

東邦大学学術リポジトリ

Toho University Academic Repository

タイトル	Cox比例ハザードモデルを用いた資料の除籍に関する要素の探索的分析
作成者（著者）	橋本, 郷史
公開者	日本医学図書館協会
発行日	2023.06
ISSN	04452429
掲載情報	医学図書館. 70(2). p.101 108.
資料種別	学術雑誌論文
内容記述	論文
著者版フラグ	publisher
その他資源識別子	https://cir.nii.ac.jp/crid/1050860764322089472
メタデータのURL	https://mylibrary.toho-u.ac.jp/webopac/TD28216134

Cox 比例ハザードモデルを用いた資料の除籍に関する要素の探索的分析

橋本 郷史*

東邦大学医学メディアセンター大橋病院図書室

I. 背景と目的

1. はじめに

汗牛充棟^{かんぎゅうじゅうとう}は図書館員の憧れかもしれないが、資料を無限に所蔵することはできない。書架の容量には限界があり、新たに資料を受け入れるためには、その分過去に受け入れた資料を除いて場所を作らなければいけない。外部に倉庫などを契約して資料を分散所蔵する機関もあるが、現実的にはどこで所蔵しようとも、どこかの時点で収容の限界が訪れる。資料の除籍は図書館員にとって避けては通れない作業である。一方で、どの資料をいつ捨て、どの資料をいつまで残すのかを判断することはとても難しいことでもある。除籍判断の難しさを軽減することができないだろうかと考え、今回の調査を行った。

2. 除籍作業の概要及び業務上の悩み

この調査は実際の除籍作業に関する業務上の悩みに端を発するものである。そこでまずは著者の勤務する東邦大学医学メディアセンター（以下、本センター）における除籍作業の概要及びその課題について説明する。

本センターは、医学部と看護学部の2つの学部図書館と、大橋と佐倉の附属病院内にある2つの病院図書室の計4つの施設からなる。大森の医学部キャンパスにある「本館」が最も規模が大きく、所蔵する図書資料数は約57,000冊で、年間2,000冊弱の図書資料を新規に受け入れている。図書資料を配架している書架の総延長距離における占有率は8割程度となっている。東邦大学（以下、本学）では資料の増加に対応するため、毎年の受け入れ相当分程度の資料を除籍することが認められている。また、本センターは主に医学系の資料を所蔵しているという性質上、経年とともに資料へのニーズが減衰しやすく、古くなった資料を除籍しやすいという事情がある。

しかし、一度資料を除籍すれば二度と利用できなくなってしまうため、除籍する資料の選定は慎重を要する。本センターでは、資料管理担当者が日々除籍対象を選定し、ほぼ隔月で除籍を行っている。選定の偏りや、不注意による除籍漏れ、あるいは除籍しすぎなども生じる恐れがあるため、除籍選定後の確認作業も必要になり、除籍がもたらす業務的な負荷は大きい。また、電子資料の増加に伴い、物理媒体の資料管理に労力を割くことは年々難しくなっている。除籍選定とそのチェックをいかに効率化できるかが課題となっている。

3. 調査の目的

しかし、除籍作業の効率化は一朝一夕にはなし得ない。そこで、今回の調査では、まずは資料の持つ要素と除籍のされやすさとの関係についての探索的な分析を試みることにした。

「この本はそろそろ除籍したほうがよさそうだ」と、図書館員が判断する。この時、資料の持つある種の要素がその判断に影響を与えていることは間違いない。除籍の判断には多くの要素が関わってくるが、その中には肌感覚や個別の特殊事情以外の汎用的かつ他と比較可能な客観的な要素も含まれるはずである。そのような要素を明らかにすることができれば、除籍を体系的に捉え、分析し、効率化するための一助となると考えた。

今回の調査の目的は「資料が持つ、除籍の判断と関わるデータの分類・計測可能な要素を抽出すること」と「それらの要素と除籍との関係を数値的に可視化すること」である。

II. 方法

1. 分析対象

過去に受け入れをした資料を対象データとし、除籍と資料の要素との関係を分析した。なお、この調査は最終的には業務の効率化に資することを前提としているため、分析には業務上利用可能で入手が容易なデータをな

*Satoshi HASHIMOTO : ヘルスサイエンス情報専門員（上級）
〒153-8515 東京都目黒区大橋2-22-36. Tel.03-3481-8478
hashimoto@mnc.toho-u.ac.jp (2023年1月27日 受理)

るべく加工せずに用いた。入手や加工が難しいデータでは継続して業務に反映させることができないためである。

前章で述べたように、本センターの資料は医学系の資料が主であり、資料が古くなるとその情報価値が下がるため、資料発行からの経過年が除籍の大きな要素であることには疑いが無い。今回の調査では発行からの経過年以外の除籍に関する要素について知るため、過去のある一定の時期に購入した資料のその後の除籍・在籍状況と資料の要素との関係を調べた。分析の対象とする資料は、本センターの本館で2011年度に受け入れした図書資料とし、対象データの取得は2021年10月14日に行った。

地区を本館に限定したのは次の理由による。本センターは前述のとおり4つの図書館・図書室からなるが、対象利用者層やサービス形態、所蔵規模が異なると所蔵資料の構成や除籍の傾向も異なり分析が難しくなるため、最も規模が大きく年間の受け入れ資料数が多い本館を分析の対象とした。研究室など図書館以外で所蔵している資料も対象外とした。

また、資料種別が異なると除籍に対する事情や考え方も変わるため、雑誌・視聴覚資料は除外し、今回は図書のみを対象とした。

2011年に「出版された」資料ではなく、「受け入れをした」資料をベースにする理由は次のとおりである。書誌データの出版年は、volがある場合など、実際の資料の出版年と異なるものも多く、機械的な抽出に利用するには精度が低く使いづらい。一方、資料を受け入れた日は資産登録に結びつくため、データ上でもれなく管理されており、抽出条件として利用しやすい。また、本センターの本館では新刊図書を見計らいで購入しており、ある年度に受け入れられる資料は基本的にその年度に出版された資料になる。ただし、利用者からの希望で購入する図書など、出版年に多少のばらつきがあるので、今回はさらに手作業でデータを確認し、2009年以前に発行された資料は除外した。

2. 分析手法：Cox比例ハザード分析

除籍は資料の死である、と仮定して、資料が除籍されるまでの期間（生存時間）と、資料の持つ要素との関係をCox比例ハザードモデル¹⁾を用いた分析手法（以下、Cox比例ハザード分析）によって確認した。この手法は生存時間分析の代表的な手法で、生存時間への各種の説明変数（ここでは資料の各要素）の影響をハザード比（Hazard Ratio。以下、HR）で表現する。HRとは、イベント（ここでは除籍）の起こりやすさを比率で表したもので、一般に、この値が1を超えていれば、その要素

はイベント発生リスクを増大させ、1より小さければリスクを減少させる、と解釈される。

Cox比例ハザード分析は観察期間の異なるデータを比較でき、加えて、イベント発生以外の理由で期間中に対象から外れるデータを「打ち切り」データとして扱うことができる。打ち切りデータとなった場合、打ち切り発生以降の計算ではその発生数を分析対象の全数から減らして計算を行うことになる。このことから、Cox比例ハザード分析は、今回の対象のように、除籍までの期間が資料ごとに異なり、かつ期間中に除籍以外の理由で観察対象が減少する可能性があるデータの分析に適していると考えた。分析には統計ソフトEZR²⁾（バージョン1.53）を用いた。

3. 一般的な除籍基準の確認

Cox比例ハザード分析にかける要素は資料の要素であれば何でもいいというものではない。関係があると思われる要素をあらかじめ選出した上で分析にかける必要がある。分析にかける要素を決めるために、まずは一般的な図書の除籍基準として、本学を含め、他の大学図書館の基準やMLAの公開している蔵書構築基準を確認した^{3)~9)}。

除籍には大きく分けると2種類ある。1つは、汚損・破損・紛失・移管など、図書館側では制御できない外部事情によってやむを得ず資料を除籍するケースである。もう1つは、図書館員がその資料を不要と判断して能動的に除籍をするケースである。この調査では、図書館員が行う除籍判断に関連する要素を調べたいため、1つ目のケースは除籍関連要素の分析対象としては扱わず、2つ目のケースのみを除籍関連要素の分析対象として扱った。2つ目のケースに該当する除籍基準としては「利用頻度が低い」「改版がある」「発行年が古い」「重複所蔵している」「媒体が古い」「代替資料がある」といったものが挙げられている。

もちろん除籍の要素はこれだけではなく、その他の複合的な要素や、歴史的に重要である、所蔵機関に関連した資料である、といった資料個別あるいは所蔵する機関に独自の要素もある。ただ、今回の調査ではまずは一般的な要素について調査することとし、独自の要素については今回の対象から除外した。

4. 分析に用いる変数

資料の生存状態の判定には、データ取得時点（2021年10月14日）の在籍状態を用いた。除籍された資料は、いつ除籍されたかに加え、不要、汚損、紛失などの除籍理由の情報も付加されている。除籍理由を確認して「不

要」として除籍したものを除籍イベント（＝資料の死）として扱った。除籍理由が「汚損」「紛失」などになっているものは、打ち切りデータとして扱った。

生存日数は、2012年4月1日から除籍されるまで、あるいは打ち切りとなるまでの日数とした。データ取得時点までに除籍されていない資料であれば、データ取得日である2021年10月14日までの日数が生存日数になる（最大3,483日）。日数のカウントのスタートを資料取得年度の翌年度の初日としたのは、本学の規定上、資料を購入と同じ年度内に除籍することは原則できないためである。

説明変数として、前節Ⅱ.3.の除籍の一般的な基準も参考にしながら、図書館システムから抽出可能な以下の7つの要素を用いることとした。①複本の有無、②登録翌年度の利用回数/10、③改訂版かどうか、④単著かどうか、⑤言語（日本語か英語か）、⑥分類、⑦各分類の蔵書数/1,000。

いくつかの要素について補足する。

除籍基準からも、「利用頻度が低い」資料は除籍の対象となりやすいと思われるため、②登録翌年度の利用回数を分析に加えた。本センターでは、貸出だけでなく館内利用も回数をカウントしているため、貸出と館内利用の両方を足したものを利用した。ただし、多いものでは利用回数が100回を超えるものもあるため、利用1回あたりのHRを計算しても影響度合いが見えづらい。そこで、数値を10分の1にして変数化することで、利用10回あたりのHRを算出する。なお、抽出の容易さを考慮して登録翌年度の利用回数を用いている。

除籍基準に挙げられている「改訂版がある」に関連して、③改訂版かどうかを分析に加えた。注意が必要なのは、ここで扱った要素は「改訂版があるかどうか」ではなく「その資料が改訂版かどうか」であるということである。改訂版があるかどうかを知るには人の手でその後発行された全ての資料をチェックしなければいけないため、なるべく入手が容易なデータを利用する、というこの調査の趣旨に反する。一方で、その資料が改訂版かどうかは、抽出したデータからそうと分かる。改訂版の判断のために、書誌の版表示を確認した。また、いわゆる「VOL積み」と言われる、同じ書誌の中で改訂を続ける資料についても書誌及び巻号の情報を確認し、改訂に該当すると判断したものは改訂版として扱った。なお「机上版」「新装版」など、内容の改訂に関係しない版は改訂版として扱っていない。

資料の購入時には、内容の偏りを判断するための一つの要素として単著か複数人著作かを確認する。その要素

の除籍への影響をはかるため、④単著かどうか、も分析に加えた。学会など、団体が発行している資料は複数人著作として扱った。

資料の分野の違いは除籍に影響している可能性があるため、⑥分類も分析に加えた。本学ではNDCをベースにした独自分類を使用しているため、大きくは医学系（NDC49番台）かそうでないかで分け、更に医学系の中身を分類の3桁目で490から499までの10個に分けた。

除籍基準に挙げられている「代替資料がある」かどうかをデータから判断することは容易ではないが、同じ分類に属する資料数の多寡はこの基準に類似した一つの指標になると考え、⑦各分類の蔵書数、についても分析に加えた。これまで挙げた①-⑥の要素は、2011年度に受け入れた資料のものだが、⑦については、2011年度開始時点の本館の全図書資料を元にして算出した。区分には要素⑥の分類を用い、蔵書数1冊あたりでHRを算出して影響度が見えづらいため、数値を1,000分の1にし、1,000冊あたりのHRを算出する。

除籍基準で挙げられている「媒体が古い」については、今回の調査は図書が対象で媒体が1種類しかないため要素化しなかった。

Ⅲ. 結果

1. 概要

分析の対象件数は1,934件で、そのうち調査時点で在籍している資料は8割弱の1,500件だった。不要と判断され除籍となった資料は約2割の419件、紛失などやむを得ない事情で除籍（＝打ち切り）となった資料は1%に満たない15件だった。資料全体及び、打ち切りを含む在籍資料と、不要と判断され除籍となった資料の基本属性と特徴は表1のようになった。

除籍された資料はそうでない資料と比べ、平均利用回数が2.4倍と多かった。また、複本がある資料の割合は4.2倍、改訂版資料の割合は2.6倍、英語資料の割合は1.7倍とそれぞれ高かった。

分類別で見ると、除籍された資料はそうでない資料と比べ「基礎医学」（1.8倍）や「薬学」（5.8倍）分野の資料の割合が高く、一方で「内科学」（0.8倍）や「婦人科・産科」（0.5倍）「眼科・耳鼻科」（0.3倍）「口腔・歯科」（0.5倍）分野の資料の割合は低かった。

2. Cox比例ハザード分析の結果

各要素の単変量Cox比例ハザード分析の結果は表2のようになった。

表1. 分析対象資料の基本属性と特徴

要素	全体 (N = 1934)	在籍 (含む打ち切り) (N = 1515)	除籍 (不要) (N = 419)
生存日数：中央値 (四分位範囲)	3483 (3483 - 3483)	3483 (3483 - 3483)	2191 (2131 - 2644)
複本有：数 (%)	104 (5.4)	48 (3.2)	56 (13.4)
利用回数：平均 (標準偏差)	1.86 (5.59)	1.42 (4.93)	3.44 (7.30)
改訂版である：数 (%)	700 (36.2)	408 (26.9)	292 (69.7)
複数人著作：数 (%)	1715 (88.7)	1329 (87.7)	386 (92.1)
英語：数 (%)	310 (16.0)	210 (13.9)	100 (23.9)
分類：数 (%)	490 (医学一般)	136 (7.0)	104 (6.9)
	491 (基礎医学)	220 (11.4)	146 (9.6)
	492 (臨床医学)	235 (12.2)	189 (12.5)
	493 (内科学)	525 (27.1)	433 (28.6)
	494 (外科学)	383 (19.8)	304 (20.1)
	495 (婦人科・産科)	42 (2.2)	37 (2.4)
	496 (眼科・耳鼻科)	68 (3.5)	63 (4.2)
	497 (口腔・歯科)	18 (0.9)	16 (1.1)
	498 (公衆衛生)	107 (5.5)	87 (5.7)
	499 (薬学)	44 (2.3)	17 (1.1)
	49以外	156 (8.1)	119 (7.9)
	37 (8.8)		

表2. 単変量Cox比例ハザード分析の結果

要素	HR (95% CI)	
複本有	3.87* (2.92 - 5.13)	
利用回数 (10回あたり)	1.26* (1.18 - 1.35)	
改訂版である	4.95* (4.01 - 6.09)	
複数人著者	1.57* (1.10 - 2.23)	
英語	1.71* (1.36 - 2.13)	
分類	490 (医学一般)	1.08 (0.75 - 1.55)
	491 (基礎医学)	1.83* (1.42 - 2.35)
	492 (臨床医学)	0.86 (0.63 - 1.17)
	493 (内科学)	0.72** (0.57 - 0.91)
	494 (外科学)	0.93 (0.73 - 1.19)
	495 (婦人・産科)	0.52 (0.22 - 1.26)
	496 (眼科・耳鼻科)	0.31** (0.13 - 0.74)
	497 (口腔・歯科)	0.48 (0.12 - 1.92)
	498 (公衆衛生)	0.83 (0.53 - 1.30)
	499 (薬学)	5.22* (3.53 - 7.71)
	49以外	1.15 (0.82 - 1.61)
蔵書数 (1,000冊あたり)	0.99 (0.97 - 1.02)	

* 除籍リスク増大

** 除籍リスク減少

この結果は、あるなしの二値要素の場合、この要素がない場合に対してのある場合のHRとなる。量的な要素の場合、数量が1増えるごとのHRとなる。なお、このHRはこの観測期間における平均的傾向を示したものである。今回の分析ではHRの数値自体を細かく解釈しない。この要素があると除籍リスクがX倍になる、というような扱いはせず、あくまで、その要素が除籍リスクの増大因子になるか減少因子になるか、と、その影響が大きいか小さいか、という程度の判断に用いる。

各要素のHRの95%信頼区間(95% CI)が1より大きい小さいかで区分した。HRにアスタリスク(*)が1つついている要素は、その信頼区間が1より大きく、リスク増大に影響していると思われる要素である。HRにアスタリスクが2つついている要素は、その信頼区間が1より小さく、リスク減少に影響していると思われる要素である。印がついていない要素は、HRの信頼区間が1をまたいでおり、リスクを増やすとも減らすともどちらとも言いがたい要素である。

リスクを増大させる要素は「複本がある」「利用回数が多い」「改訂版である」「複数の著者によって書かれている」「英語の資料である」「基礎医学、薬学分野の資料である」であった。リスクを減少させる要素は「内科学分野、眼科・耳鼻科分野の資料である」であった。

IV. 考察

1. リスク増大要素

まずはリスクを増やす要素について考える。また、必要に応じて追加で多変量のCox比例ハザード分析を行いその結果を示す。

1) 複本がある(HR:3.87)、改訂版である(HR:4.95)

複本があると除籍されやすいというのは除籍基準的にも自然で、想像していたことがデータの的にも確認できた形である。改訂版であると除籍されやすい、というのも同じく除籍基準からすると当たり前のことのようにあるが、方法で述べたようにここでは「改訂版がある」ではなく「改訂版である」ことを見ており、その後版を重ねるかどうかはわからなくても、資料それ自体の改訂情報が除籍されやすくなる要素として使えることがわかったのは大きいと考える。

2) 利用回数が多い(HR:1.26)、複数人による著作である(HR:1.57)

「利用回数が多い」「複数人による著作」だと除籍リスクが増大しているとの結果だった。理由を考えてみると、利用回数が多い資料は、定番の資料のため版を重ね

ている資料であったり、複本であったりする可能性が高そうである。また、複数人で書かれた資料は単著の資料よりも、改訂されやすかつ内容的にもスタンダードな定番の資料である可能性が高いのかもしれない。そこで、この2要素をそれぞれ、複本有無と改訂版かどうかという変数を加えた3要素でCox比例ハザード分析にかけた。この多変量分析によって複本と改訂版の要素の影響を考慮した上での分析対象要素のHRが算出できる。その結果、利用回数(1/10)と複数人著作のHR(95% CI)はそれぞれ、1.04(0.94-1.15)、1.29(0.90-1.84)となった。

いずれもHRの信頼区間が1をまたいでおり、これらの要素がリスク増大の直接的な要因であるとは判断しづらい状態になった。利用回数が多いこと、複数人著作であることがリスク増大要因のように見えていたが、その要素を持つ資料は、複本があったり、改訂版であったりすることが多く、それらの要素の影響を受けていた。

なお、今回確認した利用回数は、抽出の容易さを考慮して登録翌年度の利用回数をを用いている。もし、除籍をされる直近数年間の利用回数をカウントすれば、別の結果が得られる可能性がある。

3) 英語の資料である(HR:1.71)

「英語の資料である」ことは除籍リスクの増大につながっていた。しかし、英語の資料は網羅的な選定がしづらく、比較的有名なものを中心にカタログや旧版の購入履歴などを参考にして購入している。そのため、日本語資料より改訂版である資料比率が高い。今回の対象データでは、日本語資料の改訂版率は約3割だったが、英語資料では約7割だった。改訂版資料であることはリスク増大の大きな要因のため、改訂版かどうかを変数に追加して分析を行った結果、HR(95% CI)は、0.92(0.73-1.16)となった。

リスクは減少し、HRの信頼区間は1をまたぐ形になった。英語資料が除籍されやすいのは、改訂版資料の比率が高いことが強く影響しており、英語資料ということ自体が除籍の要因ではなかった。

4) 「基礎医学」(HR:1.83)、「薬学」(HR:5.22)分野の資料である

「基礎医学」や「薬学」分野の資料は除籍リスクが高い結果となった。

基礎医学の資料は、漠然と除籍されにくいイメージがあるが、教科書に指定されている本が多いという特徴も持っている。これまでに分かっている除籍リスクの増大要素である、複本率と改訂版率について見てみるとこ

ろ、どちらも資料全体の平均よりも高い状態だった。複本の有無と改訂版かどうかという変数を加えて分析を行った結果、そのHR (95% CI) は、1.46 (1.13 – 1.89) となった。

HRは下がったが、信頼区間はまだ1より大きく、除籍リスクの増大要因となっていた。教科書に指定される資料はメジャーな資料であることが多く、そのような資料は複本率と改訂版率が高い。加えて、カリキュラムの変更で教科書でなくなった際に、複本を除籍することがある。基礎医学分野の資料であることそれ自体の除籍への影響力は必ずしも大きくなさそうではあるが、授業との関わりなどの複合的な要素がこの分野に影響を及ぼしているのだということがわかった。

続いて薬学に関して考えてみる。本学には本センターの他に薬学系図書館があるため、本センター本館の除籍選定時に、主たる収集対象ではない薬学系の資料を除籍しやすい傾向がある。また、本館は併設されている病院の医療従事者もサービス対象としており、薬学系資料については現場でよく使われる「医薬品集」などの短期間で改訂される資料の割合が高い。改訂版かどうかを変数に加えて分析を行った結果、そのHRは下がり、3.96 (2.68 – 5.86) となったが、薬学分野の資料であることが除籍リスク増大の要因であることは変わらなかった。

2. リスク減少要素

「内科学」(HR : 0.72) や「眼科・耳鼻科」(HR : 0.31) の資料であることはリスク減少の要因となっていた。どちらの分野も、資料全体の平均よりも複本率が低かったため、複本かどうかを変数に加えて分析してみたところ、そのHR (95% CI) は、内科、眼科・耳鼻科それぞれ、0.81 (0.64 – 1.03), 0.34 (0.14 – 0.82) となった。

内科学は複本率の影響を除くと、HRの信頼区間が1をまたぎ、分野そのものが除籍減少の要因になっているとはいえない状態になった。眼科・耳鼻科は複本率の影響を除いても、分野そのものが除籍リスク減少の要因となっていた。

これは、ひとつには眼科・耳鼻科領域の所蔵規模が小さいことが影響していると思われる。そのような分野では、コレクションが細りすぎないように、除籍選定時に除籍を抑制するためである。しかし、各分野の所蔵数あたりの除籍されやすさには数値としてはこの傾向が現れていない。この点について次の節で考察する。

3. 分野の所蔵規模と除籍のされやすさについて

ある分野の所蔵規模の小ささは除籍判断に抑制的な影響を及ぼすと述べた。眼科・耳鼻科だけでなく、婦人科・産科、口腔・歯科などの所蔵規模の小さい他の分野でも (資料数が少ないために信頼区間は1をまたいでいるものの)、HRが低くなる傾向が現れているのはそのためだと思われる。

しかし、各分野の所蔵数あたりの除籍されやすさには数値としてはこの傾向が現れていない。累積蔵書数1,000冊あたりのHRは0.99でその信頼区間は0.97 – 1.02となっている。これは、その分野の蔵書数の多寡からは除籍のされやすさに関する一般的な傾向を読み取る (計算する) ことはできない、という結果である。ここで用いた分類の粒度が荒いため、より細かく、NDC分類4桁目と5桁目までで分野の再区分を行い、それぞれの100冊あたりのHRと信頼区間を算出してみたが、4桁目では0.99 (0.98 – 1.00)、5桁目では0.99 (0.95 – 1.03) となり、細分化した分野区分でも、分野の所蔵数と除籍されやすさとの間に関係を見出すことはできなかった。

なぜこのような結果になるのかについては、次のような原因が考えられる。分野の所蔵規模は確かに除籍の際の考慮事項である。しかし、その際に考慮する「分野」とは、NDC何桁目、というように機械的に区切りをつけられるものではなく、除籍検討の機会に応じて、下位 (または上位) 分野や並列分野とまとめて一つと考えたり、逆に同じ分類番号の中でも分けて考えたりすることがある流動的なものである。後者の例としては、特定の手術手技や、特定の治療薬に関する資料というものを想像してもらえるとわかりやすいと思う。また、先に挙げた薬学分野のように、各分野にはそれぞれの個別の事情が存在する。つまり、桁数などで分類番号を一律に区分したとしても、その粒度は実際の除籍判断の区分とは異なり、各区分に含まれる事情も均すことができない。このため、これらの区分間の資料数のみを比べたとしても、そこに (広義の意味での) 線形性を持った除籍傾向を見いだすことはできないのである。

NLMCのようにより分類の階層や粒度が揃っている分類体系で同じ分析を行った場合に結果がどのように変化するかは興味のあるところである。ただし、分類体系がどうであったとしても、分類には、これまでの考察内で述べたような、あるいは本稿で明確に考察しきれない要素や事情が含まれており「代替資料がある」という除籍基準を、単純にある分野の蔵書数という要素で説明することはできないものと思われる。

V. 結論

1. まとめ

所蔵データから、資料が持つ除籍判断と関わるデータ的に分類・計測可能な要素を抽出することができた。また、それらの要素と除籍との関係を数値的に可視化することもできた。

「複本があること」「改訂版であること」は除籍リスク増大の要素となっており、関連する除籍基準と矛盾しない結果だった。分類としては「基礎医学分野や薬学分野の資料であること」がリスク増大に、「眼科・耳鼻科分野の資料であること」がリスク減少に関連しているという結果で、これも本学の除籍の事情にマッチしていた。

一方で「利用頻度が低い」「代替資料がある」という除籍基準の影響をデータ的に確認することはできなかった。ただし、どちらについても、それらが結果に現れない理由をデータの性質や除籍の事情から類推可能であった。

2. 分析の限界と注意点

可視化された除籍傾向は、除籍の基準や事情と矛盾しないものだったが、可視化できていない要素もあるため、この分析を除籍の偏りチェックを目的として行う場合には注意が必要である。自機関のデータや事情に通じた者が分析に加わり、一定の統計的慎重さを持って結果を解釈すべきと考える。

今回の調査で使用した分類はNDCである。NLMCを用いて分類関係の分析を行うと結果は違ってくるものと思われる。

利用回数については、今回分析に使用した「資料を受け入れた翌年度の利用回数」では除籍に目立った影響を与えていなかったが、除籍選定時には、直近数年の利用回数を調べて参考にするとはごく普通に行われているため、この分析とは別の形で利用回数をカウントすれば、要素化できる可能性がある。

今回の調査では推測統計の手法を用いた。今回の標本からその性質を予測した母集団は、同質の資料構成と除籍基準内での除籍作業が行われている資料群ということになる。本センターの除籍に関してはこの知見を援用できるが、この結果をそのまま一般化して直接他の機関の除籍に適用することはできない。ただし、この分析自体は他の機関でも行えるため、他の機関がその機関のデータを用いて同様の分析を実行すれば、その機関としての結果が得られる。

また、今回の調査で算出されたHRはあくまでこの約10年という観測期間内の重み付け平均である。観測期間内でもHRは変動し、異なる期間で観測した場合にはHRは違った値となる。なお、HRの解釈や用法については、短くまとまった論考がJAMAに掲載されている¹⁰⁾。

3. 今後の展望

今回確認できた要素などを使って、資料の除籍可能性を予測するモデルを構築することを検討する。一般的かつ計算が容易なものとして、例えばロジスティック回帰モデル¹¹⁾などが利用できるのではと考えている。

その後、モデルを実際のデータと比較して、その適合度などを確認していき、最終的にはそのモデルを使って除籍作業を支援できるようになればと考えている。

参考文献

- 1) Cox D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972;34(2):187-220.
- 2) Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone marrow transplantation*. 2013;48(3):452-8.
- 3) 小野めぐみ, 金廣康子, 清水浩子, 安藤美紀, 島貫裕美子. 複本除籍大作戦: 明治薬科大学図書館における資料廃棄の取り組み. *大学図書館研究*. 2021;118:1-9.
- 4) 山田稔. 予算・書店・選書・除籍: 愛知淑徳大学図書館の事例. *館灯*. 2009;47:76-81.
- 5) 岡田信恵, 大野圭子, 榊原佐知子, 田原美奈子. 東海地区医学図書館協議会加盟館における除籍(廃棄)の現状調査について. *薬学図書館*. 2014;59(1):7-17.
- 6) 谷口豊. ウィーディングによる蔵書管理-書架スペース捻出のための廃棄・除籍・保管・移管. *図書館雑誌*. 1999;93(3):185-7.
- 7) 香田友美恵, 村瀬由紀子, 安藤美紀. 朝日大学図書館における除籍資料の統計的分析. *医学図書館*. 2016;63(2):147-50.
- 8) 高畑亜紗美, 吉新裕昭, 大谷裕. 「集書・除籍に関するJMLA加盟館アンケート」調査結果. *医学図書館*. 2012;59(4):284-9.
- 9) Medical Library Association Collection Development Caucus. *Health Sciences Collection Development: An Overview of Fundamental Knowledge and Practices (2nd Edition)*[internet]. <https://pressbooks.rampages.us/hscolldev/> [accessed 2023-05-21]
- 10) Stensrud Mats J., Hernán Miguel A. Why Test for Proportional Hazards?. *JAMA*. 2020;323(14):1401-2.
- 11) Cox D. R. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1958;20(2):215-42.

An Exploratory Analysis of Factors Associated with Decisions of Weeding Library Collections Using a Cox Proportional Hazards Model

Satoshi HASHIMOTO

Toho University Medical Media Center, Ohashi Medical Center Library. 2-22-36, Ohashi Meguro-ku, Tokyo 153-8515, Japan

Abstract: In this study, survival time analysis (Cox proportional hazards analysis) was conducted to identify elements of the material related to weeding decisions. To this end, the 10-year status of book materials (purchased in 2011) in the library system was examined. The number of materials analyzed was 1,934, of which 1,500 were remained in the collection, 419 were removed as unnecessary, and 15 were removed due to defacement or other reasons. After adjustments, the following elements were identified as factors associated with an increased risk of removal: “duplicate existence (HR: 3.87 [95%CI, 2.92-5.13])”, “revised edition (HR: 4.95 [95%CI, 4.01-6.09])”, and “materials in the fields of basic medicine (HR: 1.46 [95%CI, 1.13-1.89]) and pharmaceutical sciences (HR: 3.96 [95%CI, 2.68-

5.86])”. Conversely, “materials in the fields of ophthalmology and otolaryngology (HR: 0.34 [95%CI, 0.14-0.82])” was identified as being associated with a decreased risk of removal. These results are in agreement with our institution’s weeding criteria and circumstances. It should be noted that the outcomes are particular to this institution and may differ depending upon the institution conducting the examination. Further research based on this knowledge for construction of a model for predicting removal would be desirable.

Keywords: Weeding Decision, Survival Analysis, Cox proportional Hazard Model, Exploratory Research, Medical Library (*Igaku Toshokan*. 2023;70(2):101-108)