are shown in Fig. 4. The results of correlations for angle parameters are shown in Fig. 5. Preoperative planning and postoperative

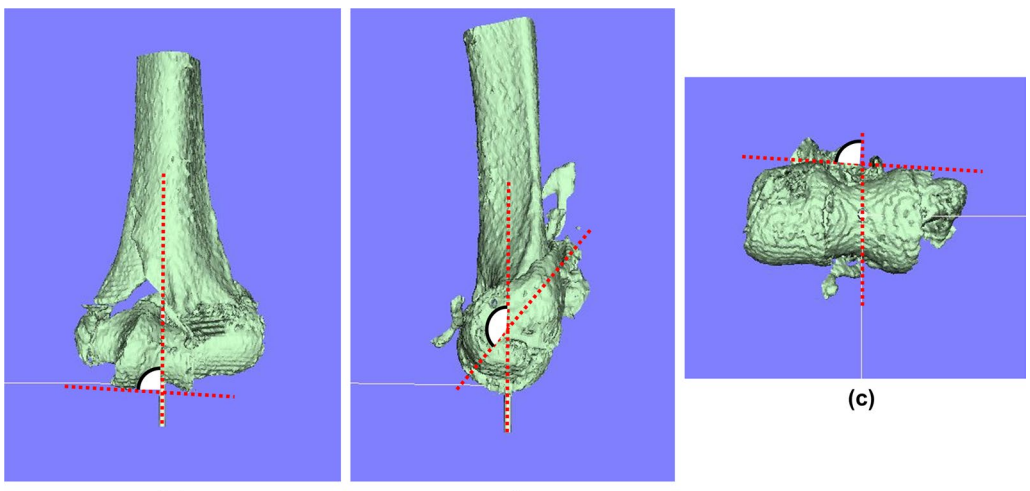


Fig. 3: 3D parameters. **a** Coronal angle (CA). **b** Sagittal angle (SA). **c** Axial angle (AA)

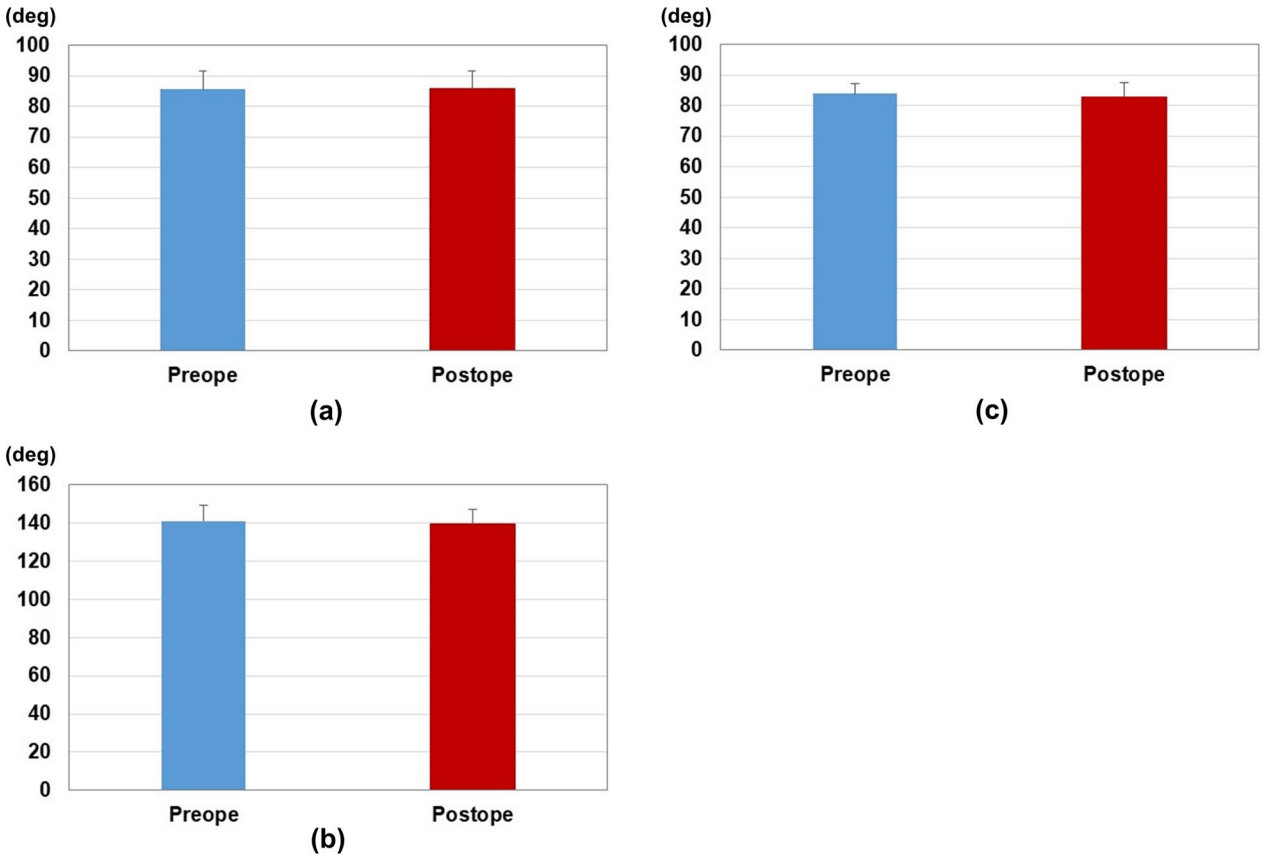


Fig. 4 Results of parameter measurements. **a** Results of the coronal angle. **b** Results of the sagittal angle. **c** Results of the axial angle. The blue bar indicates the measurements of the preoperative plan. The red bar indicates the measurements of the postoperative reduction

measurement values were CA: 85.6 ± 5.9°/85.8 ± 5.9°, SA: 140.9 ± 8.5°/139.4 ± 7.9°, and AA: 84.0 ± 3.1°/82.6 ± 4.9°, respectively. There were no significant differences between preoperative planning and postoperative measurements. ICCs were CA: 0.75 ($P < 0.01$), SA: 0.78 ($P < 0.01$), and AA: 0.34 ($P < 0.05$). There were good correlations in CA and SA, respectively. There was a poor correlation in AA. Interrater reliabilities were excellent for CA and SA with ICC values of 0.98 and 0.96, respectively. Interrater reliability was good for AA with ICC value of 0.89.

Discussion

In the surgical treatment of distal humeral fractures, stable anatomic joint reconstruction with osteosynthesis is necessary for proper bone healing and early functional recovery [11]. Since biomechanical studies have demonstrated the benefits of double plating over single plating for proximal and intra-articular fractures of the distal humerus [13–15], the placement of double plates in distal humeral fractures has been recommended in some cases. Conventional preoperative planning with X-ray images has generally been performed by transferring image data onto tracing paper. This was the standard method used by most orthopedic surgeons in clinical practice. In conventional planning, rotational reductions were difficult to assess prior to surgery. In addition, there was no method to three-dimensionally evaluate the interference of distal screws from the inside and outside plates. Advances in image processing technology have led to the development of preoperative planning systems based on the digital processing of image data [15, 18, 20, 22, 23]. The preoperative planning system was previously shown to be useful for visualizing the three-dimensional structure of fractures, judging the feasibility of reduction, and assessing the accuracy of implant selections. In addition, a virtual simulation increased the confidence of trainees and improved their decision making [24].

We previously developed a 3D surgical simulation system for distal humerus fractures. This simulation system is useful for evaluating the reduction shapes of rotational and angular deformations in distal humerus fractures. It also allows for the preoperative selection of appropriate screw directions and lengths. However, there is currently

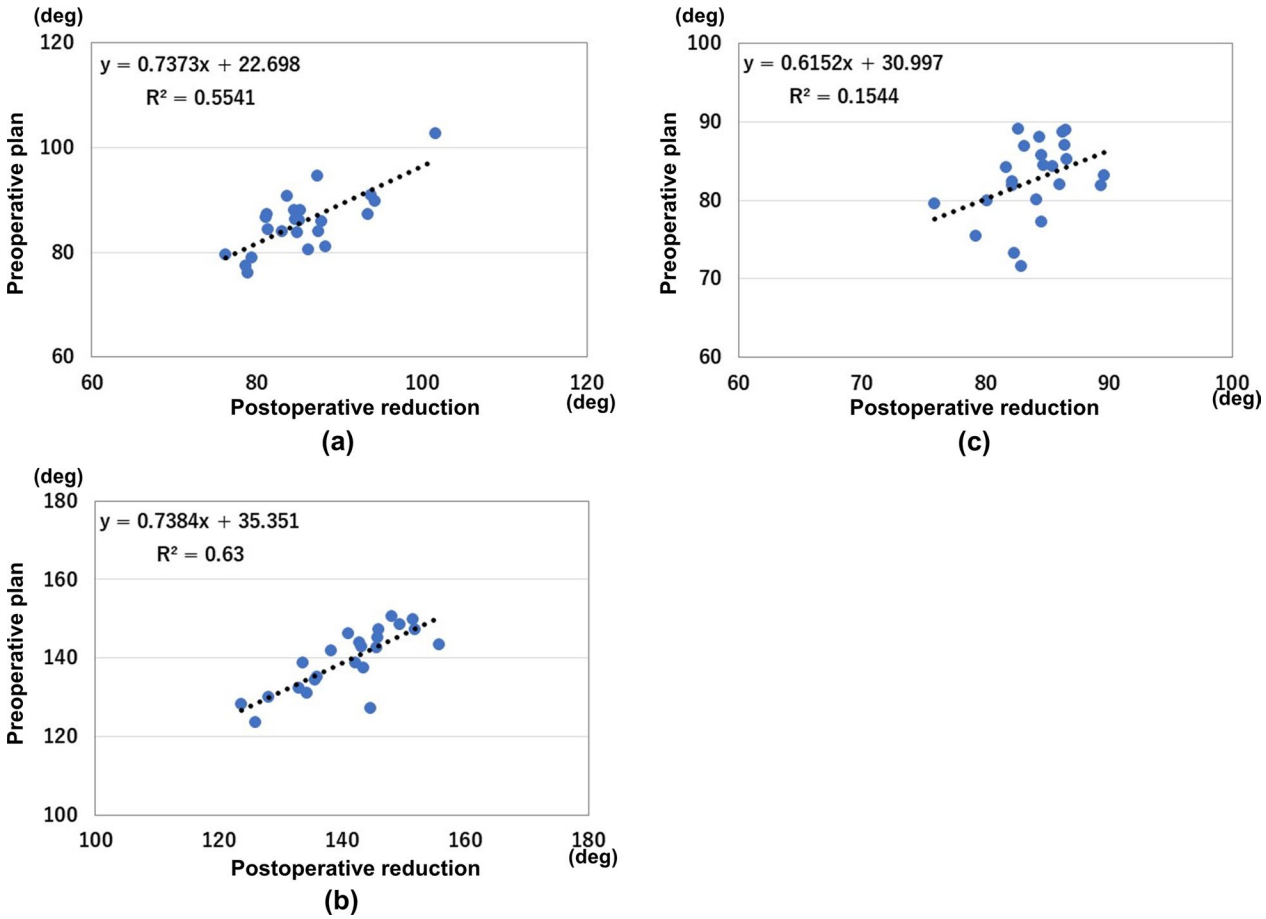


Fig. 5 Results of correlations between the preoperative plan and postoperative reduction. **a** Results of the coronal angle. **b** Results of the sagittal angle. **c** Results of the axial angle

no standard protocol to evaluate the accuracy of the reduction for the 3D preoperative planning. Therefore, we herein attempted to develop a method for evaluating the reduction accuracy in 3D digital preoperative planning for the osteosynthesis of distal humerus fractures. In the original method, the accuracy was assessed from separate images of the preoperative plan and postoperative reduction. Difficulties are sometimes associated with evaluating the accuracy for the correction of rotation using this method. Using a registration algorithm, injured, preoperative planning, and postoperative reduction images were evaluated in the same coordinate system. We propose this method to evaluate the three-dimensional reduction accuracy for the preoperative planning of distal humerus fractures.

The results obtained revealed good correlations for coronal and sagittal parameters and moderate correlation for the axial parameter. This suggests further room for improvements in rotational reduction. In this case series, the majority of incisions were placed posteriorly and the most common approaches were para-tricipital exposure or olecranon osteotomy. In this posterior approach, the posterior surface of the distal humerus bone may be clearly visualized in the surgical field. Therefore, it is relatively easy to evaluate the reduction shape viewed from the coronal direction. In addition, the reduction shape in the sagittal plane may be confirmed by fluoroscopy. On the other hand, difficulties were associated with evaluating the correction of rotation both in the surgical field and fluoroscopy. This may be one reason for the poor correlation of the axial angle. This may need to be improved by creating a reference point on the forearm for a surgical site assessment in future studies. This method and the parameters examined may be useful for confirming the three-dimensional reduction accuracy of preoperative planning in the osteosynthesis of distal humerus fractures.

There are several limitations that need to be addressed. CT scans were needed for 3D preoperative planning. CT has clear advantages in terms of excellent bone and soft tissue contrast and no geometric distortion. However, it exposes the patient to radiation. Precautions need to

be taken to reduce radiation exposure, such as scanning elbows away from the trunk. Furthermore, we did not compare the reduction shape with the unaffected side of the patient's elbow. For assessment of normal anatomical reduction, it may be better to compare the reduction shape with the unaffected side of the elbow. In addition, we did not compare clinical outcomes with the accuracy of reduction in cases without 3D preoperative planning because assessments of reduction based on 3D reference points were only possible when performing 3D preoperative planning. To demonstrate the clinical significance of 3D preoperative planning, the clinical outcomes in different preoperative planning methods need to be examined.

In conclusion, 3D preoperative planning for distal humeral fractures showed the good correlations of coronal and sagittal angles, but the relatively poor correlation of axial angle. This may be attributed to an inability to assess the rotation angle during surgery. We propose the measurement indices shown in the present study as three-dimensional evaluation parameters for the reduction of distal humerus fractures.

Abbreviations

3D	Three-dimensional
CT	Computed tomography
AO	Arbeitsgemeinschaft für Osteosynthesefragen
ICC	Intra-class correlation coefficients

Author contributions

YY contributed to research design, analysis of data, and wrote the manuscript, SI contributed to analysis and interpretation of data and wrote the manuscript, AI contributed to analysis and interpretation of data and wrote the manuscript, SK contributed to analysis and interpretation of data and wrote the manuscript, TO contributed to analysis and interpretation of data and wrote the manuscript, and TI contributed to interpretation of data, wrote the manuscript, and supervised the study. All authors were fully involved in the study and approved the final version of this manuscript.

Funding

This research was supported by grants from JSPS KAKENHI (Grant Number 19K09582), AMED (A324TS), and the National Mutual Insurance Federation of Agricultural Cooperatives. These funders were not involved in data collection, data analysis, or the preparation or editing of the manuscript.

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was performed in line with the principles of the Declaration of Helsinki. This study protocol was approved by the Institutional Review Board of Tokyo Medical University Ibaraki Medical Center. This study was registered as NCT04349319 at ClinicalTrials.gov.

Consent for publication

Written consent for publication was obtained from all study participants.

Competing interests

No benefits in any form have been received or will be received from a commercial party directly or indirectly related to the subject of this article.

Author details

[1]Department of Orthopaedic Surgery, Tokyo Medical University Ibaraki Medical Center, 3-20-1 Chuo, Ami, Inashiki, Ibaraki 300-0395, Japan. [2]Department of Orthopaedic Surgery, University of Tsukuba Hospital, Tsukuba, Ibaraki 305-8576, Japan. [3]Department of Orthopaedic Surgery, Kikkoman General Hospital, Noda, Chiba 278-0005, Japan. [4]Department of Orthopaedic Surgery, Mito Medical Center Hospital, Ibaraki, Ibaraki 311-3193, Japan.

Received: 1 February 2023 Accepted: 1 April 2023
Published online: 08 April 2023

References

1. Rose SH, Melton LJ, Morrey BF, Ilstrup DM, Riggs BL. Epidemiologic features of humeral fractures. Clin Orthop Relat Res. 1982;168:24–30.
2. Anglen J. Distal humerus fractures. J Am Acad Orthop Surg. 2005;13:291–7.
3. Jupiter JB, Mehne DK. Fractures of the distal humerus. Orthopedics. 1992;15:825–33.
4. Amir S, Jannis S, Daniel R. Distal humerus fractures: a review of current therapy concepts. Curr Rev Musculoskelet Med. 2016;9:199–206.
5. Egol K, Koval K, Zuckerman J. Handbook of fractures. Philadelphia: Lippincott Williams & Wilkins; 2010.
6. Wang Y, Zhuo Q, Tang P, Yang W. Cochrane database of systematic reviews. Chichester: Wiley; 1996.
7. Korner J, Diederichs G, Arzdorf M, Lill H, Josten C, Schneider E, Linke B. A biomechanical evaluation of methods of distal humerus fracture fixation using locking compression plates versus conventional reconstruction plates. J Orthop Trauma. 2004;18:286–93.
8. Kinzl L, Fleischmann W. The treatment of distal upper arm fractures. Unfallchirurg. 1991;94:455–60.
9. Pajarinen J, Björkenheim JM. Operative treatment of type C intercondylar fractures of the distal humerus: results after a mean follow-up of 2 years in a series of 18 patients. J Shoulder Elb Surg. 2002;11:48–52.
10. O'Driscoll SW. Optimizing stability in distal humeral fracture fixation. J Shoulder Elbow Surg. 2005;14(1 Suppl S):186S-194S.
11. Lee SK, Kim KJ, Park KH, Choy WS. A comparison between orthogonal and parallel plating methods for distal humerus fractures: a prospective randomized trial. Eur J Orthop Surg Traumatol. 2013;24:1123–31.
12. Holdsworth BJ, Mossad MM. Fractures of the adult distal humerus. Elbow function after internal fixation. J Bone Joint Surg Br. 1900;72:362–5.
13. Jacobson SR, Glisson RR, Urbaniak JR. Comparison of distal humerus fracture fixation: a biomechanical study. J South Orthop Assoc. 1997;6:241–9.
14. Self J, Viegas SF, Buford WL, Patterson RM. A comparison of double-plate fixation methods for complex distal humerus fractures. J Shoulder Elbow Surg. 2005;4(1 Pt 1):10–6.
15. Scolaro JA, Hsu JE, Svach DJ, Mehta S. Plate selection for fixation of extra-articular distal humerus fractures: a biomechanical comparison of three different implants. Injury. 2014;45:2040–4.
16. Zheng G, Nolte LP. Computer-assisted orthopedic surgery: current state and future perspective. Front Surg. 2015;2:66.
17. Hernandez D, Garimella R, Eltorai AEM, Daniels AH. Computer-assisted orthopaedic surgery. Orthop Surg. 2017;9:152–8.
18. Yoshii Y, Kusakabe T, Akita K, Tung WL, Ishii T. Reproducibility of three-dimensional digital preoperative planning for the osteosynthesis of distal radius fractures. J Orthop Res. 2017;35:2646–51.
19. Hsu CL, Chou YC, Li YT, Chen JE, Hung CC, Wu CC, Shen HC, Yeh TT. Pre-operative virtual simulation and three-dimensional printing techniques for the surgical management of acetabular fractures. Int Orthop. 2019;43:1969–76.
20. Yoshii Y, Teramura S, Oyama K, Ogawa T, Hara Y, Ishii T. Development of three-dimensional preoperative planning system for the osteosynthesis of distal humerus fractures. BioMed Eng OnLine. 2020;19:56.
21. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63.
22. Wang H, Wang F, Newman S, Lin Y, Chen X, Xu L, Wang Q. Application of an innovative computerized virtual planning system in acetabular fracture surgery: a feasibility study. Injury. 2016;47:1698–701.

23. Boudissa M, Courvoisier A, Chabanas M, Tonetti J. Computer assisted surgery in preoperative planning of acetabular fracture surgery: state of the art. Expert Rev Med Devices. 2018;15:81–9.
24. Clarke E. Virtual reality simulation—the future of orthopaedic training? A systematic review and narrative analysis. Adv Simul. 2021;6:2.

Publisher's Note

Received: 2 March 2023 | Accepted: 3 April 2023

DOI: 10.1111/1759-7714.14904

ORIGINAL ARTICLE

WILEY

Clinical benefit of platinum doublet combination therapy in older adults with advanced non-small cell lung cancer: A prospective multicenter study by the National Hospital Organization in Japan

Mototsugu Shimokawa[1,2] | Masaki Kanazu[3] | Ryusei Saito[4] | Masahide Mori[3] | Atsuhisa Tamura[5] | Yoshio Okano[6] | Yuka Fujita[7] | Takeo Endo[8] | Mitsuru Motegi[9] | Shohei Takata[10] | Toshiyuki Kita[11] | Noriaki Sukoh[12] | Fumitaka Mizuki[13] | Mitsuhiro Takenoyama[14] | Shinji Atagi[15]

[1]Department of Biostatistics, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan

[2]Clinical Research Institute, National Hospital Organization Kyushu Cancer Center, Fukuoka, Japan

[3]Department of Thoracic Oncology, National Hospital Organization Osaka Toneyama Medical Center, Toyonaka, Japan

[4]Division of Respiratory Medicine, National Hospital Organization Shibukawa Medical Center, Shibukawa, Japan

[5]Department of Respiratory Diseases, National Hospital Organization Tokyo National Hospital, Tokyo, Japan

[6]Division of Pulmonary Medicine, National Hospital Organization Kochi Hospital, Kochi, Japan

[7]Department of Respiratory Medicine, National Hospital Organization Asahikawa Medical Center, Asahikawa, Japan

[8]Department of Respiratory Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan

[9]Department of Respiratory Medicine, National Hospital Organization Takasaki General Medical Center, Takasaki, Japan

[10]Department of Respiratory Medicine, National Hospital Organization Fukuokahigashi Medical Center, Fukuoka, Japan

[11]Department of Respiratory Medicine, National Hospital Organization Kanazawa Medical Center, Kanazawa, Japan

[12]Department of Respiratory Medicine, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan

[13]Center for Clinical Research, Yamaguchi University Hospital, Yamaguchi, Japan

[14]Department of Thoracic Oncology, National Hospital Organization Kyushu Cancer Center, Fukuoka, Japan

[15]Department of Thoracic Oncology, National Hospital Organization Kinki-Chuo Chest Medical Center, Sakai, Japan

Correspondence

Mototsugu Shimokawa, Department of Biostatistics, Yamaguchi University Graduate School of Medicine Minamikogushi, Ube, Yamaguchi, 755-8505, Japan.
Email: moto@yamaguchi-u.ac.jp

Funding information
National Hospital Organization in Japan

Abstract

Background: Previous trials suggest that older adults with non-small cell lung cancer (NSCLC) derive benefit from platinum doublet combination therapy, but its superiority is controversial. Although geriatric assessment variables are used to assess the individual risk of severe toxicity and clinical outcomes in older patients, the standard first-line treatment is still debated. Therefore, we aimed to identify the risk factors for clinical outcomes in older patients with NSCLC.

Methods: Patients aged ≥75 years with advanced NSCLC treated at any of 24 National Hospital Organization institutions completed a pre-first-line chemotherapy assessment, including patient characteristics, treatment variables, laboratory test values, and geriatric assessment variables. We evaluated whether these variables were the risk factors for progression-free survival (PFS) and overall survival (OS).

Results: A total of 148 patients with advanced NSCLC were treated with combination therapy ($n = 90$) or monotherapy ($n = 58$). Median PFS was 5.3 months and OS was 13.6 months. We identified that hypoalbuminemia (hazard ratio [HR] 2.570, 95% confidence interval [CI]: 1.117–5.913, $p = 0.0264$) was a risk factor for PFS and monotherapy

(HR 1.590, 95% CI: 1.070–2.361, $p = 0.0217$), lactate dehydrogenase (HR 3.682, 95% CI: 1.013–13.39, $p = 0.0478$), and high C-reactive protein (HR 2.038, 95% CI: 1.141–3.642, $p = 0.0161$) were risk factors for OS. The median OS was significantly longer in patients treated with combination therapy than in those who received monotherapy (16.5 months vs. 10.3 months; HR 0.684, 95% CI: 0.470–0.995, $p = 0.0453$).

Discussion: Platinum doublet combination therapy may be beneficial in older patients with NSCLC. Identification of risk factors will assist in the development of a personalized treatment strategy.

KEYWORDS

non-small cell lung cancer, older patients, platinum doublet therapy, single-agent chemotherapy

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide, and the majority of patients diagnosed with the disease have non-small cell lung cancer (NSCLC).[1] About 70% of patients with NSCLC are diagnosed at an advanced stage, and the median age at diagnosis is 70 years.[2] Although systemic chemotherapy is one of the therapeutic options available for patients with advanced NSCLC, the standard first-line treatment for older patients is still debated. Previous trials have suggested that older patients with NSCLC benefit from platinum doublet combination therapy, whereas its superiority continues to be debated.[3–6] An explanation for this controversy is that there is considerable heterogeneity in the physiological changes that occur with aging. Furthermore, a low number of "fit" older patients are enrolled in clinical trials. Therefore, it is difficult to predict the tolerability of chemotherapy in "unfit" older adult patients in clinical practice, because they are more vulnerable to chemotherapy-related adverse events than "fit" older adult patients.

Age is an important factor in management decisions because of the complex interplay between normal age-related decline and comorbidities. The Karnofsky performance status or Eastern Cooperative Oncology Group performance status (ECOG-PS) is used in patients to predict treatment toxicity and survival.[7,8] However, these tools were validated in younger adults and are not suitable for predicting vulnerability to chemotherapy in older patients. Other factors, including comorbidity, nutrition, physical and cognitive function, and social support, also correlate with toxicity of therapy and cancer outcomes.[9] The Comprehensive Geriatric Assessment (CGA), which is a compilation of standardized tools for assessment of these factors, can help to predict mortality in older patients with cancer.[10–13]

Although the CGA is too complicated for use in daily clinical practice, it has been validated among oncologists.[14,15] Furthermore, several studies have investigated how to predict the risk of chemotherapy toxicity and found that a certain subgroup of older patients are more vulnerable to adverse events from chemotherapy.[16–18] We have previously reported a risk stratification system for prediction of vulnerability to chemotherapy in older patients with NSCLC.[19] In this study, we evaluated whether several variables, including patients' characteristics and the treatment variables, were the risk factors for progression-free survival (PFS) and overall survival (OS).

METHODS

Patients

In total, 354 patients from any of 24 National Hospital Organization institutions were enrolled in this prospective study between April 2013 and March 2017. A total of 148 of these patients were aged ≥75 years and had histologically or cytologically proven advanced NSCLC (according to the TNM classification, seventh edition) and were treated with cytotoxic chemotherapy (platinum doublet therapy or a single agent) as first-line therapy. Patients were excluded from the study if they had had active malignancy within the previous 5 years, a history of chemotherapy, had massive pleural or pericardial effusion or ascites, or had received radiation therapy to the lung. The study was approved by the National Hospital Organization Central Review Board and conducted in accordance with the Declaration of Helsinki and ethical guidelines for clinical research (UMIN000010384). All patients provided their written informed consent before enrollment.

Study schema

All patients completed a pre-first-line chemotherapy assessment, which included the characteristics of the cancer (tumor type, stage, and driver mutation status), treatment variables, laboratory test values, and geriatric assessment variables. The ability to perform activities of daily living was assessed using the Barthel Index.[20] Independence in everyday living and dementia were evaluated by physicians. Hearing and falls in the previous 6 months were evaluated by self-report or by the family. The patients were followed through one cycle of chemotherapy to monitor for grade 3 (severe) to grade 5 (death) adverse events according to the National Cancer Institute Common Terminology Criteria for Adverse Events, version 4.0. The antitumor response to treatment was assessed on the basis of the Response Evaluation Criteria in Solid Tumors (version 1.1) using computed tomography scans. PFS was defined as the interval between

TABLE 1 Patient demographics and clinical characteristics at baseline.

Characteristics	Monotherapy ($N = 58$)	Combination ($N = 90$)	Overall ($N = 148$)	*p*-value
Age				
Mean ± SD	80.1 ± 3.2	77.7 ± 2.4	78.6 ± 3.0	<0.0001
Median (range)	80.0 (75.0–88.0)	77.0 (75.0–86.0)	78 (75.0–88.0)	
Sex, *n* (%)				0.8551
Male	47 (81.0)	74 (82.2)	121 (81.8)	
Female	11 (19.0)	16 (17.8)	27 (18.2)	
Stage, *n* (%)				0.0699
III B	3 (5.2)	16 (17.8)	19 (12.8)	
IV	48 (82.8)	62 (68.9)	110 (74.3)	
recurrence	7 (12.1)	12 (13.3)	19 (12.8)	
Histology, *n* (%)				0.9307
Adenocarcinoma	39 (67.2)	62 (68.9)	101 (68.2)	
Non-small cell carcinoma	4 (6.9)	7 (7.8)	11 (7.4)	
Squamous cell carcinoma	15 (25.9)	21 (23.3)	36 (24.3)	
Mutation status (EGFR or ALK), *n* (%)				0.8742
Mutation	4 (6.9)	8 (8.9)	12 (8.1)	
Wild-type	29 (50.0)	42 (46.7)	71 (48.0)	
Unknown	25 (43.1)	40 (44.4)	65 (43.9)	
BMI, *n* (%)				0.2054
<22	30 (51.7)	37 (41.1)	67 (45.3)	
≥22	28 (48.3)	53 (58.9)	81 (54.7)	
ECOG-PS, *n* (%)				0.4683
0–1	53 (91.4)	85 (94.4)	138 (93.2)	
≥2	5 (8.6)	5 (5.6)	10 (6.8)	
CCI, *n* (%)				0.3559
0–1	47 (81.0)	78 (86.7)	125 (84.5)	
≥2	11 (19.0)	12 (13.3)	23 (15.5)	
Bodyweight loss, *n* (%)				0.9950
<5%	49 (84.5)	76 (84.4)	125 (84.5)	
≥5%	9 (15.5)	14 (15.6)	23 (15.5)	
Frail, *n* (%)				0.3559
1	47 (81.0)	78 (86.7)	125 (84.5)	
≥2	11 (19.0)	12 (13.3)	23 (15.5)	
Recognition, *n* (%)				0.1649
1	53 (91.4)	87 (96.7)	140 (94.6)	
≥2	5 (8.6)	3 (3.3)	8 (5.4)	
Barthel Index, *n* (%)				0.6647
<85	4 (6.9)	8 (8.9)	12 (8.1)	
≥85	54 (93.1)	82 (91.1)	136 (91.9)	
MMSE, *n* (%)				0.8265
<27	21 (36.2)	31 (34.4)	52 (35.1)	
≥27	37 (63.8)	59 (65.6)	96 (64.9)	
Dose reduction, *n* (%)				
No	47 (81.0)	24 (26.7)	71 (48.0)	<0.0001
Yes	11 (19.0)	66 (73.3)	77 (52.0)	
Anemia, *n* (%)				0.9198
0–1	55 (94.8)	85 (94.4)	140 (94.6)	

(Continues)

TABLE 1 (Continued)

Characteristics	Monotherapy ($N = 58$)	Combination ($N = 90$)	Overall ($N = 148$)	p-value
≥2	3 (5.2)	5 (5.6)	8 (5.4)	
Hypoalbuminemia, *n* (%)				0.7803
0–1	50 (86.2)	79 (87.8)	129 (87.2)	
≥2	8 (13.8)	11 (12.2)	19 (12.8)	
Creatinine, *n* (%)				0.9778
None	51 (87.9)	79 (87.8)	130 (87.8)	
≥1	7 (12.1)	11 (12.2)	18 (12.2)	
LDH, *n* (%)				0.6534
<460	56 (96.6)	88 (97.8)	144 (97.3)	
≥460	2 (3.4)	2 (2.2)	4 (2.7)	
CRP, *n* (%)				0.8509
<3	45 (77.6)	71 (78.9)	116 (78.4)	
≥3	13 (22.4)	19 (21.1)	32 (21.6)	

Abbreviations: BMI, body mass index; CCI, Charlson Comorbidity Index; CRP, C-reactive protein; ECOG-PS, Eastern Cooperative Oncology Group performance status; LDH, lactate dehydrogenase; MMSE, Mini-Mental State Examination; SD, standard deviation.

treatment and the date of the first documented tumor progression, as determined by the attending physicians, or death from any cause, whichever occurred first. For cases without computed tomography examination but wherein clinical symptoms or findings on a chest radiograph suggested progression of disease, progression disease onset was defined as the date when the physician clinically evaluated the progression of disease. OS was defined as the interval between the date of diagnosis and date of death or date of last follow-up for censored patients.

Statistical analysis

Patient characteristics are summarized using descriptive statistics or contingency tables. Associations between treatments and patient characteristics were examined using the unpaired *t*-test for continuous variables and the chi-squared test for categorical variables. PFS and OS were estimated using the Kaplan–Meier method, and the survival curves were compared with the log-rank test and a Cox proportional hazards model. The risk factors for PFS and OS were evaluated using a Cox proportional hazards model. In the multivariable analysis, all variables were evaluated for univariate analysis were selected. All statistical analyses were performed using SAS 9.4 (SAS Institute Inc.). A two-sided p-value of ≤ 0.05 was considered statistically significant.

RESULTS

Patient and treatment characteristics

A total of 148 patients were included in the analysis (Table 1). The median age was 78 years (range, 75–88). The proportion of patients with stage IIIB NSCLC was 12.8%, and those of patients including stage IV and recurrence was 87.2%. About four-fifths of the patients were male (81.8%), and the most common tumor type was adenocarcinoma (68.2%). In terms of first-line cytotoxic chemotherapy, more patients received platinum doublet combination chemotherapy (60.8%) than monotherapy (39.2%). About half of the patients (48.0%) were treated with standard doses for nonelderly patients.

Geriatric assessment

A total of 23 patients (15.5%) were at high risk of complications (Charlson Comorbidity Index[21] ≥2 points). Body mass index (BMI, calculated as kg/m^2) ranged from 15.5 to 32.9, with about half (54.7%) of the patients having a BMI ≥22. Twenty-three patients (15.5%) had experienced weight loss of ≥5% within the previous 6 months. Although most patients had good performance status (ECOG-PS 0/1, 93.2%), patients with poor performance status (ECOG-PS ≥2, 6.8%) were also enrolled. Twenty-three patients (15.5%) were limited in their ability to perform activities of daily living. Twelve patients (8.1%) had a Barthel Index score of <85 points. About one-third of the patients (35.1%) scored <27 points on the Mini-Mental State Examination.[22] Eight patients (5.4%) had trouble in everyday living because of cognitive dysfunction.

Effects of combination therapy and monotherapy on PFS and OS

The median PFS in patients treated with platinum doublet combination therapy was 5.8 months (95% CI: 4.9–7.0),

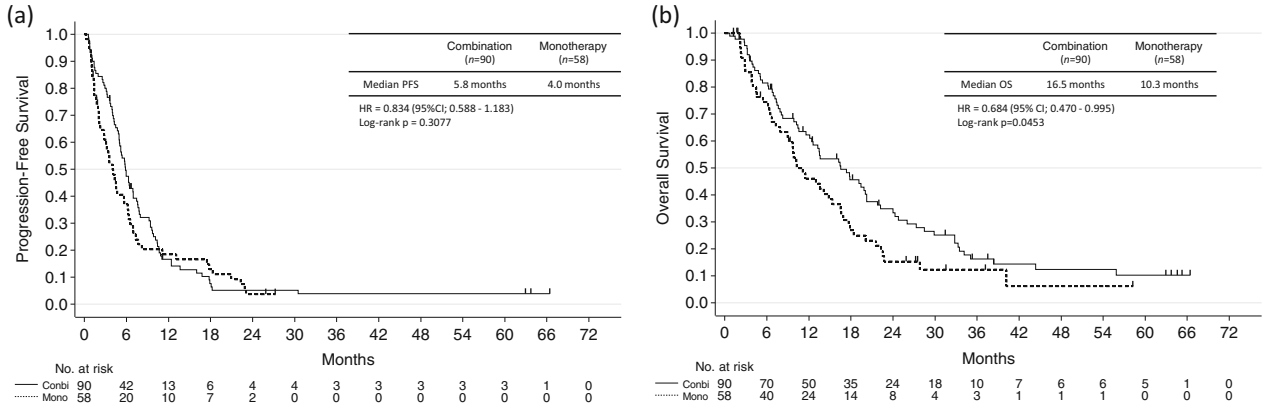


FIGURE 1 Progression-free survival and overall survival in the single-agent chemotherapy group and the platinum doublet combination therapy group. CI, confidence interval; HR, hazard ratio; OS, overall survival; PFS, progression-free survival.

which was not significantly different from the median PFS of 4.0 months (95% CI: 2.8–5.7) in patients who received monotherapy (HR 0.834, 95% CI: 0.588–1.183, $p = 0.3077$) (Figure 1a). However, the median OS in patients treated with combination therapy was 16.5 months (95% CI: 12.1–20.2), which was significantly longer than that of 10.3 months (95% CI: 7.9–15.4) in patients who received monotherapy (HR 0.684, 95% CI: 0.470–0.995, $p = 0.0453$) (Figure 1b).

The overall response rate was higher in the combination cohort than in the single-agent cohort (31.1% vs. 13.8%). The disease control rates were similar between these cohorts (Table 2).

Toxicity of chemotherapy

A total of 62 patients developed severe hematological toxicity (grade 3–5 according to the National Cancer Institute Common Terminology Criteria for Adverse Events, version 4.0). The frequency of severe hematological toxicity was higher in patients treated with single-agent chemotherapy than in those treated with platinum doublet combination therapy (51.7% vs. 35.6%). In terms of nonhematological toxicity, adverse events emerged in most cases at any grades. The frequency of severe nonhematological toxicity was not high and was similar between the combination and monotherapy groups (Table 3).

Risk factors for PFS and OS

The median follow-up duration was 11.9 months. Median PFS was 5.3 months (95% CI: 4.4–6.3) and median OS was 13.6 months (95% CI: 10.5–17.3) for all patients (Figure 2). Patient sex, Mini-Mental State Examination score, performance status, Charlson Comorbidity Index value, weight loss, frailty, recognition, Barthel Index, MMSE, anemia, hypoalbuminemia, creatinine, lactate dehydrogenase (LDH), C-reactive protein (CRP), and platinum doublet combination therapy or single-agent chemotherapy were evaluated as potential risk factors in the Cox proportional hazards

TABLE 2 Response rates in the single-agent chemotherapy group and the platinum doublet combination therapy group.

Response rate	Monotherapy ($N = 58$), *n* (%)	Combination ($N = 90$), *n* (%)	Overall ($N = 148$), *n* (%)
CR	1 (1.7)	1 (1.1)	2 (1.4)
PR	7 (12.1)	27 (30.0)	34 (23.0)
SD	34 (58.6)	42 (46.7)	76 (51.4)
PD	14 (24.1)	19 (21.1)	33 (22.3)
NE	2 (3.4)	1 (1.1)	3 (2.0)
Overall response rate			
n (%)	8 (13.8)	28 (31.1)	36 (24.4)
[95% CI]	[6.1–25.4]	[21.8–41.7]	[17.7–32.1]
Disease control rate			
n (%)	42 (72.4)	70 (77.8)	112 (75.7)
[95% CI]	[59.1–83.3]	[67.8–85.9]	[68.0–82.4]

Abbreviations: CI, confidence interval; CR, complete response; NE, not evaluable; PD, progressive disease; PR, partial response; SD, stable disease.

TABLE 3 Adverse events.

Adverse event	Monotherapy ($N = 58$)	Combination ($N = 90$)	Overall ($N = 148$)
Hematological toxicity, *n* (%)			
Any	49 (84.5)	83 (92.2)	132 (89.2)
≥grade 3	30 (51.7)	32 (35.6)	62 (41.9)
Non-hematological toxicity, *n* (%)			
Any	55 (94.8)	87 (96.7)	142 (95.9)
≥grade 3	8 (13.8)	13 (14.4)	21 (14.2)

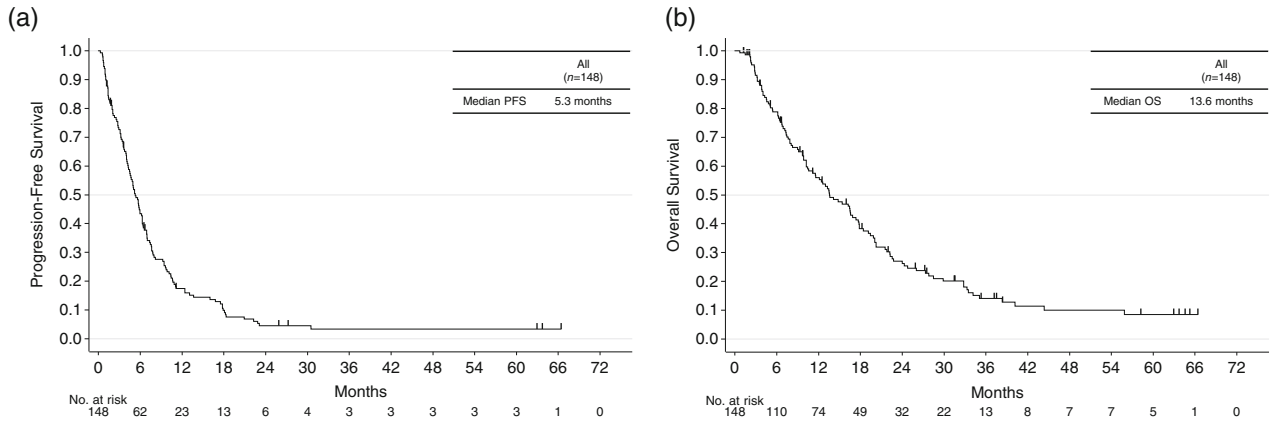


FIGURE 2 Progression-free survival and overall survival in the entire study population. OS, overall survival; PFS, progression-free survival.

model. Hypoalbuminemia was the only independent risk factor for PFS (hazard ratio [HR] 2.570, 95% CI: 1.117–5.913, $p = 0.0264$) (Table 4). Independent risk factors for OS were monotherapy (HR 1.590, 95% CI: 1.070–2.361, $p = 0.0217$), high LDH (HR 3.682, 95% CI: 1.013–13.39, $p = 0.0478$), and high CRP (HR 2.038, 95% CI: 1.141–3.642, $p = 0.0161$) (Table 4).

DISCUSSION

The correlation between CGA and clinical outcomes in older patients with cancer has previously been investigated.[9–13] There is considerable heterogeneity in older patients with cancer in terms of physiological changes, and it is difficult to identify factors that predict clinical outcomes, including the negative effects of adverse events. Previous studies have identified several risk factors that could predict the frequency of severe adverse events in older patients.[16–18] Furthermore, we have developed a risk stratification tool to predict vulnerability to chemotherapy in older patients with NSCLC.[19] However, long-term survival is the essential factor when considering the treatment strategy.

Two recent randomized Phase III trials that compared docetaxel with platinum doublet combination chemotherapy (carboplatin plus pemetrexed, and carboplatin plus nab-paclitaxel) found that combination chemotherapy was tolerable and highly effective for "fit" older adults.[4,5] A certain number of "unfit" older patients were included in our study, and our findings suggested the efficacy of combination chemotherapy in the clinical setting, especially in terms of survival. Although the difference may be more pronounced in the long-term survival than PFS, further discussion is limited as there is a bias in the first-line therapy regimen, which is chosen by the attending physicians. Furthermore, the Japanese population appears to be more susceptible to toxicities from docetaxel.[23] Docetaxel or docetaxel plus bevacizumab was administered in about half of the patients treated with monotherapy in our study (Table S1). This high number of patients in the docetaxel group may be the reason why monotherapy emerged as one of the risk factors for overall survival.

In this study, we also identified high LDH and CRP levels as risk factors for OS. However, LDH is elevated not only in patients with cancer but also in those with other diseases, so the prognostic role of LDH in patients with lung cancer is not conclusive.[24–26] In addition to survival, LDH is one of the factors predicting the risk of chemotherapy toxicity in older patients with cancer.[16] The CRP to albumin ratio (high CRP and hypoalbuminemia), reflecting prolonged exhaustion owing to inflammation, may be a potential prognostic factor in patients with cancer.[27–29] In addition to CRP, hypoalbuminemia was extracted as an independent risk factor for PFS in our study. Nutritional status as well as tumor inflammation may be more relevant to clinical outcome.

This study had several limitations. First, there was a degree of bias in the treatment variables because the decision regarding selection of the first-line chemotherapy regimen or dose de-escalation was made by the physicians, as in clinical practice. Second, patients with NSCLC treated with immune checkpoint inhibitors (ICIs) as first-line therapy were not included. As previously reported, pembrolizumab has a clinical benefit in patients with advanced NSCLC, regardless of patient age.[30,31] Although the combination of platinum doublet chemotherapy and an ICI has emerged as one of the standard therapies for patients with advanced NSCLC,[32–35] its safety in older patients is uncertain. When choosing a therapeutic strategy involving an ICI, we should understand the factors influencing clinical outcomes before starting chemotherapy.

In conclusion, our study has identified several laboratory values that reflect prolonged exhaustion owing to inflammation and might be predictors of outcomes in older patients with NSCLC. Moreover, platinum doublet combination therapy may be of benefit in this population in the clinical setting.

TABLE 4 Risk factors for progression-free survival and overall survival.

	Progression-free survival				Overall survival			
	Univariate		Multivariate		Univariate		Multivariate	
Risk factor	**OR (95% CI)**	***p*-value**	**OR (95% CI)**	***p*-value**	**OR (95% CI)**	***p*-value**	**OR (95% CI)**	***p*-value**
Sex: male vs. female	1.164 (0.735–1.843)	0.5176	1.174 (0.712–1.935)	0.5289	1.296 (0.792–2.121)	0.3023	1.403 (0.799–2.463)	0.2392
BMI: ≥22 vs. <22	1.130 (0.804–1.587)	0.4823	0.969 (0.662–1.419)	0.8733	1.237 (0.857–1.786)	0.2559	0.991 (0.650–1.511)	0.9660
ECOG-PS: ≥2 vs. 0–1	3.474 (1.719–7.022)	0.0005	2.619 (0.854–8.032)	0.0921	3.145 (1.520–6.507)	0.0020	2.353 (0.670–8.257)	0.1817
CCI: 0–1 vs. ≥2	1.017 (0.638–1.620)	0.9441	0.807 (0.464–1.404)	0.4477	1.217 (0.717–2.065)	0.4662	0.936 (0.510–1.717)	0.8306
Bodyweight loss: ≥5% vs. <5%	1.108 (0.681–1.803)	0.6793	0.752 (0.405–1.394)	0.3652	1.508 (0.899–2.532)	0.1197	1.155 (0.606–2.198)	0.6618
Frail: ≥2 vs. 1	1.239 (0.770–1.995)	0.3771	1.667 (0.741–3.752)	0.2166	1.519 (0.904–2.550)	0.1143	1.376 (0.569–3.325)	0.4787
Recognition 1 vs. ≥2	1.934 (0.851–4.395)	0.1153	2.594 (0.895–7.516)	0.0791	1.173 (0.478–2.883)	0.7275	1.209 (0.394–3.712)	0.7397
Barthel Index: <85 vs. ≥85	1.369 (0.738–2.542)	0.3195	0.908 (0.371–2.221)	0.8333	1.668 (0.893–3.114)	0.1084	0.861 (0.326–2.276)	0.7634
MMSE: ≥27 vs. <27	1.184 (0.829–1.692)	0.3533	1.063 (0.694–1.627)	0.7803	1.110 (0.758–1.626)	0.5920	1.079 (0.683–1.705)	0.7439
Anemia: ≥2 vs. 0–1	3.014 (1.457–6.231)	0.0029	0.999 (0.374–2.666)	0.9977	4.742 (2.008–11.19)	0.0004	2.745 (0.992–7.598)	0.0519
Hypoalbuminemia: ≥2 vs. 0–1	3.452 (2.029–5.876)	<0.0001	2.570 (1.117–5.913)	0.0264	2.882 (1.647–5.042)	0.0002	1.309 (0.567–3.021)	0.5279
Creatinine: None vs. ≥1	1.304 (0.742–2.291)	0.3563	1.433 (0.748–2.744)	0.2778	1.480 (0.790–2.772)	0.2208	2.158 (0.980–4.752)	0.0563
LDH: ≥460 vs. <460	2.195 (0.691–6.979)	0.1826	2.751 (0.807–9.384)	0.1059	1.868 (0.591–5.903)	0.2874	3.682 (1.013–13.39)	0.0478
CRP: ≥3 vs. <3	2.547 (1.661–3.905)	<0.0001	1.612 (0.902–2.881)	0.1068	2.525 (1.620–3.935)	<0.0001	2.038 (1.141–3.642)	0.0161
Single agent chemotherapy vs. platinum doublet combination therapy	0.834 (0.588–1.183)	0.3098	0.836 (0.572–1.222)	0.3560	0.684 (0.470–0.995)	0.0468	0.629 (0.423–0.934)	0.0217

Abbreviations: BMI, body mass index; CCI, Charlson Comorbidity Index; CI, confidence interval; CRP, C-reactive protein; ECOG-PS, Eastern Cooperative Oncology Group performance status; LDH, lactate dehydrogenase; MMSE, Mini-Mental State Examination.

AUTHOR CONTRIBUTIONS

Conceptualization: Mototsugu Shimokawa, Masaki Kanazu, Shinji Atagi. Data curation: Ryusei Saito, Masahide Mori, Atsuhisa Tamura, Yoshio Okano, Yuka Fujita, Takeo Endo, Mitsuru Motegi, Shohei Takata, Toshiyuki Kita, Noriaki Sukoh, Mitsuhiro Takenoyama, Shinji Atagi. Formal analysis: Mototsugu Shimokawa, Masaki Kanazu, Fumitaka Mizuki. Funding acquisition: Masaki Kanazu. Investigation: Mototsugu Shimokawa, Masaki Kanazu. Methodology: Mototsugu Shimokawa, Masaki Kanazu, Shinji Atagi. Supervision: Fumitaka Mizuki, Shinji Atagi. Validation: Fumitaka Mizuki, Mototsugu Shimokawa. Visualization: Mototsugu Shimokawa, Masaki Kanazu. Writing: Mototsugu Shimokawa, Masaki Kanazu. Review and editing: Ryusei Saito, Masahide Mori, Atsuhisa Tamura, Yoshio Okano, Yuka Fujita, Takeo Endo, Mitsuru Motegi, Shohei Takata, Toshiyuki Kita, Noriaki Sukoh, Fumitaka Mizuki, Mitsuhiro Takenoyama, Shinji Atagi.

ACKNOWLEDGMENTS

We thank all investigators and patients, including their families, for their participation in this study. We are also grateful to Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

FUNDING INFORMATION

This study was supported financially by the National Hospital Organization in Japan.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

Mototsugu Shimokawa https://orcid.org/0000-0001-8140-4565

Masaki Kanazu https://orcid.org/0000-0003-4405-1929

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7–30.
2. Owonikoko TK, Ragin CC, Belani CP, Oton AB, Gooding WE, Taioli E, et al. Lung cancer in elderly patients: an analysis of the surveillance, epidemiology, and end results database. J Clin Oncol. 2007; 25(35):5570–7.
3. Quoix E, Zalcman G, Oster JP, Westeel V, Pichon E, Lavolé A, et al. Carboplatin and weekly paclitaxel doublet chemotherapy compared with monotherapy in elderly patients with advanced non-small-cell lung cancer: IFCT-0501 randomised, phase 3 trial. Lancet. 2011; 378(9796):1079–88.
4. Okamoto I, Nokihara H, Nomura S, Niho S, Sugawara S, Horinouchi H, et al. Comparison of carboplatin plus pemetrexed followed by maintenance pemetrexed with docetaxel monotherapy in elderly patients with advanced nonsquamous non-small cell lung cancer: a phase 3 randomized clinical trial. JAMA Oncol. 2020;6(5): e196828.
5. Kogure Y, Iwasawa S, Saka H, Hamamoto Y, Kada A, Hashimoto H, et al. Efficacy and safety of carboplatin with nab-paclitaxel versus docetaxel in older patients with squamous non-small-cell lung cancer (CAPITAL): a randomised, multicentre, open-label, phase 3 trial. Lancet. 2021;2(12):e791–800.
6. Abe T, Takeda K, Ohe Y, Kudoh S, Ichinose Y, Okamoto H, et al. Randomized phase III trial comparing weekly docetaxel plus cisplatin versus docetaxel monotherapy every 3 weeks in elderly patients with advanced non-small-cell lung cancer: the intergroup trial JCOG0803/WJOG4307L. J Clin Oncol. 2015;33(6):575–81.
7. Albain KS, Crowley JJ, LeBlanc M, Livingston RB. Survival determinants in extensive-stage non-small-cell lung cancer: the southwest oncology group experience. J Clin Oncol. 1991;9(9):1618–26.
8. Sweeney CJ, Zhu J, Sandler AB, Schiller J, Belani CP, Langer C, et al. Outcome of patients with a performance status of 2 in eastern cooperative oncology group study E1594: a phase II trial in patients with metastatic nonsmall cell lung carcinoma. Cancer. 2001;92(10):2639–47.
9. Dale W, Mohile SG, Eldadah BA, Trimble EL, Schilsky RL, Cohen HJ, et al. Biological, clinical, and psychosocial correlates at the interface of cancer and aging research. J Natl Cancer Inst. 2012;104(8):581–9.
10. Extermann M, Hurria A. Comprehensive geriatric assessment for older patients with cancer. J Clin Oncol. 2007;25(14):1824–31.
11. Kanesvaran R, Li H, Koo KN, Poon D. Analysis of prognostic factors of comprehensive geriatric assessment and development of a clinical scoring system in elderly Asian patients with cancer. J Clin Oncol. 2011;29(27):3620–7.
12. Wildiers H, Heeren P, Puts M, Topinkova E, Janssen-Heijnen MLG, Extermann M, et al. International Society of Geriatric Oncology consensus on geriatric assessment in older patients with cancer. J Clin Oncol. 2014;32(24):2595–603.
13. Corre R, Greillier L, Le Caër H, et al. Use of a comprehensive geriatric assessment for the Management of Elderly Patients with Advanced non-Small-Cell Lung Cancer: the phase III randomized ESOGIA-GFPC-GECP 08-02 study. J Clin Oncol. 2016;34(13):1476–83.
14. Mohile SG, Dale W, Somerfield MR, Schonberg MA, Boyd CM, Burhenn PS, et al. Practical assessment and Management of Vulnerabilities in older patients receiving chemotherapy: ASCO guideline for geriatric oncology. J Clin Oncol. 2018 Aug;36(22):2326–47.
15. Gajra A, Jeune-Smith Y, Fortier S, Feinberg B, Phillips E Jr, Balanean A, et al. The use and knowledge of validated geriatric assessment instruments among US Community oncologists. JCO Oncol Pract. 2022;18(7):e1081–90.
16. Extermann M, Boler I, Reich RR, Lyman GH, Brown RH, DeFelice J, et al. Predicting the risk of chemotherapy toxicity in older patients: the chemotherapy risk assessment scale for high-age patients (CRASH) score. Cancer. 2012 Jul;118(13):3377–86.
17. Hurria A, Togawa K, Mohile SG, Owusu C, Klepin HD, Gross CP, et al. Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study. J Clin Oncol. 2011;29(25):3457–65.
18. Hurria A, Mohile S, Gajra A, Klepin H, Muss H, Chapman A, et al. Validation of a prediction tool for chemotherapy toxicity in older adults with cancer. J Clin Oncol. 2016;34(20):2366–71.
19. Kanazu M, Shimokawa M, Saito R, Mori M, Tamura A, Okano Y, et al. Predicting systemic therapy toxicity in older adult patients with advanced non-small cell lung cancer: a prospective multicenter study of National Hospital Organization in Japan. J Geriatr Oncol. 2022; 13(8):1216–22.
20. Mahoney FI, Barthel DW. Functional EVALUATION: the BARTHEL index. Md State Med J. 1965;14:61–5.
21. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83.
22. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189–98.
23. Kenmotsu H, Tanigawara Y. Pharmacokinetics, dynamics and toxicity of docetaxel: why the Japanese dose differs from the Western dose. Cancer Sci. 2015;106(5):497–504.
24. Fan S, Guan Y, Zhao G, An G. Association between plasma fibrinogen and survival in patients with small-cell lung carcinoma. Thorac Cancer. 2018;9(1):146–51.

25. Fiegl M, Pircher A, Waldthaler C, Gamerith G, Kocher F, Pall G, et al. Small steps of improvement in small-cell lung cancer (SCLC) within two decades: a comprehensive analysis of 484 patients. Lung Cancer. 2014;84(2):168–74.
26. Lee DS, Park KR, Kim SJ, Chung MJ, Lee YH, Chang JH, et al. Serum lactate dehydrogenase levels at presentation in stage IV non-small cell lung cancer: predictive value of metastases and relation to survival outcomes. Tumour Biol. 2016;37(1):619–25.
27. Zhou T, Zhan J, Hong S, Hu Z, Fang W, Qin T, et al. Ratio of C-reactive protein/albumin is An inflammatory prognostic score for predicting overall survival of patients with small-cell lung cancer. Sci Rep. 2015;5:10481.
28. Miyazaki T, Yamasaki N, Tsuchiya T, Matsumoto K, Kunizaki M, Kamohara R, et al. Ratio of C-reactive protein to albumin is a prognostic factor for operable non-small-cell lung cancer in elderly patients. Surg Today. 2017;47(7):836–43.
29. Koh YW, Lee HW. Prognostic impact of C-reactive protein/albumin ratio on the overall survival of patients with advanced nonsmall cell lung cancers receiving palliative chemotherapy. Medicine (Baltimore). 2017;96(19):e6848.
30. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med. 2016;375(19):1823–33.
31. Nosaki K, Saka H, Hosomi Y, Baas P, de Castro G Jr, Reck M, et al. Safety and efficacy of pembrolizumab monotherapy in elderly patients with PD-L1-positive advanced non-small-cell lung cancer: pooled analysis from the KEYNOTE-010, KEYNOTE-024, and KEYNOTE-042 studies. Lung Cancer. 2019 Sep;135:188–95.
32. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, de Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med. 2018;378(22):2078–92.
33. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. N Engl J Med. 2018;378(24):2288–301.
34. West H, McCleod M, Hussein M, Morabito A, Rittmeyer A, Conter HJ, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol. 2019;20(7):924–37.
35. Nishio M, Barlesi F, West H, et al. Atezolizumab plus chemotherapy for first-line treatment of nonsquamous NSCLC: results from the randomized phase 3 IMpower132 trial. J Thorac Oncol. 2021;16(4):653–64.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shimokawa M, Kanazu M, Saito R, Mori M, Tamura A, Okano Y, et al. Clinical benefit of platinum doublet combination therapy in older adults with advanced non-small cell lung cancer: A prospective multicenter study by the National Hospital Organization in Japan. Thorac Cancer. 2023;14(17):1597–605. https://doi.org/10.1111/1759-7714.14904

Intensive Care Med (2023) 49:706–707
https://doi.org/10.1007/s00134-023-07073-0

LETTER

Publication hyper-inflation in the field of intensive care

Yusuke Tsutsumi[1,2*], Asuka Tsuchiya[3,4] and Takuya Kawahara[5]

Dear Editor,

It has been about 30 years since the introduction of evidence-based medicine (EBM) [1]. Now, physicians carry out their daily clinical practice based on the newest and best available evidence. However, over the past two decades, the number of published evidence has exploded. In 2010, Bastian et al. reported that 75 randomized controlled trials (RCTs) and 11 systematic reviews (SRs) were published per day [2]. Moreover, the recent coronavirus disease 2019 (COVID-19) pandemic induced a further massive increase in the number of publications [3]. As a result, this might lead to the so called "publication hyper-inflation". Yet, it is still unclear how many publications intensivists are exposed to.

We therefore conducted a meta-epidemiological study to examine the number of publications in the field of intensive care. We systematically searched Pubmed for RCTs and SRs published between 1990 and 2021. To search potentially relevant studies for intensivists, we used three Medical Subject Headings (MeSH) terms: "Critical Care", "Intensive Care Units", "Critical Care Nursing". To retrieve RCTs and SRs, we used one of the search strategies as the study design filter which previous study used [2] (see supplementary materials for details). We then combined MeSH terms and the study design filter to examine the number of RCTs and SRs per year. We used Mann–Kendall test for trend analysis and binomial test for calculating the 95% confidence interval (CI) of the SRs to RCTs ratio. We performed two types of sensitivity analyses. First, we used Cochrane Central Register of Trials (CENTRAL) instead of Pubmed to retrieve the number of RCTs similar to the previous study [2]. Second, we excluded COVID-19-related studies to consider the influence of COVID-19-periods. We defined COVID-19-related studies based on MeSH terms (Supplement for details).

The results showed the number of RCTs per year was 66 in 1990, reached 331 in 2013, and then passed to 309 (25.8 /month) in 2021. The number of SRs, which was 2 in 1990, almost equated the number of RCTs in 2016 at 327 and reached 519 (43.3/month) in 2021 (both p for trend < 0.001) (Fig. 1, supplementary Table S1). Therefore, the ratio SRs to RCTs reached 1.68 (95%CI, 1.46 to 1.94) in 2021 (supplementary Fig. S1, Table S2). The results of sensitivity analyses also showed the explosion of publications (supplementary Figs. S2-S3, Table S1, Table S3).

The results of this study confirmed that there is an explosion in the number of publications in the field of intensive care. Ideally, intensivists need to read every article possible and catch up with newest knowledge. However, our study showed that the current number of publications far exceeds the number that can be read. The increase in the number of SRs is particularly noteworthy, with the number of SRs published relative to RCTs being greater than in any other disease area reported in previous studies [4]. Increased number of SRs may have a negative impact rather than contributing to the progress of EBM [5]. In the era of "publication hyper-inflation", we caution that intensivists should examine the quality of papers and carefully select the ones that should be read more than ever before.

*Correspondence: tsutsumi.yusuke.84x@st.kyoto-u.ac.jp
[1] Department of Emergency Medicine, National Hospital Organization Mito Medical Center, 280 Sakuranosato Ibaraki-Machi Higashi-Ibaraki-Gun, Ibaraki 311-3117, Japan
Full author information is available at the end of the article

Springer

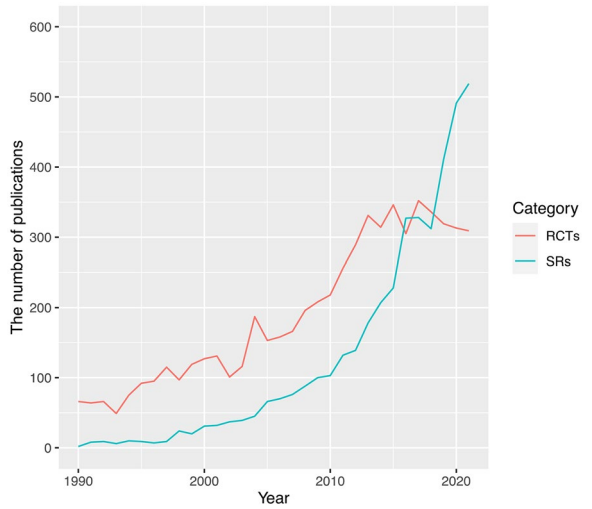


Fig.1 Annual trend in the number of published SRs and RCTs between 1990 and 2021. The number of RCTs significantly increased from 66 to 309 (25.8/month) over the periods (p for trend < 0.001). The number of SRs also significantly increased from 2 in 1990 to 519 (43.3/month) in 2021 (p for trend < 0.001)

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1007/s00134-023-07073-0.

Author details
[1] Department of Emergency Medicine, National Hospital Organization Mito Medical Center, 280 Sakuranosato Ibaraki-Machi Higashi-Ibaraki-Gun, Ibaraki 311-3117, Japan. [2] Human Health Science, Kyoto University Graduate School of Medicine, Kyoto, Japan. [3] Department of Emergency and Critical Care Medicine, Tokai University School of Medicine, Isehara, Japan. [4] Department of Clinical Epidemiology and Health Economics, School of Public Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [5] Clinical Research Promotion Center, The University of Tokyo Hospital, Tokyo, Japan.

Data availability
All data relevant to the study are uploaded as supplementary information.

Conflicts of interest
The authors declare that they have no conflict of interest.

Publisher's Note

Accepted: 12 April 2023
Published: 4 May 2023

References

1. Guyatt G, Cairns J, Churchill D et al (1992) Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA 268:2420–2425. https://doi.org/10.1001/jama.1992.03490170092032
2. Bastian H, Glasziou P, Chalmers I (2010) Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLOS Med 7:e1000326. https://doi.org/10.1371/journal.pmed.1000326
3. Ioannidis JPA, Salholz-Hillel M, Boyack KW, Baas J (2021) The rapid, massive growth of COVID-19 authors in the scientific literature. R Soc Open Sci 8:210389. https://doi.org/10.1098/rsos.210389
4. Niforatos JD, Weaver M, Johansen ME (2019) Assessment of publication trends of systematic reviews and randomized clinical trials, 1995 to 2017. JAMA Intern Med 179:1593–1594. https://doi.org/10.1001/jamainternmed.2019.3013
5. Ioannidis JPA (2016) The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 94:485–514. https://doi.org/10.1111/1468-0009.12210

American Journal of Emergency Medicine 70 (2023) 101–108

Contents lists available at ScienceDirect

American Journal of Emergency Medicine

journal homepage: www.elsevier.com/locate/ajem

Prehospital shock index predicts 24-h mortality in trauma patients with a normal shock index upon emergency department arrival

Yoshie Yamada, MD [a], Sayaka Shimizu, PhD [b], Shungo Yamamoto, Dr PH [a,d], Yoshinori Matsuoka, Dr PH [a,e], Yusuke Tsutsumi, PhD [f], Asuka Tsuchiya, PhD [g], Tsukasa Kamitani, Dr PH [a], Hajime Yamazaki, PhD [a], Yusuke Ogawa, PhD [a], Shunichi Fukuhara, PhD [c,h,i], Yosuke Yamamoto, PhD [a,*]

[a] Department of Healthcare Epidemiology, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[b] Institute for Health Outcomes & Process Evaluation Research (iHope International), Kyoto, Japan
[c] Section of Clinical Epidemiology, Department of Community Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[d] Department of Transformative Infection Control Development Studies, Osaka University Graduate School of Medicine, Suita, Japan
[e] Department of Emergency Medicine, Kobe City Medical Center General Hospital, Hyogo, Japan
[f] Department of Emergency Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan
[g] Department of Emergency and Critical Care Medicine, Tokai University School of Medicine, Kanagawa, Japan
[h] Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, MD, USA
[i] Shirakawa Satellite for Teaching And Research (STAR) for General Medicine, Fukushima Medical University, Fukushima, Japan

ARTICLE INFO

Article history:
Received 14 November 2022
Received in revised form 19 April 2023
Accepted 4 May 2023

Keywords:
Trauma
Shock index
Prehospital
Vital sign
24-h mortality

ABSTRACT

Background: The shock index (heart rate divided by systolic blood pressure) of trauma patients upon emergency department arrival predicts blood loss and death. However, some patients with normal shock indices (0.4 < shock index <0.9) upon emergency department arrival also have poor prognoses. This study aimed to determine whether abnormal prehospital shock indices in trauma patients with normal shock indices upon emergency department arrival were predictors of a high risk of mortality.

Methods: We conducted a retrospective cohort study of emergency department-admitted trauma patients from 2004 to 2017. The study included 89,495 consecutive trauma patients aged ≥16 years, with Abbreviated Injury Scale score of ≥3, who were transported to the emergency department directly from the field and had a normal shock index upon emergency department arrival. According to the prehospital shock index scores, the patients were categorized into low shock index (≤ 0.4), normal shock index, and high shock index (≥0.9) groups. Odds ratios and 95% confidence intervals were calculated using logistic regression analysis.

Results: The 89,495 patients had a median age of 64 (interquartile range: 43–79) years, and 55,484 (62.0%) of the patients were male. There were 1350 (1.5%) 24-h deaths in total; 176/4263 (4.1%), 1017/78,901 (1.3%), and 157/6331 (2.5%) patients were in the low, normal, and high prehospital shock index groups, respectively. The adjusted odds ratios for 24-h mortality compared with the normal shock index group were 1.63 (95% confidence interval: 1.34–1.99) in the low shock index group and 1.62 (95% confidence interval: 1.31–1.99) in the high shock index group.

Conclusion: Trauma patients with abnormal prehospital shock indices but normal shock indices upon emergency department arrival are at a higher risk of 24-h mortality. Identifying these indices could improve triage and targeted care for patients.

1. Background

Approximately 4.4 million people die from traumatic injuries each year [1] accounting for approximately 8% of all annual deaths worldwide. The main causes of trauma deaths include road accidents, suicides, homicides, and falls. There are nearly 70,000 yearly trauma-related deaths in Japan, accounting for about 5% of all deaths [2], and their main causes are unintentional accidents and suicide. According

* Corresponding author at: Department of Healthcare Epidemiology, School of Public Health, Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.
E-mail address: yamamoto.yosuke.5n@kyoto-u.ac.jp (Y. Yamamoto).

https://doi.org/10.1016/j.ajem.2023.05.008
0735-6757/

to the Tokyo Fire and Disaster Management Agency, trauma accounts for 27% of emergency transportation cases [3]. Therefore, identifying trauma patients with a higher risk of mortality or patients who need immediate diagnosis and treatment in the emergency department (ED) is crucial.

Vital signs are used to predict the severity of injury and prognosis of trauma patients [4,5]. The shock index (SI; heart rate divided by systolic blood pressure) [6] is a predictor of visible and hidden blood loss, a need for blood transfusion, injury severity, and mortality [7]. The SI is useful in patients without obvious vital sign abnormalities [8,9], has a greater predictive ability than any vital sign [5,10,11], and is easier to calculate than the other indices, such as the age shock index [10,12], the reverse shock index [13], or the Trauma and Injury Severity Score [14]. It is known that a high SI (e.g., ≥ 0.9) at prehospital or upon ED arrival is associated with increased mortality risk [15-17], whereas a low SI (e.g., ≤ 0.4; high blood pressure and low heart rate) has been suggested as a predictor of a serious head injury, leading to an increased mortality risk [18,19]. However, a low SI cut-off value has rarely been considered in previous studies.

The middle SI range ($0.4 < SI < 0.9$) could be considered "normal" and may indicate a good prognosis. However, some patients may have a normal SI range upon ED arrival but have a poor prognosis. Vital sign changes over time due to physiological compensatory mechanisms could help detect patients at high risk of mortality [20]. The SI may also vary depending on when the vital signs are measured [21]. Sometimes, an abnormal SI measured immediately after an injury could become normal upon ED arrival. Therefore, prehospital SI might be useful to further stratify mortality risk among patients with normal SI upon ED arrival. However, the association between prehospital SI and prognosis in such patients has not been examined. Therefore, this study aimed to determine whether an abnormal prehospital SI (SI ≥ 0.9 or ≤0.4) was associated with a higher risk of 24-h mortality than normal prehospital SI ($0.4 < SI < 0.9$) among trauma patients with normal SI upon ED arrival.

2. Methods

2.1. Study design and participants

We performed a retrospective cohort study of ED-admitted trauma patients using anonymized data from the Japan Trauma Data Bank (JTDB). The JTDB was approved by the ethics committee of the National Defence Medical College. The ethics committee of Kyoto University approved our research (approval number: R2601). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement was used to ensure the proper reporting of methods, results, and discussion.

The eligibility criterion was as follows: trauma patients aged ≥16 years, transported to the ED directly from the field and had normal SI upon arrival. The exclusion criteria were as follows: burn injury [4,22], prehospital cardiopulmonary arrest, prehospital fluid infusion, hypotension (systolic blood pressure < 90 mmHg) [11,23] upon ED arrival, bradycardia (heart rate ≤ 40 bpm) [24] upon ED arrival, as well as missing data for the systolic blood pressure, heart rate, or outcome. We excluded patients with burns according to the methods of previous studies because their treatment differs from that for other causes of trauma [4,22]. Patients who presented with hypotension and bradycardia upon ED arrival were excluded because they usually needed prompt examination or intervention.

2.2. Setting and data sources

We used the JTDB registry data from 2004 to 2017 [25]. The JTDB is a nationwide prospective registry of trauma cases in Japan, established by the Japanese Association for the Surgery of Trauma and the Japanese Association for Acute Medicine. A total of 264 emergency hospitals across Japan participate in the registry, comprising approximately 70% of the government-certified tertiary emergency and critical care centers [26]. This registry enrolled approximately 300,000 trauma patients who presented to an ED with an Abbreviated Injury Scale (AIS) score of ≥3 for any part of their body. Paramedics and medical staff measured the prehospital vital signs and those upon ED arrival, treatments, diagnoses, injury severity, and in-hospital mortality, and compiled to form the registry data.

2.3. Measurement

2.3.1. Exposure

The exposure was prehospital SI measured by the emergency medical services in the field. The SI was categorized into low (≤0.4), normal ($0.4 < SI < 0.9$), and high (≥0.9) SI groups. These categories were selected because the association between SI upon ED arrival and mortality has been reported to follow a U-shaped curve, with SIs of 0.4 and 0.9 being associated with approximately equal mortality rates with the lowest mortality rate found in between these values [19].

2.3.2. Outcome

The primary outcome was mortality within 24 h of ED arrival. The secondary outcomes were invasive hemostatic interventions (thoracoabdominal surgery, endoscopic surgery, surgical hemostasis, angiostomy, and transcatheter arterial embolisation), blood transfusion within 24 h, head surgery, and in-hospital mortality.

2.3.3. Other factors to be adjusted

In the multivariable analyses, adjustment was performed using the factors below: the patients' sex, age, Glasgow Coma Scale (GCS) upon ED arrival, respiratory rate upon ED arrival, year of ED arrival, transportation time (time of departure from the field to ED arrival), type of injury (blunt, penetrating, unknown, and other injuries), cause of injury (unintentional accident, occupational accident, suicide attempt, assault by others, unknown, and other causes), and comorbidities (respiratory, cardiovascular, digestive, metabolic, central nervous system, mental, or immunodeficiency diseases and cancer). All the variables were recorded by paramedics and medical staff at each participating hospital.

2.4. Statistical analysis

2.4.1. Descriptive analysis

The eligible patient characteristics were summarized for the entire cohort and each SI group. The continuous variables were presented as the medians and interquartile ranges (IQRs), while the categorical variables were presented as numbers and percentages.

2.4.2. Primary analysis

We calculated the odds ratios (ORs) and 95% confidence intervals (CIs) for the 24-h mortality of the prehospital low and high SI groups and compared them with those of the normal SI group using a logistic regression analysis after adjusting for the abovementioned factors. Model 1 included the covariates usually available at the time of patients' ED arrival: sex, age, GCS, respiratory rate, year of ED arrival, transportation time, and type of injury. The other factors, such as the cause of injury and comorbidities, were not necessarily obtainable during ED arrival. These respective factors were added in Model 2 and Model 3.

2.4.3. Secondary analysis

We used logistic regression analyses to calculate the ORs for the secondary outcomes: invasive hemostatic interventions, blood transfusion within 24 h, head surgery, and in-hospital mortality. The same independent variables used in the primary analysis were used to determine whether the prehospital SI could predict the secondary outcomes.

2.4.4. Subgroup analysis

We also performed the same analysis performed in the primary analysis for the following subgroups: patients with isolated serious head injuries (defined as a head AIS score ≥ 3 and AIS score < 3 in other body parts) and patients without serious head injuries (AIS score ≥ 3 in other body parts and a head AIS score < 3).

2.4.5. Sensitivity analysis

We performed three sensitivity analyses. Firstly, we changed the normal SI definition to 0.4 < SI < 1.0 and 0.5 < SI < 0.7. No definite criterion exists for a normal SI range; however, an SI "below 1.0" or "0.5 to 0.7" has also been considered to be a normal range in clinical settings and previous studies [8,23,27]. Secondly, we added the adjustment for SI upon ED arrival as a covariate into Model 1, 2 and 3 of the primary multivariable logistic regression analysis. This analysis was undertaken in consideration of the possibility that the level of SI upon ED arrival may be a stronger predictor than prehospital SI, despite the study population falling within the normal range of SI upon ED arrival. Thirdly, we performed a multiple imputation (MI) for the missing adjusted variables in the primary multivariable logistic regression analysis, based on the assumption that the data were missing at random. The MI procedure imputed the missing values using chained equations with factors of all the variables used in Model 1. We created 20 imputed datasets and performed a logistic regression analysis for each. The results were integrated using Rubin's rule.

All the analyses were conducted using STATA/MP, Version 15.1 (StataCorp, TX, USA). Except for the sensitivity analysis with MI, only the cases without any missing covariate values were included in the analyses (complete case analysis).

3. Results

3.1. Patient characteristics

A total of 113,494 adult (age ≥ 16) trauma patients were transported directly to the ED from the field and had an SI within the normal range upon ED arrival (Fig. 1). The eligibility criteria were met by 89,495 (78.9%) patients, of whom 55,484 (62.0%) were male. The median age was 64 (IQR: 43–79) years (Table 1). Blunt and penetrating injuries accounted for 96.7% and 1.9% of all trauma cases, respectively. In

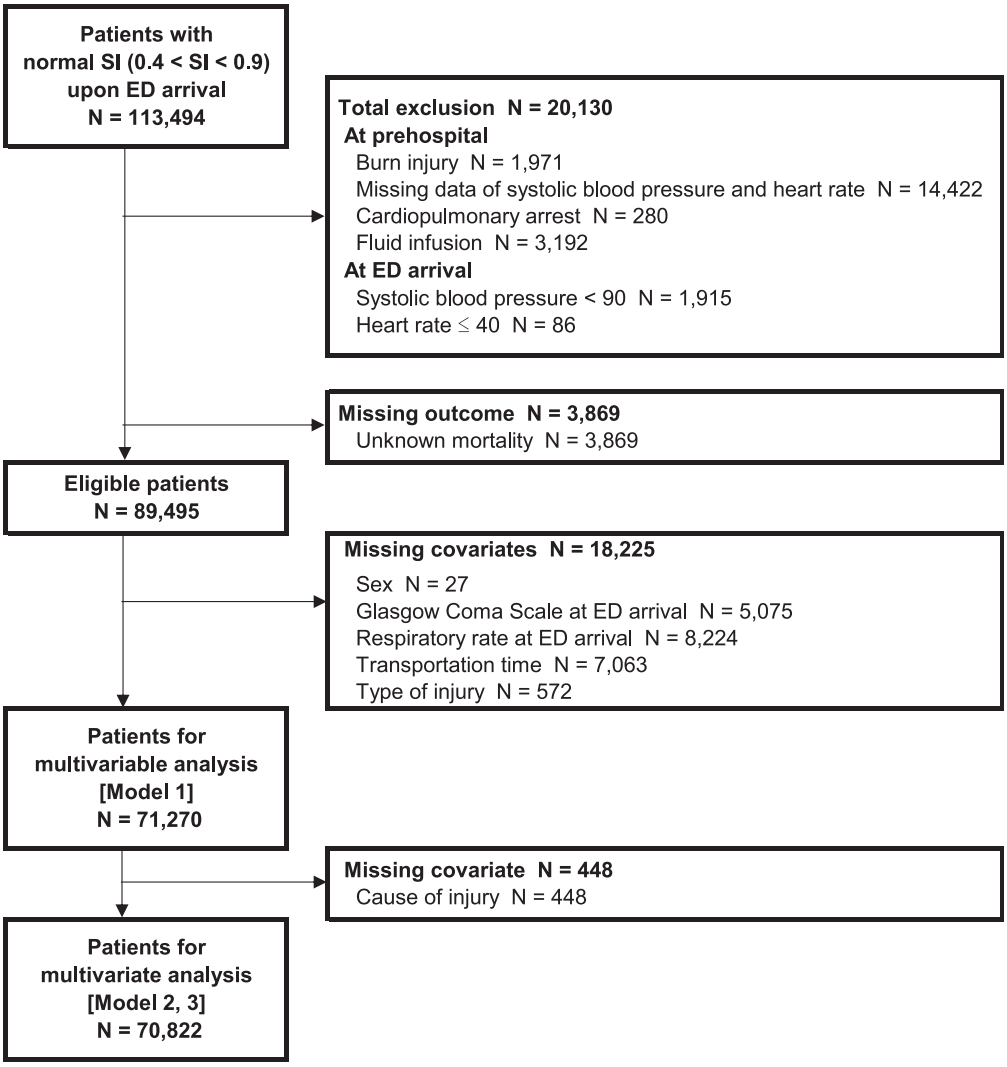


Fig. 1. Patient selection flowchart.
Abbreviations: ED, emergency department.
[Model 1] was adjusted for age, sex, Glasgow Coma Scale, respiratory rate, year of emergency department arrival, transportation time, and type of injury.
[Model 2] was adjusted for age, sex, Glasgow Coma Scale, respiratory rate, year of emergency department arrival, transportation time, type of injury, and cause of injury.
[Model 3] was adjusted for age, sex, Glasgow Coma Scale, respiratory rate, year of emergency department arrival, transportation time, type of injury, cause of injury, and comorbidities.

Table 1
Patient characteristics.

Parameter	Total		Prehospital SI					
			SI ≤ 0.4		0.4 < SI < 0.9		SI ≥ 0.9	
	N = 89,495		N = 4263		N = 78,901		N = 6331	
Male	55,484	(62.0)	2456	(57.6)	48,821	(61.9)	4207	(66.5)
Missing	27	(0.0)	1	(0.0)	23	(0.0)	3	(0.0)
Age [year][a]	64	[43–79]	74	[62–83]	65	[44–79]	50	[31–69]
At prehospital								
Shock index [bpm/mmHg][a]	0.60	[0.51–0.72]	0.37	[0.34–0.39]	0.60	[0.52–0.70]	1.00	[0.94–1.10]
Heart rate [bpm][a]	84	[73–95]	66	[60–72]	84	[74–93]	102	[90–115]
Systolic blood pressure [mm Hg][a]	139	[120–160]	180	[161–200]	140	[121–159]	100	[87–111]
At ED arrival								
Shock index [bpm/mmHg][a]	0.58	[0.50–0.69]	0.48	[0.43–0.55]	0.58	[0.50–0.68]	0.74	[0.64–0.82]
Heart rate [bpm][a]	82	[73–93]	75	[67–86]	82	[73–92]	92	[81–104]
Systolic blood pressure [mm Hg][a]	140	[124–159]	155	[136–173]	140	[124–159]	129	[115–144]
Glasgow Coma Scale[a]	15	[14–15]	15	[13–15]	15	[14–15]	14	[13–15]
Missing	5075	(5.7)	245	(5.7)	4565	(5.8)	265	(4.2)
Respiratory rate [/min][a]	20	[17–24]	20	[16–24]	20	[17–24]	21	[18–26]
Missing	8224	(9.2)	425	(10.0)	7353	(9.3)	446	(7.0)
Year								
2004–2009	14,842	(16.6)	687	(16.1)	12,870	(16.3)	1285	(20.3)
2010–2014	48,751	(54.5)	2324	(54.5)	42,979	(54.5)	3448	(54.5)
2015–2017	25,902	(28.9)	1252	(29.4)	23,052	(29.2)	1598	(25.2)
Transportation time [min][a]	13.1	[6.6–19.7]	13.1	[8.7–19.7]	10.9	[6.6–19.7]	13.1	[8.7–19.7]
Missing	7063	(7.9)	347	(8.1)	6244	(7.9)	472	(7.5)
Type of injury								
Blunt injury	86,582	(96.7)	4152	(97.4)	76,523	(97.0)	5907	(93.3)
Penetrating injury	1672	(1.9)	44	(1.0)	1282	(1.6)	346	(5.5)
Unknown	533	(0.6)	35	(0.8)	468	(0.6)	30	(0.5)
Others	136	(0.2)	3	(0.1)	119	(0.2)	14	(0.2)
Missing	572	(0.6)	29	(0.7)	509	(0.6)	34	(0.5)
Cause of injury								
Unintentional accident	78,252	(87.4)	3828	(89.8)	69,477	(88.1)	4947	(78.1)
Occupational accident	5015	(5.6)	227	(5.3)	4436	(5.6)	352	(5.6)
Suicide attempt	2716	(3.0)	46	(1.1)	1965	(2.5)	705	(11.1)
Assault by others	993	(1.1)	23	(0.5)	824	(1.0)	146	(2.3)
Unknown	1161	(1.3)	65	(1.5)	997	(1.3)	99	(1.6)
Others	644	(0.7)	39	(0.9)	575	(0.7)	30	(0.5)
Missing	714	(0.8)	35	(0.8)	627	(0.8)	52	(0.8)
Comorbidities								
Respiratory	4147	(4.6)	157	(3.7)	3655	(4.6)	335	(5.3)
Cardiovascular	27,116	(30.3)	1837	(43.1)	24,112	(30.6)	1167	(18.4)
Digestive	7055	(7.9)	378	(8.9)	6210	(7.9)	467	(7.4)
Metabolic	12,094	(13.5)	704	(16.5)	10,801	(13.7)	589	(9.3)
Central nervous system / mental	15,340	(17.1)	782	(18.3)	13,349	(16.9)	1209	(19.1)
Immunodeficiency / cancer	9151	(10.2)	508	(11.9)	8159	(10.3)	484	(7.6)
Body part with serious injury								
Isolated serious head injury	24,925	(27.9)	1560	(36.6)	22,119	(28.0)	1246	(19.7)
Without serious head injury	55,032	(61.5)	2235	(52.4)	48,714	(61.7)	4083	(64.5)
Head and other body part injury	9538	(10.7)	468	(11.0)	8068	(10.2)	1002	(15.8)

Abbreviations: SI, shock index; ED, emergency department.
n (%), unless otherwise specified.
"Isolated serious head injury" was defined as a head Abbreviated Injury Scale (AIS) score ≥ 3 and AIS score < 3 for other body parts.
"Without serious head injury" was defined as a head AIS score < 3 and AIS score ≥ 3 for other body parts.
"Head and other body part injury" were defined as AIS score ≥ 3 for both the head and other body parts.
[a] Median [interquartile range].

total, 87.4% of the patients had unintentional accidents, 5.6% had occupational accidents, and 3.0% attempted suicide. The proportions of the low, normal, and high SI groups were 4.8%, 88.2%, and 7.1%, respectively.

3.2. Primary analysis: association between prehospital SI and 24-h mortality

Overall, 1350 (1.5%) 24-h deaths occurred, including 176/4263 (4.1%), 1017/78,901 (1.3%), and 157/6331 (2.5%) in the low, normal, and high SI groups, respectively (Fig. 2A). Compared with the normal SI group, the unadjusted ORs for 24-h mortality in the low and high SI groups were 3.30 (95% CI: 2.80–3.88) and 1.95 (95% CI: 1.64–2.31), respectively. The corresponding adjusted ORs for 24-h mortality in the low and high SI groups were 1.63 (95% CI: 1.34–1.99) and 1.62 (95% CI: 1.31–1.99) in Model 1 (Fig. 2B), 1.65 (95% CI: 1.35–2.01) and 1.50 (95% CI: 1.21–1.85) in Model 2, and 1.63 (95% CI: 1.34–2.00) and 1.49 (95% CI: 1.21–1.85) in Model 3. The ORs and 95% CIs for the adjusted factors are shown in the Supplementary (Table S1).

3.3. Secondary analysis: association between prehospital SI and invasive hemostatic interventions, blood transfusion within 24 h, head surgery, and in-hospital mortality

Similar to the primary analysis, the low and high SI groups showed higher ORs for in-hospital mortality than the normal SI group (Fig. S1). Compared with the normal SI group, the ORs for invasive hemostatic interventions were higher in the high SI group but lower in the low SI group (Fig. 3A). The OR for blood transfusion within 24 h was higher in the high SI group than in the normal SI group (Fig. 3B).

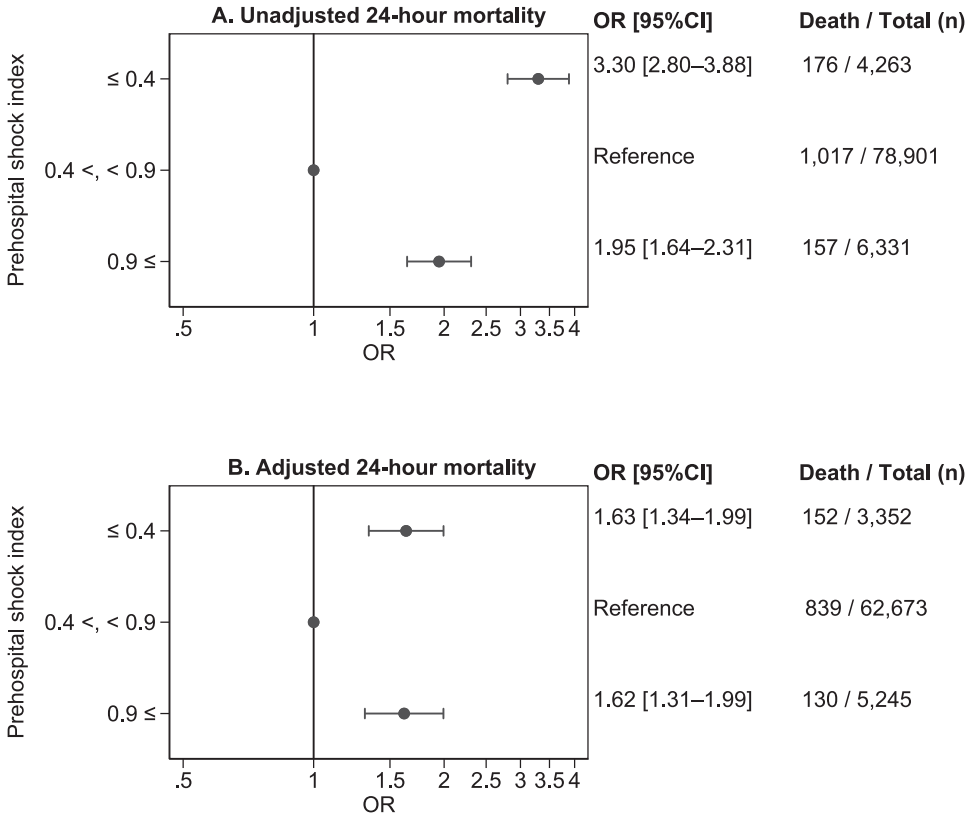


Fig. 2. Primary analysis: odds ratios of the prehospital shock index for 24-h mortality based on the logistic regression analysis.
Abbreviations: OR, odds ratio; CI, confidence interval.
Graph A is from the unadjusted model, and Graph B is from the model adjusted for age, sex, Glasgow Coma Scale, respiratory rate, year of emergency department arrival, transportation time, and type of injury [Model 1].

The low SI group showed a higher OR for head surgery, and the high SI group showed a lower OR than the normal SI group (Fig. 3C).

3.4. Subgroup analysis: association between prehospital SI and 24-h mortality in patients with and without head injury

The low SI group in the subpopulation of patients with isolated serious head injuries was associated with a higher risk of 24-h mortality than the normal SI group; however, there were no substantial differences between the high and normal SI groups in this subgroup analysis (Fig. 4A). Among the patients without serious head injuries, the high but not the low SI group showed a higher risk of 24-h mortality than the normal SI group (Fig. 4B).

3.5. Sensitivity analysis: prehospital SI and 24-h mortality with altered cut-off values, additional adjustment for SI upon ED arrival as a covariate, and MI

Sensitivity analyses with the normal range of the SI set to 0.4–1.0 or 0.5–0.7, additional adjustment for the SI upon ED arrival, and MI for the missing values showed results similar to those of the primary analysis (Fig. S2-S4). In the analysis of the entire cohort with the MI, 20.4% of the patients had missing values, with the most common factors being respiratory rate (9.2%), transportation time (7.9%), and GCS (5.7%; Table 1).

4. Discussion

In this retrospective cohort study of 89,495 patients with a normal SI upon ED arrival recorded in the JTDB, we found an association between prehospital SI abnormalities (SI ≤ 0.4 or ≥0.9) and 24-h mortality (Fig. 2). We confirmed the robustness of the results using three sensitivity analyses; alternative definitions of the normal SI, additional adjustment for the SI upon ED arrival, and analysis with MI showed results similar to those of the primary analysis (Fig. S2–S4). Invasive hemostatic interventions and blood transfusion within 24 h were performed more frequently in the high SI group than in the normal SI group, while head surgery was more frequent in the low SI group (Fig. 3). The subgroup analysis of patients with isolated serious head injuries showed that the low SI group was associated with a higher risk of 24-h mortality than the normal SI group, while the high SI group without serious head injuries had a higher risk of 24-h mortality (Fig. 4).

The two leading causes of trauma-related deaths were bleeding and neurological damage. When injuries in the body trunk or limbs cause massive bleeding, blood pressure drops due to hypovolaemia. Compensatory mechanisms work to maintain cardiac output and blood flow to vital organs [20]. The sympathetic nervous system is activated to increase the heart rate and, consequently, the cardiac output. This reaction increases the SI (high heart rate and low blood pressure), indicating an increased risk of death from haemorrhagic shock. Additionally, the sympathetic nervous system constricts the peripheral blood vessels to raise the blood pressure, resulting in normal SI values in some cases. In contrast, a serious head injury with intracranial haemorrhage could increase intracranial pressure, resulting in bradycardia and high blood pressure (low SI), known as the Cushing reflex [18], potentially explaining the secondary analysis results, in which the low SI group was associated with a higher rate of head surgeries than the normal SI group (Fig. 3). These mechanisms may temporarily alter the vital signs and SI [21], sometimes normalizing the SI value [20], leading clinicians to misestimate the risk of death.

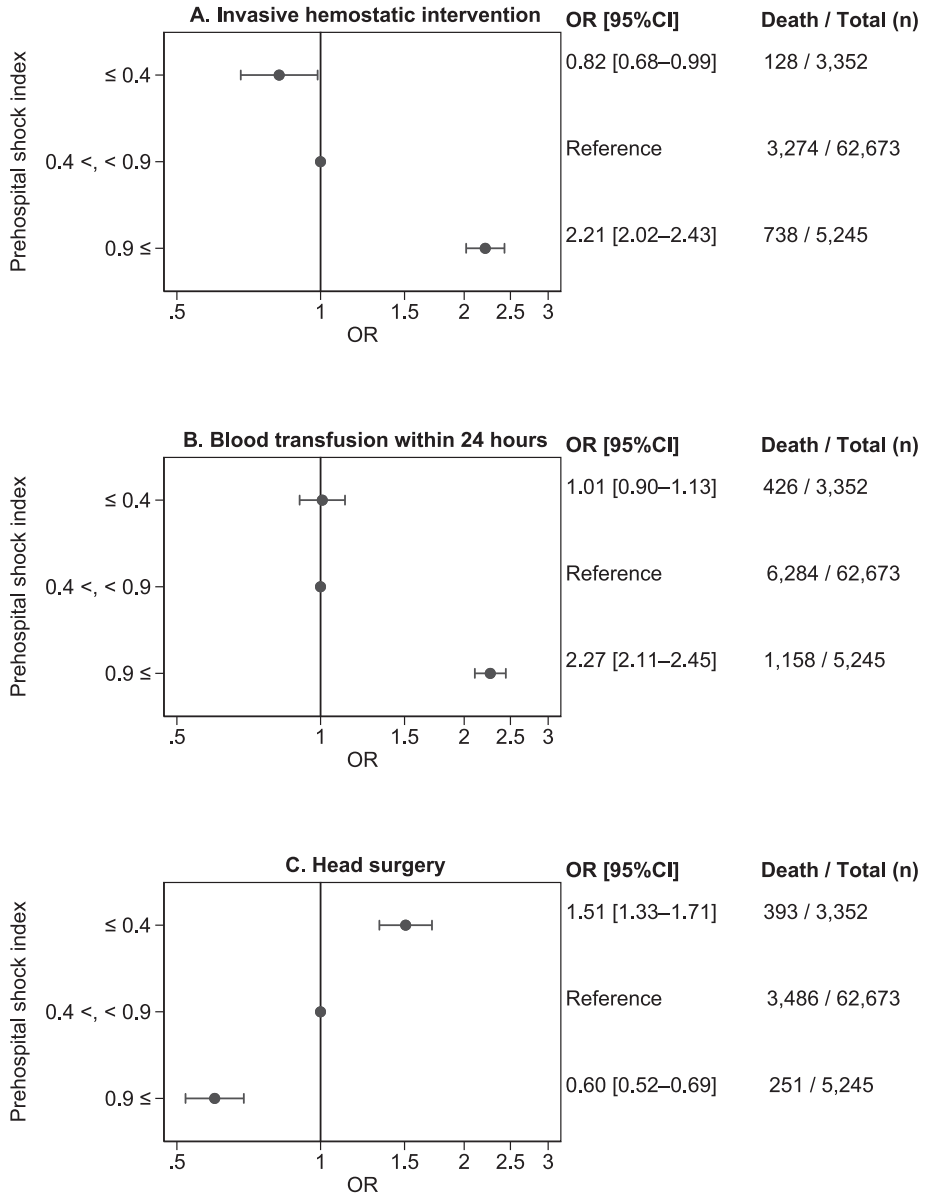


Fig. 3. Secondary Analysis: adjusted odds ratios of the prehospital shock index for each treatment based on the logistic regression analysis.
Abbreviations: OR, odds ratio; CI, confidence interval.
Adjusted for age, sex, Glasgow Coma Scale, respiratory rate, year of emergency department arrival, transportation time, and type of injury [Model 1].
"Invasive hemostatic interventions" comprised thoracoabdominal surgery, endoscopic surgery, surgical hemostasis, angiostomy, and transcatheter arterial embolisation.
Graph A, B and C are from independent logistic regression models, see "Primary analysis" in Method section.

Previous studies have attempted to improve the prognostic ability of the SI by considering the difference between the prehospital SI and ED arrival SI [15,28,29]. Patients with different prognoses may have been classified into the same group in these studies. Briefly, the prognosis in patients with a similar increase in SI but different prehospital SI values could be different. For example, the prognosis of patients with low prehospital SI and normal SI upon ED arrival might differ from that of patients with normal prehospital SI and high SI upon ED arrival. In clinical settings, physicians rush to treat patients with high SI upon ED arrival, regardless of their prehospital SI. Our study focused on patients with a normal SI upon ED arrival as they are generally considered to have a good prognosis. The results allowed us to identify patients at high risk of death, which may require therapeutic interventions based on their prehospital SI.

The SI was used to predict blood loss, indicating the need for blood transfusions following trauma [7,17], and the risk of death [7,11,23]. In our study, patients with a high prehospital SI had a higher mortality risk, as reported previously [17]; furthermore, they had a higher risk of undergoing an invasive hemostatic intervention or blood transfusion within 24 h than patients who had normal prehospital SI, possibly due to severe blood loss following organ injuries (Fig. 3). Without a serious head injury, the 24-h mortality in the low SI group was rare, and there were no substantial differences between the low and normal SI groups (Fig. 4). Therefore, after excluding trauma patients with a serious head injury, a traditional single high SI cut-off value (e.g., 0.9) might be sufficient to predict mortality and the need for therapeutic interventions. However, patients in the low SI group required head surgery more frequently than those in the normal SI group. When trauma was confined to a serious head injury, mortality in the low but not high SI group was more frequent than in the normal SI group. This finding suggested that a low SI cut-off value should be set for trauma patients with isolated serious head injuries.

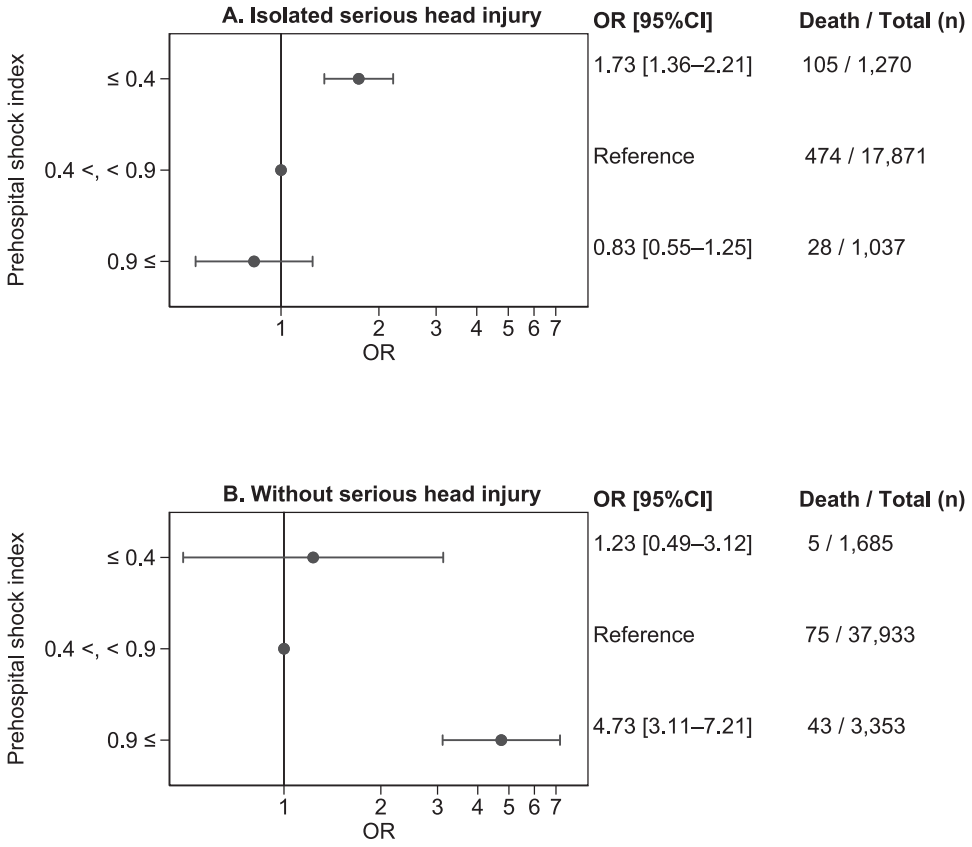


Fig. 4. Subgroup analysis: adjusted odds ratios of the prehospital shock index for 24-h death with/without serious head injury based on the logistic regression analysis.
Abbreviations: OR, odds ratio; CI, confidence interval.
Adjusted for age, sex, Glasgow Coma Scale, respiratory rate, year of emergency department arrival, transportation time, and type of injury [Model 1].
Graph A and B are from different populations; Graph A. is from the population with isolated serious head injury defined as a head Abbreviated Injury Scale (AIS) score ≥ 3 and AIS score < 3 for the other body parts, and Graph B is from the population without serious head injury defined as a head AIS score < 3 and AIS score ≥ 3 for the other body parts.

This study had several strengths. First, we used one of the largest multicenter trauma registries in the world, resulting in an adequately large sample. Second, we confirmed the robustness of the results associated the prehospital SI with 24-h mortality using sensitivity analyses. Finally, unlike previous studies, we excluded patients with normal SI who were in shock states upon ED arrival. This eligibility criterion could identify patients whose prognoses were uncertain and who needed further risk estimations in clinical settings.

This study also had several limitations. First, the extrapolation of our results to populations with less serious injuries could be challenging because this registry enrolled only trauma patients with serious injuries (AIS score ≥ 3). Second, it was unclear whether our results may be applied to patients in countries where penetrating injuries are more common [30]. It is often difficult for physicians to determine whether patients require further examinations or interventions after a blunt injury with visible and hidden blood loss. Our findings may provide useful information for clinical decision-making in patients with blunt injuries. Third, there were missing covariates and outcomes in this study. However, we found no apparent differences in the patient characteristics among the eligible patients, those with missing covariate data, and those who were excluded due to missing outcomes (Table S2). Fourth, we did not have data on the use of drugs intimately associated with SI (e.g., vasoactive drugs such as beta blockers or calcium blockers). As a substitute, adjustment was performed using comorbidities, for which patients were likely to use such kinds of drugs. Last, this was an observational study; therefore, we cannot exclude the effect of unknown factors which may have affected the observed relationship.

5. Conclusion

Among the trauma patients with normal SI upon ED arrival (0.4 < SI < 0.9), abnormal prehospital SI (SI ≥ 0.9 or SI ≤ 0.4) was associated with higher 24-h mortality than normal prehospital SI. This study contributes to a more effective triage of trauma patients with normal SI upon ED arrival.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit organizations.

Ethics approval

The ethics committee of Kyoto University approved this study (approval number: R2601). This study was conducted in accordance with Japan's ethical guidelines for medical and biological research involving human subjects.

Patient consent for publication

Not applicable.

Prior presentations

Annual Meeting of the Japanese Association for Acute Medicine, Gifu Japan, November 18, 2020.

CRediT authorship contribution statement

Yoshie Yamada: Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Conceptualization. **Sayaka Shimizu:** Writing – review & editing, Validation, Methodology, Formal analysis. **Shungo Yamamoto:** Writing – review & editing, Methodology, Conceptualization. **Yoshinori Matsuoka:** Writing – review & editing, Validation, Conceptualization. **Yusuke Tsutusmi:** Writing – review & editing, Resources, Conceptualization. **Asuka Tsuchiya:** Writing – review & editing, Resources, Conceptualization. **Tsukasa Kamitani:** Writing – review & editing, Resources, Conceptualization. **Hajime Yamazaki:** Writing – review & editing, Methodology, Conceptualization. **Yusuke Ogawa:** Writing – review & editing, Methodology, Conceptualization. **Shunichi Fukuhara:** Writing – review & editing, Supervision, Conceptualization. **Yosuke Yamamoto:** Writing – review & editing, Supervision, Conceptualization.

Data availability

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

Y. Yamamoto has received consultancy fees from Nippon Shinyaku Co., Ltd, and personal fees from Sun Pharma, Asahi Kasei Pharma, TORAY, and Ono, outside the submitted work. The other authors state that they have no conflict of interest with the present work.

Acknowledgments

We would like to thank the members of the JTDB for the data of this research and Editage (www.editage.com) for English language editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ajem.2023.05.008.

References

[1] World Health Organization. Injuries and Violence. https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence; 2021. accessed October 28 2022.
[2] Ministry of Health LaW. Fact Sheet of Monthly Demographic Statistics. https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/geppo/nengai19/index.html; 2019. [in Japanese] (accessed October 28 2022).
[3] Tokyo Fire Department. Current Activities of the Emergency Medical Service. https://www.tfd.metro.tokyo.lg.jp/hp-kyuukanka/katudojitai/R01.html; 2020. [in Japanese] (accessed October 28 2022).
[4] Shiraishi A, Otomo Y, Yoshikawa S, Morishita K, Roberts I, Matsui H. Derivation and validation of an easy-to-compute trauma score that improves prognostication of mortality or the trauma rating index in age, Glasgow coma scale, respiratory rate and systolic blood pressure (TRIAGES) score. Crit Care. 2019;23(1):365.
[5] Totten AM, Cheney TP, O'Neil ME, et al. Physiologic predictors of severe injury: systematic review. AHRQ Comparative Effectiveness Reviews; 2018 18-EHC008-EF.
[6] Allgöwer M, Burri C. Schockindex. DMW. 1967;92(43):1947–50. [in German].
[7] King RW, Plewa MC, Buderer NM, Knotts FB. Shock index as a marker for significant injury in trauma patients. Acad Emerg Med. 1996;3(11):1041–5.
[8] Vandromme MJ, Griffin RL, Kerby JD, McGwin Jr G, Rue 3rd LW, Weinberg JA. Identifying risk for massive transfusion in the relatively normotensive patient: utility of the prehospital shock index. J Trauma. 2011;70(2):384–8.
[9] Birkhahn RH, Gaeta TJ, Terry D, Bove JJ, Tloczkowski J. Shock index in diagnosing early acute hypovolemia. Am J Emerg Med. 2005;23(3):323–6.
[10] Zarzaur BL, Croce MA, Magnotti LJ, Fabian TC. Identifying life-threatening shock in the older injured patient: an analysis of the National Trauma Data Bank. J Trauma. 2010;68(5):1134–8.
[11] Haider AA, Azim A, Rhee P, et al. Substituting systolic blood pressure with shock index in the National Trauma Triage Protocol. J Trauma Acute Care Surg. 2016;81(6):1136–41.
[12] Bruijns SR, Guly HR, Bouamra O, Lecky F, Lee WA. The value of traditional vital signs, shock index, and age-based markers in predicting trauma mortality. J Trauma Acute Care Surg. 2013;74(6):1432–7.
[13] Wu SC, Rau CS, Kuo SCH, Chien PC, Hsieh HY, Hsieh CH. The reverse shock index multiplied by Glasgow coma scale score (rSIG) and prediction of mortality outcome in adult trauma patients: a cross-sectional analysis based on registered trauma data. Int J Environ Res Public Health. 2018;15(11):2346.
[14] Tohira H, Jacobs I, Mountain D, Gibson N, Yeo A. Systematic review of predictive performance of injury severity scoring tools. Scand J Trauma Resusc Emerg Med. 2012;20:63.
[15] Cannon CM, Braxton CC, Kling-Smith M, Mahnken JD, Carlton E, Moncure M. Utility of the shock index in predicting mortality in traumatically injured patients. J Trauma. 2009;67(6):1426–30.
[16] McNab A, Burns B, Bhullar I, Chesire D, Kerwin A. A prehospital shock index for trauma correlates with measures of hospital resource use and mortality. Surgery. 2012;152(3):473–6.
[17] Olaussen A, Blackburn T, Mitra B, Fitzgerald M. Review article: shock index for prediction of critical bleeding post-trauma: a systematic review. Emerg Med Australas. 2014;26(3):223–8.
[18] Yumoto T, Mitsuhashi T, Yamakawa Y, et al. Impact of Cushing's sign in the prehospital setting on predicting the need for immediate neurosurgical intervention in trauma patients: a nationwide retrospective observational study. Scand J Trauma Resusc Emerg Med. 2016;24(1):147.
[19] Odom SR, Howell MD, Gupta A, Silva G, Cook CH, Talmor D. Extremes of shock index predicts death in trauma patients. J Emerg Trauma Shock. 2016;9(3):103–6.
[20] Abou-Khalil B, Scalea TM, Trooskin SZ, Henry SM, Hitchcock R. Hemodynamic responses to shock in young trauma patients: need for invasive monitoring. Crit Care Med. 1994;22(4):633–9.
[21] Chen L, Reisner AT, Gribok A, Reifman J. Exploration of prehospital vital sign trends for the prediction of trauma outcomes. Prehosp Emerg Care. 2009;13(3):286–94.
[22] Oyetunji TA, Chang DC, Crompton JG, et al. Redefining hypotension in the elderly: normotension is not reassuring. Arch Surg. 2011;146(7):865–9.
[23] Singh A, Ali S, Agarwal A, Srivastava RN. Correlation of shock index and modified shock index with the outcome of adult trauma patients: a prospective study of 9860 patients. N Am J Med Sci. 2014;6(9):450–2.
[24] Kusumoto FM, Schoenfeld MH, Barrett C, et al. 2018 ACC/AHA/HRS guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines, and the Heart Rhythm Society. J Am Coll Cardiol. 2019;74(7):932–87.
[25] The Japanese Association for the Surgery of Trauma. Japan Trauma Data Bank. https://www.jtcr-jatec.org/traumabank/index.htm; 2017. [in Japanese] (accessed October 28 2022).
[26] Ministry of Health LaW. Evaluation of the Quality of Critical Care Medical Center. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000188907.html; 2017. [in Japanese] (accessed October 28 2022).
[27] Jehan F, Con J, McIntyre M, et al. Pre-hospital shock index correlates with transfusion, resource utilization and mortality; the role of patient first vitals. Am J Surg. 2019;218(6):1169–74.
[28] Schellenberg M, Strumwasser A, Grabo D, et al. Delta shock index in the emergency department predicts mortality and need for blood transfusion in trauma patients. Am Surg. 2017;83(10):1059–62.
[29] Joseph B, Haider A, Ibraheem K, et al. Revitalizing vital signs: the role of delta shock index. Shock. 2016;46:50–4.
[30] Søreide K. Epidemiology of major trauma. Br J Surg. 2009;96(7):697–8.

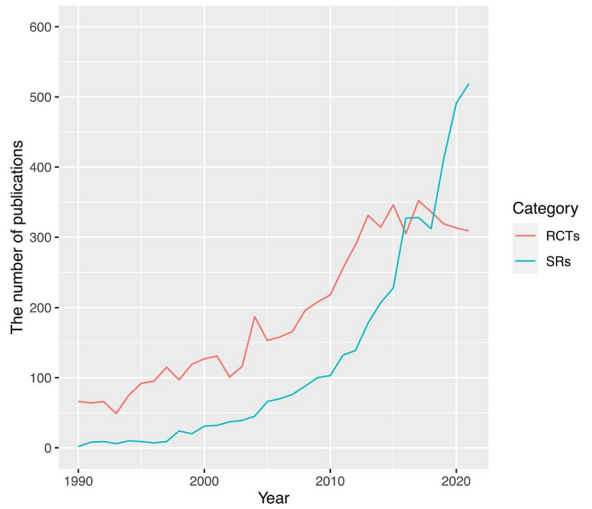


Fig.1 Annual trend in the number of published SRs and RCTs between 1990 and 2021. The number of RCTs significantly increased from 66 to 309 (25.8/month) over the periods (p for trend < 0.001). The number of SRs also significantly increased from 2 in 1990 to 519 (43.3/month) in 2021 (p for trend < 0.001)

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1007/s00134-023-07073-0.

Author details
[1] Department of Emergency Medicine, National Hospital Organization Mito Medical Center, 280 Sakuranosato Ibaraki-Machi Higashi-Ibaraki-Gun, Ibaraki 311-3117, Japan. [2] Human Health Science, Kyoto University Graduate School of Medicine, Kyoto, Japan. [3] Department of Emergency and Critical Care Medicine, Tokai University School of Medicine, Isehara, Japan. [4] Department of Clinical Epidemiology and Health Economics, School of Public Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [5] Clinical Research Promotion Center, The University of Tokyo Hospital, Tokyo, Japan.

Data availability
All data relevant to the study are uploaded as supplementary information.

Conflicts of interest
The authors declare that they have no conflict of interest.

Publisher's Note

Accepted: 12 April 2023
Published: 4 May 2023

References

1. Guyatt G, Cairns J, Churchill D et al (1992) Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA 268:2420–2425. https://doi.org/10.1001/jama.1992.03490170092032
2. Bastian H, Glasziou P, Chalmers I (2010) Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLOS Med 7:e1000326. https://doi.org/10.1371/journal.pmed.1000326
3. Ioannidis JPA, Salholz-Hillel M, Boyack KW, Baas J (2021) The rapid, massive growth of COVID-19 authors in the scientific literature. R Soc Open Sci 8:210389. https://doi.org/10.1098/rsos.210389
4. Niforatos JD, Weaver M, Johansen ME (2019) Assessment of publication trends of systematic reviews and randomized clinical trials, 1995 to 2017. JAMA Intern Med 179:1593–1594. https://doi.org/10.1001/jamainternmed.2019.3013
5. Ioannidis JPA (2016) The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 94:485–514. https://doi.org/10.1111/1468-0009.12210

ORIGINAL ARTICLE

Survival Impact of Second-Line Immune Checkpoint Inhibitors in Older Patients With Advanced Squamous-Cell NSCLC: Post Hoc Analysis of the CAPITAL Study

Yoshihito Kogure, MD, PhD,[a,*] Akiko Kada, MPH,[b] Hiroya Hashimoto, PhD,[c] Shinji Atagi, MD,[d] Yuichi Takiguchi, MD,[e] Hideo Saka, MD,[a] Noriyuki Ebi, MD,[f] Akira Inoue, MD,[g] Takayasu Kurata, MD,[h] Yuka Fujita, MD,[i] Yoichi Nishii, MD,[j] Hidetoshi Itani, MD,[k] Takeo Endo, MD,[l] Akiko M. Saito, MD,[b] Takuo Shibayama, MD,[m] Nobuyuki Yamamoto, MD,[n] Akihiko Gemma, MD[o]

[a]*Department of Respiratory Medicine, National Hospital Organization Nagoya Medical Center, Nagoya, Japan*
[b]*Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan*
[c]*Core Laboratory, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan*
[d]*Japan Community Health Care Organization Yamatokoriyama Hospital Health Management Center, Yamatokoriyama, Japan*
[e]*Department of Medical Oncology, Chiba University Hospital, Chiba, Japan*
[f]*Department of Respiratory Medicine, Fukuoka University Hospital, Fukuoka, Japan*
[g]*Department of Palliative Medicine, Tohoku University School of Medicine, Sendai, Japan*
[h]*Department of Thoracic Oncology, Kansai Medical University Hospital, Osaka, Japan*
[i]*Department of Respiratory Medicine, National Hospital Organization Asahikawa Medical Center, Asahikawa, Japan*
[j]*Respiratory Center, Matsusaka Municipal Hospital, Matsusaka, Japan*
[k]*Department of Respiratory Medicine, Japanese Red Cross Ise Hospital, Ise, Japan*
[l]*Division of Respiratory Medicine, National Hospital Organization Mito Medical Center, Mito, Japan*
[m]*Department of Respiratory Medicine, National Hospital Organization Okayama Medical Center, Okayama, Japan*
[n]*Department of Pulmonary Medicine and Medical Oncology, Wakayama Medical University, Wakayama, Japan*
[o]*Department of Pulmonary Medicine and Oncology, Graduate School of Medicine, Nippon Medical School, Tokyo, Japan*

Received 18 January 2023; revised 23 March 2023; accepted 3 April 2023
Available online - 6 April 2023

*Corresponding author.

Disclosure: Dr. Kogure reports receiving a research grant from Taiho Pharmaceutical and Merck Sharp & Dohme and an honoraria from Taiho Pharmaceutical, Merck Sharp & Dohme, Chugai Pharmaceutical, AstraZeneca, Boehringer Ingelheim, Ono Pharmaceutical, Eli Lilly, and GlaxoSmithKline. Dr. Hashimoto reports receiving honoraria from Chugai Pharmaceutical. Dr. Atagi reports receiving honoraria from Chugai Pharmaceutical, Eli Lilly, Taiho Pharmaceutical, Novartis Pharma, AstraZeneca, Boehringer Ingelheim, Ono Pharmaceutical, Boehringer Ingelheim, Pfizer, Bristol-Myers Squibb, Merck Sharp & Dohme, Kyowa-Kirin, Novartis, and Merck. Dr. Takiguchi reports receiving a research grant from Ono Pharmaceutical, Taiho Pharmaceutical, Chugai Pharmaceutical, Boehringer Ingelheim, Eli Lilly, and Daiichi Sankyo and an honoraria from Ono Pharmaceutical, Taiho Pharmaceutical, Chugai Pharmaceutical, Boehringer Ingelheim, Eli Lilly, Daiichi Sankyo, AstraZeneca, Pfizer, Bristol-Myers Squibb, and Novartis. Dr. Saka reports receiving honoraria from AstraZeneca, Chugai Pharmaceutical, and GlaxoSmithKline. Dr. Inoue reports receiving honoraria from AstraZeneca, Chugai Pharmaceutical, Eli Lilly, and Daiichi Sankyo. Dr. Kurata reports receiving a research grant from Merck Sharp & Dohme, Bristol-Myers Squibb, AstraZeneca, Takeda, and Amgen and an honoraria from Ono Pharmaceutical, Eli Lilly, Bristol-Myers Squibb, AstraZeneca, Merck Sharp & Dohme, Pfizer, Chugai Pharmaceutical, Nippon Kayaku, and Boehringer Ingelheim. Dr. Itani reports receiving a research grant from Chugai Pharmaceutical and an honoraria from Eli Lilly. Dr. Shibayama reports receiving honoraria from Chugai Pharmaceutical, Boehringer Ingelheim, AstraZeneca, Daiichi Sankyo, and TERUMO Corporation. Dr. Yamamoto reports receiving honoraria and a research grant from Taiho Pharmaceutical, Merck Sharp & Dohme, Chugai Pharmaceutical, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Amgen, Tumura, Sanofi, Daiichi Sankyo, Nippon Kayaku, Novartis, Bristol-Myers Squibb, and Janssen; receiving honoraria from Ono Pharmaceutical, Otsuka, Guardant Health Japan, Kyowa-Kirin, GlaxoSmithKline, Takeda, Pfizer, Miyarisan, and Merck; receiving research grant from Shionogi, Asahikasei-pharma, Astellas, Esai, AbbVie, and Tosoh; and serving on the advisory board of AstraZeneca, Eli Lilly, and Takeda. Dr. Gemma reports receiving honoraria from Ono Pharmaceutical. The remaining authors declare no conflict of interest.

Address for correspondence: Yoshihito Kogure, MD, PhD, Department of Respiratory Medicine, National Hospital Organization Nagoya Medical Center, 4-1-1 Sannomaru, Naka-ku, Nagoya 460-0001, Japan. E-mail: yoshihito.kogure@nnh.go.jp

Cite this article as: Kogure Y, Kada A, Hashimoto H, et al. Survival impact of second-line immune checkpoint inhibitors in older patients with advanced squamous-cell NSCLC: post hoc analysis of the CAPITAL study. *JTO Clin Res Rep*. 2023;4:100514.

ISSN: 2666-3643

https://doi.org/10.1016/j.jtocrr.2023.100514

ABSTRACT

Introduction: In the CAPITAL study, a randomized phase 3 study, wherein carboplatin plus nab-paclitaxel treatment was compared with docetaxel treatment for older patients with squamous-cell lung cancer, the former became the new standard of care for such patients. Our study aimed to evaluate whether the efficacy of second-line immune checkpoint inhibitors (ICIs) affected the primary analysis of overall survival (OS).

Methods: Herein, we performed a post hoc analysis of the impact of second-line ICIs on OS, safety in each group of participants aged more than 75 years, and intracycle nab-paclitaxel skip status.

Results: Patients were randomly allocated to the carboplatin plus nab-paclitaxel (nab-PC) arm (n = 95) or the docetaxel (D) arm (n = 95). Of these patients, 74 of 190 (38.9%) were transferred to ICIs for second-line treatment (nab-PC arm: 36, D arm: 38). A survival benefit was numerically observed only for patients for whom first-line therapy was terminated owing to disease progression (median OS [nab-PC arm]: with and without ICIs, 321 and 142 d, respectively; median OS [D arm]: with and without ICIs, 311 and 256 d, respectively). The OS among patients who received ICI after adverse events was similar in the two arms. In the D arm, a significantly higher frequency of grade greater than or equal to 3 adverse events was observed among patients aged more than or equal to 75 years (86.2%) than among those aged less than 75 years (65.6%, $p = 0.041$), including a significantly higher frequency of neutropenia (84.6% versus 62.5%, $p = 0.032$); no such differences were observed in the nab-PC arm.

Conclusions: We found that second-line ICI treatment seemed to have a little impact on OS.

Keywords: Squamous-cell non–small cell lung cancer; Elderly; Immune checkpoint inhibitors; Nab-paclitaxel

Introduction

In Japan, cytotoxic monotherapy has been the standard therapy for older patients with advanced squamous NSCLC in the past few decades.[1,2] Carboplatin plus nab-paclitaxel was suggested to be effective for squamous NSCLC and to improve overall survival (OS) for older patients in the subgroup analysis of CA-031 study.[3,4] On the basis of these results, we conducted a CAPITAL study to evaluate the efficacy and safety of carboplatin plus nab-paclitaxel compared with docetaxel for older patients with advanced squamous-cell NSCLC.[5] In this study, compared with docetaxel, carboplatin plus nab-paclitaxel yielded better OS (16.9 versus 10.0 mo, $p < 0.001$), progression-free survival (PFS) (5.8 versus 4.0 mo, $p < 0.001$), and objective response rate (ORR, 66% versus 28%, $p < 0.001$).[6]

Currently, immune checkpoint inhibitors (ICIs) are the mainstay of treatment for advanced NSCLC. OS improves in second-line treatment with ICI compared with that in cytotoxic monotherapy.[7–10] If the efficacy of the second-line treatment in each group was different, the difference may have influenced the interpretation of OS in this study. Therefore, we conducted this post hoc analysis of the aforementioned phase 3 study to evaluate the impact of second-line ICIs on survival. As an additional analysis, we evaluated the intracycle nab-paclitaxel skip status and the safety of each group of patients aged above 75 years and below 75 years.

Materials and Methods

This multicenter, open-label, phase 3, randomized trial was performed at 92 institutions in Japan. The inclusion criteria were as follows: advanced squamous-cell NSCLC with no prior systemic chemotherapy, age above or equal to 70 years, and Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0 or 1. Patients were randomly allocated, in a 1:1 ratio, to the nab-PC arm (carboplatin [area under the free carboplatin plasma concentration versus time curve], 6 mg/mL/min plus nab-paclitaxel, 100 mg/m^2 weekly) or the D arm (docetaxel, 60 mg/m^2) for each 21-day cycle and stratified according to ECOG PS (0 versus 1), age (<75 y versus ≥75 y), clinical stage (IIIB versus IV versus recurrent), sex (male versus female), and institution. OS was the primary end point of this study. The secondary end points were PFS, ORR, safety, and quality of life. The study was approved by the Clinical Research Ethics Committee of each participating institution, and written informed consent was obtained from each patient before their participation in the study. This study was conducted in accordance with the principles of the Declaration of Helsinki, 2013.

Post Hoc Analysis

In the post hoc analysis, the primary objective was to evaluate whether the efficacy of second-line ICIs had an impact on the primary analysis of OS in the CAPITAL study. The secondary objectives were to evaluate the intracycle nab-paclitaxel skipping status (the number of patients who had to skip nab-paclitaxel on day 8 or 15) and to compare safety between patients aged above 75 years and those aged below 75 years. For ICI on survival, after confirming the reasons for discontinuation

Table 1. Second-Line ICI Use According to the Reason for Termination of Study Treatment

	D Arm (n = 95)		Nab-PC Arm (n = 95)	
	Second-Line ICI		Second-Line ICI	
Reason for Termination of Study Treatment	Yes	No	Yes	No
PD, n (%)	29 (60.4)	19 (39.6)	22 (55.0)	18 (45.0)
AE, n (%)	8 (26.7)	22 (73.3)	8 (25.0)	24 (75.0)
Others,[a] n (%)	1 (5.9)	16 (94.1)	6 (26.1)	17 (73.9)

[a]Patient request, died, or lost to follow-up.
AE, adverse event; D, docetaxel; ICI, immune checkpoint inhibitor; Nab-PC, carboplatin plus nab-paclitaxel; PD, progressive disease.

of the study treatment, we evaluate whether ICI has an impact on survival.

Statistical Analysis

The full analysis set was used for the efficacy analysis. The safety analysis included patients who received at least one cycle of the trial treatment. ICIs were used as posttreatment options after discontinuation of the study treatment. Therefore, logistic regression was performed, controlling for the variables treatment group, age (<75 y or ≥75 y), ECOG PS, cancer stage (IIIB, IV, or postoperative recurrence), reason for discontinuation, and best response before discontinuation, to determine who was more likely to use posttreatment ICIs. The reasons for discontinuation were categorized as follows: adverse events (AEs), disease progression, and others. As the reason for discontinuation was largely related to the use of ICIs, OS was determined according to the reason, ICI use, and group. OS was defined as the time from termination of the study treatment to the day of death of any cause. Survivors were censored on the final day of their confirmed survival. OS was estimated using the Kaplan–Meier method, and the median OS and 95% confidence intervals (CIs) were calculated. The frequency of AEs by age group was compared using Fisher's exact test. All statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC).

Results

From February 2016 to August 2020, 196 patients were enrolled and randomly allocated to either the nab-PC arm (n = 98) or the D arm (n = 98). Of these, three patients in each group were excluded from the full analysis set because they did not receive any treatment, withdrew consent, or became ineligible after random assignment. A flow diagram is found in Supplementary Figure 1.

The reasons for termination of the study treatment are found in Table 1. In the D arm, 48 patients (50.5%) experienced progressive disease (PD) and 30 patients (31.6%) terminated the study treatment for AEs. In the nab-PC arm, 40 patients (42.1%) experienced PD and 32 patients (33.7%) terminated the study treatment for AEs. In total, 38 patients (40.0%) in the D arm and 36 patients (37.9%) in the nab-PC arm received ICI as a second-line therapy.

After PD, 29 (60.4%) and 22 (55.0%) patients received ICI as a second-line therapy in the D and nab-PC arms, respectively. Among the patients who terminated the study treatment for AEs, eight in each arm were treated with ICI as a second-line therapy (D arm, 26.7%; nab-PC arm, 25.0%). Second-line ICI treatment was more often necessitated by AEs and PD than by other reasons (OR = 1.2 [95% CI: 0.4–3.6] and 4.8 [1.8–12.9], respectively; Table 2).

In the D arm, the median OS was 311 days (95% CI: 219–428) among the patients receiving the ICI treatment after PD and 256 days (95% CI: 75–478) among the patients who did not receive the ICI treatment after PD; the corresponding median OS for patients with AEs were 186.5 (95% CI: 85–501) days and 154 (95% CI: 116–314) days, respectively (Fig. 1*A*). A numerical survival benefit was observed after the treatment with ICIs only for patients who developed PD during the first-line therapy. In the nab-PC arm, the median OS was 321 (95% CI: 163–804) days among the patients receiving ICIs after PD and 142 (95% CI: 51–293) days among those not receiving ICIs after PD; the corresponding values among the patients with AEs were 224 (95% CI: 150–not applicable) and 373 (95% CI: 206–639) days, respectively. The median OS was similar between patients with and those without post-AE ICI treatment, whereas a numerical survival benefit of ICI was observed among the patients with PD (Fig. 1*B*).

The overall intracycle skipping of nab-paclitaxel in this study is illustrated in Figure 2*A*. In cycle 1, nab-paclitaxel was skipped by 11 of 95 patients (11.6%) on day 8 and by 36 of 95 (37.9%) on day 15. In cycle 2, 13 of 83 (15.7%) and 28 of 83 patients (33.7%) skipped the administration on days 8 and 15, respectively. Finally, only 12 of 27 patients received the drug as planned on day 8 and 14 of 27 on day 15, in cycle 6.

The frequency of skipping nab-paclitaxel on day 8 did not differ according to age. Nevertheless, the skipping

Table 2. ORs for Second-Line ICI Against Reference Categories

Variables	Second-Line ICI Yes (n)	Total (n)	Second-Line ICI Yes (%)	OR	95% CI Lower	95% CI Upper
Group A (D)	38	95	40.0	1.0	—	—
Group B (nab-PC)	36	95	37.9	1.1	0.5	2.2
Reason for discontinuation, PD	51	88	58.0	4.8	1.8	12.9
AEs	16	62	25.8	1.2	0.4	3.6
Other	7	40	17.5	1.0	—	—
Age, 0–74 y	23	65	35.4	1.0	—	—
≥75 y	51	125	40.8	1.5	0.7	3.0
ECOG PS, 0	26	63	41.3	1.0	—	—
1	48	127	37.8	0.9	0.4	1.8
Stage, III B	14	32	43.8	2.3	0.6	7.9
IV	52	134	38.8	1.7	0.6	4.6
Postoperative recurrence	8	24	33.3	1.0	—	—
Best response before discontinuation, missing	2	19	10.5	—	—	—
Not CR/PR	41	87	47.1	1.3	0.6	2.7
CR/PR	31	84	36.9	1.0	—	—
Total	74	190	38.9	—	—	—

AE, adverse event; CI, confidence interval; CR, complete response; D, docetaxel; ECOG, Eastern Clinical Oncology Group; ICI, immune checkpoint inhibitor; nab-PC, carboplatin plus nab-paclitaxel; PD, progressive disease; PR, partial response; PS, performance status.

rate was higher on day 15 among patients aged above or equal to 75 years than that among younger patients throughout all cycles (Fig. 2*B* and *C*).

There was no difference in toxicity between patients aged above 75 years and below 75 years in the nab-PC group; the frequency remained approximately 80%.

In the D arm, a significantly higher frequency of grade 3 or higher any AEs (86.2% versus 65.6%, $p = 0.041$) was observed in patients aged above or equal to 75 years than in those aged below 75 years, including a significantly higher frequency of neutropenia (84.6% versus 62.5%, $p = 0.032$) (Table 3). No other statistically significant differences in individual AEs were observed in the D arm and none at all in the nab-PC arm (82.4% versus 83.6%, $p = 1.00$).

Dose intensities (DIs) of the drugs administered in the study are listed in Table 4. The relative DI of docetaxel was 0.92 and that of carboplatin was 0.94, whereas that of nab-paclitaxel was only 0.72.

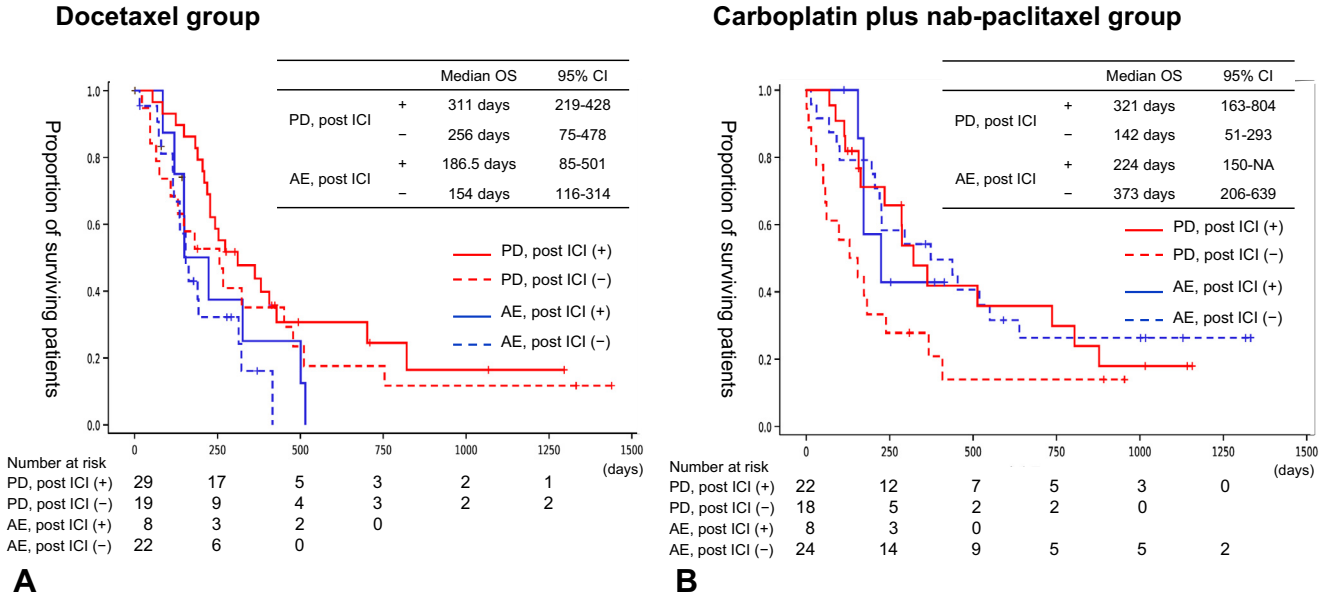


Figure 1. OS from termination of study. Kaplan–Meier plots of OS for patients were classified according to the combination of reason for discontinuation and ICI use. AE, adverse event; CI, confidence interval; NA, not applicable; OS, overall survival; PD, progressive disease.

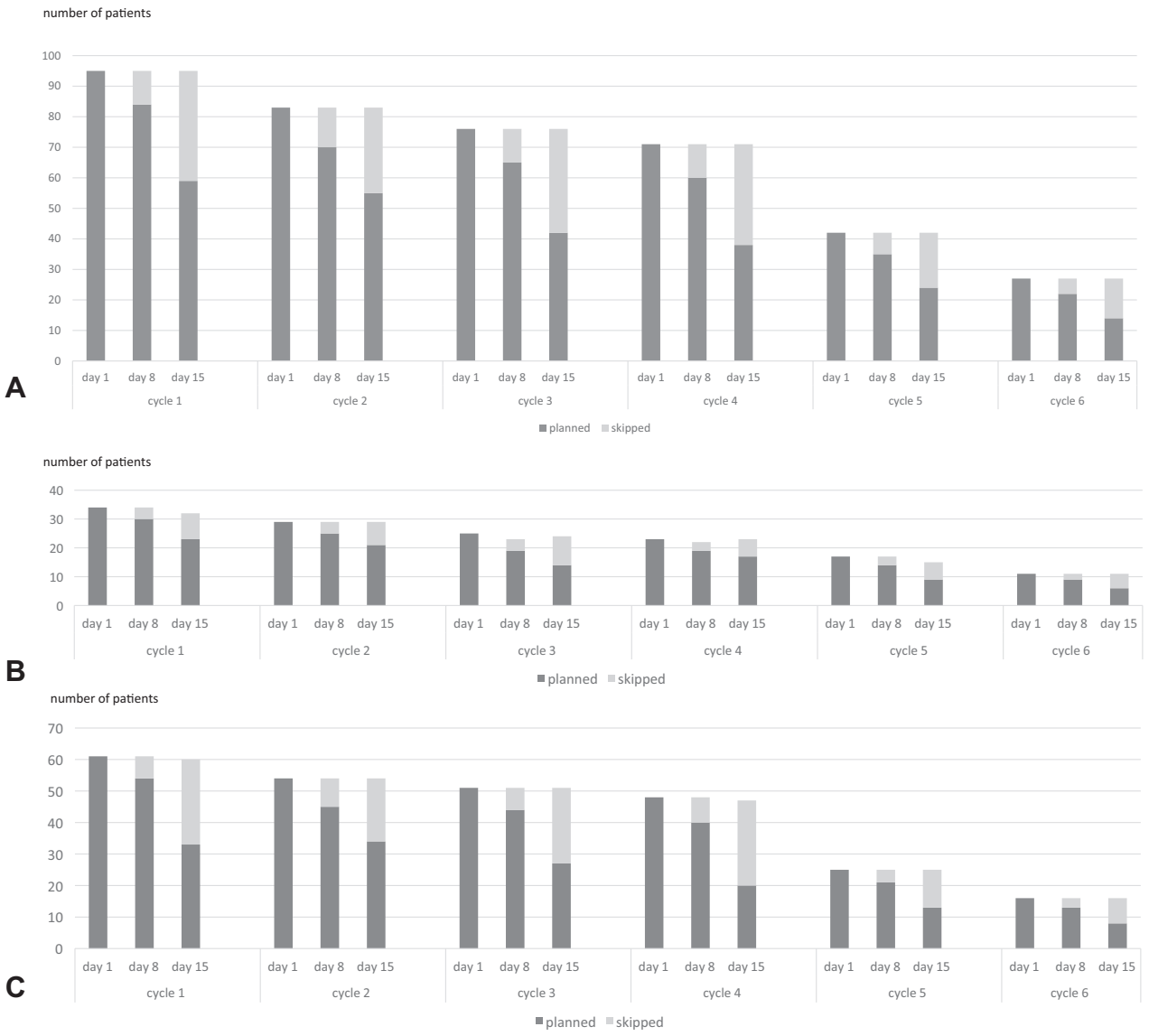


Figure 2. Intracycle skipping of nab-paclitaxel. Intracycle skipping of nab-paclitaxel among (*A*) patients in total, (*B*) patients aged <75 years, and (*C*) those aged ≥75 years.

Discussion

Approximately 40% of patients in both arms of the study received ICI therapy as a second-line therapy. Among these patients, approximately 60% received this therapy owing to PD. The median OS for patients receiving ICI treatment after PD was 55 days longer than that for patients not receiving ICI treatment after PD in the D arm; similarly, the median OS was 179 days longer in the nab-PC arm. The improvement of OS was higher in the nab-PD arm. The nab-PC arm had a better response rate (28.0% in the D-arm versus 66.3% in the nab-PC arm, $p < 0.0001$) and no additional toxicity in patients aged above or equal to 75 years. These results could lead to better general conditions in the nab-PC arm at the time of PD, and this difference may have led to a disparity in the benefits between the two arms. Nevertheless, as the OS of patients receiving ICI treatment after PD was comparable between the arms of the study, we do not believe that such differences affected the OS in the main analysis of this study. Furthermore, little survival benefit was associated with ICI treatment after AEs. This can be explained by the definition of survival time in this analysis: from the end of first-line treatment to the day of death of any cause. Therefore, the time from discontinuation of toxicity to the start of second-line ICI therapy could have varied widely between the patients. This is a limitation of the present study. Another limitation of this study was that data on the date of post–first-line treatment, including those with second-line ICIs, were not collected. This prohibited us from performing an analysis, in which the timing of the

Table 3. AEs of Grade Greater Than or Equal to 3 in Each Group According to Treatment and Age

	D Arm (n = 97)			Nab-PC Arm (n = 95)		
AE	<75 y (n = 32)	≥75 y (n = 65)	*p* Value	<75 y (n = 34)	≥75 y (n = 61)	*p* Value
WBC	15 (46.9)	40 (61.5)	0.25	16 (47.1)	28 (45.9)	1.00
ANC	20 (62.5)	55 (84.6)	0.032	20 (58.8)	28 (45.9)	0.66
Hb	1 (3.1)	1 (1.5)	1.00	11 (32.4)	26 (42.6)	0.45
PLT	0	1 (1.5)	1.00	3 (8.8)	9 (14.8)	0.62
FN	3 (9.4)	16 (24.6)	0.12	3 (8.8)	6 (9.8)	1.00
Neuropathy[a]	0	0	NA	0	1 (1.6)	1.00
Fatigue	2 (6.3)	1 (1.5)	0.50	3 (8.8)	9 (14.8)	0.62
Myalgia	0	0	NA	0	1 (1.6)	1.00
Arthralgia	0	1 (1.5)	1.00	0	2 (3.3)	0.82
Edema	1 (3.1)	0	0.66	0	1 (1.6)	1.00
Any	21 (65.6)	56 (86.2)	0.04	28 (82.4)	51 (83.6)	1.00

Note: Data are presented as counts (%) of patients experiencing each AE at a grade of ≥3.
[a]Grade ≥ 2.
AE, adverse event; ANC, absolute neutrophil count; D, docetaxel; FN, febrile neutropenia; Hb, hemoglobin; NA, not applicable; Nab-PC, carboplatin plus nab-paclitaxel; PLT, platelet; WBC, white blood cell.

second-line ICI treatment was incorporated. From these limitations, we may not be able to dissect whether survival benefit was associated with the ICI treatment after AEs or not.

In the D-arm, grade 3 or higher AEs were significantly more common in the subgroup of patients aged above or equal to 75 years than in those aged below 75 years (86.2% versus 65.6%, $p = 0.041$), but no such difference was observed in the nab-PC arm (83.6% versus 82.4%, $p = 1.00$). A possible reason for this difference is the dosing schedule used. Docetaxel was administered every 3 weeks, whereas nab-paclitaxel was administered weekly. Weekly administration enables dose modification within the cycle. As revealed by the lower relative DI for nab-paclitaxel than that for docetaxel, intracycle skipping was a frequent occurrence with nab-paclitaxel. Indeed, approximately 15% and 35% of the patients in that arm skipped the treatment on days 8 and 15, respectively, up to cycle 2. After cycle 3, more than half of the patients had to skip the nab-paclitaxel treatment on day 8 and 45% on day 15. We believe that such intracycle skipping allowed patients aged above or equal to 75 years to continue the treatment without increased toxicity. As a considerable survival benefit was observed despite a relative DI of 0.72 for nab-paclitaxel, clinicians need not hesitate to perform dose modifications, such as skipping or dose reduction, in clinical practice. Langer et al.[11] reported on the safety and efficacy of nab-paclitaxel plus carboplatin every 3 weeks or with a 1-week break between cycles in older patients with advanced NSCLC. Although the 1-week break between cycles did not significantly reduce occurrence of the primary end point (the percentage of grade ≥2 peripheral neuropathy and grade ≥3 myelosuppression), it improved the ORR and PFS. On the basis of their results, schedule modification may also improve the prognosis.

An investigation of whether the reasons for discontinuation of first-line treatment are related to the effectiveness of ICI in second-line treatment is warranted. Recently, ICI has been used mainly in the first-line setting for patients with advanced NSCLC, and it is difficult to validate this clinical question.

In conclusion, the impact of second-line ICI treatment on survival was similar between the two arms of our study. ICI as a post–first-line treatment seemed to have little impact on the interpretation of OS in the CAPITAL study. The percentage of patients for whom nab-paclitaxel regimens were skipped on days 8 and 15 changed as treatment cycles progressed. Although the

Table 4. Dose Intensity of the Drugs Administered in the Study

Study Drug	Median Cycles (Range)	Theoretical DI/wk	DI/wk (Range)	Relative DI (Range)
Docetaxel	3 (1-15)	20 mg/m^2	18.3 mg/m^2 (6.2-20.0)	0.92 (0.31-1.00)
Nab-paclitaxel	4 (1-26)	100 mg/m^2	72.0 mg/m^2 (33.3-100.0)	0.72 (0.33-1.00)
Carboplatin	4 (1-26)	AUC: 2 mg/mL/min	AUC: 1.9 mg/mL/min (0.2-2.0)	0.94 (0.10-1.00)

AUC, area under the concentration curve; DI, dose intensity.

toxicity of carboplatin plus nab-paclitaxel did not differ according to age, that of docetaxel was significantly worse in patients aged above or equal to 75 years than in those aged less than 75 years.

CRediT Authorship Contribution Statement

Yoshihihito Kogure: Conceptualization, Methodology, Writing—original draft, Writing—review and editing, Final manuscript approval.

Akiko Kada: Methodology, Data analysis, Writing—original draft, Writing—review and editing, Final manuscript approval.

Hiroya Hashimoto: Methodology, Data analysis, Writing—original draft, Writing—review and editing, Final manuscript approval.

Shinji Atagi: Conceptualization, Writing—review and editing, Final manuscript approval.

Yuichi Takiguchi: Conceptualization, Writing—review and editing, Final manuscript approval.

Hideo Saka: Conceptualization, Writing—review and editing, Final manuscript approval.

Noriyuki Ebi: Conceptualization, Writing—review and editing, Final manuscript approval.

Akira Inoue: Conceptualization, Writing—review and editing, Final manuscript approval.

Takayasu Kurata: Conceptualization, Writing—review and editing, Final manuscript approval.

Yuka Fujita: Writing—review and editing, Final manuscript approval.

Yoichi Nishii: Writing—review and editing, Final manuscript approval.

Hidetoshi Itani: Writing—review and editing, Final manuscript approval.

Takeo Endo: Writing—review and editing, Final manuscript approval.

Akiko M. Saito: Data verification, Final manuscript approval.

Takuo Shibayama: Writing—review and editing, Final manuscript approval.

Nobuyuki Yamamoto: Writing—review and editing, Supervision, Final manuscript approval.

Akihiko Gemma: Writing—review and editing, Supervision, Final manuscript approval.

Acknowledgments

The authors thank the patients, their families, and all the investigators who participated in this study. This study was supported in part by the National Cancer Center Research and Development Fund (26-A-22, 29-A-15, and 2020-A-13). This study was conducted by the National Hospital Organization Nagoya Medical Center under a funding contract with Taiho Pharmaceutical Co., Ltd., Tokyo, Japan.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *JTO Clinical and Research Reports* at www.jtocrr.org and at https://doi.org/10.1016/j.jtocrr.2023.100514.

References

1. Kudoh S, Takeda K, Nakagawa K, et al. Phase III study of docetaxel compared with vinorelbine in elderly patients with advanced non-small-cell lung cancer: results of the West Japan Thoracic Oncology Group Trial (WJTOG 9904). *J Clin Oncol*. 2006;24:3657–3663.
2. Abe T, Takeda K, Ohe Y, et al. Randomized phase III trial comparing weekly docetaxel plus cisplatin versus docetaxel monotherapy every 3 weeks in elderly patients with advanced non-small cell lung cancer: the intergroup trial JCOG0803/WJOG4307L. *J Clin Oncol*. 2015;33:575–581.
3. Socinski MA, Okamoto I, Hon JK, et al. Safety and efficacy analysis by histology of weekly nab-paclitaxel in combination with carboplatin as first-line therapy in patients with advanced non-small-cell lung cancer. *Ann Oncol*. 2013;24:2390–2396.
4. Socinski MA, Langer CJ, Okamoto I, et al. Safety and efficacy of weekly nab®-paclitaxel in combination with carboplatin as first-line therapy in elderly patients with advanced non-small-cell lung cancer. *Ann Oncol*. 2013;24:314–321.
5. Kogure Y, Saka H, Takiguchi Y, et al. A randomized phase III study comparing carboplatin plus nab-paclitaxel versus docetaxel for elderly patients with squamous cell lung cancer: study protocol. *Clin Lung Cancer*. 2018;19:e711–e715.
6. Kogure Y, Iwasawa S, Saka H, et al. Efficacy and safety of carboplatin with nab-paclitaxel versus docetaxel in older patients with squamous non-small-cell lung cancer (CAPITAL): a randomised, multicentre, open-label, phase 3 trial. *Lancet Healthy Longev*. 2021;2:e791–e800.
7. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med*. 2015;373:123–135.
8. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373:1627–1639.
9. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet*. 2017;389:255–265.
10. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375:1823–1833.
11. Langer CJ, Kim ES, Anderson EC, et al. Nab-paclitaxel-based therapy in underserved patient populations: the ABOUND.70+ study in elderly patients with advanced NSCLC. *Front Oncol*. 2018;8:262.

BMJ Health & Care Informatics

Twenty-year follow-up of promising clinical studies reported in highly circulated newspapers: a meta-epidemiological study

Aran Tajika ,[1] Yasushi Tsujimoto,[1,2] Akira Onishi,[3] Yusuke Tsutsumi ,[4,5] Satoshi Funada ,[1] Yusuke Ogawa,[6] Nozomi Takeshima,[1] Yu Hayasaka,[1] Naotsugu Iwakami,[1] Toshi A Furukawa [1]

To cite: Tajika A, Tsujimoto Y, Onishi A, *et al*. Twenty-year follow-up of promising clinical studies reported in highly circulated newspapers: a meta-epidemiological study. *BMJ Health Care Inform* 2023;**30**:e100768. doi:10.1136/bmjhci-2023-100768

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2023-100768).

Received 23 March 2023
Accepted 29 May 2023

For numbered affiliations see end of article.

Correspondence to
Dr Aran Tajika;
aran.tajika28@gmail.com

ABSTRACT

Objectives Researchers have identified cases in which newspaper stories have exaggerated the results of medical studies reported in original articles. Moreover, the exaggeration sometimes begins with journal articles. We examined what proportion of the studies quoted in newspaper stories were confirmed.

Methods We identified newspaper stories from 2000 that mentioned the effectiveness of certain treatments or preventions based on original studies from 40 main medical journals. We searched for subsequent studies until June 2022 with the same topic and stronger research design than each original study. The results of the original studies were verified by comparison with those of subsequent studies.

Results We identified 164 original articles from 1298 newspaper stories and randomly selected 100 of them. Four studies were not found to be effective in terms of the primary outcome, and 18 had no subsequent studies. Of the remaining studies, the proportion of confirmed studies was 68.6% (95% CI 58.1% to 77.5%). Among the 59 confirmed studies, 13 of 16 studies were considered to have been replicated in terms of effect size. However, the results of the remaining 43 studies were not comparable.

Discussion In the dichotomous judgement of effectiveness, about two-thirds of the results were nominally confirmed by subsequent studies. However, for most confirmed results, it was impossible to determine whether the effect sizes were stable.

Conclusions Newspaper readers should be aware that some claims made by high-quality newspapers based on high-profile journal articles may be overturned by subsequent studies within the next 20 years.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ When newspapers cite the results of clinical research articles, they sometimes misrepresent the results based on exaggerated expectations.
⇒ Studies with higher levels of evidence may overturn the results of clinical research.

WHAT THIS STUDY ADDS

⇒ The results of clinical research articles were relatively stable in papers in which the citation source was properly listed in the newspaper article.
⇒ However, the results of approximately one-third of the papers were overturned in the following two decades.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Journalists should be careful in accurately reporting clinical research articles and stating the sources of their citations.
⇒ Readers should be aware that more than a few claims made in highly circulated newspapers based on high-profile journal articles may still be overturned by subsequent studies.

INTRODUCTION

As people's health awareness has increased, newspapers have covered more stories about health and medicine. These stories feature many diseases, including cancer, stroke, infectious diseases and mental disorders. Some sensationalise the fear and frustration of the disease, while others provide hope for new treatments or preventative measures. These stories are often based on articles published in medical journals. The important points of these articles are summarised and presented clearly in newspaper stories for the general public.

However, the media coverage often exaggerates fear and hopes.[1] For example, a phase I uncontrolled study of a new cancer drug published in the *New England Journal of Medicine* showed some effects in one subgroup. Newspapers reported that this treatment produced highly promising results.[1] However, studies cited in newspaper stories are sometimes overturned. Gonon[2] investigated the 'top 10' most frequently reported studies on attention deficit hyperactivity disorder (by newspapers) and compared these results with

those of subsequent studies. Two studies were confirmed, four attenuated, three refuted and one was neither confirmed nor refuted.

When the strength of the research design is considered, randomised controlled trials (RCTs) and their meta-analyses provide the strongest evidence for treatment decisions. However, newspapers are more likely to report observational studies (OSs) than RCTs.[3] Notably, exaggeration often begins with medical journal articles themselves.[1] One problem with studies with weak evidence is that the reproducibility of the results is low. Ioannidis conducted a simulation study and noted that a meta-analysis of good-quality RCTs and adequately powered RCTs assumed a reproducibility of 85%, but only 23% for underpowered RCTs and approximately 20% for adequately powered OSs.[4] Ioannidis[5] identified studies cited more than 1000 times in high-impact factor (IF) journals in general and internal medicine. When these studies were compared with subsequent studies that theoretically had better-controlled designs, only half of the RCTs and none of the OSs were replicated. Furthermore, when statistically significant and extremely favourable initial reports of intervention effects were examined, it was found that the majority of such large treatment effects emerged from small studies. When additional trials were performed, the effect sizes typically became much smaller.[6] When newspapers report and overestimate the results of these initially promising studies, the information that reaches the public may be doubly overstated.

This study investigated the trustworthiness of medical news. We examined whether newspaper reports were confirmed through subsequent studies that examined the same clinical questions. In other words, we examined how much caution general readers need to exercise when reading newspaper reports on medical research.

METHODS

Selection of newspaper stories and original studies

We selected four quality papers (two from the USA and two from the UK) and four non-quality papers (two from the USA and two from the UK) with the highest circulation according to the Audit Bureau of Circulations[7] and Alliance for Audited Media.[8] We examined these two newspaper types for several reasons. Generally, quality papers are believed to have higher quality reporting than non-quality papers,[9] which tend to focus on readers' emotions rather than on the veracity of the reports.[10] However, when we consider the respective circulations of the two types of papers, non-quality papers have as many readers as quality papers; they sometimes have more power to lead public opinion.[11]

We selected newspaper articles that quoted main medical journals. First, we selected medical journals from the following two fields: 'general and internal medicine' and 'public, environmental and occupational health' according to their journal IF on Journal Citation Reports. In addition, we selected the 20 journals in each field with the highest IFs for 2000. We ultimately selected 40 medical journals as an ad hoc set of representative medical journals that might meet the public interest. Next, we searched the LexisNexis database,[12] which contains stories from prominent newspapers worldwide. We used the names of 40 medical journals as search words and selected newspaper stories:

- Printed in 2000 in the four above-mentioned quality and four non-quality newspapers.
- That quoted articles that were published in the above-mentioned 40 journals.
- In which we could identify the original medical journal article.
- That mentioned the effectiveness, recommendation of treatment or prevention at that time.

Pairs of independent investigators (AT, YO, NT, YH and NI) selected eligible newspaper articles for analysis. Disagreements were resolved through discussions between the two investigators and, when necessary, in consultation with a third author (TAF). We found the original articles quoted in these newspapers. When two or more articles were quoted in a newspaper story, we selected all the articles. When the number of eligible studies was greater than 100, 100 studies were randomly selected. Original articles were classified into the following categories:

- Animal or laboratory study.
- Clinical study.
  - Case reports or case series.
  - OS.
  - RCT.
  - Systematic review (SR) of OSs with or without meta-analysis.
  - SR of RCTs with or without meta-analysis.
  - Other reviews (eg, narrative reviews).
- Others (eg, comment, letter).

We excluded studies in which specific clinical questions were not identifiable (eg, health economics studies) because we could not search for corresponding subsequent studies in the next step.

Selection of subsequent studies on the same clinical questions

For each original article, we searched for subsequent studies that examined the same clinical questions using 'stronger' research designs. The evidence levels of all the studies were classified according to the following hierarchy:

1. SR of RCTs.
2. Single RCT.
3. SR of OSs/single OS.
4. Case series/a case study.

The characteristics of 'stronger design' are as follows[5 13]:

- The subsequent study used a design with a higher level of evidence hierarchy than the original study.
- If studies had the same level of evidence hierarchy, a study with a larger sample size constituted stronger evidence.

- If the design of the original study was an SR of an RCTs, we searched for the latest SR for the RCTs.
- If the design of the original study was the SR of OSs or other reviews, we searched for the largest RCT or the latest meta-analysis of RCTs. If we could not find these studies, we searched for the latest OS meta-analyses.
- If the original study was an animal or laboratory study, we searched for the most appropriate clinical study asking the same clinical question according to the evidence hierarchy.

First, two authors (AT, YaT, AO, YuT and SF) independently searched the Web of Science for potential new papers in which the original paper was cited through December 2021. Subsequently, to prevent search omissions, AT conducted a PubMed search through June 2022 to search for anything more valid than the candidates' new articles on the Web of Science. If new candidate papers were found, the authors discussed them in pairs to identify the new papers. The PubMed search was conducted using the most comprehensive terms possible, and the search formula was documented.

Comparisons of original and subsequent studies

We extracted the data when the original study authors presented their primary outcomes. If the authors failed to designate their primary outcome(s), the outcome described first was considered the primary outcome. Next, we extracted the outcomes of the subsequent studies, which were as similar as possible to those of the original studies.

We conducted the following two-step comparison. First, we compared the effectiveness of the original studies with that of newer studies and classified each comparison into one of three categories: 'unchallenged', 'contradicted' or 'confirmed'.[5 13]

- Unchallenged: when there was no subsequent study with a higher level of evidence.
- Contradicted: when a subsequent study denied the effectiveness of the original study.
- Confirmed: The original and subsequent studies concluded that the intervention was effective, regardless of the effect size difference.

When we could not compare these outcomes, we compared the benefits and applicability of both studies and made qualitative judgements.

Furthermore, among 'confirmed' cases, when the outcomes of both original and subsequent studies were exactly comparable (ie, when a new paper was a meta-analysis, the original paper was included in the funnel plot of the new paper, and accurate effect size comparison was possible), we compared the effect sizes of both studies. Outcomes were extracted as continuous or dichotomous data such as standardised mean difference (SMD), OR, risk ratio (RR) or HR. We gave preference to continuous data. We compared these values when the SMD was shown in the subsequent meta-analysis, and when the SMD of the original paper was shown in that study. When studies showed effectiveness using only dichotomous data, the OR was calculated first. We then converted OR into SMD using the following formula[14]:

$$\mathrm{SMD} = \frac{\sqrt{3}}{\pi} \ln \mathrm{OR}$$

We classified 'confirmed' cases into one of two categories: 'initially stronger effects' or 'replicated'.[13]

- Initially stronger effects: when the point estimate of the original study was not included in the 95% CI of the SMD of the subsequent study or the SMD of the original study was 0.2 SD units or greater than that of the subsequent study (0.2 SD units would signify a small effect difference according to Cohen's rule of thumb).[15]
- Replicated: when the point estimate of the original study was included in the 95% CI of the SMD of the subsequent study, and the two SMDs were within 0.2 SD units apart, or the effect size of the subsequent study was larger than that of the original study.

When the SMD could not be calculated from the RR or the study showed only the HR, as it could not be converted into SMD, we directly compared only the RRs or HRs. Their 95% CI was presented in the papers without considering the difference of 0.2 SD units of SMDs.

Outcomes

Primary outcome

We defined the primary outcome, 'the proportion of confirmed studies', as follows:

$$\text{Proportion of confirmed studies} = \frac{\text{Confirmed studies}}{\text{Total studies} - \text{Unchallenged studies}} \times 100\ (\%)$$

Secondary outcomes

We classified the original studies according to their research design and medical fields and examined the differences between quality and non-quality papers. The proportion of confirmed studies in each subgroup was calculated.

Analyses

Statistical analyses were performed using STATA V.17.0. Statistical differences among subgroup categories were tested using the χ^2 test, and SMD was compared using the Wilcoxon signed-rank test. The level of significance was set at $p<0.05$ (two tailed).

Patient and public involvement

No patients or public members were involved in conducting this research.

RESULTS

Characteristics of newspaper stories, original studies and subsequent studies

Figure 1 illustrates the details of the search. The eight newspapers selected were the *New York Times* (USA, quality), *Washington Post* (USA, quality), *Daily Telegraph* (UK, quality), *Times* (UK, quality), *USA Today* (USA,

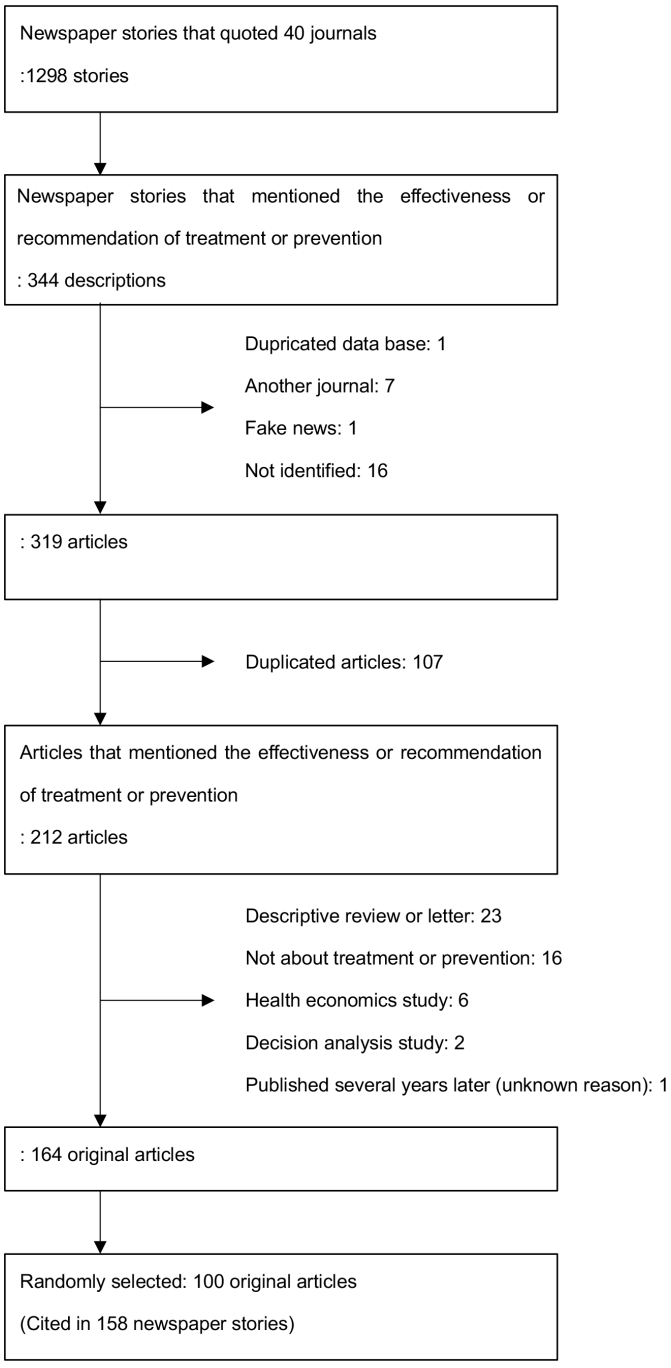


Figure 1 Flow chart of original study identification process.

non-quality), *Daily News* (USA, non-quality), *Daily Mail* (UK, non-quality) and *Daily Mirror* (UK, non-quality). When searching for journal names in newspaper stories, we found 1298 newspaper stories, of which 344 described the effectiveness of or recommended certain treatments or preventive measures (kappa=0.73) (table 1). Online supplemental eTable 1 lists the names of 40 medical journals.

A total of 344 newspaper stories were referred to in 319 scientific journal articles. After excluding duplicates, we identified 212 articles that mentioned the effectiveness of the recommended treatment or prevention. We excluded 48 articles because the research questions could not be identified. Finally, we identified 164 original articles and randomly selected 100 of them. These were cited in 158 newspaper articles. The journals in which the 100 original articles were published were as follows: *New England Journal of Medicine* (*NEJM*), 39; *Journal of the American Medical Association* (*JAMA*), 21; *Lancet*, 16; *British Medical Journal* (*BMJ*), 9; *Archives of Internal Medicine*, 8; *Annals of Internal Medicine*, 3; *American Journal of Epidemiology*, 1; *American Journal of Public Health*, 1; *Infection Control and Hospital Epidemiology*, 1; and *Mayo Clinic Proceedings*, 1. Approximately three-quarters of these articles were published in three major journals (*NEJM, JAMA, Lancet*).

Of the 100 articles, 58 were RCTs and 31 OSs. A few other designs corresponded to various ICD-10 categories. Of the 158 newspaper stories, two-thirds were in quality papers and the rest in non-quality.

For four of the 100 original studies, the newspapers stated their effectiveness, but the primary outcome of those studies did not indicate their effectiveness. Therefore, these were excluded from this study. In the remaining 96 studies, 104 effective treatments were identified. Subsequent studies on each treatment were searched. We identified relevant subsequent studies for 86 of these 104 treatments. The 18 others remained unchallenged (table 2). Of the 86 subsequent studies, 83 were SR (SR of RCTs, n=45; SR of OSs, n=23; SR of RCTs and OSs, n=15), followed by RCT (n=2) and OS (n=1). The PubMed search formulae are listed in online supplemental eTable 2.

Comparisons of original and subsequent studies

Table 2 shows the proportions of the confirmed studies. A total of 69% (59/86) (95% CI 58.1 to 77.5) of the original studies were confirmed in subsequent studies. Furthermore, of the 59 confirmed original studies, 16 were comparable to subsequent studies in terms of effect size. Among these 16, 13 were replicated and three reported effect sizes larger than the corresponding subsequent studies. Of these 16 studies, 11 compared SMDs. The median SMDs of the original and subsequent studies were 0.23 (0.18, 0.45) and 0.25 (0.15, 0.32), respectively (p=0.34, Wilcoxon signed-rank test). However, for the remaining 43 studies, strict comparisons of effect sizes were not possible because the outcomes were not fully matched between the original and subsequent studies. Details of the original and subsequent studies are presented in online supplemental eTable 3.

We conducted subgroup analyses on the proportions of confirmed studies for each research design in the original articles (online supplemental eTable 4). The proportions of confirmed OS and RCT studies (of which there was a relatively large number) were 61.3% (19/31) and 70.5% (31/44), respectively. Other designs included fewer studies, and we found no significant differences in the research design (p=0.74, χ^2 test). For the ICD-10 categories, the differences according to disease were not significant (p=0.67, χ^2 test). The proportion of confirmed studies cited in quality papers (56/88, 63.6%) was lower

Table 1 Characteristics of included newspaper stories

Newspaper	Country	Newspaper type	Newspaper stories that quoted 20 general and internal medicine journals	Newspaper stories that quoted 20 public, environmental and occupational health journals	Total
New York Times	USA	Quality	258	13	271
Washington Post	USA	Quality	279	22	301
Daily Telegraph	UK	Quality	28	5	33
Times	UK	Quality	191	18	209
USA Today	USA	Non-quality	122	11	133
Daily News	USA	Non-quality	65	7	72
Daily Mail	UK	Non-quality	173	9	182
Daily Mirror	UK	Non-quality	91	6	97
Total			1207	91	1298

than that in non-quality papers (31/44, 70.5%); however, the difference was not statistically significant (p=0.42, χ^2 test).

Example 1: contradicted

A prospective cohort study published in *BMJ* in 2000, covered by *Daily Mail*, suggested that drinking fluoridated water significantly reduced hip fractures.[16] Neither the subsequent matching study, meta-analysis of 14 observational studies, nor the original study[17] found any significant risk reduction in hip fractures.

Example 2: confirmed

One RCT published in the *JAMA* in 2000 and covered by the *Washington Post* suggested that sertraline was more effective than a placebo in patients with post-traumatic stress disorder (PTSD). The subsequent matching study was a meta-analysis comparing pharmacotherapies for PTSD, published in 2022.[18] In the subgroup analysis, which included the original RCT, sertraline was compared with placebo. The authors concluded that sertraline was effective. Therefore, the effectiveness reported in the original study was confirmed in a subsequent study. Furthermore, the point estimate of the original study's RR described in the subsequent study's forest plot was 0.70, and the point estimate and 95% CI of the RR of the new article was 0.68 (0.56 to 0.81). After calculating the SMD from these values, the original study had an SMD of 0.26, and the new study had a value of 0.27 (95% CI 0.15 to 0.40). We categorised this finding as not only 'confirmed' but also 'replicated'.

Example 3: unchallenged

Examples included in the unchallenged studies are as follows: Most studies have investigated unique interventions (eg, short nails for preventing infection, anti-digoxin fab for cardiac arrhythmia, horse chestnut seed extract for chronic venous insufficiency, beta-sheet breaker peptides for prion-related disorders, the Krukenberg procedure for double-hand amputees and yoga for carpal tunnel syndrome), and several studies have examined the effects of special drug use (eg, ondansetron for bulimia nervosa, growth hormone for Crohn's disease and combination therapy with old antidepressants, nefazodone and psychotherapy for chronic depression). However, these findings are difficult to validate using well-designed studies. The details are shown in online supplemental eTable 3.

DISCUSSION

This is the first study to examine a 20-year course of treatment or prevention recommended by newspaper articles in various medical fields. We selected newspaper stories that recommended certain treatments or preventions in 2000 and compared their results with those in the original research articles and compared the original studies with newer ones with better-controlled designs. Sixty-nine per cent (59/86) of the original studies were confirmed by subsequent studies. Among the confirmed studies, 13 of the 16 studies replicated both the direction and magnitude of the treatment effect. In studies in which the effects were confirmed, the effect sizes were relatively stable. However, the results of the remaining 43 studies were not comparable.

Table 2 Main analyses of the proportion of confirmed studies

	Total	Unchallenged	Contradicted	Confirmed	Proportion of confirmed studies, 95% CI (%)
Original studies	104*	18	27	59	68.6 (58.1 to 77.5)

*104 comparisons from 96 original studies (including duplicates).

As far as we know, few studies investigated the replicability of articles quoted in daily newspapers.[2 19] One is about attention deficit hyperactivity disorder studies, and the other is about risk factor studies; the proportions of 'confirmed' studies according to their definitions were 20% and 49%, respectively. The proportion of confirmed cases in our study (68.6%) was higher than those in these studies. The reasons for this may be as follows. Previous studies have not focused on treatment or prevention. Therefore, these proportions could not be compared. Furthermore, the definition of 'confirmed' in these studies was stricter than in our study. However, even in well-known newspapers, one-third of the stories may have been overturned by subsequent studies. Several studies have reported that the reporting standard in quality newspapers is significantly higher than that in non-quality papers.[9 20 21] In this study, the proportion of confirmed studies in quality newspapers was slightly lower than that in non-quality newspapers; however, this difference was not statistically significant. There may not be much of a difference between highly circulated quality papers and low-quality papers.

This study had some limitations. First, newspaper story authors often do not provide details about their information sources. It is often claimed that the best journalists are those with the most sources'.[22] In these cases, we could not find any articles quoted in newspapers. Therefore, for convenience, we used the journal names as search words. Consequently, only better-quality newspaper stories, in which journal names were written, were included. This may have led to the discovery of higher quality stories. Consequently, the proportion of quoted RCT may be higher than that of other standard newspaper stories. The credibility of studies cited in newspaper articles that do not list the sources of citations remains unclear. Second, an increasing number of SRs have been published in recent years, and several similar SRs can often be found on any research topic. Therefore, it is difficult to select the most appropriate option. To find the optimal subsequent study, two independent researchers checked the full paper and selected the best study from among several candidates. This reduced the number of arbitrary choices as much as possible. Third, we assumed that most subsequent study designs would be SR. Therefore, we searched the Web of Science for new studies that cited the original paper, and compared them with the effect sizes shown in the forest plot. However, the authors of subsequent SRs did not always cite the original articles for various reasons (eg, subtle differences in the type of outcome or timing of measurement). If cited, they were excluded from forest plots. Only 11 studies compared SMDs and 43 studies, although found to be effective, were unable to compare effect sizes. It is possible that the original studies reported a very large effect size, while the subsequent studies were only marginally significant. Based on these results, it is impossible to determine whether the SMDs are stable. Future studies should rigorously compare effect sizes by aligning outcomes. Fourth, 18 unchallenged studies focused on unique topics. Our definition of primary outcome excluded these numbers from the denominator, which makes the proportion of confirmed studies appear higher than it is. If these were included in the denominator, the proportion of confirmed cases would have been much lower.

However, this study has several strengths. This is the first study to examine the veracity of newspaper stories on treatment and prevention in various medical fields. Second, we followed up on each treatment over a 20-year period and took relevant subsequent studies with stronger designs as the gold standard. Although we cannot rule out the possibility that the results of subsequent studies may be reversed in the future, we believe that the results obtained over the past 20 years are generally robust. Third, to find the most appropriate subsequent study, we reviewed and discussed many SRs using the Web of Science and PubMed. We spent a lot of time carefully going through this process to make sure we did not miss any relevant papers.

CONCLUSION

The results for clinical research articles were relatively stable for papers in which the citation source was properly listed in newspaper articles. Journalists should provide information on the source studies to enable researchers to identify them. However, the results of approximately one-third of these studies were overturned over the following two decades. Readers should be aware that more than a few claims made in highly circulated newspapers based on high-profile journal articles may be overturned in subsequent studies.

Author affiliations
[1]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan
[2]Oku Medical Clinic, Osaka, Japan
[3]Department of Advanced Medicine for Rheumatic diseases, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4]Department of Emergency Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan
[5]Department of Human Health Science, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[6]Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

Contributors AT and TAF conceived the study. AT and TAF designed the study. AT, YaT, AO, YuT, SF, YO, NT, YH and NI did the literature search and extracted the data. AT did the analyses. AT and TAF wrote the first draft of the manuscript. All authors contributed to the interpretation of the findings and subsequent edits of the manuscript. TAF provided overall supervision to the project. AT is the guarantor and accepts full responsibility for the work and the conduct of the study, had access to the data and controlled the decision to publish.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests AT received lecture fees from Sumitomo Dainippon Pharma, Eisai, Janssen Pharmaceutical, Meiji-Seika Pharma, Mitsubishi Tanabe Pharma, Otsuka and Takeda Pharmaceutical. AO received research grants and/or speaker fees from Pfizer, Bristol-Myers Squibb, Advantest, Asahi Kasei Pharma, Chugai Pharmaceutical, Eli Lilly Japan K. K., Ono Pharmaceutical Co., UCB Japan Co., Mitsubishi Tanabe Pharma Co., Eisai, Abbvie, Takeda Pharmaceutical, and Daiichi

Sankyo. SF received a research grant from JSPS KAKENHI (Grant Number JP 20K18964), the KDDI Foundation and the Pfizer Health Research Foundation. TAF reports personal fees from Boehringer-Ingelheim, DT Axis, Kyoto University Original, Shionogi, and SONY, and a grant from Shionogi outside the submitted work; TAF has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iDs
Aran Tajika http://orcid.org/0000-0003-3926-8867
Yusuke Tsutsumi http://orcid.org/0000-0002-9160-0241
Satoshi Funada http://orcid.org/0000-0002-8925-2348
Toshi A Furukawa http://orcid.org/0000-0003-2159-3776

REFERENCES

1 Woloshin S, Schwartz LM, Kramer BS. Promoting healthy skepticism in the news: helping journalists get it right. *J Natl Cancer Inst* 2009;101:1596–9.
2 Gonon F, Konsman J-P, Cohen D, *et al*. Why most biomedical findings echoed by newspapers turn out to be false: the case of attention deficit hyperactivity disorder. *PLoS One* 2012;7:e44275.
3 Selvaraj S, Borkar DS, Prasad V, *et al*. Media coverage of medical journals: do the best articles make the news *PLoS ONE* 2014;9:e85355.
4 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
5 Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218.
6 Pereira TV, Horwitz RI, Ioannidis JPA. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 2012;308:1676.
7 Circulations ABO. n.d. Available: http://www.auditbureau.org
8 Media AfA. n.d. Available: http://www.auditedmedia.com
9 Wilson A, Bonevski B, Jones A, *et al*. Media reporting of health interventions: signs of improvement, but major problems persist. *PLoS One* 2009;4:e4831.
10 Bell L, Seale C. The reporting of cervical cancer in the mass media: a study of UK newspapers. *Eur J Cancer Care (Engl)* 2011;20:389–94.
11 Dean M. Tabloid campaign forces UK to reconsider sex-offence laws. *Lancet* 2000;356:745.
12 Lexisnexis. n.d. Available: https://www.lexisnexis.com/en-us/home.page
13 Tajika A, Ogawa Y, Takeshima N, *et al*. Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *Br J Psychiatry* 2015;207:357–62.
14 Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19:3127–31.
15 Cohen J. *Statistical power analysis in the behavioral science*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
16 Phipps KR, Orwoll ES, Mason JD, *et al*. Community water fluoridation, bone mineral density, and fractures: prospective study of effects in older women. *BMJ* 2000;321:860–4.
17 Yin X-H, Huang G-L, Lin D-R, *et al*. Exposure to fluoride in drinking water and hip fracture risk: a meta-analysis of observational studies. *PLoS One* 2015;10:e0126488.
18 Williams T, Phillips NJ, Stein DJ, *et al*. Pharmacotherapy for post traumatic stress disorder (PTSD). *Cochrane Database Syst Rev* 2022;3:CD002795.
19 Dumas-Mallet E, Smith A, Boraud T, *et al*. Poor replication validity of biomedical association studies reported by newspapers. *PLoS One* 2017;12:e0172650.
20 Cooper BEJ, Lee WE, Goldacre BM, *et al*. The quality of the evidence for dietary advice given in UK national newspapers. *Public Underst Sci* 2012;21:664–73.
21 Robinson A, Coutinho A, Bryden A, *et al*. Analysis of health stories in daily newspapers in the UK. *Public Health* 2013;127:39–45.
22 Weitkamp E. British newspapers privilege health and medicine topics over other science news. *Public Relations Review* 2003;29:321–33.

THERAPEUTIC ADVANCES in
Medical Oncology

Original Research

Prospective exosome-focused translational research for afatinib (EXTRA) study of patients with nonsmall cell lung cancer harboring *EGFR* mutation: an observational clinical study

Ther Adv Med Oncol

2023, Vol. 15: 1–14

DOI: 10.1177/17588359231177021

Saori Takata, Kei Morikawa, Hisashi Tanaka, Hidetoshi Itani, Masashi Ishihara, Kazuya Horiuchi, Yasuhiro Kato, Shinnosuke Ikemura, Hideyuki Nakagawa, Yoshiro Nakahara, Yoshitaka Seki, Akihiro Bessho, Nobumasa Takahashi, Kentaro Hayashi, Takeo Endo, Kiyoshi Takeyama, Toshiya Maekura, Nagio Takigawa, Akikazu Kawase, Makoto Endoh, Kenji Nemoto, Kazuma Kishi, Kenzo Soejima, Yusuke Okuma, Kenichi Yoshimura, Daisuke Saigusa, Yae Kanai, Koji Ueda, Akira Togashi, Noriyuki Matsutani and Nobuhiko Seki

Correspondence to:
Nobuhiko Seki
Division of Medical Oncology, Department of Internal Medicine, Teikyo University School of Medicine, 2-11-1 Kaga, Itabashi-ku, Tokyo 173-8605, Japan.
nseki@med.teikyo-u.ac.jp

Saori Takata
Department of Respiratory Medicine, Kyorin University School of Medicine, Tokyo, Japan

Kei Morikawa
Division of Respiratory Medicine, Department of Internal Medicine, St. Marianna University School of Medicine, Kanagawa, Japan

Hisashi Tanaka
Department of Respiratory Medicine, Hirosaki University Graduate School of Medicine, Aomori, Japan

Hidetoshi Itani
Department of Respiratory Medicine, Ise Red Cross Hospital, Mie, Japan

Masashi Ishihara
Division of Medical Oncology, Department of Internal Medicine, Teikyo University School of Medicine, Tokyo, Japan

Kazuya Horiuchi
Respiratory Disease Center, Showa University Northern Yokohama Hospital, Kanagawa, Japan

Yasuhiro Kato
Yusuke Okuma
Department of Thoracic Oncology and Respiratory Medicine, Tokyo Metropolitan Cancer and Infectious Diseases Center, Komagome Hospital, Tokyo, Japan

Shinnosuke Ikemura
Kenzo Soejima
Clinical and Translational Research Center, Keio University Hospital, Tokyo, Japan

Abstract

Background: The exosome-focused translational research for afatinib (EXTRA) study is the first trial to identify novel predictive biomarkers for longer treatment efficacy of afatinib in patients with epidermal growth factor receptor (*EGFR*) mutation-positive nonsmall cell lung cancer (NSCLC) via a comprehensive association study using genomic, proteomic, epigenomic, and metabolomic analyses.

Objectives: We report details of the clinical portion prior to omics analyses.

Design: A prospective, single-arm, observational study was conducted using afatinib 40 mg/day as an initial dose in untreated patients with *EGFR* mutation-positive NSCLC. Dose reduction to 20 mg every other day was allowed.

Methods: Progression-free survival (PFS), overall survival (OS), and adverse events (AEs) were evaluated.

Results: A total of 103 patients (median age 70 years, range 42–88 years) were enrolled from 21 institutions in Japan between February 2017 and March 2018. After a median follow-up of 35.0 months, 21% remained on afatinib treatment, whereas 9% had discontinued treatment because of AEs. The median PFS was 18.4 months, with a 3-year PFS rate of 23.3%. The median afatinib treatment duration in patients with final doses of 40 ($n = 27$), 30 ($n = 23$), and 20 mg/day ($n = 35$), and 20 mg every other day ($n = 18$) were 13.4, 15.4, 18.8, and 18.3 months, respectively. The median OS was not reached, with a 3-year OS rate of 58.5%. The median OS in patients who did ($n = 25$) and did not ($n = 78$) receive osimertinib during the entire course of treatment were 42.4 months and not reached, respectively ($p = 0.654$).

Conclusions: As the largest prospective study in Japan, this study confirmed favorable OS following first-line afatinib in patients with *EGFR* mutation-positive NSCLC in a real-world setting. Further analysis of the EXTRA study is expected to identify novel predictive biomarkers for afatinib.

Trial registration: UMIN-CTR identifier (UMIN000024935, https://center6.umin.ac.jp/cgi-open-bin/ctr/ctr_his_list.cgi?recptno=R000028688

Keywords: afatinib, biomarker, EGFR–TKI, exosome, nonsmall-cell lung cancer, omics

Received: 10 February 2023; revised manuscript accepted: 2 May 2023.

Hideyuki Nakagawa
Department of Respiratory Medicine, National Hospital Organization Hirosaki Hospital, Aomori, Japan

Yoshiro Nakahara
Department of Respiratory Medicine, Kitasato University School of Medicine, Kanagawa, Japan

Yoshitaka Seki
Department of Internal Medicine, The Jikei University Daisan Hospital, Tokyo, Japan

Akihiro Bessho
Department of Respiratory Medicine, Japanese Red Cross Okayama Hospital, Okayama, Japan

Nobumasa Takahashi
Department of General Thoracic Surgery, Saitama Cardiovascular and Respiratory Center, Saitama, Japan

Kentaro Hayashi
Division of Respiratory Medicine, Department of Internal Medicine, Nihon University School of Medicine, Tokyo, Japan

Takeo Endo
Department of Respiratory Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan

Kiyoshi Takeyama
Department of Respiratory Medicine, Tokyo Women's Medical University School of Medicine, Tokyo, Japan

Toshiya Maekura
Department of Respiratory Medicine, Hoshigaoka Medical Center, Osaka, Japan

Nagio Takigawa
Department of General Internal Medicine 4, Kawasaki Medical School, Okayama, Japan

Akikazu Kawase
First Department of Surgery, Hamamatsu University School of Medicine, Shizuoka, Japan

Makoto Endoh
Department of Thoracic Surgery, Yamagata Prefectural Central Hospital, Yamagata, Japan

Kenji Nemoto
Department of Respiratory Medicine, National Hospital Organization, Ibarakihigashi National Hospital, Ibaraki, Japan

Introduction

Recent advances in epidermal growth factor receptor (EGFR)-targeted therapy for nonsmall cell lung cancer (NSCLC) have improved survival in precision medicine. As a result, three generations of EGFR tyrosine kinase inhibitors (TKIs) have been approved in Japan as first-line treatments for patients with *EGFR* mutation-positive NSCLC: first-generation reversible TKIs (erlotinib and gefitinib), second-generation irreversible TKIs (afatinib and dacomitinib), and third-generation mutant-selective TKIs (osimertinib).[1] The FLAURA phase III study recently demonstrated significantly prolonged overall survival (OS) in patients with *EGFR* mutation-positive NSCLC treated with first-line osimertinib ($n = 279$) compared with first-generation EGFR–TKIs [gefitinib ($n = 183$) or erlotinib ($n = 94$)].[2] Osimertinib has thus been established as the standard treatment for previously untreated common *EGFR* mutation-positive NSCLC. However, its efficacy in the FLAURA study was not definitive, with hazard ratios (HRs) for OS of 1.00 [95% confidence interval (CI): 0.75–1.32] and 1.00 (95% CI: 0.71–1.40) in Asian and *EGFR* L858R-mutated patients, respectively, suggesting limited benefit of osimertinib over first-generation EGFR–TKIs in these subgroups.[2] Considering that race and *EGFR* mutation subtypes were just two stratified factors in the randomization of the FLAURA study, the results could indicate that these two factors had independent negative impacts on the clinical benefit of osimertinib.

Regarding first-line afatinib, the LUX-Lung 7 phase IIb study showed a trend toward better OS in patients treated with afatinib ($n = 160$) compared with gefitinib ($n = 159$).[3] Furthermore, the retrospective Gio-Tag study, which included real-world clinical patients treated with first-line afatinib followed by second-line osimertinib, found that the median duration of sequential afatinib and osimertinib treatment was 37.1 months, and the median OS was 44.8 months in Asian patients ($n = 50$), compared with 27.6 and 36.7 months, respectively, in non-Asians ($n = 137$).[4] Similarly, the Up-SwinG study, which had a similar study design, found a median duration for sequential afatinib and osimertinib treatment of 28.8 months and median OS of 42.3 months in Asian patients ($n = 118$), compared with 25.5 and 31.3 months, respectively, in non-Asians ($n = 73$).[5] These data thus indicated that first-line afatinib followed by second-line osimertinib might prolong the total duration of EGFR–TKI therapy, especially in Asian patients, thus improving OS. However, the efficacy of first-line afatinib, like other EGFR–TKIs, varies, with some patients benefiting from long-term efficacy while others do not. There is thus a need to identify biomarkers for afatinib efficacy in actual clinical settings.[6,7]

As previously reported, the EXTRA (EXosome-focused Translational Research for Afatinib) study protocol aims to explore novel biomarkers for afatinib efficacy by matching data from multi-omics analyses of peripheral blood samples in patients treated with first-line afatinib to clinical efficacy data (Supplemental Figure S1).[8] We conducted a prognostic survey and locked the clinical data 3 years after the final enrollment in this trial, and then started to carry out proteomic, genomic, metabolomic, and epigenomic analyses. Prior to the publication of the results of these four omics analyses, the current study aimed to analyze the clinical efficacy data based on first-line afatinib treatment. This report may be considered to reflect the latest real-world data on first-line afatinib, which has been used in Japan since 2014.

Patients and methods

Study design

The EXTRA study was designed as a prospective, single-arm, observational study to identify novel predictive biomarkers associated with longer OS in patients treated with first-line afatinib, via comprehensive genomic, proteomic, epigenomic, and metabolomic association analyses using serial peripheral blood samples (free molecules in serum/plasma and exosome-packaged molecules) (Supplemental Figure S1).[8] We planned to enroll 60 patients as the discovery cohort and 40 patients as the independent validation cohort.

Patient eligibility

The main inclusion criteria for registration were: age ⩾20 years; histologically or cytologically confirmed metastatic or locally advanced NSCLC; *EGFR* mutation (common or uncommon); Eastern Cooperative Oncology Group performance status (PS) of 0 or 1; adequate bone marrow, renal, and liver functions; and chemotherapy-naive. The main exclusion criteria were: interstitial pneumonia or pulmonary fibrosis; active infection or uncontrolled disease; and other active malignant disease.

Study treatment

Enrolled patients were initially treated with afatinib 40 mg/day, and the dose was adjusted according to toxicities observed by the investigators. Patients who developed drug-related grade ⩾2 adverse events (AEs) temporarily discontinued afatinib until recovery to grade 1, and then resumed afatinib treatment with a 10 mg dose reduction. The dose could be reduced by a further 10 mg in patients who developed drug-related grade ⩾2 AEs again despite the initial dose decrease. A total of three dose reductions were allowed, with a minimum dose of afatinib of 20 mg every other day.

Treatment was discontinued in patients who developed afatinib-induced grade ⩾1 interstitial lung disease and in patients who required a fourth dose reduction of afatinib. Treatment was continued until disease progression, unacceptable toxicity, or withdrawal of consent by the patient.

Assessment

Tumor response was assessed by thoracoabdominal and head computed tomography or head magnetic resonance imaging. Tumor assessment was performed every 8 weeks for the first 24 weeks of treatment and every 12 weeks thereafter until progressive disease (PD), treatment discontinuation, withdrawal of consent, or death, with the date of treatment initiation defined as the reference date. The tumor response was evaluated according to RECIST, version 1.1.

AEs were classified by the Medical Dictionary for Regulatory Activities, and their severities were assessed according to the Common Terminology Criteria for Adverse Events, version 4.0.

Statistical analysis

The primary endpoint was the identification of novel predictive biomarkers of afatinib efficacy associated with longer OS. The secondary endpoints were the following clinical indicators to be matched with the generated omics data: objective response rate (ORR), disease control rate (DCR), progression-free survival (PFS), OS, and AEs.

ORR was defined as the percentage of patients who had a complete or partial radiological response. DCR was defined as the percentage of patients who had a complete or partial radiological response, or stable disease. PFS was defined as the time from each registration to confirmation of PD or death from any cause. OS was defined as the time from the registration to death from any cause.

The 95% CIs for the proportions of ORR and DCR were calculated using the Clopper–Pearson method. Median PFS and OS and their 95% CIs were estimated using the Kaplan–Meier method. Between-group comparisons were performed using log-rank tests. The analyses were carried out using SAS ver. 9.4 (SAS Institute Inc., Cary, NC, USA).

Ethics

This study complied with all the principles in the Declaration of Helsinki (as revised in 2013), and was approved by the Ethical Review Board for Medical and Health Research Involving Human Subjects at Teikyo University (Approval No. 16-066). All enrolled patients provided written informed consent. This trial was registered with the University Hospital Medical Information Network clinical trial registry (No. UMIN000024935).

Results

Patient characteristics

A total of 103 patients (60 patients in the discovery cohort, 43 patients in the validation cohort) were enrolled from 21 institutions in Japan between February 25, 2017, and March 30, 2018. The patient characteristics are shown in Table 1. The median age was 70 years, with 32% aged ⩾75 years, 74% were female, and 50% were PS 1. About a quarter of patients (27%) had postsurgery recurrence, about a fifth (22%) had brain metastasis before afatinib treatment, and all patients (100%) had adenocarcinoma. Most patients (90%) had common *EGFR* mutations (exon 19 deletion and exon 21 L858R), and the other 10% had uncommon *EGFR* mutations.

Patient flow

Patient flow is summarized in Figure 1. Among all 103 enrolled patients, treatment was discontinued in 81 patients after a median follow-up of 35.0 months (range: 0.5–44.4). The reasons for discontinuation were: PD in 70 patients (68%); AEs in nine patients (9%); and physician's decision in two patients (2%), including cognitive impairment in one patient and new onset of

Kazuma Kishi
Department of Respiratory Medicine, Respiratory Center, Toranomon Hospital, Tokyo, Japan

Yusuke Okuma
Department of Thoracic Oncology and Respiratory Medicine, Tokyo Metropolitan Cancer and Infectious Diseases Center, Komagome Hospital, Tokyo, Japan

Kenichi Yoshimura
Department of Biostatistics, Medical Center for Translational and Clinical Research, Hiroshima University Hospital, Hiroshima, Japan

Daisuke Saigusa
Laboratory of Biomedical and Analytical Sciences, Faculty of Pharma-Science, Teikyo University, Tokyo, Japan

Yae Kanai
Department of Pathology, Keio University School of Medicine, Tokyo, Japan

Koji Ueda
Project for Realization of Personalized Cancer Medicine, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan

Akira Togashi
Nippon Boehringer Ingelheim Co., Ltd., Tokyo, Japan

Noriyuki Matsutani
Department of Surgery, Teikyo University Hospital, Mizonokuchi, Kanagawa, Japan

Table 1. Patient characteristics.

Characteristic	No. of patients	%
Total	103	100
Age (years)		
Median (range)	70 (42–88)	
<70	50	49
70–74	21	20
75–79	16	16
⩾80	16	16
Sex		
Male	27	26
Female	76	74
PS		
0	52	50
1	51	50
Stage		
IIIB	3	3
IV	72	70
Postsurgery recurrence[a]	28	27
Brain metastasis		
Present	23	22
Absent	80	78
Histology		
Adenocarcinoma	103	100
EGFR mutation		
Exon 19 deletion	52	50
Exon 21 L858R	41	40
Others	10	10
Exon 18	3	3
Exon 20 insertion	3	3
Exon 20 T790M	1	1
Exon 21 L861Q	1	1
Exon 20 S768I + Exon 18 G719X	2	2

[a]These patients were not amenable to local therapy.
EGFR, epidermal growth factor receptor; PS, performance status.

thyroid cancer in one patient. A total of 22 patients (21%) finally remained on afatinib treatment.

Treatment efficacy

Tumor responses are summarized in Supplemental Table S1. The ORR and DCR were 60.2% (95% CI: 50.1–69.7) and 87.4% (95% CI: 79.4–93.1), respectively.

The Kaplan–Meier curve of PFS and a forest plot of median PFS are presented in Figure 2(a) and (b), respectively. The median PFS was 18.4 months (95% CI: 13.8–22.1), with a 3-year PFS rate of 23.3%. Subgroup analyses of median PFS indicated trends toward a longer PFS in patients with PS 0 (25.0 months, 95% CI: 18.8–28.4) *versus* 1 (13.6 months, 95% CI: 9.3–16.4), patients with postsurgery recurrence [27.7 months, 95% CI: 18.8—not calculable (NC)] *versus* stage IIIB/IV (15.4 months, 95% CI: 12.2–20.2), patients without brain metastasis (20.6 months, 95% CI: 15.4–24.7) *versus* those with brain metastasis (13.8 months, 95% CI: 8.4–18.0), and patients with *EGFR* exon 19 deletion mutation (21.2 months, 95% CI: 15.4–24.8) *versus* uncommon *EGFR* mutations (14.3 months, 95% CI: 0.3–31.0).

The Kaplan–Meier curve of OS is presented in Figure 2(c). The median OS was not reached (95% CI: 34.9—NC), with a 3-year OS rate of 58.5%. Subgroup analyses of median OS could, therefore, only be calculated for patients aged ⩾75 years (42.4 months, 95% CI: 24.8—NC), patients with PS 1 (31.6 months, 95% CI: 24.8–42.4), stage IIIB/IV (35.4 months, 95% CI: 31.6—NC), brain metastasis (32.2 months, 95% CI: 17.1—NC), and patients with uncommon *EGFR* mutations (34.9 months, 95% CI: 9.8—NC).

Toxicity analysis

All AEs are summarized in Table 2. Among 103 patients, grade 3, grade 4, and grade 5 AEs occurred in 21 (20%), 2 (2%), and 1 (1%) patients, respectively. The most frequent grade ⩾3 AEs were diarrhea in 12 patients (12%), anorexia in 8 patients (8%), and rash acneiform in 6 patients (6%). In addition, pneumonitis was observed in three patients (3%), comprising one case each of grade 2 (1%), grade 3 (1%), and grade 5 (1%).

Nine patients discontinued afatinib treatment because of AEs, comprising three discontinuations due to pneumonitis (one each grade 2, grade 3, and grade 5), rash acneiform in three patients (grade 2 in two patients and grade 4 in one patient), diarrhea in two patients (grade 1 in each patient), and anorexia in one patient (grade 4).

Treatment duration according to final dose

The treatment duration according to the final dose of afatinib is presented in Figure 3. The median afatinib treatment durations in patients with final doses of 40 ($n=27$), 30 ($n=23$), and 20 mg/day ($n=35$), and 20 mg every other day ($n=18$) were 13.4, 15.4, 18.8, and 18.3 months, respectively.

Regarding the nine patients who discontinued treatment because of AEs, five patients were in the group with a final dose of 40 mg/day, and no patient was in the group with a final dose of 20 mg every other day. In contrast, the 22 patients who remained on afatinib treatment were distributed equally among the four dosage groups.

Poststudy treatment

Eighty-one patients discontinued treatment with afatinib during the follow-up period (Figure 1), including 62 patients who received poststudy treatment (77%) (Supplemental Table S2) and 53 patients who underwent re-biopsy (65%) before second-line treatment, resulting in the detection of *EGFR* T790M mutation in 16/53 patients (30%).

A total of 25 of the 81 patients (31%) received osimertinib as poststudy treatment. Osimertinib was administered as second-line therapy in 19 patients (23%), comprising 11 patients with *EGFR* T790M mutation-positive status and 8 patients with *EGFR* T790M mutation-unknown status. Similarly, osimertinib was administered after second-line therapy in six patients (7%), comprising three patients with *EGFR* T790M mutation-positive status and three patients with *EGFR* T790M mutation-unknown status.

Impact of osimertinib on OS

The Kaplan–Meier curves of OS for patients who did ($n=25$) and did not ($n=78$) use osimertinib during the entire course of treatment are presented in Figure 4. The median OS was

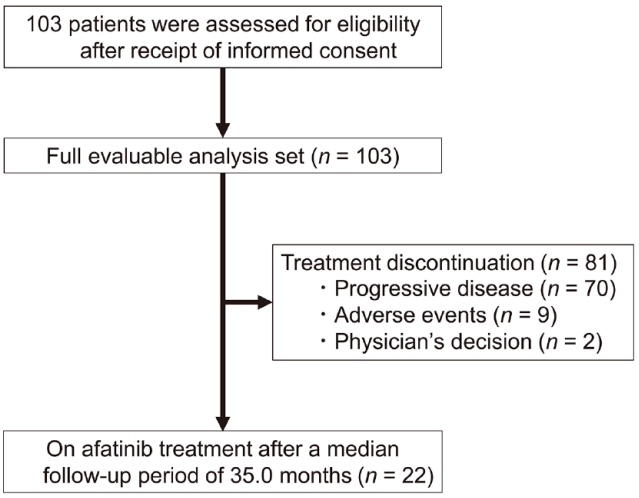


Figure 1. Patient flow in the EXTRA study.
EXTRA, exosome-focused translational research for afatinib.

42.4 months (95% CI: 30.1—NC) and not reached (95% CI: 34.2—NC), respectively, resulting in no significant difference between the groups (log-rank test, $p=0.654$).

The median treatment durations of osimertinib in all 25 patients, 14 patients with *EGFR* T790M mutation-positive status, and 11 patients with *EGFR* T790M mutation-unknown status were 7.9 months (95% CI: 5.3–10.5), 8.0 months (95% CI: 5.1–10.9), and 7.8 months (95% CI: 2.9–12.7), respectively.

Discussion

We are currently conducting genomic, proteomic, epigenomic, and metabolomic analyses of peripheral blood samples (free molecules in serum/plasma and exosome-packaged molecules) collected from patients before, during, and after treatment in the EXTRA study. Furthermore, a comprehensive association study based on the ORR, DCR, PFS, OS, and AEs reported here is also in progress. The results for these clinical indicators, based on a sufficient observation period of 35.0 months in a clinical study of patients with advanced NSCLC, are considered to reflect the latest real-world data for first-line afatinib, which has been used in Japan since 2014.

In the EXTRA study, the median PFS in patients receiving afatinib was 18.4 months (95% CI: 13.8–22.1), and subgroup analyses indicated trends toward longer PFS for patients with PS 0 (25.0 months, 95% CI: 18.8–28.4), postsurgery

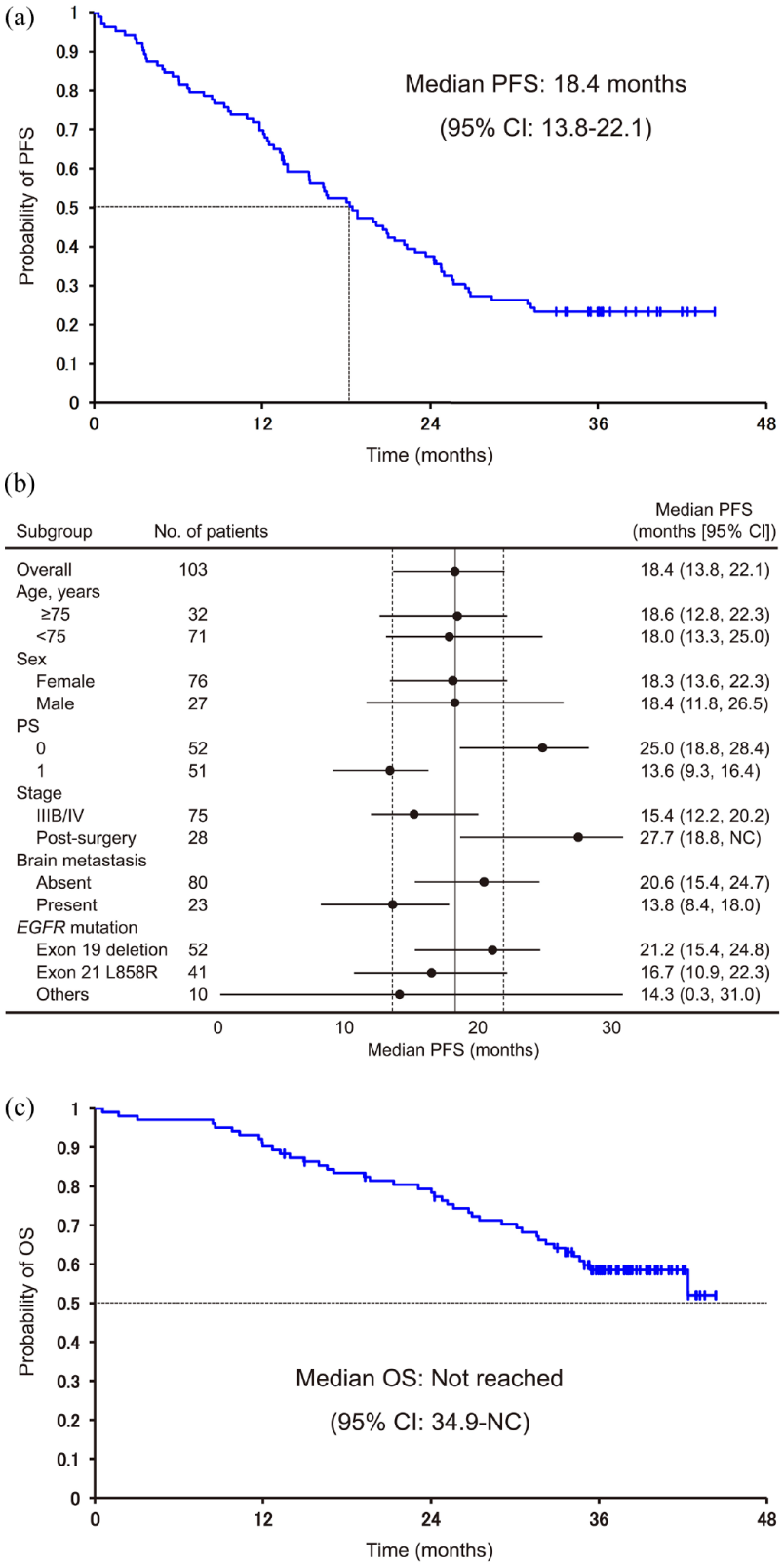


Figure 2. Kaplan–Meier curve of PFS (a), forest plot of median PFS (b), and Kaplan–Meier curve of OS (c) in patients treated with afatinib.
CI, confidence interval; EGFR, epidermal growth factor receptor; NC, not calculable; OS, overall survival; PFS, progression-free survival; PS, performance status.

Table 2. Summary of all AEs.

	No. of patients ($n = 103$)					
AE	**All grades**	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**
Total	101	36	41	21	2	1
Rash acneiform	68	44	18	5	1	0
Diarrhea	63	36	15	11	1	0
Paronychia	51	30	21	0	0	0
Mucositis oral	41	26	13	2	0	0
Anorexia	20	11	1	7	1	0
ALT increased	13	12	0	0	1	0
AST increased	9	8	0	0	1	0
Dysgeusia	5	5	0	0	0	0
Vomiting	4	1	1	2	0	0
Anemia	3	2	0	1	0	0
Pneumonitis	3	0	1	1	0	1
Creatinine increased	2	1	1	0	0	0
Conjunctivitis	2	2	0	0	0	0
Leukopenia	1	1	0	0	0	0
Rhinitis	1	1	0	0	0	0
Nasal bleeding	1	1	0	0	0	0
Constipation	1	1	0	0	0	0
Deep vein thrombosis	1	0	0	1	0	0
Hyponatremia	1	0	0	1	0	0

AE, adverse event; ALT, alanine aminotransferase; AST, aspartate transaminase.

recurrence (27.7 months, 95% CI: 18.8—NC), without brain metastasis (20.6 months, 95% CI: 15.4–24.7), and with *EGFR* exon 19 deletion mutation (21.2 months, 95% CI: 15.4–24.8). Historically, the median PFS of 18.4 months (95% CI: 13.8–22.1) in the EXTRA study seemed to be better than the median PFS reported in the LUX-Lung 3 study [$n = 230$; 11.1 months (95% CI: unpublished)] and its Japanese subset [$n = 54$; 13.8 months (95% CI: 11.0–19.1)], the LUX-Lung 6 study [$n = 242$; 11.0 months (95% CI: 9.7–13.7)], and the LUX-Lung 7 study [$n = 160$; 11.0 months (95% CI: 10.6–12.9)].[9–12] The potentially better PFS in the EXTRA study compared with these previous studies might be attributable to the inclusion of patients with postsurgery recurrence, while the LUX-Lung 3, 6, and 7 studies only included stage IIIB/IV patients. Indeed, about a quarter of patients (27%) in the EXTRA study had postsurgery recurrence and demonstrated a median PFS of 27.7 months (95% CI: 18.8—NC); however, even if the analysis was limited to stage IIIB/IV patients, the median PFS was 15.4 months (95% CI: 12.2–20.2), which still seemed better than in the previous studies. Given that there was little difference in other patient characteristics, including PS, brain metastasis, and *EGFR* mutation status, the potentially better PFS in stage IIIB/IV patients in the EXTRA study may be related to the

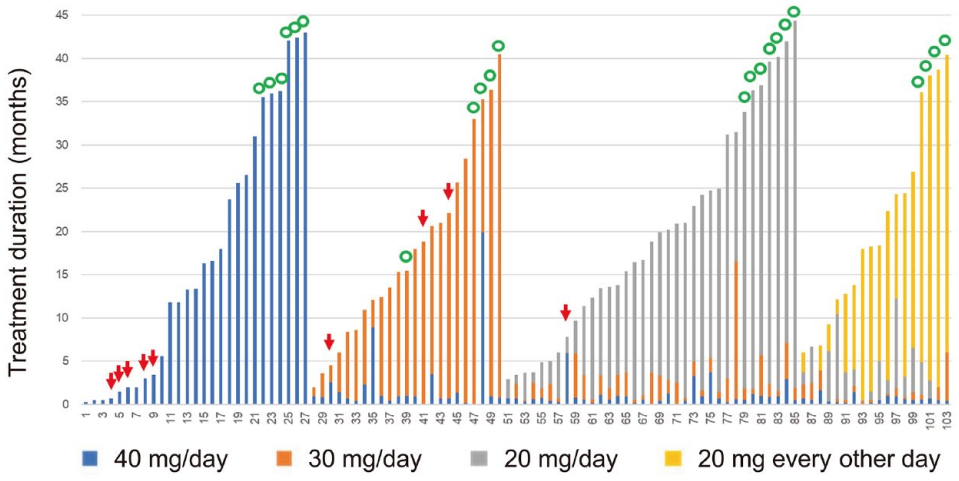


Figure 3. Treatment duration by the final reduced dosage of afatinib.
Circles represent patients remaining on treatment; arrows represent treatment discontinuation due to AEs.
AE, adverse event.

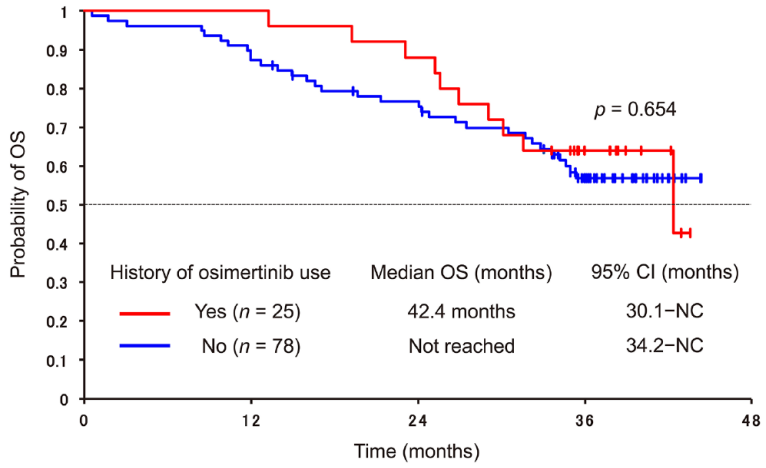


Figure 4. Kaplan–Meier curves of OS in patients who did and did not receive osimertinib during the course of treatment.
CI, confidence interval; NC, not calculable; OS, overall survival.

difference in rates of treatment discontinuation due to AEs.[9–12]

Differences in the body-surface area (BSA) and/or liver metabolic functions mean that the incidence of AEs following administration of a fixed-dose EGFR–TKI is generally higher in Japanese compared with Western patients.[13] Notably, a BSA $\leqslant 1.7\,m^2$ was significantly associated with severe afatinib-related AEs,[14] although the strategy of using a fixed-dose of afatinib was decided based on the results of the phase I study, demonstrating only a weak correlation between total body clearance of afatinib and BSA ($r^2 = 0.06$).[15] The rate of afatinib discontinuation due to AEs in the Japanese subset of the LUX-Lung 3 study was 19%, compared with 8%, 6%, and 6% in the global LUX-Lung 3, 6, and 7 studies, respectively.[9–12] A similar trend was observed in the FLAURA study, with osimertinib discontinuation rates of 26% in Japanese patients ($n = 65$) *versus* 13% in global patients ($n = 279$), and discontinuation rates of first-generation EGFR–TKIs of 35% in Japanese patients ($n = 55$) *versus* 18% in global patients ($n = 277$).[16,17] As a result, the actual median treatment durations in Japanese

patients in the osimertinib (15.0 months) and first-generation EGFR–TKI groups (10.3 months) were shorter compared with the median PFS values in the two groups (19.1 and 13.8 months, respectively), whereas the actual median treatment durations in all patients in the osimertinib (20.7 months) and first-generation EGFR–TKI groups (11.5 months) were almost the same as the median PFS values in the respective groups (18.9 and 10.2 months, respectively).[16–18] Therefore, if the rate of treatment discontinuation due to AEs in Japanese patients in the FLAURA study had been lower, their median PFS might have been much better. Given these results, the fact that the rate of treatment discontinuation due to AEs (9%) in the EXTRA study was about half of that (19%) in the Japanese subset of the LUX-Lung 3 study is presumed to be one factor responsible for the favorable PFS in the EXTRA study.

The reason for the low rate of treatment discontinuation due to AEs in the EXTRA study, despite the Japanese ethnicity, might be that the minimum dose of afatinib specified in the LUX-Lung 3, 6, and 7 study protocols was 20 mg/day, compared with 20 mg every other day in the EXTRA study. A growing body of Japanese evidence supports this hypothesis. At least six phase II studies of first-line afatinib in patients with *EGFR* mutation-positive NSCLC have been conducted in Japan, including three studies ($n = 30$, 40, and 38, respectively) with a protocol-specified minimum dose of afatinib of 20 mg/day,[19–21] and three studies ($n = 53$, 46, and 35, respectively) with 20 mg every other day.[22–24] The rate of treatment discontinuation due to AEs was lower in the latter (8, 11, and 11%, respectively) compared with the former studies (17, 20, and 21%, respectively). Furthermore, the median PFS seemed better in the latter (12.6, 15.2, and 15.6 months, respectively) than in the former studies (11.8, 12.9, and 14.2 months, respectively), although the patients' characteristics were not necessarily the same in all studies. These results suggest that a dose of 20 mg every other day may be more appropriate for some Japanese patients than 20 mg/day in terms of tolerability and preserved efficacy. In addition, plasma afatinib concentration was shown to be positively correlated with grade 3 AEs and negatively with BSA but not with treatment duration in Japanese patients.[24] We therefore agree with previous reports indicating that tolerability-guided dose reduction of afatinib had no impact on treatment duration.[25,26] Most patients in the EXTRA study with a final dose of 20 mg every other day showed a durable response, with tolerability-guided dose reduction to 20 mg every other day within several months after initiating a dose of 20 mg/day.

Considering the use of EGFR–TKI monotherapy in terms of OS, the HR of osimertinib over first-generation EGFR–TKIs in the FLAURA study was 1.00 (95% CI: 0.75–1.32) in Asian and 1.39 (95% CI: 0.83–2.34) in Japanese patients; although the latter result was from an exploratory posthoc analysis, suggesting that EGFR–TKIs other than osimertinib may also be a treatment option, especially in Japanese patients.[2,18] Notably, the Kaplan–Meier OS curve for first-generation EGFR–TKIs in the Japanese subset in the FLAURA study was initially inferior to that of osimertinib, crossing over at approximately month 27, after which the gap widened.[18] In contrast, the Kaplan–Meier OS curves in the EXTRA study and for afatinib in the Japanese subset of the LUX-Lung 3 study were almost identical to that for the first-generation EGFR–TKI in the Japanese subset of the FLAURA study.[10,18] As a result, the 3-year OS rates were also similar across the studies: 59% in the EXTRA study, approximately 61% (estimated from Kaplan–Meier curve in the published literature) for afatinib in the Japanese subset of the LUX-Lung 3 study, and 63% for first-generation EGFR–TKI in the Japanese subset of the FLAURA study.[10,18] Moreover, these results were based on patient characteristics with little overall difference between the three studies, especially the treatment rates with second-line osimertinib.[10,18]

The potentially better OS associated with first- and second-generation EGFR–TKIs compared with osimertinib in Japanese patients might be attributable to better postprogression survival (PPS) after first-line treatment with these EGFR–TKIs compared with osimertinib. In the Japanese subset of the FLAURA study, PPS after treatment with osimertinib and a first-generation EGFR–TKI was 20 months and NC, respectively,[18] while PPS after treatment with afatinib in the Japanese subset of the LUX-Lung 3 study and the EXTRA study was 33 months and NC, respectively.[10] We hypothesized that the potentially better PPS of Japanese patients treated with first- or second-generation EGFR–TKIs compared with osimertinib might be because rechallenge therapy with EGFR–TKIs may be less effective after osimertinib. In contrast, rechallenge with osimertinib after first- or second-generation EGFR–TKIs will be effective in patients

with *EGFR* T790M mutation, and rechallenge with EGFR–TKIs other than osimertinib at any treatment line may be effective even in patients without *EGFR* T790M mutation. In a phase II study ($n = 12$) of rechallenge therapy with dacomitinib after osimertinib, the median PFS was only 1.8 months, with limited results even in patients with second-site *EGFR* mutations (C797S or G724S).[27] The reason for this phenomenon is considered to be the frequent development of resistance to osimertinib with the co-occurrence of two or more mutations, making EGFR–TKI monotherapy less effective. In the Osiris study ($n = 50$), the co-mutation rate after osimertinib was 42%, and PFS was comparable between patients treated with cytotoxic chemotherapy and individualized treatment with molecularly targeted therapy.[28] However, two phase II studies of rechallenge therapy with first- or second-generation EGFR–TKIs other than osimertinib in patients with *EGFR* T790M mutation-negative status showed median PFS values of 4.2 months ($n = 12$) and 4.7 months ($n = 32$), respectively.[29,30] Additionally, in a retrospective study ($n = 1603$) of rechallenge therapy with first- or second-generation EGFR–TKIs before osimertinib became available in Japan, rechallenge was performed once in 28% of patients and twice or more in 12% of patients.[31]

We, therefore, inferred that, even if osimertinib is not available after first-line afatinib, multiple rechallenge therapy with EGFR–TKIs may still be effective. The EXTRA study found no significant difference in OS between patients treated with and without osimertinib throughout the treatment. However, a difference might eventually be observed because patients currently not receiving osimertinib may subsequently receive osimertinib if an *EGFR* T790M mutation is detected during long-term follow-up. Nonetheless, the lack of any difference in patients with advanced NSCLC after 35.0 months is clinically meaningful.

Notably, the current frequency of 30% for detecting *EGFR* T790M mutation in the EXTRA study does not seem to be satisfactory. However, several studies have shown that the *EGFR* T790M mutation-positivity rate increased with increasing treatment duration with first-line EGFR–TKI.[32–34] Thus, there is a high probability that the 21% of patients still receiving treatment with first-line afatinib will become *EGFR* T790M mutation-positive in the future. In contrast, there may be some situations in the real-world setting where osimertinib is expected, taking into consideration the treatment duration with first-line EGFR–TKI, to patients with *EGFR* T790M mutation-unknown status because of difficulty in performing re-biopsy for various reasons. In fact, the EXTRA study, reflecting real-world clinical practice, included 11 patients who received osimertinib despite their *EGFR* T790M mutation-unknown status, resulting in a relatively favorable median treatment duration of 7.8 months. However, we have to take care that it is considered the current standard of care in Japan as well to offer chemotherapy to patients with *EGFR* T790M mutation-negative or mutation-unknown status If these patients are chemotherapy-naive.

To date, only one retrospective cohort study has directly compared afatinib and osimertinib in terms of OS in Japanese patients.[35] Consecutive patients were treated with afatinib ($n = 224$) or osimertinib ($n = 326$) as first-line therapy, resulting in median OS after propensity score matching of 36.2 and 25.1 months, respectively (HR 1.47, 95% CI: 1.07–2.02), and median PFS of 16.5 and 20.5 months, respectively (HR 1.02, 95% CI: 0.81–1.28). The median PPS values in the two groups were 19.7 and 4.6 months, respectively, indicating better PPS after first-line treatment with afatinib compared with osimertinib in Japanese patients in a real-world setting. However, further studies are needed to determine the optimal first-line EGFR–TKI.

Our study had several limitations. First, this was a single-arm study with no comparison group. We are therefore now conducting a randomized phase II study comparing first-line afatinib and osimertinib in patients with *EGFR* mutation-positive NSCLC, with 3 year OS rate as the endpoint (the Heat on Beat study).[36] Patient accrual ($n = 100$) was completed on September 7, 2021, and the results will be published in the future. Second, we could not analyze the median treatment duration for sequential afatinib and osimertinib, like the Gio-Tag retrospective and Up-SwinG studies, nor the details of the rechallenge therapy with EGFR–TKI, because we did not schedule these items for analyses at the start of the study. We, therefore, aim to collect these data, together with new OS data, after a minimum follow-up period of 5 years. We also anticipate the results of the ongoing and prospective Gio-Tag Japan study (UMIN000037452). Third, we did not assess serum afatinib concentrations to monitor its pharmacokinetic profile, despite the importance

of these data for validating tolerability-guided dose reduction. However, our comprehensive association study with multi-omics analyses, including metabolomics, will provide useful toxicity predictors for afatinib.

In conclusion, the EXTRA study is the largest prospective study reflecting current real-world data for the use of first-line afatinib in patients with *EGFR* mutation-positive NSCLC in Japan, in an era when osimertinib almost exclusively monopolizes first-line treatment. The results confirmed the favorable OS following first-line afatinib, possibly because of favorable PFS based on a low rate of treatment discontinuation due to AEs and favorable PPS independent of treatment with osimertinib. In the near future, the EXTRA study will identify novel predictive biomarkers for longer OS associated with first-line treatment with afatinib via a comprehensive association study using genomic, proteomic, epigenomic, and metabolomic analyses.

Declarations

Ethics approval and consent to participate

This study complied with all the principles in the Declaration of Helsinki (as revised in 2013), and it was approved by the Ethical Review Board for Medical and Health Research Involving Human Subjects at Teikyo University (approval no. 16-066). All enrolled patients provided written informed consent. This trial was registered with the University Hospital Medical Information Network clinical trial registry (no. UMIN000024935).

Consent for publication

Not applicable

Author contributions

Saori Takata: Resources; Writing – review & editing.

Kei Morikawa: Resources; Writing – original draft; Writing – review & editing.

Hisashi Tanaka: Resources; Writing – review & editing.

Hidetoshi Itani: Resources; Writing – review & editing.

Masashi Ishihara: Resources; Writing – review & editing.

Kazuya Horiuchi: Resources; Writing – review & editing.

Yasuhiro Kato: Resources; Writing – review & editing.

Shinnosuke Ikemura: Resources; Writing – review & editing.

Hideyuki Nakagawa: Resources; Writing – review & editing.

Yoshiro Nakahara: Resources; Writing – review & editing.

Yoshitaka Seki: Resources; Writing – original draft; Writing – review & editing.

Akihiro Bessho: Resources; Writing – review & editing.

Nobumasa Takahashi: Resources; Writing – review & editing.

Kentaro Hayashi: Resources; Writing – review & editing.

Takeo Endo: Resources; Writing – review & editing.

Kiyoshi Takeyama: Resources; Writing – review & editing.

Toshiya Maekura: Resources; Writing – review & editing.

Nagio Takigawa: Resources; Writing – original draft; Writing – review & editing.

Akikazu Kawase: Resources; Writing – review & editing.

Makoto Endoh: Resources; Writing – review & editing.

Kenji Nemoto: Resources; Writing – review & editing.

Kazuma Kishi: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Kenzo Soejima: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Yusuke Okuma: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Kenichi Yoshimura: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing.

Daisuke Saigusa: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Yae Kanai: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Koji Ueda: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Akira Togashi: Conceptualization; Methodology; Project administration; Writing – review & editing.

Noriyuki Matsutani: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Nobuhiko Seki: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Validation; Writing – original draft; Writing – review & editing.

Acknowledgements
The authors would like to thank all the patients who participated in this study and their families, as well as the investigators. The authors express their gratitude to ASCA Corporation for editing a draft of this article funded by Boehringer Ingelheim.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by Boehringer Ingelheim.

Competing interests
Hisashi Tanaka: personal fees as honoraria from Ono Pharmaceutical, Bristol Myers Squibb, AstraZeneca, Chugai Pharmaceutical, Boehringer Ingelheim, and Pfizer; Yoshitaka Seki: personal fees as honoraria from AstraZeneca, Eli Lilly, MSD, Ono Pharmaceutical, Novartis, Taiho Pharmaceutical, Chugai Pharmaceutical, Boehringer Ingelheim, Shionogi, Sanofi, Kyowa Kirin, Kyorin Pharmaceutical, and GSK; Kazuma Kishi: research funding and personal fees as honoraria from Boehringer Ingelheim; Yusuke Okuma: research funding and personal fees as honoraria from AstraZeneca and Chugai Pharmaceutical; research funding from AbbVie; and personal fees as honoraria from Eisai, Eli Lilly, MSD, Novartis, Ono Pharmaceutical, Takeda Pharmaceutical, and Taiho Pharmaceutical; Akira Togashi: employee of Boehringer Ingelheim; Nobuhiko Seki: research funding and personal fees as honoraria from Eli Lilly, Ono Pharmaceutical, Boehringer Ingelheim, Taiho Pharmaceutical, Chugai Pharmaceutical, Takeda Pharmaceutical, Nippon Kayaku, Pfizer, and Daiichi Sankyo; research funding from Eisai and Shionogi; and personal fees as honoraria from AstraZeneca, MSD, Bristol Myers Squibb, Novartis, and Kyowa Kirin. No potential conflicts of interest were disclosed by the other authors.

Availability of data and materials
All data generated or analyzed during this study are included in this published article.

ORCID iDs
Kei Morikawa https://orcid.org/0000-0002-0745-5118

Yasuhiro Kato https://orcid.org/0000-0001-8585-1995

Akihiro Bessho https://orcid.org/0000-0002-0394-7377

Nagio Takigawa https://orcid.org/0000-0002-3835-5128

Supplemental material
Supplemental material for this article is available online.

References
1. The Japan Lung Cancer Society. Guidelines for diagnosis and treatment of lung cancer/malignant pleural mesothelioma/thymic tumor, https://www.haigan.gr.jp/guideline/2022/ (2022, accessed 1 December 2022). (Japanese)
2. Ramalingam SS, Vansteenkiste J, Planchard D, *et al.* Overall survival with osimertinib in untreated, *EGFR*-mutated advanced NSCLC. *New Engl J Med* 2020; 382: 41–50.

3. Paz-Ares L, Tan EH, O'Byrne K, *et al.* Afatinib versus gefitinib in patients with EGFR mutation-positive advanced non-small-cell lung cancer: overall survival data from the Phase IIb LUX-Lung 7 trial. *Ann Oncol* 2017; 28: 270–277.
4. Hochmair MJ, Morabito A, Hao D, *et al.* Sequential afatinib and osimertinib in patients with *EGFR* mutation-positive non-small-cell lung cancer: final analysis of the GioTag study. *Future Oncol* 2020; 16: 2799–2808.
5. Popat S, Jung HA, Lee SY, *et al.* Sequential afatinib and osimertinib in patients with EGFR mutation-positive NSCLC and acquired T790M: a global non-interventional study (UpSwinG). *Lung Cancer* 2021; 162: 9–15.
6. Trams EG, Lauter CJ and Salem N Jr, *et al.* Exfoliation of membrane ecto-enzymes in the form of micro-vesicles. *Biochim Biophys Acta* 1981; 645: 63–70.
7. Valadi H, Ekström K, Bossios A, *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* 2007; 9: 654–659.
8. Okuma Y, Morikawa K, Tanaka H, *et al.* Prospective exosome-focused translational research for afatinib study of non-small cell lung cancer patients expressing EGFR (EXTRA study). *Thorac Cancer* 2019; 10: 395–400.
9. Sequist LV, Yang JC, Yamamoto N, *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013; 31: 3327–3334.
10. Kato T, Yoshioka H, Okamoto I, *et al.* Afatinib versus cisplatin plus pemetrexed in Japanese patients with advanced non-small cell lung cancer harboring activating EGFR mutations: subgroup analysis of LUX-Lung 3. *Cancer Sci* 2015; 106: 1202–1211.
11. Wu YL, Zhou C, Hu CP, *et al.* Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014; 15: 213–222.
12. Park K, Tan EH, O'Byrne K, *et al.* Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. *Lancet Oncol* 2016; 17: 577–589.
13. Seki N, Uematsu K, Shibakuki R, *et al.* Promising new treatment schedule for gefitinib responders after severe hepatotoxicity with daily administration. *J Clin Oncol* 2006; 24: 3213–3214.
14. Arrieta O, De la Torre-Vallejo M, López-Macías D, *et al.* Nutritional status, body surface, and low lean body mass/body mass index are related to dose reduction and severe gastrointestinal toxicity induced by afatinib in patients with non-small cell lung cancer. *Oncologist* 2015; 20: 967–974.
15. Yap TA, Vidal L, Adam J, *et al.* Phase I trial of the irreversible EGFR and HER2 kinase inhibitor BIBW 2992 in patients with advanced solid tumors. *J Clin Oncol* 2010; 28: 3965–3972.
16. Soria JC, Ohe Y, Vansteenkiste J, *et al.* Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *New Engl J Med* 2018; 378: 113–125.
17. Ohe Y, Imamura F, Nogami N, *et al.* Osimertinib versus standard-of-care EGFR-TKI as first-line treatment for EGFRm advanced NSCLC: FLAURA Japanese subset. *Jpn J Clin Oncol* 2019; 49: 29–36.
18. Nogami N, Ramalingam SS, Imamura F, *et al.* Osimertinib as first-line therapy for EGFRm advanced NSCLC (FLAURA): Final OS in Japanese subset. In: Presented at the 60th Annual Meeting of the Japan Lung Cancer Society, Osaka, Japan, 6–8 December 2019, Paper no. PS-1. https://www.haigan.gr.jp/journal/am/2019a/19a_pdsy0000PS-1.html
19. Nakamura A, Tanaka H, Saito R, *et al.* Phase II study of low-dose afatinib maintenance treatment among patients with EGFR-mutated non-small cell lung cancer: North Japan Lung Cancer Study Group Trial 1601 (NJLCG1601). *Oncologist* 2020; 25: e1451–e1456.
20. Imai H, Kaira K, Suzuki K, *et al.* A phase II study of afatinib treatment for elderly patients with previously untreated advanced non-small-cell lung cancer harboring EGFR mutations. *Lung Cancer* 2018; 126: 41–47.
21. Minegishi Y, Yamaguchi O, Sugawara S, *et al.* A phase II study of first-line afatinib for patients aged ⩾75 years with EGFR mutation-positive advanced non-small cell lung cancer: North East Japan Study Group trial NEJ027. *BMC Cancer* 2021; 21: 208.
22. Noro R, Igawa S, Bessho A, *et al.* A prospective, phase II trial of monotherapy with low-dose afatinib for patients with EGFR, mutation-positive, non-small cell lung cancer: Thoracic Oncology Research Group 1632. *Lung Cancer* 2021; 161: 49–54.
23. Yokoyama T, Yoshioka H, Fujimoto D, *et al.* A phase II study of low starting dose of afatinib

as first-line treatment in patients with EGFR mutation-positive non-small-cell lung cancer (KTORG1402). *Lung Cancer* 2019; 135: 175–180.

24. Mizugaki H, Oizumi S, Fujita Y, *et al.* Pharmacokinetic and pharmacogenomic analysis of low-dose afatinib treatment in elderly patients with EGFR mutation-positive non-small cell lung cancer. *Eur J Cancer* 2022; 160: 227–234.
25. Yang JC, Sequist LV, Zhou C, *et al.* Effect of dose adjustment on the safety and efficacy of afatinib for EGFR mutation-positive lung adenocarcinoma: post hoc analyses of the randomized LUX-Lung 3 and 6 trials. *Ann Oncol* 2016; 27: 2103–2110.
26. Schuler M, Paz-Ares L, Sequist LV, *et al.* First-line afatinib for advanced EGFRm+ NSCLC: analysis of long-term responders in the LUX-Lung 3, 6, and 7 trials. *Lung Cancer* 2019; 133: 10–19.
27. Choudhury NJ, Makhnin A, Tobi YY, *et al.* Pilot study of dacomitinib for patients with metastatic ***EGFR***-mutant lung cancers with disease progression after initial treatment with osimertinib. *JCO Precis Oncol* 2021; 5: PO.21.00005.
28. Van Der Wel T, Jebbink M, Boelens Mc, *et al.* MA07.03 paired liquid and tissues biopsies to guide treatment for patients that progress on 2nd line osimertinib treatment. *J Thorac Oncol* 2021; 16: S906.
29. Oda N, Hotta K, Ninomiya K, *et al.* A phase II trial of EGFR-TKI readministration with afatinib in advanced non-small-cell lung cancer harboring a sensitive non-T790M EGFR mutation: Okayama Lung Cancer Study Group trial 1403. *Cancer Chemother Pharmacol* 2018; 82: 1031–1038.
30. Song Y, Wu YL, Cao LJ, *et al.* Efficacy and safety of gefitinib as third-line treatment in NSCLC patients with activating EGFR mutations treated with first-line gefitinib followed by second-line chemotherapy: a single-arm, prospective, multicenter phase II study (RE-CHALLENGE, CTONG1304). *Am J Clin Oncol* 2019; 42: 432–439.
31. Inoue A, Yoshida K, Morita S, *et al.* Characteristics and overall survival of EGFR mutation-positive non-small cell lung cancer treated with EGFR tyrosine kinase inhibitors: a retrospective analysis for 1660 Japanese patients. *Jpn J Clin Oncol* 2016; 46: 462–467.
32. Matsuo N, Azuma K, Sakai K, *et al.* Association of EGFR exon 19 deletion and EGFR-TKI treatment duration with frequency of T790M mutation in EGFR-mutant lung cancer patients. *Sci Rep* 2016; 6: 36458.
33. Oya Y, Yoshida T, Kuroda H, *et al.* Association between EGFR T790M status and progression patterns during initial EGFR-TKI treatment in patients harboring EGFR mutation. *Clin Lung Cancer* 2017; 18: 698–705.e2.
34. Kawamura T, Kenmotsu H, Omori S, *et al.* Clinical factors predicting detection of T790M mutation in rebiopsy for EGFR-mutant non-small-cell lung cancer. *Clin Lung Cancer* 2018; 19: e247–e252.
35. Ito K, Morise M, Wakuda K, *et al.* A multicenter cohort study of osimertinib compared with afatinib as first-line treatment for EGFR-mutated non-small-cell lung cancer from practical dataset: CJLSG1903. *ESMO Open* 2021; 6: 100115.
36. Morikawa K, Tanaka H, Itani H, *et al.* Hypothesis generative head-to-head study comparing efficacy of afatinib and osimertinib based on immunological biomarkers in Japanese NSCLC patients with EGFR mutations (Heat on Beat study). *Ther Adv Med Oncol* 2020; 12: 1758835920967254.

Follow-up focused on psychological intervention initiated after intensive care unit in adult patients and informal caregivers: a systematic review and meta-analysis

Shodai Yoshihiro[1], Shunsuke Taito[2,3], Kota Yamauchi[4], Shunsuke Kina[5], Takero Terayama[6], Yusuke Tsutsumi[3,7,8], Yuki Kataoka[3,9,10,11] and Takeshi Unoki[12]

[1] Department of Pharmaceutical Services, Hiroshima University Hospital, Hiroshima, Japan
[2] Division of Rehabilitation, Department of Clinical Practice and Support, Hiroshima University Hospital, Hiroshima, Japan
[3] Scientific Research WorkS Peer Support Group (SRWS-PSG), Osaka, Japan
[4] Division of Rehabilitation, Steel Memorial Yawata Hospital, Fukuoka, Japan
[5] Division of Rehabilitation, Nakagami Hospital, Okinawa, Japan
[6] Department of Psychiatry, School of Medicine, National Defense Medical College, Tokorozawa, Japan
[7] Department of Emergency Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan
[8] Department of Human Health Science, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[9] Department of Internal Medicine, Kyoto Min-iren Asukai Hospital, Kyoto, Japan
[10] Section of Clinical Epidemiology Section, Department of Community Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
[11] Department of Healthcare Epidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan
[12] Department of Acute and Critical Care Nursing, School of Nursing, Sapporo City University, Sapporo, Japan

ABSTRACT

Psychological dysfunction is one of the considerable health-related outcomes among critically-ill patients and their informal caregivers. Follow-up of intensive care unit (ICU) survivors has been conducted in a variety of different ways, with different timing after discharge, targets of interest (physical, psychological, social) and measures used. Of diverse ICU follow-up, the effects of follow-ups which focused on psychological interventions are unknown. Our research question was whether follow-up with patients and their informal caregivers after ICU discharge improved mental health compared to usual care. We published a protocol for this systematic review and meta-analysis in https://www.protocols.io/ (https://dx.doi.org/10.17504/protocols.io.bvjwn4pe). We searched PubMed, Cochrane Library, EMBASE, CINAHL and PsycInfo from their inception to May 2022. We included randomized controlled trials for follow-ups after ICU discharge and focused on psychological intervention for critically ill adult patients and their informal caregivers. We synthesized primary outcomes, including depression, post-traumatic stress disorder (PTSD), and adverse events using the random-effects method. We used the Grading of Recommendations Assessment, Development and Evaluation approach to rate the certainty of evidence. From the 10,471 records, we identified 13 studies ($n = 3,366$) focusing on patients and four ($n = 538$) focusing on informal caregivers. ICU follow-up for patients resulted in little to no difference in the

Submitted 17 October 2022
Accepted 28 March 2023
Published 9 June 2023

Corresponding author
Shunsuke Taito,
shutaitou@hiroshima-u.ac.jp

Academic editor
Michael Simon

Additional Information and Declarations can be found on page 19

DOI **10.7717/peerj.15260**

How to cite this article Yoshihiro S, Taito S, Yamauchi K, Kina S, Terayama T, Tsutsumi Y, Kataoka Y, Unoki T. 2023. Follow-up focused on psychological intervention initiated after intensive care unit in adult patients and informal caregivers: a systematic review and meta-analysis. *PeerJ* **11**:e15260 http://doi.org/10.7717/peerj.15260

prevalence of depression (RR 0.89, 95% CI [0.59–1.34]; low-certainty evidence) and PTSD (RR 0.84, 95% CI [0.55–1.30]; low-certainty evidence) among patients; however, it increased the prevalence of depression (RR 1.58 95% CI [1.01–2.46]; very low-certainty evidence), PTSD (RR 1.36, 95% CI [0.91–2.03]; very low-certainty evidence) among informal caregivers. The evidence for the effect of ICU follow-up on adverse events among patients was insufficient. Eligible studies for informal caregivers did not define any adverse event. The effect of follow-ups after ICU discharge that focused on psychological intervention should be uncertain.

Subjects Emergency and Critical Care, Nursing, Psychiatry and Psychology, Mental Health, Rehabilitation
Keywords Intensive care units, Critical care, Mental disorders, Post intensive care syndrome

INTRODUCTION

Adult patients who are admitted to intensive care units (ICU) and their informal caregivers may experience psychological dysfunction, which can persist following discharge (*Needham et al., 2012*). Psychological dysfunction of critically-ill adult patients and their informal caregivers is called post intensive care syndrome (PICS) and PICS-Family (PICS-F), respectively. Other symptoms of PICS include cognitive and physical impairments. Previous studies found that the prevalence of these patients with depression, post-traumatic stress disorder (PTSD), and anxiety was approximately 29% (*Rabiee et al., 2016*), 34% (*Parker et al., 2015*), and 34% (*Nikayin et al., 2016*) after one year of ICU discharge. Studies have also reported that the prevalence of acquired psychological dysfunction among informal caregivers was similar to that among patients (*Johnson et al., 2019*). Therefore, psychological dysfunction is a considerable health-related outcome among critically-ill patients and their informal caregivers.

According to the current guidelines and a systematic review (SR), follow-up with patients who have been admitted to the ICU is comprised of a variety of contents, targets, and times of initiation (*National Institute for Health and Care Excellence, 2009*; *Rosa et al., 2019*). The National Institute for Health and Clinical Excellence guidelines for follow-ups recommended providing enhanced or individualized physical intervention from early mobilization to home rehabilitation (*National Institute for Health and Care Excellence, 2009*). One SR found that the intervention that was initiated in the ICU and continued after ICU discharge included diary and physical rehabilitation (*Rosa et al., 2019*). In addition, the SR did not separately investigate patients and informal caregivers. Similarly, the counterplan for PICS-F was the ICU diary and communication in the ICU. Another SR showed that care providers and informal caregivers regarded the ICU diary as beneficial (*Brandao Barreto et al., 2021*), while another SR asserted that communication in the ICU might reduce symptoms of depression and PTSD (*DeForge et al., 2022*). It would be obvious that these interventions which initiated in the ICU reduced psychological problems of patients and informal caregivers. Moreover, a recent SR studied psychological intervention for patients' informal caregivers, but did not separately investigate adult patients and

pediatric patients (*Cherak et al., 2021*). In a pediatric randomized controlled trial (RCT), interventions were specifically designed for children such as skin-to-skin contact (*Mörelius et al., 2015*), kangaroo care (*Ettenberger et al., 2017*), or guidance for baby care (*Fotiou et al., 2016*). There was clinical heterogeneity among the included studies in the previous SR. Hence, the effects of follow-ups for adult patients and informal caregivers that focused on psychological interventions after ICU discharge have remained unknown.

Thus, the objective of this systematic review and meta-analysis (SR/MA) was to investigate the following research question: does follow-up with adult patients and their informal caregivers following ICU discharge improve mental health compared to usual care?

MATERIALS & METHODS

Protocol and registration

We published a protocol for this SR/MA in http://www.protocols.io (*Yoshihiro et al., 2021*). We conducted this SR/MA in accordance with guidelines prescribed by the Cochrane Handbook for Systematic Reviews of Interventions (*Higgins et al., 2020*) and Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (*Page et al., 2021*). The principles listed in the PRISMA statement formed the basis of our SR/MA report (*Page et al., 2021*) (Table S1).

Eligibility criteria

Studies

We included randomized controlled trials that assessed the effects of follow-up after ICU discharge on mental health outcomes among adult patients and informal caregivers. We analyzed papers including published and unpublished articles, abstracts of conferences, and condolence letters. We excluded studies with cluster randomized or quasi-randomized trials, cohort studies, case-control studies, and case series. Furthermore, while including studies for this SR/MA, we did not apply restrictions pertaining to language, country, observation period, or publication year.

In May 2021, we searched the following databases: MEDLINE (PubMed), the Cochrane Central Register of Controlled Trials (Cochrane Library), EMBASE (Dialog), the Cumulative Index to Nursing and Allied Health Literature (CINAHL) (accessed via EBSCO), and APA PsycInfo (Ovid). In May 2021, we searched for ongoing and unpublished trials in trial registers such as ClinicalTrials.gov and the World Health Organization International Clinical Trials Platform Search Portal (WHO ICTRP), respectively. Details of these searches have been listed in the protocol (*Yoshihiro et al., 2021*). We conducted a 'snowball' search to identify studies that used reference lists of publications eligible for full-text review (including international guidelines) (*National Institute for Health and Care Excellence, 2009*; *Nolan et al., 2021*) and used Google Scholar to identify and screen those studies. We reconducted these searches in May 2022. Additionally, we contacted the authors of the original studies for unpublished or additional data.

Population

We included trials with adult patients (age ≥18 years) admitted to ICUs and their informal caregivers; these trials were randomized during both ICU and hospital discharge. We included studies involving informal caregivers regardless of whether the admitted patient survived. We excluded studies involving patients and their caregivers who were younger than 18 years, did not provide consent for participation, or showed cognitive impairment. Furthermore, studies involving patients or caregivers who had experienced myocardial infarction or were in their perioperative period were excluded. In this article, we have referred to our target population of ''critically-ill adult patients'' as ''patients.'' if not necessary.

Interventions

We defined *intervention* as a service or program initiated after ICU discharge (within one month after hospital discharge), including multidisciplinary interventions, follow-up clinics, and other programs. In the included studies, we recognized counseling such as cognitive-behavioral therapy, that interventions target mental health conditions. In addition, we included psychological intervention performed as needed after monitoring. We incorporated all intervention periods by all professionals. In the included studies, nurses and physicians intervening in therapies had been trained for each study.

We excluded studies involving interventions in the ICU that were comprised of participant-led initiatives like ICU diaries and ICU records, interventions that provided general information pertaining to post-intensive care syndrome using web tools or video materials, or that compared enhanced physical rehabilitation with usual care. We did not predefine the details of the psychological interventions because we wanted to verify interventions that improved psychological outcomes other than physical and diary interventions.

Outcomes

We included trials with defined clinical outcomes, such as symptoms of depression and PTSD, and all adverse events were considered primary outcomes among patients and caregivers (*Marra et al., 2018*). Additional outcomes among patients included anxiety, health-related quality of life (HR-QoL), pain, readmission, and long-term mortality; additional outcomes among caregivers included anxiety and HR-QoL. We followed core outcome sets (*Angus & Carlet, 2003*; *Major et al., 2016*; *Needham et al., 2017*). We selected outcomes for mental health as primary outcomes. We defined depression, PTSD, and anxiety as the prevalence rate of significant symptoms based on definitions by the included studies' authors, measured between three months and one year after randomization or ICU discharge. We defined adverse events using the incidence proportion of all adverse events set by the original authors during the follow-up period of included studies. We defined HR-QoL using a mental component summary of the Medical Health Survey Short-Form 36 (SF-36), measured between three months and one year after randomization or ICU discharge. SF-36 was used for self-reported evaluation scales for the evaluation of HR-QoL (*Angus & Carlet, 2003*; *Needham et al., 2017*). If the outcome of HR-QoL was measured by other self-reported evaluation scales in included studies, we assessed whether the scales

could be synthesized with SF-36. We defined pain using self-reported evaluation scales for pain set by the original authors, measured between three months and one year after randomization or ICU discharge. We defined readmission as the proportion of readmission (at least once) during the follow-up period of the included studies. For long-term mortality, we collected the reported mortality at the longest timepoint available in the study, which ranged between 3 and 12 months after randomization.

Search strategy

Selection process

Three reviewers (SY, YK, and KS) independently screened the titles and abstracts of records during the initial screening. We assessed records—included in the initial screening—for eligibility based on the inclusion criteria by reading the full texts. We resolved disagreements between two reviewers *via* discussion with a third reviewer (TS) to achieve consensus. We combined machine learning classifiers during the selection process (*Marshall et al., 2018*).

Data collection process

Three reviewers (SY, YK, and KS) independently extracted data from the included studies using a standardized data collection form. We pre-checked the form by using 10 randomly selected studies. We extracted the following characteristics:

Methods: Study design, study follow-up period, and study country;

Participants: Country, setting, mental condition (depression, PTSD, and anxiety), sample size, age, relationship of informal caregivers with patients, and attrition;

Interventions: type, intervention about the psychological problem, providers, media, initiation, duration, and frequency;

Outcomes: primary and additional outcomes specified and collected, and the timepoints reported.

Data items

Study risk-of-bias assessment

Two to three reviewers (SY, YK, and KS) independently classified the risk of bias as "low", indicating "some concerns", or "high" based on the Risk-of-Bias 2.0 (*Sterne et al., 2019*). We resolved disagreements between two reviewers *via* discussion with the third reviewer (TS) to achieve consensus. As participants could not be blinded to the intervention owing to its nature, we assessed the overall risk-of-bias using four domains, which excluded the estimation of measurement-of-outcome.

Effect measures

We analyzed the dichotomous variables by calculating risk ratios (RR) with 95% confidence intervals (CIs). We analyzed the continuous variables using standard mean differences (SMD) with 95% CI.

Synthesis methods

We synthesized the collected variables (except for adverse events) using the random-effects method; data for patients and informal caregivers were synthesized separately. We used the Review Manager software (RevMan 5.4.2) for quantitative synthesis.

Dealing with missing data

We used available data published and inquired to authors. We performed (modified) intention-to-treat data for all dichotomous data as much as possible. For continuous data, we did not impute missing data and performed a meta-analysis of the available data in the original studies and the converted data from available data based on the method in the Cochrane handbook (*Higgins et al., 2020*).

Assessment of heterogeneity

We assessed heterogeneity by visual inspection of the forest plot and I^2 statistics (I^2 values of 0% to 40%: might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may represent substantial heterogeneity; 75% to 100%: considerable heterogeneity). We performed Cochrane Chi^2 test(Q-test) for I^2 statistic and defined *P* values less than 0.10 as statistically significant.

Sensitivity analysis and subgroup analysis

We conducted the sensitivity analysis and subgroup analysis for the primary outcomes where sufficient data were available. We conducted sensitivity analysis of patients using studies measured by the Depression subscale of the Hospital Anxiety and Depression Scale(HADS-D) score for depression, studies measured by the Impact of Event Scale-Revised (IES-R) score for PTSD, and exclusion of imputed data. We conducted the sub-group analyses by timing for initiation of follow-up(in-hospital, out-hospital, or in- and out-hospital). For analysis for informal caregivers, we conducted sensitivity analysis using studies measured by IES-R scores for PTSD. We divided the ICU survivors and non-survivors in the sub-group analyses for informal caregivers.

Reporting bias assessment

We identified the number of studies that had not been published on ClinicalTrials.gov and WHO ICTRP. We assessed outcome reporting bias by comparing the outcomes defined in trial protocols with the outcomes reported in the publications. We assessed the publication bias of outcomes by visual inspection of the funnel plots.

Certainty assessment

Two reviewers (SY and TU) evaluated the certainty of evidence based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (*Hultcrantz et al., 2017*). We resolved disagreements between two reviewers *via* discussion with the third reviewer (KY) to achieve consensus. We generated a table to summarize the findings of the seven outcomes (except for long-term mortality) using GRADE Pro GDT (https://gradepro.org) based on the Cochrane Handbook (*Higgins et al., 2020*). We selected the following outcomes for patients: (1) depression, (2) PTSD, (3) all adverse events, (4) anxiety, (5) HR-QoL, (6) pain, and (7) readmission. We selected the following outcomes for informal caregivers: (1) depression, (2) PTSD, (3) all adverse events, (4) anxiety, and (5) HR-QoL.

Difference between protocol and review

We did not conduct Egger's test as we synthesized data from fewer than 10 studies. We could not conduct planned sensitivity and sub-group analyses for PTSD and adverse events among patients and depression and adverse events among informal caregivers. We added a sub-group analysis for the endpoints of the measured outcomes, dividing them into 6 months and 12 months.

RESULTS

Study selection

We identified 10,425 records from databases and registers, and 46 records from citation searches and guidelines (*National Institute for Health and Care Excellence, 2009*; *Nolan et al., 2021*). After excluding duplicates, we could not retrieve the full text for one record from the Cochrane Library and confirmed that the record was an error through author inquiry. We assessed 240 full texts for eligibility and identified 119 studies. The flow diagram for study selection is presented in Fig. 1.

We identified six ongoing studies and one no-information study with patients, and one ongoing study with informal caregivers *via* ClinicalTrials.gov and WHO ICTRP. The details of all studies without results are outlined in Table S2. We excluded 92 studies after conducting full-text reviews; the reasons for their exclusion are listed in Table S3.

Since 12 of the included studies did not include results (*Chen et al., 2022*; *Ewens et al., 2019*; *Friedman et al., 2022*; *Gawlytta et al., 2020*; *Gawlytta et al., 2017*; *Haines et al., 2019*; *Khan et al., 2018*; *Moulaert et al., 2015*; *Ojeda et al., 2021*; *Rohr et al., 2021*) (NCT03431493, NCT03926533, NCT04329702), we included 15 studies for quantitative analysis. Of these 15 studies, 11 focused on patients (*Abdelhamid et al., 2021*; *Bloom et al., 2019*; *Cox et al., 2018*; *Cox et al., 2019*; *Cuthbertson et al., 2009*; *Daly et al., 2005*; *Douglas et al., 2005*; *Douglas et al., 2007*; *Hernández et al., 2014*; *Kredentser et al., 2018*; *McWilliams, Benington & Atkinson, 2016*; *Schmidt et al., 2016*; *Schmidt et al., 2020*; *Valsøet al., 2020*; *Vlake et al., 2021*), two focused on informal caregivers (*Ågren et al., 2019*; *Kentish-Barnes et al., 2017*), and two focused on both patients and informal caregivers (*Bohart et al., 2019*; *Jensen et al., 2016*; *Jones et al., 2004*; *Jones et al., 2003*). One study (*Cox et al., 2018*) was conducted with both patients and informal caregivers, but we could not retrieve outcome data for the informal caregivers. The details of these studies are outlined in Table 1.

Study characteristics

We selected 13 studies that included 3,366 patients (Table 1A). These studies were conducted in eight countries: the USA ($n = 4$), the UK ($n = 3$), and Denmark, Germany, Norway, Netherlands, Canada, and Australia ($n = 1$ in each country). Patients in two studies had sepsis, and patients in six studies were provided mechanical ventilation. One study included patients with moderate PTSD symptoms after ICU discharge. Interventions in six studies focused on psychological problems among patients following critical illness. Interventions in seven studies included rehabilitation programs, multidisciplinary programs, and case management for monitoring and therapy for psychological problems.

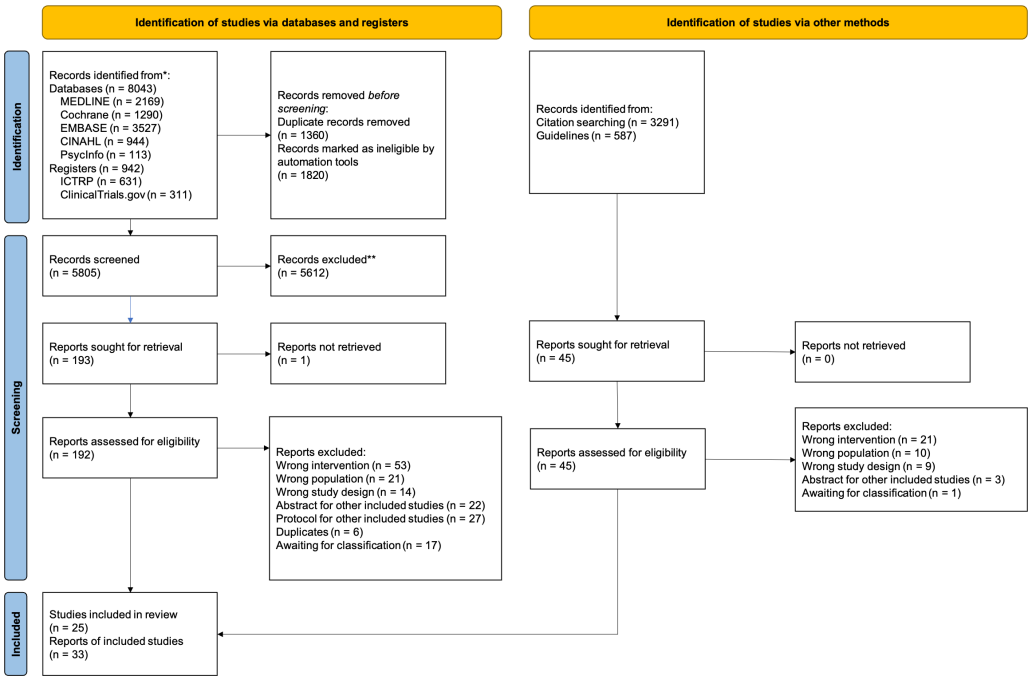


Figure 1 **PRISMA flow.**

Full-size DOI: 10.7717/peerj.15260/fig-1

We selected four studies, which included 538 informal caregivers (Table 1B). These studies were conducted in four countries: the UK, Denmark, France, and Sweden ($n = 1$ in each country). Most caregivers were spouses (47.8%), followed by children (16.8%), parents (9.3%), and siblings (1.3%) of the patients. All the studies included informal caregivers with or without psychological problems. Follow-ups were conducted on patients and caregivers in three studies, while one study conducted interventions on caregivers of the ICU non-survivor.

Risk of bias in studies

The domains and overall risk of bias for each outcome are outlined in Fig S1. On the assessment of the randomization process, we found that one study (*Daly et al., 2005*) showed risk-of-bias concerns owing to no description of the details of concealment, and two studies (*Ågren et al., 2019*; *Bloom et al., 2019*) showed high risk of bias owing to an imbalance of patient characteristics. On the assessment of deviation from the intended interventions, we found that three studies (*Ågren et al., 2019*; *Cox et al., 2019*; *Daly et al., 2005*) showed some risk-of-bias concerns owing to the difference of drop-outs between each group, and one study (*Kredentser et al., 2018*) had a high risk of bias owing to no information and no conduct of modified intention for treatment. On the assessment of the missing outcome data, we found that four studies (*Cox et al., 2018*; *Cox et al., 2019*; *Jensen et al., 2016*; *Vlake et al., 2021*) had a low risk of bias for implementation of missing values; however, 10.2–52.1% of the participants dropped out in all eligible studies. The assessment of the outcome measurement indicated that all studies had a high risk of bias for outcomes estimated *via* self-reported questionnaires as patients could not be blinded to the interventions owing to their nature.

Table 1 Included studies.

(A) Patients Authors year	Registry Number Country Observational period	No of participants Age, years Intervention/-Control	Mental condition Intervention/Control	Attrition, %	Type of intervention	Type of intervention against psychological problem	Professionals/ sources of intervention	Timing, duration, and/or frequency of intervention
Jones et al. (2003)	Not stated about registration the United Kingdom six months after ICU discharge	69/57 Mean ± SD, 57 ± 17/59 ± 16	Depression not stated; PTSD not stated; Anxiety not stated	19	Semi-structured programs for psychological, psychosocial, and physical problems	Provision of coping skills	Print media	After ICU discharge six weeks from one week
Daly et al. (2005)	No detail of registration the United States of America two months after hospital discharge	231/103 Mean ± SD, 60.7 ± 16.6/ 61.4 ± 16.1	Depression not stated; PTSD not stated; Anxiety not stated	26	Multidisciplinary intervention by nurse with support from a physician	Provision of coping skills	Nurse	After hospital discharge Two months
Cuthbertson et al. (2009)	ISRCTN24294750 The United Kingdom 12 months after ICU discharge	143/143 Median (IQR), 59 (46–49)/60 (46–71)	Depression not stated; PTSD not stated; Anxiety not stated	32.9	Multidisciplinary intervention by nurse with support from an intensivist	Psychological intervention required after monitoring	Nurse	After hospital discharge Two times at 3 months and 9 months
Jensen et al. (2016)	NCT01721239 Denmark 12 months after ICU discharge	190/196 Median (IQR), 66 (57.75–73.5)/67.5 (58–75)	Depression not stated; PTSD not stated; Anxiety not stated	39.1	Individualized, semi-structured program for psychological problem	Therapy: Cognitive behavioral therapy	Nurse	After ICU discharge Three times at 1–3, 5, and 10 months
McWilliams, Benington & Atkinson (2016)	NCT02491021 The United Kingdom seven weeks after hospital discharge	37/36 Mean ± SD, 55.0 ± 12.9, 60.8 ± 12.3	Depression not stated; PTSD not stated; Anxiety not stated	13.7	Rehabilitation program consisted of exercise and education component	Education	Nurse; Facilitators other than physician and nurse	After hospital discharge Total 6 educational sessions, 1 h per session, for 7 weeks
Schmidt et al. (2016)	ISRCTN61744782 Germany 12 months after ICU discharge	148/143 Mean ± SD, 62.1 ± 14.1/ 61.2 ± 14.9	Depression not stated; PTSD not stated; Anxiety not stated	30.6	Case management, telephone monitoring, and education of behavioral activation for patients, which consisted of general practitioner, case manager, and liaison physician	Provision of coping skills	Nurse; Physician	After ICU discharge Monthly for 6 months, and once every 3 months for the final 6 months
Cox et al. (2018)	NCT01983254 The United States of America 12 months after randomization (within two weeks after hospital discharge)	39/47 Mean ± SD, 49.7 ± 13.8/53.7 ± 13.5	*Patients* Depression 27/20 PTSD 4/6 Anxiety 24/17	*Patients* 25.1	Training for psychological problems, combined with Telephone and web	Provision of coping skills	Facilitators other than physician and nurse; Digital media	After hospital discharge six telephone sessions for thirty minutes, once per week
Bloom et al. (2019)	NCT03124342 The United States of America 30 days after hospital discharge	145/157 Median (IQR), 56 (44–67), $n = 111$/56 (48–66), $n = 121$	Depression Not stated; PTSD not stated; Anxiety Not stated	27.5	Multidisciplinary case management based on ICU recovery program	Psychological intervention required after monitoring	Nurse; Physician; Facilitators other than physician and nurse	After hospital discharge At least 30 days
Cox et al. (2019)	NCT02701361 The United States of America Three months after hospital discharge	1) Telephone-based mindfulness training, 31/18 Mean ± SD, 48.1 ± 16.1/53.3 ± 12.6	1) Depression 4/1 PTSD 1/1 Anxiety 6/1	1) 10.2	1) Telephone-based training for psychological problems	1) Provision of coping skills	1) Facilitator other than physician and nurse	After hospital discharge Four sessions each week for one month
		2) Self-directed mindfulness training by mobile app, 31/18 Mean ± SD, 48.7 ± 15.3/53.3 ± 12.6	2) Depression 1/1 PTSD 2/0 Anxiety 2/1	2) 22.4	2) Self-directed training for psychological problems	2) Provision of coping skills	2) Digital media	
Kredentser et al. (2018)	NCT02067559 Canada 90 days after ICU discharge	Sample size of usual care and psychoeducation in four arms 14/14 Mean ± SD, 59.3 ± 15.5/49.9 ± 16.9	Depression not stated; PTSD not stated; Anxiety not stated	60.7	Education for psychological problem	Provision of coping skills	Print media	After ICU discharge or after return of the ability to provide consent

(*continued on next page*)

Table 1 (*continued*)

(A) Patients

Authors year	Registry Number Country Observational period	No of participants Age, years Intervention/-Control	Mental condition Intervention/Control	Attrition, %	Type of intervention	Type of intervention against psychological problem	Professionals/ sources of intervention	Timing, duration, and/or frequency of intervention
Valsøet al. (2020)	NCT02077244 Kingdom of Norway Twelve months after ICU discharge	111/113 Mean ± SD, 53 ± 16/50 ± 18	Depression not stated; PTSD 111/113; Anxiety not stated	23.7	Individualized, semi-structured program for psychological and psychosocial problems	Therapy: Cognitive behavioral therapy	Nurse	After ICU discharge three times in the first week, one and two months later
Abdelhamid et al. (2021)	ACTRN12616000206426 Australia six months after hospital discharge	21/21 Mean ± SD, 64 ± 11/68 ± 8	Depression not stated; PTSD not stated; Anxiety not stated	38.1	Multidisciplinary intervention by an intensivist and endocrinologist	Psychological intervention required after monitoring	Physician	After hospital discharge At least one time, repeated as needed for six months from one month
Vlake et al. (2021)	NL6611 Netherlands six months after ICU discharge	25/25 Median (95% range), 61 (23-75)/59 (59–80)	Depression 6/12 PTSD 12/13; Anxiety not stated	16	ICU-specific virtual reality for psychological problem	Therapy: Virtual reality exposure therapy	Digital media	After ICU discharge The number of desired sessions was offered daily

(B) Informal caregivers

Authors year	Registry number Country Observational period	No of participants Age, years Intervention/Control	Mental condition Intervention/Control	Relationship of caregivers with patients				Attrition, %	Type of intervention	Type of intervention against psychological problem	Professionals/sources of intervention;	Timing, duration, and/or frequency of intervention
				Spouse, %	Child, %	Parent, %	Sibling, %					
Jones et al. (2004)	No detail of registration the United Kingdom six months after ICU discharge	*Caregivers* 58/46 Mean ± SD, 62 ± 17/60 ± 15.4	Depression 13/14 PTSD not stated; Anxiety 34/29	51.9	19.2	18.3	6.7	19.2	Training for psychological problems	Provision of coping skills	Print media;	After ICU discharge six weeks from one week
Jensen et al. (2016)	NCT03264365 Denmark 12 months after ICU discharge	87/94 Median (IQR), 57.4 (50–67)/61 (41.8–69)	Depression not stated; PTSD not stated; Anxiety not stated	71.3	Not stated	17.1	Not stated	38.7	Individualized, semi-structured program for psychological problems	Therapy: Cognitive behavioral therapy	Nurse;	After ICU discharge Once at 1–3 months
Kentish-Barnes et al. (2017)	NCT02325297 France six months after in the 24 h following the death of the patient	109/99 Median (Range), 57 (46–65.5) /56 (44–64.5)	Depression not stated; PTSD not stated; Anxiety not stated	35.6	39.9	Not stated	Not stated	22.3	Condolence letters	Empathy: Condolence letters	Print media;	After patient's death Once at 15 days
Ågren et al. (2019)	NCT03325049 The Kingdom of Sweden 12 months after ICU discharge	Seven families (17 individuals) /10 families (28 individuals) Mean ± SD, 60 ± 19/61 ± 17	Depression not stated; PTSD not stated; Anxiety not stated	Not stated	Not stated	Not stated	Not stated	51.1	Health-promoting conversation forced on experience of the current situation	Empathy: Counseling	Nurse;	After ICU discharge Two weeks interval, within approximately 4 to 8 weeks after hospital discharge

Notes.

IQR, Interquartile range; ICU, intensive care unit; SD, standard deviation; PTSD, post-traumatic stress disorder.

Patient outcomes

Depression

As shown in Fig. 2 and Table 2, ICU follow-ups resulted in little to no differences in the prevalence rate of depressive symptoms among patients (RR 0.89, 95% CI [0.59, 1.34]; $I^2 = 1\%$; four studies, 758 patients; low-certainty evidence) (*Abdelhamid et al., 2021*; *Jensen et al., 2016*; *Jones et al., 2003*; *Schmidt et al., 2016*; *Vlake et al., 2021*); we detected slight heterogeneity. Planned sensitivity analyses of studies using the Depression subscale scores of the Hospital Anxiety and Depression Scale (HADS-D) yielded similar findings (RR 0.90, 95% CI [0.50–1.63]). Planned sensitivity analysis that excluded the imputed data showed a similar trend (RR 1.08, 95% CI [0.55–2.09]). Sub-group analysis for the timing of follow-up initiation showed a similar trend in the group of initiation from both ICU discharge and hospital discharge. In the sub-group analysis, there was no difference in the endpoint to measure depressive symptoms between 6 months and 12 months. Details of the analysis are provided in Fig S2.

Post-traumatic stress disorder

ICU follow-ups resulted in little to no differences in the prevalence rate of PTSD symptoms among patients (RR 0.84, 95% CI [0.55–1.30]; $I^2 = 53\%$; four studies, 732 patients; low-certainty evidence) (*Jensen et al., 2016*; *Jones et al., 2003*; *Schmidt et al., 2016*; *Vlake et al., 2021*); we detected moderate heterogeneity (Fig. 2 and Table 2). Planned sensitivity analysis of studies using the Impact of Event Scale- Revised scores (IES-R) yielded similar results (RR 0.51, 95% CI [0.08–3.23]). The planned sensitivity analysis that excluded the imputed data generated similar findings (RR 1.06, 95% CI [0.75–1.50]; Fig. S2). Sub-group analysis for the endpoint to measure PTSD symptoms showed a similar trend in the endpoint to measure depressive symptoms between 6 months and 12 months. Details of the analysis are provided in Fig. S2.

Adverse events

Although evidence indicates considerable uncertainty, ICU follow-ups resulted in little to no differences in the occurrence of adverse events (*Vlake et al., 2021*) (Table 2). Two studies included adverse events as outcome measures (*Bloom et al., 2019*; *Vlake et al., 2021*). One published article (*Bloom et al., 2019*) did not report the results pertaining to adverse events, and we could not obtain information about adverse events from its authors. This study defined adverse events as the need for intervention to prevent events such as mortality, prolonged hospitalization, acquisition of disability, congenital anomalies, and birth defects. Another study (*Vlake et al., 2021*) defined adverse events as incidents of cybersickness, delirium, or the use of haloperidol. Considering the clinical heterogeneity in studies, we included all types of adverse events except for cybersickness.

Anxiety

ICU follow-ups resulted in little to no differences in the prevalence rate of anxiety symptoms among patients (RR 1.04, 95% CI [0.68–1.60]; $I^2 = 0\%$; two studies, 488 patients; low certainty of evidence) (*Jensen et al., 2016*; *Jones et al., 2003*); no significant heterogeneity was detected (Table 2 and Fig. S3).

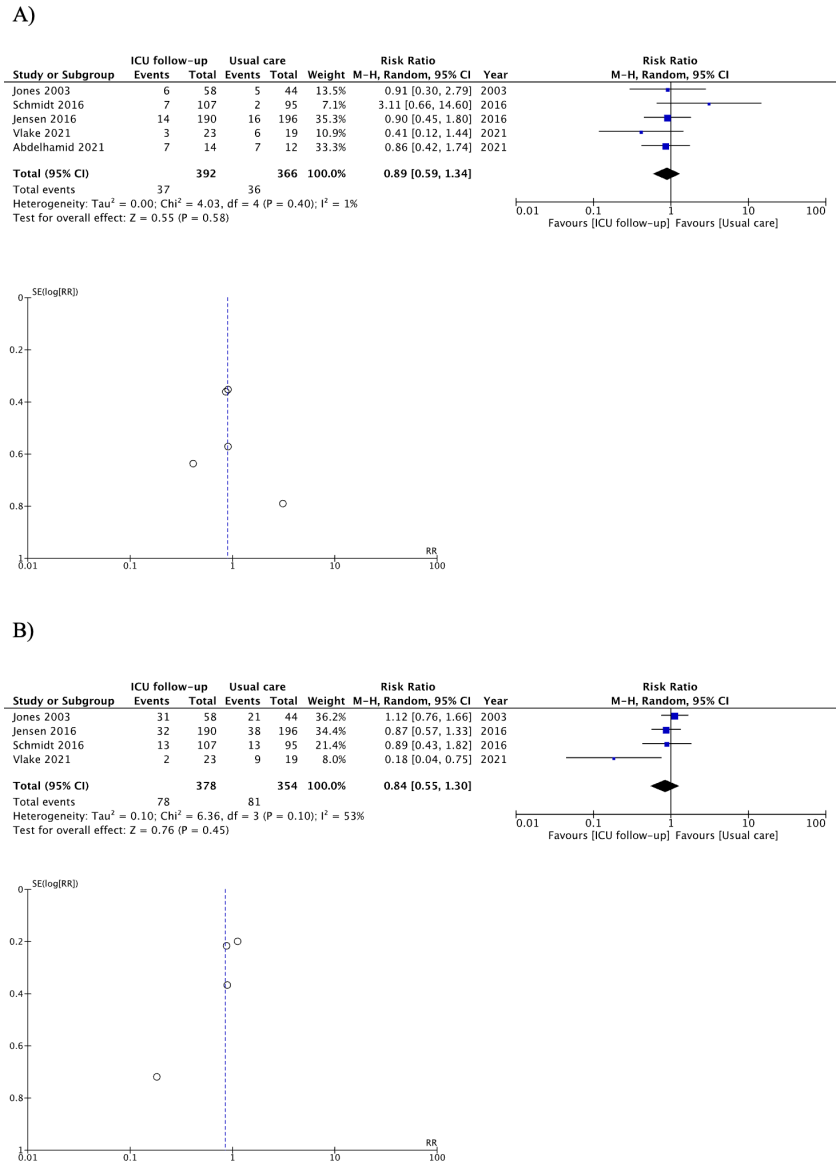


Figure 2 **Forest plot and funnel plot of primary outcomes for patients.** (A) Depression, (B) Post-traumatic stress disorder. Adverse events were not pooled.

Full-size DOI: 10.7717/peerj.15260/fig-2

Health-related quality of life

ICU follow-ups resulted in little to no differences in the HR-QoL scores among patients (SMD 0.05, 95% CI [−0.08–0.18]; $I^2 = 0\%$; seven studies, 905 patients; low-certainty evidence) (*Abdelhamid et al., 2021*; *Cox et al., 2018*; *Cox et al., 2019*; *Cuthbertson et al., 2009*; *Jensen et al., 2016*; *Schmidt et al., 2016*; *Vlake et al., 2021*); no significant heterogeneity was detected (Table 2 and Fig. S3). Of the seven studies, four measured the HR-QoL using the Mental Component Summary (MCS) of the Short-Form-36 (SF-36) (*Abdelhamid et al., 2021*; *Cuthbertson et al., 2009*; *Jensen et al., 2016*; *Schmidt et al., 2016*), one study used the MCS of the SF-12 (*Vlake et al., 2021*), and two studies used the EuroQoL Visual Analogue

Table 2 **Summary of findings for patients.**

ICU follow-up compared to usual care for critically ill patients

Patient or population: Critically ill patients

Setting:

Intervention: ICU follow-up

Comparison: Usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative (95% CI)	No of (studies)	Certainty of (GRADE)	Comments
	Risk with Usual care	**Risk with ICU follow-up**				
Proportion of patients with depression	Median 114 per 1,000	101 per 1,000 (67 to 152)	RR 0.89 (0.59 to 1.34)	758 (5 RCTs)	⨁⨁◯◯ Low[a,b]	
Proportion of patients with PTSD	Median 145 per 1,000	122 per 1,000 (80 to 188)	RR 0.84 (0.55 to 1.30)	732 (4 RCTs)	⨁⨁◯◯ Low[a,b]	
All adverse events	Median 0 per 1,000	0 per 1,000 (0 to 0)	Not estimable	42 (1 RCT)	⨁◯◯◯ Very low[a,c]	
Proportion of patients with anxiety	Median 206 per 1,000	214 per 1,000 (140 to 329)	RR 1.04 (0.68 to 1.60)	488 (2 RCTs)	⨁⨁◯◯ Low[a,b]	
HR-QoL	–	SMD 0.05 higher (0.08 lower to 0.18 higher)	–	905 (8 RCTs)	⨁⨁◯◯ Low[a,b]	
Pain	–	SMD 0.08 lower (0.32 lower to 0.17 higher)	–	258 (3 RCTs)	⨁⨁◯◯ Low[a,b]	
Readmission	Median 274 per 1,000	261 per 1,000 (211 to 318)	RR 0.95 (0.77 to 1.16)	1016 (8 RCTs)	⨁⨁◯◯ Low[a,b]	

*__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).

Confidence interval, CI; health-related quality of life; HR-QoL; intensive care unit, ICU; odds ratio; OR; risk ratio RR; standardized mean difference, SMD; post-traumatic stress disorder, PTSD; randomized controlled trial, RCT.

GRADE Working Group grades of evidence

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Notes.

[a]Downgrade for a high risk of bias: Some included studies assessed presented some concerns.

[b]Downgrade for imprecision: The sample size was small.

[c]Downgrade for imprecision: Outcome was reported in only 1 study.

Scale (EQ-VAS) (*Cox et al., 2018*; *Cox et al., 2019*). The analysis of studies using the MCS of the SF-36 and the SF-12 yielded similar findings (SMD 0.04, 95% CI [−0.11–0.19]).

Pain

ICU follow-ups resulted in little to no differences in the pain scores among patients (SMD −0.08, 95% CI [−0.32, 0.17]; $I^2 = 0\%$; three studies, 258 patients; low-certainty evidence) (*Abdelhamid et al., 2021*; *Schmidt et al., 2016*; *Vlake et al., 2021*); no significant heterogeneity was detected (Table 2 and Fig. S3). One study (*Schmidt et al., 2016*) measured pain intensity using the Graded Chronic Pain Scale; one study (*Abdelhamid et al., 2021*)

used the pain comportment of the SF-36. For one study (*Vlake et al., 2021*), we obtained data for the pain comportment of the SF-12 which was converted to the VAS 100 scale *via* author inquiry.

Readmission

ICU follow-ups resulted in little to no significant in the proportion of patients readmitted to the hospital during follow-up periods (RR 0.95, 95% CI [0.77–1.16]; $I^2 = 18\%$; seven studies, 1,016 patients; low certainty evidence) (*Abdelhamid et al., 2021*; *Bloom et al., 2019*; *Cox et al., 2018*; *Cox et al., 2019*; *Daly et al., 2005*; *Jensen et al., 2016*; *McWilliams, Benington & Atkinson, 2016*); no significant heterogeneity was detected (Table 2 and Fig. S3).

Long term mortality

ICU follow-ups resulted in little to no differences in long-term mortality among patients (RR 0.95, 95% CI [0.74–1.21]; $I^2 = 0\%$; nine studies, 1,608 patients) (*Abdelhamid et al., 2021*; *Cox et al., 2018*; *Cuthbertson et al., 2009*; *Jensen et al., 2016*; *Jones et al., 2003*; *Kredentser et al., 2018*; *Schmidt et al., 2016*; *Valsøet al., 2020*; *Vlake et al., 2021*) (Fig. S3); no significant heterogeneity was detected.

Informal caregiver outcomes

Depression

Although the evidence indicated considerable uncertainty, ICU follow-ups increased the prevalence rate of depressive symptoms—measured using the HADS-D—among informal caregivers (RR 1.58 95% CI [1.01–2.46]; one study, 188 caregivers; very low-certainty evidence) (*Kentish-Barnes et al., 2017*) (Table 3). However, the other two studies (*Bohart et al., 2019*; *Cox et al., 2018*) did not report the proportion of informal caregivers with depressive symptoms, but instead provided their HADS-D scores. The point estimate of HADS-D score was higher in the ICU follow-up groups than control; thus, no inconsistencies were observed.

Post-traumatic stress disorder

Although the evidence indicated considerable uncertainty, ICU follow-ups increased the prevalence rate of PTSD symptoms—measured using the IES-R—among informal caregivers (RR 1.36, 95% CI [0.91–2.03]; $I^2 = 19\%$; two studies, 303 caregivers; very low certainty of evidence) (*Bohart et al., 2019*; *Kentish-Barnes et al., 2017*) (Fig. 3 and Table 3); we detected slight heterogeneity. Planned sensitivity analysis of studies using the IES-R showed that ICU follow-ups significantly increased the proportion of patients with PTSD(RR 1.51, 95% CI [1.09–2.09]) (Fig. S4). One study (*Cox et al., 2018*) measured the IES-R scores and not the proportion of informal caregivers with PTSD; the point estimate of the IES-R scores was higher for the ICU follow-up group. In a sub-analysis, we found that only caregivers with non-survivors developed PTSD owing to ICU follow-ups (Fig. S4). In another sub-analysis, there was no difference in the endpoint to measure PTSD symptoms between 6 and 12 months.

Adverse events

Eligible studies with informal caregivers did not define any adverse events (Table 3).

Table 3 **Summary of findings for informal caregivers.**

ICU follow-up compared to usual care for caregivers of critically ill patients

Patient or population: Caregivers of critically ill patients

Setting:

Intervention: ICU follow-up

Comparison: Usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative (95% CI)	No of (Studies)	Certainty of (GRADE)	Comments
	Risk with Usual care	**Risk with ICU follow-up**				
Proportion of caregivers with depression	Median 242 per 1,000	382 per 1,000 (244 to 595)	RR 1.58 (1.01 to 2.46)	188 (1 RCT)	⨁◯◯◯ Very low[a,b]	
Proportion of caregivers with PTSD	Median 352 per 1,000	478 per 1,000 (320 to 714)	RR 1.36 (0.91 to 2.03)	303 (2 RCTs)	⨁◯◯◯ Very low[a,b]	
All adverse events	Not pooled	Not pooled	Not pooled	(0 RCTs)	–	
Proportion of caregivers with anxiety	Median 318 per 1,000	372 per 1,000 (264 to 518)	RR 1.17 (0.83 to 1.63)	272 (2 RCTs)	⨁◯◯◯ Very low[a,b]	
HR-QoL	–	SMD 0.07 lower (0.41 lower to 0.27 higher)	-	133 (2 RCTs)	⨁◯◯◯ Very low[a,c]	

*__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).

Confidence interval, CI; health-related quality of life; HR-QoL; intensive care unit, ICU; risk ratio RR; standardized mean difference, SMD; post-traumatic stress disorder, PTSD; randomized controlled trial, RCT.

GRADE Working Group grades of evidence

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Notes.

[a]Downgrade for a high risk of bias: This intervention was not able to blind the assessors because of both the nature of intervention and the use of self-reported outcomes.

[b]Downgrade for imprecision: The sample size was small.

[c]Downgrade for imprecision: CI included possibility of both reasonable benefit and harm.

Anxiety

Although the evidence indicated considerable uncertainty, ICU follow-ups increased the prevalence rate of anxiety symptoms, measured using the Anxiety subscale of the HADS (HADS-A), among informal caregivers (RR 1.17, 95% CI 0.83 to 1.63; two studies, 272 caregivers; very low-certainty evidence) (*Jones et al., 2004*; *Kentish-Barnes et al., 2017*) (Table 3 and Fig. S5); no significant heterogeneity was detected ($I^2 = 0\%$). One study (*Cox et al., 2018*) measured the HADS-A scores and not the proportion of caregivers with anxiety; the point estimate of the HADS-A scores was higher for the ICU follow-up group.

Health-related quality of life

Although the evidence indicated considerable uncertainty, ICU follow-ups had little to no effect on the HR-QoL measured using the MCS of the SF-36 among informal caregivers (MD −0.70, 95% CI [−4.51, 3.11]; $I^2 = 0\%$; two studies, 133 caregivers; very low certainty

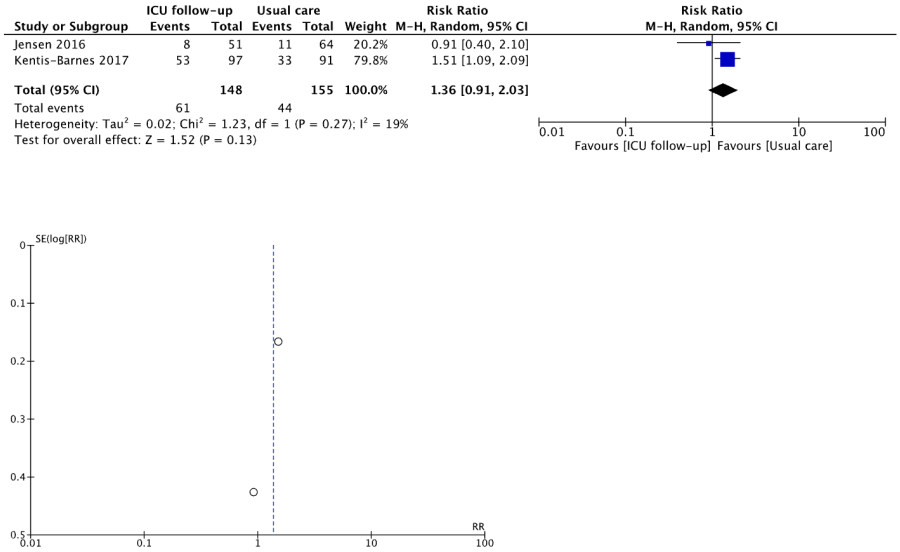


Figure 3 **Forest plot and funnel plot of primary outcomes for informal caregivers.** Post-traumatic stress disorder. Since the outcome of depression was reported in only 1 RCT, we do not show the forest plot and funnel plot. Adverse events were not pooled.

Full-size DOI: 10.7717/peerj.15260/fig-3

of evidence); no significant heterogeneity was detected (*Ågren et al., 2019*; *Bohart et al., 2019*) (Table 3 and Fig. S5).

DISCUSSION

Our SR/MA revealed that ICU follow-ups did not decrease the prevalence of depression, PTSD, and anxiety among patients. On the contrary, ICU follow-ups increased the prevalence of depression and PTSD among informal caregivers; however, there was low certainty of evidence. Furthermore, sensitivity and sub- analyses yielded similar results. Although the certainty of the evidence was low, the ICU follow-up did not decrease pain among patients.

The follow-up initiated after ICU discharge did not reduce psychological dysfunction among critically-ill patients. A Cochrane SR focusing on ICU survivors included four RCTs and concluded that the evidence for the efficacy of post-ICU follow-ups was insufficient (*Schofield-Robinson et al., 2018*). Our SR/MA revealed the ineffectiveness of post-ICU follow-ups for depression and anxiety with greater certainty than the Cochrane SR (*Schofield-Robinson et al., 2018*). The National Institute for Health and Clinical Excellence guidelines (*National Institute for Health and Care Excellence, 2009*) suggested that medical staff should conduct psychological intervention to monitor and develop preventive or treatment strategies for psychological dysfunction. However, our findings contradicted this guideline. Two reasons may explain this finding. First, the intervention content differed. The guideline (*National Institute for Health and Care Excellence, 2009*) was based on interventions comprised of enhanced or individualized physical rehabilitation; however, we focused on psychological intervention and excluded interventions pertaining to

mobilization. Second, the timings of initiation of interventions were different. The guideline (*National Institute for Health and Care Excellence, 2009*) suggested that medical staff might be suitable to assess the need for patient rehabilitation before ICU discharge; however, we focused on interventions initiated after ICU discharge and interventions for psychological dysfunction. Considering our findings, follow-ups focusing on psychological intervention initiated after ICU discharge need not be conducted for patients.

The current approaches to psychological intervention after ICU discharge were not helpful for patients and led to increased depression, PTSD, and anxiety in informal caregivers. Patients and informal caregivers have high levels of depression, anxiety, and PTSD, and the current approaches fail to address this, though it is important to screen for all components of PICS. The guidelines published by the European Resuscitation Council and the European Society of Intensive Care Medicine pertained to cardiac arrests among adults (*Nolan et al., 2021*). Based on qualitative synthesis, the guideline panel suggested that medical staff should monitor and provide information about psychological problems among informal caregivers following patients' hospital discharge (*Nolan et al., 2021*). Our SR scoped the only RCTs as a more rigorous study design with narrower eligible criteria than that of the previous SR (*Rosa et al., 2019*). As for the effect of ICU follow-up on psychological symptoms, our meta-analysis conclusions contradicted that of the previous SRs accordingly (*Cherak et al., 2021*; *Rosa et al., 2019*). This could be because of the differences in the target informal caregivers as well as the different design used in the two SRs. A recent SR showed that mental health interventions after ICU discharge may alleviate psychological problems among informal caregivers (*Cherak et al., 2021*). The primary relationship between informal caregivers and patients in the previous SR was that of parents of children. The primary informal caregivers of critically ill adults in our SR/MA were spouses, so the intervention to reduce psychological modulation in our SR was different from that of the previous SR. Moreover, the SR included quasi-experimental and uncontrolled trials and did not conduct sub-analyses of the relationship with patients. These reasons could lead to negative results. Although it is necessary to monitor psychological dysfunction among informal caregivers, follow-ups might have both positive and harmful effects on depression, PTSD, and anxiety among informal caregivers (after the ICU discharge) of adult patients.

Further research must generate a risk assessment model and other interventions to reduce psychological dysfunction and alleviate the intensity of risk factors among patients and their informal caregivers in the high-risk group. The prevalence of depression and PTSD among patients in the usual care group in our SR/MA was lower after 12 months from ICU discharge compared to patients in previous reviews (*Parker et al., 2015*; *Rabiee et al., 2016*). Furthermore, although the guidelines (*National Institute for Health and Care Excellence, 2009*) suggested the need for risk assessment of psychological dysfunction among critically-ill patients, we find no risk assessment model suitable for psychological dysfunction. Previous studies showed that pain was associated with psychological dysfunction among patients in the ICU (*Puntillo et al., 2018*) and persisted after ICU discharge (*Kemp et al., 2019*); thus, pain could be one of the risk factors for psychological dysfunction. It is unclear whether follow-up would reduce pain or the risk (of psychological dysfunction)

associated with factors like pain. Additionally, our eligible studies excluded patients with cognitive impairments due to the nature of the intervention. One cohort study reported that symptoms of PICS overlapped (*Marra et al., 2018*). Patients and their informal caregivers with cognitive impairments might not be able to find and avoid psychological intervention by themselves. Thus, in a future study, we should develop an effective intervention for participants with a high-risk of PICS.

Our SR/MA had several strengths. First, we searched databases like APA PsycInfo (Ovid), which covered the psychiatric domain, in addition to guidelines and citations *via* Google Scholar. Second, we conducted sensitivity and sub-analysis based on pre-registered protocols, yielding interesting findings. However, we could not verify the results for all primary outcomes owing to the small number of eligible studies. Third, several studies included in this SR/MA were well-designed except for the nature of the intervention. Finally, our definitions for the critical outcome measures were based on core outcomes among critically ill patients.

However, several limitations of our SR/MA need to be acknowledged. First, our search strategy involved using keywords for outcome measures instead of intervention strategies. Searches using outcome keywords might result in more favorable outcomes for intervention (*Tsujimoto et al., 2021*). Nevertheless, our SR/MA found negative results for the effectiveness of ICU follow-ups. Second, the attrition of participants in all eligible studies was higher than 20%. As participants who developed psychological dysfunction tended to withdraw from the studies, the compliance of participants with the needs of follow-ups decreased. Finally, there were several issues that require further investigation. Most reviewed studies did not report adverse events, which was a critical outcome measure for ICU survivors and their families. We could not verify the effective initiation, period, and type of intervention as they were outside the scope of our SR/MA. Similarly, the researchers' experiences were unknown.

CONCLUSION

We conducted a systematic review and meta-analysis for ICU follow-ups initiated after ICU discharge, focusing on psychological intervention. We found that ICU follow-ups did not decrease the risk of psychological dysfunction and readmission among patients. The evidence of the effect of ICU follow-up on adverse events among patients was insufficient. Similarly, there was insufficient evidence for the effect of ICU follow-ups among informal caregivers. Future studies should focus on ICU follow-ups for high-risk patients and informal caregivers of surviving patients to monitor in order to prevent the development of psychological dysfunction.

ACKNOWLEDGEMENTS

We greatly appreciate Sara L. Douglas, Ronald L. Hickman Jr., Johan H. Vlake, V.R.M. Moulaert, Kimberley Haines, Konrad Schmidt, Nancy Kentish-Barnes, Christina Jones, and Janet F Jensen for providing additional information about their studies. We also appreciate

Koichi Mino for providing information for this study. The authors greatly appreciate Editage for English language editing.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by JSPS KAKENHI Grant Number JP18K17719. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
JSPS KAKENHI: JP18K17719.

Competing Interests

Shunsuke Taito, Yusuke Tsutsumi, and Yuki Kataoka are affiliated Scientific Research WorkS Peer Support Group (SRWS-PSG), Osaka, JAPAN, which is an academic research group. Kota Yamauchi is employed by the Steel Memorial Yawata Hospital. Yuki Kataoka is employed by the Kyoto Min-iren Asukai Hospital. Yusuke Tsutsumi is employed by the National Hospital Organization Mito Medical Center. The authors declare there are no competing interests.

Author Contributions

- Shodai Yoshihiro conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Shunsuke Taito conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Kota Yamauchi conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Shunsuke Kina conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Takero Terayama conceived and designed the experiments, prepared figures and/or tables, and approved the final draft.
- Yusuke Tsutsumi conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Yuki Kataoka conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Takeshi Unoki conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw measurements are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.15260#supplemental-information.

REFERENCES

Ågren S, Eriksson A, Fredrikson M, Hollman-Frisman G, Orwelius L. 2019. The health promoting conversations intervention for families with a critically ill relative: a pilot study. *Intensive and Critical Care Nursing* **50**:103–110 DOI 10.1016/j.iccn.2018.04.007.

Abdelhamid YAli, Phillips LK, White MG, Presneill J, Horowitz M, Deane AM. 2021. Survivors of intensive care with type 2 diabetes and the effect of shared-care follow-up clinics: the SWEET-AS Randomized Controlled Pilot Study. *Chest* **159**:174–185 DOI 10.1016/j.chest.2020.08.011.

Angus DC, Carlet J. 2003. Surviving intensive care: a report from the 2002 Brussels Roundtable. *Intensive Care Medicine* **29**:368–377 DOI 10.1007/s00134-002-1624-8.

Bloom SL, Stollings JL, Kirkpatrick O, Wang L, Byrne DW, Sevin CM, Semler MW. 2019. Randomized clinical trial of an ICU recovery pilot program for survivors of critical illness. *Critical Care Medicine* **47**:1337–1345 DOI 10.1097/ccm.0000000000003909.

Bohart S, Egerod I, Bestle MH, Overgaard D, Christensen DF, Jensen JF. 2019. Reprint of Recovery programme for ICU survivors has no effect on relatives' quality of life: secondary analysis of the RAPIT-study. *Intensive and Critical Care Nursing* **50**:111–117 DOI 10.1016/j.iccn.2018.11.009.

Brandao Barreto B, Luz M, Do Amaral Lopes SAV, Rosa RG, Gusmao-Flores D. 2021. Exploring family members' and health care professionals' perceptions on ICU diaries: a systematic review and qualitative data synthesis. *Intensive Care Medicine* **47**:737–749 DOI 10.1007/s00134-021-06443-w.

Chen Y, Wang R, Yu J, Zhu L, Lu Y, Deng X. 2022. Effects of MBSR therapy on negative emotions, fatigue, and sleep quality in post-ICU patients: a randomized controlled clinical trial protocol. *Medicine* **101**:e28331 DOI 10.1097/md.0000000000028331.

Cherak SJ, Rosgen BK, Amarbayan M, Wollny K, Doig CJ, Patten SB, Stelfox HT, Fiest KM. 2021. Mental health interventions to improve psychological outcomes in informal caregivers of critically ill patients: a systematic review and meta-analysis. *Critical Care Medicine* **49**:1414–1426 DOI 10.1097/ccm.0000000000005011.

Cox CE, Hough CL, Carson SS, White DB, Kahn JM, Olsen MK, Jones DM, Somers TJ, Kelleher SA, Porter LS. 2018. Effects of a telephone- and web-based coping skills training program compared with an education program for survivors of critical illness and their family members. A randomized clinical trial. *American Journal of Respiratory and Critical Care Medicine* **197**:66–78 DOI 10.1164/rccm.201704-0720OC.

Cox CE, Hough CL, Jones DM, Ungar A, Reagan W, Key MD, Gremore T, Olsen MK, Sanders L, Greeson JM, Porter LS. 2019. Effects of mindfulness training programmes delivered by a self-directed mobile app and by telephone compared

with an education programme for survivors of critical illness: a pilot randomised clinical trial. *Thorax* **74**:33–42 DOI 10.1136/thoraxjnl-2017-211264.

Cuthbertson BH, Rattray J, Campbell MK, Gager M, Roughton S, Smith A, Hull A, Breeman S, Norrie J, Jenkinson D, Hernández R, Johnston M, Wilson E, Waldmann C. 2009. The PRaCTICaL study of nurse led, intensive care follow-up programmes for improving long term outcomes from critical illness: a pragmatic randomised controlled trial. *The BMJ* **339**:b3723 DOI 10.1136/bmj.b3723.

Daly BJ, Douglas SL, Kelley CG, O'Toole E, Montenegro H. 2005. Trial of a disease management program to reduce hospital readmissions of the chronically critically ill. *Chest* **128**:507–517 DOI 10.1378/chest.128.2.507.

DeForge CE, George M, Baldwin MR, South K, Beauchemin M, McHugh ME, Smaldone A. 2022. Do interventions improve symptoms among ICU surrogates facing end-of-life decisions? A prognostically-enriched systematic review and meta-analysis. *Critical Care Medicine* **50**:e779-e790 DOI 10.1097/ccm.0000000000005642.

Douglas SL, Daly BJ, Kelley CG, O'Toole E, Montenegro H. 2005. Impact of a disease management program upon caregivers of chronically critically ill patients; A cost-effectiveness analysis of a health education programme for elderly persons with age-related macular degeneration: a longitudinal study. *Chest* **128(6)**:3925–3936 DOI 10.1378/chest.128.6.3925.

Douglas SL, Daly BJ, Kelley CG, O'Toole E, Montenegro H. 2007. Chronically critically ill patients: health-related quality of life and resource use after a disease management intervention. *American Journal of Critical Care* **16**:447–457 DOI 10.4037/ajcc2007.16.5.447.

Ettenberger M, Rojas Cárdenas C, Parker M, Odell-Miller H. 2017. Family-centred music therapy with preterm infants and their parents in the Neonatal Intensive Care Unit (NICU) in Colombia—a mixed-methods study. *Nordic Journal of Music Therapy* **26**:207–234 DOI 10.1080/08098131.2016.1205650.

Ewens B, Myers H, Whitehead L, Seaman K, Sundin D, Hendricks J. 2019. A web-based recovery program (ICUTogether) for intensive care survivors: protocol for a randomized controlled trial. *JMIR Research Protocols* **8**:e10935 DOI 10.2196/10935.

Fotiou C, Vlastarakos PV, Bakoula C, Papagaroufalis K, Bakoyannis G, Darviri C, Chrousos G. 2016. Parental stress management using relaxation techniques in a neonatal intensive care unit: a randomised controlled trial. *Intensive and Critical Care Nursing* **32**:20–28 DOI 10.1016/j.iccn.2015.08.006.

Friedman D, Grimaldi L, Cariou A, Aegerter P, Gaudry S, Ben Salah A, Oueslati H, Megarbane B, Meunier-Beillard N, Quenot JP, Schwebel C, Jacob L, Robin Lagandré S, Kalfon P, Sonneville R, Siami S, Mazeraud A, Sharshar T. 2022. Impact of a Postintensive Care Unit Multidisciplinary Follow-up on the Quality of Life (SUIVI-REA): protocol for a Multicenter Randomized Controlled Trial. *JMIR Research Protocols* **11**:e30496 DOI 10.2196/30496.

Gawlytta R, Kesselmeier M, Scherag A, Niemeyer H, Böttche M, Knaevelsrud C, Rosendahl J. 2020. Internet-based cognitive-behavioral writing therapy for reducing posttraumatic stress after severe sepsis in patients and their spouses

(REPAIR): results of a randomized controlled trial. *Reseach Square* **12(3)**:e050305 DOI 10.21203/rs.3.rs-74259/v1.

Gawlytta R, Niemeyer H, Böttche M, Scherag A, Knaevelsrud C, Rosendahl J. 2017. Internet-based cognitive-behavioural writing therapy for reducing post-traumatic stress after intensive care for sepsis in patients and their spouses (REPAIR): study protocol for a randomised-controlled trial. *The BMJ Open* **7**:e014363 DOI 10.1136/bmjopen-2016-014363.

Haines KJ, Holdsworth C, Cranwell K, Skinner EH, Holton S, MacLeod-Smith B, Bates S, Iwashyna TJ, French C, Booth S, Carmody J, Henningham L, Searle G, Shackell M, Maher L. 2019. Development of a peer support model using experience-based co-design to improve critical care recovery. *Critical Care Explorations* **1**:e0006 DOI 10.1097/cce.0000000000000006.

Hernández RA, Jenkinson D, Vale L, Cuthbertson BH. 2014. Economic evaluation of nurse-led intensive care follow-up programmes compared with standard care: the PRaCTICaL trial. *The European Journal of Health Economics* **15**:243–252 DOI 10.1007/s10198-013-0470-7.

Higgins J, Thomas J, Chandler J, LTC M, Page M, Welch V (eds.) 2020. *Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated 2020). Available at https://training.cochrane.org/handbook/current* (accessed on 2020).

Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, Alper BS, Meerpohl JJ, Murad MH, Ansari MT, Katikireddi SV, Östlund P, Tranæus S, Christensen R, Gartlehner G, Brozek J, Izcovich A, Schünemann H, Guyatt G. 2017. The GRADE Working Group clarifies the construct of certainty of evidence. *Journal of Clinical Epidemiology* **87**:4–13 DOI 10.1016/j.jclinepi.2017.05.006.

Jensen JF, Egerod I, Bestle MH, Christensen DF, Elklit A, Hansen RL, Knudsen H, Grode LB, Overgaard D. 2016. A recovery program to improve quality of life, sense of coherence and psychological health in ICU survivors: a multicenter randomized controlled trial, the RAPIT study. *Intensive Care Medicine* **42**:1733–1743 DOI 10.1007/s00134-016-4522-1.

Johnson CC, Suchyta MR, Darowski ES, Collar EM, Kiehl AL, Van J, Jackson JC, Hopkins RO. 2019. Psychological sequelae in family caregivers of critically iii intensive care unit patients. A systematic review. *Annals of the American Thoracic Society* **16**:894–909 DOI 10.1513/AnnalsATS.201808-540SR.

Jones C, Skirrow P, Griffiths RD, Humphris GH, Ingleby S, Eddleston J, Waldmann C, Gager M. 2003. Rehabilitation after critical illness: a randomized, controlled trial. *Critical Care Medicine* **31**:2456–2461 DOI 10.1097/01.Ccm.0000089938.56725.33.

Jones C, Skirrow P, Griffiths RD, Humphris G, Ingleby S, Eddleston J, Waldmann C, Gager M. 2004. Post-traumatic stress disorder-related symptoms in relatives of patients following intensive care. *Intensive Care Medicine* **30**:456–460 DOI 10.1007/s00134-003-2149-5.

Kemp HI, Laycock H, Costello A, Brett SJ. 2019. Chronic pain in critical care survivors: a narrative review. *British Journal of Anaesthesia* **123**:e372-e384 DOI 10.1016/j.bja.2019.03.025.

Kentish-Barnes N, Chevret S, Champigneulle B, Thirion M, Souppart V, Gilbert M, Lesieur O, Renault A, Garrouste-Orgeas M, Argaud L, Venot M, Demoule A, Guisset O, Vinatier I, Troché G, Massot J, Jaber S, Bornstain C, Gaday V, Robert R, Rigaud JP, Cinotti R, Adda M, Thomas F, Calvet L, Galon M, Cohen-Solal Z, Cariou A, Azoulay E. 2017. Effect of a condolence letter on grief symptoms among relatives of patients who died in the ICU: a randomized clinical trial. *Intensive Care Medicine* **43**:473–484 DOI 10.1007/s00134-016-4669-9.

Khan S, Biju A, Wang S, Gao S, Irfan O, Harrawood A, Martinez S, Brewer E, Perkins A, Unverzagt FW, Lasiter S, Zarzaur B, Rahman O, Boustani M, Khan B. 2018. Mobile critical care recovery program (m-CCRP) for acute respiratory failure survivors: study protocol for a randomized controlled trial. *Trials* **19**:94 DOI 10.1186/s13063-018-2449-2.

Kredentser MS, Blouw M, Marten N, Sareen J, Bienvenu OJ, Ryu J, Beatie BE, Logsetty S, Graff LA, Eggertson S, Sweatman S, Debroni B, Cianflone N, Arora RC, Zarychanski R, Olafson K. 2018. Preventing posttraumatic stress in icu survivors: a single-center pilot randomized controlled trial of ICU diaries and psychoeducation. *Critical Care Medicine* **46**:1914–1922 DOI 10.1097/ccm.0000000000003367.

Major ME, Kwakman R, Kho ME, Connolly B, McWilliams D, Denehy L, Hanekom S, Patman S, Gosselink R, Jones C, Nollet F, Needham DM, Engelbert RH, Schaaf Mvander. 2016. Surviving critical illness: what is next? An expert consensus statement on physical rehabilitation after hospital discharge. *Critical Care* **20**:354 DOI 10.1186/s13054-016-1508-x.

Marra A, Pandharipande PP, Girard TD, Patel MB, Hughes CG, Jackson JC, Thompson JL, Chandrasekhar R, Ely EW, Brummel NE. 2018. Co-occurrence of post-intensive care syndrome problems among 406 survivors of critical illness. *Critical Care Medicine* **46**:1393–1401 DOI 10.1097/ccm.0000000000003218.

Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. 2018. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. *Research Synthesis Methods* **9**:602–614 DOI 10.1002/jrsm.1287.

McWilliams DJ, Benington S, Atkinson D. 2016. Outpatient-based physical rehabilitation for survivors of prolonged critical illness: a randomized controlled trial. *Physiotherapy: Theory and Practice* **32**:179–190 DOI 10.3109/09593985.2015.1137663.

Mörelius E, Örtenstrand A, Theodorsson E, Frostell A. 2015. A randomised trial of continuous skin-to-skin contact after preterm birth and the effects on salivary cortisol, parental stress, depression, and breastfeeding. *Early Human Development* **91**:63–70 DOI 10.1016/j.earlhumdev.2014.12.005.

Moulaert VR, Van Heugten CM, Winkens B, Bakx WG, De Krom MC, Gorgels TP, Wade DT, Verbunt JA. 2015. Early neurologically-focused follow-up after cardiac arrest improves quality of life at one year: a randomised controlled trial. *International Journal of Cardiology* **193**:8–16 DOI 10.1016/j.ijcard.2015.04.229.

National Institute for Health and Care Excellence. 2009. Rehabilitation after critical illness in adults Clinical guideline [CG83]. *Available at https://www.nice.org.uk/guidance/cg83* (accessed on 26 May 2021).

Needham DM, Davidson J, Cohen H, Hopkins RO, Weinert C, Wunsch H, Zawistowski C, Bemis-Dougherty A, Berney SC, Bienvenu OJ, Brady SL, Brodsky MB, Denehy L, Elliott D, Flatley C, Harabin AL, Jones C, Louis D, Meltzer W, Muldoon SR, Palmer JB, Perme C, Robinson M, Schmidt DM, Scruth E, Spill GR, Storey CP, Render M, Votto J, Harvey MA. 2012. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. *Critical Care Medicine* **40**:502–509 DOI 10.1097/CCM.0b013e318232da75.

Needham DM, Sepulveda KA, Dinglas VD, Chessare CM, Friedman LA, Bingham 3rd CO, Turnbull AE. 2017. Core outcome measures for clinical research in acute respiratory failure survivors. An international modified delphi consensus study. *American Journal of Respiratory and Critical Care Medicine* **196**:1122–1130 DOI 10.1164/rccm.201702-0372OC.

Nikayin S, Rabiee A, Hashem MD, Huang M, Bienvenu OJ, Turnbull AE, Needham DM. 2016. Anxiety symptoms in survivors of critical illness: a systematic review and meta-analysis. *General Hospital Psychiatry* **43**:23–29 DOI 10.1016/j.genhosppsych.2016.08.005.

Nolan JP, Sandroni C, Böttiger BW, Cariou A, Cronberg T, Friberg H, Genbrugge C, Haywood K, Lilja G, Moulaert VRM, Nikolaou N, Olasveengen TM, Skrifvars MB, Taccone F, Soar J. 2021. European resuscitation council and european society of intensive care medicineicine guidelines 2021: post-resuscitation care. *Intensive Care Medicine* **47**:369–421 DOI 10.1007/s00134-021-06368-4.

Ojeda A, Calvo A, Cuñat T, Artigas RM, Comino-Trinidad O, Aliaga J, Arias M, Ahuir M, Ferrando C, Dürsteler C. 2021. Rationale and study design of an early care, therapeutic education, and psychological intervention program for the management of post-intensive care syndrome and chronic pain after COVID-19 infection (PAIN-COVID): study protocol for a randomized controlled trial. *Trials* **22**:486 DOI 10.1186/s13063-021-05463-7.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *The BMJ* **372**:n71 DOI 10.1136/bmj.n71.

Parker AM, Sricharoenchai T, Raparla S, Schneck KW, Bienvenu OJ, Needham DM. 2015. Posttraumatic stress disorder in critical illness survivors: a metaanalysis. *Critical Care Medicine* **43**:1121–1129 DOI 10.1097/ccm.0000000000000882.

Puntillo KA, Max A, Timsit JF, Ruckly S, Chanques G, Robleda G, Roche-Campo F, Mancebo J, Divatia JV, Soares M, Ionescu DC, Grintescu IM, Maggiore SM, Rusinova K, Owczuk R, Egerod I, Papathanassoglou EDE, Kyranou M, Joynt GM, Burghi G, Freebairn RC, Ho KM, Kaarlola A, Gerritsen RT, Kesecioglu J, Sulaj MMS, Norrenberg M, Benoit DD, Seha MSG, Hennein A, Pereira FJ, Benbenishty JS, Abroug F, Aquilina A, Monte JRC, An Y, Azoulay E. 2018. Pain

distress: the negative emotion associated with procedures in ICU patients. *Intensive Care Medicine* **44**:1493–1501 DOI 10.1007/s00134-018-5344-0.

Rabiee A, Nikayin S, Hashem MD, Huang M, Dinglas VD, Bienvenu OJ, Turnbull AE, Needham DM. 2016. Depressive symptoms after critical illness: a systematic review and meta-analysis. *Critical Care Medicine* **44**:1744–1753 DOI 10.1097/ccm.0000000000001811.

Rohr M, Brandstetter S, Bernardi C, Fisser C, Drewitz KP, Brunnthaler V, Schmidt K, Malfertheiner MV, Apfelbacher CJ. 2021. Piloting an ICU follow-up clinic to improve health-related quality of life in ICU survivors after a prolonged intensive care stay (PINA): study protocol for a pilot randomised controlled trial. *Pilot and Feasibility Studies* **7**:90 DOI 10.1186/s40814-021-00796-1.

Rosa RG, Ferreira GE, Viola TW, Robinson CC, Kochhann R, Berto PP, Biason L, Cardoso PR, Falavigna M, Teixeira C. 2019. Effects of post-ICU follow-up on subject outcomes: a systematic review and meta-analysis. *Journal of Critical Care* **52**:115–125 DOI 10.1016/j.jcrc.2019.04.014.

Schmidt KF, Schwarzkopf D, Baldwin LM, Brunkhorst FM, Freytag A, Heintze C, Reinhart K, Schneider N, Von Korff M, Worrack S, Wensing M, Gensichen J. 2020. Long-term courses of sepsis survivors: effects of a primary care management intervention. *The American Journal of Medicine* **133**:381–385.e385 DOI 10.1016/j.amjmed.2019.08.033.

Schmidt K, Worrack S, Korff MVon, Davydow D, Brunkhorst F, Ehlert U, Pausch C, Mehlhorn J, Schneider N, Scherag A, Freytag A, Reinhart K, Wensing M, Gensichen J. 2016. Effect of a primary care management intervention on mental health-related quality of life among survivors of sepsis: a randomized clinical trial. *JAMA* **315**:2703–2711 DOI 10.1001/jama.2016.7207.

Schofield-Robinson OJ, Lewis SR, Smith AF, McPeake J, Alderson P. 2018. Follow-up services for improving long-term outcomes in intensive care unit (ICU) survivors. *Cochrane Database of Systematic Reviews* **11**:Cd012701 DOI 10.1002/14651858.CD012701.pub2.

Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett MS, Eldridge SM, Emberson JR, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. 2019. RoB 2: a revised tool for assessing risk of bias in randomised trials. *The BMJ* **366**:l4898 DOI 10.1136/TheBMJ.l4898.

Tsujimoto Y, Tsutsumi Y, Kataoka Y, Banno M, Furukawa TA. 2021. Around ten percent of most recent Cochrane reviews included outcomes in their literature search strategy and were associated with potentially exaggerated results: a research-on-research study. *Journal of Clinical Epidemiology* **141**:74–81 DOI 10.1016/j.jclinepi.2021.08.030.

Valsø Å, Rustøen T, Småstuen MC, Ekeberg Ø, Skogstad L, Schou-Bredal I, Myhren H, Sunde K, Tøien K. 2020. Effect of nurse-led consultations on post-traumatic

stress and sense of coherence in discharged icu patients with clinically relevant post-traumatic stress symptoms—a randomized controlled trial. *Critical Care Medicine* **48**:e1218-e1225 DOI 10.1097/ccm.0000000000004628.

Vlake JH, Van Bommel J, Wils EJ, Korevaar TIM, Bienvenu OJ, Klijn E, Gommers D, Van Genderen ME. 2021. Virtual reality to improve sequelae of the postintensive care syndrome: a multicenter, randomized controlled feasibility study. *Critical Care Explorations* **3**:e0538 DOI 10.1097/cce.0000000000000538.

Yoshihiro S, Taito S, Yamauchi K, Kina S, Terayama T, Tsutsumi Y, Kataoka Y, Unoki T. 2021. Efficacy of follow-up after intensive care unit (ICU) discharge for improving long-term outcomes and mental health in ICU patients and informal caregivers: a systematic review and meta-analysis protocol. *Available at https://dx.doi.org/10.17504/protocols.io.bvjwn4pe* (accessed on 8 June 2021).

Received: 14 December 2022 | Revised: 5 June 2023 | Accepted: 7 June 2023

DOI: 10.1002/jrsm.1649

RESEARCH ARTICLE

Research Synthesis Methods WILEY

An open competition involving thousands of competitors failed to construct useful abstract classifiers for new diagnostic test accuracy systematic reviews

Yuki Kataoka[1,2,3,4] | Shunsuke Taito[2,5] | Norio Yamamoto[2,6,7] | Ryuhei So[2,8,9] | Yusuke Tsutsumi[2,4,10] | Keisuke Anan[2,11,12] | Masahiro Banno[2,13,14] | Yasushi Tsujimoto[2,15,16] | Yoshitaka Wada[2,17] | Shintaro Sagami[18,19] | Hiraku Tsujimoto[20] | Takashi Nihashi[21] | Motoki Takeuchi[22] | Teruhiko Terasawa[23] | Masahiro Iguchi[24] | Junji Kumasawa[25,26] | Takumi Ichikawa[27] | Ryuki Furukawa[27] | Jun Yamabe[28] | Toshi A. Furukawa[16]

Correspondence
Toshi A. Furukawa, Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan.
Email: furukawa@kuhp.kyoto-u.ac.jp

Funding information
Fujifilm Corporation

Abstract

There are currently no abstract classifiers, which can be used for new diagnostic test accuracy (DTA) systematic reviews to select primary DTA study abstracts from database searches. Our goal was to develop machine-learning-based abstract classifiers for new DTA systematic reviews through an open competition. We prepared a dataset of abstracts obtained through database searches from 11 reviews in different clinical areas. As the reference standard, we used the abstract lists that required manual full-text review. We randomly splitted the datasets into a train set, a public test set, and a private test set. Competition participants used the training set to develop classifiers and validated their classifiers using the public test set. The classifiers were refined based on the performance of the public test set. They could submit as many times as they wanted during the competition. Finally, we used the private test set to rank the submitted classifiers. To reduce false exclusions, we used the Fbeta measure with a beta set to seven for evaluating classifiers. After the competition, we conducted the external validation using a dataset from a cardiology DTA review. We received 13,774 submissions from 1429 teams or persons over 4 months. The top-honored classifier achieved a Fbeta score of 0.4036 and a recall of 0.2352 in the external validation. In conclusion, we were unable to develop an abstract classifier with sufficient recall for immediate application to new DTA systematic reviews. Further studies are needed to update and validate classifiers with datasets from other clinical areas.

For affiliations refer to page 715

KEYWORDS

diagnostic test accuracy, machine learning, open competition, search filter, systematic review

Highlights

What Is Already Known

- For updating systematic reviews, there are some machine learning (ML)-based abstract classifiers to reduce human workload.
- There is no abstract classifier with a reasonable degree of recall required for new diagnostic test accuracy (DTA) systematic reviews to select primary DTA studies from a database search.

What Is New

- We conducted an open competition to develop abstract classifiers for new diagnostic test accuracy (DTA) systematic reviews.
- The top three best-performing classifiers showed poor recall in the external validation set with different clinical areas from the development set.

Potential Impact for *Research Synthesis Methods* Readers

- The performance of the "design-specific classifier" of DTA studies developed with a limited clinical area dataset was poor compared to the "review-specific classifier" used in the update DTA reviews.
- Open competitions can be a solution to develop machine learning classifiers for general researchers who are not necessarily familiar with machine learning.

1 | INTRODUCTION

Machine learning (ML) models are increasingly being used in the medical field. Applications of ML models in clinical practice have been prominent in the field of diagnostic imaging, such as the diagnosing bone fractures,[1] detection of breast cancer,[2] and COVID-19 diagnosis.[3] Clinical implementation is not limited to diagnosis but is expanding into alerts,[4] patient education,[5] and many other areas.

In recent years, ML models have been tested in the systematic review process.[6] In particular, ML models are being actively applied in the screening of titles and abstracts process to reduce human resources. Most approaches are done in the updating intervention reviews.[7,8] This situation is the same in diagnostic test accuracy (DTA) systematic reviews.[9]

DTA systematic reviews summarize the accuracy of diagnostic tests systematically and transparently, yet lack universally accepted study design labels, unlike randomized controlled trials (RCTs).[10] For instance, the terms "cohort study" and "case–control study" are commonly used to label DTA studies in the abstract. These terms also used for studies to investigate the association between exposures and outcomes. This will lead difficulty for classification.[11] In addition, indexing terms for RCTs are available in both MEDLINE and Embase. However, for DTA studies, the indexing term is only available in Embase, and it is known to be inadequate.[12] Due to these reasons, none of the current abstract classifiers for identifying primary DTA studies have a sufficient combination of high recall with reasonable precision required for systematic reviews. This absence necessitates a dedicated classifier to minimize the number of abstracts to read when conducting a systematic database search for new DTA systematic reviews.

In open ML competitions, the participants compete to develop the best ML model for a specific goal. Open competitions are open to anyone and have succeeded in solving various problems in the medical field.[13–15] Therefore, we conducted an open ML competition to develop DTA abstract classifiers. Our goal with the "FILtering of diagnostic TEst accuracy studies" (FILTER) challenge was to develop ML-based abstract classifiers for new DTA systematic reviews through an open competition (https://signate.jp/competitions/595).

2 | METHODS

We show the whole schema of the FILTER challenge in the Figure 1. Each dataset was a comma-separated value files (CSV), including serial numbers, titles, abstracts, and binary reference labels of true and false values. We used titles and abstracts as predictors. As the reference standard, we used the abstract lists that required manual full-text review when the original DTA systematic review was conducted. Hereafter, the term "record" includes the title and abstract, but not other bibliographic information.

Participants trained their ML-based classifiers using a training dataset. Subsequently, they submitted their classifiers to the leaderboard. On the leaderboard, the classifiers were evaluated using the public test dataset. Based on this evaluation, we released a tentative ranking. The participants tuned their classifiers based on their scores and rankings. They could submit as many times as they wanted during the competition. Finally, we evaluated classifiers using the private test dataset and determined the final ranking.

2.1 | Preparation of datasets

We defined a DTA study as an original study that evaluated a test against a clinical reference standard for humans.[11] We accepted multivariable diagnostic prediction model studies, but excluded prognostic prediction model studies, that measured predictors and outcomes at different time points.[16] We excluded modeling studies, studies that assessed diagnostic training for medical professionals, and case series (e.g., studies without controls, such as following polymerase chain reaction results of specific patients). We also excluded studies in which the abstracts were written only in languages other than English.

We downloaded the EPPI-Centre COVID-19 "living" systematic map of research on 25 December 2020.[17] The dataset included systematically indexed 33,008 records of COVID-19 studies. Among them, 1769 records were indexed as "Diagnosis". Because the "Diagnosis" category included records that meet our exclusion criteria such as modeling studies, we independently reviewed 1769 records and selected 1070 records. Our research purpose

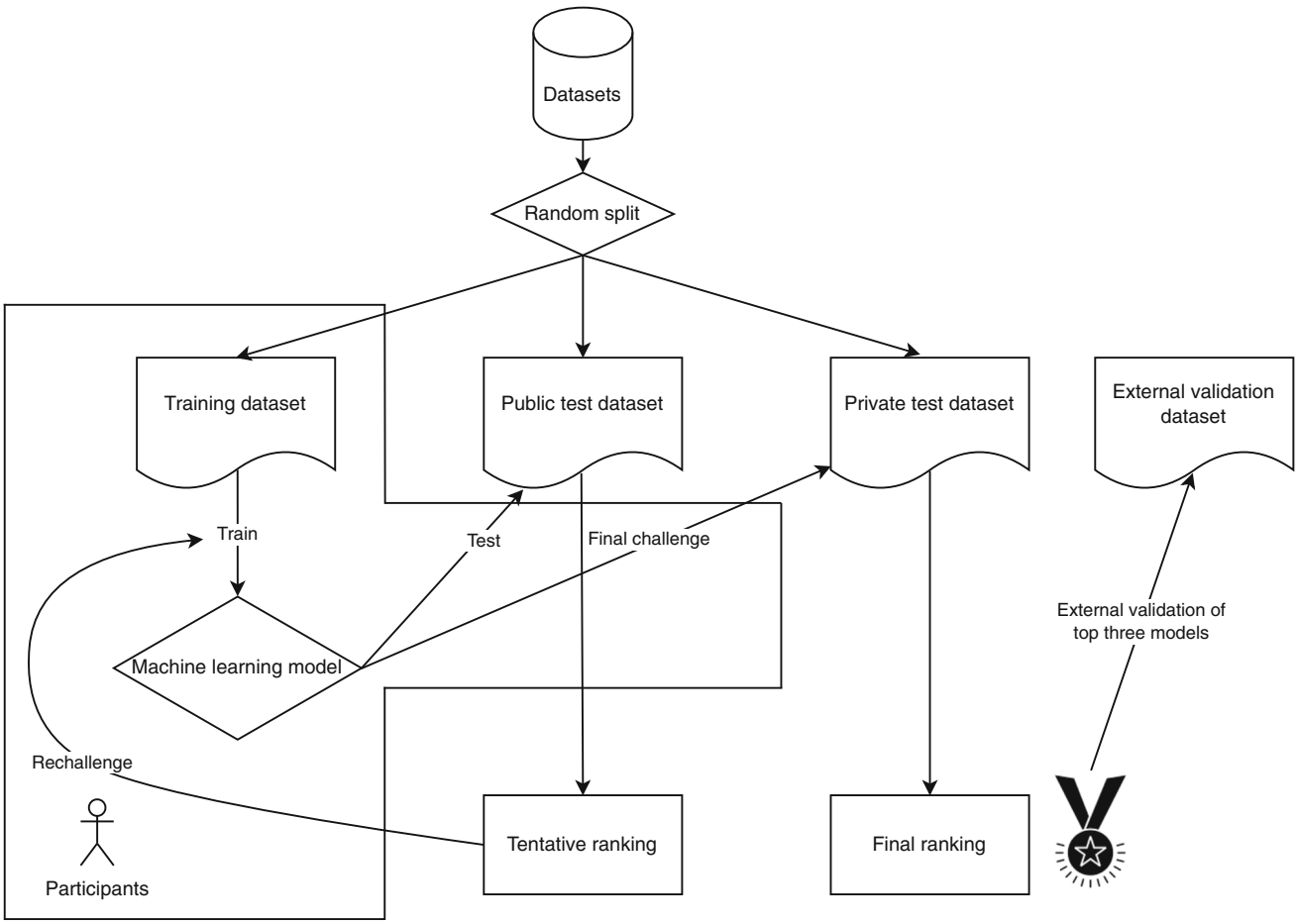


FIGURE 1 Whole schema of the FILTER challenge.

TABLE 1 Characteristics of diagnostic test accuracy systematic reviews.

Author	Population or target condition	Index test	Search date
EPPI 2020	COVID-19	Various tests	Dec 2020
Shiroshita 2020	Malignant pleural effusion	Ultrasonography	Dec 2019
Sagami 2021	Inflammatory bowel diseases	Ultrasonography	Mar 2019
Shiroshita 2020	Bird fancier's lung	Inhalation challenge tests	Nov 2019
Tsujimoto 2017	Gastric tube placement	Ultrasonography	Mar 2016
Tsutsumi 2020	Aortic dissection	Risk score	Dec 2018
Nihashi 2013	Glioma	PET	Jun 2011
Nihashi 2020	Dementia with Lewy bodies	DAT-SPECT and MIBG scintigraphy	Mar 2018
Mishima 2016	Dementia with Lewy bodies	Biomarkers	Mar 2015
Takeuchi 2016	Cause of fever of unknown origin	Nuclear imaging	Oct 2015
Iguchi 2020	Acute meningitis	Jolt accentuation	Apr 2020
Tsujimoto 2022[a]	Pulmonary hypertension	Ultrasonography	Aug 2021

[a]Used for the external validation.

was to create a filter with high recall, therefore, if we could not judge whether the record is about the DTA study or not, we judged it as a DTA record.

To address topics other than COVID-19, we used several previous DTA systematic reviews, including those on malignancy,[18] gastrointestinal disorders,[19] respiratory disorders,[20] emergency care,[21,22] neurology,[23–25] and infectious disease.[26,27] Some search data were not stored in a reusable form, and fewer records were included than those retrieved for the original studies. We included 82,359 records. Among these, 1822 records were labeled as DTA studies. The characteristics of DTA systematic reviews is shown in Table 1.

2.2 | Coding Challenge

We used the titles and records of prediction sources. After excluding duplications or missing records, we randomly split into using a 4:3:3 ratio. These sets included a train dataset (n = 27,145, labeled DTA n = 632), a public test dataset (n = 20,417, labeled DTA n = 474), and a private test dataset (n = 20,417, labeled DTA n = 469) (Figure 2).

From the competition website, the participants downloaded a CSV file containing the training dataset with true or false labels and the test dataset without labels. The test dataset consisted of a combination of both public and private test datasets. They developed classifiers in their local environments using only the training dataset. They were limited to using open and free tools for classifier training, such as Python and R. Upon predicting labels for the test dataset, they uploaded their results in CSV format to the competition website for validation. We displayed the name of participants and Fbeta scores for the test dataset portion on a publicly accessible bulletin board. The participants iteratively fine-tuned their classifiers based on the results. Finally, we used the private test dataset to rank the submissions. We required the top three winners to submit their source code, and we verified the reproducibility of their results. These datasets are publicly available at (https://osf.io/bmfne/).

2.3 | Tasks and evaluation metrics

We used the Fbeta score to evaluate the classifiers. The Fbeta score is a performance metric used in machine learning, particularly in classification tasks, to evaluate the effectiveness of a classifier in terms of both precision and recall. Precision is the proportion of true positives among records retrieved by the filter. The recall is the proportion of true positives among the records that should be retrieved by the filter. The Fbeta score is defined as: Fbeta = ((1 + beta^2) * Precision * Recall) / (beta^2 * Precision + Recall). In the Fbeta score, beta represents the balance between precision and recall. The scores range from 0 to 1, with 0 being the worst and 1 being the best.[28] Beta is the factor representing precision versus recall in the composite score. When beta = 0, only precision is considered, and as beta increases, recall is weighted at beta >1, giving recall priority over precision.

To determine beta, we set the recall at 0.96, in accordance with a previous study on methodology filters by Cochrane.[29] We ran simulations with varying precision from 0.6 to 0.9 and simultaneously varying beta. We adopted seven as the value of beta when the value of

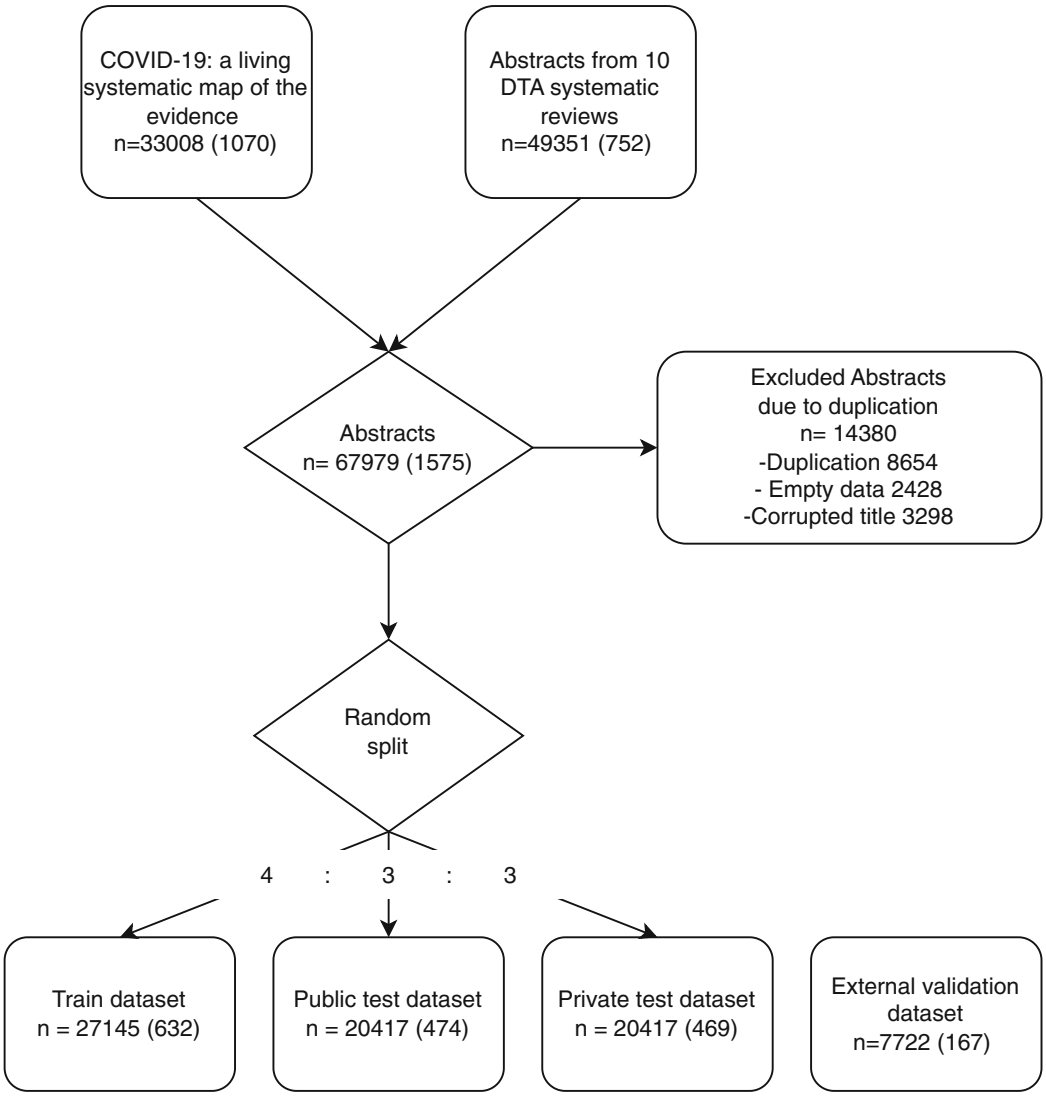


FIGURE 2 Flowchart of abstracts processing (Number of true labels).

Fbeta reached the plateau (Figure 3). We honored the top three classifiers with high Fbeta in the private dataset.

2.4 | External validation

For the external validation, we used a DTA review for cardiology (n = 7722, labeled DTA n = 167).[30] We tested the three classifiers in the same manner in the competition. We evaluated the classifiers using Fbeta, precision, and recall the using cut-off used by each classifier when submitted to competition, area under the curve (AUC), and Brier score. The AUC is a score to evaluate discrimination, which means ability to distinguish between the positive and negative classes. It is calculated as the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate against the false positive rate at various classification cut-offs. An AUC of 1.0 indicates a perfect classifier, while an AUC of 0.5 indicates a classifier that performs no better than random. In general, the higher the AUC means better discrimination.[31] The Brier score is a score to evaluate calibration, which means consistency between the predicted probabilities and the true binary outcomes. It was calculated as the average of the squared differences between the predicted probability and the true binary outcome for all instances in the test set. In general, the lower Brier score means better calibration.[32] Even if a classifier is noninformative, the maximum score decreases when the outcome proportion is lower.[33]

We depicted receiver operating characteristic curves and calibration plots. In constructing the calibration plot, we used deciles of predicted probabilities and applied a logarithmic scale on the horizontal axis due to the narrow range of predicted probabilities. For statistical analysis, we used Google Collaboratory, a Python-based data analysis and machine learning tool that can be executed in a web browser[34] and Rstudio.[35]

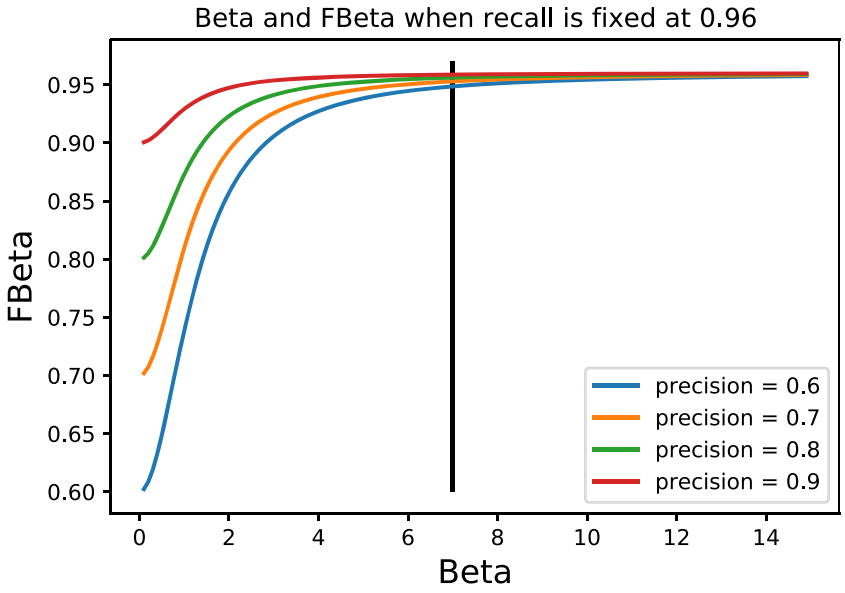


FIGURE 3 Simulation of Fbeta and beta. The black vertical bar corresponds with 7. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Honored top three classifiers.

	The original machine learning models	Fbeta in the public test dataset	Fbeta in the private test dataset	Fbeta in the external validation dataset	Precision in the external validation dataset[a]	Recall in the external validation dataset[a]
1st classifier	PubMed BERT BioM Electra SapBERT	0.9261	0.9390	0.4036	0.4096	0.2352
2nd classifier	PubMed BERT	0.9219	0.9384	0.3262	0.1858	0.3313
3rd classifier	PubMed BERT SciBERT BlueBERT BioRedditBERT	0.9285	0.9367	0.3891	0.1908	0.3976

[a]Using the predefined cut-offs at the time of classifier submission.

3 | RESULTS

3.1 | Overview of submissions

From July 28 to October 4, 2021, we held the challenge. We announced this on the competition page in Japanese (https://signate.jp/competitions/471) and advertised it through social-networking sites. We received 13,774 submissions from 1429 teams or individuals.

3.2 | Algorithm Performances

We honored the top three classifiers evaluated using Fbeta in the private test dataset (Table 2). The award winners received cash prizes of 100,000 yen for first place, 50,000 yen for second place, and 50,000 yen for third place. All three classifiers were based on the Bidirectional Encoder Representations from Transformers (BERT) and learned vocabulary related to medicine.[36] The source codes of the three classifiers are available at (https://github.com/signatelab/paper-classification-challenge-winners-solutions). The Fbeta scores, recall, AUC, and Brier score in the external validation dataset were 0.4036, 0.2352, 0.9330, and 0.0193 for the 1st classifier, 0.3262, 0.3313, 0.9263, 0.0206 for the 2nd classifier, and 0.3891, 0.3976, 0.9263, 0.0207 for the 3rd classifier, respectively (Figure 4). The calibration plots showed agreement with the predicted probability of the 1st classifier at the low probability (Figure 5).

4 | DISCUSSION

We conducted an open competition to develop a search filter studied to be included in new DTA systematic reviews. In 3 months, 1429 teams or persons participated in the FILTER challenge. The three honored classifiers

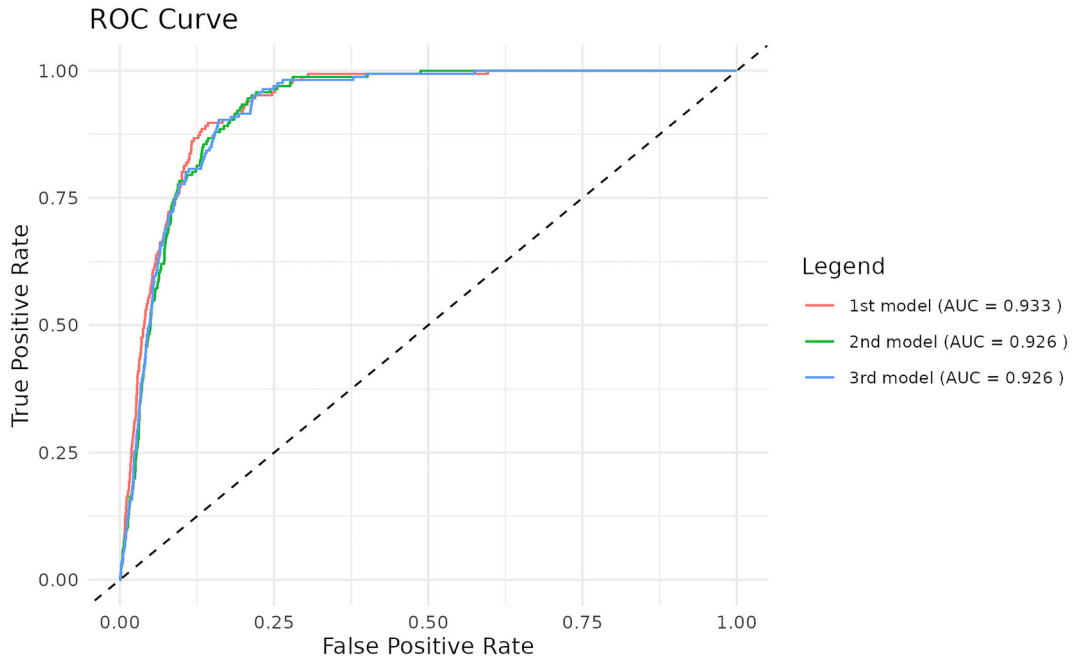


FIGURE 4 Receiver operating characteristic curves in the external validation. Classifier name (area under the curve). [Colour figure can be viewed at wileyonlinelibrary.com]

used BERT to learn medicine-related vocabulary. The three classifiers showed poor precision and recall in the external validation dataset, including records from different clinical areas, compared with the training dataset. The calibration plots showed agreement with the predicted probability of the 1st classifier in low-probability areas.

We were able to attract more than 1400 participants. Previously published medical competitions attracted less than a thousand participants in as much time as we have.[14,15,37] The number of competition participants was comparable to that of medical competitions on Kaggle, the world's largest open competition site.[38] This may be because we set up the task of natural language processing in the medical fields. Most existing competitions in the medical fields dealt with image processing. To protect personal information, it is difficult to prepare a text dataset for open competition with answers for machine learning in the medical fields. On the other hand, SRs use publicly available information, hence, privacy protection is not an issue. In addition, the prize money would have also motivated the competitors. Open competitions can be a solution for researchers who are not necessarily familiar with machine learning, if they can successfully set up an interesting challenge for a machine learning task, and prepare a dataset and prizes.

Several studies have attempted to deploy ML to update DTA reviews to reduce the classification workload and achieved a reduction of approximately 80% in screening burden.[8,39–41] Previous studies developed "review-specific classifier". In other words, both the DTA study design and the clinical areas were considered in the ML classifiers. In contrast, we intended to develop a "design-specific classifier" to be used in any DTA reviews in the FILTER challenge. However, on conducting external validation, we found that the classifier's performance was suboptimal in the challenge dataset.

The result raises concerns about overfitting,[42] and calls for a more in-depth discussion of the factors affecting the generalizability. Firstly, the classifiers might learn features dependent on the frequent words in the competition dataset. For instance, polymerase chain reaction would be frequently seen in a DTA study of infectious diseases, while pulmonary artery pressure would be frequent in a cardiology DTA study. To address this issue, the classifiers need to be updated using datasets from other clinical areas that were not used in the challenge. Publishing the search results and decisions as a dataset in an easy-to-use format would facilitate future updates of the filters. Secondary, class imbalance would cause the overfitting. Our challenge dataset contained 2.3% true labels, such imbalance could influence the learning results.[43] With regard to this issue, further research is needed to improve the learning process. Thirdly, attempts to improve reporting quality, such as the Standards for Reporting of Diagnostic Accuracy Studies statement,[11] would alter descriptive styles in included studies over time. This change could potentially affect the decisions of

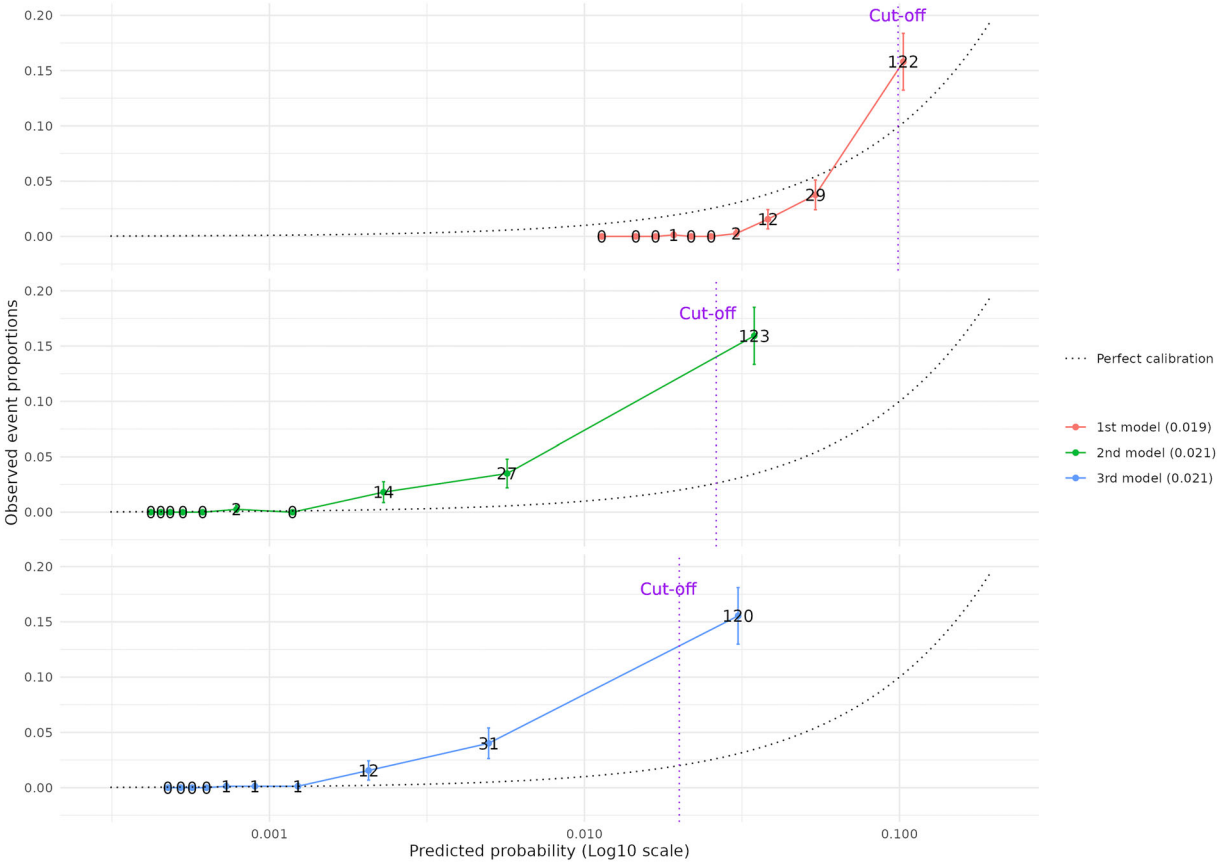


FIGURE 5 Calibration plots in the external validation. Classifier name (Brier score). Predicted probabilities were divided into deciles. The numbers represent actual event counts within deciles, and error bars indicate the 95% confidence intervals for observed event proportions. Cut-offs were adopted from the competition submission. Note that the predicted probabilities are continuous variables and the number displayed above the decile to the right of cut-off is more than the number that the classifier determined to be True. [Colour figure can be viewed at wileyonlinelibrary.com]

classifiers. In machine learning, an intrinsic issue called the "black box" problem exists, where it is fundamentally unclear which factors classifiers emphasize in their decision-making process.[44] Further research is necessary to address this issue.

Systematic reviewers will not be able to use honored classifiers as are with current cutoffs. In the external validation, a conservative approach is recommended to prevent overestimation.[45] Thus, we used predefined cut-offs during competition submission. This choice resulted in the honored classifiers performing lower than the predefined recall. Our results of external validation indicate the need to recalibrate new cutoff values for use in other datasets. Referring to the recalibrated cutoffs, systematic reviewers might be able to use the 1st classifier to sort the records in the order of predictive probability, perform title and abstract reviews, and stop the review when the predictive probability becomes quite low.

Our study had several limitations. First, we considered records that we could not determine as DTA to be DTA abstracts. There may have been some misclassifications related to the selection of records. However, the purpose of this study was to replace the work done by humans, and we will have to allow for the existence of certain errors in it. Second, other ML classifiers may provide optimal solutions for the current dataset. New natural language processing ML classifiers are developed every year and achieving the best performance.[46] Researchers will be able to use a different classifier than BERT when updating.

5 | CONCLUSIONS

We conducted an open ML competition with more than one thousand challengers. We could not develop a search filter with immediate applicability to new DTA reviews. Further studies are needed to update and validate filters with datasets from other clinical areas.

AUTHOR CONTRIBUTIONS

Shunsuke Taito: Conceptualization; data curation; writing – review and editing. **Norio Yamamoto:** Data

curation; writing – review and editing. **Ryuhei So:** Conceptualization; data curation; writing – review and editing. **Yusuke Tsutsumi:** Conceptualization; data curation; formal analysis; writing – review and editing. **Keisuke Anan:** Conceptualization; data curation; writing – review and editing. **Masahiro Banno:** Conceptualization; data curation; writing – review and editing. **Yasushi Tsujimoto:** Conceptualization; data curation; writing – review and editing. **Yoshitaka Wada:** Data curation; writing – review and editing. **Shintaro Sagami:** Data curation; writing – review and editing. **Hiraku Tsujimoto:** Data curation; writing – review and editing. **Takashi Nihashi:** Data curation; writing – review and editing. **Motoki Takeuchi:** Data curation; writing – review and editing. **Teruhiko Terasawa:** Conceptualization; data curation; methodology; visualization; writing – review and editing. **Masahiro Iguchi:** Data curation; writing – review and editing. **Junji Kumasawa:** Data curation; writing – review and editing. **Takumi Ichikawa:** Data curation; formal analysis; methodology; writing – review and editing. **Ryuki Furukawa:** Data curation; formal analysis; methodology; software; writing – review and editing. **Jun Yamabe:** Data curation; formal analysis; methodology; software; writing – review and editing.

AFFILIATIONS

[1]Department of Internal Medicine, Kyoto Min-iren Asukai Hospital, Kyoto, Japan
[2]Scientific Research Works Peer Support Group (SRWS-PSG), Osaka, Japan
[3]Section of Clinical Epidemiology, Department of Community Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
[4]Department of Healthcare Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
[5]Division of Rehabilitation, Department of Clinical Practice and Support, Hiroshima University Hospital, Hiroshima, Japan
[6]Department of Orthopedic Surgery, Miyamoto Orthopedic Hospital, Okayama, Japan
[7]Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan
[8]Department of Psychiatry, Okayama Psychiatric Medical Center, Okayama, Japan
[9]CureApp, Inc., Tokyo, Japan
[10]Department of Emergency Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan
[11]Division of Respiratory Medicine, Saiseikai Kumamoto Hospital, Kumamoto, Japan
[12]Department of Healthcare Epidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan
[13]Department of Psychiatry, Seichiryo Hospital, Nagoya, Japan
[14]Department of Psychiatry, Nagoya University Graduate School of Medicine, Nagoya, Japan
[15]Oku Medical Clinic, Osaka, Japan
[16]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto University, Kyoto, Japan
[17]Department of Rehabilitation Medicine, School of Medicine, Fujita Health University, Toyoake, Japan
[18]Center for Advanced IBD Research and Treatment, Kitasato University Kitasato Institute Hospital, Tokyo, Japan
[19]Department of Gastroenterology and Hepatology, Kitasato University Kitasato Institute Hospital, Tokyo, Japan
[20]Hospital Care Research Unit, Hyogo Prefectural Amagasaki General Medical Center, Amagasaki, Japan
[21]Department of Radiology, Komaki City Hospital, Komaki, Japan
[22]Department of Emergency and General Internal Medicine, Fujita Health University School of Medicine, Toyoake, Japan
[23]Section of General Internal Medicine, Department of Emergency and General Internal Medicine, Fujita Health University School of Medicine, Toyoake, Japan
[24]Department of Neurology, Fukushima Medical University, Fukushima, Japan
[25]Human Health Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan
[26]Department of Critical Care Medicine, Sakai City Medical Center, Sakai, Japan
[27]Yahoo Japan Corporation, Tokyo, Japan
[28]Self-employed Engineer, Tokyo, Japan

ACKNOWLEDGMENTS

We would like to thank the EPPI for access to the COVID-19 'living' systematic map of research. We acknowledge Mr. Yohei Ikenoue, Dr. Naoki Nishiyama, Mr. Yukihiro Toma, and Dr. Shigeru Saito in SIGNATE Corporation provided the platform and advice to set the rules to hold the competition. We acknowledge Dr. Munehiko Sasajima and Dr. Naoki Katoh at the University of Hyogo for their assistance with public relations. We also acknowledge Dr. Aki Mishima for retrieving the data.

FUNDING INFORMATION

We acknowledge the Fujifilm Corporation for providing prizes for excellence in the competition. The funder

played no role in the design and conduct of the study, or in the decision to prepare and submit the manuscript for publication.

CONFLICT OF INTEREST STATEMENT

Yuki Kataoka: none known. Shunsuke Taito: none known. Norio Yamamoto: none known. Ryuhei So: grants from Osake-no-Kagaku Foundation, speaker's honoraria from Otsuka Pharmaceutical Co., Ltd., Nippon Shinyaku Co., Ltd., and Takeda Pharmaceutical Co., Ltd., outside the submitted work. Yusuke Tsutsumi: none known. Masahiro Banno: none known. Keisuke Anan: none known. Yasushi Tsujimoto: none known. Yoshitaka Wada: none known. Shintaro Sagami: SS served as a speaker for Janssen Pharmaceutical, AbbVie, EA Pharma, Kyorin Pharmaceutical, Mochida Pharmaceutical, Takeda Pharmaceutical, Mitsubishi Tanabe Pharma, Nippon Kayaku, and Zeria Pharmaceutical and as an endowed chair for AbbVie, JIMRO, Zeria Pharmaceutical, Kyorin Pharmaceutical, Mochida Pharmaceutical, and EA Pharma. Hiraku Tsujimoto: HT is an employee of Rege Nephro Co., Ltd., which aims to research and develop various solutions for renal diseases. Takashi Nihashi: none known. Motoki Takeuchi: none known. Teruhiko Terasawa: none known. Masahiro Iguchi: none known. Junji Kumasawa: none known. Takumi Ichikawa: none known, this research is unconnected to his employer and does not reflect the views of his employer. Ryuki Furukawa: none known, this research is unconnected to his employer and does not reflect the views of his employer. Jun Yamabe: none known. Toshi A. Furukawa: TAF reports personal fees from DT Axis, Kyoto University Original, MSD and SONY, and a grant from Shionogi, outside the submitted work; In addition, Toshi A. Furukawa has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at (https://github.com/signatelab/paper-classification-challenge-winners-solutions) and (https://osf.io/bmfne/). Additional data that support the findings of this study are available on request from the corresponding author.

ORCID

Yuki Kataoka https://orcid.org/0000-0001-7982-5213
Shunsuke Taito https://orcid.org/0000-0003-1218-4225
Norio Yamamoto https://orcid.org/0000-0002-7902-9994
Ryuhei So https://orcid.org/0000-0002-9838-350X
Yusuke Tsutsumi https://orcid.org/0000-0002-9160-0241
Keisuke Anan https://orcid.org/0000-0002-2701-7334
Masahiro Banno https://orcid.org/0000-0002-2539-1031
Yasushi Tsujimoto https://orcid.org/0000-0002-7214-5589
Yoshitaka Wada https://orcid.org/0000-0003-2191-3629
Shintaro Sagami https://orcid.org/0000-0002-1202-8161
Hiraku Tsujimoto https://orcid.org/0000-0002-5847-4132
Junji Kumasawa https://orcid.org/0000-0003-4619-945X
Toshi A. Furukawa https://orcid.org/0000-0003-2159-3776

REFERENCES

1. Regnard N-E, Lanseur B, Ventre J, et al. Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays. *Eur J Radiol*. 2022;154(110447):110447.
2. Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med*. 2021;27(2):244-249.
3. Kataoka Y, Baba T, Ikenoue T, et al. Development and external validation of a deep learning-based computed tomography classification system for COVID-19. *Ann Clin Epidemiol*. 2022;4:110-119. advpub(22014):22014.
4. Henry KE, Adams R, Parent C, et al. Factors driving provider adoption of the TREWS machine learning-based early warning system and its effects on sepsis treatment timing. *Nat Med*. 2022;28(7):1447-1454.
5. Kataoka Y, Takemura T, Sasajima M, Katoh N. Development and early feasibility of chatbots for educating patients with lung cancer and their caregivers in Japan: mixed methods study. *JMIR Cancer*. 2021;7(1):e26911.
6. Tercero-Hidalgo JR, Khan KS, Bueno-Cavanillas A, et al. Artificial intelligence in COVID-19 evidence syntheses was underutilized, but impactful: a methodological study. *J Clin Epidemiol*. 2022;148:124-134.
7. Qin X, Liu J, Wang Y, et al. Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. *J Clin Epidemiol*. 2021;133:121-129.
8. Stansfield C, Stokes G, Thomas J. Applying machine classifiers to update searches: analysis from two case studies. *Res Synth Methods*. 2022;13(1):121-133.
9. Lange T, Schwarzer G, Datzmann T, Binder H. Machine learning for identifying relevant publications in updates of systematic reviews of diagnostic test studies. *Res Synth Methods*. 2021;12(4):506-515.
10. Beynon R, Leeflang MMG, McDonald S, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database Syst Rev*. 2013;9:MR000022.
11. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.

12. Thompson G, Zhelev Z, Hunt H, Hyde C. It was not easy to identify the study design from the title and abstract of articles indexed as diagnostic (test) accuracy studies in EMBASE in 2012 and 2019. *J Clin Epidemiol*. 2022;144:102-110.
13. Data science bowl 2017 [Internet]. [cited 2022 Sep 1]. Available from: https://www.kaggle.com/c/data-science-bowl-2017
14. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*. 2017;318(22):2199.
15. Knoll F, Murrell T, Sriram A, et al. Advancing machine learning for MR image reconstruction with an open competition: overview of the 2019 fastMRI challenge. *Magn Reson Med*. 2020;84(6):3054-3070.
16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.
17. [cited 2022 Sep 9]. Available from: http://eppi.ioe.ac.uk/COVID19_MAP/covid_map_v37.html
18. Shiroshita A, Nozaki S, Tanaka Y, Luo Y, Kataoka Y. Thoracic ultrasound for malignant pleural effusion: a systematic review and meta-analysis. *ERJ Open Res*. 2020;6(4): 00464–2020.
19. Sagami S, Kobayashi T, Miyatani Y, et al. Accuracy of ultrasound for evaluation of colorectal segments in patients with inflammatory bowel diseases: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2021;19(5):908-921.e6.
20. Shiroshita A, Jin Z, Tanaka Y, Kataoka Y. Diagnostic accuracy and safety of inhalation challenge tests for bird fancier's lung-systematic review and meta-analysis. *Clin Exp Allergy*. 2020;50(9):1007-1016.
21. Tsujimoto H, Tsujimoto Y, Nakata Y, Akazawa M, Kataoka Y. Ultrasonography for confirmation of gastric tube placement. *Cochrane Database Syst Rev*. 2017;4(4):CD012083.
22. Tsutsumi Y, Tsujimoto Y, Takahashi S, et al. Accuracy of aortic dissection detection risk score alone or with D-dimer: a systematic review and meta-analysis. *Eur Heart J Acute Cardiovasc Care*. 2020;9(3_suppl):S32-S39.
23. Nihashi T, Dahabreh IJ, Terasawa T. Diagnostic accuracy of PET for recurrent glioma diagnosis: a meta-analysis. *AJNR Am J Neuroradiol*. 2013;34(5):944-950.
24. Nihashi T, Ito K, Terasawa T. Diagnostic accuracy of DAT-SPECT and MIBG scintigraphy for dementia with Lewy bodies: an updated systematic review and Bayesian latent class model meta-analysis. *Eur J Nucl Med Mol Imaging*. 2020;47(8):1984-1997.
25. Mishima A, Nihashi T, Ando Y, et al. Biomarkers differentiating dementia with Lewy bodies from other dementias: a meta-analysis. *J Alzheimers Dis*. 2016;50(1):161-174.
26. Takeuchi M, Dahabreh IJ, Nihashi T, Iwata M, Varghese GM, Terasawa T. Nuclear imaging for classic fever of unknown origin: meta-analysis. *J Nucl Med*. 2016;57(12):1913-1919.
27. Iguchi M, Noguchi Y, Yamamoto S, Tanaka Y, Tsujimoto H. Diagnostic test accuracy of jolt accentuation for headache in acute meningitis in the emergency setting. *Cochrane Database Syst Rev*. 2020;6(6):CD012824.
28. Powers DMW. What the F-measure doesn't measure: features, flaws, fallacies and fixes [Internet]. arXiv [cs.IR]. 2015 Available from: http://arxiv.org/abs/1503.06410
29. Glanville J, Kotas E, Featherstone R, Dooley G. Which are the most sensitive search filters to identify randomized controlled trials in MEDLINE? *J Med Libr Assoc*. 2020;108(4):556-563.
30. Tsujimoto Y, Kumasawa J, Shimizu S, et al. Doppler transthoracic echocardiography for detection of pulmonary hypertension in adults. *Cochrane Database Syst Rev*. 2022;5(5):CD012809.
31. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans Knowl Data Eng*. 2005;17(3):299-310.
32. Rufibach K. Use of brier score to assess binary predictions. *J Clin Epidemiol Elsevier BV*. 2010;63(8):938-939. author reply 939.
33. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models. *Epidemiology*. 2010;21(1):128-138.
34. Bisong E. Google Colaboratory. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress; 2019:59-64.
35. Posit Team. Posit Cloud [Internet]. Posit Cloud: Integrated Development for R. RStudio, PBC, Boston, MA. 2022 Available from: https://posit.cloud/
36. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. 2018 Available from: 10.48550/ARXIV.1810.04805.
37. Le T, Winsnes CF, Axelsson U, et al. Analysis of the human protein atlas weakly supervised single-cell classification competition. *Nat Methods*. 2022;19(10):1221-1229.
38. Kaggle competitions [Internet]. [cited 2022 Oct 24]. Available from: https://www.kaggle.com/competitions?sortOption=numTeams&searchQuery=medicine
39. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328.
40. Dinnes J, Sharma P, Berhane S, et al. Rapid, point-of-care antigen tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev*. 2022;7:CD013705.
41. Struyf T, Deeks JJ, Dinnes J, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst Rev*. 2022;5(5):CD013665.
42. Volovici V, Syn NL, Ercole A, Zhao JJ, Liu N. Steps to avoid overuse and misuse of machine learning in clinical research. *Nat Med*. 2022;28(10):1996-1999.
43. Li Z, Kamnitsas K, Glocker B. Overfitting of neural nets under class imbalance: analysis and improvements for segmentation. *Lecture Notes in Computer Science*. Springer International Publishing; 2019:402-410 (Lecture notes in computer science).
44. Zednik C. Solving the black box problem: a normative framework for explainable artificial intelligence. *Philos Technol*. 2021;34(2):265-288.
45. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698.
46. Luo R, Sun L, Xia Y, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining [Internet]. arXiv [cs.CL]. 2022 Available from: http://arxiv.org/abs/2210.10341

How to cite this article: Kataoka Y, Taito S, Yamamoto N, et al. An open competition involving thousands of competitors failed to construct useful abstract classifiers for new diagnostic test accuracy systematic reviews. *Res Syn Meth*. 2023;14(5):707-717. doi:10.1002/jrsm.1649

Received: 29 March 2023 | Revised: 11 July 2023 | Accepted: 24 August 2023

DOI: 10.1002/jor.25684

RESEARCH ARTICLE

Journal of Orthopaedic Research®

Pathophysiology of sex difference in refractoriness in lateral epicondylitis: Biomechanical study of wrist torque

Kazuhiro Ikeda[1,2] | Yuichi Yoshii[3] | Sho Kohyama[1] | Akira Ikumi[2] | Takeshi Ogawa[4] | Reimi Ikeda[5] | Masashi Yamazaki[2]

[1]Department of Orthopedic Surgery, Kikkoman General Hospital, Noda City, Japan

[2]Department of Orthopedic Surgery, Faculty of Medicine, University of Tsukuba, Tsukuba City, Japan

[3]Department of Orthopedic Surgery, Tokyo Medical University Ibaraki Medical Center, Ami Town, Japan

[4]Department of Orthopedic Surgery, Mito Medical Center, Ibarakimachi, Japan

[5]Department of Orthopedic Surgery, Moriya Daiichi General Hospital, Moriya City, Japan

Correspondence

Yuichi Yoshii, Department of Orthopedic Surgery, Tokyo Medical University Ibaraki Medical Center, 3-20-1 Chuo, Ami, Inashiki, Ibaraki 300-0395, Japan.
Email: yyoshii@tokyo-med.ac.jp

Abstract

Eccentric contractions of the wrist extensors worsen lateral epicondylitis (LE), whose pathophysiology may involve sex differences in wrist torque. This study aimed to investigate sex differences in wrist torque in patients with LE. The wrist extension and flexion torques of 22 patients with LE (11 males and 11 females) were measured. Maximum muscle output over time was measured for 20 s, initial torque was defined as muscle strength, and the degree of eccentric contraction was quantified and defined as the eccentric contraction index (ECI). The affected/unaffected side ratio of the wrist extensor, extensor/flexor ratio of muscle strength, and affected/unaffected side difference of ECI between sexes were statistically analyzed. Furthermore, correlations between wrist extensor torque, ECI, and Visual Analog Scale of pain during the examination were evaluated. Females were found to display lower affected/unaffected side ratios of the wrist extensor and wrist extension/flexion ratios for the affected side, compared with males; however, no differences were found in the wrist extension/flexion ratios for the unaffected side in both sexes. Additionally, females presented with larger differences between the affected and unaffected sides in the ECI. Based on correlations between wrist torques, ECI, and pain, females tended to suppress muscle output to prevent pain from eccentric contraction of wrist extensors more than males, which would induce an imbalance in muscle strength of the wrist extensors and flexors. This imbalance may result in chronic eccentric contraction of the wrist extensors with gripping, exacerbating LE.

KEYWORDS

biomechanics, elbow, tendon

1 | INTRODUCTION

1.1 | Background

Lateral epicondylitis (LE) of the humerus is the tendinopathy of the forearm extensors.[1,2] The prevalence of LE is 1%–3% of the general population[2,3] and 2%–14.5% of manual workers who perform repetitive gripping movements with the forearm in pronation.[4,5] Females have a higher prevalence of LE and a 2.7 times higher risk of refractoriness than males.[3,4,6] Pathophysiologically, repetitive eccentric contraction of wrist extensors is a factor in the development and exacerbation of LE, which induces micro rupture at the enthesis.[7,8] Uncontrolled conditions worsen histological severity with vascular fibroblast proliferation, mucoid degeneration, calcification, and ectopic ossification at the enthesis.[7,9,10] Decreased muscle strength and endurance of the wrist extensors are considered to be risk factors for the development of LE, because it induces an eccentric contraction of the wrist extensors in gripping with forearm

pronation.[11] Although LE is considered a self-limiting disease, about 10% of patients do not respond to conservative treatment and sometimes require surgery for refractory LE.[9]

As refractory LE produces a socioeconomic burden resulting from labor difficulties, overcoming it is a social issue; the economic loss reportedly reaches £27 million annually in England.[12] Although heavy labor, smoking, and females are epidemiologically known as risk factors for refractoriness,[4] the pathophysiology leading to refractoriness is poorly understood. Particularly, no literature provides pathophysiological evidence that females are more prone to develop refractory LE than males. The lack of pathophysiological knowledge leads to inappropriate treatment, which results in a difference in refractoriness rates between sex. Therefore, it is essential to understand the pathophysiology of the difference in refractoriness rates between sex to overcome refractory LE.

1.2 | Rationale

As LE exacerbates with repetitive eccentric contractions of wrist extensors,[7,8] female patients with LE would be expected to be more exposed to eccentric contractions than males. Muscles are prone to eccentric contraction when muscle performance declines, including low muscle strength, contraction velocity, and endurance.[11] In patients with LE, muscle contraction velocity of the extensor carpi radialis delays,[13] and the strength of the wrist extensor reduces relative to the strength of the flexor.[14,15] As wrist extensors and flexors co-contract in gripping,[16,17] this muscle strength imbalance of the wrist may result in repetitive eccentric contractions with each gripping motion. Therefore, we hypothesized that the loss of wrist extensor strength due to LE is more significant in females than in males, resulting in a more significant muscle imbalance between the wrist extensors and flexors. This study aimed to reveal sex differences in wrist extension and flexion torques in patients with LE at their initial examination.

2 | METHODS

2.1 | Study design

The study protocol conforms to the principles outlined in the 1964 Declaration of Helsinki. This study was approved by our Institutional Review Board. Written informed consent was obtained from all patients.

This case–control study investigated the wrist extension and flexion torques of patients with LE cross-sectionally (Level of evidence III).

2.2 | Participants

Participants were consecutive patients with unilateral LE whom the author examined in our hospital between February 2022 and January 2023. The exclusion criteria were patients with any shoulders or wrists symptoms, a history of elbow trauma, and arthritic diseases: osteoarthritis (Kellgren–Lawrence classification 2 or higher),[18] osteochondritis dissecans, and rheumatoid arthritis. Furthermore, we excluded mild patients to clarify the characteristics of wrist torque in patients with LE. As previous studies reported quick disabilities of the arm, shoulder, and hand (QuickDASH) scores of 11–15 points at the end of conservative treatment for LE,[19,20] we set a cutoff score of QuickDASH for exclusion criteria of patients with mild LE as 11.

LE diagnosis was based on physical examination findings, positive Thomsen or Maudsley's test, and tenderness at the lateral epicondyle. All patients underwent an X-ray and MRI to rule out elbow arthritis and trauma.

Thirty patients with LE received the author's examination during the study period. We excluded two patients with bilateral symptoms, four with a QuickDASH score of fewer than 11 points, one with a history of elbow trauma, and one who refused to participate in this study. Subsequently, this study included 22 patients with LE (11 males and females each, average age and SD: 55.6 ± 13.4 years, age range: 39–84 years).

2.3 | Physical examination and clinical evaluation

The following data were recorded on the subject background: height, weight, body mass index (BMI), forearm length and circumference, lever arm length for obtaining wrist torque, and grip strength. BMI is reported to be positively associated with grip strength[21] and wrist torque.[22] The grip strength was measured with a hand dynamometer in the upper limb hanging downward, elbow extension, and the forearm in a neutral position. The forearm length was from the lateral epicondyle of the humerus to the styloid process of the radius. The forearm circumference was around the thickest part of the forearm. These forearm lengths and circumferences were measured as indicators of the forearm muscle volume.[23] The lever arm length was from the tip of the radial styloid to the grip center. This measurement of the lever arm length was evaluated by gripping the torque measuring device with their wrist and forearm in a neutral position.

As the patient-based self-evaluation for LE, disability/symptom score of QuickDASH, Visual Analog Scale (VAS) for pain at rest, during daily activities, and wrist torque measurement were evaluated. Thereafter, we calculated the VAS difference as follows: VAS difference = (VAS during daily activity) − (VAS during torque measurement).

2.4 | Measurement of wrist torque

Wrist extension and flexion muscle strengths were measured with a wrist torque measuring device (Three-One Design Inc.) (Figure 1).[24] This apparatus measures wrist torques every 10 ms and averages

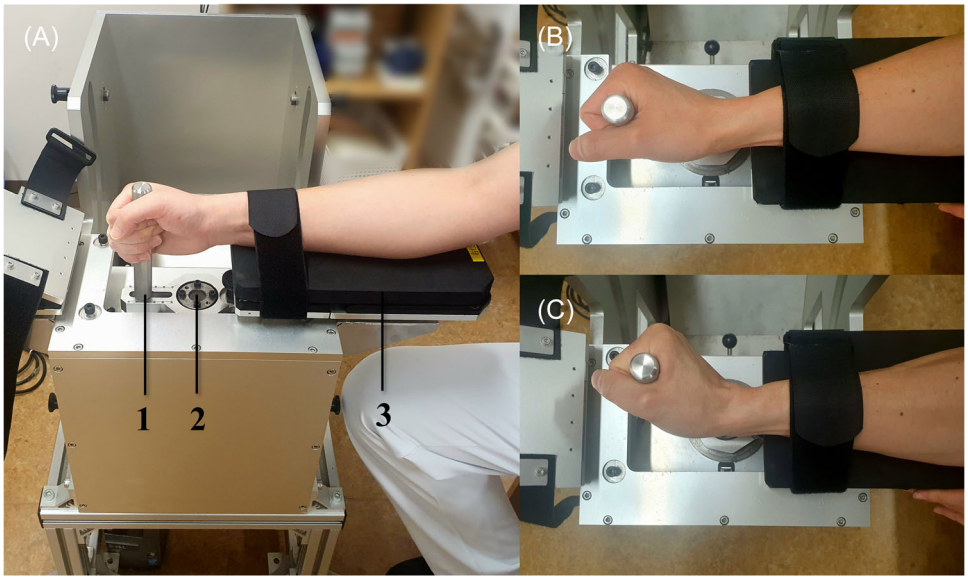


FIGURE 1 (A) The wrist torque measuring equipment consists of a handle (1), a rotation center to measure the wrist torque (2), and a table to place the forearms (3). (B) Participants adjusted the moving handle to ensure the rotation center of the wrist and the rotation center of the equipment together. (C) The handle rotates around the wrist axis in response to the input torque.

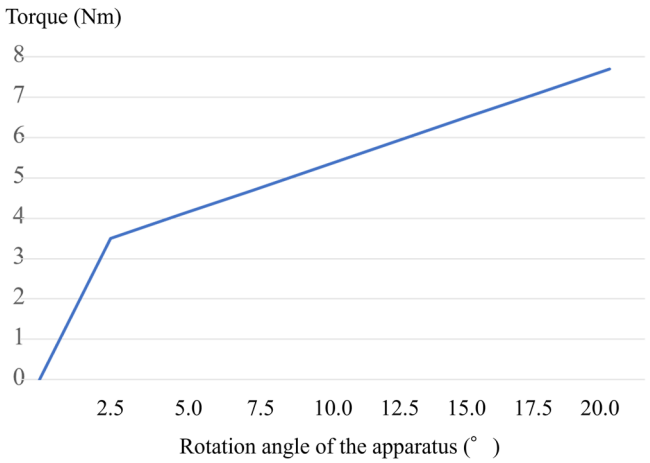


FIGURE 2 Input torque required to rotate the torque-measuring device, measured every 2.5°. The measuring machine required 1.5 Nm/° for the initial movement. Thereafter, the required force was 0.24 Nm/° to rotate the handle.

them every second, which enables us to evaluate reliable torque over time. Participants sat on a chair and placed their forearms on the apparatus table, with their elbows slightly flexed at 30°–45° and the forearm in a neutral position. The rotation centers of the wrist, which is the axis connecting the radial and ulnar styloid,[24] were adjusted to the rotational center of the device by moving the handle position. The gripping handle of this measurement device rotates around the wrist rotational axis in response to the output of wrist torque. Figure 2 shows the torque output required to rotate this torque-measuring device every 2.5°. Patients performed a maximal voluntary wrist extension or flexion contraction for 20 s. As wrist torque measurement in this study evaluates dynamic muscle contraction, decreased muscle output over time during examination represents eccentric contraction. On the affected side, the patient was encouraged to exert as much force as possible within the self-acceptable pain. Although we explained that the examination could be interrupted if the pain was severe, none of the patients were interrupted.

2.5 | Data analysis

We assessed maximum muscle strength and the degree of eccentric contraction that occurred during wrist torque examination.

The wrist torque over time was evaluated for 20 s from its first peak waveform. The wrist torque of the first peak waveform was defined as muscle strength at the starting point of the measurement (MS[S]). MS(S) was used for further evaluation as the maximum muscle strength of the patients. Muscle strength of 20 s after the start point was used as muscle strength at the end of the measurement (MS[E] [Nm]). Then, the muscle strength loss (MS[L]) per second was calculated with the following equation: MS(L) = (MS[S] − MS[E])/20 [Nm/s]). MS(L) reflects the degree of eccentric contraction that occurred during the measurement. Finally, we normalized MS(L) by dividing MS(S) and defined this as eccentric contraction index (ECI): ECI = MS(L)/MS(S).

The primary parameters in this study were the affected/unaffected side ratio of the MS(S), the wrist extension/flexion ratio of the MS(S), and the affected/unaffected side difference of the ECI.

2.6 | Statistical analysis

We performed the Shapiro–Wilk test for each evaluated item as a normality test. All collected parameters, including clinical characteristics and measured values, were compared. We used the χ^2 test for categorical parameters, the Student's *t* test for continuous parameters between sexes, and the paired *t* test to compare the affected and unaffected sides with a normal distribution. We used the Mann–Whitney *U* test for continuous parameters between sexes and Wilcoxon signed-rank test to compare the affected and unaffected sides with an irregular distribution.

We analyzed Spearman's correlation coefficients to assess the association between the clinical characteristics and wrist torque. Correlation coefficients of $\pm 0.3 < r < \pm 0.7$ were considered moderate correlations, and those of $r > \pm 0.7$ were considered strong correlations.

For statistical comparisons in the present study, we assumed that comparisons of the extension/flexion ratio of the muscle strength within each sex would have the most power, as previous studies have already shown significant differences.[15] Therefore, in determining sample size, we calculated the sample size that would satisfy an effect size of 0.8, a type I error of 0.05, and a power of 0.8 for the paired *t* test. Subsequently, we determined a sample size of 11 for each sex. $p < 0.05$ was considered to be significant. All statistical analyses were performed using Bellcurve for Excel version 3.20 (SSRI Co.n).

3 | RESULTS

3.1 | Demographic and clinical characteristics

Table 1 summarizes the demographic and clinical characteristics of males and females. There were no significant differences in the distribution of age. Concerning physical characteristics, height, weight, forearm length, and circumference, lever arm length to obtain wrist torque and grip strength were larger in males than in females. There was no significant difference in BMI or in any of the items related to clinical characteristics or severity of LE between sexes, that is, comparisons of dominant hand incidence, duration of disease, affected/unaffected side ratio of grip strength, QuickDASH score, or VAS.

TABLE 1 Demographic and clinical characteristics between males and females.

	Group		
	Males ($n = 11$)	Females ($n = 11$)	*p*
Age (years)[a]	54.1 ± 16.3	57.1 ± 10.4	0.61
Height (cm)[a]	169.3 ± 7.5	157.0 ± 5.4	<0.001
Weight (kg)[a]	71.6 ± 12.2	55.9 ± 13.5	0.010
BMI[a]	24.9 ± 2.9	22.7 ± 5.5	0.26
Forearm length (cm)[a]			
Affected side	26.3 ± 1.5	23.7 ± 1.5	<0.001
Unaffected side	26.3 ± 1.8	23.7 ± 1.5	0.001
Forearm circumference (cm)[a]			
Affected side	27.2 ± 2.2	23.5 ± 2.7	0.002
Unaffected side	27.2 ± 2.1	23.4 ± 2.9	0.002
Lever arm length for obtaining the wrist torque (mm)[a]	66.4 ± 3.2	60.0 ± 2.3	<0.001
Grip strength (kgf)[a]			
Affected side	27.2 ± 10.7	13.5 ± 7.6	0.002
Unaffected side	38.2 ± 10.4	23.5 ± 8.9	0.002
Affected/unaffected side ratio[a]	0.75 ± 0.28	0.57 ± 0.28	0.15
Affects the dominant hand	8 (72.7%)	9 (81.8%)	0.61
Duration of disease (days)[b]	57 (35–74.5)	62 (19.5–89.5)	0.53
QuickDASH score[a]	31.6 ± 18.3	42.0 ± 20.6	0.25
VAS during			
At rest[b]	0 (0–25.5)	0 (0–21.0)	0.97
Daily activity[a]	50.4 ± 18.9	54.2 ± 15.3	0.61
Wrist torque measurement[a]	27.1 ± 26.4	34.5 ± 21.6	0.48

Note: Underline, $p < 0.05$.

Abbreviations: BMI, body mass index; DASH, disabilities of arm, shoulder and hand; VAS, Visual Analog Scale.

[a]Data are normally distributed, described with average ± SD.

[b]Data are irregularly distributed, described with median (25–75 percentile).

3.2 | Wrist torque comparison

Figure 3 shows the results of the average values for wrist extension torque over time in males and females, with Table 2 summarizing the analysis. MS(S) was significantly increased in males than females on both the affected and unaffected sides. The affected/unaffected side ratio of MS(S) was significantly increased in males than in females. On the affected side, MS(L) and ECI were significantly increased in males than that in females. ECI difference between the affected and unaffected sides was more significantly increased in females than males. The ECI of the affected side was lower than that of the unaffected side in females ($p = 0.002$), whereas there was no significant difference in males ($p = 0.67$).

Figure 4 shows the results of the average values for wrist flexion torque over time, with Table 3 summarizing the analysis. MS(S) was increased in males than in females on both the affected and unaffected sides. There were no significant differences in the affected/unaffected side ratio of the MS(S), MS(L), ECI, and the

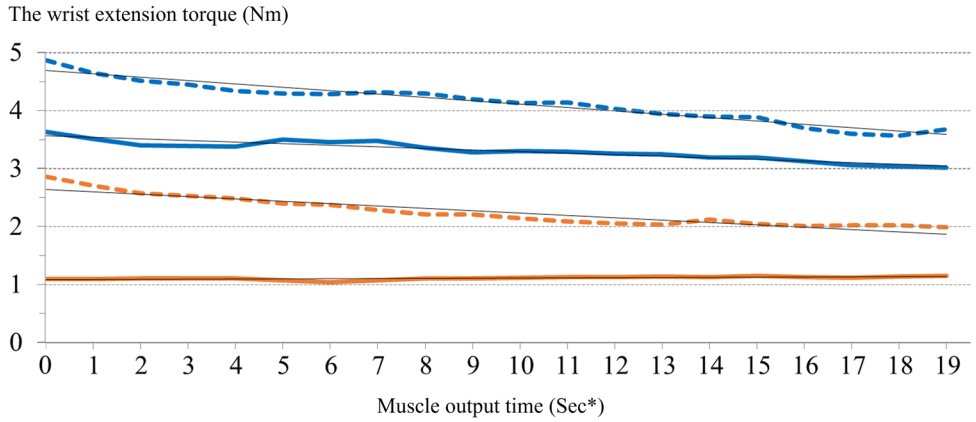


FIGURE 3 The average wrist extension torque for males and females over time. The intercept of the regression line approximates the muscle strength at the starting point of the measurement and the slope approximates the decrease in muscle strength per second. The regression line for each graph was as follows: Blue dotted line, the unaffected side in males ($y = -0.058x + 4.7$, $R^2 = 0.95$); blue solid line, the affected side in males ($y = -0.027x + 3.6$, $R^2 = 0.89$); orange dotted line, the unaffected side in females ($y = -0.041x + 2.7$, $R^2 = 0.89$); orange solid line; affected side in females ($y = -0.0031x + 1.1$, $R^2 = 0.49$). *The apparatus measures wrist torques every 10 ms and averages them every second; the torque at x seconds in the graph indicates the average torque of x to $x + 1$ s during the measurement.

TABLE 2 The result of wrist extension torques measurement.

	Males ($n = 11$)	Females ($n = 11$)	p
MS(S)			
Affected side (Nm)[b]	4.1 (2.6–4.3)	0.7 (0.5–1.5)	<0.001
Unaffected side (Nm)[a]	4.9 ± 1.5	2.7 ± 1.5	0.003
Affected/unaffected side ratio[a]	0.76 ± 0.24	0.40 ± 0.19	0.001
MS(L)			
Affected side ($\times 10^{-2}$ Nm/s)[b]	2.2 (0.2–6.5)	−0.4 (−0.9–0.0)	0.028
Unaffected side ($\times 10^{-2}$ Nm/s)[b]	4.8 (2.6–8.8)	2.2 (1.3–6.4)	0.18
ECI			
Affected side ($\times 10^{-3}$)[a]	9.6 ± 14	−9.2 ± 14	0.006
Unaffected side ($\times 10^{-3}$)[a]	11.8 ± 6.7	12.7 ± 7.8	0.77
Affected/unaffected side difference ($\times 10^{-3}$)[b]	−3.8 (−10.4 to 4.3)	−19.9 (−27.1 to −11.8)	0.016

Note: Underline, $p < 0.05$.

Abbreviations: ECI, eccentric contraction index; MS(L), muscle strength loss; MS(S), muscle strength at the starting point for the measurement.

[a]Data are normally distributed, described with average ± SD.

[b]Data are irregularly distributed, described with median (25–75 percentile).

affected/unaffected side difference of the ECI between males and females.

Figure 5 shows the results of the wrist extension/flexion ratio of the MS(S). In males, the wrist extension/flexion ratios of the MS(S) were 0.48 (0.43–0.67) and 0.60 (0.43–0.76) on the affected and unaffected sides, respectively. In females, the wrist extension/flexion ratios of the MS(S) were 0.29 (0.29–0.41) and 0.62 (0.44–0.72) on the affected and unaffected sides, respectively. On the affected side, the wrist extension/flexion ratio of the MS(S) was increased in males than in females ($p = 0.020$), whereas there was no significant difference between sexes on the unaffected side ($p = 0.95$). The wrist extension/flexion ratio of the MS(S) was increased on the unaffected side than that on the affected side in females ($p = 0.013$), whereas there was no significant difference in males ($p = 0.86$).

The resistible weight of the wrist extensors and flexors, which are calculated from the lever arm length and torque results, are shown in Supporting Information: Table S1.

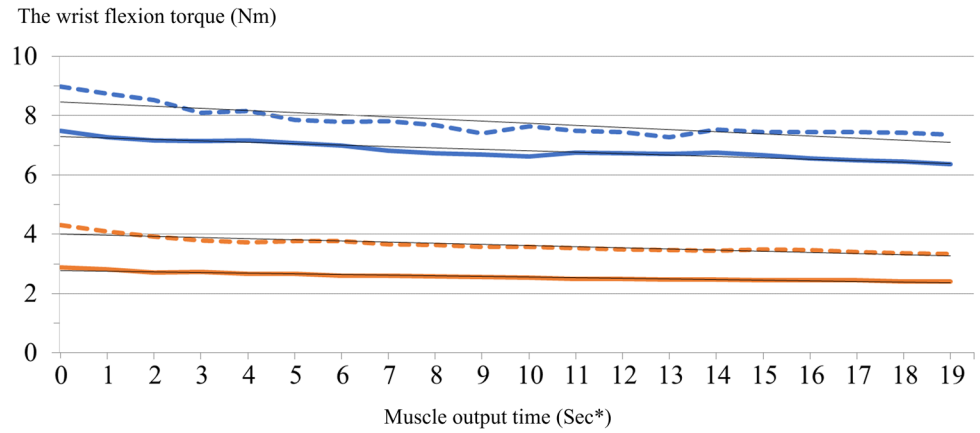


FIGURE 4 Wrist flexion torque for males and females over time. The regression line for each graph was as follows: Blue dotted line, the unaffected side in males ($y = -0.071x + 8.5$, $R^2 = 0.76$); blue solid line, the affected side in males ($y = -0.048x + 7.3$, $R^2 = 0.89$); orange dotted line, the unaffected side in females ($y = -0.039x + 4.0$, $R^2 = 0.85$); orange solid line; affected side in females ($y = -0.022x + 2.8$, $R^2 = 0.91$). *The torque at x seconds in the graph indicates the average torque of x to $x + 1$ s during the measurement.

TABLE 3 The result of wrist flexion torques and endurances in males and females.

	Males ($n = 11$)	Females ($n = 11$)	p
MS(S)			
Affected side (Nm)[b]	7.8 (4.2–9.0)	2.5 (1.7–4.5)	<0.001
Unaffected side (Nm)[b]	7.3 (6.5–9.6)	3.7 (3.0–5.3)	0.0014
Affected/Unaffected side ratio[a]	0.83 ± 0.30	0.69 ± 0.28	0.24
MS(L)			
Affected side ($\times 10^{-2}$ Nm/s)[b]	3.1(−0.1–11.0)	2.0 (−1.6–3.1)	0.49
Unaffected side ($\times 10^{-2}$ Nm/s)[b]	7.7 (3.4–11.4)	4.9 (2.6–6.4)	0.20
ECI			
Affected side ($\times 10^{-3}$)[b]	7.0 (−3.0–12.4)	8.5 (5.3–12.3)	0.34
Unaffected side ($\times 10^{-3}$)[a]	9.6 ± 6.3	11.1 ± 7.4	0.60
Affected/Unaffected side difference ($\times 10^{-3}$)[b]	−2.4 (−9.6–3.1)	−2.8 (−10.1–4.3)	0.81

Note: Underline, $p < 0.05$.

Abbreviations: ECI, eccentric contraction index; MS(L), muscle strength loss; MS(S), muscle strength at the starting point for the measurement.

[a]Data are normally distributed, described with average ± SD.

[b]Data are irregularly distributed, described with median (25–75 percentile).

3.3 | Correlations between body characteristics, clinical characteristics, and wrist extension torque

The correlations between body characteristics data and wrist extension torque of the unaffected side are shown in Supporting Information: Table S2. Table 4 shows correlations between clinical characteristics and wrist extension torque in males and females. The correlations between the affected/unaffected side ratio of MS(S) and each parameter were as follows: affected/unaffected side ratio of grip strength showed a moderate positive correlation in both sexes; VAS difference showed a moderate negative correlation in females; and QuickDASH score showed strong negative correlation in females. The correlations between the affected/unaffected side difference in ECI and each parameter were as follows: affected/unaffected side ratio of grip strength showed a moderate positive correlation in both sexes; VAS difference and QuickDASH score showed moderate negative correlations in both sexes; and affected/unaffected side ratio of MS (S) showed a strong positive correlation in females.

4 | DISCUSSION

The present study demonstrated that the affected/unaffected side ratio of the MS(S) of the wrist extensor was significantly lower in females than in males. This suggests that the wrist extensor strength in females decreased more remarkably than that of males because of

LE. In contrast, there was no difference in the loss of wrist flexor strength between the sexes. Consequently, the extension/flexion ratio of the MS(S) in females was lower on the affected side than on either their unaffected side or the affected side in males. As there was a co-contraction of wrist flexors and extensors during gripping in the wrist extension position,[16,17] this weakness of wrist extensors relative to flexors may induce chronic eccentric contraction of wrist extensors with gripping, which exacerbates the condition of LE. Therefore, this significant loss of wrist extensor strength in female patients with LE may be a risk of refractoriness.

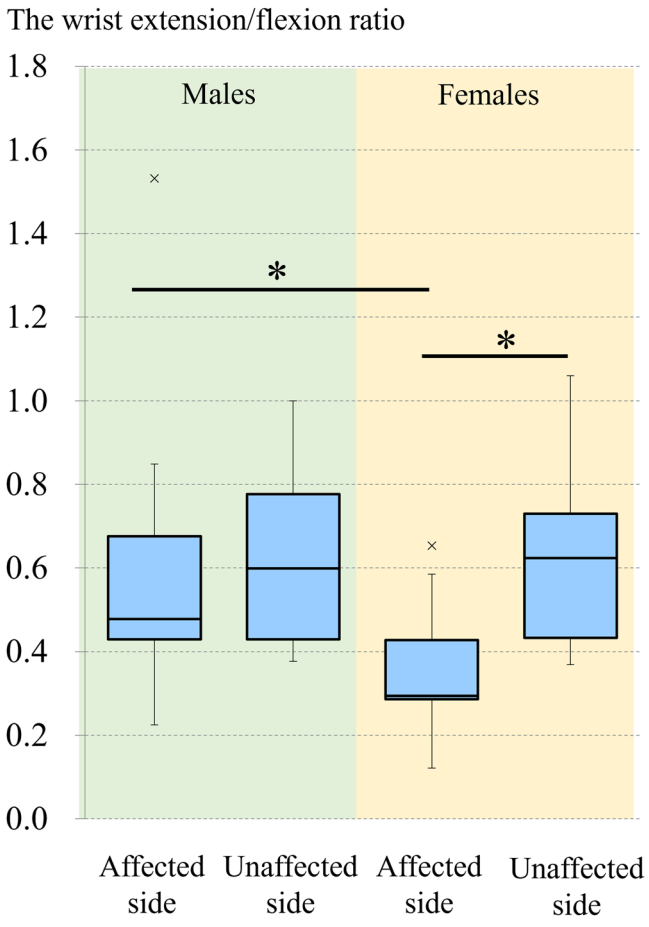


FIGURE 5 The wrist extension/flexion ratio of the muscle strength at the starting point for the measurement. The boxes show the median and the interquartile range, the whiskers with values within 1.5 times the interquartile range, and *x* marks of the outlier's values. *$p < 0.05$.

The present study identified the cause of remarkable muscle weakness in female patients with LE by quantifying the eccentric contractions and pain during the wrist torque examination. The correlation among MS(S), ECI, and VAS showed that patients with less muscle output experienced less pain and less eccentric contraction. Patients with LE experience pain due to eccentric contraction of the wrist extensors, which is clinically used in manual tests such as Thomsen and Maudsley's tests.[25] The pain caused by eccentric contraction activates the sensorimotor system and decreases muscle output through negative feedback.[13] This pain-induced suppression of muscle output was stronger in females. Although the muscle output of unaffected wrist extensors decreased almost linearly over time, the coefficient of determination of the regression line was lower in the affected wrist extensors of females. This indicates that there was unstable muscle output in the affected wrist extensors. Furthermore, the ECI of the wrist extensor in females was smaller than that in males and smaller ECI results in a smaller affected/unaffected side ratio of the MS(S) of the wrist extensor. These results indicate that females tended to suppress muscle output to prevent pain from eccentric contraction of wrist extensors more than males. Furthermore, patients who suppressed muscle output to prevent eccentric contractions had a lower affected/unaffected side ratio of grip strength in both males and females. Patients with markedly decreased grip strength should be treated with an understanding that they are at high risk for refractory LE. Aggressive conservative treatment should be considered for such patients, for example, extensive orthotic therapy of tennis elbow brace with a cock-up splint[26] and physical therapy, including controlled eccentric contraction exercises.[27]

Similar to the present study, previous studies of experimentally induced pain have consistently shown that females exhibit greater pain sensitivity than males.[28] These sex differences in pain are

TABLE 4 Correlation between clinical data and the affected wrist extension torque.

	Affected/unaffected side ratio of MS(S)		Affected/unaffected side difference in ECI	
	Males	Females	Males	Females
Affected/Unaffected side ratio of grip strength	0.34	0.37	0.52	0.42
VAS difference[a]	0.00	−0.61	−0.30	−0.39
QuickDASH score	−0.10	−0.72	−0.40	−0.47
Affected/unaffected side ratio of MS(S)			−0.19	0.83

Note: 0.3 < < 0.7; 0.7 < ; −0.7 < <−0.3; <−0.7.

Abbreviations: DASH score, disabilities of the arm, shoulder and hand score; ECI, eccentric contraction index; MS(S), muscle strength at the starting point for the measurement; VAS, visual analog scale.

[a]VAS difference = (VAS during daily activity) − (VAS during torque measurement).

reported as multiple biopsychosocial mechanisms, including sex hormones, endogenous opioid function, and genetic factors.[28] We have shown that severe pain in female patients with LE is not only the consequence of a severe condition but also the cause of morbid muscle output suppression, which leads to refractory LE. In LE refractoriness, severe pain triggers the vicious cycle of worsening LE and further reduces the muscle output of wrist extensors.

The limitation of the present study was that this was a cross-sectional study. We did not directly evaluate the prognosis of the patients. Resistance to LE treatment is multifactorial, including biological, environmental, and psychological factors.[5,11,29] A lower affected/unaffected side ratio of the MS(S) of the wrist extensor or an extension/flexion ratio of the MS(S) is only one of the biological risk factors for refractoriness. A longitudinal study with multivariate analysis would identify the risks more quantitatively. The present study provides basic research results on wrist torque characteristics in males and females of patients with LE, which may be able to use in future longitudinal studies.

In conclusion, wrist extensor strength on the affected side of LE was decreased more remarkably in females than in males. Females tended to suppress muscle output to prevent pain from eccentric contraction of wrist extensors more than males. This morbid condition resulted in wrist extensor/flexor strength imbalance. Female patients with LE are prone to eccentric contraction of the wrist extensors in each gripping, which is a risk of refractoriness of LE.

AUTHOR CONTRIBUTIONS

Conceptualization: Ikeda K. *Data curation*: Ikeda K. *Formal analysis*: Ikeda K., Ikeda R. *Investigation*: Ikeda K. *Methodology*: Ikeda K., Yoshii Y. *Project administration*: Ikumi A. *Resource*: Ikeda K., Kohyama S. *Software*: Yoshii Y. *Supervision*: Yoshii Y., Yamazaki M. *Validation*: Ikeda R., Ogawa T. *Visualization*: Ikeda K. *Writing—original draft*: Ikeda K. *Writing—review and editing*, Yoshii Y. All authors have read and approved the final submitted manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ETHICS STATEMENT

Our Institutional Review Board approved the present study (Approval No. KC-H26, Date 27th. May. 2022).

ORCID

Kazuhiro Ikeda http://orcid.org/0000-0002-7112-6511
Yuichi Yoshii http://orcid.org/0000-0003-1447-4664

REFERENCES

1. Maffulli N, Longo UG, Denaro V. Novel approaches for the management of tendinopathy. *J Bone Joint Surg Am*. 2010;92:2604-2613.
2. Shiri R, Viikari-Juntura E, Varonen H, Heliovaara M. Prevalence and determinants of lateral and medial epicondylitis: a population study. *Am J Epidemiol*. 2006;164:1065-1074.
3. Coombes BK, Bisset L, Vicenzino B. A new integrative model of lateral epicondylalgia. *Br J Sports Med*. 2009;43:252-258.
4. Fan ZJ, Silverstein BA, Bao S, et al. Quantitative exposure-response relations between physical workload and prevalence of lateral epicondylitis in a working population. *Am J Ind Med*. 2009;52:479-490.
5. Shiri R, Viikari-Juntura E. Lateral and medial epicondylitis: role of occupational factors. *Best Pract Res Clin Rheumatol*. 2011;25:43-57.
6. Wolf JM, Mountcastle S, Burks R, Sturdivant RX, Owens BD. Epidemiology of lateral and medial epicondylitis in a military population. *Mil Med*. 2010;175:336-339.
7. Bhabra G, Wang A, Ebert JR, Edwards P, Zheng M, Zheng MH. Lateral elbow tendinopathy: development of a pathophysiology-based treatment algorithm. *Orthop J Sports Med*. 2016;4:232596711667063.
8. Ozone K, Kokubun T, Takahata K, et al. Structural and pathological changes in the enthesis are influenced by the muscle contraction type during exercise. *J Orthop Res*. 2022;40:2076-2088.
9. Ikeda K, Ogawa T, Ikumi A, et al. Individual evaluation of the common extensor tendon and lateral collateral ligament improves the severity diagnostic accuracy of magnetic resonance imaging for lateral epicondylitis. *Diagnostics*. 2022;12:1871.
10. Kraushaar BS, Nirschl RP. Current concepts review - tendinosis of the elbow (Tennis Elbow). clinical features and findings of histological, immunohistochemical, and electron microscopy studies*. *J Bone Joint Surg*. 1999;81:259-278.
11. Ikeda K, Yoshii Y, Kohyama S, et al. Sex differences in wrist torque and endurance-biomechanical factors associated with developing lateral epicondylitis of the humerus. *J Orthop Res*. 2023;41(8):1670-1677.
12. Hopkins C, Fu SC, Chua E, et al. Critical review on the socio-economic impact of tendinopathy. *Asia Pac J Sports Med Arthrosc Rehabil Technol*. 2016;4:9-20.
13. Chourasia AO, Buhr KA, Rabago DP, Kijowski R, Irwin CB, Sesto ME. Effect of lateral epicondylosis on grip force development. *J Hand Ther*. 2012;25:27-37.; Quiz 37.
14. Pienimäki TT, Siira PT, Vanharanta H. Chronic medial and lateral epicondylitis: a comparison of pain, disability, and function. *Arch Phys Med Rehabil*. 2002;83:317-321.
15. Unyó C, Chaler J, Rojas-Martínez M, et al. A cross-sectional study comparing strength profile of dorsal and palmar flexor muscles of the wrist in epicondylitis and healthy men. *Eur J Phys Rehabil Med*. 2013;49:507-515.
16. Lee JA, Sechachalam S. The effect of wrist position on grip endurance and grip strength. *J Hand Surg*. 2016;41:e367-e373.
17. Mogk JPM, Keir PJ. Crosstalk in surface electromyography of the proximal forearm during gripping tasks. *J Electromyography Kinesiol*. 2003;13:63-71.
18. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis*. 1957;16:494-502.
19. Guler T, Yildirim P. Comparison of the efficacy of kinesiotaping and extracorporeal shock wave therapy in patients with newly diagnosed lateral epicondylitis: a prospective randomized trial. *Niger J Clin Pract*. 2020;23:704-710.
20. Smith-Forbes EV, Howell DM, Willoughby J, Pitts DG, Uhl TL. A retrospective cohort study of QuickDASH scores for three hand therapy acute upper limb conditions. *Mil Med*. 2018;183:522-529.
21. Pasdar Y, Darbandi M, Mirtaher E, Rezaeian S, Najafi F, Hamzeh B. Associations between muscle strength with different measures of obesity and lipid profiles in men and women: results from RaNCD cohort study. *Clin Nutr Res*. 2019;8:148-158.
22. Seo NJ, Armstrong TJ, Ashton-Miller JA, Chaffin DB. 2008. Wrist strength is dependent on simultaneous power grip intensity. *Ergonomics*. 51: 1594-1605.
23. Vanmechelen IM, Shortland AP, Noble JJ. Lower limb muscle volume estimation from maximum cross-sectional area and muscle length in

cerebral palsy and typically developing individuals. *Clin Biomech*. 2018;51:40-44.
24. Yoshii Y, Yuine H, Kazuki O, Tung W, Ishii T. Measurement of wrist flexion and extension torques in different forearm positions. *Biomed Eng Online*. 2015;14:115.
25. Zwerus EL, Somford MP, Maissan F, Heisen J, Eygendaal D, van den Bekerom MP. Physical examination of the elbow, what is the evidence? A systematic literature review. *Br J Sports Med*. 2018;52:1253-1260.
26. Ma KL, Wang HQ. Management of lateral epicondylitis: a narrative literature review. *Pain Res Manag*. 2020;2020:1-9.
27. Lee J, Kim T, Lim K. Effects of eccentric control exercise for wrist extensor and shoulder stabilization exercise on the pain and functions of tennis elbow. *J Phys Ther Sci*. 2018;30:590-594.
28. Bartley EJ, Fillingim RB. Sex differences in pain: a brief review of clinical and experimental findings. *Br J Anaesth*. 2013;111:52-58.
29. Haahr JP. Physical and psychosocial risk factors for lateral epicondylitis: a population based case-referent study. *Occup Environ Med*. 2003;60:322-329.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ikeda K, Yoshii Y, Kohyama S, et al. Pathophysiology of sex difference in refractoriness in lateral epicondylitis: biomechanical study of wrist torque. *J Orthop Res*. 2024;42:277-285. doi:10.1002/jor.25684

in vivo *37*: 2203-2209 (2023)
doi:10.21873/invivo.13320

Atezolizumab Monotherapy for Non-small Cell Lung Cancer Patients: An Observational Study in Ibaraki Group (ATTENTION-IBARAKI)

SHINICHIRO OKAUCHI[1], GEN OHARA[1], TOSHIHIRO SHIOZAWA[2], HIROKO WATANABE[3], TAKESHI NUMATA[4], RYOTA NAKAMURA[4], TOMOHIRO TAMURA[5], NORIHIRO KIKUCHI[6], KUNIHIKO MIYAZAKI[7], SHIGEN HAYASHI[8], HIROFUMI SAKURAI[8], TAKAAKI YAMASHITA[9], KOICHI KURISHIMA[10], MASAHARU INAGAKI[11], TAKEO ENDO[1], HIROICHI ISHIKAWA[10], TAKAYUKI KABURAGI[5], HIROAKI SATOH[4], TORU SAKAMOTO[3] and NOBUYUKI HIZAWA[2]

[1]*Division of Respiratory Medicine, Mito Medical Center, University of Tsukuba, Mito, Japan;*
[2]*Division of Respiratory Medicine, Faculty of Clinical Medicine, University of Tsukuba, Tsukuba, Japan;*
[3]*Division of Respiratory Medicine, Tsukuba Memorial Hospital, Tsukuba, Japan;*
[4]*Departments of Respiratory Medicine and Surgery, National Hospital Organization Mito Medical Center, Ibarakimachi, Japan;*
[5]*Respiratory Center, Ibaraki Prefectural Central Hospital, Kasama, Japan;*
[6]*Division of Respiratory Medicine, National Hospital Organization Kasumigaura Medical Center, Tsuchiura, Japan;*
[7]*Division of Respiratory Medicine, Ryugasaki Saiseikai Hospital, Ryugasaki, Japan;*
[8]*Division of Respiratory Medicine, Ibaraki Seinan Medical Center Hospital, Sakai, Japan;*
[9]*Division of Respiratory Medicine, JA Toride Medical Center Hospital, Toride, Japan;*
[10]*Division of Respiratory Medicine, Tsukuba Medical Center Hospital, Tsukuba, Japan;*
[11]*Division of Thoracic Surgery, Tsuchiura Kyodo General Hospital, Tsuchiura, Japan*

Abstract. *Background/Aim: Atezolizumab is a monoclonal antibody that targets programmed death-ligand 1 (PD-L1) expressed on cancer cells derived from various organs and antigen-presenting cells and is currently commonly used in combination with chemotherapy. We conducted a study to clarify the current status of response to atezolizumab monotherapy in clinical practice and clarify the factors that contribute to long-term response and survival. Patients and Methods: We conducted a retrospective review of patients with advanced non-small cell lung cancer (NSCLC) treated with atezolizumab monotherapy from April 2018 to March 2023 at 11 Hospitals. Results: The 147 patients evaluated had a progression-free survival (PFS) of 3.0 months and an overall survival of 7.0 months. Immune-related adverse events of any grade were observed in 13 patients (8.8%), grade 3 or higher in nine patients (6.1%), and grade 5 with pulmonary toxicity in one patient (0.7%). Favorable factors related to PFS were 'types of NSCLC other than adenocarcinoma'. Favorable factors for overall survival were 'performance status 0-1' and 'treatment lines up to 3'. There were 16 patients (10.9%) with PFS >1 year. No characteristic clinical findings were found in these 16 patients compared to the remaining 131 patients. Conclusion: Efficacy and immune-related adverse events of NSCLC patients associated with atezolizumab monotherapy were comparable to those of previous clinical trial results. Knowledge of characteristics of patients who are most likely to benefit from atezolizumab monotherapy is a crucial step towards implementing appropriate prescribing.*

Correspondence to: Hiroaki Satoh, MD, Ph.D., Division of Respiratory Medicine, Mito Medical Center, University of Tsukuba, 3100015, Mito, Ibaraki, Japan. Tel: +81 292312371, Fax: +81 292215137, e-mail: hirosato@md.tsukuba.ac.jp

Key Words: Atezolizumab, chemotherapy, progression-free survival, overall survival, immune-related adverse events.

Programmed death-1 (PD-1) is expressed on activated T cells (1, 2). By binding PD-L1 and PD-L2 expressed on cancer cells and antigen-presenting cells, this PD-L1 suppresses T cell activation, resulting in immune escape of cancer cells (1, 2). Anti-PD-1 antibodies bind to PD-1 on T cells and block the binding of PD-1 to PD-L1/PD-L2, thereby

Table I. *Clinical features of NSCLC patients treated with atezolizumab monotherapy.*

	Number of patients
Total number of patients	147
Sex, Male:Female	109:38
Performance status, 0-1:2 or more	103:44
Age, median (range) years	68 (43-87) years
70 years or older:69 years or less	58:89
75 years or older:74 years or less	28:119
Pathology, adenocarcinoma:squamous cell cancer:large cell cancer:others	105:27:2:13
Stage, IIIA-C:IVA-B	21:126
Driver gene, negative:positive	106:41
EGFR:ALK:others	33:3:5

NSCLC: Non-small cell lung cancer; EGFR: epidermal grwoth factor receptor; ALK: anaplastic lymphoma kinase.

blocking the transmission of inhibitory signals and activating T cells, and restoring the antitumor effect (1, 2). On the other hand, anti-PD-L1 antibodies block the interaction with PD-1 on T cells by binding to PD-L1 expressed by cancer cells and antigen-presenting cells (1, 2). As a result, inhibitory signaling to T cells is reduced and T cell activation is maintained (1, 2).

Atezolizumab is an anti-PD-L1 monoclonal antibody that targets PD-L1 expressed on cancer cells and antigen-presenting cells and inhibits its interaction with PD-1 on T cells, thus exerting its anti-tumor effect (3). PD-L1 expression has been observed in carcinomas arising from many organs, including non-small cell lung cancer (NSCLC) (3). For NSCLC, atezolizumab was shown to be useful in the OAK and IMpower110 trials (3, 4). Atezolizumab, like other anti-PD-1 antibodies, was initially used as a monotherapy, but has become popular in combination with chemotherapy (5, 6). Real-world clinical data for immune checkpoint inhibitor (ICI) monotherapy for NSCLC have been reported from several institutes (7-18). However, most were reports on anti-PD-1 antibodies, such as nivolumab and pembrolizumab (8-11, 17, 18). Although there have been studies on ICI monotherapy, including atezolizumab (7, 12-16), detailed clinical outcomes for patients treated with atezolizumab monotherapy have been limited (7, 12, 16). To the best of our knowledge, there have been only two reports on clinical outcomes for atezolizumab monotherapy, both with cohorts of less than 50 patients (12, 16).

ICI monotherapy might be prescribed as a third line of therapy or later, or as ICI re-challenge therapy in patients who have already received ICI combination chemotherapy. Although it seems difficult to expect a response, the clinical significance of this therapy is unclear. The aim of the study was to clarify the significance of atezolizumab monotherapy in clinical practice and to clarify the factors that contribute to long-term response and survival. We believe that this information will provide useful information for future treatment with atezolizumab.

Patients and Methods

This retrospective study reviewed patients with pathologically-diagnosed NSCLC from April 2018 to March 2023 at 11 Hospitals in our prefecture (Ibaraki Prefecture: 6,095 km^2). Among these patients, information from all patients who received atezolizumab monotherapy were assembled with no exclusion criteria. Pathologic diagnosis of NSCLC was determined according to the WHO classification. Prior to initiation of atezolizumab treatment, all patients underwent TNM classification (19). For imaging, head computed tomography or magnetic resonance imaging, bone scans, and ultrasonography and/or computed tomography of the abdomen were performed. Suitable patients were identified in each hospital's clinical database, and information on patient demographics [age, sex, Eastern Cooperative Oncology Group performance status (PS), histology, clinical stage, *etc*.] was extracted from databases. Tumor response was assessed as complete response, partial response, stable disease, progressive disease, or not evaluable according to Response Evaluation Criteria in Solid Tumors (RECIST) (20). Information was also extracted on response to treatment, duration of response and survival from initiation of atezolizumab. Adverse events were classified using the National Cancer Institute Common Terminology Criteria for Adverse Events (version 5.0) (21).

For statistical comparison between the two groups, the Chi-squared test and Mann-Whitney *U*-test were used. Survival probability was estimated with the Kaplan–Meier method and evaluated using the log-rank test and Cox's proportional hazard model. Multivariate analysis was performed using factors that had $p<0.02$ by univariate analysis. A p-value of <0.01 was considered to indicate a significant difference.

This study was approved by the Institutional Review Boards of the University of Tsukuba Mito Medical Center, Mito Kyodo General Hospital (NO-22-42) and each participating hospital.

Results

Patient characteristics. Clinical information on all the 147 patients who received atezolizumab monotherapy during the study period was compiled. Table I shows patient characteristics. Median age was 68 years (range=43-87 years) and 109 (74.1%) were men; 103 patients (70.1%) had PS 0-1

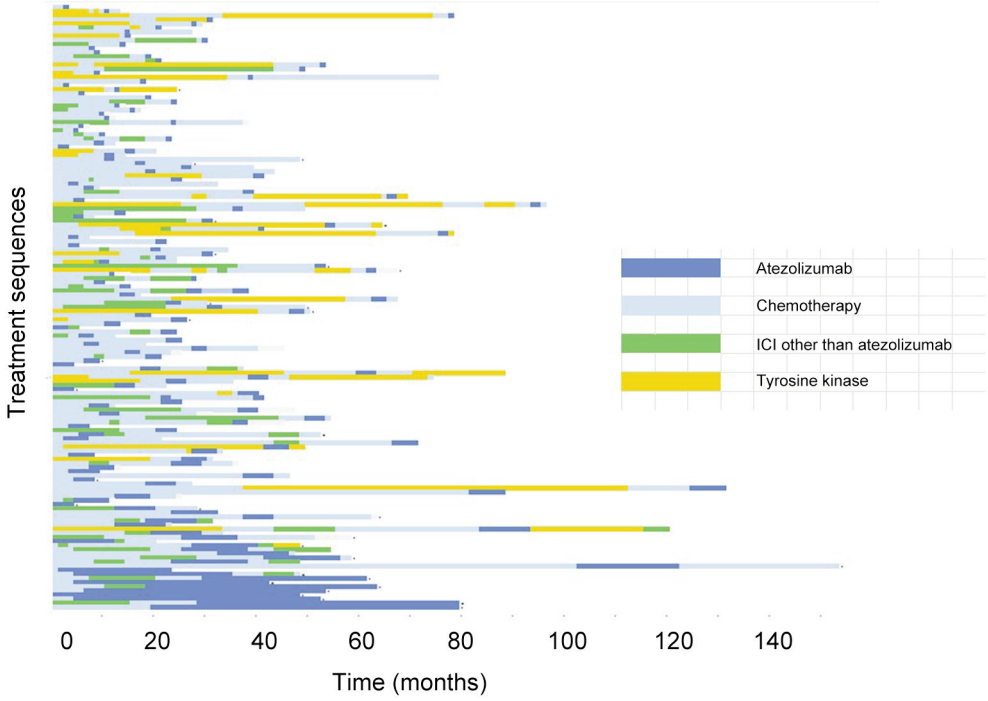


Figure 1. *Treatment sequences for the 147 patients with non-small cell lung cancer who were treated with atezolizumab monotherapy.*

and 105 patients (71.4%) had adenocarcinoma. The median 'treatment line' for atezolizumab therapy was third-line of therapy (range=1-10 lines). Seventy-four patients (50.3%) received first- to third-line therapy, and 73 patients (49.7%) had atezolizumab therapy in the fourth or later lines. Six patients (4.1%) received atezolizumab as first-line therapy.

Response to treatment. Figure 1 shows the specific treatment sequences for the 147 patients. The response rate of atezolizumab monotherapy was 15.0% (three complete response, 19 partial response). Thirty-five patients (23.8%) were observed to have stable disease, giving a disease control rate of 38.8%. There was no significant difference in response rate among sex (men, 9.2% *vs*. women, 7.9%; p=0.1936), PS (PS 0-1, 18.4% *vs*. PS 2-3, 6.8%; p=0.0810), or atezolizumab treatment line (lines 1-3, 18.9% *vs*. line 4 or later, 11.0%; p=0.2476). However, compared to patients with adenocarcinoma, patients with non-adenocarcinoma histology had a higher response rate (28.6% *vs*. 9.5%; p=0.001).

Survival analysis. Of the 147 patients evaluated, 114 (77.6%) had died at the time of analysis. The median follow-up time was 6.0 months [95% confidence interval (CI)=10.0-14.0 months] (Figure 2). One- and two-year overall survival (OS) was 34.7% (95% CI=26.9%-42.5%) and 15.0% (95% CI=9.1%-20.8%), respectively. Median progression-free survival (PFS) was 3.0 months (95% CI=2.3-3.7 months) and median OS was 7.0 months (95% CI=4.7-9.3 months).

In order to identify favorable factors affecting PFS and OS, univariate and multivariate analyses were performed using sex, PS, age, PD-L1, stage, NSCLC histology, driver genes, irAEs, and treatment line as variables. As shown in Table II, 'types of NSCLC other than adenocarcinoma' was a favorable prognostic factor in multivariate analysis for PFS. In multivariate analysis for OS, both 'good PS (0-1)' and 'treatment line up to third-line' were favorable prognostic factors for OS (Table II).

Sixteen patients (10.9%) had PFS >1 year. Comparing these 16 patients with 131 patients who had PFS <1 year, no characteristic clinical differences were found (Table III).

Toxicity. Table IV shows immune-related adverse events (irAEs). irAEs were observed in 13 cases (8.8%), of which grade 3 or higher was observed in nine cases (6.1%). Pulmonary toxicity was observed in three patients, two of whom were grade 3 and one was grade 5. Other irAEs included hepatobiliary toxicity in three patients (one grade 1, one grade 2, one grade 4), two with thyroid dysfunction (two grade 2), and one with pituitary dysfunction. There was also one case of pneumothorax (grade 2), one case of hemoptysis, one case of hyperglycemia (grade 3), and one case of arthralgia (grade 3).

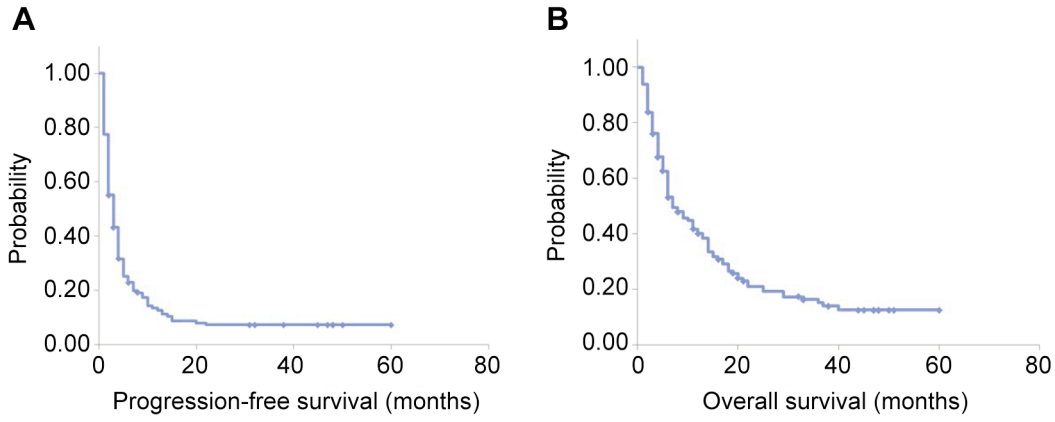


Figure 2. *In the 147 patients who were treated with atezolizumab monotherapy, median progression-free survival was 3.0 months [95% confidence interval (CI)=2.3-3.7 months] (A) and median overall survival was 7.0 months (95% CI=4.7-9.3 months) (B).*

Table II. *Uni- and multivariate analysis of survival from the initiation of atezolizumab monotherapy.*

	Progression free survival				Overall survival			
	Univariate analysis (*p*-Value)	Multivariate analysis			Univariate analysis (*p*-Value)	Multivariate analysis		
		Odds ratio	95% CI	*p*-Value		Odds ratio	95% CI	*p*-Value
Sex, male	0.0467	0.6817	0.456-1.020	0.0624	0.2095			
Performance status, 0-1	0.1165	1.4266	0.974-2.090	0.0683	0.0069	1.7582	1.163-2.659	0.0075
Age, less than 70 years	0.9479				0.4624			
PD-L1, 50% or more	0.9369				0.7323			
Stage, IIIA-C	0.0265				0.1489			
Pathology, non-AD	0.0027	0.6522	0.433-0.982	0.0404	0.7086			
Driver gene, negative	0.2898				0.7717			
irAEs, grade 1 or more	0.6443				0.4489			
Treatment line, 1-3	0.0175	1.3175	0.923-1.881	0.1288	0.0188	1.4596	1.004-2.122	0.0477
Prior ICI before atezolizumab, present	0.4510				0.1597			

PD-L1: Programmed cell death ligand 1; AD: adenocarcinoma; irAE: immune-related adverse events; ICI: immune checkpoint inhibitors; CI: confidence interval.

Discussion

The 147 patients evaluated had a PFS of 3.0 months and an OS of 7.0 months. IrAEs were observed in 13 patients (8.8%), with grade 3 or higher in nine patients (6.1%), and grade 5 with pulmonary toxicity in one patient (0.7%). Favorable factors related to PFS were 'types of NSCLC other than adenocarcinoma'. Favorable factors for OS were 'PS 0-1' and 'treatment line up to 3'. There were 16 patients (10.9%) with PFS >1 year. No characteristic clinical findings were found in these 16 patients compared to the remaining 131 patients.

Anti-PD-1 antibodies, such as nivolumab and pembrolizumab, suppress immune checkpoints by binding to PD-1 on T cells, a type of immune cell (1, 2). On the other hand, the PD-L1 antibody, atezolizumab, suppresses immune

Table III. *Comparison of clinical features in patients who had progression-free survival more than one year (Group A) and those who did not have (Groups B).*

	Group A	Group B	*p*-Value
Number of patients	16	131	
Sex, male:female	15:1	94:37	0.0709
PS, 0-1:2-4	13:3	90:41	0.3938
Age, <69:≥70 years	7:9	74:57	0.4268
PD-L1, <50%:>50%	13:3	105:26	0.9999
Stage, IIIC:IVA-B	4:12	17:114	0.2462
Pathology, adenocarcinoma:others	8:8	97:34	0.0743
Driver gene, positive:negative	2:14	39:92	0.2364
irAEs, absent:present	4:12	16:115	0.2360

PS: Performance status; PD-L1: programmed cell death ligand 1; irAE: immune-related adverse events.

Table IV. *Immune-related adverse events.*

	Any grades	Grade 3-4	Grade 5
Pulmonary toxicity	3	2	1
Hepatobiliary toxicity	3	2	0
Thyroid dysfunction	2	2	0
Pituitary toxicity	1	1	0
Hemoptysis	1	1	0
Hyperglycemia	1	1	0
Arthralgia	1	1	0
Pneumothorax	1	0	0
Hyperamylasemia	1	0	0

checkpoints by binding to PD-L1 on cancer cells and antigen-presenting cells (1, 2). Anti-PD-1 antibodies are known to bind to PD-L1 and PD-L2, and anti-PD-L1 antibodies are known to bind to PD-1 and B7-1. This means that the inhibitory binding is slightly different: anti-PD-1 antibodies can, therefore, block both PD-1/PD-L1 binding and PD-1/PD-L2 binding (1, 2). However, PD-L1 antibodies block PD-1/PD-L1 binding, but not PD-1/PD-L2 binding. Alternatively, PD-L1 antibodies can block B7-1/PD-L1 binding (1, 2). It has been speculated that this difference might affect clinical outcomes, as well as differences in efficacy and irAEs (3). Although it is generally accepted that there might be little difference in the efficacy and irAEs between anti-PD-1 and anti-PL-L1 antibodies, information regarding atezolizumab monotherapy for NSCLC in clinical practice is insufficient.

In clinical trials of ICI monotherapy for NSCLC in patients on second and subsequent lines of therapy, PFS and OS for anti-PD-1 antibodies are reported to be 2-7 months and 5.8-18 months, respectively (12). A clinical trial of single-agent PD-L1 antibody for NSCLC patients (OAK trial) reported a PFS of 2.8 months and an OS of 12.7 months (3). This trial did not include patients with poor PS and was limited to patients receiving atezolizumab on treatment lines 2-3. Although some reports included pembrolizumab, most of the clinical results of ICI monotherapy for NSCLC patients were for nivolumab monotherapy. The PFS and OS were 1.8-3.3 months and 5.9-14.6 months, respectively (10).

There are only two reports of PFS and OS with atezolizumab monotherapy in clinical practice. In them, PFS and OS were 1.4 to 2.0 months and 6.5 to 12.8 months, respectively (11, 16). However, the number of patients investigated in these two studies was 38 and 43, respectively, and they were not fully evaluated (11, 16). The PFS of the present study was similar to those of clinical trials of anti-PD-1 and anti-PD-L1 antibodies in clinical practice. The median age of patients in this study was 68 years, and poor PS was 29.9%, resulting in an OS of 7.0 months. These results were similar to those of Weis *et al*. who administered atezolizumab to 43 second-line patients (PFS, 2.0 months; OS, 6.5 months) (16). The median age of their patients was 67.2 years, and 20.7% of the cohort had poor PS.

Several studies have performed multivariate analyses on prognostic factors in ICI monotherapy. Most were studies on anti-PD-1 antibodies, and adverse factors cited included poor PS, low PD-L1, epidermal growth factor receptor positivity, tumor size, increased platelets, and bone metastasis (7-9). Only the report by Furuya *et al*. evaluated prognostic factors for an anti-PD-L1 antibody. They reported that 'good PS' was a favorable factor for both PFS and OS in 38 patients who received atezolizumab monotherapy after anti-PD-1 antibodies (11). Good PS was also a favorable factor in OS in our study. However, if patient characteristics and treatment sequence and other clinical conditions are different, different results might be obtained. As such, caution is required and additional data in this area is needed. To our knowledge, long-term treatment with atezolizumab monotherapy has not been defined, and there have been no reports of patients receiving long-term treatment. Furuya *et al*. reported that seven of 38 patients were able to receive atezolizumab monotherapy for at least 4.0 months (11). Among our patients, 16 (10.9%) had >1 year of atezolizumab monotherapy. Although favorable clinical factors could not be identified, it is noteworthy that there were clearly individual patients who maintained a long-term response.

A total of 60.5% of patients in the OAK trial had irAEs of any grade (3). In clinical trials of single-agent ICI therapy, Sonpavde *et al*. investigated the irAEs that developed in the 8,730 patients in the trials (22). Their review showed a lower frequency of irAEs with the anti-PD-1 antibody, atezolizumab, than those with anti-PD-L1 antibodies. They speculated that these results were due to differences in the way anti-PD-L1 and anti-PD-1 antibodies acted (1, 2). On the other hand, Mencoboni *et al*. reviewed atezolizumab monotherapy irAEs in clinical practice. In their review, most patients were treated with the anti-PD-1 antibody, nivolumab, and they reported an incidence of irAEs of any grade of 7%-71%, and of grades 3-4 of 0-25% (12). Patients with grade 5 lung injury were also reported (23). In our study, irAEs of any grade were observed in only 8.8% of patients, of which grade 3 or higher was observed in 6.1%. One patient developed grade 5 pulmonary toxicity. Although the incidence of irAEs of any grade appeared to be very low, the possibility of under-evaluation in retrospective studies cannot be ruled out.

There are other limitations in this study that should be mentioned. Although the study included the largest number of patients reported for such a study, it was a retrospective study of patients with a wide range of background characteristics. It should be noted that our results are not definitive and do not allow final conclusions. Novel driver gene examinations such

as Kristen rat sarcoma viral oncogene homolog (KRAS), c-ros oncogene 1 (ROS1) and rearranged during transfection (RET) were introduced into clinical practice during the study period. Identification of these driver genes as well as EGFR and anaplastic lymphoma kinase (ALK) and progress of corresponding therapeutic agents have extended life and provided palliation for lung cancer-patients positive for these mutations (24). As shown in Table I, this study included 8 patients positive for driver genes other than EGFR [4 KRAS positive, 3 ALK fusion gene-positive, 1 B-Raf proto-oncogene, serine/threonine kinase (BRAF) positive]. Those genes could not be evaluated equally in all patients. Therefore, it was possible that this situation affected the results. Because of the small number of these patients and because it was impossible to newly investigate the driver genes that were undetectable at the beginning of this study, we treated them as EGFR gene-negative patients in this study. It might have been better to investigate the prognosis separately for driver gene-positive and -negative patients for these driver genes. Since there was a high proportion of epidermal growth factor receptor gene-positive patients, it might be possible to analyze this cohort separately from this report.

The results of this study indicate that real-world atezolizumab monotherapy in patients with unfavorable clinical conditions, might achieve PFS and OS similar to those reported in previous clinical trials. The frequency and severity of irAEs were similar to those of previous reports in trials and practice with ICI monotherapy, confirming the existence of cases in which long-term administration is possible. Although ICI has entered the era of combination therapy with chemotherapy, the selection and appropriate management of patients who can benefit from atezolizumab monotherapy is still important. Efforts to identify patients who could benefit from atezolizumab monotherapy will continue to be of value.

Conflicts of Interest

The Authors have no conflicts of interest to declare in relation to this study.

Authors' Contributions

SO, KM, TS, SH, HSaku, and HS designed the study. TN, HY, TS, YW, SO, TT, NK, KM, SH, HirofS, TY, KK, MI, HS, HI, TK, TE, and TS collected the data. SH, SO, RN and HS analyzed the data. SH, SO, HS, and NH prepared the manuscript. AN, HS, and NH supervised the study. All Authors approved the final version for submission.

References

1 Khadela A, Postwala H, Rana D, Dave H, Ranch K, Boddu SHS: A review of recent advances in the novel therapeutic targets and immunotherapy for lung cancer. Med Oncol 40(5): 152, 2023. DOI: 10.1007/s12032-023-02005-w

2 Lahiri A, Maji A, Potdar PD, Singh N, Parikh P, Bisht B, Mukherjee A, Paul MK: Lung cancer immunotherapy: progress, pitfalls, and promises. Mol Cancer 22(1): 40, 2023. DOI: 10.1186/s12943-023-01740-y

3 Gandara DR, Von Pawel J, Mazieres J, Sullivan R, Helland Å, Han J, Ponce Aix S, Rittmeyer A, Barlesi F, Kubo T, Park K, Goldschmidt J, Gandhi M, Yun C, Yu W, Matheny C, He P, Sandler A, Ballinger M, Fehrenbacher L: Atezolizumab treatment beyond progression in advanced NSCLC: Results from the randomized, phase III OAK study. J Thorac Oncol 13(12): 1906-1918, 2018. DOI: 10.1016/j.jtho.2018.08.2027

4 Herbst RS, Giaccone G, De Marinis F, Reinmuth N, Vergnenegre A, Barrios CH, Morise M, Felip E, Andric Z, Geater S, Özgüroğlu M, Zou W, Sandler A, Enquist I, Komatsubara K, Deng Y, Kuriki H, Wen X, Mccleland M, Mocci S, Jassem J, Spigel DR: Atezolizumab for First-Line Treatment of PD-L1–Selected Patients with NSCLC. N Engl J Med 383(14): 1328-1339, 2020. DOI: 10.1056/NEJMoa1917346

5 Ryu R, Ward KE: Atezolizumab for the first-line treatment of non-small cell lung cancer (NSCLC): Current status and future prospects. Front Oncol 8: 277, 2018. DOI: 10.3389/fonc.2018.00277

6 Boussageon M, Swalduz A, Chouaïd C, Bylicki O: First-line treatment of advanced non-small-cell lung cancer with immune-checkpoint inhibitors: New combinations and long-term data. BioDrugs 36(2): 137-151, 2022. DOI: 10.1007/s40259-022-00515-z

7 Knetki-Wróblewska M, Tabor S, Piórek A, Płużański A, Winiarczyk K, Zaborowska-Szmit M, Zajda K, Kowalski DM, Krzakowski M: Nivolumab or atezolizumab in the second-line treatment of advanced non-small cell lung cancer? A prognostic index based on data from daily practice. J Clin Med 12(6): , 2023. DOI: 10.3390/jcm12062409

8 Juarez-Garcia A, Sharma R, Hunger M, Kayaniyil S, Penrod JR, Chouaïd C: Real-world effectiveness of immunotherapies in pre-treated, advanced non-small cell lung cancer Patients: A systematic literature review. Lung Cancer 166: 205-220, 2022. DOI: 10.1016/j.lungcan.2022.03.008

9 Choi MG, Choi CM, Lee DH, Kim SW, Yoon S, Ji W, Lee JC: Impact of gender on response to immune checkpoint inhibitors in patients with non-small cell lung cancer undergoing second- or later-line treatment. Transl Lung Cancer Res 11(9): 1866-1876, 2022. DOI: 10.21037/tlcr-22-146

10 Ivanović M, Knez L, Herzog A, Kovačević M, Cufer T: Immunotherapy for metastatic non-small cell lung cancer: real-world data from an academic central and eastern european center. Oncologist 26(12): e2143-e2150, 2021. DOI: 10.1002/onco.13909

11 Furuya N, Nishino M, Wakuda K, Ikeda S, Sato T, Ushio R, Tanzawa S, Sata M, Ito K: Real-world efficacy of atezolizumab in non-small cell lung cancer: A multicenter cohort study focused on performance status and retreatment after failure of anti-PD-1 antibody. Thorac Cancer 12(5): 613-618, 2021. DOI: 10.1111/1759-7714.13824

12 Mencoboni M, Ceppi M, Bruzzone M, Taveggia P, Cavo A, Scordamaglia F, Gualco M, Filiberti RA: Effectiveness and safety of immune checkpoint inhibitors for patients with advanced non small-cell lung cancer in real-world: Review and meta-analysis. Cancers (Basel) 13(6): 1388, 2021. DOI: 10.3390/cancers13061388

13 Ayers KL, Mullaney T, Zhou X, Liu JJ, Lee K, Ma M, Jones S, Li L, Redfern A, Jappe W, Liu Z, Goldsweig H, Yadav KK, Hahner N, Dietz M, Zimmerman M, Prentice T, Newman S,

Veluswamy R, Wisnivesky J, Hirsch FR, Oh WK, Li SD, Schadt EE, Chen R: Analysis of real-world data to investigate the impact of race and ethnicity on response to programmed cell death-1 and programmed cell death-ligand 1 inhibitors in advanced non-small cell lung cancers. Oncologist 26(7): e1226-e1239, 2021. DOI: 10.1002/onco.13780

14 Geiger-gritsch S, Olschewski H, Kocher F, Wurm R, Absenger G, Flicker M, Hermann A, Heininger P, Fiegl M, Zechmeister M, Endel F, Wild C, Pall G: Real-world experience with anti-PD-1/PD-L1 monotherapy in patients with non-small cell lung cancer. Wien Klin Wochenschr 133(21-22): 1122-1130, 2021. DOI: 10.1007/s00508-021-01940-w

15 Stinchcombe TE, Miksad RA, Gossai A, Griffith SD, Torres AZ: Real-world outcomes for advanced non-small cell lung cancer patients treated with a PD-L1 inhibitor beyond progression. Clin Lung Cancer 21(5): 389-394.e3, 2020. DOI: 10.1016/j.cllc.2020.04.008

16 Weis TM, Hough S, Reddy HG, Daignault-Newton S, Kalemkerian GP: Real-world comparison of immune checkpoint inhibitors in non-small cell lung cancer following platinum-based chemotherapy. J Oncol Pharm Pract 26(3): 564-571, 2020. DOI: 10.1177/1078155219855127

17 Morita R, Okishio K, Shimizu J, Saito H, Sakai H, Kim YH, Hataji O, Yomota M, Nishio M, Aoe K, Kanai O, Kumagai T, Kibata K, Tsukamoto H, Oizumi S, Fujimoto D, Tanaka H, Mizuno K, Masuda T, Kozuki T, Haku T, Suzuki H, Okamoto I, Hoshiyama H, Ueda J, Ohe Y: Real-world effectiveness and safety of nivolumab in patients with non-small cell lung cancer: A multicenter retrospective observational study in Japan. Lung Cancer 140: 8-18, 2020. DOI: 10.1016/j.lungcan.2019.11.014

18 Barlesi F, Dixmier A, Debieuvre D, Raspaud C, Auliac JB, Benoit N, Bombaron P, Moro-Sibilot D, Audigier-Valette C, Asselain B, Egenod T, Rabeau A, Fayette J, Sanchez ML, Labourey JL, Westeel V, Lamoureux P, Cotte FE, Allan V, Daumont M, Dumanoir J, Reynaud D, Calvet CY, Ozan N, Pérol M: Effectiveness and safety of nivolumab in the treatment of lung cancer patients in France: preliminary results from the real-world EVIDENS study. Oncoimmunology 9(1): 1744898, 2020. DOI: 10.1080/2162402X.2020.1744898

19 Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, Nicholson AG, Groome P, Mitchell A, Bolejack V, Goldstraw P, Rami-Porta R, Asamura H, Ball D, Beer DG, Beyruti R, Bolejack V, Chansky K, Crowley J, Detterbeck F, Erich Eberhardt WE, Edwards J, Galateau-Sallé F, Giroux D, Gleeson F, Groome P, Huang J, Kennedy C, Kim J, Kim YT, Kingsbury L, Kondo H, Krasnik M, Kubota K, Lerut A, Lyons G, Marino M, Marom EM, Van Meerbeeck J, Mitchell A, Nakano T, Nicholson AG, Nowak A, Peake M, Rice T, Rosenzweig K, Ruffini E, Rusch V, Saijo N, Van Schil P, Sculier J, Shemanski L, Stratton K, Suzuki K, Tachimori Y, Thomas CF, Travis W, Tsao MS, Turrisi A, Vansteenkiste J, Watanabe H, Wu Y, Baas P, Erasmus J, Hasegawa S, Inai K, Kernstine K, Kindler H, Krug L, Nackaerts K, Pass H, Rice D, Falkson C, Filosso PL, Giaccone G, Kondo K, Lucchi M, Okumura M, Blackstone E, Abad Cavaco F, Ansótegui Barrera E, Abal Arca J, Parente Lamelas I, Arnau Obrer A, Guijarro Jorge R, Ball D, Bascom G, Blanco Orozco A, González Castro M, Blum M, Chimondeguy D, Cvijanovic V, Defranchi S, De Olaiz Navarro B, Escobar Campuzano I, Macía Vidueira I, Fernández Araujo E, Andreo García F, Fong K, Francisco Corral G, Cerezo González S, Freixinet Gilart J, García Arangüena L, García Barajas S, Girard P, Goksel T, González Budiño M, González Casaurrán G, Gullón Blanco J, Hernández Hernández J, Hernández Rodríguez H, Herrero Collantes J, Iglesias Heras M, Izquierdo Elena J, Jakobsen E, Kostas S, León Atance P, Núñez Ares A, Liao M, Losanovscky M, Lyons G, Magaroles R, De Esteban Júlvez L, Mariñán Gorospe M, Mccaughan B, Kennedy C, Melchor Íñiguez R, Miravet Sorribes L, Naranjo Gozalo S, Álvarez De Arriba C, Núñez Delgado M, Padilla Alarcón J, Peñalver Cuesta J, Park J, Pass H, Pavón Fernández M, Rosenberg M, Ruffini E, Rusch V, Sánchez De Cos Escuín J, Saura Vinuesa A, Serra Mitjans M, Strand T, Subotic D, Swisher S, Terra R, Thomas C, Tournoy K, Van Schil P, Velasquez M, Wu Y, Yokoi K: The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol 11(1): 39-51, 2016. DOI: 10.1016/j.jtho.2015.09.009

20 Eisenhauer E, Therasse P, Bogaerts J, Schwartz L, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). 1) Eur J Cancer 45(2): 228-247, 2009. DOI: 10.1016/j.ejca.2008.10.026

21 National Cancer Institute Common Terminology Criteria for Adverse Events (version 5.0). Available at: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fctep.cancer.gov%2FprotocolDevelopment%2Felectronic_applications%2Fdocs%2FCTCAE_v5.0.xlsx&wdOrigin=BROWSELINK [Last accessed on February 16, 2023]

22 Spagnolo P, Chaudhuri N, Bernardinello N, Karampitsakos T, Sampsonas F, Tzouvelekis A: Pulmonary adverse events following immune checkpoint inhibitors. Curr Opin Pulm Med 28(5): 391-398, 2022. DOI: 10.1097/MCP.0000000000000895

23 Sonpavde GP, Grivas P, Lin Y, Hennessy D, Hunt JD: Immune-related adverse events with PD-1 versus PD-L1 inhibitors: a meta-analysis of 8730 patients from clinical trials. Future Oncol 17(19): 2545-2558, 2021. DOI: 10.2217/fon-2020-1222

24 Check JH, Poretta T, Check D, Srivastava M: Lung cancer – standard therapy and the use of a novel, highly effective, well tolerated, treatment with progesterone receptor modulators. Anticancer Res 43(3): 951-965, 2023. DOI: 10.21873/anticanres.16240

Received June 7, 2023
Revised July 1, 2023
Accepted July 7, 2023

ANTICANCER RESEARCH *43*: xxx-xxx (2023)
doi:10.21873/anticanres.11xxx

Atezolizumab for EGFR-mutated Non-small Cell Lung Cancer Patients: An Observation Study in Ibaraki Group (ATTENTION-IBARAKI)

TOMOHIRO TAMURA[1], TAKESHI NUMATA[2], HIDETOSHI YANAI[2], RYOTA NAKAMURA[2], TOSHIHIRO SHIOZAWA[3], HIROKO WATANABE[4], SHINICHIRO OKAUCHI[5], NORIHIRO KIKUCHI[6], KUNIHIKO MIYAZAKI[7], SHIGEN HAYASHI[8], TAKAAKI YAMASHITA[9], KOICHI KURISHIMA[10], MASAHARU INAGAKI[11], HIROAKI SATOH[5], HIROICHI ISHIKAWA[10], TAKAYUKI KABURAGI[11], TAKEO ENDO[2], TORU SAKAMOTO[4] and NOBUYUKI HIZAWA[3]

[1]Respiratory Center, Ibaraki Prefectural Central Hospital, Kasama, Japan;
[2]Departments of Respiratory Medicine and Surgery, National Hospital Organization Mito Medical Center, Ibarakimachi, Japan;
[3]Division of Respiratory Medicine, Faculty of Clinical Medicine, University of Tsukuba, Tsukuba, Japan;
[4]Division of Respiratory Medicine, Tsukuba Memorial Hospital, Tsukuba, Japan;
[5]Division of Respiratory Medicine, Mito Medical Center, University of Tsukuba, Mito, Japan;
[6]Division of Respiratory Medicine, National Hospital Organization Kasumigaura Medical Center, Tsuchiura, Japan;
[7]Division of Respiratory Medicine, Ryugasaki Saiseikai Hospital, Ryugasaki, Japan;
[8]Division of Respiratory Medicine, Ibaraki Seinan Medical Center Hospital, Sakai, Japan;
[9]Division of Respiratory Medicine, JA Toride Medical Center Hospital, Toride, Japan;
[10]Division of Respiratory Medicine, Tsukuba Medical Center Hospital, Tsukuba, Japan;
[11]Division of Thoracic Surgery, Tsuchiura Kyodo General Hospital, Tsuchiura, Japan

Abstract. *Background/Aim: Atezolizumab, an anti-programed death-ligand 1 monoclonal antibody, targets programed death-ligand 1 expressed on cancer cells and antigen-presenting cells and is now commonly used in combination with chemotherapy. We conducted a study to clarify the efficacy of atezolizumab in epidermal growth factor receptor (EGFR)-mutated patients who are considered less responsive to immune checkpoint inhibitors. Patients and Methods: A retrospective review of patients with advanced non-small cell lung cancer (NSCLC) who received atezolizumab-containing therapy at 11 hospitals from April 2018 to March 2023 was performed. Results: Median progression-free survival and overall survival in 33 EGFR-mutated patients treated with atezolizumab monotherapy were 2.0 and 9.0 months, respectively, and those in 19 patients who received combined atezolizumab plus chemotherapy were 12.0 and 17.0 months, respectively. When comparing EGFR-mutated and EGFR-negative patients after propensity score matching, there were no significant differences in progression-free survival and overall survival between the two groups, whether atezolizumab monotherapy or combined atezolizumab plus chemotherapy. Among EGFR-mutated patients, being male was a significant favorable factor in both atezolizumab treatment groups. None of the EGFR-mutated patients had grade 5 immune-related adverse events. Conclusion: Efficacy of atezolizumab in EGFR-mutated NSCLC patients could be comparable to that for EGFR-negative patients. To prolong the survival of EGFR-mutated NSCLC patients, appropriate selection and sequencing of EGFR for tyrosine kinase inhibitors, as well as immune checkpoint inhibitors, anti-tumor agents, and anti-angiogenic agents are important.*

Correspondence to: Hiroaki Satoh, MD, Ph.D., Division of Respiratory Medicine, Mito Medical Center, University of Tsukuba, 3100015, Mito, Ibaraki, Japan. Tel: +81 292312371, Fax: +81 292215137, e-mail: hirosato@md.tsukuba.ac.jp

Key Words: Atezolizumab, epidermal growth factor receptor, progression-free survival, overall survival, immune-related adverse events.

Immune checkpoint inhibitors (ICIs) that have revolutionized the treatment of advanced non-small cell lung cancer (NSCLC) include anti-programed death 1 (PD-1) and anti-programed death-ligand 1 (PD-L1) antibodies that possess different mechanisms of action (1). As with other anticancer drugs, clinical trials of ICI monotherapy were conducted initially,

followed by a clinical trial of combined ICI and chemotherapy (1). In addition, not only the results of clinical trials but also real-practice data have been accumulated (2), and factors related to favorable response, especially in NSCLC without driver genes, have been investigated (3, 4). In patients with epidermal growth factor receptor (EGFR)-mutated NSCLC, EGFR tyrosine kinase inhibitors (TKIs) have been the standard drugs and are usually the first choice of treatment. Perhaps for this reason, EGFR-mutated patients have not been included in clinical trials of ICIs. EGFR-mutated NSCLC is, however, considered difficult to cure with EGFR-TKIs (5). As such, treatment regimens including ICIs might be one of the treatment options after EGFR-TKIs are no longer effective. ICIs might also be a treatment option for patients who could not continue EGFR-TKIs due to adverse events. Against this background, clinical trials of ICI-containing therapies for EGFR-mutated NSCLC have begun (6, 7). Based on results from these trials, it seems that EGFR-mutated patients are less responsive to ICIs and have a shorter duration of response than EGFR-negative patients (6, 7). However, to the best of our knowledge, none of these studies used propensity matching to adjust for background factors or evaluated the efficacy of ICIs in EGFR-mutated patients in real clinical settings.

An additional factor in the treatment of patients with EGFR-mutated NSCLC is whether a patient who has become resistant to an EGFR-TKI may respond to TKI re-challenge after treatment with different agents. As such, the treatment sequence of EGFR-TKIs, as well as ICIs and conventional anticancer drugs, are attracting attention in the treatment of EGFR-mutated NSCLC patients (8-10). Aside from treatment efficacy, a caution was issued regarding the occurrence of serious pulmonary injury as an immune-related adverse event (irAE) when a TKI, osimertinib, was administered after the anti-PD-1 antibody, nivolumab (11). To avoid the onset of serious irAEs, there is increasing interest in the treatment sequence of therapeutic agents, including ICIs, after TKIs (11, 12). Clinical trials on treatment sequences could be difficult to conduct in patients with EGFR-mutated NSCLC due to the long duration of response. Therefore, it is important to accumulate and share information from real clinical practice.

Here, we focused on atezolizumab, an anti-PD-L1 antibody, and investigated its clinical usefulness in EGFR-mutated NSCLC patients. The reason why we selected this ICI among others was that there are treatment results for atezolizumab in EGFR-mutated patients in clinical trials and real clinical practice, and we thought that more accurate information could be obtained by comparison with our results. The purpose of this study was to investigate the therapeutic results of this drug in actual clinical practice, clarify the current state of treatment, and contribute to the future medical care of EGFR-mutated NSCLC patients.

Patients and Methods

This retrospective study included patients pathologically diagnosed with NSCLC who underwent atezolizumab monotherapy or combined atezolizumab plus chemotherapy from April 2018 to March 2023 at 11 hospitals in our prefecture (Ibaraki Prefecture, 6,095 km^2). Pathological diagnosis was based on the WHO classification. All patients underwent TNM classification (13) using head computed tomography or magnetic resonance imaging, bone scans, or 2-deoxy-2-[18F]fluoro-D-glucose positron emission tomography/computed tomography (PET/CT), and ultrasonography and/or CT of the abdomen prior to initiation of therapy. Suitable patients were identified in clinical databases in each hospital, and the following information extracted from their records: patient demographics at the time of commencing atezolizumab monotherapy or combined atezolizumab plus chemotherapy [age, sex, Eastern Cooperative Oncology Group performance status (PS), histology, stage], objective tumor response, EGFR mutation status, PD-L1 tumor proportion score (TPS), and objective tumor response, progression-free survival (PFS), overall survival (OS), and immune-related adverse events (irAEs). Patient survival time was calculated from the initiation date of atezolizumab to the date of event or latest follow-up contact. According to the Response Evaluation Criteria in Solid Tumors, the tumor response was evaluated as complete response, partial response, stable disease, progressive disease, or not evaluable (14). Adverse events were classified using the National Cancer Institute Common Terminology Criteria for Adverse Events (version 5.0) (15).

The chi-squared test and Mann-Whitney *U*-test were used for statistical comparison between groups. To ensure that the group of EGFR-mutated patients and EGFR-negative patients were as similar as possible, we used propensity score matching. Selected covariates included sex, PS, age, stage, PD-L1 treatment, and grade 1-4 irAEs. Matching was carried out using a ratio of 1:1, and a caliper distance of 0.030, without replacement. Using the log-rank test and Cox's proportional hazard model, survival probability was estimated with the Kaplan-Meier method. A multivariate analysis was carried out with the significant factors identified in the univariate analysis. Multivariate analysis analyzed factors with *p*-values less than 0.2 by univariate analysis. A p-value <0.01 was considered to indicate a significant difference.

This research was approved by the University of Tsukuba Mito Medical Center-Mito Kyodo General Hospital (NO-22-42) and each Institutional Review Board.

Results

Patient characteristics. Clinical information on 265 patients was assembled during the study period. In all, 147 patients were treated with atezolizumab monotherapy, and 118 patients were treated with combined atezolizumab plus chemotherapy. Across the cohort, there were 52 EGFR-mutated and 213 EGFR-negative patients. Among the 52 EGFR-mutated patients, 33 were treated with atezolizumab monotherapy, and 19 were treated with combined atezolizumab plus chemotherapy. Of the 19 patients, 17 received combination therapy with carboplatin, paclitaxel, and bevacizumab. Table I shows the background characteristics of these patients. In

Table I. *Backgrounds of clinical features in non-small cell carcinoma patients with or without EGFR.*

	EGFR positive	EGFR negative	*p*-value
Atezolizumab monotherapy			
Total number of patients	33	114	
Sex Male: Female	18:15	91:23	0.0061
Performance status, 0-1:2 or more	24:9	79:35	0.8302
Age, less than 70, 70 or more (years)	21:12	60:54	0.3220
Pathology, adenocarcinoma:other than adenocarcinoma	32:1	73:41	0.0001
Stage, IIIA-C:IVA-B	4:29	17:97	0.7852
PD-L1, more than 50: 50 ore less (%)	3:30	17:97	0.5662
irAEs grade 1-4, present:absent	4:29	25:89	0.3200
Combined atezolizumab and chemotherapy			
Total number of patients	19	99	
Sex Male: Female	7:12	81:18	0.0002
Performance status, 0-1:2 or more	17:2	88:11	0.9999
Age, less than 70, 70 or more (years)	9:10	51:48	0.8054
Pathology, adenocarcinoma:other than adenocarcinoma	19:0	73:26	0.0123
Stage, IIIA-C:IVA-B	2:17	12:87	0.9999
PD-L1, more than 50: 50 or less (%)	6:13	21:78	0.3732
irAEs grade 1-4, present:absent	3:16	24:75	0.5578

EGFR: Epidermal growth factor receptor; PD-L1: programmed cell death-ligand 1; irAE: immune-related adverse events.

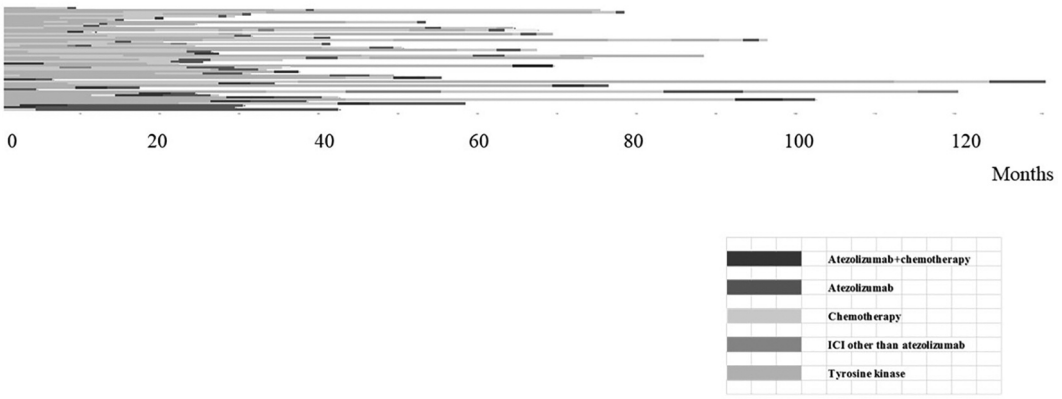


Figure 1. *The specific treatment sequences are shown for the 52 epidermal growth factor receptor-mutated patients with non-small cell lung cancer who were treated with atezolizumab monotherapy and combined atezolizumab plus chemotherapy.*

patients receiving atezolizumab monotherapy, there were significant differences in sex and histology between EGFR-mutated and EGFR-negative patients. Similarly, there were also significant differences in sex and histology between EGFR-mutated and EGFR-negative patients treated with combined atezolizumab plus chemotherapy.

Survival analysis. Of 147 patients treated with atezolizumab monotherapy (EGFR-mutated and EGFR-negative combined), 114 (77.6%) had died at the time of analysis. The median follow-up time was 6.0 months [95% confidence interval (CI)=10.0-14.0 months]. In 33 EGFR-mutated patients treated with atezolizumab monotherapy, the median PFS was 2.0 months (95%CI=1.0-3.0 months) and median OS was 9.0 months (95%CI=2.7-15.3 months). Of 118 patients treated with combined atezolizumab plus chemotherapy (EGFR-mutated and EGFR-negative combined), 58 (49.2%) had died at the time of analysis. The median follow-up time was 12.0 months (95%=13.0-16.9 months). In 19 EGFR-mutated patients treated with combined atezolizumab plus chemotherapy, the median PFS was 12.0 months (95%CI=6.3-17.7 months) and the median OS was 17.0 months (95%CI=8.3-25.7 months). Figure 1 shows the specific treatment sequences for the patient cohort.

As shown in Table I, EGFR-mutated and EGFR-negative patients had different characteristics. EGFR-mutated patients

Table II. *Backgrounds of clinical features in propensity matched non-small cell carcinoma patients with or without EGFR.*

	EGFR positive	EGFR negative	*p*-value
Atezolizumab monotherapy			
Total number of patients	33	33	
Sex Male: Female	18:15	18:15	0.9999
Performance status, 0-1:2 or more	24:9	26:7	0.7746
Age, less than 70, 70 or more (years)	21:12	21:12	0.9999
Pathology, adenocarcinoma:other than adenocarcinoma	1:32	0:33	0.9999
Stage, IIIA-C:IVA-B	4:29	3:30	0.9999
PD-L1, more than 50:50 ore less (%)	3:30	9:24	0.1081
irAEs grade 1-4, present:absent	4:29	2:31	0.6724
Combined atezolizumab and chemotherapy			
Total number of patients	19	19	
Sex Male: Female	7:12	7:12	0.9999
Performance status, 0-1:2 or more	17:2	17:2	0.9999
Age, less than 70, 70 or more (years)	9:10	8:11	0.9999
Pathology, adenocarcinoma:other than adenocarcinoma	19:0	19:0	0.9999
Stage, IIIA-C:IVA-B	2:17	1:18	0.9999
PD-L1, more than 50: 50 or less (%)	6:13	3:16	0.4470
irAEs grade 1-4, present:absent	3:16	8:11	0.1510

EGFR: Epidermal growth factor receptor; PD-L1: programmed cell death-ligand; irAE: immune-related adverse events.

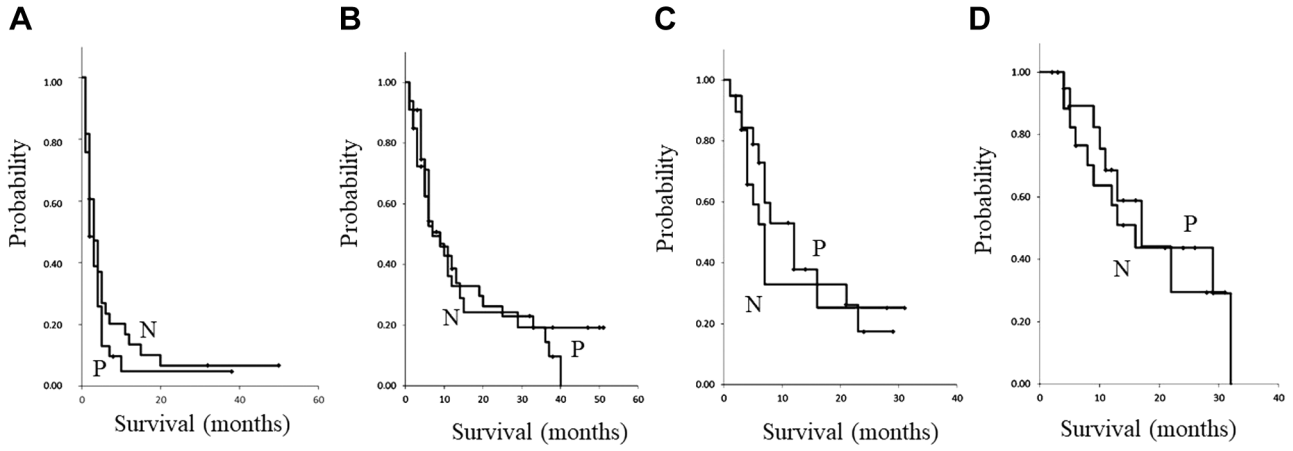


Figure 2. *Progression-free survival (PFS) and overall survival (OS) for each treatment regimen in the study. PFS in patients treated with atezolizumab monotherapy (A), PFS in patients treated with combined atezolizumab plus chemotherapy (B), OS in patients treated with atezolizumab monotherapy (C), and OS in patients treated with combined atezolizumab plus chemotherapy (D). P: EGFR positive; N: EGFR negative.*

were younger and had a higher proportion of adenocarcinoma compared with EGFR-negative patients. To examine the PFS and OS among EGFR-mutated and EGFR-negative patients, 1:1 propensity matching was performed. Table II presents the characteristics of propensity-matched patients treated with atezolizumab monotherapy and those treated with combined atezolizumab plus chemotherapy. After confirming that there were no differences for these characteristics, PFS and OS were compared between EGFR-mutated and EGFR-negative patients. Figure 2 shows PFS and OS curves for patients with and without EGFR mutations. In patients treated with atezolizumab monotherapy, there was no significant difference in PFS and OS between EGFR-mutated and EGFR-negative patients (p=0.3095 and p=0.7712, respectively). In patients treated with combined atezolizumab plus chemotherapy, no significant difference was observed in PFS and OS between EGFR-mutated and EGFR-negative patients (p=0.4671 and p=0.7856, respectively).

Table III. *Uni- and multivariate analysis of survival from the initiation of atezolizumab therapy.*

	Progression-free survival				Overall survival			
	Univariate	Multivariate			Univariate	Multivariate		
	p-Value	Odds ratio	95% CI	*p*-Value	*p*-value	Odds ratio	95% CI	*p*-Value
Atezolizumab monotherapy								
Sex, male	0.724				0.199	2.60	0.809-8.375	0.109
Performance status, 0-1	0.165				0.034	4.32	1.270-14.665	0.019
Age, less than 70 years	0.876				0.273			
PD-L1, 50% or more	0.722				0.658			
Stage, IIIA-C	0.590				0.846			
EGFR, negative	0.746				0.265			
irAEs, grade 1-4	0.341				0.818			
Combined atezolizumab and chemotherapy								
Sex, male	0.718				0.860			
Performance status, 0-1	0.289				0.346			
Age, less than 70 years	0.242				0.508			
PD-L1, 50% or more	0.827				0.697			
Stage, IIIA-C	0.505				0.725			
EGFR, negative	0.978				0.240			
irAEs, grade 1-4	0.726				0.416			

PD-L1: Programmed cell death-ligand 1; EGFR: epidermal growth factor receptor; irAE: immune-related adverse events.

To identify favorable factors affecting PFS and OS, univariate and multivariate analyses were performed using sex, PS, age, PD-L1, stage, types of EGFR mutation, and irAEs as variables. In patients treated with atezolizumab monotherapy, male sex was a favorable factor for PFS and OS (Table III). In patients treated with combined atezolizumab plus chemotherapy, male sex, and PS (0-1) were favorable factor for PFS, and male sex, age less than 70 years, and presence of irAEs (grade 1-4) were favorable factors for OS (Table III). Univariate and multivariate analyzes were performed to investigate prognostic factors using patient data after propensity matching. The results are shown in Table III. In patients treated with atezolizumab monotherapy, the presence or absence of EGFR mutation was not a significant prognostic factor for either PFS or OS. PS in OS was the only significant factor in those treated with atezolizumab monotherapy. In patients treated with combined atezolizumab and chemotherapy, the presence or absence of EGFR mutation was not a significant prognostic factor for either PFS or OS. There were no significant factors in those treated with the combination therapy.

EGFR-TKI before atezolizumab and duration of response to atezolizumab. The relationship between type of TKI and duration of response to atezolizumab in 46 patients, including 29 who received TKIs before atezolizumab monotherapy and 17 who received TKIs before atezolizumab plus chemotherapy, was examined. In the 29 patients who received TKIs prior to atezolizumab monotherapy (seven with osimertinib and 22 with other TKIs), there were no significant differences in PFS or OS between these two groups by type of TKI (p=0.3047 and p=0.6636, respectively). No significant difference was seen in PFS or OS between these two groups by type of TKI in the 17 patients who received TKIs prior to combined atezolizumab plus chemotherapy (10 with osimertinib and seven with other TKIs; p=0.5705 and p=0.2597, respectively). One patient received atezolizumab monotherapy immediately after osimertinib, and the PFS and OS for that patient was 2 months and 11 months, respectively. This patient developed grade 3 liver injury. Nine patients received combined atezolizumab plus chemotherapy immediately after osimertinib, and these patients had a median PFS of 12 months. Although OS was not reached, five of the nine patients survived for more than 12 months. Three of the 9 patients who received atezolizumab immediately after osimertinib had 5 cases of irAEs, including one case of colitis (grade 3), pulmonary toxicity (grade 2), thyroid dysfunction (grade 2), diarrhea (grade 2), and skin toxicity (grade 2).

Of the 29 patients who received TKIs before atezolizumab monotherapy, 10 had TKI PFS <1 year and 19 had TKI PFS >1 year. There was no significant difference in PFS or OS for atezolizumab monotherapy in these patients (p=0.2794 and p=0.9820, respectively). Of the 17 patients who received TKIs before combined atezolizumab plus chemotherapy, five

had TKI PFS <1 year and 12 had TKI PFS >1 year. No significant difference in PFS or OS for combined atezolizumab plus chemotherapy in these patients was found (p=0.4375 and p=0.4717 respectively).

Treatment after administration of atezolizumab. EGFR-TKI rechallenge was performed in seven of 33 EGFR-mutated patients treated with atezolizumab monotherapy. The median PFS upon TKI re-administration in these patients was 9 months (95%CI=2-19 months), with three of them having a PFS of 12 months or longer. Four of these 7 patients received EGFR-TKI immediately after atezolizumab monotherapy, and the median PFS on TKI readmission was 11 months. TS-1, an antimetabolite, was prescribed in four of 33 patients treated with atezolizumab monotherapy. The median PFS upon TKI re-administration in those patients was 11 months (95%CI=1-22 months), with two of them having a PFS of 12 months or longer. Docetaxel+ramucirumab was administered in eight patients after atezolizumab administration. The median PFS for these patients was 1 month (95%CI=1-6 months).

Toxicity. irAEs were observed in seven patients (13.5%), with four of these patients (7.7%) having grade ≥3. One patient had pulmonary toxicity (grade 2). Hepatobiliary toxicity was observed in two patients (2 grade 3), thyroid dysfunction in two patients (two grade 2), arthralgia in one patient (grade 3), colitis in one patient (grade 3), diarrhea in one patient (grade 2), and skin toxicity in one patient (grade 2). Grade 5 irAEs were not observed.

Discussion

The main results of our investigation were as follows. Median PFS and OS in 33 EGFR-mutated patients treated with atezolizumab monotherapy were 2 and 9 months, respectively, and those in 19 patients who received combined atezolizumab plus chemotherapy were 12 and 17 months, respectively. When comparing EGFR-mutated and EGFR-negative patients after propensity score matching, there were no significant differences in PFS and OS between the two groups, whether treated with atezolizumab monotherapy or combined atezolizumab plus chemotherapy. Among EGFR-mutated patients, being male was a significant favorable factor in both these two atezolizumab treatment groups. In uni- and multivariate analyses in propensity matched patients treated with atezolizumab monotherapy and those treated with combined atezolizumab and chemotherapy, we confirmed that the presence or absence of EGFR mutations was not a significant prognostic factor for either PFS or OS. The type of TKI given prior to atezolizumab and the duration of response of TKI did not influence the PFS on atezolizumab treatment. For atezolizumab treatment after TKI treatment, there were patients who responded to TKI retreatment and TS-1 treatment for a relatively long period of time. None of the 52 EGFR-mutated patients had grade 5 irAEs.

A recent review of anti-PD-1 monotherapy for EGFR-mutated patients reported a median PFS and OS of around 3.9 months and 10.7 months, respectively (16). As for atezolizumab monotherapy, median PFS and OS of EGFR-mutated patients were 2.1-3.2 months and 10.2-13.0 months, respectively (8, 17). Among ICIs, whether anti-PD-1 or anti-PD-L1, monotherapy appeared to be commonly used for EGFR-mutated patients with poor PS or in whom multiple therapies had failed. We surmised that the short PFS and OS, including in our patients, were attributable to these backgrounds (8, 17). On the other hand, for combined ICI plus chemotherapy, trials including EGFR-mutated patients were conducted with anti-PD-L1 antibodies only (18-20). Clinical trials of combined atezolizumab plus chemotherapy included 4.2%-7.2% of EGFR-mutated patients (18-20). All patients enrolled in these trials had a PS of 0-1 and a median age of 63-64 years. The median PFS and median OS were 7.0-10.2 months and 14.4-26.1 months, respectively (18-20). In real clinical practice, the median PFS and median OS of the combined therapies in EGFR-mutated patients were 5.2-13.6 months and 10.9-22.9 months, respectively (21-25). In these studies, patients with PS2 accounted for 5%-25% of patients, with a median age of 56-64 years (21-25). In our study, 10.5% of patients had PS2, with a median age of 72 years, and median PFS and OS of 12.0 and 17.0 months, respectively. PFS and OS in real clinical practice studies, including ours, were similar to those in clinical trials, even though the proportion of elderly patients and those with poor PS were higher than those in the clinical trials (18-25).

In general, it seems to be accepted that EGFR-mutated patients are less responsive to ICIs and have a shorter duration of response than EGFR-negative patients (6, 7). EGFR-mutated patients are more likely to be female, young, and have adenocarcinoma compared to EGFR-negative patients (26). To the best of our knowledge, for ICI monotherapy or combined ICI plus chemotherapy, there have been no previous reports on the comparison of PFS and OS between propensity-matched EGFR-mutated and EGFR-negative NSCLC patients. Although with a small number of patients, this was the first report comparing PFS and OS between propensity-matched EGFR-mutated and EGFR-negative NSCLC patients. It has been widely accepted that males with NSCLC responded better in pivotal ICI trials (27-29). Recently, Choi *et al*. reported that sex was not an independent prognostic factor for immunotherapy in real-world data, although various factors affected immunotherapy response, such as wild-type EGFR and high expression of PD-L1, which frequently occur in males (30). Taking these into consideration, it is important

to compare survival with matching patient backgrounds. Studies with large numbers of patients to confirm the findings of our study are necessary.

Recently, the sequence of treatment with ICIs in EGFR-mutated patients has attracted attention, and several studies have been reported (8-11). Wu *et al*. reported that patients who received osimertinib prior to an ICI had a shorter PFS and OS than those who received a TKI other than osimertinib (8). In these cases, patients received nivolumab and pembrolizumab (8). Neither PFS nor OS was different in our atezolizumab-treated patients. Yamaguchi *et al*. reported that osimertinib immediately after nivolumab significantly increased the frequency of grade 3 or higher hepatotoxicity (12). Uchida *et al*. investigated drug-induced lung injury in patients treated with a TKI immediately after an ICI, and with an ICI immediately after TKI treatment (11). Based on the results, they called attention to the onset of pulmonary injury with TKI treatment immediately after an ICI, and with ICI treatment immediately after a TKI (11). In the present study, hepatobiliary toxicity was observed in two patients (both grade 3), and a pulmonary irAE was observed in only one patient (grade 2) who received a TKI immediately followed by combination atezolizumab plus chemotherapy. It is unclear whether these results were due to differences in the administered ICIs, or because the number of patients studied was small, but these results are noteworthy. Si *et al*. showed prolongation of PFS and OS of ICI treatment with angiogenesis inhibitors (9). In the present study, bevacizumab was administered in 17 of 19 patients who received concurrent chemotherapy. Although comparisons with and without bevacizumab were not possible, this result should be noted. Another noteworthy treatment sequence is TKI-rechallenge after ICI. Kaira *et al*. reported that EGFR-TKI rechallenge immediately after ICI therapy was identified as an effective therapy for NSCLC patients with resistance to EGFR-TKIs (10). Our patients had a median PFS of 11 months. Also, noteworthy was the relatively long period of PFS on TS-1 treatment in our patients.

As for irAEs, additional caution is required when ICIs are used either as monotherapy or in combination with chemotherapy. In our study, irAEs of any grade were observed in seven of 52 patients (13.5%). Although the incidence of irAEs appeared to be low, the possibility of under-evaluation for retrospective studies could not be ruled out. There were other limitations in this study. Although the survey included a sufficient number of patients for statistical processing, it was a retrospective study of patients with a wide range of background characteristics. TTF-1, LKB1 and KEAP1 could affect treatment efficacy (31), but we were unable to investigate these in this study. Therefore, it should be noted that our results are not definitive and do not allow final conclusions.

Conclusion

Even in EGFR-mutated patients, ICIs are considered important therapeutic agents and must be included in the treatment sequence. Considering the possibility of supplementing clinical trial results or providing a raised awareness, it is important to carefully examine valuable data from real clinical practice. To obtain longer survival in EGFR-mutated NSCLC patients, it will be important to consider the appropriate selection and treatment sequence of EGFR-TKIs, as well as ICIs, conventional anti-tumor drugs, and anti-angiogenic drugs.

Conflicts of Interest

The Authors have no conflicts of interest to declare in relation to this study.

Authors' Contributions

TT, KM, TS, SO, HS: Study design. TN, HY, TS, YW, SO, SO, NK, KM, SH, TY, KK, MI, HS, HI, TK, TE, TS: Data collection. TT, SH, SO, RN, HS: Data analysis. TT, SH, SO, HS, NH: Manuscript preparation. AN, HS, NH: Supervision. All Authors have seen and approved the final version before submission.

References

1 Reck M, Remon J, Hellmann MD: First-line immunotherapy for non–small-cell lung cancer. J Clin Oncol 40(6): 586-597, 2022. DOI: 10.1200/JCO.21.01497

2 Sham NO, Zhao L, Zhu Z, Roy TM, Xiao H, Bai Q, Wakefield MR, Fang Y: Immunotherapy for non-small cell lung cancer: Current agents and potential molecular targets. Anticancer Res 42(7): 3275-3284, 2022. DOI: 10.21873/anticanres.15816

3 Inomata M, Matsumoto M, Hayashi K, Seto Z, Hirai T, Tokui K, Taka C, Okazawa S, Kambara K, Imanishi S, Miwa T, Hayashi R, Matsui S, Tobe K: Association between the severity of immune-related adverse events and the prognosis in patients with non-small cell lung cancer receiving treatment with immune checkpoint inhibitors. Anticancer Res 43(7): 3241-3246, 2023. DOI: 10.21873/anticanres.16498

4 Nakahama K, Izumi M, Yoshimoto N, Fukui M, Sugimoto A, Nagamine H, Ogawa K, Sawa K, Tani Y, Kaneda H, Mitsuoka S, Watanabe T, Asai K, Kawaguchi T: Influence of smoking history on the effectiveness of immune-checkpoint inhibitor therapy for non-small cell lung cancer: Analysis of real-world data. Anticancer Res 43(5): 2185-2197, 2023. DOI: 10.21873/anticanres.16381

5 Christopoulos P: Can we cure metastatic EGFR-mutated lung cancer today? Transl Cancer Res 12(2): 209-212, 2023. DOI: 10.21037/tcr-22-2926

6 Liao D, Yu L, Shangguan D, Zhang Y, Xiao B, Liu N, Yang N: Recent advancements of monotherapy, combination, and sequential treatment of EGFR/ALK-TKIs and ICIs in non-small cell lung cancer. Front Pharmacol 13: 905947, 2022. DOI: 10.3389/fphar.2022.905947

7 Qian X, Guo X, Li T, Hu W, Zhang L, Wu C, Ye F: Efficacy of immune checkpoint inhibitors in EGFR-Mutant NSCLC patients with EGFR-TKI resistance: A systematic review and meta-analysis. Front Pharmacol 13: 926890, 2022. DOI: 10.3389/fphar.2022.926890

8 Wu SG, Chiang CL, Wang CC, Hung JY, Hsia TC, Kuo CH, Shih JY: Clinical outcomes of atezolizumab therapy for previously-treated advanced-stage non-small cell lung cancer: a real-world study in Taiwan. J Cancer 13(9): 2922-2932, 2022. DOI: 10.7150/jca.74617

9 Si J, Hao Y, Wei J, Xiang J, Xu C, Shen Q, Song Z: Clinical outcomes of immune checkpoint inhibitors to treat non-small cell lung cancer patients harboring epidermal growth factor receptor mutations. BMC Pulm Med 23(1): 158, 2023. DOI: 10.1186/s12890-023-02466-9

10 Kaira K, Kobayashi K, Shiono A, Yamaguchi O, Hashimoto K, Mouri A, Shinomiya S, Miura Y, Imai H, Kagamu H: Effectiveness of EGFR-TKI rechallenge immediately after PD-1 blockade failure. Thorac Cancer 12(6): 864-873, 2021. DOI: 10.1111/1759-7714.13864

11 Uchida T, Kaira K, Yamaguchi O, Mouri A, Shiono A, Miura Y, Hashimoto K, Nishihara F, Murayama Y, Kobayashi K, Kagamu H: Different incidence of interstitial lung disease according to different kinds of EGFR-tyrosine kinase inhibitors administered immediately before and/or after anti-PD-1 antibodies in lung cancer. Thorac Cancer 10(4): 975-979, 2019. DOI: 10.1111/1759-7714.13039

12 Yamaguchi O, Kaira K, Kawasaki T, Mouri A, Hashimoto K, Shiono A, Shinomiya S, Miura Y, Nishihara F, Murayama Y, Kobayashi K, Mochida S, Kagamu H: Severe hepatotoxicity due to osimertinib after nivolumab therapy in patients with non-small cell lung cancer harboring EGFR mutation. Thorac Cancer 11(4): 1045-1051, 2020. DOI: 10.1111/1759-7714.13363

13 Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, Nicholson AG, Groome P, Mitchell A, Bolejack V, International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee, Advisory Boards, and Participating Institutions, International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee Advisory Boards and Participating Institutions: The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. J Thorac Oncol 11(1): 39-51, 2016. DOI: 10.1016/j.jtho.2015.09.009

14 Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 45(2): 228-247, 2009. DOI: 10.1016/j.ejca.2008.10.026

15 National Cancer Institute Common Terminology Criteria for Adverse Events (version 5.0). Available at: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fctep.cancer.gov%2FprotocolDevelopment%2Felectronic_applications%2Fdocs%2FCTCAE_v5.0.xlsx&wdOrigin=BROWSELINK [Last accessed on February 16, 2023]

16 Luciani A, Ghidini A, Borgonovo K, Parati MC, Petrelli F: Outcome of non-small-cell lung cancer with driven mutations treated with anti-PD-(L)1 agents: A systematic review. Tumori: 030089162211226, 2022. DOI: 10.1177/03008916221122601

17 Miura S, Nishio M, Akamatsu H, Goto Y, Hayashi H, Gemma A, Yoshino I, Misumi T, Hata A, Hataji O, Fujita K, Seike M, Yanagitani N, Nishino K, Hara S, Saito R, Mori M, Tsuda T, Iwasawa S, Nakagawa S, Mitsudomi T: Effectiveness and safety of atezolizumab monotherapy in previously treated Japanese patients with unresectable advanced or recurrent NSCLC: a multicenter, prospective, observational study (J-TAIL). JTO Clin Res Rep 4(3): 100484, 2023. DOI: 10.1016/j.jtocrr.2023.100484

18 Reck M, Wehler T, Orlandi F, Nogami N, Barone C, Moro-Sibilot D, Shtivelband M, González Larriba JL, Rothenstein J, Früh M, Yu W, Deng Y, Coleman S, Shankar G, Patel H, Kelsch C, Lee A, Piault E, Socinski MA: Safety and patient-reported outcomes of atezolizumab plus chemotherapy with or without bevacizumab *versus* bevacizumab plus chemotherapy in non-small-cell lung cancer. J Clin Oncol 38(22): 2530-2542, 2020. DOI: 10.1200/JCO.19.03158

19 West H, McCleod M, Hussein M, Morabito A, Rittmeyer A, Conter HJ, Kopp H, Daniel D, Mccune S, Mekhail T, Zer A, Reinmuth N, Sadiq A, Sandler A, Lin W, Ochi Lohmann T, Archer V, Wang L, Kowanetz M, Cappuzzo F: Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol 20(7): 924-937, 2019. DOI: 10.1016/S1470-2045(19)30167-6

20 Nishio M, Barlesi F, West H, Ball S, Bordoni R, Cobo M, Longeras PD, Goldschmidt J Jr, Novello S, Orlandi F, Sanborn RE, Szalai Z, Ursol G, Mendus D, Wang L, Wen X, McCleland M, Hoang T, Phan S, Socinski MA: Atezolizumab plus chemotherapy for first-line treatment of nonsquamous NSCLC: results from the randomized Phase 3 IMpower132 trial. J Thorac Oncol 16(4): 653-664, 2021. DOI: 10.1016/j.jtho.2020.11.025

21 Bylicki O, Tomasini P, Radj G, Guisier F, Monnet I, Ricordel C, Bigay-Game L, Geier M, Chouaid C, Daniel C, Swalduz A, Toffart AC, Doubre H, Peloni JM, Moreau D, Subtil F, Grellard JM, Castera M, Clarisse B, Martins-Lavinas PH, Decroisette C, Greillier L: Atezolizumab with or without bevacizumab and platinum-pemetrexed in patients with stage IIIB/IV non-squamous non-small cell lung cancer with EGFR mutation, ALK rearrangement or ROS1 fusion progressing after targeted therapies: A multicentre phase II open-label non-randomised study GFPC 06-2018. Eur J Cancer 183: 38-48, 2023. DOI: 10.1016/j.ejca.2023.01.014

22 Trummer A, Bethge A, Dickgreber N, Dittrich I, Golpon H, Hoffknecht P, Overbeck TR, Wesseler C, Reck M: NSCLC with uncommon EGFR mutations treated with atezolizumab plus bevacizumab and chemotherapy. Lung Cancer 174: 141-145, 2022. DOI: 10.1016/j.lungcan.2022.11.006

23 White MN, Piper-Vallillo AJ, Gardner RM, Cunanan K, Neal JW, Das M, Padda SK, Ramchandran K, Chen TT, Sequist LV, Piotrowska Z, Wakelee HA: Chemotherapy plus immunotherapy *versus* chemotherapy plus bevacizumab *versus* chemotherapy alone in EGFR-mutant NSCLC after progression on osimertinib. Clin Lung Cancer 23(3): e210-e221, 2022. DOI: 10.1016/j.cllc.2021.11.001

24 Hata T, Sakaguchi C, Hirano K, Kobe H, Ishida M, Nakano T, Tachibana Y, Tamiya N, Shiotsu S, Takeda T, Yamada T, Yokoyama T, Tsuchiya M, Nagasaka Y: Exploratory analysis of immunochemotherapy compared to chemotherapy after EGFR-TKI in non-small cell lung cancer patients with EGFR mutation:

A multicenter retrospective study. Thorac Cancer 14(11): 1004-1011, 2023. DOI: 10.1111/1759-7714.14836

25 Ikeuchi N, Igata F, Kinoshita E, Kawabata T, Tan I, Osaki Y, Otsuka R, On R, Ikeda T, Nakao A, Sasaki T, Aoyama T, Hirano R, Harada T, Ebi N, Fujita M, Inoue H: Real-world efficacy and safety of atezolizumab plus bevacizumab, paclitaxel and carboplatin for first-line treatment of Japanese patients with metastatic non-squamous non-small cell lung cancer. Anticancer Res 43(2): 713-724, 2023. DOI: 10.21873/anticanres.16210

26 Lim C, Tsao MS, Le LW, Shepherd FA, Feld R, Burkes RL, Liu G, Kamel-Reid S, Hwang D, Tanguay J, da Cunha Santos G, Leighl NB: Biomarker testing and time to treatment decision in patients with advanced nonsmall-cell lung cancer. Ann Oncol 26(7): 1415-1421, 2015. DOI: 10.1093/annonc/mdv208

27 Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, Leiby MA, Lubiniecki GM, Shentu Y, Rangwala R, Brahmer JR, KEYNOTE-024 Investigators: Pembrolizumab *versus* chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med 375(19): 1823-1833, 2016. DOI: 10.1056/NEJMoa1606774

28 Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, Barlesi F, Kohlhäufl M, Arrieta O, Burgio MA, Fayette J, Lena H, Poddubskaya E, Gerber DE, Gettinger SN, Rudin CM, Rizvi N, Crinò L, Blumenschein GR Jr, Antonia SJ, Dorange C, Harbison CT, Graf Finckenstein F, Brahmer JR: Nivolumab *versus* docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med 373(17): 1627-1639, 2015. DOI: 10.1056/NEJMoa1507643

29 Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, Antonia S, Pluzanski A, Vokes EE, Holgado E, Waterhouse D, Ready N, Gainor J, Arén Frontera O, Havel L, Steins M, Garassino MC, Aerts JG, Domine M, Paz-Ares L, Reck M, Baudelet C, Harbison CT, Lestini B, Spigel DR: Nivolumab *versus* docetaxel in advanced squamous-cell non-small-cell lung cancer. N Engl J Med 373(2): 123-135, 2015. DOI: 10.1056/NEJMoa1504627

30 Choi MG, Choi CM, Lee DH, Kim SW, Yoon S, Ji W, Lee JC: Impact of gender on response to immune checkpoint inhibitors in patients with non-small cell lung cancer undergoing second- or later-line treatment. Transl Lung Cancer Res 11(9): 1866-1876, 2022. DOI: 10.21037/tlcr-22-146

31 Di Federico A, De Giglio A, Parisi C, Gelsomino F: STK11/LKB1 and KEAP1 mutations in non-small cell lung cancer: Prognostic rather than predictive? Eur J Cancer 157: 108-113, 2021. DOI: 10.1016/j.ejca.2021.08.011

Received July 23, 2023
Revised August 20, 2023
Accepted August 25, 2023

JJCO Japanese Journal of Clinical Oncology

Japanese Journal of Clinical Oncology, 2024, 54(2), 153–159
https://doi.org/10.1093/jjco/hyad156
Advance Access Publication Date: 20 November 2023
Original Article

OXFORD

Original Article

Development and evaluation of a rapid one-step high sensitivity real-time quantitative PCR system for minor *BCR-ABL* (e1a2) test in Philadelphia-positive acute lymphoblastic leukemia (Ph+ ALL)

Michihiro Hidaka[1,*], Koiti Inokuchi[2], Nobuhiko Uoshima[3], Naoto Takahashi[4], Nao Yoshida[5], Shuichi Ota[6], Hirohisa Nakamae[7], Hiromi Iwasaki[8], Kenichiro Watanabe[9], Yoshiyuki Kosaka[10], Norio Komatsu[11], Kuniaki Meguro[12], Yuho Najima[13], Tetsuya Eto[14], Takeshi Kondo[15], Shinya Kimura[16], Chikashi Yoshida[17], Yuichi Ishikawa[18], Masashi Sawa[19], Tomoko Hata[20], Keizo Horibe[21], Hiroatsu Iida[22], Takeshi Shimomura[23], Nobuaki Dobashi[24], Isamu Sugiura[25], Junya Makiyama[26], Naoyuki Miyagawa[27], Asuka Sato[28], Ryuta Ito[28], Itaru Matsumura[29], Yuzuru Kanakura[30] and Tomoki Naoe[31]

[1]Department of Hematology, National Hospital Organization Kumamoto Medical Center, Kumamoto, Japan, [2]Department of Hematology, Nippon Medical School, Tokyo, Japan, [3]Department of Hematology, Japanese Red Cross Kyoto Daini Hospital, Kyoto, Japan, [4]Department of Hematology, Nephrology and Rheumatology, Akita University Graduate School of Medicine, Akita, Japan, [5]Department of Hematology and Oncology, Children's Medical Center, Japanese Red Cross Aichi Medical Center Nagoya First Hospital, Aichi, Japan, [6]Department of Hematology, Sapporo Hokuyu Hospital, Hokkaido, Japan, [7]Department of Hematology, Graduate School of Medicine, Osaka City University, Osaka, Japan, [8]Department of Hematology, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan, [9]Department of Hematology and Oncology, Shizuoka Children's Hospital, Shizuoka, Japan, [10]Department of Hematology and Oncology, Center of Childhood Cancer, Hyogo Prefectural Kobe Children's Hospital, Hyogo, Japan, [11]Department of Hematology, Juntendo University School of Medicine, Tokyo, Japan, [12]Department of Hematology, National Hospital Organization Sendai Medical Center, Miyagi, Japan, [13]Hematology Division, Tokyo Metropolitan Cancer and Infectious Diseases Center, Komagome Hospital, Tokyo, Japan, [14]Department of Hematology, Hamanomachi Hospital, Fukuoka, Japan, [15]Blood Disorders Center, Aiiku Hospital, Hokkaido, Japan, [16]Department of Hematology, Respiratory Medicine and Oncology, Saga University, Saga, Japan, [17]Department of Hematology, National Hospital Organization Mito Medical Center, Ibaraki, Japan, [18]Department of Hematology and Oncology, Nagoya University Graduate School of Medicine, Aichi, Japan, [19]Department of Hematology and Oncology, Anjo Kosei Hospital, Aichi, Japan, [20]Department of Hematology, Nagasaki University Hospital, Nagasaki, Japan, [21]Department of Pediatrics, National Hospital Organization Nagoya Medical Center, Aichi, Japan, [22]Department of Hematology, National Hospital Organization Nagoya Medical Center, Aichi, Japan, [23]Department of Hematology, National Hospital Organization Hiroshimanishi Medical Center, Hiroshima, Japan, [24]Department of Clinical Oncology and Hematology, The Jikei University Daisan Hospital, Tokyo, Japan, [25]Department of Hematology and Oncology, Toyohashi Municipal Hospital, Aichi, Japan, [26]Department of Hematology/Oncology, Institute of Medical Science, The University of Tokyo Hospital, Tokyo, Japan, [27]Division of Hematology/Oncology, Kanagawa Children's Medical Center, Kanagawa, Japan, [28]Diagnostic Division, Otsuka Pharmaceutical Co. Ltd., Tokushima, Japan, [29]Department of Hematology and Rheumatology, Kindai University

Faculty of Medicine, Osaka, Japan, [30]Sumitomo Hospital, Osaka, Japan and [31]National Hospital Organization Nagoya Medical Center, Aichi, Japan

*For reprints and all correspondence: Michihiro Hidaka, National Hospital Organization Kumamoto Medical Center, 1-5 Ninomaru, Chuo-ku, Kumamoto 860-0008 Japan. E-mail: hidaka.michihiro.rp@mail.hosp.go.jp

Received 19 June 2023; Editorial Decision 23 October 2023; Accepted 25 October 2023

Abstract

Objective: Minimal residual disease assessment of *BCR-ABL* messenger ribonucleic acid levels is crucial in Philadelphia chromosome-positive acute lymphoblastic leukemia for prognosis and treatment planning. However, accurately quantifying minor *BCR-ABL* transcripts, which comprise 70% of Philadelphia chromosome-positive acute lymphoblastic leukemia cases, lacks a national-approved method.

Methods: We developed the "Otsuka" minor BCR-ABLmessenger ribonucleic acid assay kit with exceptional precision (0.00151%). Minor BCR-ABL messenger ribonucleic acid levels were analyzed in 175 adults, 36 children with acute lymphoblastic leukemia and 25 healthy individuals to evaluate the kit's performance.

Results: The "Otsuka" kit showed high concordance with a commonly used chimeric gene screening method, indicating reliable detection of positive cases. Quantitative results demonstrated a robust correlation with both a laboratory-developed test and a diagnostic research product. The "Otsuka" kit performs comparably or even surpass to conventional products, providing valuable insights into Philadelphia chromosome-positive acute lymphoblastic leukemia pathology.

Conclusions: The 'Otsuka" minor BCR-ABL messenger ribonucleic acid assay kit exhibits excellent performance in quantifying minor *BCR-ABL* transcripts in Philadelphia chromosome-positive acute lymphoblastic leukemia patients. Our results align well with established screening methods and show a strong correlation with laboratory-developed tests and diagnostic research products. The "Otsuka" kit holds great promise as a valuable tool for understanding Philadelphia chromosome-positive acute lymphoblastic leukemia pathology and guiding effective treatment strategies.

Key words: minor *BCR-ABL*, Philadelphia-positive acute lymphoblastic leukemia, minimal residual disease

Introduction

Acute lymphoblastic leukemia (ALL) is a hematopoietic tumor characterized by neoplastic changes in immature lymphoid cells and bone marrow (BM) infiltration. The reported annual incidence is ~1 in 100 000 adults and ~3 in 100 000 children (1). The Philadelphia chromosome is present in 30–40% of adult patients (2) and 3–5% of pediatric patients (3), making its confirmation crucial for diagnosis and treatment. Philadelphia chromosome-positive (Ph+) ALL is classified into two types of *Breakpoint cluster region-abelson* (*BCR-ABL*) fusion genes: minor *BCR-ABL* and major *BCR-ABL*. Minor *BCR-ABL* accounts for ~70% of adult patients (4) and ~90% of pediatric patients (5).

The standard treatment for Ph+ ALL patients now involves chemotherapy with tyrosine kinase inhibitors (TKIs), followed by allogeneic hematopoietic stem cell transplantation. The combination of chemotherapy with TKIs has shown significant improvement in survival rates and high rates of complete remission with negative minimal residual disease (MRD) (6–9). Negative MRD is the most significant prognostic factor in Ph+ ALL (10–13). Detecting MRD after 3 months of remission induction therapy indicates a high likelihood of relapse and reduced survival (11). Furthermore, it has been observed that patients with negative MRD before transplantation have lower post-transplant relapse rates (12). Negative MRD is now widely used as a marker for treatment strategies. However, standardized evaluation methods for MRD in minor *BCR-ABL*, which accounts for 70% of adults and 90% of children with Ph+ ALL, have not been established.

Until recently, Japan lacked approved diagnostic agents for determining minor *BCR-ABL* positivity and MRD levels, resulting in the absence of a standardized measurement method. Measurement of minor *BCR-ABL* levels relied on various diagnostic reagents, including laboratory-developed tests from registered laboratories. Typically, a sample would be considered MRD negative if quantitative polymerase chain reaction (PCR) did not detect BCR-ABL.

We have recently developed an *in vitro* diagnostic system called the "Otsuka" minor BCR-ABL messenger ribonucleic acid (mRNA) assay kit. This kit is designed to specifically target the e1a2 breakpoint that is predominantly found in Ph+ ALL. In addition, this kit utilizes *ABL* mRNA instead of *GAPDH* mRNA as the reference gene. This adaptation facilitates comparison of MRD data with results obtained in other countries. The assay procedure is remarkably simple, requiring only 2 hours and a small amount of BM and peripheral blood (PB). The reverse transcription quantitative PCR (RT-qPCR) reaction is performed in a single tube, facilitating comparison and evaluation of measurements across different laboratories.

In this study, we validated the kit using individual patients' samples and compared it with currently used diagnostics in research. The data obtained clearly demonstrated the assay kit's efficacy as a powerful tool for diagnosing and monitoring the pathological conditions of Ph+ ALL patients.

Patients and methods

Ethical conduct of the study

This study (Protocol No. ODK-1601-CLN-001) adhered to the ethical principles outlined in the Declaration of Helsinki, Ethical Guidelines for Medical and Health Research Involving Human Subjects, Japan's Act on Securing Quality, Efficacy and Safety of Products Including Pharmaceuticals and Medical Devices, and other relevant ministerial ordinances, notifications and the study protocol. No instances of ethical misconduct were reported throughout the study.

The National Hospital Organization (NHO) Central Review Board approved the implementation of this study for the 12 NHO-operated institutions. In addition, each hospital's institutional review board reviewed and approved the study based on the protocol, informed consent form and case report forms for the remaining 41 institutions.

Before participating in the study, the principal investigator or subinvestigator provided a detailed explanation to each patient using the informed consent form, allowing sufficient time for the patient to make an informed decision. Written voluntary informed consent was obtained from each patient, ensuring their comprehensive understanding of the information. For pediatric patients, written informed consent was obtained from either the patients themselves or their legally authorized representatives.

Patients and samples

This multicenter, open-label study was conducted from June 2017 to March 2019, involving 53 participating institutions. A total of 236 participants were enrolled, including 175 adults and 36 pediatric patients with suspected ALL who met the inclusion criteria and none of the exclusion criteria. In addition, 25 healthy individuals were included. Due to the limited number of children with Ph+ ALL, pediatric patients undergoing remission induction therapy or consolidation therapy were eligible for enrollment. Furthermore, existing samples stored at the institutions were utilized.

Among the adult patients with suspected ALL (n = 175), there were 33 patients with Ph+ ALL harboring minor *BCR-ABL*, 8 patients with Ph+ ALL harboring major *BCR-ABL*, 46 patients with Ph- ALL and 88 patients with other diseases. The pediatric patients with suspected ALL (n = 36) comprised 9 patients with Ph+ ALL harboring minor *BCR-ABL*, 22 patients with Ph- ALL and 5 patients with other diseases.

All adult patients with suspected ALL were enrolled before initiating treatment, and samples of BM and PB were collected. Patients diagnosed with Ph- ALL or non-ALL malignancies based on the test results were withdrawn from the study. Patients diagnosed with Ph+ ALL remained in the study, and BM and/or PB samples were collected at nine time points: Days 0, 8, 15, 22 and 29 after remission induction therapy, as well as weeks 0, 2, 4 and at the end of consolidation therapy.

Study design

RNA was extracted from BM and PB samples for analysis. The levels of minor *BCR-ABL* mRNA were measured using the minor BCR-ABL mRNA Assay Kit "Otsuka" (referred to as "Otsuka"), an *in vitro* diagnostic product by Otsuka Pharmaceutical Co., Japan. To evaluate the performance of Otsuka, "a screening test for leukemia-related chimeric genes", one of the tests used to diagnose Ph+ ALL, was used as a control for the pre-treatment samples (Sample No. 1). Using the screening test as a control, the positive concordance rate, negative concordance rate and overall concordance rate for Otsuka were calculated.

We then evaluated the correlation between Otsuka and two control reagents to validate the performance of Otsuka in monitoring treatment efficacy (Sample Nos. 1–9). The two control reagents were a laboratory-developed test from a registered laboratory (minor *BCR/ABL* assay (14), hereafter referred to as control reagent A) and the Ipsogen *BCR-ABL1* mbcr kit (15), hereafter referred to as control reagent B). In addition, minor *BCR-ABL* levels in patients with minor *BCR-ABL* and major *BCR-ABL* were compared, as well as other hematologic disorders. Results from BM samples were compared with PB samples, and the changes in minor *BCR-ABL* mRNA levels over the clinical course were graphically plotted. Statistical analyses were performed using SAS system version 9.4 (SAS Institute Japan).

Diagnostic kits

Otsuka is designed to detect minor *BCR-ABL* mRNA fusion transcript e1a2. The kit can perform reverse transcription reaction and quantitative PCR of minor *BCR-ABL* mRNA and *ABL* mRNA in a single reaction solution simultaneously and continuously. The components and detailed procedure of this kit are attached as Supplement. Briefly, in assays using the kit, 15 μL of PCR mix was added to 10 μL of RNA sample, and measurement was performed using an ABI™ 7500 Fast Dx system (Applied Biosystems, Foster City, CA, USA). Based on the obtained results, the ratio of minor *BCR-ABL* mRNA copies to *ABL* mRNA copies was calculated and normalized.

The limit of detection for this kit was determined according to Clinical and Laboratory Standards Institute guideline EP17-A2. Based on the results, for an *ABL* mRNA copy number of $\geq$10 000, the minimum detectable sensitivity was reported as a minor *BCR::ABL1* mRNA copy number of 13.58 copies/test and a minor *BCR::ABL1* mRNA to *ABL* mRNA ratio of 0.00151%.

Fusion gene transcripts

At initial diagnosis, fusion gene transcripts were screened for in patient BM samples using RT-qPCR, including minor *BCR-ABL* and major *BCR-ABL*. Once detected, the fusion gene transcripts were assayed, and minor *BCR-ABL* mRNA levels in BM or PB samples were determined.

Results

Concordance of test results between Otsuka and existing screening test for chimeric genes related to leukemia

Concordance between Otsuka and the existing screening test for leukemia-related chimeric genes was examined using samples collected from patients with suspected ALL (Sample No. 1), obtained before treatment initiation. Positive results in the control test, at or above 250 copies/μg RNA (minimum detectable sensitivity), were considered positive, while results below this threshold were considered negative. Similarly, positive results obtained with Otsuka at or above the minimum detectable sensitivity were classified as positive, and results below that threshold were considered negative.

Table 1A presents the results for 180 (153 adult and 27 pediatric) BM samples: the positive agreement rate was 97.6% (40/41), negative agreement rate was 95.0% (132/139) and overall agreement rate was 95.6% (172/180). The kappa coefficient, which measures diagnostic agreement, was 0.88, with a 95% confidence interval (CI) of 0.80–0.96.

Table 1. Consistency rate between Otsuka and control test

			Control test				
			Positive	Negative	Total		
(A) BM							
	Otsuka	Positive	40	7	47	Positive consistency rate:	97.6% (40/41)
		Negative	1	132	133	Negative consistency rate:	95.0% (132/139)
		Total	41	139	180	Total consistency rate:	95.6% (172/180)
(B) PB							
	Otsuka	Positive	38	8	46	Positive consistency rate:	97.4% (38/39)
		Negative	1	114	115	Negative consistency rate:	93.4% (114/122)
		Total	39	122	161	Total consistency rate:	94.4% (152/161)

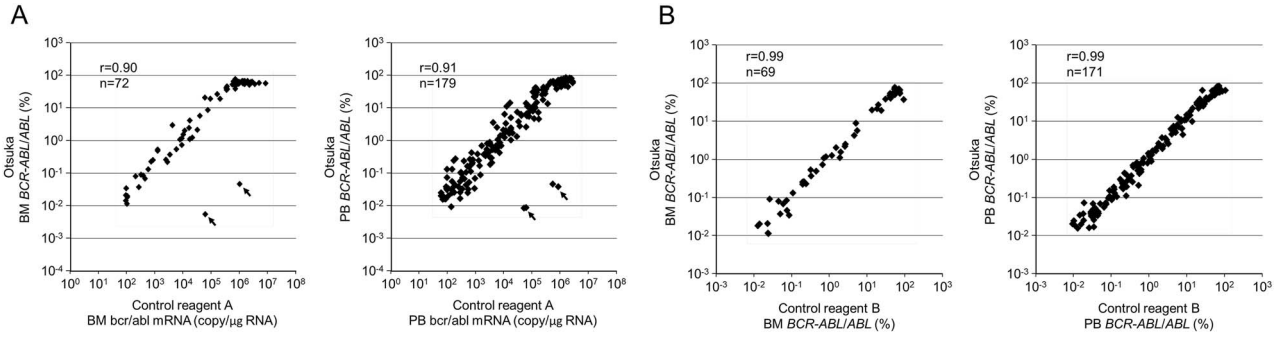


Figure 1. Correlation between Otsuka and control reagent A or control reagent B. (A) A strong correlation was observed between *BCR-ABL* levels measured by Otsuka and those by control reagent A for both BM and PB samples from Ph+ ALL patients. Samples with *BCR-ABL* subtype e1a3 are indicated by arrows. (B) A strong correlation was observed between *BCR-ABL* levels measured by Otsuka and those by control reagent B for both BM and PB samples from Ph+ ALL patients.

Table 1B displays the results for 161 adult PB samples (pediatric PB was not collected): the positive agreement rate was 97.4% (38/39), negative agreement rate was 93.4% (114/122) and overall agreement rate was 94.4% (152/161). The kappa coefficient was 0.86, with a 95% CI of 0.77–0.95.

To evaluate performance in monitoring therapeutic effects, the correlation between measurements obtained by Otsuka and control reagent A or B was assessed. All samples above the detection threshold were examined for correlation: 72 (63 adult and 9 pediatric) BM and 179 (155 adult and 24 pediatric) PB samples were tested with Otsuka and control reagent A, showing a strong correlation with a correlation coefficient of r = 0.90 for BM samples and r = 0.91 for PB samples (Fig. 1A).

Similarly, using Otsuka and control reagent B, 69 (61 adult and 8 pediatric) BM and 171 (149 adult and 22 pediatric) PB samples above the detection threshold were evaluated, demonstrating a strong correlation with a correlation coefficient of r = 0.99 for both BM and PB samples (Fig. 1B).

Stratified analyses by disease

Figure 2 presents elegant scatter plots showing the minor *BCR-ABL* mRNA/ABL mRNA ratio (%) for BM samples (n = 153) in Fig. 2A and PB samples (n = 193) in Fig. 2B. These samples were collected from patients with ALL, non-ALL malignancies and healthy individuals (PB only) before treatment initiation. Measurements below the limit of detection were represented as 0.0001%.

In Fig. 2A, all patients with Ph+ ALL and minor BCR-ABL (n = 29) tested positive (maximum: 78.2%, mean: 55.0%), whereas patients with Ph- ALL (n = 42) tested negative. Interestingly, although patients with Ph+ ALL and major *BCR-ABL* (n = 6) tested positive, the minor *BCR-ABL* mRNA/ABL mRNA ratios (%) were low (maximum: 0.0822%, mean: 0.0356%). Among patients with non-ALL malignancies (n = 76), positive cases included one with acute myeloid leukemia (AML) harboring minor *BCR-ABL* (7.64%) and one with chronic myeloid leukemia (CML) harboring minor *BCR-ABL* (37.9%). The remaining cases involved patients with CML

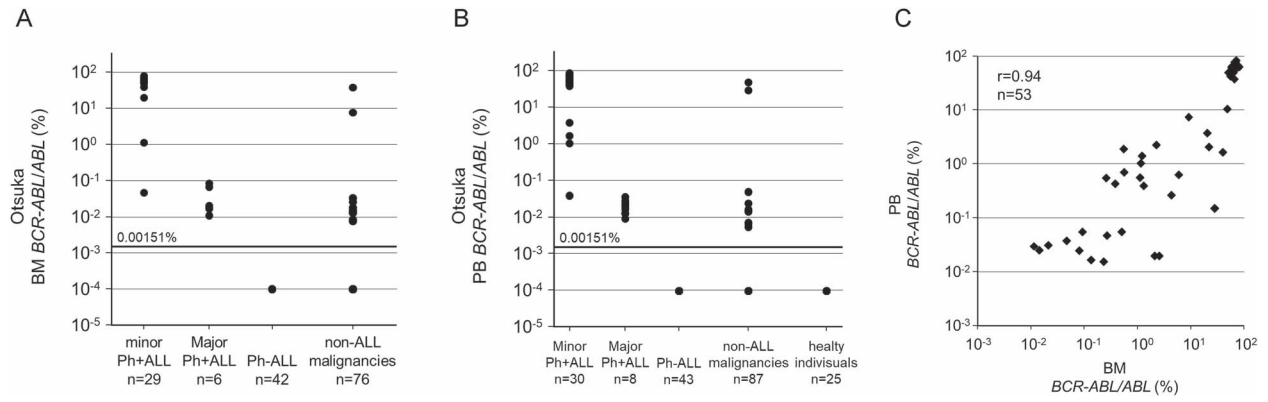


Figure 2. Distribution of minor *BCR-ABL/ABL* (%) measured by Otsuka (BM). (A) Comparison of *BCR-ABL* levels in BM samples from patients with Ph+ ALL harboring minor *BCR-ABL*, Ph+ ALL harboring major *BCR-ABL*, Ph− ALL and non-ALL malignancies. (B) Comparison of *BCR-ABL* levels in PB samples from patients with Ph+ ALL harboring minor *BCR-ABL*, Ph+ ALL harboring major *BCR-ABL*, Ph− ALL and non-ALL malignancies and healthy individuals. (C) Correlation of *BCR-ABL* levels between BM and PB samples.

and major *BCR-ABL*, where minor *BCR-ABL* mRNA/*ABL* mRNA ratios (%) were low (maximum: 0.0326%, mean: 0.0166%).

In Fig. 2B, all patients with Ph+ ALL and minor *BCR-ABL* (n = 30) tested positive (maximum: 83.8%, mean: 53.5%), whereas patients with Ph- ALL (n = 43) and healthy individuals (n = 25) tested negative. Similarly, although patients with Ph+ ALL and major *BCR-ABL* (n = 8) tested positive, the minor *BCR-ABL* mRNA/*ABL* mRNA ratios (%) were low (maximum: 0.0355%, mean: 0.0199%). Among patients with non-ALL malignancies (n = 87), positive cases included one with AML harboring minor *BCR-ABL* (28.4%) and one with CML harboring minor *BCR-ABL* (46.9%). The remaining cases involved patients with CML and major *BCR-ABL*, where minor *BCR-ABL* mRNA/*ABL* mRNA ratios (%) were low (maximum: 0.0490%, mean: 0.0173%).

Comparison of BM and PB

A total of 82 sample pairs were available for patients with Ph+ ALL harboring minor *BCR-ABL*, allowing measurement comparison in both BM and PB samples before treatment initiation and before/after consolidation therapy (Sample Nos. 1, 6 and 9, respectively). Out of these pairs, 53 exhibited BM and PB data exceeding the limit of detection. Correlation analysis between BM and PB data, specifically for BCR-ABL/ABL (%) (Fig. 2C), revealed a strong correlation with r = 0.94. However, three pairs displayed a ≥2-log difference, consistently showing higher values in BM compared with PB.

Eighteen of the 82 sample pairs had both BM and PB data below the detection limit. In the remaining 11 pairs, BM data were above the detection limit and PB data were below the detection limit. Notably, all these samples were collected after treatment initiation (Samples Nos. 6 and 9). This disparity is likely attributed to the treatment response, which eliminates leukemia cells from PB earlier than from the BM (16) (data not shown, Supplementary Fig. 1).

Time course of *BCR-ABL*/*ABL* in BM and PB during therapy in individual patients

We continued sample collection from 31 out of 33 patients with Ph+ ALL harboring minor *BCR-ABL*. Figure 3A presents representative data from the 31 sample pairs. Measurements below the minimum detectable sensitivity were plotted as 0.0001%.

The minor *BCR-ABL/ABL* (%) results obtained using Otsuka throughout the study period ranged from 0.00485 to 78.2% for BM and from 0.00761 to 83.8% for PB.

Comparison of the time course of minor *BCR-ABL* mRNA levels measured by Otsuka and existing research diagnostic kits for research use

By comparing the time course of minor *BCR-ABL* mRNA levels measured by Otsuka with those obtained using control reagents A and B, we analyzed 30 patients with Ph+ ALL harboring minor *BCR-ABL* who had data available for at least 2 consecutive time points. Figure 3B presents data from BM and PB samples of two patients as representative examples. The time course results obtained using the Otsuka kit closely aligned with those obtained using control reagents A and B.

Data measured by control reagent A were plotted as 1 copy/μg RNA if they fell below the minimum detectable sensitivity (50 copies/μg RNA). If data were below the minimum detectable sensitivity of Otsuka or were undetectable or not calculable using control reagent B, they were plotted as 0.0001%.

Discussion

In Japan, there was no approved diagnostic agent to identify minor *BCR-ABL* MRD levels, and no standard measurement method had been established. In this study, we evaluated the efficacy of Otsuka, a novel assay kit focusing on the e1a2 breakpoint of *BCR-ABL* mRNA, which is frequently found in Ph+ ALL. Notably, the kit uses *ABL* mRNA as the reference gene instead of *GAPDH* mRNA. As a result, this kit makes it possible to compare MRD data with those from other countries. The study revealed a significant concordance between Otsuka and the conventional control test in detecting minor *BCR-ABL* mRNA in both BM and PB samples across multiple centers. Quantitative analysis also established a strong correlation between Otsuka and the two control reagents. Otsuka specifically detected minor *BCR-ABL* mRNA and exhibited comparable time course data to the control reagents during the therapeutic process, indicating its favorable performance.

Although the results demonstrated a high agreement between Otsuka and the control assay, a few false positive cases were reported (7/115 in BM samples and 8/122 in PB samples). This was thought

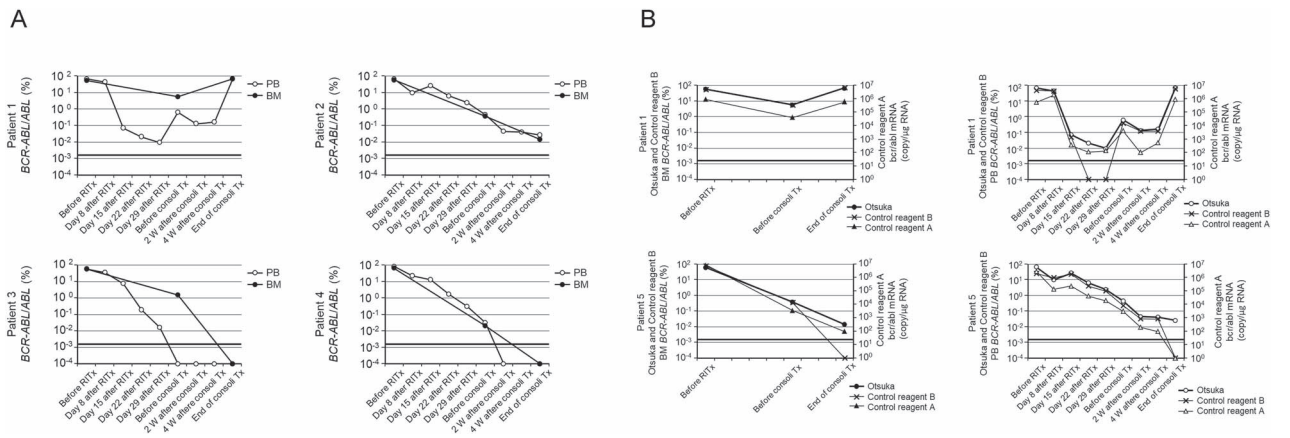


Figure 3. Time course of *BCR-ABL* levels in BM and PB samples over the therapeutic process. (A) Patient 1; The *BCR-ABL* level in PB was lowest 29 days after remission induction therapy but did not decrease to at or below the minimum detectable sensitivity. Later, this patient experienced hematological relapse. Patient 2; In both BM and PB samples, *BCR-ABL* levels declined with treatment, but did not decrease to at or below the minimum detectable sensitivity. Patient 3; *BCR-ABL* levels in PB samples were at or below the minimum detectable sensitivity before the start of consolidation therapy. After the completion of consolidation therapy, *BCR-ABL* levels in BM samples also decreased to at or below the minimum detectable sensitivity. Patient 4; *BCR-ABL* levels in PB samples decreased to at or below the minimum detectable sensitivity within 2 weeks of the start of consolidation therapy. After the completion of consolidation therapy, *BCR-ABL* levels in BM samples also decreased at or below the minimum detectable sensitivity. (B) Time course of *BCR-ABL* levels in Patient 1 (BM and PB) and Patient 5 (BM and PB), measured by Otsuka, control reagent A and control reagent B. Time courses of *BCR-ABL* levels were similar regardless of diagnostic reagents.

to be due to the difference in minimum detectable sensitivity between Otsuka and the control assay [13.58 copies/test (1 test is equivalent to 1 μgRNA) and 250 copies/μgRNA, respectively]. All of these cases occurred in patients with ALL or CML who had major *BCR-ABL*. This can be attributed to infrequent alternative splicing events that occur when major BCR-ABL mRNA is abundantly transcribed, a phenomenon often observed in Ph+ leukemia patients with major BCR-ABL (17). Stratified analyses by disease revealed positive test results for samples from patients with Ph+ ALL or CML harboring major BCR-ABL, but with a low minor BCR-ABL mRNA/ABL mRNA ratio, which can also be explained by alternative splicing (Fig. 2A and B).

In the correlation plots between Otsuka and control reagent A (Fig. 1A), a few samples displayed inconsistent results between the two methods (indicated by arrows in Fig. 1A), all originating from a single patient (two BM and four PB samples). Sanger sequencing analysis determined the presence of *BCR-ABL* e1a3, a rare variant subtype accounting for 1 to 2% of Ph+ ALL cases (18). In our study, the e1a3 subtype occurred in 1 out of 33 adult patients and 1 out of 3 pediatric patients. Otsuka was unable to detect e1a3 due to its primer design targeting the second exon of the *ABL1* gene (data not shown). Consequently, after excluding data from this patient, we analyzed the correlation between Otsuka and control reagent A in BM and PB samples, revealing a strong correlation with r = 0.98 (n = 61) for BM samples and r = 0.97 (n = 151) for PB samples.

The correlation analysis of *BCR-ABL/ABL* (%) between BM and PB (Fig. 2C) revealed that BM values exceeded PB data in some samples collected post-treatment initiation. However, Fig. 3A demonstrated good agreement between BM and PB data in samples collected pre-treatment (Sample No. 1), suggesting that either BM or PB samples can be used for measurements before treatment commencement. As mentioned previously, the 11 points where the detection results differed between PB and BM were all patterns where BM exceeded the detection sensitivity and PB was below the detection sensitivity. Conversely, there was no pattern where PB sensitivity exceeded the detection sensitivity and BM sensitivity was below the detection sensitivity. In other words, when PB exceeded the detection sensitivity or showed an increasing trend, the sensitivity of BM exceeded the detection sensitivity without exception. Therefore, the PB measurement of this assay is at least useful for predicting relapse or non-remission in BM. (19).

While major *BCR-ABL* testing has been standardized for clinical use, the evaluation of minor *BCR-ABL* fusion genes lacks such standardization, particularly in Japan where in-house or laboratory-developed reagents are employed. The Europe Against Cancer program has emphasized the importance of sensitive and accurate MRD quantification, leading to efforts to standardize qRT-PCR analysis. In the CML community, international standardization efforts have been ongoing since 2003. Similar endeavors have been reported for Ph+ ALL, where the e1a2 breakpoint, prevailing in this disease, was prioritized for qRT-PCR standardization, differing from the e13a2 and e14a2 breakpoints typical of CML (20).

To summarize, Otsuka represents the first clinical assay kit in Japan that accurately detects minor *BCR-ABL* mRNA levels, matching or surpassing the performance of conventionally used research reagents. Based on these findings, the assay kit obtained approval from the Ministry of Health, Labour, and Welfare in Japan in June 2021, with health insurance coverage commencing in November 2021. Regular monitoring of minor *BCR-ABL* mRNA in Ph+ ALL BM or PB samples at clinics will provide valuable insights to healthcare professionals for assessing individual patients' disease status.

Acknowledgements

We thank all participating institutions and physicians, not only the authors, but also the following investigators, for their support for the study.

The following investigators participated in this study: Kazutaka Sunami (National Hospital Organization Okayama Medical Center), Kenichi Ishizawa (Yamagata University Hospital), Yasunori Ueda (Kurashiki Central Hospital), Emiko Sakaida (Chiba University Hospital), Hirokazu Tanaka (Kindai University Faculty of Medicine), Junya Kuroda (Kyoto Prefectural University of Medicine), Yuki Yuza (Tokyo Metropolitan Children's Medical Center), Shingo Yano (The Jikei University School of Medicine), Mitsutoshi Kurosawa (National Hospital Organization Hokkaido Cancer Center), Koichi Miyamura (Japanese Red Cross Nagoya First Hospital), Jiro Fujita (Osaka

University Graduate School of Medicine), Hironori Ueno (National Hospital Organization Tokyo Medical Center), Tomohiko Kamimura (Harasanshin Hospital), Kaichi Nishiwaki (The Jikei University Kashiwa Hospital), Kazuma Oyashiki (Tokyo Medical University Hospital), Katsuyoshi Koh (Saitama Children's Medical Centre), Shinichiro Yoshida (National Hospital Organization Nagasaki Medical Center), Takuo Ito (National Hospital Organization Kure Medical Center), Yasushi Takamatsu (Fukuoka University Hospital), Masaya Okada (Hyogo College of Medicine Hospital), Tsuyako Iwai (Shikoku Medical Center for Children and Adults) and Kayo Taira (Hakata Clinic).

Supplementary material

Supplementary material is available at *Japanese Journal of Clinical Oncology* online.

Funding

This study was funded by Otsuka Pharmaceutical Co., Ltd., Tokyo, Japan.

Conflict of interest statement

Michihiro Hidaka, Masashi Sawa, Nobuaki Dobashi and Tomoki Naoe received research funding from Otsuka. Nobuaki Dobashi also received honoraria and payment for expert testimony from Otsuka. Naoto Takahashi received research funding and honoraria from Otsuka. Itaru Matsumura, Yuzuru Kanakura and Tomoki Naoe received personal fees from Otsuka during the conduct of the study. Asuka Sato and Ryuta Ito are employees of Otsuka Pharmaceutical Co., Ltd.

References

1. Redaelli A, Laskin L, Stephens M, Botteman F, Pashos L. A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL). *Eur J Cancer Care* 2005;14:53–62.
2. Liu-Dumlao T, Kantarjian H, Thomas A, O'Brien S, Ravandi F. Philadelphia-positive acute lymphoblastic leukemia: current treatment options. *Curr Oncol Rep* 2012;14:387–94.
3. Kaczmarska A, Śliwa P, Zawitkowska J, Lejman M. Genomic analyses of pediatric acute lymphoblastic leukemia Ph+ and Ph-like-recent progress in treatment. *Int J Mol Sci* 2021;22:6411–27.
4. Kantarjian M, Talpaz M, Dhingra K, et al. Significance of the P210 versus P190 molecular abnormalities in adults with Philadelphia chromosome-positive acute leukemia. *Blood* 1991;78:2411–8.
5. Bernt M, Hunger P. Current concepts in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia. *Front Oncol* 2014;4:1–21.
6. Yanada M, Takeuchi J, Sugiura I, et al. High complete remission rate and promising outcome by combination of imatinib and chemotherapy for newly diagnosed BCR-ABL-positive acute lymphoblastic leukemia: a phase II study by the Japan Adult Leukemia Study Group. *J Clin Oncol* 2006;24:460–6.
7. Hatta Y, Mizuta S, Matsuo K, et al. Final analysis of the JALSG Ph+ALL202 study: tyrosine kinase inhibitor-combined chemotherapy for Ph+ALL. *Ann Hematol* 2018;97:1535–45.
8. Sugiura I, Doki N, Hata T, et al. Dasatinib-based 2-step induction for adults with Philadelphia chromosome–positive acute lymphoblastic leukemia. *Blood Adv* 2022;6:624–36.
9. Rousselot P, Coudé M, Gokbuget N, et al. Dasatinib and low-intensity chemotherapy in elderly patients with Philadelphia chromosome-positive ALL. *Blood* 2016;128:774–82.
10. Berry A, Zhou S, Higley H, et al. Association of minimal residual disease with clinical outcome in pediatric and adult acute lymphoblastic leukemia: a meta-analysis. *JAMA Oncol* 2017;3:e170580.
11. Ravandi F, Jorgensen L, Thomas A, et al. Detection of MRD may predict the outcome of patients with Philadelphia chromosome-positive ALL treated with tyrosine kinase inhibitors plus chemotherapy. *Blood* 2013;122:1214–21.
12. Nishiwaki S, Imai K, Mizuta S, et al. Impact of MRD and TKI on allogeneic hematopoietic cell transplantation for Ph+ALL: a study from the adult ALL WG of the JSHCT. *Bone Marrow Transplant* 2016;51:43–50.
13. Akahoshi Y, Arai Y, Nishiwaki S, et al. Minimal residual disease (MRD) positivity at allogeneic hematopoietic cell transplantation, not the quantity of MRD, is a risk factor for relapse of Philadelphia chromosome-positive acute lymphoblastic leukemia. *Int J Hematol* 2021;113:832–9.
14. Osumi K, Fukui T, Kiyoi H, et al. Rapid screening of leukemia fusion transcripts in acute leukemia by real-time PCR. *Leuk Lymphoma* 2002;43:2291–9.
15. *ipsogen BCR-ABL1 mbcr Controls Kit [Internet]*. Venlo, The Netherlands; [cited 2022 Aug 27]. ipsogen BCR-ABL1 mbcr Kit; Available from: https://www.qiagen.com/us/products/diagnostics-and-clinical-research/oncology/ipsogen-leukemia/ipsogen-bcr-abl1-mbcr-p190-kit.
16. Scheuring J, Pfeifer H, Wassmann B, et al. Early minimal residual disease (MRD) analysis during treatment of Philadelphia chromosome/Bcr-Abl–positive acute lymphoblastic leukemia with the Abl-tyrosine kinase inhibitor imatinib (STI571). *Blood* 2003;101:85–90.
17. Lee S, LeMaistre A, Kantarjian M, et al. Detection of two alternative BCR/ABL mRNA junctions and minimal residual disease in Philadelphia chromosome positive chronic myelogenous leukemia by polymerase chain reaction. *Blood* 1989;73:2165–70.
18. Burmeister T, Schwartz S, Taubald A, et al. Atypical BCR-ABL mRNA transcripts in adult acute lymphoblastic leukemia. *Haematologica* 2007;92:1699–702.
19. Muffly L, Sundaram V, Chen C, et al. Concordance of peripheral blood and bone marrow measurable residual disease in adult acute lymphoblastic leukemia. *Blood Adv* 2021;5:3147.
20. Pfeifer H, Cazzaniga G, van der Velden VHJ, et al. Standardisation and consensus guidelines for minimal residual disease assessment in Philadelphia-positive acute lymphoblastic leukemia (Ph + ALL) by real-time quantitative reverse transcriptase PCR of e1a2 BCR-ABL1. *Leukemia* 2019;33:1910–22.

Received: 5 July 2023 | Revised: 12 October 2023 | Accepted: 12 November 2023

DOI: 10.1002/cnr2.1938

ORIGINAL ARTICLE

Cancer Reports | WILEY

Female and preserved platelet count subgroups of myelodysplastic syndrome patients benefit from standard-dose azacitidine

Shinichi Ogawa[1] | Tatsuhiro Sakamoto[2] | Ryota Matsuoka[3] | Kantaro Ishitsuka[4] | Yasuko Ogino[1] | Ayano Sootome[2] | Kenichi Makishima[4] | Chikashi Yoshida[5] | Yufu Ito[6] | Seiichi Shimizu[6] | Takuya Suyama[7] | Atsushi Shinagawa[7] | Takayoshi Ito[1] | Naoshi Obara[2] | Manabu Kusakabe[2] | Mamiko Sakata-Yanagimoto[2] | Yasushi Miyazaki[8] | Yasuhito Nannya[9] | Shigeru Chiba[2]

[1]Division of Hematology, JA Toride General Medical Center, Toride, Ibaraki, Japan

[2]Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan

[3]Department of Pathology, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan

[4]Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

[5]Division of Hematology, National Hospital Organization Mito Medical Center, Mito, Ibaraki, Japan

[6]Division of Hematology, Tsuchiura Kyoudou General Hospital, Tsuchiura, Ibaraki, Japan

[7]Division of Hematology, Hitachi General Hospital, Hitachi, Ibaraki, Japan

[8]Department of Hematology, Atomic Bomb Disease Institute, Nagasaki University, Nagasaki, Japan

[9]Department of Hematology, Institute of Medical Science, University of Tokyo, Tokyo, Japan

Correspondence
Shinichi Ogawa, Division of Hematology, JA Toride General Medical Center, 302-0022, Hongo 2-1-1, Toride, Ibaraki, Japan.
Email: shin199901598ogawa@hotmail.co.jp

Funding information
Japan Agency for Medical Research and Development, Grant/Award Number: cm0106505h; Japan Society for the Promotion of Science, Grant/Award Number: 21K16261

Abstract

Background: Hypomethylating agents, including azacytidine (AZA), are standard therapeutics for patients with high-risk myelodysplastic syndromes (MDS), a group of myeloid neoplasms. However, treatment schedules are not unified in real-world practice; in addition to the standard 7-day (standard-dose) schedule, shortened (reduced-dose) schedules are also used.

Aims: The aim of this study was to discover the patient group(s) which show differential efficacy between standard-and reduced-dose AZA to MDS.

Methods and Results: The outcome of different AZA doses in a cohort of 151 MDS patients were retrospectively analyzed. Overall survival (OS) was not significantly different between standard- and reduced-dose AZA groups by multivariate analysis. However, an interaction was found between either the sex (female vs. male), the platelet counts (< 40 $\times$ 10^3/μl vs. ≥ 40 $\times$ 10^3/μl), or the karyotype risk (< poor vs. ≥ poor) and standard-dose AZA for longer OS. Subgroup analyses revealed better OS with standard- over reduced-dose AZA in female patients (HR, 0.27 [95% CI,

0.090-0.79]; $p = 0.011$), and those with platelet counts ≥ 40 × 10^3/μl (HR, 0.51 [95% CI, 0.26-0.99]; $p = 0.041$). The union of female and preserved platelet count subgroups also benefited from standard-dose AZA. With this as a test cohort, we next analyzed patients registered in the JALSG MDS212 study, for whom 7-day and 5-day AZA treatment strategies were prospectively compared, as a validation cohort ($N = 172$). That cohort showed the same tendency as the retrospective results.

Conclusion: We identified the union of female and preserved platelet count subgroups which benefited from standard-dose AZA, imparting crucial information to physicians planning treatment regimens in MDS patients.

KEYWORDS
azacitidine, dose, myelodysplastic syndrome, platelet counts, sex

1 | INTRODUCTION

Myelodysplastic syndromes (MDS) are a group of chronic myeloid neoplasms characterized by pancytopenia, dysplasia, and predisposition to acute myeloid leukemia (AML).[1,2] The mainstay of therapy includes hypomethylating agents (HMAs), including azacitidine (AZA), and hematopoietic stem cell transplantation (HSCT) if eligible.[3–6] In the standard protocol, AZA is given at 75 mg/m^2 per day for 7 consecutive days every 28 days based on a Phase III study demonstrating prolonged overall survival (OS).[5] High-risk MDS patients are also reported to respond to shortened schedules (reduced-doses) of AZA although OS benefit differences between standard- and reduced-doses are controversial.[7–11] Based on these reports, the 5-day protocol is often clinically used because of convenience and better tolerability.[12] If the standard protocol is superior to the reduced protocol, the patients who receive the reduced protocol may lengthen their OS by changing their administration protocol. On the other hand, if the reduced protocol is equal to the standard protocol, it needs to be considered whether the standard protocol is reconsidered to lighten the adverse effects on patients and reduce economic burden to both patients and the society. Therefore, it is important to exhaustively compare the efficacy and difference between the protocols. Because AZA is a backbone of new combinatorial therapies for MDS and AML with venetoclax, magrolimab, APR246, and so on,[13–17] detailed data between the standard and the reduced doses may influence clinical studies and resulting new therapeutic regimens. Although a Phase III clinical trial was conducted to prove the superiority of the 7-day over the 5-day protocol, it was never completed and statistically significant OS differences between the 7- and 5-day protocols were not proven.[18]

We had a community-based information that reduced-dose AZA is prescribed to a significant proportion of MDS patients in Ibaraki Prefecture in Japan. Thus, we conducted a multicenter retrospective study to disclose real-world dosing schedules and investigate any potential differences in OS between patients receiving AZA at standard- or reduced-doses. Furthermore, we intended to identify subcohorts in which AZA dose delineated OS. To define such subcohorts, interaction analyses between cumulative AZA dose and each clinical parameter were performed as screening before subgroup analyses were performed for selected parameters. Sex and platelet count were each related to AZA dose dependency.

After validation with a prospective cohort registered in the JALSG MDS212 study,[18] it was suggested that the standard- or near standard-dose of AZA, in comparison with the reduced-dose, improved OS in female patients and those with preserved platelet counts.

2 | METHODS

2.1 | Patients and inclusion criteria for clinical analyses

One-hundred and eighty-six patients were enrolled, all diagnosed with MDS according to either the FAB[19] or the WHO 2016 criteria,[1] and treated by AZA from March 2011 to May 2019 at 5 hospitals in Ibaraki Prefecture. Two patients with a history of HSCT before AZA administration and one with a shortage of clinical data were removed (Table S1).

To investigate the influence of differences in the dose of AZA on hematological improvement (HI) and OS, we further removed 32 patients who died sooner than day 112 after the commencement of AZA. Then, resulting 151 patients who survived for 112 (28 days × 4 courses) days or longer (survivor112) were determined as a main target of our analysis. This was because we planned to exclude the short survivors dying sooner than day 112 based on our understanding represented by the following reports. First, the median number of courses required for the initial response was three, and 90% of responses were seen by 6 courses in MDS.[20] Second, AZA should be continued for at least 4–6 courses to judge whether the patients respond to AZA or not in AML patients.[21] Therefore, we collected the cumulative AZA dose at day 112, as well as data on the total number of AZA treatment course and the mean period of AZA administration in each course (6 days or shorter, or longer than 6 days). This retrospective study was approved

by the institutional review board in each hospital. This retrospective study was based on the medical records. Obeying the approval of each institutional review board, we performed opt-out in each hospital instead of written informed consent.

2.2 | Definitions of hematological improvement, survival, and cumulative AZA dose

Hematological improvement (HI) to AZA was defined according to the revised IWG 2019 hematological response criteria.[22] OS was defined as the time from the day of the first administration of AZA to the day of death caused by any reasons. Living patients were censored at the last contact and those patients receiving stem cell transplantation were censored at the day of the stem cell infusion. Cumulative doses of AZA (mg/m^2) in the first 4 courses were calculated by dividing the sum of AZA given on or before day 112 by the body surface area at the first administration of AZA. If AZA was administered at 75 mg/m^2/day for 5 days and body surface area was unchanged, the cumulative AZA dose was considered to be 1500 mg/m^2. Based on this calculation, cumulative AZA doses equal to or less than 1500 mg/m^2 were defined as reduced-dose while over 1500 mg/m^2 was defined as the standard-dose.

2.3 | Statistics

Fisher's exact test was used for univariate analyses of binary variables for response to AZA while the Mann–Whitney U test was used for univariate analysis of continuous variables. Logistic regression modeling was used for multivariate analyses of binary variables for response to AZA. OS was evaluated using the Kaplan–Meier method. The log-rank test was used to compare the survival curves between the patient groups of interest. The Cox proportional hazard model was used to estimate hazard ratios (HR) and 95% confidence intervals (95% CI) of HR in univariate and multivariate analyses of OS. In multivariate analyses of hematological improvement rate and OS, age (<75 vs. ≥75), sex, karyotype risk defined by the revised international prognostic scoring system criteria (IPSS-R) (<poor vs. ≥poor),[23] bone marrow blast percentage (<10% vs. ≥10%), neutrophile counts (<800/μl vs. ≥800/μl), hemoglobin levels (<8 g/dL vs. ≥8 g/dL), platelet counts (<40 × 10^3/μl vs. ≥40 × 10^3/μl), and cumulative AZA doses (reduced-dose vs. standard-dose) were included as explanatory variables, irrespective of p values. Factors with p values <.05 were additionally included in explanatory variables. To obtain the propensity score (PS), the probability to receive the standard-dose was calculated using a logistic regression model in which explanatory variables were age, sex, bone marrow blast percentage, WHO 2016 diagnosis, karyotype-risk defined by IPSS-R, with or without transplantation, neutrophile counts (Neu; /μL), hemoglobin levels (Hb; g/dL), and platelet counts (Plt; × 10^4/μL) at the first administration of AZA. The PS matching was performed using 1:1 caliper matching (caliper 0.2). Statistical analyses were performed using EZR.[24]

3 | RESULTS

3.1 | Patient characteristics

Characteristics of the 183 patients are shown in Table S1. The median age was 72 years (range, 29–90) with a male/female ratio of 2.05. Myelodysplastic syndrome with excess blasts 1 (MDS-EB1) and MDS-EB2 were the most prevalent (62.8%), followed by MDS with multiple lineage dysplasia (MDS-MLD; 19.1%) and AML with myelodysplasia-related changes (AML-MRC; 10.4%), according to the WHO 2016 criteria. All AML-MRC cases corresponded to refractory anemia with excess blasts (RAEB) in transformation (RAEB-t) according to the FAB classification. Based on IPSS-R, 71.1% were judged to have high- or very high-risk prognosis. The median number of AZA courses was 6 (range, 1–61).

Of Survivors112, the standard- and reduced-doses were given to 91 and 60 patients, respectively (Table 1). Median cumulative AZA doses at day 112 were 2074 mg/m^2 (10–90 percentile, 1575–2100 mg/m^2) and 1232 mg/m^2 (853–1500 mg/m^2) in the standard- and reduced-dose groups, respectively (Figure S1). The median Hb concentrations at the first administration of AZA were significantly higher in the standard-dose group than the reduced-dose group (8.6 g/dL vs. 7.8 g/dL, $p = .04$, Table 1), which potentially influenced the choice of the AZA dose. All other factors, including age, sex, diagnosis, IPSS-R-risk, karyotype-risk, Neu and Plt at the first administration of AZA, and bone marrow blast percentage at diagnosis or within 3 months before AZA start, were not significantly different between the two groups.

3.2 | Hematological improvement

The hematological improvement (HI) rate in any parameter by AZA was 54.1% in Survivors112 (95% CI, 45.7%–62.4%) (Table 1). In univariate analyses, cumulative AZA dose and sex significantly affected the HI rate; these rates were greater with regard to standard-dose and male sex (Tables 1 and S2). All other factors, such as age (<75 or ≥75), bone marrow blast percentage (<10% or ≥10%), or IPSS-R-risk (< high or ≥high), karyotype-risk (< poor or ≥poor), Neu, Hb, and Plt, did not significantly affect the HI rate. In our multivariate analysis, sex and the cumulative AZA dose were again the significant parameter affecting the HI rate (Table S2). Response to AZA based on bone marrow evaluations could not be investigated because these data at the appropriate time points after AZA initiation were missing in a substantial number of patients.

3.3 | Survival in the entire Survivors112 cohort

Median survival time (MST) was 509 days (95% CI, 445–640 days), while 1-year OS was 72.7% (95% CI, 64.2%–79.6%) in the Survivors112 cohort (Table 1). MST and 1-year OS in the whole cohort (183 patients) were described in Table S1. In univariate analyses, high or very high IPSS-R (HR, 1.85 [95% CI, 1.10–3.12]; $p = .019$), poor or

TABLE 1 Patient characteristics divided by cumulative dose of AZA at day 112.

	All patients (%)	Cumultive dose of AZA at day 112		*p* value
		≤1500 mg/m^2 (%)	>1500 mg/m^2 (%)	
N	151	60	91	
Age, median [range]	72 [29, 90]	74 [29, 90]	72 [42, 86]	.47
Sex				1
Male	102 (67.5)	41 (68.3)	61 (67.0)	
Female	49 (32.5)	19 (31.7)	30 (33.0)	
WHO 2016 criteria				.30
MDS-SLD	4 (2.6)	1 (1.7)	3 (3.3)	
MDS-MLD	27 (17.9)	10 (16.7)	17 (18.7)	
MDS-EB1	50 (33.1)	25 (41.7)	25 (27.5)	
MDS-EB2	46 (30.4)	15 (25.0)	31 (34.1)	
AML-MRC	17 (11.3)	6 (10.0)	11 (12.1)	
MDS with isolated del(5q)	1 (0.7)	1 (1.7)	0 (0.0)	
MDS-RS	2 (1.3)	0 (0.0)	2 (2.2)	
CMML	1 (0.7)	0 (0.0)	1 (1.1)	
tMN	2 (1.3)	2 (3.3)	0 (0.0)	
MDS-U	1 (0.7)	0 (0.0)	1 (1.1)	
IPSS-R risk group				.30
Very low	2 (1.3)	0 (0.0)	2 (2.2)	
Low	16 (10.6)	7 (11.7)	9 (9.9)	
Intermediate	26 (17.2)	6 (10.0)	20 (22.0)	
High	50 (33.1)	22 (36.7)	28 (30.8)	
Very high	5 (37.1)	25 (41.7)	31 (34.1)	
NA	1 (0.7)	0 (0.0)	1 (1.1)	
IPSS-R karyotype group				.17
Very good	3 (2.0)	0 (0.0)	3 (3.3)	
Good	61 (40.4)	18 (30.0)	43 (47.3)	
Intermediate	33 (21.9)	15 (25.0)	18 (19.8)	
Poor	11 (7.3)	6 (10.0)	5 (5.5)	
Very poor	39 (25.8)	19 (31.7)	20 (22.0)	
NA	4 (2.6)	2 (3.3)	2 (2.2)	
Transplantation				1
No	137 (90.7)	55 (91.7)	82 (90.1)	
Yes	14 (9.3)	5 (8.3)	9 (9.9)	
Bone marrow blast %, median [range]	7.6 [0.0, 29.8]	7.0 [0.4, 26.5]	8.4 [0.0, 29.8]	.56
Neutrophile count (/μL), median [range]	888 [47, 22 243]	904 [110, 22 243]	880 [47, 19 757]	.76
Hemoglobin (g/dL), median [range]	8.1 [2.4, 12.9]	7.8 [4.2, 12.9]	8.6 [2.4, 12.9]	.040
Platelet count ($\times 10^3$/μL), median [range]	63 [5, 629]	64 [10, 629]	62 [5, 364]	.32
Hematological improvement rate, % (95% CI)	54.1 (45.7–62.4)	36.2 (24.0–49.9)	65.9 (55.0–75.7)	.00064
Median follow up time, days (95% CI)	427 (364–449)	349.5 (263–428)	445 (400–522)	.35
Median survival time, days (95% CI)	509 (445–640)	427 (321–584)	623 (482–850)	.010
OS at 1 year, % (95% CI)	72.7 (64.2–79.6)	59.1 (44.6–71.0)	82.0 (71.4–88.9)	
OS at 2 year, % (95% CI)	35.9 (26.8–45.0)	24.3 (13.3–37.2)	44.7 (31.9–56.7)	

Note: Patients were included whoose overall survival was 112 days or longer.

Abbreviations: 95% CI, 95% confidence interval; AML-MRC, acute myeloid leukemia with myelodysplasia-related changes; CMML, chronic myelomonocytic leukemia; IPSS-R, revised international prognostic scoring system; MDS-EB1, myelodysplastic syndrome with excess blasts 1; MDS-EB2, myelodysplastic syndrome with excess blasts 2; MDS-MLD, myelodysplastic syndrome with multilineage dysplasia; MDS-SLD, myelodysplastic syndrome with single lineage dysplasia; MDS-RS, myelodysplastic syndrome with ring sideroblasts; MDS-U, myelodysplastic syndrome, unclassifieable; OS, overall survival; NA, not available; tMN, therapy related myeloid neoplasms.

very poor karyotype-risk (HR, 3.29 [95% CI, 2.04–5.31]; $p = 2.5 \times 10^{-7}$), no HI by AZA (HR, 2.21 [95% CI, 1.43–3.42]; $p = 2.5 \times 10^{-4}$), Hb <8 g/dL (HR, 0.62 [95% CI, 0.40–0.95]; $p = .025$), and the reduced-dose (HR, 0.58 [95% CI, 0.38–0.88]; $p = .010$) were significant factors for poor prognosis (Table S3 and Figure S2). In the multivariate Cox proportional hazards model, poor or very poor karyotype-risk (HR, 3.14 [95% CI, 1.88–5.23]; $p = 1.2 \times 10^{-5}$) and no HI (HR, 1.89 [95% CI, 1.13–3.18]; $p = .016$) significantly shortened the OS. The cumulative AZA dose was not an independent significant prognostic factor (Table S3).

Consequently, poor or very poor karyotype risk and no HI from AZA were negative prognostic factors in both univariate and multivariate analyses for the entire Survivors112 cohort, similarly to previously verified reports.

TABLE 2 The *p* values of interaction with cumulative AZA dose in the propensity score-matched analysis.

Age (<75 vs. ≥75)	.91
Sex (female vs. male)	.040
Bone marrow blast percentage (<10% vs. ≥10%)	.51
Neutrophile count (<800/μL vs. ≥800/μL)	.83
Hemoglobin (<8 g/dL vs. ≥8 g/dL)	.98
Platelet counts (<40 × 10^3/μL vs. ≥40 × 10^3/μL)	.28
Karyotype-risk (<poor vs. ≥poor)	.25

3.4 | Survival in the subcohorts

Because the univariate analysis showed OS differences between the AZA doses, we hypothesized that the benefit of the standard-dose AZA would be clearer if confounding factors were excluded. To remove such confounding factors and delineate subcohorts in which the standard-dose AZA prolonged OS than the reduced-dose AZA, we selected 94 patients by propensity score matching from Survivors112 (Table S4, Figure S3) and performed interaction analyses between each clinical parameter and cumulative AZA dose. We picked up any interactions providing *p* values of interaction less than .30. Sex (female vs. male), platelet counts (<40 × 10^3/μl vs. ≥40 × 10^3/μl), and karyotype risk (<poor vs. ≥poor) matched the criteria and were selected as

TABLE 3 Subgroup analyses of overall survival in the propensity score-matched analysis.

	Cumulative dose of AZA at day 112	*N*	Median survival time (95% CI) (Day)	HR (95% CI)	*p* value
Female					.011
	≤1500 mg/m^2	13	321 (129–422)		
	>1500 mg/m^2	14	850 (479–1737)	0.27 (0.090–0.79)	
Male					.78
	≤1500 mg/m^2	34	484 (371–707)		
	>1500 mg/m^2	33	482 (438–812)	0.91 (0.48–1.72)	
Platelet count <40 × 10^3/μL					.84
	≤1500 mg/m^2	15	429 (164–NA)		
	>1500 mg/m^2	15	473.5 (192–850)	0.91 (0.36–2.30)	
Platelet count ≥40 × 10^3/μL					.041
	≤1500 mg/m^2	32	412 (222–594)		
	>1500 mg/m^2	32	682 (458–940)	0.51 (0.26–0.99)	
Karyotype-risk < poor					.66
	≤1500 mg/m^2	27	640 (422–1072)		
	>1500 mg/m^2	28	812 (482–1737)	0.85 (0.41–1.76)	
Karyotype-risk ≥ poor					.062
	≤1500 mg/m^2	20	222 (141–427)		
	>1500 mg/m^2	19	448 (248–NA)	0.48 (0.22–1.07)	
Female or Platelet counts ≥40 × 10^3/μL					.0078
	≤1500 mg/m^2	36	371 (222–505)		
	>1500 mg/m^2	36	682 (458–940)	0.43 (0.23–0.82)	
Male and Platelet counts <40 × 10^3/μL					.38
	≤1500 mg/m^2	11	640 (132–NA)		
	>1500 mg/m^2	11	443 (167–NA)	1.60 (0.55–4.64)	

Abbreviations: 95% CI, 95% confidence interval; HR, hazard ratio; NA, not available.

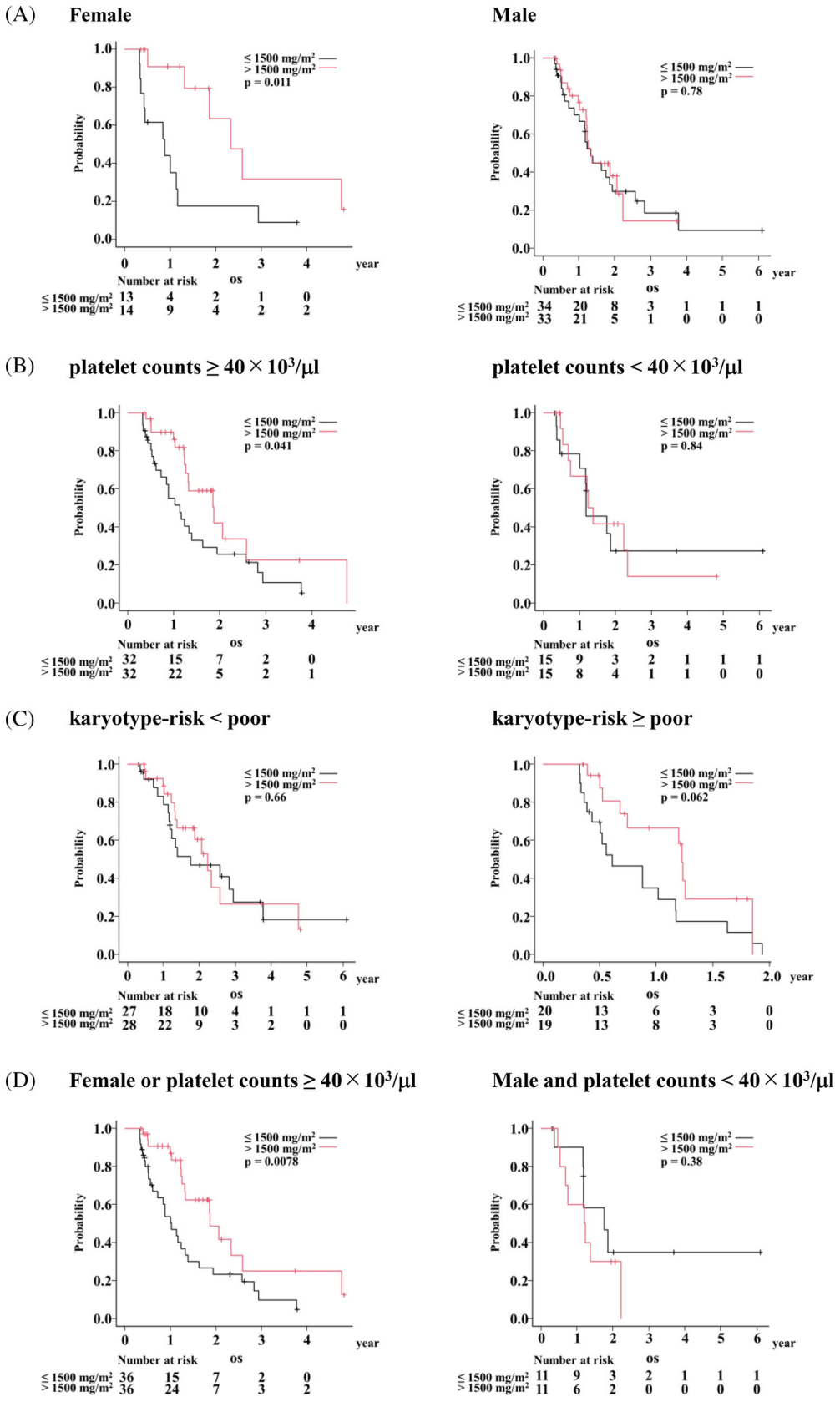


FIGURE 1 Subgroup analyses by sex, platelet counts, or karyotype in the propensity score-matched analysis. Female and male (A), platelet counts <40 × 10^3/μl and ≥40 × 10^3/μl (B), and karyotype-risk <poor and ≥poor (C), and the union of female and platelet counts ≥40 × 10^3/μl, and patients other than the union (D).

candidates for the subcohorts (Table 2). Then, univariate analyses were performed in each subcohort to investigate whether cumulative AZA dose influenced OS. The standard-dose significantly prolonged OS in the female (HR, 0.27 [95% CI, 0.090–0.79]; $p = .011$) and platelet counts ≥40 × 10^3/μl (HR, 0.51 [95% CI, 0.26–0.99]; $p = .041$) subcohorts (Table 3 and Figure 1). In the karyotype-risk ≥ poor

TABLE 4 Subgroup analyses of overall survival in 172 patients from the JALSG MDS212 cohort.

	Dose of AZA	*N*	Median survival time (95% CI) (Day)	HR (95% CI)	*p* value
Female					.26
	5 days	28	484 (438–732)		
	7 days	26	756 (420–1104)	0.68 (0.35–1.34)	
Male					.57
	5 days	61	497 (443–652)		
	7 days	57	537 (349–710)	0.88 (0.56–1.37)	
Karyotype-risk < poor					.49
	5 days	54	652 (483–848)		
	7 days	50	756 (569–1028)	0.84 (0.5–1.39)	
Karyotype-risk ≥ poor					.45
	5 days	35	438 (307–474)		
	7 days	33	378 (309–463)	0.81 (0.47–1.4)	
Platelet counts <40 × 10^3/μL					.85
	5 days	27	457 (288–695)		
	7 days	24	378 (238–455)	1.07 (0.55–2.07)	
Platelet counts ≥40 × 10^3/μL					.10
	5 days	62	509 (458–652)		
	7 days	59	710 (489–911)	0.69 (0.44–1.08)	
Female or platelet counts ≥40 × 10^3/μL					.067
	5 days	71	509 (458–652)		
	7 days	67	673 (463–868)	0.68 (0.44–1.03)	
Male and platelet counts <40 × 10^3/μL					.31
	5 days	18	457 (288–NA)		
	7 days	16	329 (169–455)	1.52 (0.68–3.4)	

Abbreviations: 95% CI, 95% confidence interval; HR, hazard ratio; NA, not available.

subcohort, there was a tendency that the standard-dose prolonged OS but without statistical significance (HR, 0.48 [95% CI, 0.22–1.07]; $p = .062$). The union of the female and platelet counts ≥40 × 10^3/μl subcohorts, in other words, the patients other than the male with platelet counts <40 × 10^3/μl, was delineated as the subcohort in which the standard-dose AZA improved OS than the reduced-dose AZA (HR, 0.43 [95% CI, 0.23–0.82]; $p = .0078$) (Table 3 and Figure 1).

To validate the results of our retrospective cohort, the OS of 172 patients who were prospectively treated with 7- and 5-day AZA (which correlates to the standard-and the reduced-doses, respectively, of the retrospective analysis) and survived 112 days or longer in the JALSG MDS212 study was compared.[18] In this entire JALSG day 112 survivor cohort, OS was not significantly different between 7- and 5-day AZA groups (HR, 0.80 [95% CI, 0.55–1.16]; $p = .24$). We then compared the OS between 7- and 5-day AZA arms of the following three subcohorts: female patients, those with platelet counts ≥40 × 10^3/μl, and those with karyotype-risk ≥ poor. In the female and the platelet counts ≥40 × 10^3/μl subcohorts, there was a tendency that OS was better in the 7-day AZA arm (HR, 0.68 [95% CI 0.35–1.34], $p = .26$; and 0.69 [95% CI, 0.44–1.08], $p = .10$; respectively). These OS differences were not observed in the male patients and those with platelet counts <40 × 10^3/μl. The difference between the two dose groups was marginal irrespective of karyotype-risk in the JALSG day 112 survivor cohort (Table 4 and Figure 2). In the patients other than the male with platelet counts <40 × 10^3/μl, there was a strong tendency that 7-day AZA prolonged OS (Table 4 and Figure 2; $p = .067$).

Taken together, our results suggested that the standard-dose AZA provided female patients and those with preserved platelet counts with better OS.

4 | DISCUSSION

By a retrospective analysis of 151 MDS patients who survived 112 days or longer after the starting of AZA, we found that OS in the female and the platelet counts ≥40 × 10^3/μl subcohorts significantly benefitted from the standard- rather than the reduced-dose. In the cohort of the Phase III clinical trial comparing the 7- and 5-day AZA scheduling,[18] the OS tended to be better with 7-day scheduling in the female patients and those with platelet counts ≥40 × 10^3/μl.

In real-world practice, either the standard- (7-day) or the reduced-dose (5-day) regimen is chosen without prognostic stratification. Our results showed that both regimens may have equal efficiency for OS prolongation in the male MDS patients with platelet

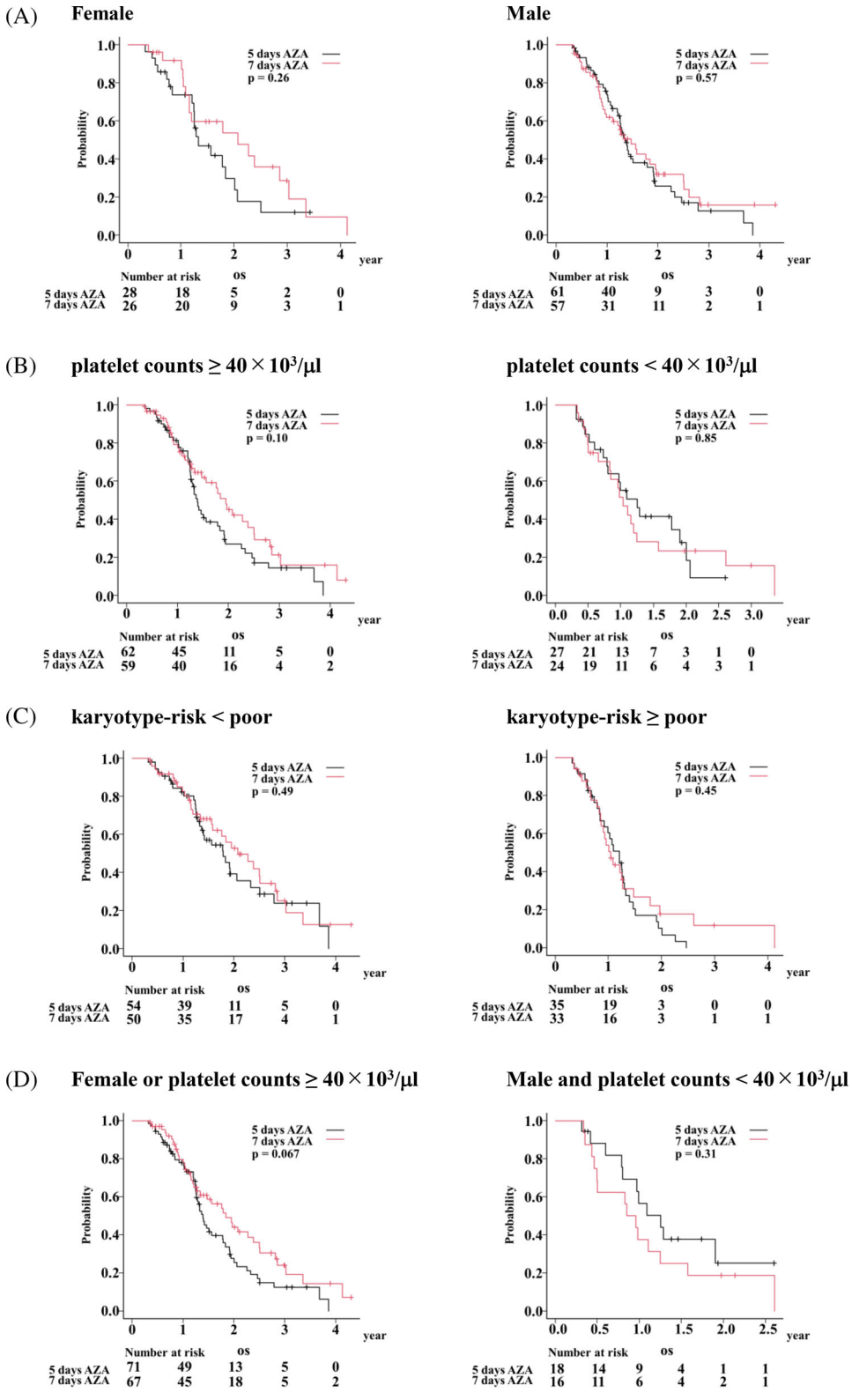


FIGURE 2 Subgroup analyses by sex, platelet counts, or karyotype in the validation cohort. Female and male (A), platelet counts <40 × 10^3/μl and ≥40 × 10^3/μl (B), karyotype-risk <poor and ≥poor (C), and the union of female and platelet counts ≥40 × 10^3/μl, and patients other than the union (D).

counts <40 × 10^3/μl. On the other hand, the standard-dose regimen reduced the risk of mortality by 57% and prolonged OS in the patients other than the male with platelet counts <40 × 10^3/μl (Table 3 and Figure 1). According to these results, AZA is recommended to be administered as the standard-dose, if the patients are not the male with platelet counts <40 × 10^3/μl. This provides novel and crucial information for the physicians treating high-risk MDS patients in choosing a treatment protocol, contributing to better quality of life and health economics.

A trend of shortened OS in patients receiving the reduced-dose, compared to the standard-dose, was previously described in a retrospective large cohort study.[25] Such a trend was also described in a

prospective study, albeit in a small number ($N = 22$) of patients based on a comparison with the AZA-001 study.[9] In a Phase III, JALSG MDS212 trial comparing 7- and 5-day AZA for RAEB and RAEB-t, although prematurely terminated because of poor recruitment, the 7-day protocol showed a statistically insignificant but visible trend of better OS (MST 538 [95% CI, 396–711] days) than the 5-day protocol (MST 477 [95% CI, 456–554] days).[18] Failure to demonstrate statistical OS differences in that study was attributed by the authors to insufficient statistical power, given that time to leukemia trasnformation was significantly longer with the 7-day protocol by multivariate analysis if only the centrally reviewed patients were investigated.[18]

In other previous reports comparing 5- and 7-day protocols, the conclusions have been controversial. Laribi et al. introduced relative dose intensity (RDI) of AZA (the relative dose intensity is the percentage of the dose received by the patient on the dose that theoretically should have been administered) to investigate how the dose of AZA influence the outcome of 93 high-risk MDS patients retrospectively. The OS and PFS were not different significantly with or without <80% RDI. The patients who responded to AZA were retrospectively divided into two groups; one group with <80% RDI and the other group without <80% RDI at the time when response was achieved. Dose reduction after the response was not considered. Then, they concluded that the group without <80% RDI showed significantly longer OS than the group with <80% RDI.[8] The time of response could be approximated by the day112,[20,21] and the doses administered to standard-dose group patients in our analysis resembled the group without <80% RDI in the report by Laribi, et al. Thus, the result of univariate analysis, but not multivariate analysis, in the current study may recapitulate the conclusion of Laribi et al. García-Delgado et al. retrospectively compared three regimens, 5 days (AZA5), 7 days including 2-day break (AZA 5-2-2), and 7 days (AZA7) in 200 patients with both high- and low-risk MDS patients, with majority with the latter. In this analysis, AZA 5-2-2 had significantly better response rate than AZA 5 or AZA 7, but the OS was not different among three regimens.[11] Fujimaki et al. compared the HI rate of their high-risk MDS patients on the 5-day AZA protocol with the HI rate of the high-risk MDS patients on the 7-day AZA protocol, and concluded that the HI rate was similar in both protocol.[10]

In the present study, 32 patients who died sooner than 112 days were removed from the landmark analysis for the Survivors112, but this removal could affect the conclusion. We explored potential differences between these short survivors and the Survivors112 by comparing the characteristics of patients. Although information on performance status and comorbidities was missing, we did not detect significant differences in other characteristics of patients between the two groups such as age, sex, diagnosis, risks on IPSS-R and karyotype, hematological parameters at the first administration of AZA, and bone marrow blast percentage at diagnosis or within 3 months before AZA start (Table S1).

In the Survivors112, the response to AZA by standard-dose was significantly better than reduced-dose in univariate and multivariate analyses. OS differences within Survivors112 were found between the two cumulative AZA dose groups in the univariate analysis. However, multivariate analysis did not show a significant difference. This could be due to biases or the dilution effect, which obscured findings in a specific subgroup of patients by other patients, according to the results of subgroup analyses in our study.

We reported that the standard-dose improved OS in female and the platelet counts ≥40 $\times$ 10^3/μl subcohorts within Survivors112. While we found an association between standard-dose and longer OS in specific subcohorts, there might be factors that influenced the results other than the standard-dose, given that the nature of retrospective analysis.

Preserved platelet count is an important component for prognosis prediction in IPSS-R[23] and, thus, should be selective for a subcohort with better OS. While this could be correlated to tolerability, clear explanations on why platelet, but not other blood cell lineages, affect the AZA dose preference remain elusive.

It was unexpected that the OS advantage imparted by the standard-dose was seen in female but not male patients. The activity of cytidine deaminase that inactivates AZA is known to be lower in females than males in a murine model.[26] Likewise, as cytarabine clearance from blood is known to be faster in males than females (as reported in a clinical trial),[27] AZA metabolism could differ between females and males and activity could persist in females if the dose is the same. It is, however, unclear whether and how this knowledge can explain the differences in observed outcomes.

There were several limitations in our study. First, the patients of our cohort came from 5 hospitals in Ibaraki prefecture, Japan, thus, there might be a geographical bias. Second, performance status and comorbidities were missing in our study. Third, the cumulative dose of AZA was surrogate index of 5-day dose or 7-day dose, but not equal to those. Fourth, the inclusion criteria between the current retrospective study and the JALSG MDS212 study were different. Of note, 26.8% of patients in the current retrospective cohort were MDS with low blasts, in contrast to the JALSG MDS212 cohort that included only RAEB and RAEB-t patients. Fifth, availability of the mutation profiles was incomplete in our retrospective cohort and not useful for the analysis.

Prospective study including large numbers of MDS patients is ideal to confirm results from our retrospective cohort. Given the premature termination of JALSG MDS212 prospective study due to poor recruitment, however, it might not be easy to perform a new prospective study comparing AZA doses in the future when new drugs would be equipped. In another way, better-designed retrospective analysis which takes the limitations of our cohort into account may be feasible. Simultaneously, it is warranted to elucidate the mechanism how the gender and platelet count influence OS under different AZA doses.

In conclusion, we identified by retrospective analysis that female and platelet counts ≥40 $\times$ 10^3/μl subcohorts of MDS (including oligoblastic AML), receive OS benefits from standard-dose rather than reduced-dose AZA. The same tendency was observed in the validation cohort independent of our cohort, although statistical significance was not seen.

AUTHOR CONTRIBUTIONS

Shinichi Ogawa: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); resources

(lead); writing – original draft (lead). **Tatsuhiro Sakamoto:** Conceptualization (equal); data curation (equal); formal analysis (equal); methodology (equal); resources (equal). **Ryota Matsuoka:** Formal analysis (equal). **Kantaro Ishitsuka:** Formal analysis (supporting). **Yasuko Ogino:** Resources (equal). **Ayano Sootome:** Resources (equal). **Kenichi Makishima:** Formal analysis (supporting). **Chikashi Yoshida:** Resources (equal). **Yufu Ito:** Resources (equal). **Seiichi Shimizu:** Resources (equal). **Takuya Suyama:** Resources (equal). **Atsushi Shinagawa:** Resources (equal). **Takayoshi Ito:** Resources (equal). **Naoshi Obara:** Validation (equal). **Manabu Kusakabe:** Resources (equal). **Mamiko Sakata-Yanagimoto:** Validation (equal). **Yasushi Miyazaki:** Validation (equal). **Yasuhito Nannya:** Validation (lead). **Shigeru Chiba:** Supervision (lead); writing – review and editing (lead).

ACKNOWLEDGMENTS

We thank all patients who participated in the study. We also thank physicians participating in the JALSG MDS212 study. This work was supported in part by AMED under Grant Number cm0106505h (to SC). This work was supported in part by a grant from JSPS to TS (KAKENHI #21K16261). The authors would also like to thank Dr. Bryan J. Mathis of the University of Tsukuba Hospital International Medical Center for language revision.

CONFLICT OF INTEREST STATEMENT

S.Ogawa received honorarium from Novartis, Janssen Pharmaceutical, Nippon Shinyaku, Chugai Pharmaceutical, Daiichi Sankyo, and Bristol-Myers Squibb outside the submitted work. The husband of A. Sootome has stock of Otsuka Holdings. K.Makishima received honorarium from Ono Pharmaceutical outside the submitted work. C. Yoshida received honorarium from Novartis, Otsuka Pharmaceutical, AbbVie GK, Janssen Pharmaceutical, Nippon Shinyaku, Chugai Pharmaceutical., and honoraria and research funding from Bristol-Myers Squibb outside the submitted work. Y. Ito received honorarium from Nippon Shinyaku outside the work. T. Suyama received honorarium from Abbvie, Meiji-Seika, Sanofi, Janssen Pharmaceutical, Kyowa-Kirin and Nippon Shinyaku outside the work, and has stock of Takeda Pharmaceutical. T. Ito received honorarium from Janssen Pharmaceutical, Ono Pharmaceutical, Nippon Shinyaku, Chugai Pharmaceutical, Sanofi, Meiji-Seika, Bristol-Myers Squibb, Takeda Pharmaceutical, Novartis, and Otsuka Pharmaceutical outside the work. N.Obara received research funding from Kyowa-Kirin outside the submitted work. M.Kusakabe received honorarium from Janssen Pharmaceutical, Astellas, Takeda Pharmaceutical, Chugai Pharmaceutical, Astrazeneca, Meiji-Seika outside the submitted work. M.Sakata-Yanagimoto received research funding from Eisai, Bristol-Myers Squibb, Otsuka and Mundipharma outside the submitted work, and honorarium from Nippon Shinyaku, Kyowa-Kirin, Chugai Pharmaceutical, Astellas, Meiji-Seika, Zenyaku Kogyo, Mundipharma outside the submitted work. Y.Miyazaki received honorarium from Nippon-Shinyaku, Bristol-Myers Squibb, Novartis, Sumitomo Pharma, Kyowa-Kirin, Abbvie, Daiichi-Sankyo, Takeda Pharmaceutical, Janssen Pharmaceutical, Astellas, Pfizer, Chugai Pharmaceutical, SymBio, Otsuka Pharmaceutical, and Research funding from Sumitomo-Dainippon outside the submitted work. Y.Nannya received honorarium from Takeda Pharmaceutical, Pfizer, Chugai Pharmaceutical, Sumitomo Pharma, Astrazeneca, Kyowa-Kirin, Asahi Kasei Pharma, Nippon Shinyaku, Fuji Pharma, Janssen Pharmaceutical, Filgen, Otsuka Pharmaceutical and Bristol Myers Squibb outside the submitted work, and research funding from Daiichi Sankyo RD Novare outside the submitted work. Y. Nannya participates in advisory boards in Novartis and Otsuka Pharmaceutical outside the submitted work. S.Chiba reports research funding from Ono Pharmaceutical, Chugai Pharmaceutical, Eisai, Kyowa Kirin, Astellas, Bayer, and Thyas outside the submitted work, and honorarium from Nippon Shinyaku, Chugai Pharmaceutical, Eisai, Kyowa-Kirin, Astellas, Janssen Pharmaceutical, Asahi Kasei, Sanofi, Meiji-seika, Takeda Pharmaceutical, AMGEN, Ono Pharmaceutical, MSD, Astrazeneca, Abbvie, Daiichisankyo, Pfizer, and Otsuka Pharmaceutical outside the submitted work. Any of the authors of this manuscript is not a current Editor or Editorial Board Member of Cancer Science. All authors had full access to all the data in this study and had final responsibility for the decision to submit for publication.

DATA AVAILABILITY STATEMENT

The raw data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

Approval of research protocol by an institutional review board. This retrospective study was approved by the institutional review board in each hospital.

INFORMED CONSENT

N/A. This retrospective study was based on the medical records. Obeying the approval of each institutional review board, we performed opt-out in each hospital instead of written informed consent.

ORCID

Shinichi Ogawa https://orcid.org/0000-0002-7416-0926

REFERENCES

1. Swerdlow SH, Campo E, Harris NL, et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, WHO Classification of Tumours*. Vol 2. 4th ed. International Agency for Research on Cancer; 2017.
2. Cazzola M. Myelodysplastic syndromes. *N Engl J Med*. 2020;383(14):1358-1374.
3. Kantarjian H, Issa JP, Rosenfeld CS, et al. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. *Cancer*. 2006;106(8):1794-1803.
4. Silverman LR, Demakos EP, Peterson BL, et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol*. 2002;20(10):2429-2440.
5. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol*. 2009;10(3):223-232.
6. Palacios-Berraquero ML, Alfonso-Piérola A. Current therapy of the patients with MDS: walking towards personalized therapy. *J Clin Med*. 2021;10(10).

7. Lyons RM, Cosgriff TM, Modi SS, et al. Hematologic response to three alternative dosing schedules of azacitidine in patients with myelodysplastic syndromes. *J Clin Oncol*. 2009;27(11):1850-1856.
8. Laribi K, Bolle D, Alani M, et al. Impact of the relative dose intensity on survival of patients with high-risk myelodysplastic syndromes treated with Azacitidine. *Cancer Med*. 2019;8(5):2188-2195.
9. Martin MG, Walgren RA, Procknow E, et al. A phase II study of 5-day intravenous azacitidine in patients with myelodysplastic syndromes. *Am J Hematol*. 2009;84(9):560-564.
10. Fujimaki K, Miyashita K, Kawasaki R, Tomita N. Efficacy and safety of a 5-day regimen of azacitidine for patients with high-risk myelodysplastic syndromes. *Eur J Haematol*. 2016;97(3):228-231.
11. García-Delgado R, de Miguel D, Bailén A, et al. Effectiveness and safety of different azacitidine dosage regimens in patients with myelodysplastic syndromes or acute myeloid leukemia. *Leuk Res*. 2014;38(7):744-750.
12. Grinblatt DL, Sekeres MA, Komrokji RS, Swern AS, Sullivan KA, Narang M. Patients with myelodysplastic syndromes treated with azacitidine in clinical practice: the AVIDA registry. *Leuk Lymphoma*. 2015;56(4):887-895.
13. DiNardo CD, Jonas BA, Pullarkat V, et al. Azacitidine and Venetoclax in previously untreated acute myeloid leukemia. *N Engl J Med*. 2020;383(7):617-629.
14. Sallman DA, Al Malki MM, Asch AS, et al. Magrolimab in combination with Azacitidine in patients with higher-risk myelodysplastic syndromes: final results of a phase Ib study. *J Clin Oncol*. 2023;41(15):2815-2826.
15. Cluzeau T, Sebert M, Rahmé R, et al. Eprenetapopt plus Azacitidine in TP53-mutated myelodysplastic syndromes and acute myeloid leukemia: a phase II study by the Groupe francophone des Myélodysplasies (GFM). *J Clin Oncol*. 2021;39(14):1575-1583.
16. Sallman DA, DeZern AE, Garcia-Manero G, et al. Eprenetapopt (APR-246) and Azacitidine in TP53-mutant myelodysplastic syndromes. *J Clin Oncol*. 2021;39(14):1584-1594.
17. Garcia-Manero G, Goldberg AD, Winer ES, et al. Eprenetapopt combined with venetoclax and azacitidine in TP53-mutated acute myeloid leukaemia: a phase 1, dose-finding and expansion study. *Lancet Haematol*. 2023;10(4):e272-e283.
18. Miyazaki Y, Kiguchi T, Sato S, et al. Prospective comparison of 5- and 7-day administration of azacitidine for myelodysplastic syndromes: a JALSG MDS212 trial. *Int J Hematol*. 2022;116:228-238.
19. Bennett JM, Catovsky D, Daniel MT, et al. Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol*. 1982;51(2):189-199.
20. Silverman LR, McKenzie DR, Peterson BL, et al. Further analysis of trials with azacitidine in patients with myelodysplastic syndrome: studies 8421, 8921, and 9221 by the cancer and leukemia group B. *J Clin Oncol*. 2006;24(24):3895-3903.
21. Santini V, Ossenkoppele GJ. Hypomethylating agents in the treatment of acute myeloid leukemia: a guide to optimal use. *Crit Rev Oncol Hematol*. 2019;140:1-7.
22. Platzbecker U, Fenaux P, Adès L, et al. Proposals for revised IWG 2018 hematological response criteria in patients with MDS included in clinical trials. *Blood*. 2019;133(10):1020-1030.
23. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*. 2012;120(12):2454-2465.
24. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant*. 2013;48(3):452-458.
25. Itzykson R, Thépot S, Quesnel B, et al. Prognostic factors for response and overall survival in 282 patients with higher-risk myelodysplastic syndromes treated with azacitidine. *Blood*. 2011;117(2):403-411.
26. Mahfouz RZ, Jankowska A, Ebrahem Q, et al. Increased CDA expression/activity in males contributes to decreased cytidine analog half-life and likely contributes to worse outcomes with 5-azacytidine or decitabine therapy. *Clin Cancer Res*. 2013;19(4):938-948.
27. Fleming RA, Capizzi RL, Rosner GL, et al. Clinical pharmacology of cytarabine in patients with acute myeloid leukemia: a cancer and leukemia group B study. *Cancer Chemother Pharmacol*. 1995;36(5):425-430.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ogawa S, Sakamoto T, Matsuoka R, et al. Female and preserved platelet count subgroups of myelodysplastic syndrome patients benefit from standard-dose azacitidine. *Cancer Reports*. 2023;e1938. doi:10.1002/cnr2.1938

Tohoku J. Exp. Med., 2023, **261**, 267-272

Small Intestinal Adenocarcinoma Arising at the Anastomotic Site after Kasai Operation for Biliary Atresia: A Case Report and Literature Review

Yuki Ishikawa,[1] Ryusuke Saito,[1] Keigo Murakami,[2] Atsushi Fujio,[1] Koji Miyazawa,[1] Kengo Sasaki,[1] Muneyuki Matsumura,[1] Hiroaki Mitsugashira,[1] Kazuki Degawa,[1] Yoshinobu Kobayashi,[3] Ryo Muto,[3] Kazuaki Tokodai,[1] Toru Furukawa,[2] Michiaki Unno[1] and Takashi Kamei[1]

[1]Department of Surgery, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan
[2]Department of Investigative Pathology, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan
[3]Department of Surgery, National Hospital Organization Mito Medical Center, Higashiibaraki-gun, Ibaraki, Japan

Biliary atresia is an obliterative cholangiopathy of unknown etiology. Hepatic portoenterostomy, in which obliterated extrahepatic bile ducts are resected and bile flow is restored, known as Kasai operation, is performed within 3 months after birth. While this operation enhances long-term survival of patients, the occurrence of primary malignant hepatic tumors has been increasing. We report a case of small intestinal adenocarcinoma arising at the anastomotic site after Kasai operation. A 49-year-old man, who underwent Kasai operation for biliary atresia when he was 2 months old, experienced rapidly progressive jaundice and liver dysfunction. Deceased-donor liver transplantation was performed for liver failure. Macroscopically, there was a white-yellow tumor located at the anastomotic site of hepatic portoenterostomy of the resected liver. Pathological examination revealed a well-differentiated adenocarcinoma with some Paneth cells in the neoplastic lesion. Immunohistochemically, the tumor cells were negative for cytokeratin 7 (CK7) but positive for cytokeratin 20 (CK20) and a homeobox domain-containing transcription factor (CDX2). Mucin expression in tumor cells was negative for mucin 1 (MUC1) and mucin 6 (MUC6) and positive for mucin 2 (MUC2) and mucin 5AC (MUC5AC). The pathological diagnosis was small intestinal adenocarcinoma originating from the jejunum. The patient was discharged 48 days after the operation. The patient had not experienced recurrence at 10 months after the operation. This is the first report of small intestinal adenocarcinoma arising at the anastomotic site after Kasai operation for biliary atresia. Special care should be taken for the patients after Kasai operation with acute progressive jaundice and liver dysfunction because there is a possibility of malignancy in their native liver.

Keywords: biliary atresia; Kasai operation; liver transplantation; malignant tumor; small intestinal adenocarcinoma
Tohoku J. Exp. Med., 2023 December, **261** (4), 267-272.
doi: 10.1620/tjem.2023.J080

Introduction

Biliary atresia (BA) is an idiopathic cholangiopathy that develops in neonates and causes the progressive inflammatory obliteration of extrahepatic bile ducts. Untreated BA will lead to biliary cirrhosis and liver failure in the first or second year of their lives. Hepatic portoenterostomy, known as Kasai operation, is necessary for survival. Kasai operation was first developed for the treatment option of BA in 1950 and significantly improved patients' prognosis. However, after Kasai operation, half of the patients require liver transplantation (LT) before adulthood because of liver failure, hepatopulmonary syndrome, portal hypertension, and repeated cholangitis (Kasahara et al. 2017; Nio 2017; Miyagi et al. 2022b). There are some reports of primary malignant hepatic tumors after Kasai operation. Hepatocellular carcinoma (HCC) is the most common, followed by cholangiocarcinoma and hepatoblastoma (Van

Received July 27, 2023; revised and accepted September 19, 2023; J-STAGE Advance online publication September 28, 2023
Correspondence: Ryusuke Saito, M.D., Ph.D., Department of Surgery, Tohoku University Graduate School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-0872, Japan.
e-mail: ryusuke009009@gmail.com

Wyk et al. 1972; Tamura et al. 1993; Kohno et al. 1995; Tatekawa et al. 2001; Azuhata et al. 2003; Taat et al. 2004; Brunati et al. 2007; Hol et al. 2008; Hadžić et al. 2011; Aggarwal et al. 2012; Kim et al. 2012; Vera et al. 2012; Fukuda et al. 2013; Yoon et al. 2014; Arai et al. 2016; Nio et al. 2019; Uno et al. 2020). Here, we report the first case of small intestinal adenocarcinoma arising at the anastomotic site after Kasai operation for BA.

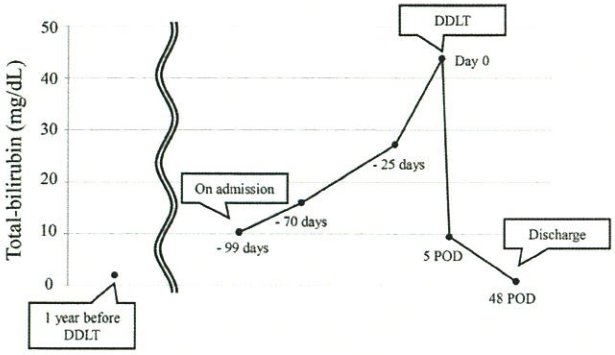


Fig. 1. Preoperative time course changes of the serum level of total bilirubin.
One year before liver transplantation (LT), the serum level of total bilirubin was 2.1 mg/dL. On admission, the serum level of total bilirubin level increased to 10.4 mg/dL and dramatically increased to 43.8 mg/dL just before LT. It immediately decreased to the normal level after the operation. DDLT, deceased-donor liver transplantation; POD, post operative day(s).

Case Presentation

The patient was a man who underwent Kasai operation for BA 2 months after birth. The postoperative course was good, and he only experienced several episodes of cholangitis in his adulthood without admission. When he became 49 years old, the sudden onset of jaundice and the progressive elevation of hepatobiliary enzymes occurred. He was hospitalized with cholangitis, and his jaundice and liver dysfunction progressed rapidly. Serum total bilirubin was 10.4 mg/dL (reference range 0.4-1.5 mg/dL) on admission and dramatically increased up to 43.8 mg/dL (Fig. 1). The elevation of other hepatobiliary enzymes and the decrease of synthetic capacity of the liver was detected. The elevation of carcinoembryonic antigen (CEA) and protein induced by vitamin K absence or antagonist-II (PIVKA-II) was also observed (CEA 7.0 ng/mL, reference range < 5.0 ng/mL; PIVKA-II 12,351 mAU/mL, reference range < 40 mAU/mL).

Enhanced computed tomography (CT) showed the rapid progression of intrahepatic bile duct dilation within two months after the onset of jaundice (Fig. 2A, B), and the abdominal X-ray showed the appearance of massive ascites (Fig. 2C, D). There was no apparent stone or mass detected by CT or magnetic resonance imaging. LT was necessary

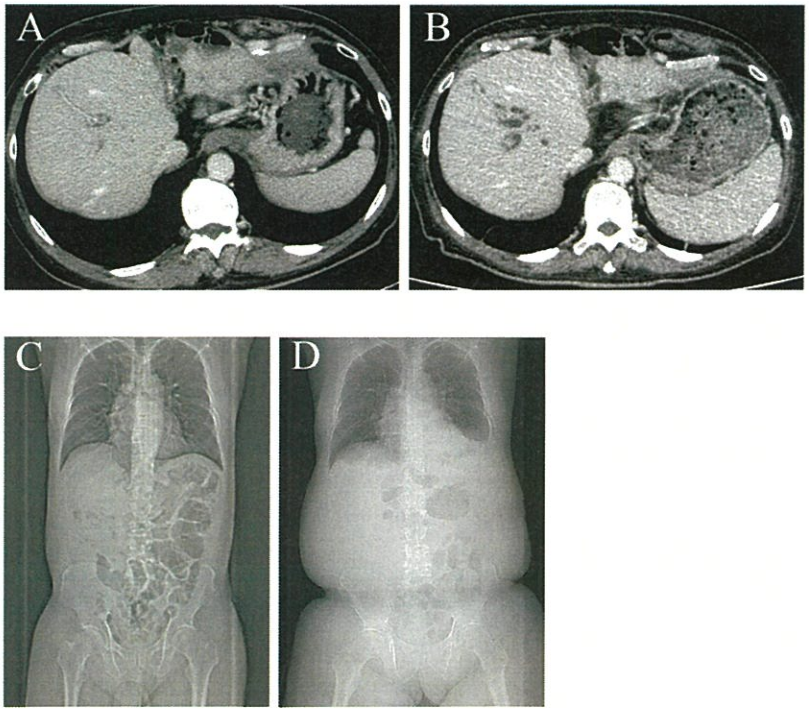


Fig. 2. Preoperative time course changes of imaging findings.
(A, B) Enhanced computed tomography imaging on admission (A) and before liver transplantation (LT) (B). Rapid progression of dilatation of the bile duct was observed. (C, D) Abdominal X-ray on admission (C) and before LT (D). Massive ascites was observed just before LT.

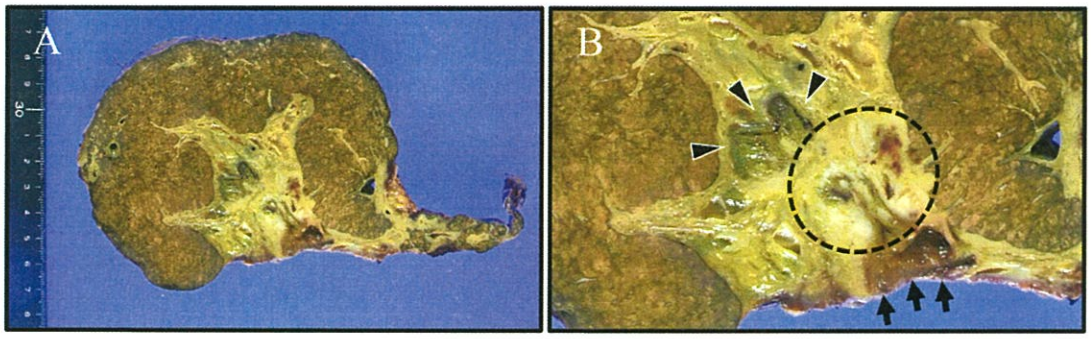


Fig. 3. Macroscopic findings of the resected liver.
(A) The atrophy of the left lobe and biliary cirrhosis was observed. (B) A white-yellow tumor with indistinct boundaries (inside of the dotted line) was located at the anastomotic site of the hepatic portoenterostomy (arrows, jejunal loop; arrowheads, bile duct).

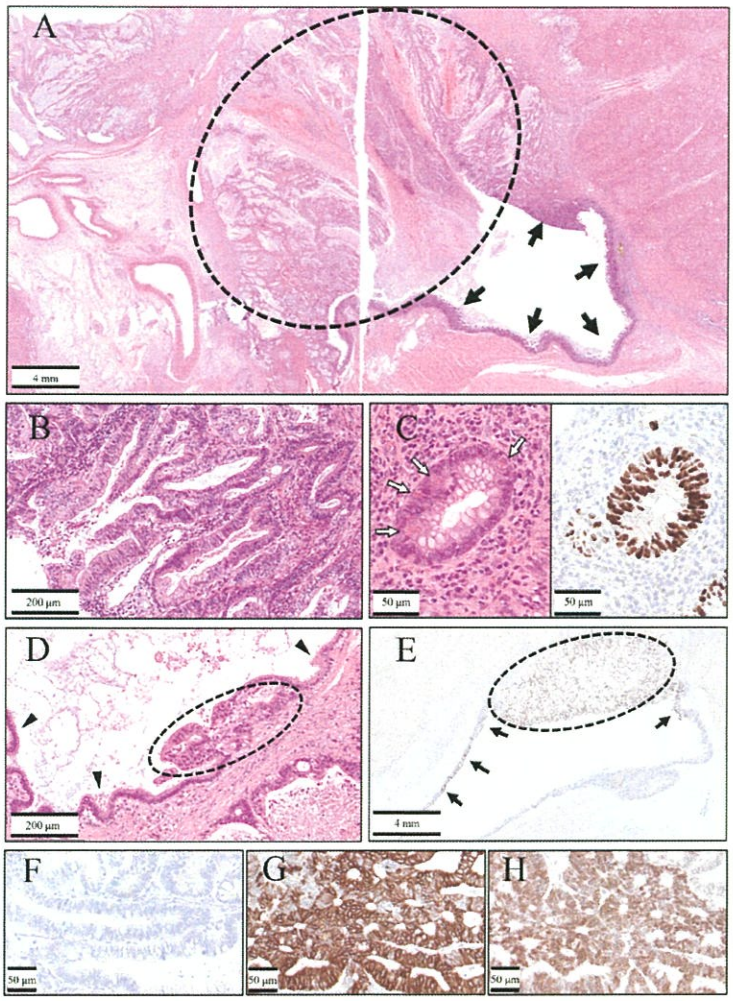


Fig. 4. Pathological findings of the lesion.
(A) Predominant localization of the lesion (inside of the dotted line) at hepatic portoenterostomy (arrows, jejunal loop) revealed by hematoxylin and eosin (H&E) staining. (B) Observation of well-differentiated adenocarcinoma. (C) H&E staining (left side) and p53 immunostaining (right side) of the identical tumor gland obtained from serial sections. Paneth cells in the neoplasm (white arrows). (D) Carcinoma cells (inside of the dotted line) displaced the bile duct epithelium (arrowheads). (E) p53-positive carcinoma in situ (arrows) in the jejunal loop continuously linked to the advanced lesion (inside of the dotted line). (F, G, H) Immunohistochemistry reveals CK7-negative staining in carcinoma cells (F). Diffuse positivity for CK20 (G) and CDX2 (H).

Table 1. Primary malignant tumors after hepatic Kasai operation for biliary atresia.

Tumor type	Small intestinal adenocarcinoma	Hepatocellular carcinoma	Cholangiocarcinoma	Hepatoblastoma
Number of reports	1	17	6 (3 PCCs and 3 ICCs)	2
Age (range)	49	7 (0-39)	23.5 (13-63)	2 and 4
Sex (M:F)	1:0	7:10	2:4	0:2
Elevation of tumor markers (%)	1 (100%)	7 (41%)	3 (50%)	2 (100%)
Diagnosis before operation	None	15	3 (One PCC and 2 ICCs)	2
The number of LT	1	11	4	1
Follow-up period, month (range)	10	18 (2-72)	9.5 (6-180)	Not addressed
Mortality	0	5	4	1
References	The current case	Van Wyk et al. 1972 Tamura et al. 1993 Kohno et al. 1995 Tatekawa et al. 2001 Azuhata et al. 2003 Brunati et al. 2007 Hol et al. 2008 Hadžić et al. 2011 Aggarwal et al. 2012 Kim et al. 2012 Yoon et al. 2014 Arai et al. 2016	Vera et al. 2012 Fukuda et al. 2013 Yoon et al. 2014 Arai et al. 2016 Nio et al. 2019 Uno et al. 2020	Tatekawa et al. 2001 Taat et al. 2004

ICC, intrahepatic cholangiocarcinoma; LT, liver transplantation; PCC, perihilar cholangiocarcinoma.

for lifesaving because of the progressive jaundice and liver failure (The Model for End-Stage Liver Disease score was 30 and Child-Pugh classification was categorized as class C). Deceased-donor LT (DDLT) was performed. The recipient's abdominal cavity was highly adhesive with clear ascites. The recipient's liver was removed, and the liver graft was transplanted. With gross observation, his left lobe of the liver was atrophic, and the cut surface revealed biliary cirrhosis (Fig. 3A). A white-yellow tumor measuring 30 mm in diameter with indistinct boundaries was observed at the anastomotic site of the hepatic portoenterostomy (Fig. 3B).

The pathological examination revealed well-differentiated adenocarcinoma containing Paneth cells, and the tumor was located mainly in the anastomotic site of hepatic portoenterostomy (Fig. 4A-C). On the hepatic side, the tumor invaded the hepatic hilum and the hepatic parenchyma. The tumor cells displaced the bile duct epithelium, while intraductal papillary neoplasm of bile duct was not observed (Fig. 4D). On the jejunal side, the tumor invaded the muscularis propria at its deepest point, and p53-positve carcinoma in situ was also observed continuously linked to the advanced lesion (Fig. 4E). To determine the origin of the tumor, immunohistochemical staining was performed. The tumor cells were diffusely positive for cytokeratin 20 (CK20) and a homeobox domain-containing transcription factor (CDX2), whereas negative for cytokeratin 7 (CK7) (Fig. 4F-H). In addition, the cells were diffusely positive for mucin 2 (MUC2), partially positive for mucin 5AC (MUC5AC), and negative for mucin 1 (MUC1) and mucin 6 (MUC6). From these results, the pathological diagnosis was small intestinal adenocarcinoma originating from the jejunal loop, pT4 cN0 cM0 Stage IIB (UICC-TNM, 8th edition) (Brierley et al. 2017). Although CT-guided drainage was performed for the intra-abdominal abscess, the postoperative course was uneventful and the patient discharged 48 days after the operation. No recurrence has been detected for 10 months after the operation without any adjuvant chemotherapy.

Discussion

BA is an idiopathic cholangiopathy of neonates and early infancy characterized by obstruction of the extrahepatic bile ducts, and its incidence rate is about 1/10,000 live births (Kasahara et al. 2017; Nio 2017). Kasai operation is performed for BA within 3 months, and the earlier the operation is performed, the better the prognosis and native liver function would be (Jimenez-Rivera et al. 2013). As the prognosis of the patients after Kasai operation improves, the occurrence of postoperative primary malignant hepatic tumors has been increasing. Although there are some reports of intrahepatic malignancy after Kasai operation, there is no other report of small intestinal adenocarcinoma originating from the jejunal loop at the anastomotic site.

It is sometimes difficult to determine the origin of the tumor of hepatic hilum. In our case, the tumor was located mainly in the bile duct and caused obstructive jaundice, so cholangiocarcinoma was highly suspected at first (Fig. 3). On pathological examination, the tumor contained Paneth cells, which are observed in the small intestinal tissue. In addition, the tumor continued from the surface of the jejunum to hilum of the liver. Immunohistochemical staining

also showed that the tumor was CDX2-positive, CK7-negative, and CK20-positive staining patterns, which indicates the diagnosis of intestinal tumors (Park et al. 2007). MUC1-negative, MUC2-positive, and MUC5AC-positive expression patterns also align with the characteristics of intestinal tumors (Lau et al. 2004).

Small intestinal adenocarcinoma is a rare malignant tumor, and its incidence rate is reported to be less than 0.001% (Li et al. 2016). Inflammatory bowel diseases such as ulcerative colitis and Crohn's disease have been reported as the risk factors for small intestinal adenocarcinoma (Goodman et al. 2013). In this case, the chronic inflammation at the anastomotic site caused by repeated cholangitis could be associated with developing the tumor after Kasai operation. On the other hand, intrahepatic tumors including HCC and intrahepatic cholangiocarcinoma, could be caused by cholestatic liver cirrhosis (Arai et al. 2016).

In this case, the tumor was coincidentally found by the pathological examination after LT. Retrospectively, there were two preoperative findings suspicious for the presence of the tumor. Firstly, the obstructive jaundice and the dilatation of the intrahepatic duct had progressed rapidly in a short period (Fig. 1). Secondly, the level of CEA, which is a nonspecific tumor marker for small intestinal adenocarcinoma, was elevated (Chen and Vaccaro 2018). There are some reports that balloon-assisted endoscopic retrograde cholangiopancreatography (ERCP) was useful for obstructive jaundice after Roux-en-Y reconstruction (Liu et al. 2017). Although the success rate of ERCP after Kasai operation is low, there are some reports that ERCP was effective in controlling cholangitis after Kasai operation (Liu et al. 2017; Hyun et al. 2018). Therefore, the balloon-assisted ERCP could be one of the options for controlling obstructive jaundice and the diagnosis of the tumor in our case. Table 1 shows the summary of the reports searched by Pubmed with the keywords "Biliary atresia," "Kasai operation" and "Malignant tumor," and the references of those reports.

The tumor markers were also elevated in half of the cases of malignancy after Kasai operation (Table 1). Since primary malignancies after Kasai operation occurred at younger age compared with natural populations, careful follow up should be done using imaging examinations and tumor markers.

LT for patients with malignancy is controversial. The effectiveness of LT for some malignancies, including HCC, hepatoblastoma, and hilar cholangiocarcinoma are reported (Mantel et al. 2016; Shimamura et al. 2019; Miyagi et al. 2022a; Srinivasan et al. 2023). In our case, the preoperative diagnosis was difficult because the imaging tests could not detect any tumor, and it was difficult to distinguish the obstructive jaundice from the jaundice caused by liver failure. The accuracy of preoperative diagnosis of the tumor located in the liver hilum was extremely low after Kasai operation (Table 1). We are using everolimus as an immunosuppressant, which was reported to reduce the risk of recurrence after LT for HCC (ALoun et al. 2023). Postoperative management of small intestinal carcinoma was unclear because of the small number of cases (Li et al. 2016). Further studies are warranted to establish the treatment and predict the prognosis of primary malignant tumors arising after Kasai operation.

In conclusion, we experienced a case of small intestinal adenocarcinoma arising at the anastomotic site after Kasai operation. The number of malignancies after Kasai operation is increasing because of the improving prognosis. Special care should be taken for the patient with rapidly progressive jaundice and the elevation of tumor markers.

Acknowledgments

We would like to thank Editage (https://www.editage.com) for the English language editing.

Conflict of Interest

The authors declare no conflict of interest.

References

ALoun, A., Abu-Zeid, E.E.D. & Garzali, I.U. (2023) Does mtori base immunosuppression offer survival advantage after liver transplantation for hepatocellular carcinoma? Systematic review and meta-analysis of randomized controlled trials. *Hepatol. Forum*, **4**, 82-88.

Aggarwal, S., Vadada, D. & Sharma, V. (2012) A rare complication in an adult patient after Kasai portoenterostomy for biliary atresia. *Arab J. Gastroenterol.*, **13**, 148-149.

Arai, Y., Kubota, M., Kobayashi, T., Ohyama, T. & Yokota, N. (2016) A case of synchronous combined cancer with intrahepatic cholangiocarcinoma and hepatocellular carcinoma 39 years after operation for biliary atresia. *J. Jpn. Soc. Pediatr. Surg.*, **52**, 1303-1308.

Azuhata, T., Koshinaga, T., Tanaka, M., Ikeda, T., Hagiwara, N., Takekawa, M., Nonaka, S., Fukazawa, M., Sugitani, M. & Nemoto, N. (2003) Hepatocellular carcinoma complicating biliary atresia caused by HCV infection: a case report. *J. Jpn. Soc. Pediatr. Surg.*, **39**, 765-770.

Brierley, J.D., Gospodarowicz, M.K. & Wittekind, C. (2017) *TNM Classification ofMalignant Tumors*, 8th ed., John Wiley-Blackwell, Hoboken, NJ, USA.

Brunati, A., Feruzi, Z., Sokal, E., Smets, F., Fervaille, C., Gosseye, S., Clapuyt, P., de Ville de Goyet, J. & Reding, R. (2007) Early occurrence of hepatocellular carcinoma in biliary atresia treated by liver transplantation. *Pediatr. Transplant.*, **11**, 117-119.

Chen, E.Y. & Vaccaro, G.M. (2018) Small bowel adenocarcinoma. *Clin. Colon Rectal Surg.*, **31**, 267-277.

Fukuda, A., Sakamoto, S., Kanazawa, H., Shigeta, T., Karaki, C., Hamano, I., Uchida, H., Kitagawa, H., Okuse, C., Miyazaki, O., Nosaka, S., Nakazawa, A. & Kasahara, M. (2013) Incidentally detected cholangiocarcinoma in an explanted liver with biliary atresia after Kasai operation. *Pediatr. Transplant.*, **17**, E62-66.

Goodman, M.T., Matsuno, R.K. & Shvetsov, Y.B. (2013) Racial and ethnic variation in the incidence of small-bowel cancer subtypes in the United States, 1995-2008. *Dis. Colon Rectum*, **56**, 441-448.

Hadžić, N., Quaglia, A., Portmann, B., Paramalingam, S., Heaton, N.D., Rela, M., Mieli-Vergani, G. & Davenport, M. (2011) Hepatocellular carcinoma in biliary atresia: King's College Hospital experience. *J. Pediatr.*, **159**, 617-622. e611.

Hol, L., van den Bos, I.C., Hussain, S.M., Zondervan, P.E. & de

Man, R.A. (2008) Hepatocellular carcinoma complicating biliary atresia after Kasai portoenterostomy. *Eur. J. Gastroenterol. Hepatol.*, **20**, 227-231.

Hyun, J.J., Irani, S.S. & Kozarek, R.A. (2018) Endoscopic retrograde cholangiopancreatography in adult patients with biliary atresia: PROCESS-compliant case series. *Medicine (Baltimore)*, **97**, e0603.

Jimenez-Rivera, C., Jolin-Dahel, K.S., Fortinsky, K.J., Gozdyra, P. & Benchimol, E.I. (2013) International incidence and outcomes of biliary atresia. *J. Pediatr. Gastroenterol. Nutr.*, **56**, 344-354.

Kasahara, M., Umeshita, K., Sakamoto, S., Fukuda, A., Furukawa, H. & Uemoto, S. (2017) Liver transplantation for biliary atresia: a systematic review. *Pediatr. Surg. Int.*, **33**, 1289-1295.

Kim, J.M., Lee, S.K., Kwon, C.H., Joh, J.W., Choe, Y.H. & Park, C.K. (2012) Hepatocellular carcinoma in an infant with biliary atresia younger than 1 year. *J. Pediatr. Surg.*, **47**, 819-821.

Kohno, M., Kitatani, H., Wada, H., Kajimoto, T., Matuno, H., Tanino, M., Nakagawa, T. & Takarada, A. (1995) Hepatocellular carcinoma complicating biliary cirrhosis caused by biliary atresia: report of a case. *J. Pediatr. Surg.*, **30**, 1713-1716.

Lau, S.K., Weiss, L.M. & Chu, P.G. (2004) Differential expression of MUC1, MUC2, and MUC5AC in carcinomas of various sites: an immunohistochemical study. *Am. J. Clin. Pathol.*, **122**, 61-69.

Li, J., Wang, Z., Liu, N., Hao, J. & Xu, X. (2016) Small bowel adenocarcinoma of the jejunum: a case report and literature review. *World J. Surg. Oncol.*, **14**, 177.

Liu, K., Joshi, V., Saxena, P. & Kaffes, A.J. (2017) Predictors of success for double balloon-assisted endoscopic retrograde cholangiopancreatography in patients with Roux-en-Y anastomosis. *Dig. Endosc.*, **29**, 190-197.

Mantel, H.T., Westerkamp, A.C., Adam, R., Bennet, W.F., Seehofer, D., Settmacher, U., Sanchez-Bueno, F., Fabregat Prous, J., Boleslawski, E., Friman, S. & Porte, R.J.; European Liver and Intestine Transplant Association (ELITA) (2016) Strict selection alone of patients undergoing liver transplantation for hilar cholangiocarcinoma is associated with improved survival. *PLoS One*, **11**, e0156127.

Miyagi, S., Fujio, A., Nakagawa, K., Tokodai, K., Kashiwadate, T., Miyazawa, K., Sasaki, K., Matsumura, M., Nishimaki, H., Hibi, T., Kamei, T. & Unno, M. (2022a) A new strategy of liver transplantation for locally advanced unresectable perihilar cholangiocarcinoma using living grafts with simultaneous resection of recipients' hepatic artery and portal vein without neoadjuvant radiation: a case report. *Transplant. Proc.*, **54**, 1643-1647.

Miyagi, S., Sasaki, K., Tokodai, K., Fujio, A., Kashiwadate, T., Miyazawa, K., Matsumura, M., Kamei, T. & Unno, M. (2022b) Adult living donor liver transplantation for the carry-over patients after Kasai operation who have a history of multiple laparotomies. *Transplant. Proc.*, **54**, 430-434.

Nio, M. (2017) Japanese biliary atresia registry. *Pediatr. Surg. Int.*, **33**, 1319-1325.

Nio, M., Wada, M., Sasaki, H., Tanaka, H., Hashimoto, M. & Nakajima, Y. (2019) Correctable biliary atresia and cholangiocarcinoma: a case report of a 63-year-old patient. *Surg. Case Rep.*, **5**, 185.

Park, S.Y., Kim, B.H., Kim, J.H., Lee, S. & Kang, G.H. (2007) Panels of immunohistochemical markers help determine primary sites of metastatic adenocarcinoma. *Arch. Pathol. Lab. Med.*, **131**, 1561-1567.

Shimamura, T., Akamatsu, N., Fujiyoshi, M., Kawaguchi, A., Morita, S., Kawasaki, S., Uemoto, S., Kokudo, N., Hasegawa, K., Ohdan, H., Egawa, H., Furukawa, H. & Todo, S.; Japanese Liver Transplantation Society (2019) Expanded living-donor liver transplantation criteria for patients with hepatocellular carcinoma based on the Japanese nationwide survey: the 5-5-500 rule - a retrospective study. *Transpl. Int.*, **32**, 356-368.

Srinivasan, S., Prasad, M., Parambil, B.C., Shrimal, A., Gollamudi, V.R.M., Subramani, V., Ramadwar, M., Khanna, N., Baheti, A.D., Gala, K., Patil, V., Laskar, S., Qureshi, S. & Chinnaswamy, G. (2023) Treatment outcomes and prognostic factors in children with hepatoblastoma using a risk-stratified approach. *Pediatr. Blood Cancer*, **70**, e30302.

Taat, F., Bosman, D.K. & Aronson, D.C. (2004) Hepatoblastoma in a girl with biliary atresia: coincidence or co-incidence. *Pediatr. Blood Cancer*, **43**, 603-605.

Tamura, M., Hata, Y., Sasaki, F., Takahashi, H., Hamada,H., Uchino, J., Nakazawa, O., Mima, S., Anen, K. & Fujioka, Y. (1993) A case report of biliary atresia associated with hepatocellular carcinoma. *J. Jpn. Soc. Pediatr. Surg.*, **29**, 131-135.

Tatekawa, Y., Asonuma, K., Uemoto, S., Inomata, Y. & Tanaka, K. (2001) Liver transplantation for biliary atresia associated with malignant hepatic tumors. *J. Pediatr. Surg.*, **36**, 436-439.

Uno, S., Kataoka, T.R., Okajima, H., Taura, K., Sakurai, T. & Haga, H. (2020) Perihilar cholangiocarcinoma in an explanted liver after Kasai operation for biliary atresia: a case report and literature review. *Pathol. Int.*, **70**, 888-892.

Van Wyk, J., Halgrimson, C.G., Giles, G., Lilly, J., Martineau, G. & Starzl, T.E. (1972) Liver transplantation in biliary atresia with concomitant hepatoma. *S. Afr. Med. J.*, **46**, 885-889.

Vera, A., Villaveces, D. & Lopez, R. (2012) Orthotopic liver transplantation for biliary atresia complicated by incidental cholangiocarcinoma. *J. Pediatr. Gastroenterol. Nutr.*, **55**, 336-337.

Yoon, H.J., Jeon, T.Y., Yoo, S.Y., Kim, J.H., Eo, H., Lee, S.K. & Kim, J.S. (2014) Hepatic tumours in children with biliary atresia: single-centre experience in 13 cases and review of the literature. *Clin. Radiol.*, **69**, e113-119.