

P-04 文献データベースにおける 医学ジャーナルの収録状況

～ LANCET, NEJM, JAMA の
MEDLINE, Embase, Scopus, Web of Science,
Caplus, Dimensions, JSTPlusでの収録を見る ～

廣谷映子

一般社団法人 情報科学技術協会OUG ライフサイエンス分
科会 (<https://www.infosta.or.jp/research/oug-life/>)



背景・目的

■背景

- 文献データベース（以下、文献DB）には収録対象などの特徴があり、調査の際には目的に合わせて文献DBを選択する必要がある
- 特徴の1つに収録対象誌があるが、実際に収録対象誌がどの程度収録されているかを確認したことがなかった

■目的

文献DBの収録対象誌のうち特定のジャーナルについて調査・検証して今後の活用に繋げる

調査対象

- 調査対象の文献DB （ ）は提供システム
MEDLINE、Embase、Lancet Titles、
New England Journal of Medicine （以上 Dialogを使用）
Scopus
Web of Science
Dimensions
CAplus（STNを使用）
JSTPlus（JDreamⅢを使用）
J-GLOBAL

- 調査したジャーナル
LANCET、New England Journal of Medicine、JAMA

調査内容

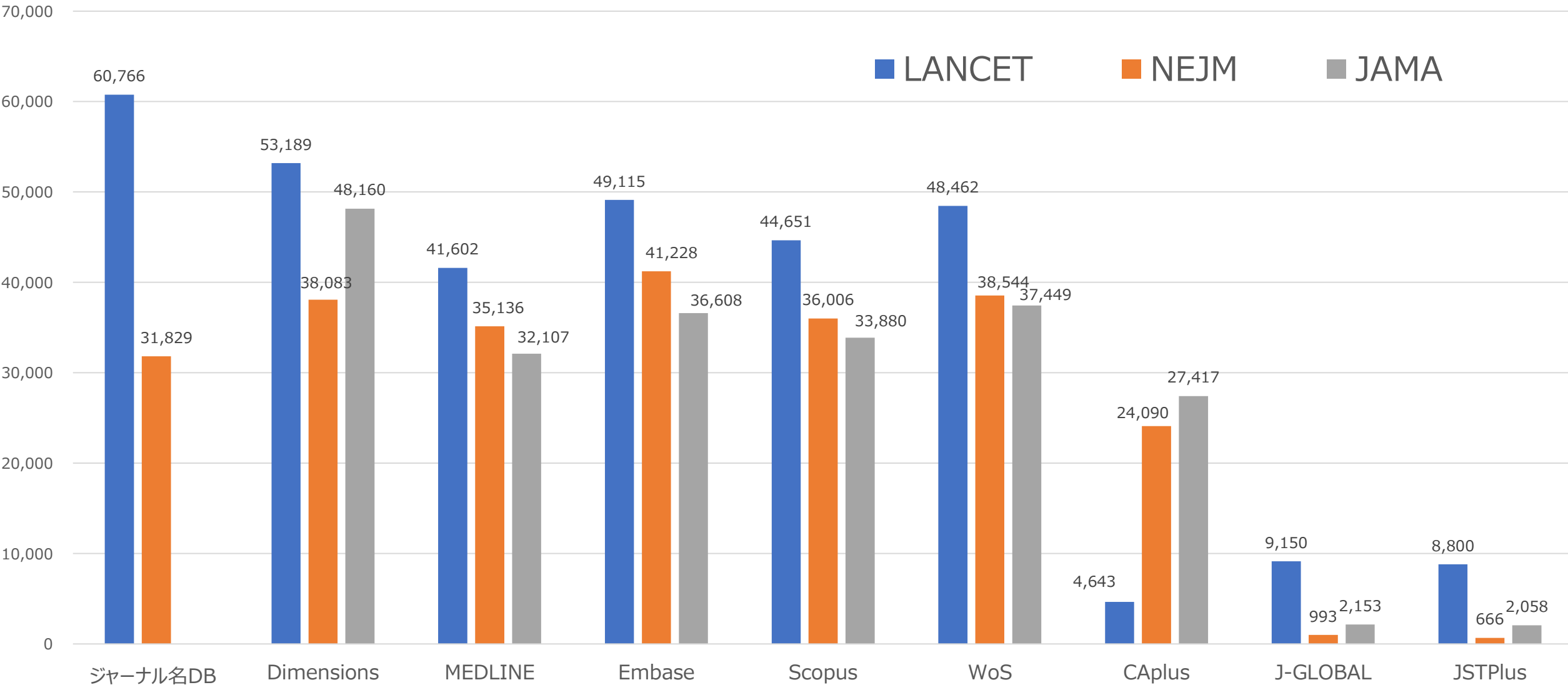
■検索内容

- ① **予備調査**（2024年2月～5月にかけて実施）
出版年を2000～2024年、pISSNとeISSNで検索
- ② **検証のための調査**（2024年5月20日、27日、6月19日に実施）
出版年を2022年、ジャーナルをLANCETに限定して各DBで検索

■検証内容

- ①にて各DBにおける3ジャーナルの収録状況を確認する
- ②の検索結果について重複状況やデータの中身を調査し検証する

①-1 各DBのジャーナル収録状況 2020~2024年

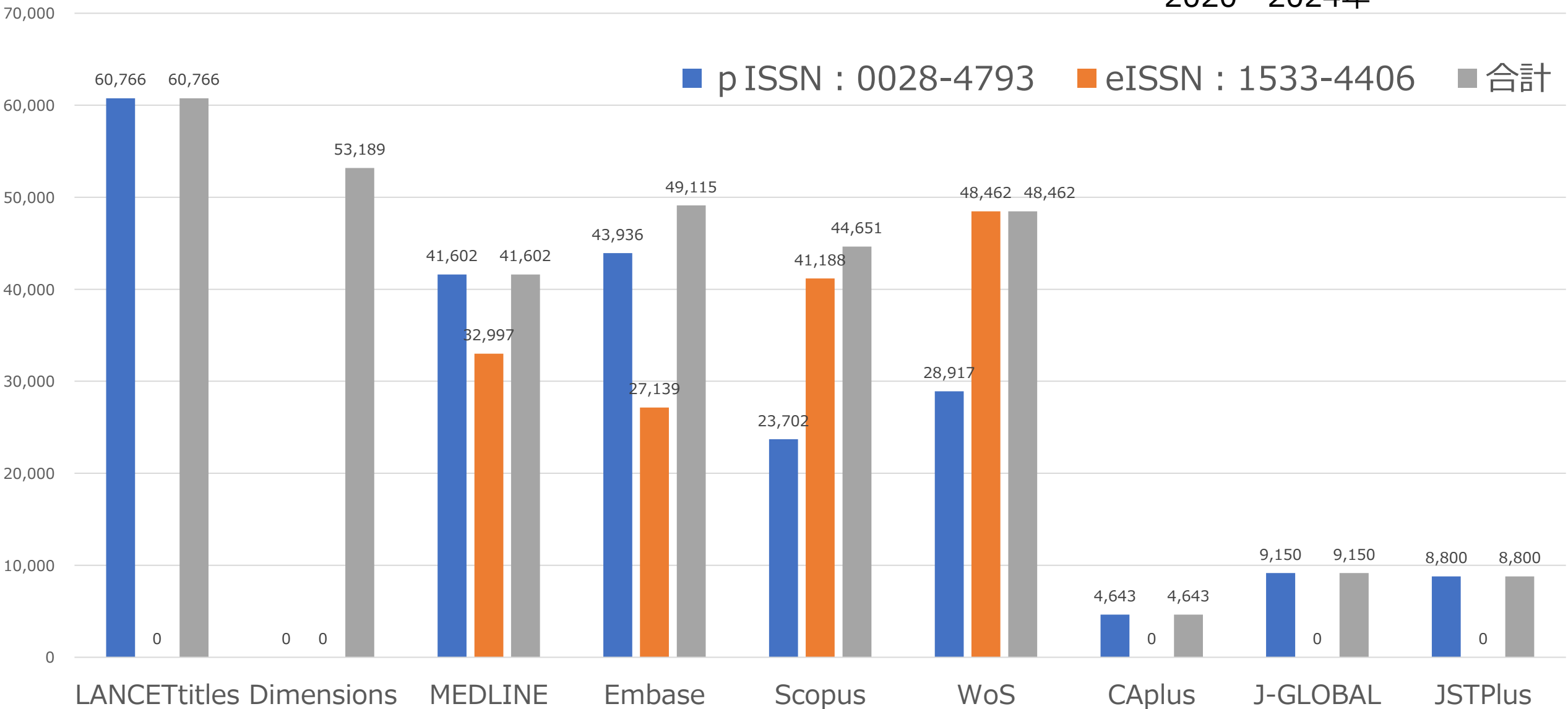


ジャーナル名DB
LANCET / NEJM

※JAMAはジャーナル名のDBがない

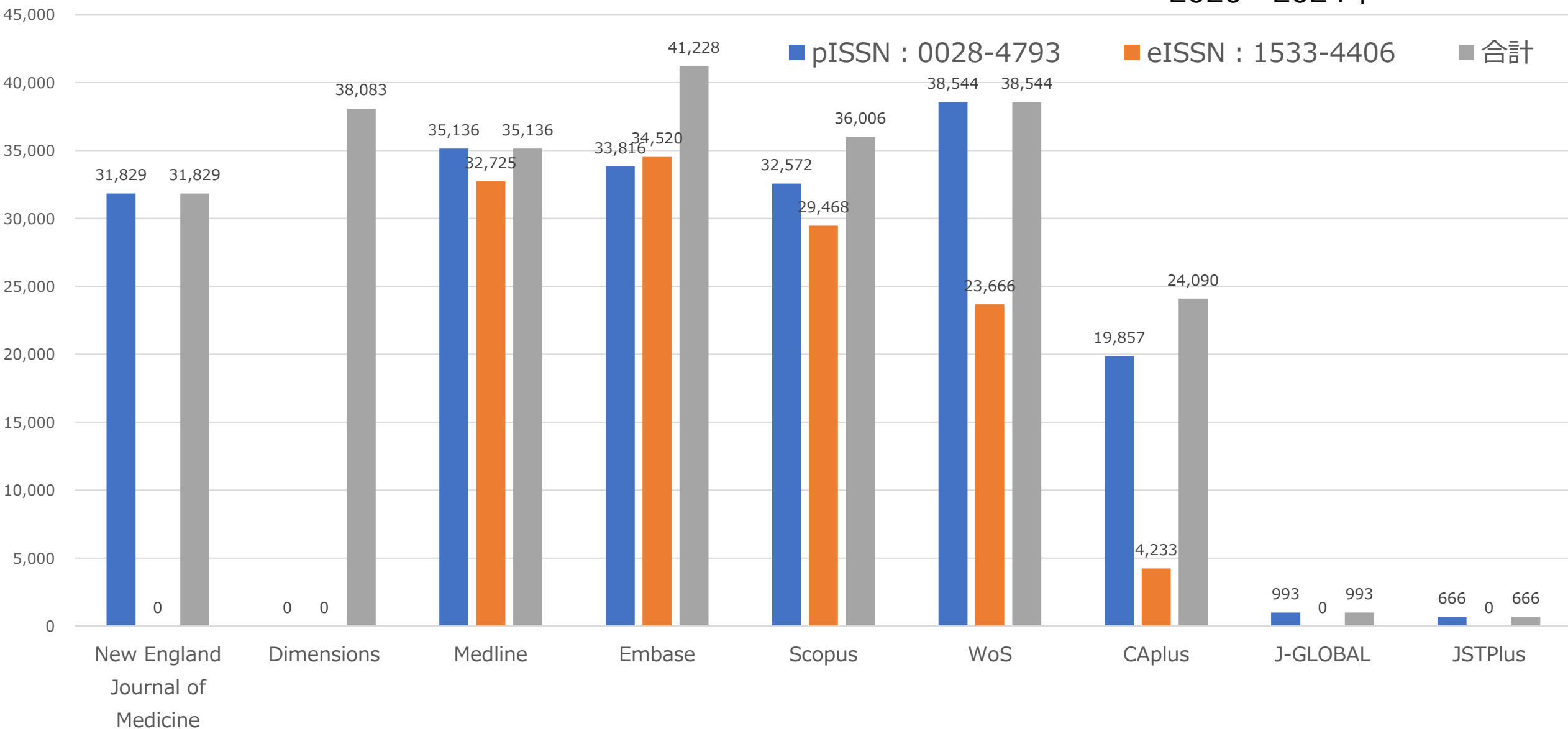
①-2 LANCETのISSN別の検索結果件数

2020～2024年



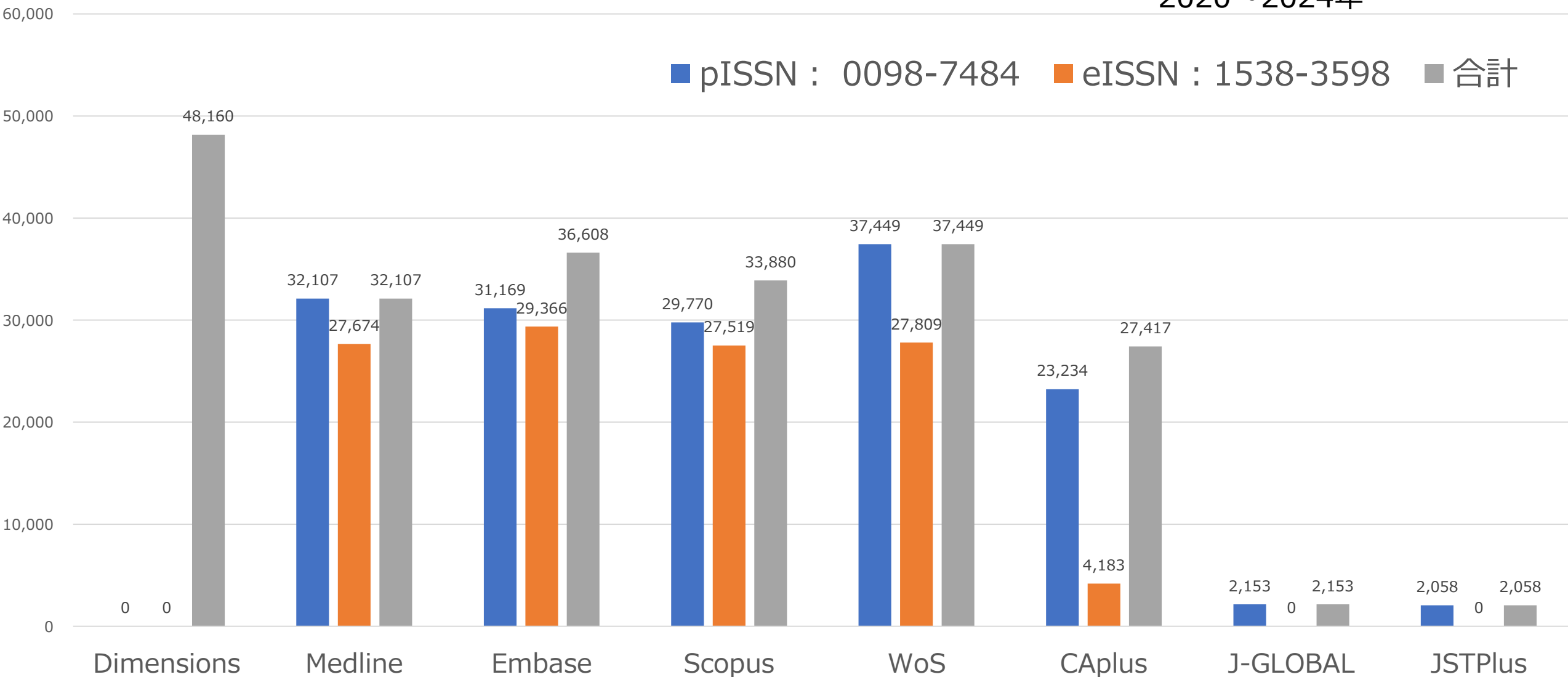
①-2 NEJMのISSN別の検索結果件数

2020～2024年



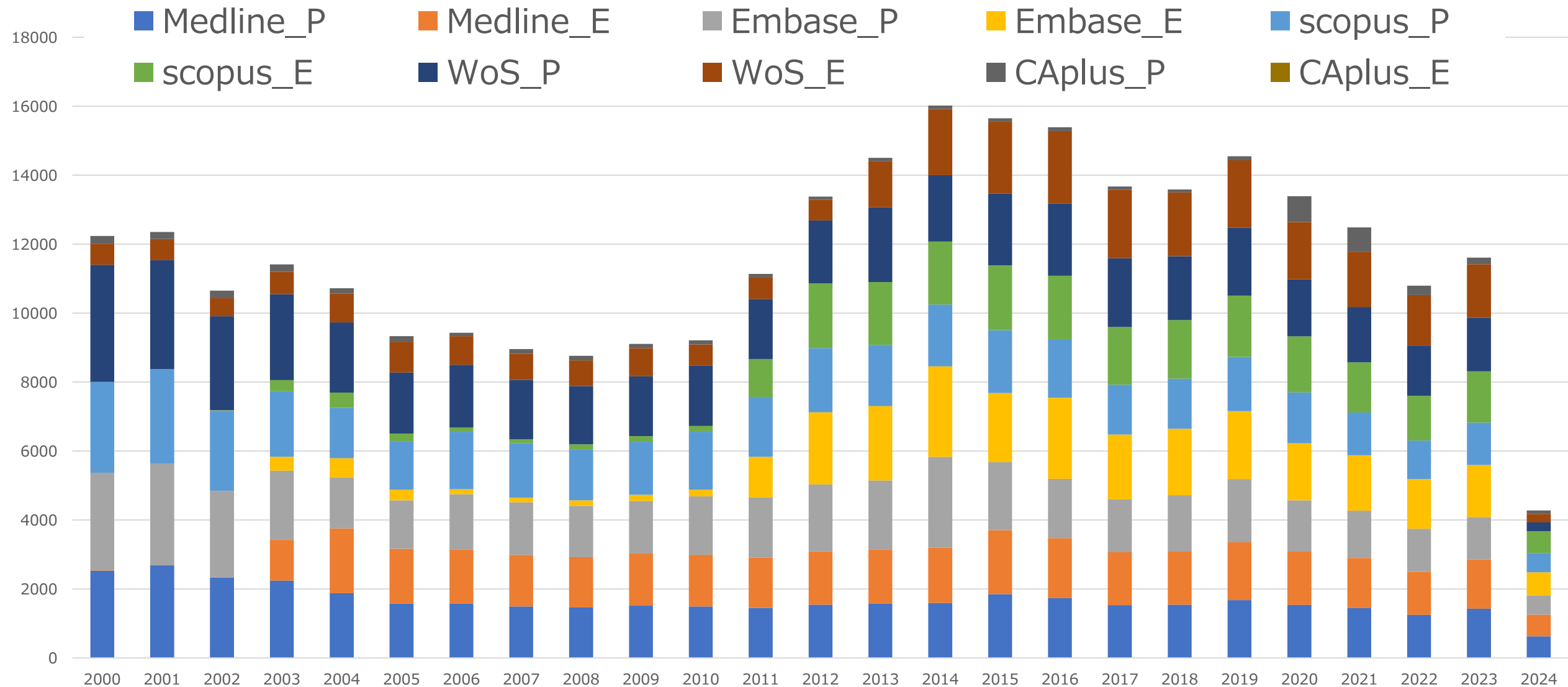
①-2 JAMAのISSN別の検索結果件数

2020～2024年



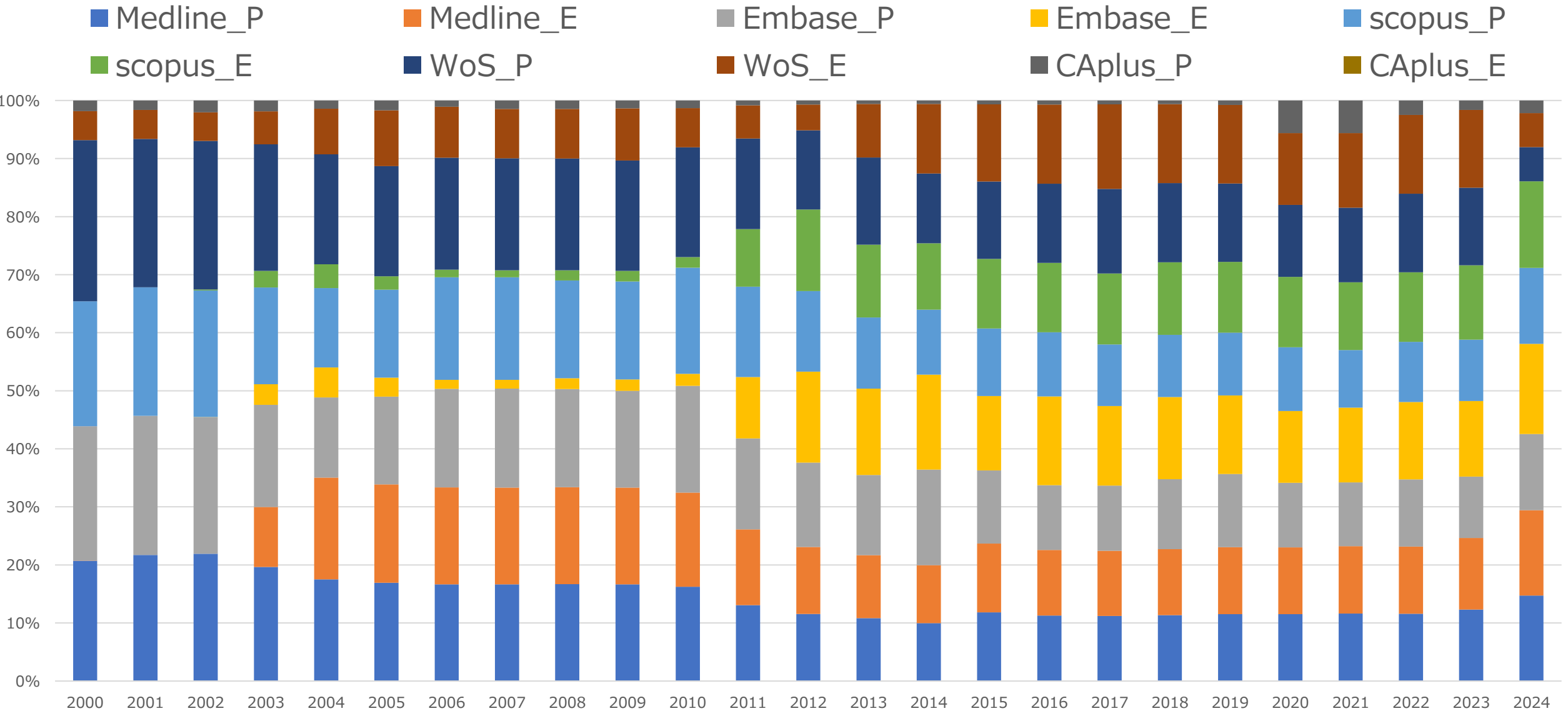
※JAMAはジャーナル名のDBがない

①-3 LANCET (p-/e-) ISSN別検索結果件数の年次推移

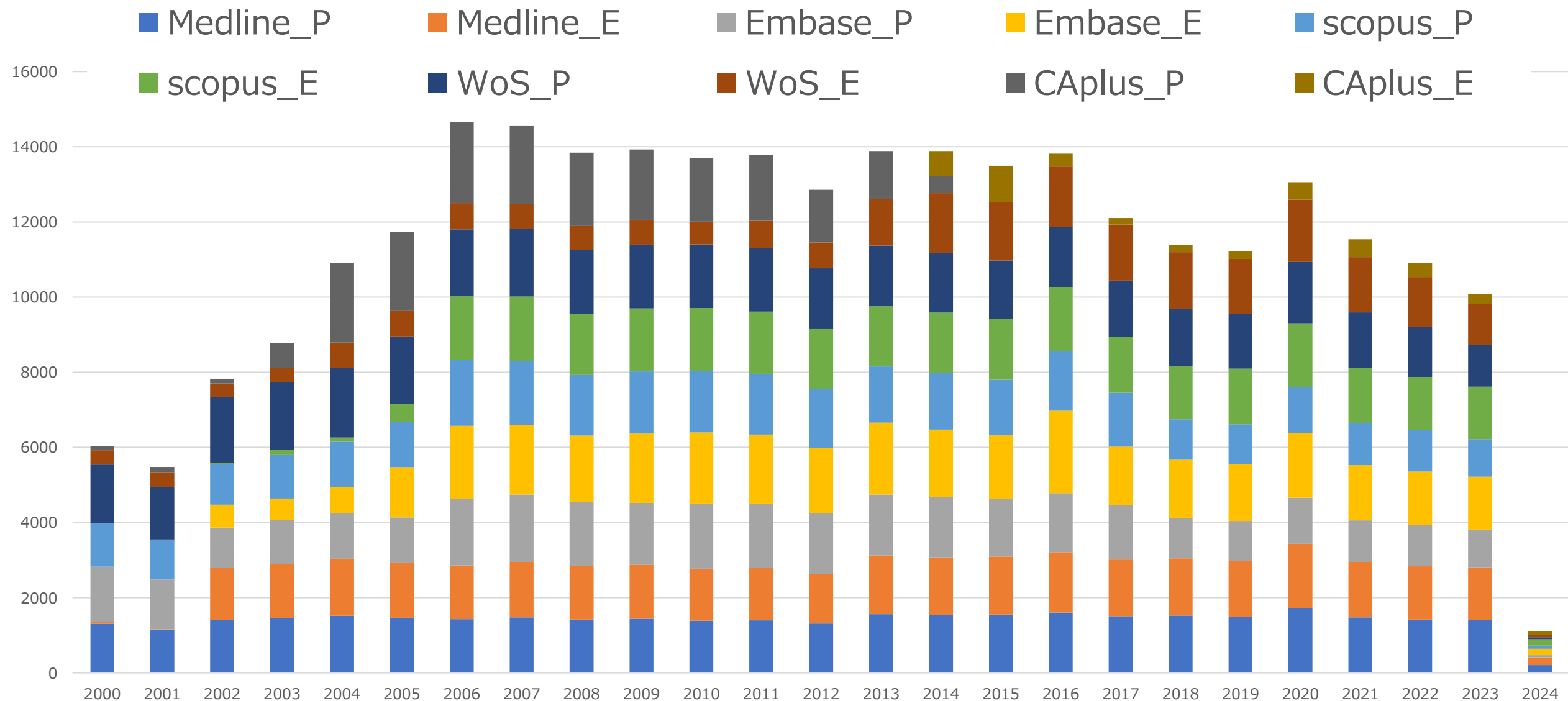


①-3 LANCET

(p-/e-) ISSN別検索結果件数の年次推移 (%)



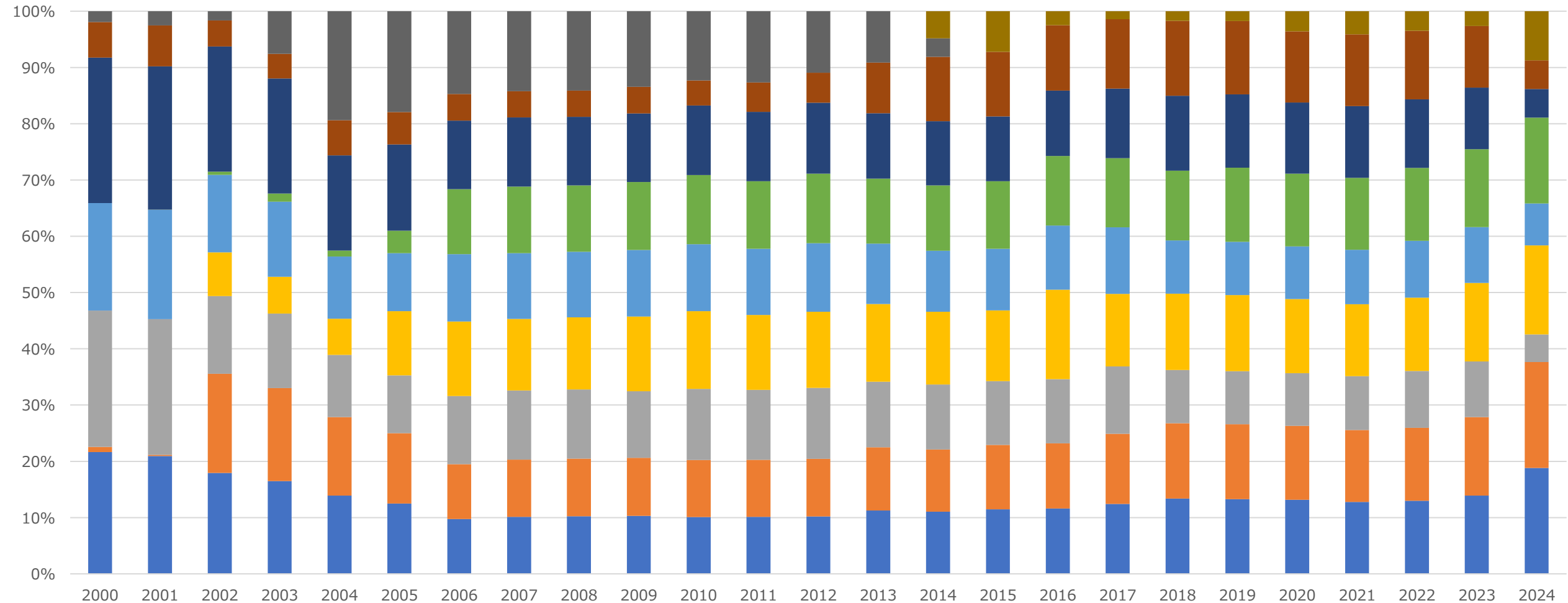
①-3 NEJM (p-/e-) ISSN別検索結果件数の年次推移



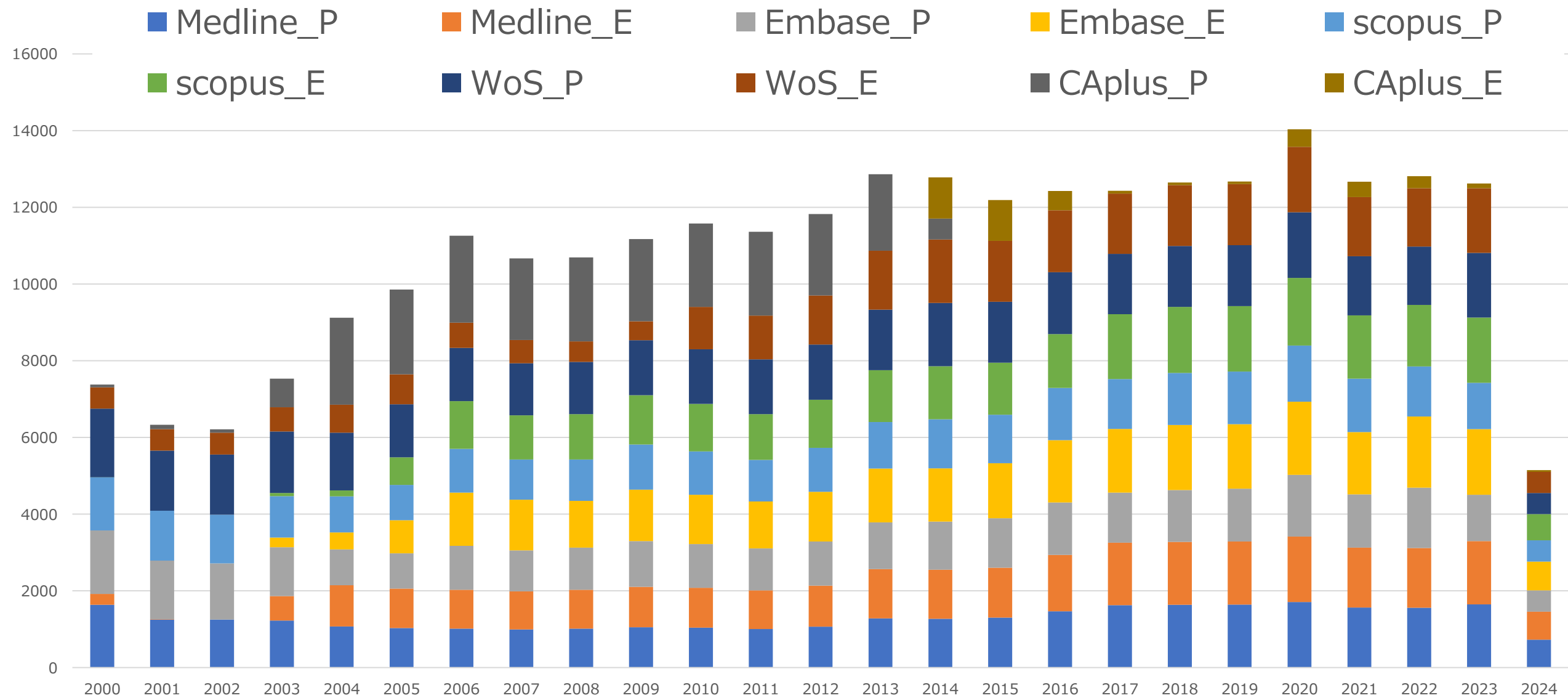
①-3 NEJM

(p-/e-) ISSN別検索結果件数の年次推移 (%)

Medline_P Medline_E Embase_P Embase_E scopus_P
scopus_E WoS_P WoS_E CAplus_P CAplus_E



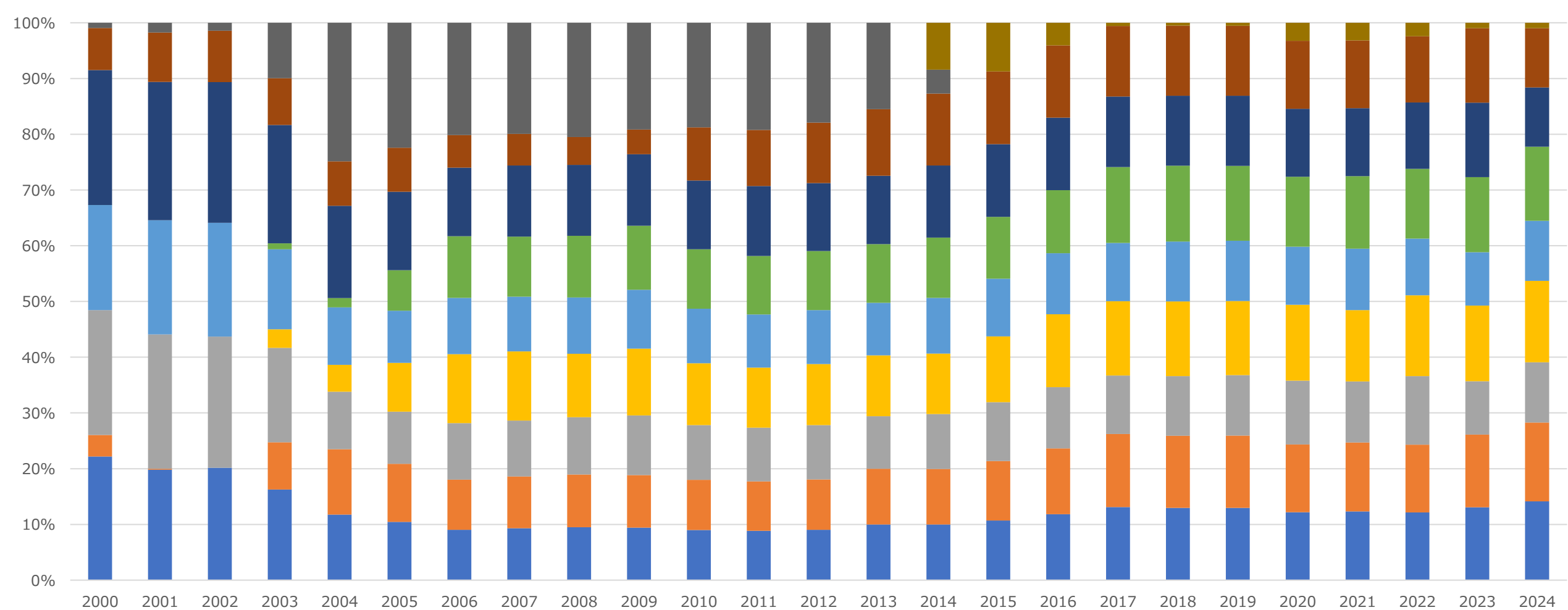
①-3 JAMA (p-/e-) ISSN別検索結果件数の年次推移



①-3 JAMA

(p-/e-) ISSN別検索結果件数の年次推移 (%)

Medline_P Medline_E Embase_P Embase_E scopus_P
scopus_E WoS_P WoS_E CAplus_P CAplus_E



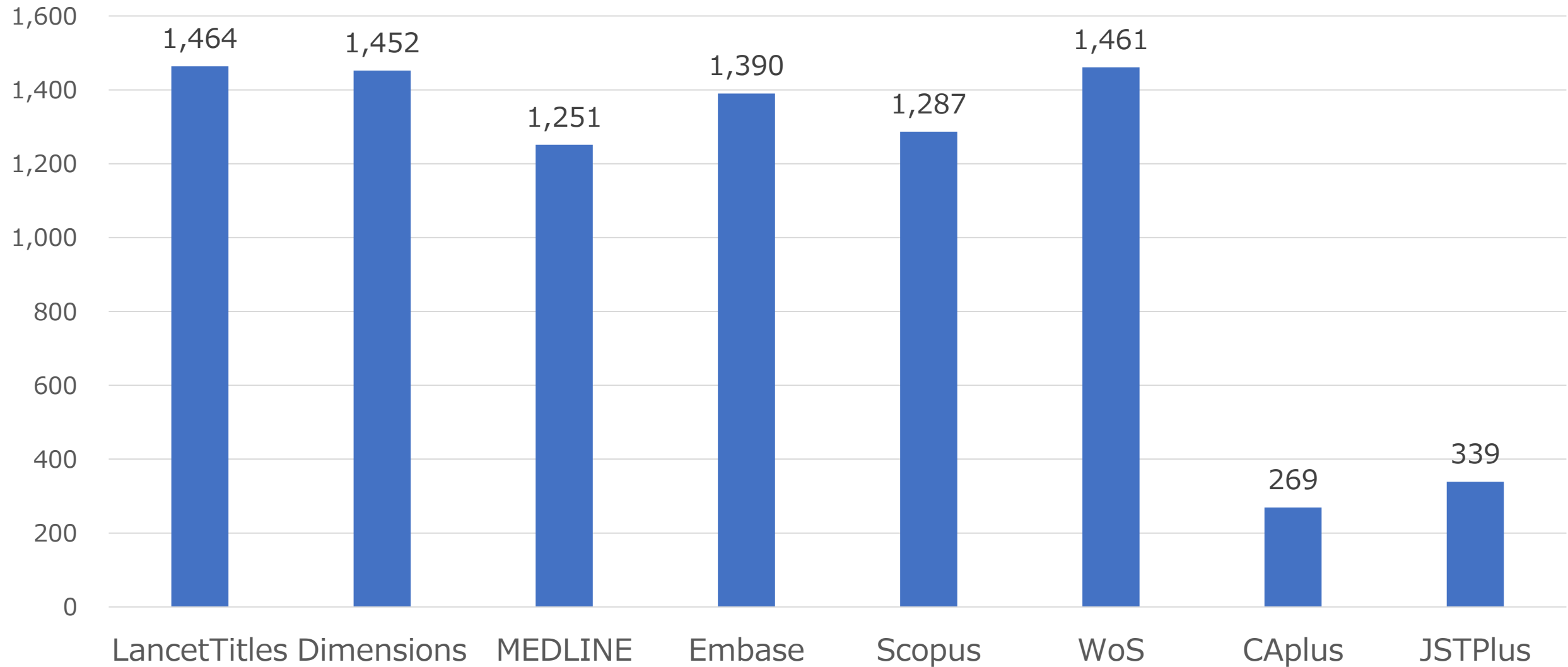
①-3 pISSNとeISSNでの検索結果

	LANCET	NEJM	JAMA
MEDLINE	2004年以降 p-とe-が一致	2003年以降 p-とe-が一致	2004年以降 p-とe-が一致
Embase	2012年以降 e-ISSNが多い	2005年以降 e-ISSNが多い	2006年以降 e-ISSNが多い
Scopus	2012年以降 e-ISSNが多い	2007年以降 e-ISSNが多い	2006年以降 e-ISSNが多い
WoS	2014年以降p-ISSNとe-ISSNの検索結果が一致		
CAplus	p-ISSNのみで 検索される	2013年まではp-ISSNのみ、 2014年はp-e-両方、 2015年以降e-のみで検索される	

結果① ジャーナル収録状況 2000～2024年

- LANCETはジャーナル名のDBの件数が最多だったが、NEJMはジャーナル名のDBが最多ではなかった。ジャーナル名DBのNEJMには除外記事があり収録方針によるものと思われた。
- CAplus、J-GLOBAL、JSTPlusは収録方針通り結果件数は少なかった。
- CAplusは3誌のうちJAMAとNEJMを多く収録していた。
- 全体的にDimensions、Embase、WoS、Scopus、MEDLINEの順に収録データが多い傾向が見られた。
- ジャーナルやDBごとにISSN（p-またはe-）の検索結果に違いが見られた。
- LANCETに関してはジャーナル名DBのLANCET Titlesの件数が最も多く、ジャーナル記事を網羅的に収録していると考えられたことからLANCETを取り上げて詳細を検証することとした。

②-1 LANCET 出版年2022年の検索結果件数



②-2 LANCET 2022年の検索結果の内訳

DB	ISSN	件数	実行日	計	備考
LancetTitles	pISSN	1,464	2024/06/19	1,464	(pISSNのみヒットしeISSNはヒットしない)
	eISSN	-	-		
Dimensions	-	1,452	2024/05/16	1,452	
MEDLINE	pISSN	1,251	2024/06/19	1,251	pISSNとeISSNは同じ結果
	eISSN	1,251	2024/06/19		
Embase	pISSN	1,247	2024/06/19	1,390	eISSNのみ 143件 (eの1,439件中49件が重複、pにもあり)
	eISSN	1,439	2024/06/19		
Scopus	pISSN	1,119	2024/05/20	1,287	eISSNのみ 168件 (eの1,293件中6件が重複、pにもあり)
	eISSN	1,293	2024/05/20		
WoS	pISSN	1,463	2024/05/20	1,461	pISSNで2件重複、eISSNで2件重複 doi無し1,030件、doiのPrefixの誤り：5件
	eISSN	1,463	2024/05/20		
CAplus	pISSN	269	2024/05/27	269	化学を重点的に収録される
	eISSN	0	2024/05/27		
JSTPlus/ JGLOBAL	pISSN	339	2024/05/13	339	著者抄録があるものののみ収録される
	eISSN	-	-		

②-3 LANCET 2022年の収録状況

収録状況	件数	収録DB と 論文タイプ (() はデータの件数)
①1つのDBで収録	45	Dimensions(45) : 401巻(43)402巻(2)で出版年は2023年
②2つのDBで収録	3	Dimensions/MEDLINE(3) : <i>Department of Error(2) Letter(1)</i>
③3つのDBで収録	179	Lancet/Dimensions(179) Embase(104) WoS(74) : <i>Conference Abstract(104)</i> はEmbaseで全件収録。論文タイプなし(73)
④4つのDBで収録	22	Lancet(22) Embase/Scopus(20) Dimensions/MEDLINE(12) WoS(2): <i>Note(16) Comment(7) Journal Article(6)</i> など
⑤5つのDBで収録	132	Embase/Scopus(132) LANCET(129) MEDLINE(101) WoS(85) Dimensions(79) : <i>Journal Article(87) Note(67)</i> など
⑥全DBで収録	1,134	<i>Journal Article(712) Note(362) Letter(255) Editorial(80)</i> など
全レコード	1,515	

照合はdoiにより行った。なおWoSでDOI無し／誤りのレコードは、巻号ページ情報を使用して照合した。
 ※JSTPlusとCAplusは収録方針によりデータが限定されるため本確認では除いた。

②-3 LANCET 2022年の各DB収録状況

LancetTitles	Dimensions	MEDLINE	Embase	Scopus	WoS	件数
	①					45
	②	②				3
③	③				③	74
③	③			③		1
③	③		③			104
④	④	④			④	2
④		④	④	④		10
④	④		④	④		10
★ 1	⑤	⑤	⑤	⑤	⑤	3
⑤	⑤	★ 3	⑤	⑤	⑤	30
⑤	⑤	⑤	⑤	⑤	★ 4	46
⑤	★ 2	⑤	⑤	⑤	⑤	53
⑥	⑥	⑥	⑥	⑥	⑥	1,134

①1つのDBで収録	45
②2つのDBで収録	3
③3つのDBで収録	179
④4つのDBで収録	22
⑤5つのDBで収録	132
⑥全DBで収録	1,134
合計	1,515

LancetTitles	Dimensions	MEDLINE	Embase	Scopus	WoS	件数
★ 1	⑤	⑤	⑤	⑤	⑤	3
⑤	⑤	★ 3	⑤	⑤	⑤	30
⑤	⑤	⑤	⑤	⑤	★ 4	46
⑤	★ 2	⑤	⑤	⑤	⑤	53

- ★ 1 LANCETに無かった3件：書誌に間違いはなかったが収録されておらず理由は不明だった。
- ★ 2 Dimensionsに無かった53件：すべて399巻。LANCET399巻の出版年は2022年だがDimensionsでは早期公開の年を出版年としている（Crossrefをデータソースとしている）ためヒットしなかった。
- ★ 3 MEDLINEに無かった30件：大半が記事タイプNoteだった。他のDB上でPMIDがあるが、MEDLINE上に収録が無いレコードが1件あった。（PubMed上にもなし）。
- ★ 4 WoSに無かった46件：記事タイプに傾向はみられず、PMIDは全件で付与されていた。

②の検証結果

- pISSNとeISSNの検索結果はDBごとに重複状況が異なった
- WoSはdoiの無いデータが多く、また、doiの誤りもあった
- LANCETに無いデータが他DBに収録されている場合のあることを確認した
- Dimensionsは出版年データが他と異なるデータがあった
- DB収録に特徴のある記事タイプが存在した
 - Conference Abstractは LANCET、Dimensions、Embaseに収録、MEDLINEとScopusは限定的な収録、WoSは0件だった
 - MEDLINEのみに収録の無いデータはNoteが多かった
- PMIDがあるがMEDLINEで検索されないデータを確認した

考察

- 収録誌とされているジャーナルの各DB収録の違いを確認した
- ジャーナルDBのLANCETで必ずしも全データが収録されているものではないことを確認した
- 論文タイプでの収録のバラつき、出版年と巻号データの不一致、PMID付与データのMEDLINE収録なし、などの状況を確認した
- 調査時には収録誌の確認とともに、その収録方針にも留意が必要である
- 検索結果は検索条件とともに、使用したDBとシステム名および検索日時を記録し保管することが望ましい

今後に向けて

- DBの収録データは収録対象誌から予測できるが想定外の未収録の可能性もあることに留意したい
- そのため複数DBを参照して検索結果を確認することが望ましい
- 今回の結果には示さなかったが、重複除去に関して、DBシステムによってはタイトルや資料名をキーに除去されることを確認している
(例：タイトル“Department of Error”がほぼ除去される、など)
重複除去はdoiなどのIDを活用して必要以上の除去は無くすべきもの
と考える
- 収録データと出力結果の関係性について今回の検証では検索条件と出力結果とを照合して精査できたが、今後、AIベースの検索では検証も困難と思われ、検索プロセスの開示が望ましいと考える

収録方針が限定的なDBについて

- JSTPlus、J-GLOBALは内容が同一で、海外文献の収録対象を記事の種類（原著など）で選定した後に著者抄録のあるもののみが収録されるため収録が限定される
- CAplusは化学を中心に収録され、臨床医学文献は収録されない特徴がある

謝辞：

今回の検証にあたり、データベース作成・提供機関のJST様、ジー・サーチ様、DigitalScience様など関係の皆様にご協力をいただきましたことを感謝申し上げます。

本発表について：

本ポスター発表はINFOPRO2024の報告※で未発表の調査結果を加えたものである。

※情報科学技術協会INFOSTA. “OUG ライフサイエンス分科会 発表資料／公開資料”. https://www.infosta.or.jp/wp/wp-content/uploads/2024/08/INFOPRO2024_OUG_LIFE_2.pdf, (参照 2025-07-25)