

# 科学技術文献情報への索引語付与例

科学技術文献情報に対して、主題を表す索引語（メインヘディング）を統制語語彙（JST科学技術用語シソーラス）、統制語と関連づけした拡大語彙（大規模辞書）より選択して  
※下記では統制語語彙をシソーラス用語、拡大語彙を準シソーラス用語として表示

和文標題：広範な血栓に対する線溶療法

抄録：症例は40歳男性で、ST上昇型心筋梗塞と診断され、冠動脈血管造影で広汎な血栓が確認された。冠動脈血管造影で広範な血栓を伴うアテローム動脈硬化性プラークが見られた。モンテプラゼによる線溶療法、抗血小板薬2剤併用療法を行ったところ、内腔径の拡大と血栓の減少を確認できた。

分類コード：GJ01030V (616.1-071)

シソーラス用語：ヒト, 症例報告, 成人, 男性, \*急性心筋梗塞/\*薬物療法(DT)/  
診断(DI), \*薬物療法, t-PA【酵素】/治療利用(TU), 血栓溶解薬/治療利用  
(TU), X線CT, 血小板凝集阻害薬/多剤併用(DC), 併用効果, 冠動脈造影

準シソーラス用語：CTA, モンテプラゼ/治療利用(TU), \*線溶療法

索引

- 科学技術分類コード：上位分類から下位分類まで8桁で表す科学技術分野の分類
- メインヘディング（MH）：JSTの科学技術用語辞書に基づく索引語
- サブヘディング（SH）：国内の医学分野に特有の索引語でメインヘディングと組み合わせて索引（26種類）

# メインヘディングへのサブヘディング付与

医学・薬学分野については、文献情報の主題を表すメインヘディングに加えて、メインヘディングと組み合わせることでメインヘディングの視点を表現するサブヘディングを付与

- 骨粗鬆症が見つかり、**アレンドロネート**5mg経口の毎日製剤の投薬を開始し、その後35mgの週1製剤に変更し継続的に投薬。
  - アレンドロネート/治療利用 …アレンドロネートを疾患の治療に使用したことを表現
- 当センターにおける筋骨格系外傷**術後深部感染症**例の検討。成功例と失敗例に分けて、起因菌、内科的合併症、感染症状出現後から初回デブリドマンまでの期間、使用した抗菌薬の種類などについて調査を行い比較した。
  - 術後感染症/薬物療法 …術後感染症を薬物により治療したことを表現
- 良性疾患であっても本疾患のような重篤な病態を来すことがあり、**腹腔鏡下手術**時の腫瘍の回収には腫瘍を飛散させない工夫が必要であると考えられた。
  - 腹腔鏡下手術/有害作用 …腹腔鏡下手術の副作用を扱っていることを表現

# サブヘディングの種類

サブヘディングには、主に疾患名のメインヘディングにつくもの、主に医薬品・治療法のメインヘディングにつくものがあります。緑色セルの5種類はJSTが評価用の正解データを保有しているサブヘディング。

No	サブヘディング	表現内容	主な対象MH	MHの 카테고리
1	薬物療法	疾患への対処	疾患名	LS51
2	食事療法	疾患への対処	疾患名	LS51
3	放射線療法	疾患への対処	疾患名	LS51
4	外科的療法	疾患への対処	疾患名	LS51,LS16
5	治療	疾患への対処	疾患名	LS51
6	予防	疾患への対処	疾患名	LS51
7	診断	疾患への対処	疾患名	LS51
8	リハビリテーション	疾患への対処	疾患名	LS51
9	化学誘発	疾患の原因	疾患名	LS51
10	病因	疾患の原因	疾患名	LS51
11	合併症	疾患の状態	疾患名	LS51
12	予後	疾患の状態	疾患名	LS51
13	転移性	疾患の状態	疾患名	LS51
14	疫学	疾病の分析	疾患名	LS51
15	病理	疾病の分析	疾患名	LS51,LS16
16	死亡率	疾病の分析	疾患名	LS51
17	遺伝学	疾病の分析	疾患名	LS51

No	サブヘディング	表現内容	主要MH	MH 카테고리
18	治療利用	薬物の利用目的	薬物・化学物質	CA03,05,06,LS06,07,LS33,34,35,36,37,38,42,44
19	多剤併用	薬物の利用目的	薬物・化学物質	CA03,05,06,LS33,34,35,36,37,38,42,44
20	有害作用	薬物の影響	薬物・化学物質	CA03,05,06,LS06,07,33,34,35,36,37,38,42,44,52,54,65,72
21	薬物相互作用	薬物の作用	薬物・化学物質	CA03,05,06,LS32,33,34,35,36,37,38,42,44,72
22	薬理学	薬物の作用	薬物・化学物質	CA03,CA05,CA06,LS33,LS34,LS35,LS36,LS37,LS38,LS42,LS44
23	薬物動力学	薬物の作用	薬物・化学物質	CA03,05,06,LS33,34,35,36,37,38,42,44
24	内因性	化学物質の由来	薬物・化学物質	CA03,05,06,LS11,15,33,34,35,36,37,38,42,44
No	サブヘディング	表現内容	主な対象MH	MH 카테고리
25	分析	疾患と体内物質の関係	薬物・化学物質/疾患名	CA03,05,06,LS32,33,34,35,36,37,38,44,51
26	血液分析	疾患と体内物質の関係	薬物・化学物質/疾患名	CA03,05,06,LS32,33,34,35,36,37,38,44,51

※表中、メインヘディングはMHと略記

# メインヘディングのカテゴリーコード

全てのメインヘディングについて最大6種類のカテゴリーコードが定義されています。

## 基礎科学分野

コード	カテゴリー名称
CA01	化学一般
CA03	元素
CA04	分析化学
CA05	無機化合物
CA06	有機化合物

## ライフサイエンス分野

コード	カテゴリー名称
LS01	生物学一般
LS03	生物学実験方法・技術・装置
LS04	生物
LS05	動物
LS06	植物
LS07	微生物・ウィルス
LS11	生理学一般
LS12	動物生理
LS14	発生・成長

コード	カテゴリー名称
LS15	血液学
LS16	組織・器官
LS17	細胞学
LS18	遺伝学
LS19	免疫学
LS31	生化学一般
LS32	生体成分一般
LS33	蛋白質・ペプチド
LS43	核酸
LS35	脂質
LS36	ビタミン・補酵素
LS37	ホルモン
LS38	酵素

コード	カテゴリー名称
LS41	薬学一般
LS42	製剤
LS43	薬理
LS44	薬物
LS45	毒物学
LS51	病気・病理・症状
LS52	医学・医療行為
LS53	医療制度・設備
LS54	医療用品
LS55	公衆衛生
LS56	獣医学
LS65	農薬学
LS72	食品
LS73	栄養学
LS74	発酵

# サブヘディング付与の重要な課題(その1)

サブヘディング自動付与に向けたJSTでの取り組みにおいて、顕在化している重要な課題として、少なくとも以下A～Cのようなものが挙げられます。

課題A) 疾患名のメインヘディングおよび同義語（大規模辞書にて定義）と一致する表記が文献データに含まれていない場合についても当該の疾患名に対するサブヘディング付与を適切に判定する。

下記の文例では、メインヘディング「細菌感染症」と一致する表記が含まれていないが、「細菌感染症」にサブヘディング「薬物療法」を付与する。

29歳男性。受診3日前より発熱を認め、近医を受診した。S5-8に78.3×57.4mmの腫瘤性病変を指摘され、当院紹介受診した。精査にて、肝膿瘍と診断された。同日、エコーガイド下経皮経肝膿瘍ドレナージを施行し、膿汁と血液培養より、*Streptococcus intermedius*を検出した。炎症所見認め、投薬とドレナージにて第10病日、発熱や炎症反応が改善し退院となり再燃はしていない。

注：赤文字はサブヘディング付与の判定に重要と想定される文脈中の用語

適切なサブヘディング付与のパターン

■ 細菌感染症/薬物療法、抗細菌薬/治療利用

※メインヘディングに組み合わせて付与するサブヘディングをメインヘディング/サブヘディングの形式で表記  
(以下、同様)

# サブヘディング付与の重要な課題(その2)

課題B) 薬効名のメインヘディングおよび同義語（大規模辞書にて定義）と一致する表記が文献データに含まれていない場合についても、当該の薬効名に対するサブヘディング付与を適切に判定する。

下記の文例では、薬物名のメインヘディング「テノビルDF」と一致する表記が含まれているが、薬効名の「抗ウイルス薬」と一致する表記が含まれていない。薬物名「テノビルDF」と薬効名「抗ウイルス薬」にサブヘディング「治療利用」を付与する。

**B型慢性肝疾患**に対する **テノビルDF** における HBs 抗原低下作用について検討することを目的とした。B型慢性肝疾患と診断され、テノビルDFが導入された365例の中で、テノビルDFが24週以上投与された新規導入症例24例とテノビルDFへ切替え導入された169例を対象とし、初期治療効果とHBs抗原量低下予測因子について検討した。テノビルDFにおけるHBs抗原量低下予測因子として、新規導入では有意な因子を認めず、切替え導入群ではテノビルDF切替え時のHBsAg量高値とALT高値とが関連していた。また、短期ではあるがテノビルDFの投与期間中に重篤な有害事象の出現はなく安全に投与可能であった。

適切なサブヘディング付与のパターン

- B型慢性肝炎/薬物療法、テノビルDF/治療利用、抗ウイルス薬/治療利用  
(参考) テノビルDFの薬効名は関係語 (RT) として位置づけられている。  
テノビルDF            RT 抗ウイルス薬  
                             RT 抗エイズ薬



# サブヘディング付与の重要な課題(その3)

課題C) 1文献のなかで複数の疾患（合併症、続発性疾患、独立な複数疾患）が扱われており、薬物が投与されている場合についても、薬物と対象疾患の関係性を踏まえて、疾患名のメインヘディングと薬物名・薬効名のメインヘディングに対応するサブヘディング付与を適切に判定する。

下記の文例では、疾患名「B型肝炎」に薬物名「エンテカビル」、薬効名「肝臓病薬」、「抗ウイルス薬」を投与。疾患名「直腸癌」に薬物名「カペシタビン」、「オキサリプラチン」、薬効名「抗腫瘍代謝拮抗物質」、「抗腫瘍アルキル化剤」を投与している。

癌化学療法を実施する上で、B型肝炎ウイルスの再活性化が問題となっており予防策を講じる必要がある。今回、固形癌治療でのde novo B型肝炎を未然に防ぐことができた症例を経験したので報告する。77歳男性、直腸癌術後、補助化学療法のためCapeOX療法（カペシタビン、オキサリプラチン）が開始された。定期的にHBV-DNA定量を行い、2015年6月、HBV-DNAが2.4 log copies/mlとなり主治医に連絡、エンテカビル（ETV）が開始となった。同年8月に化学療法終了したが、ETVは継続しHBV-DNAは検出されず、現在に至っている。

適切なサブヘディング付与のパターン

- B型肝炎/予防、エンテカビル/薬物療法、肝臓病薬/治療利用、抗ウイルス薬/治療利用  
直腸癌/薬物療法、カペシタビン/多剤併用、オキサリプラチン/多剤併用、抗腫瘍代謝拮抗物質/多剤併用、抗腫瘍アルキル化剤/多剤併用（抗腫瘍薬/多剤併用）
- ※直腸癌に対してカペシタビン、オキサリプラチンが多剤併用されているため、サブヘディングとして治療利用ではなく、多剤併用を付与する。

# 精度評価

JSTにて実施している精度評価の方法と精度目標値を参考例として以下に記載します。

## (1) 学習精度

下記の○、×を判定してF値（適合率、再現率）を算出して交差検証を行う。

- ：必要なサブヘディングが付与された
- ×：必要なサブヘディングが付与されなかった  
不要なサブヘディングが付与された

## (2) 判定精度

評価用データに対してサブヘディング付与を行い、正解データと比較して、下記の○、×を判定してF値（適合率、再現率）を算出する。

- ：必要なサブヘディングが付与された
- ×：必要なサブヘディングが付与されなかった  
不要なサブヘディングが付与された

JSTではサブヘディングの自動付与を文献情報のデータ作成において利用するためにはF値80%以上を達成すること、さらに評価用データに特化した精度でなく、医学・薬学文献一般について同等の精度を達成することが必要であると考えています。



# 貸与データの種類と規模 (2021年10月6日現在)

- ① 標題・抄録データ  
1981年より蓄積している国内医学薬学文献の標題、標題＋抄録: 1,052万文献
- ② サブヘディング索引データ  
上記①のうち349万文献
- ③ 医学薬学分野の全文テキストデータ  
2019年1月以降に会議要旨集の全文テキストデータ化の蓄積を開始: 19.1万文献  
うち、人手によりサブヘディング付与済: 8.7万文献  
※従来からの人手によるサブヘディング付与は標題、抄録だけでなく、全文も参照しながら行っている。
- ④ JST科学技術用語シソーラス  
科学技術分野を網羅する約3万語規模の日本語シソーラスで、表記の統制のほか、上位・下位語関係、関連語関係などを保持
- ⑤ JST大規模辞書  
シソーラスにてカバーしきれない専門性の高い用語、利用頻度の低い用語、表記ゆれなどを保持する約100万語規模の辞書
- ⑥ 精度評価用データ  
スライド「サブヘディング一覧」のうち、「薬物療法」、「放射線療法」、「化学誘発」、「治療利用」、「血液分析」のサブヘディングについて、それぞれ100文献程度の正解データ

# シソーラスと大規模辞書の関係

シソーラスは、意味概念の広さ、厳密さで用語間に階層をもたせています。大規模辞書は、そのシソーラス用語も含みますが、同義語やシソーラス用語よりも狭い概念などを整備した辞書です。シソーラス用語の同義語ではない用語群は、直近上位語のシソーラス用語と紐付けています。

