JAIST Repository
https://dspace.jaist.ac.jp/

Title	トピックモデル分析による深層学習を用いた橋渡し研究の動向
Author(s)	岩城，洋子；AKDEMIR，Arda；黒河，昭雄；菊地，乃依瑠；隅蔵，康一
Citation	年次学術大会講演要旨集，35：727-732
Issue Date	2020-10-31
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/17383
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

２Ｆ０８

トピックモデル分析による深層学習を用いた橋渡し研究の動向

○岩城洋子（政策研究大学院大学），Arda AKDEMIR（東京大学），
黒河昭雄（神奈川県立保健福祉大学），菊地乃依瑠（政策研究大学院大学），
隅藏康一（政策研究大学院大学）

1. はじめに

COVID-19 の流行が拡大している昨今、世界各国で医薬・医療品の研究開発(R&D)の加速と迅速な実用化が増々注目を集めている。日本でも状況は同じく、質の高い基礎研究から迅速な医薬品、医療機器開発へどのように繋いでいくのか、社会経済の持続的な発展という側面からも、国としての体制づくりが求められている(Huang,Liou & Iwaki 2020)*1。我が国でのこの種の取り組みはライフサイエンス関連予算の中で措置されており、中でも平成 19 年に厚生労働省が文部科学省、経済産業省と取りまとめた「医薬品・医療機器創出のための５か年戦略」はその端緒といえる*2。この５か年戦略により「橋渡し研究拠点」における臨床への橋渡し研究の取組強化が掲げられた。その後、平成２６年には健康・医療戦略推進法に基づき、「健康・医療戦略」「医療分野研究開発推進計画」が策定され、より臨床研究や治験実施環境の整備向上の必要性が重視された*3。現在はこれらの戦略や計画の後継として「橋渡し研究戦略的推進プログラム」(第 3 期プログラム)が進行中である。このプログラムでは、北海道大学・東北大学・筑波大学・千葉大学・東京大学・慶應義塾大学・国立がん研究センター中央病院/東病院・名古屋大学・京都大学・大阪大学・岡山大学・九州大学の１３の拠点が設けられている*4。

日本での橋渡し研究（トランスレーショナルリサーチ、以下 TR）は、「学術機関等でのアカデミアが基礎研究から応用・臨床研究を経て最終的に実際の医療へ用い、製薬・医療品等の研究開発に参画する上で製品・実用化を目指した研究」と定義づけられている*5。各橋渡し拠点はアカデミアからの革新的な基礎研究の成果を実用化に繋ぐ体制づくりが求められており、そのために人材確保・育成を含めた拠点機能の強化やネットワーク化を目的とした複数の基盤整備事業、シーズを育成し実用化を目指す橋渡し研究や医師主導治験を支援する研究開発課題、が並行して進められている*6。

これら体系的な橋渡し研究戦略による医薬・医療機器の研究開発促進や研究体制構築の重要性が増す一方、オープンイノベーションが加速化する中で研究開発タイムラインの構造やそれぞれの段階で持たせるべき研究機能が複雑化している。基礎整備事業の中でもシーズに複数のフェーズが存在しており、各フェーズで求められる段階的な成果に各研究課題がどう位置付けられているか明確でない為、橋渡し拠点の持つ各機能にどのような有効性があるのかは評価が難しいといえる。特に医薬・医療機器の研究の第一成果である論文では、これまで行っていた研究知識が存在し、関わってきた機関等の関係性によりイノベーション促進が期待できるにもかかわらず、この部分の評価はあまり重要視されていない。この初段階における論文成果の評価のためには、研究を行うことで得られた学識傾向や他施設との関わり構造を理解する必要がある。よって、橋渡し研究開発における第一段階の論文成果を基に、深層学習での言語処理を用いて研究課題や内容の構造を明確にする。

2. 先行研究

上記のように、橋渡し研究における拠点の体制は１）基礎研究の成果から実用化に繋ぐために人材確保・育成とネットワーク化を含めた基礎整備事業と、２）シーズ育成により実用化を目指すことである。しかし、拠点における基礎整備事業の評価をするにあたり、基礎研究から実用化までの研究プロセスでの研究知識・研究ネットワークがどのような構造であるのか理解されていない。戸高（2015）では、拠点の運営に当たり基礎研究から実用化までの多彩なプロセスが存在することを指摘したうえ

で、アカデミア人材の育成や、ニーズや傾向に応じた開発に課題があるとしている。加えて、橋渡しの成功に至るシーズの育成には高い品質が求められること、論文にはなかなかつながらないため研究者としての関心が高まらないことを問題点として挙げている[*7]。実際に橋渡し研究戦略ではどのような目的で行われているのか、研究はどのように貢献しているのかといった具体性を問われる意見もある。上記の先行研究や言説からも、研究の目的設定や貢献度を認識するためには、研究におけるニーズや傾向を理解しアカデミアや研究者等のネットワークディフューションを確認することが有効であると思われる。既存の事業戦略を基盤に持続可能な研究体制を構築するためにも、論文成果から行われてきた研究の構造を理解するといった評価を行う必要性がある。

2.1 橋渡し研究における論文成果の国際比較

先行研究により、TR 研究では一般的に論文成果に結びつきづらい事例も多いため、論文に関する目標設定自体が重要視されない点が指摘されている。そこで第一に、論文成果の国際比較で現状を理解する。国際論文が掲載されている Scopus のデータベースを使用し、日本で橋渡し研究が始まったとされる 2007 年から 2017 年までの論文を対象とした。検索キーワードは、国際的にも橋渡し研究とされている「Translational research, Translational Science」とし、比較対象はアメリカ、イギリス、ドイツ、オーストラリアとした（下記表１）。

表１：論文成果の国際分布

	国（所属機関）	Query（検索クエリ）	期間	論文件数	割合	検索日
1	日本	(TITLE-ABS-KEY ("Translational research" OR "Translational Science")	2007-2017	593	3.3%	Sep-20
2	アメリカ			9,359	52.1%	
3	イギリス			2,020	11.2%	
4	オーストラリア			743	4.1%	
5	ドイツ			1,457	8.1%	
	全て			17,971		

表１より、11 年間のアメリカにおける橋渡し研究の論文成果は上位（9,359 件/52.1%）を占め、続いてイギリス（2,020 件/11.2%）、ドイツ（1,457 件/8.1%）、オーストラリア（7,34 件/4.1%）、日本（5,93 件/3.3%）となっていた。量的結果から考えると、日本は他国より少ない論文成果であることがわかる。加えて、どのような分野で論文成果が構成されているのかを知るために、分野別の論文数の分析を行った（図１）。

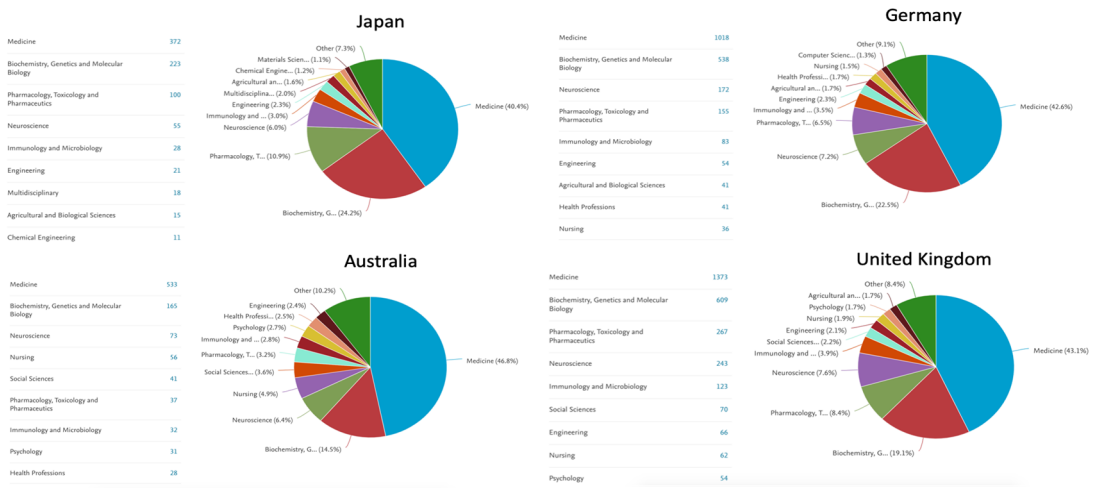

図１：分野別の文献数

図１に示すように､各国とも一番多く占めるキーワードは､Medicine であり､次いで Biochemistry, Gene 関連であることがわかる。製薬企業が多いイギリス・ドイツ・日本では Pharmacology と続いている。対照的にオーストラリアは Nursing, Social Science が頻出している。国ごとの論文数の差は大きいが、分野別でみるとイギリス・ドイツ・日本では製薬・医療品等の研究開発を重要視していることがわかった。論文数の量的尺度で貢献度を見ると、日本の論文数は他国と比較し少ない。アカデミアや研究者の人材・人的資源における規模・マンパワー・資金等や英語を母国語としないことから低くなった可能性はあるが、英語を公用語としないドイツと比べても日本の論文数は少ない傾向にある。

評価手法は各国で異なっており、イギリスやオーストラリアでは、量だけでなく質的な貢献度も評価する傾向となっている。例えば、オーストラリア政府における National Health and Medical Research Council（NHMRC）での TR 事業では、「TR 研究を通した知識向上は、オーストラリアの健康や公衆衛生上の問題を下支えする」と明確に説明されている。この研究を通した知識生産は単独では成功せず、他機関や他の研究者との協力や努力によりこの知識はより良いヘルスサービスへの貢献へ結びつくことができるという（Australian Government, NHMRC）[*8]。言い換えると、橋渡し研究における成果とは、医薬・医療機器を市場に出すだけではなく、それまでの研究プロセスから生まれる研究傾向や知識波及も評価する際には重要要素である。日本ではこのようなプロセスを対象とした評価はされていない。よって、日本の橋渡し研究の貢献度を一元的に見るだけではなく、論文成果より得られた多角的視点から学識・研究傾向や他施設との関わり構造を理解するための評価分析を行う必要性がある。

3. 主たる研究の概要

研究の主目的は、計量書誌学（bibliometric）アプローチを基に日本における橋渡し研究開発を、深層学習での言語処理（トピックモデル）を用いて研究内容の構造を理解することにある。副次的目的には、各拠点での１）研究を開始してからどういった研究トレンドがあったのか。２）研究分野の分類を明確にし、どのような研究に強みがあるのか。３）参画・協力もしくは共同している研究機関はどのようなネットワークであるのか、を可視化する。データは、日本の橋渡し研究に関わる論文を Scopus よりダウンロードし、データセットを作成した。分析は、日本でのTR第一期事業開始は2007年であり、それから 11 年間。分析のターゲットは、各拠点と、国立研究開発法人日本医療研究開発機構(AMED)での革新的医療技術創出拠点プロジェクトの助成金からなる橋渡し研究であり、それぞれの表題・アブストラクト・検索キーワードのデータを集約した。分析方法は、AI による言語処理（Latent Dirichlet Allocation（LDA））を行い、各論文中の単語の「トピック」を確率的に求める言語モデルを Python でコーディングした。医療分野で LDA 分析を使用した論文は、日本ではまだない。この全集約されたデータから、トピックモデルを用いて各研究トピックの集合体が確率的に分類され、研究トレンドを把握することができる。さらに、研究開発での疾患・人口・ターゲット・フェーズの分類を行った上、各拠点での研究インパクトを確認する。

4. 結果

4.1 橋渡し研究に関わる各拠点での論文成果

まず初めに、各拠点と AMED における橋渡し研究の論文件数・割合を算出した。表２にあるように、論文総数は 3,765 件であった。論文件数が一番高いのは京都大学で 892 件（23.7%）であった。その次に、東京大学 602 件（16%）大阪大学 526 件（14%）と続いていた。

表２：論文数とその割合

	国 （所属機関）	Query（検索クエリ）	期間	論文件数	割合(%)	検索日
1	北海道大学	(All ("Translational research" OR "Translational Science") AND AFFIL	2007-2017	237	6.3%	Sep-20
2	東北大学			240	6.4%	
3	筑波大学			119	3.2%	
4	千葉大学			125	3.3%	
5	東京大学			602	16.0%	
6	慶應義塾大学			214	5.7%	
7	国立がん研究センター中央病院/東病院			153	4.1%	
8	名古屋大学			176	4.7%	
9	京都大学			892	23.7%	
10	大阪大学			526	14.0%	
11	岡山大学			119	3.2%	
12	九州大学			355	9.4%	
13	AMED	FUND-ALL ("Project of Translational and Clinical Research Core Centers") AND FUND-ALL ("Japan Agency for Medical Research and Development"))		7	0.2%	
	Total			3765	100%	

4.2 橋渡し研究に関わる研究分野の分類（トピック分類）

次に、集約した橋渡し研究の表題・アブストラクト・検索キーワードのデータを深層学習による言語処理を行い、各論文中の単語の「トピック」を確率的に求めた。トピックは 10 に集約され、それぞれのクラスターにより頻度の高い語を 20 語集められた。さらに、これらトピックと頻度の高い語から考えられる、疾患・人口・ターゲット・フェーズを分類した。フェーズ分類に関しては、AMED の「開発フェーズ」（研究開発支援を行っている各事業での医薬・医療機器等の対象となるフェーズの設定）を使用した。この開発フェーズでは、医薬・医療機器での大分類を採用し、基礎的・応用・非臨床・臨床試験/治験の第四分類を基準とした*9。トピック分類の結果を表３で示している。もっとも研究傾向頻度が高かった分野は、トピック１（910 件）（がん（悪性新生物）が疾患ターゲットとされ、患者における男女比較を伴った臨床研究や治験）であった。第二番目に頻度が高かった分野は、トピック８（527 件）（同じくがんを疾患ターゲットとし、動物実験において細胞におけるタンパクの発現等の基礎研究）であると特定された。対照的に少なかった分野は、トピック９（113 件）（成長ホルモンやプラズマ細胞等の免疫学的発現に関連する分野であり、医薬品の用法・容量の効果測定に関わる前臨床から臨床・治験）であると分類できる。総合的に全てのトピックから考えられる研究分野の傾向としては、医薬品関連が医療機器関連に対し多い傾向と理解できる。医療機器研究が少ないことが原因で語が取れていない可能性が示唆される。

表３：トピック分類

Topic	Common 20 words	Target disease and science	Target Population	Target part or outcome	Categorized by field	No of Paper
Topic 1	patient aged cancer clinical treatment survival disease risk trial blood diabetes adult female group male analysis drug level japan	Cancer, blood, diabetes	Aged patient, male and female	Clinical treatment or trial, survival	Clinical trial	910
Topic 2	protein cell kinase drug human lung gene mutation receptor tumor peptide cancer binding molecular small proteins acid response stress EGFR	lung, cancer, kinase, peptide	Human	Protein, Cell kinase, Gene or molecular mutation (EGFR)	Basic or applied	220
Topic 3	cell animal tissue bone stem heart muscle protein mouse factor transplantation cells cardiac rat model human mice animals induced	Born, heart, cardiac, transplantation	Animal	Stem cell, tissue, induce	Basic	516
Topic 4	disease patient heart coronary artery aged risk male female cardiovascular clinical treatment outcome group therapy failure cholesterol acute article	Coronary, artery, cardiovascular	Aged patient, male and female	Risk, clinical treatment or therapy, outcome	Pre-clinical or clinical trial	278
Topic 5	research clinical medical translational health data japan based using drug human medicine analysis development model information review mass	NA	NA	Clinical or medical research using data and drug, development model	Clinical trial	301
Topic 6	human cell polymorphism disease protein genetic gene association hla single nucleotide analysis antigen locus structure interleukin factor genotype molecular	Polymorphism, interleukin	Human	Cell, genetic gene protein, HLA, nucleotide interleukin genotype	Basic or applied	153
Topic 7	gene expression human cancer dna analysis mutation disease genetic brain protein factor microrna tumor wa mir receptor female aged male	Cancer, brain, protein	Human	Gene expression, DNA genetic	Basic or applied	416
Topic 8	cell protein mouse expression animal mice factor cells gene tumor receptor human cancer line growth induced model proliferation transcription	Cancer	Animal, mouse and mice	Cell protein, expression, cell gene growth factor	Basic	527
Topic 9	drug dose ghrelin clinical male administration level plasma effect phase treatment concentration mg day trial schizophrenia controlled cisplatin liver	Ghrelin, plasma, schizophrenia, cisplatin, liver	Male	Drug dose, clinical administration, effect level, treatment	Pre-clinical to clinical trial	113
Topic 10	cancer tumor cell drug carcinoma survival human therapy neoplasms antigen patient clinical aged acid metastasis expression female article liver	Cancer, carcinoma, neoplasms, liver	Human, aged patients female	Cell drug carcinoma survival therapy	Clinical trial	302

4.3 橋渡し研究拠点のトピック別論文数と割合

上記のトピック分析結果により研究分野の分類が明確になったが、さらに拠点別の強みを理解するために各拠点のトピック別論文数と割合を算出した。前結果と同様（下記表４）、トピック１における研究開発が京都大学（285 件/32%）を含め、ほとんどの拠点が行っている傾向であるが、千葉大学はトピック８がもっとも多い。さらに、大阪大学（99 件/19%）・東北大学 （65 件/27%）は他拠点と違い、トピック３（動物実験を伴った基礎研究範囲での細胞・組織学における誘導・誘発因子の研究）に重きを置いている傾向である。論文成果からなるデータを言語処理することにより、各拠点の橋渡し研究における研究開発の構造が推測でき、トピック分類を行うことで論文成果に基づいたそれぞれの研究の強みを類推できた。

表４：各拠点における各トピック別論文数と割合

	Kyoto Uni		Chiba Uni		Hokkaido Uni		Keio Uni		Kyushu Uni		Nagoya Uni	
	論文数	割合(%)	論文数	割合(%)	論文数	割合(%)	論文数	割合(%)	論文数	割合(%)	論文数	割合(%)
Topic_1	285	32%	18	15%	74	31%	58	27%	76	22%	54	31%
Topic_2	36	4%	6	5%	12	5%	13	6%	20	6%	9	5%
Topic_3	116	13%	10	8%	18	8%	26	12%	31	9%	31	5%
Topic_4	72	8%	8	6%	23	10%	19	9%	15	4%	6	3%
Topic_5	62	7%	9	7%	19	8%	17	8%	61	17%	5	3%
Topic_6	27	3%	4	3%	12	5%	7	3%	7	2%	9	5%
Topic_7	61	7%	21	17%	26	11%	7	3%	34	10%	26	15%
Topic_8	133	15%	31	25%	25	11%	36	17%	55	16%	18	10%
Topic_9	29	3%	6	5%	8	3%	14	7%	14	4%	6	3%
Topic_10	68	8%	11	9%	19	8%	15	7%	38	11%	9	5%
	889	100%	124	100%	236	100%	212	100%	351	100%	173	100%

	National center		Okayama Uni		Osaka Uni		Tohoku Uni		Tokyo Uni		Tsukuba Uni	
Topic_1	58	38%	26	22%	78	15%	37	16%	99	17%	46	39%
Topic_2	8	5%	5	4%	41	8%	18	8%	46	8%	5	4%
Topic_3	2	1%	24	20%	99	19%	65	27%	79	13%	13	11%
Topic_4	3	2%	7	6%	32	6%	19	8%	63	11%	11	9%
Topic_5	7	5%	5	4%	37	7%	18	8%	56	9%	5	4%
Topic_6	2	1%	6	5%	29	6%	15	6%	29	5%	6	5%
Topic_7	41	27%	12	10%	50	10%	28	12%	100	17%	9	8%
Topic_8	15	10%	15	13%	76	15%	25	11%	81	14%	15	13%
Topic_9	3	2%	6	5%	10	2%	5	2%	10	2%	1	1%
Topic_10	14	9%	12	10%	67	13%	7	3%	34	6%	8	7%
	153	100%	118	100%	519	100%	237	100%	597	100%	119	100%

5. 著者・共著者間のネットワーク

最後に、橋渡し研究が始まった 2007 年からの 11 年間で研究ネットワークがどのように構成されているのか可視化するため VOSviewer を使用した。Scopus のデータセットより、著者・共著者所属の出現語句のクラスター分析を試みた。語句の頻度は 15 とし、関連性の高い上位 60%を条件として抽出された。下記の図２の結果によると、赤いクラスターは国内における研究ネットワークが強いことを示している。しかしながら、国際共著となっている青や緑のクラスターは距離も遠く、あまり国際連携されていないことがわかる。

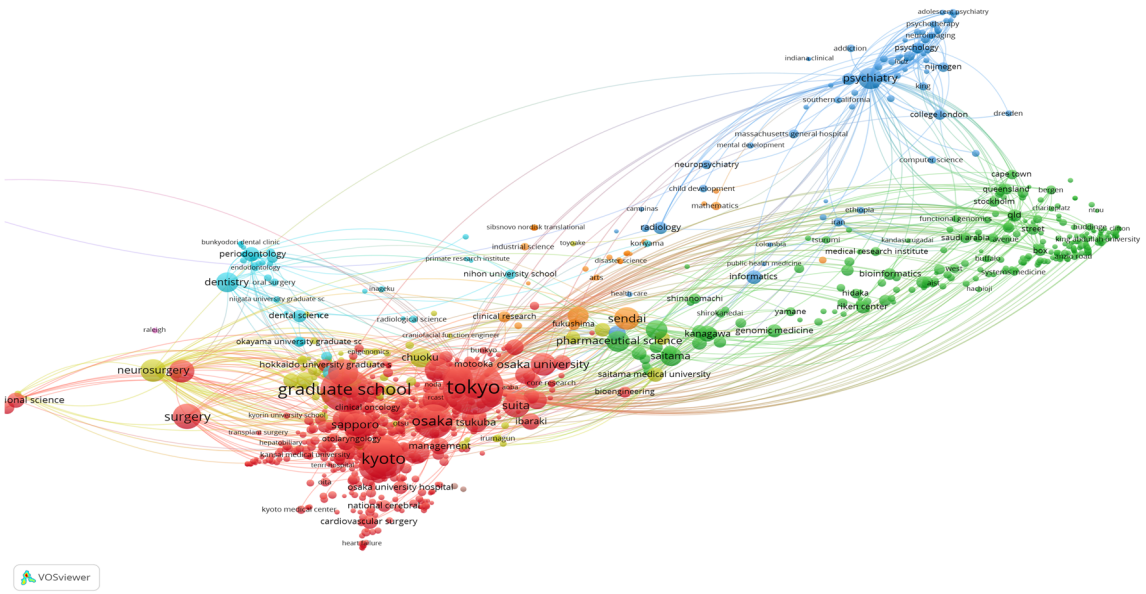

図２：著者・共著者のネットワーク

6. 議論

主目的である日本における橋渡し研究開発での計量書誌学（bibliometric）アプローチを基に、深層学習での言語処理（トピックモデル）を用いて研究内容の構造を理解できた。副次的分析により、橋渡し研究第一期事業から 11 年間での英語論文成果の量的評価から、国際的に比較すると論文成果は少ない傾向であることがわかった。さらに、深層学習による言語処理を行うことで各拠点での研究の強みや傾向が研究開発トレンドやパターンがクラスターに分かれ、分類が可能となった。問題提起にあったような第一段階である論文成果からなる貢献度を多元的視点から確認でき、研究開発を通した知識累積が得られ学際的発展は進んでいると考えられる。しかし、研究ネットワーク分析では、国内共同研究は多いものの、国際連携は少ない傾向が見られた。橋渡し研究の目的にあるように、拠点機能の強化やネットワーク化を目的とした基盤整備事業を支えるといった点からは、国内だけに止まらず国際的に研究成果を発信していくためにも、人材確保・育成は必要であると考えられる。

参考文献

[1] Huang, MC, Liou, MH and Iwaki, Y. 2020, “The impact of R&D and innovation on global supply chain transition: GTAP analysis on Japan’s public R&D investment”, J. Soc. Econ. Dev. https://doi.org/10.1007/s40847-020-00113-1.

[2] 厚生労働省医政局経済課，革新的医薬品・医療機器創出のための５か年戦略，４月（2007）/ https://www.mhlw.go.jp/houdou/2007/04/h0427-3.html

[3] 首相官邸，健康・医療戦略，７月（2014） https://www.kantei.go.jp/jp/singi/kenkouiryou/suisin/ketteisiryou/dai2/siryou1.pdf

[4] 日本医療研究開発機構，革新的医療技術創出拠点プロジェクト，拠点一覧，１１月（2018）

[5] 文部科学省，先端医科学研究の臨床への応用の推進に関する懇談会検討報告書：懇談会報告書，８月（2019）

[6] 文部科学省，研究振興局ライフサイエンス課，橋渡し研究支援総合戦略，９月（2019）/ https://www.mext.go.jp/component/b_menu/shingi/toushin/__icsFiles/afieldfile/2019/09/20/1421363_003_1.pdf

[7] 戸高 浩司，臨床研究と開発：Academic Research organization の役割，11 月 (2015)

[8] Australian Government, National Health and Medical Research Council (NHMRC), Research Policy, Research Translation/ https://www.nhmrc.gov.au/research-policy/research-translation-and-impact

[9] 国立研究開発法人日本医療研究開発機構(AMED), 橋渡し研究：開発フェーズ（参考資料） https://www.amed.go.jp/content/000022342.pdf