

令和3年度 厚生労働行政推進調査事業費補助金
(医薬品・医療機器等レギュラトリーサイエンス政策研究事業)
分担研究報告書

医療情報データベースの活用推進に関する研究

研究分担者 中山 雅晴・東北大学大学院医学系研究科 教授

研究要旨：

目的：効率的なアウトカム定義の作成手法の検討及び実用化可能なアウトカム定義を確立するため、「薬物治療を要する甲状腺機能低下症(原因は問わない)」および「治療を要する間質性肺炎」対象とした本研究を実施した。

方法：甲状腺機能低下症に関する初期条件では、2020/12/14～2021/3/31に該当病名があり、甲状腺機能低下症治療薬の投与がある患者とした。一方、間質性肺炎に関しては、間質性肺炎に関連する病名の開始日があり、シアル化糖鎖抗原(KL-6)または肺サーファクタントプロテイン D(SP-D)を測定し、間質性肺炎に対する薬物治療を行っている患者を対象とし、データ期間は2018/12/7～2019/3/31とした。いずれも診療録を確認して真のケースかどうかを判定し、陽性的中度(PPV)を算出した。更に、甲状腺機能低下症では、機械学習によりPPVに影響する要因を検討し、PPV改善を目的に改良型アウトカムを作成した。

結果：甲状腺機能低下症では、初期条件により抽出された194例を調査したが、真のケースは56例、PPV28.9%であった。機械学習では当初目立った要素が認められなかったため、検査値異常(FT4低値)の条件を追加したところ、FT4低値が非常に強く影響しており、真のケース55例中44例がFT4低値であるとともに、FT4低値はすべて真のケースだった。この条件により、陽性的中度は100%、感度は80%となることを見込まれた。間質性肺炎では、初期ルールに該当した例は125例(男性56.8%、平均年齢63.8±14.2歳)であった。真のケースは53例、疑わしいケース27例、その他のケース45例であり、PPVは42.4%であった。疑わしいケースを真のケースとして扱った場合は、PPVは64.0%であった。

結論：甲状腺機能低下症における初期ルールではPPVが低かったが、機械学習によりFT4低値を抽出条件に追加すると陽性的中度は顕著に改善した。他の要素によらず、FT4低値のみを抽出条件とすることを提案する。間質性肺炎に関しては、治療を要する間質性肺炎に対する初期の抽出条件では、真のケースは半数程度であったが、治療薬等の抽出条件の調整により、PPVの上昇が見込まれた。

A. 研究目的

製造販売後の医薬品安全性評価は、従来、副作用報告、使用成績調査等の結果に基づくことが主であったが、医療情報データベースの整備等によりリアルワールドでの大規模データに基づく評価が可能となりつつある。

MID-NET[®]は、厚生労働省の医療情報データベース基盤整備事業により構築されたデータベースで、独立行政法人医薬品医療機器総合機構(以下「PMDA」という。)による運用管理の下、平成30年度から運用を開始している。MID-NET[®]は、診療情報明細書(レセプト)、診断群分類(DPC)及び電子カルテの情報が利用可能で、現在では行政による活用の他、臨床研究や製造販売後データベース調査にも活用されており、製造販売後の医薬品安全対策の質の向上が期待されている。

医療情報データベースに基づく医薬品の安全

性等の評価では、データベースに含まれる情報から対象とする有害事象(アウトカム)を適切に特定するために、信頼できるアウトカム定義を作成する必要があるが、本邦においてアウトカム定義のバリデーションが実施された例はほとんどない。「MID-NET[®]データの特性解析及びデータ抽出条件・解析手法等に関する研究」(日本医療研究開発機構 医薬品等規制調和・評価研究事業)(以下「先行研究」という。)において、MID-NET[®]を対象にアウトカム定義の作成及びその妥当性の評価を効率的に実施するための検討を実施し、基礎的な検討手法を確立するとともに、バリデーションされた複数のアウトカム定義が作成された。しかしながら、実用化可能なアウトカム定義を増やすためには、更なる検討手法の効率化や具体的なアウトカム定義の作成を継続的に実施する必要がある。

本研究は、先行研究の成果を踏まえて、研究

の流れを見直し All possible cases の定義を決定した上で以降の検討を行う等の検討手法の改善及び実用化可能なアウトカム定義の確立を目指し、医薬品安全性評価における医療情報データベースの活用促進と、より科学的な根拠に基づく安全対策の実現に繋げることを目的とした。

B. 研究方法

安全対策上の必要性や重要性を考慮し、検討の対象とするアウトカムを複数選定した上で、各アウトカムについて、従来法又は機械学習の手法を取り入れて作成したアウトカム定義について、複数医療機関にて妥当性の評価を行った。

研究の流れは以下のとおりである。

- 1)複数拠点で検討対象とする All possible cases の定義及びアウトカム定義の検討
- 2)評価基準の作成
- 3)対象アウトカムについてカルテレビューにより真の症例を特定
- 4)機械学習及び従来法によるアウトカム定義の作成
- 5)各アウトカム定義について、陽性的中度(以下「PPV」という。)及び感度の算出並びに評価
- 6)複数医療機関の PPV の比較及び医療機関間の差異の要因検討

(倫理面への配慮)

医療機関が行う作業は、文部科学省・厚生労働省.人を対象とする医学系研究に関する倫理指針(平成26年文部科学省・厚生労働省告示第3号)を遵守して行った。

C. 研究結果

一つ目のテーマは「薬物治療を要する甲状腺機能低下症」であるが、以下の理由により、検討の対象は「薬物治療を要する甲状腺機能低下症のうち、臨床所見の有無を問わず臨床検査値異常が認められる症例」とした。理由①薬物治療を要しない未診断かつ無症候性の潜在性甲状腺機能低下症を医療情報データベースで網羅的に捕捉することは困難であること。理由②薬剤性の甲状腺機能低下症には、臨床所見を伴わず臨床検査値異常が認められる症例も含まれること。理由③薬剤性の甲状腺機能低下症の発現を評価するデータベース調査においては、甲状腺機能低下症に対する薬物治療により臨床検査値がコントロールされている症例(臨床検査値異常が認められない症例)はアウトカムの評価対象にならないと現時点で想定されること。

データ抽出対象期間は、2020年12月14日か

ら2021年3月31日とした。

対象疾患は表1の通りで、原因は問わない。例としては、原発性甲状腺機能低下症(橋本病等)、二次性甲状腺機能低下症(下垂体性、薬剤性、甲状腺手術等)などが該当する。表2にはアウトカム定義、図1に判定フローをそれぞれに記した。

表1 対象疾患分類・ICD-10コード

ICD10コード	分類
E01	ヨード欠乏による甲状腺障害及び類縁病態
E02	無症候性ヨード欠乏性甲状腺機能低下症
E03	その他の甲状腺機能低下症
E04	その他の非中毒性甲状腺腫
E06	甲状腺炎
E07	その他の甲状腺障害
E350	他に分類される疾患における甲状腺障害
E890	治療後甲状腺機能低下症
O905	分娩後甲状腺炎
O992	妊娠、分娩及び産じょくく褥>に合併する内分泌疾患、栄養疾患及び代謝疾患
R946	甲状腺機能検査の異常所見
R947	その他の内分泌機能検査の異常所見
T381	ホルモン類、その合成代替薬及び拮抗薬による中毒、他に分類されないもの、甲状腺ホルモン類及びその代替薬
T382	ホルモン類、その合成代替薬及び拮抗薬による中毒、他に分類されないもの、抗甲状腺薬
Y421	治療上の使用により有害作用を引き起こした薬物、薬剤及び生物学的製剤、ホルモン類及びその合成代替薬及び拮抗薬、他に分類されないもの、甲状腺ホルモン及びその代替薬
Y422	治療上の使用により有害作用を引き起こした薬物、薬剤及び生物学的製剤、ホルモン類及びその合成代替薬及び拮抗薬、他に分類されないもの、抗甲状腺薬

表2 アウトカム定義【甲状腺機能低下症】

定義No	アウトカム定義	参考(定義のスクリーン構成)
APC	[病名1]、[病名2]又は[病名3]の([病名1] or [病名2] or [病名3]) いずれかに該当し、index_dateの30日前から病名の終了日又は退院年月日の30日後以内に[医薬品1]に該当する	時系列and [医薬品1]
病名1	DPC傷病情報(6項目*のいずれか)に甲状腺機能低下症に関する病名がある。	
病名2	SS-MIX2傷病情報(疑い含む)に甲状腺機能低下症に関する病名がある。	
病名3	レセプト傷病情報に甲状腺機能低下症に関する病名がある。	
医薬品1	SS-MIX2処方・注射情報、DPC医薬品情報又はレセプト医薬品情報にレボチロキシンナトリウム、リオチロンナトリウム又は乾燥甲状腺の処方がある。	

*主傷病名、入院の契機となった傷病名、医療資源を最も投入した傷病名、医療資源を2番目に投入した傷病名、入院時併存症名又は入院後発症疾患名。

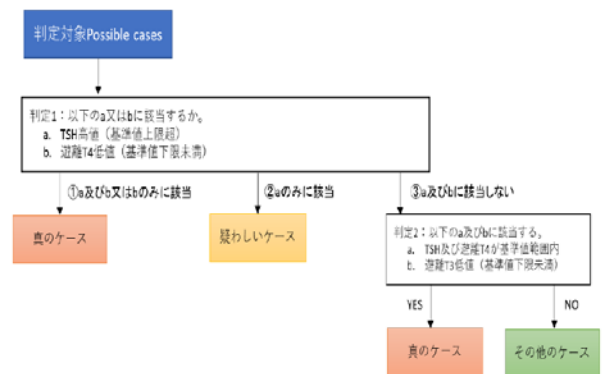


図1 判定基準-判定フロー【甲状腺機能低下症】

図 2 機械学習の結果

初期条件により抽出された 194 例を調査したが、真のケースは 56 例、PPV28.9%であった(表 3)。その他のケースとなった例には、チラーヂン等内服により検査値安定例(81 例)、Index date 周辺で検査がない例(24 例)、甲状腺機能亢進症でメルカゾールとチラーヂンを併用している例(3 例)がそれぞれ認められた。

表 3 PPV 算出と結果まとめ

甲状腺機能低下症 (APC)	真のケース	疑わしいケース	その他のケース	判定対象ポテンシャルケース
	件数	件数	件数	件数
	56	30	108	194

甲状腺機能低下症 (APC)	真のケース	疑わしいケース+その他のケース	判定対象ポテンシャルケース	陽性的中度
	件数	件数	件数	割合 (%)
	56	138	194	28.9

甲状腺機能低下症 (APC)	真のケース+疑わしいケース	その他のケース	判定対象ポテンシャルケース	陽性的中度
	件数	件数	件数	割合 (%)
	86	108	194	44.3

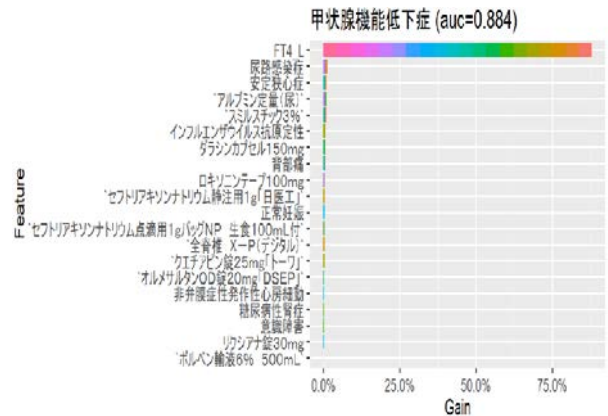


図 3 機械学習の結果(検査値異常を追加)

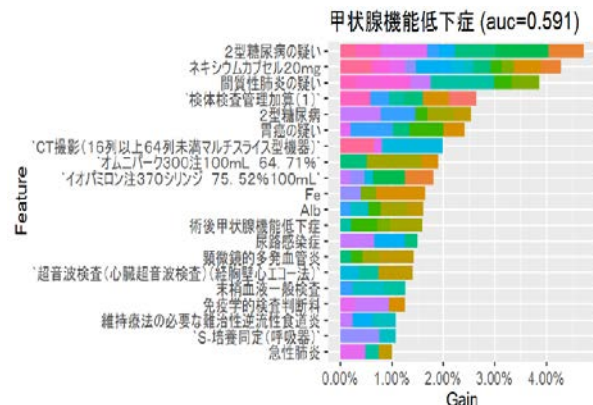
機械学習の結果から、FT4 低値が非常に強く影響しており、真のケース 55 例中 44 例が FT4 低値であるとともに、FT4 低値はすべて真のケースであることが判明した。この条件により、陽性的中度は 100%、感度は 80%となることが見込まれた(表 4)。

表 4 機械学習の感度

PPV	100%
感度	80.0%
特異度	100%

次いで全データを投入して機械学習を実施した。機械学習アルゴリズムとしては、勾配ブースティング(Gradient Boosting Decision Tree, GBDT)を用いた。具体的には、統計解析ソフト R バージョン 3. 5.1 のパッケージの XGboost を用いて機械学習を行った。ROC 解析については、pROC パッケージを使用し、Area Under the Curve (AUC) を評価した。

病名、薬剤、診療請求項目を使用したところ、AUC は 0.591 であり、目立った要素が認められなかった(図 2)。そこで、検査値異常(FT4 低値)の条件を追加したところ、AUC は 0.884 となり、ほぼ FT4 低値で占められた(図 3)。



二つ目のテーマは「間質性肺炎」であるが、間質性肺炎が発症した場合、間質性肺炎に対する治療がされることが想定されるため、検討の対象とする具体的な事象は、「治療を要する間質性肺炎」とした。ただし、放射線肺炎、及び放射線性肺線維症については、放射線を原因とすることが明らかであるため、対象からは除外した。臨床ガイドライン等を参考に、間質性肺炎を発現した症例を網羅的に抽出するための条件を作成した。

データ期間は 2018 年 12 月 7 日から 2019 年 3 月 31 日とした。初期の検索条件(初期ルール)は、間質性肺炎に関連する病名の開始日があり、シアル化糖鎖抗原(KL-6)または肺サーファクタントプロテイン D(SP-D)を測定し、間質性肺炎に対する薬物治療を行っている患者とした。

より詳細には、表 5 にある、【 [病名 1]、[病名 2] 又は [病名 3] のいずれかに該当し、かつ、indexdate の 30 日前から病名の終了日又は退院

年月日の30日後以内に、[医薬品1]、[医薬品2]又は[医薬品3]のいずれかに該当する、かつ、indexdateの30日前から病名の終了日又は退院年月日の30日後以内に、[診療行為1]又は[診療行為2]のいずれかに該当する。】とした。

Indexdateは、表1の病名1~3の条件に該当した日付(SS-MIX2傷病情報(病名オーダ)に記録された開始日、DPC傷病情報に記録された入院年月日又はレセプト傷病情報に登録された診療開始日)とした。

表5 検索条件詳細

条件名	条件の内容
病名1	SSMIX傷病情報(疑い含む)に間質性肺炎に関する病名がある。
病名2	DPC傷病情報(主傷病名、入院の契機となった傷病名、医療資源を最も投入した傷病名、医療資源を2番目に投入した傷病名、入院時併存症名又は入院後発症疾患名のいずれか)に間質性肺炎に関する病名がある。
病名3	レセプト傷病情報に間質性肺炎に関する病名がある。
医薬品1	ステロイドの処方がある。
医薬品2	免疫抑制剤の処方がある。
医薬品3	抗線維化薬の処方がある。
検体検査1	シアル化糖鎖抗原(KL-6)の結果値が500U/mL以上である。
検体検査2	シアル化糖鎖抗原(KL-6)の結果値が450U/mL以上である。
検体検査3	肺サーファクタントプロテインD(SP-D)の検査値が110ng/mL以上である。
診療行為1	CT検査の実施がある。
診療行為2	CT読影料の加算がある。

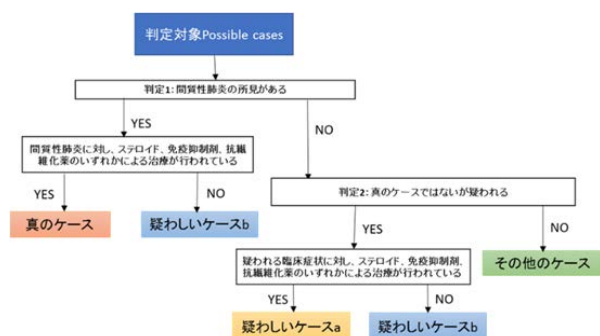


図4 真のケースの判定基準

MID-NET[®]から当該条件を用いて抽出された症例について、医師が診療録等の記録を確認し、臨床ガイドライン等を参考に作成した判定アル

ゴリズム(図4)に基づいて真のケースを判定し、PPV等を算出した。治療を必要としない間質性肺炎は、疑わしいケースに分類された。

表6 判定結果と陽性的中度

間質性肺炎	真のケース	疑わしいケース	その他のケース	判定対象ポテンシャルケース	
	件数	件数	件数	件数	割合(%)
	53	27	45	125	
間質性肺炎	真のケース	疑わしいケース+その他のケース	判定対象ポテンシャルケース	陽性的中度	
	件数	件数	件数	割合(%)	
	53	72	125	42.4	
間質性肺炎	真のケース+疑われるケース	その他のケース	判定対象ポテンシャルケース	陽性的中度	
	件数	件数	件数	割合(%)	
	80	45	125	64.0	

初期ルールに該当した例は125例(男性56.8%、平均年齢63.8±14.2歳)であった。真のケースは53例、疑わしいケース27例、その他のケース45例であり、PPVは42.4%であった(表6)。疑わしいケースを真のケースとして扱った場合は、PPVは64.0%であった(表6)。

真のケースには、特発性間質性肺炎だけでなく、薬剤性間質性肺炎や膠原病等に合併した間質性肺炎で、治療を追加した例も含まれていた。

その他のケースに該当した45例の内訳を調査したところ、本来は抽出対象外の放射性肺炎が5例、細菌性肺炎等の肺胞性肺炎が3例であった。これらにはいずれも間質性肺炎に関する病名が付与されていた。KL-6等を検査した際の疑い病名が付与された例も15例存在した。他の21例は、検査値の上昇は認められたものの、画像所見が認められない例であった。これらの21例では、治療薬は間質性肺炎治療以外の目的で使用されており、特に悪性腫瘍に対する化学療法施行中で、プロトコル内にステロイドが含まれている例が多く含まれていた。

疑わしいケースでは、新規あるいは追加の治療を必要としない間質性肺炎(図4)の疑わしいケースb)が23例と多かった。この中にも、間質性肺炎を合併した悪性腫瘍で、化学療法のプロトコル内にステロイドが含まれている例が多く含まれていた。他は、膠原病等に合併した間質性肺炎で、原疾患の治療中であり、間質性肺炎に対する追加の治療は不要と判断された例であった。

次いで、PPVを改善する条件を検討した。上

述の通り、疑わしいケースあるいはその他のケースには間質性肺炎を合併した悪性腫瘍の例が多く含まれていた。東北大学病院の化学療法プロトコルに含まれるステロイドはすべてデキサメタゾン注射であった。一方、間質性肺炎の治療に使用されるステロイドは、プレドニゾロンやヒドロコルチゾン等であり、デキサメタゾン注射は単独では用いられていなかった。化学療法に伴うデキサメタゾン使用例は、疑わしいケースで9例、その他のケースで21例あり、対象外とすることとした。この場合、真のケース53例、疑わしいケース18例、その他のケース24例となった。その結果、PPVは疑わしいケースをその他のケースとして扱うと55.8%、疑わしいケース a および b を真のケースとして扱うと74.7%となることが見込まれた。

一方、膠原病等の治療には間質性肺炎と同じくプレドニゾロン等が用いられていた。真のケースにも膠原病に合併した間質性肺炎の例が含まれており、病名や薬剤の種類による除外が困難であった。

D. 考察

薬物治療を要する甲状腺機能低下症の病名に関連する初期ルールでは ICD-10 コードを条件として抽出したが、PPVは28.9%と低く、多くの例がその他のケースとなっていた。機械学習により、病名や薬剤、診療請求項目を使用してみたが、目立った要素が認められなかった。そこで検査値に注目し、検査値異常(FT4 低値)の条件を追加したところ、AUCは0.884となり、この条件が有効であることが明らかとなった。今回の疾患の定義から考えても、検査値に着目することは合理的であり、その他のケースとなった大半は甲状腺機能低下症ではあるが数値が安定していた例であったことから、実際の効果も出ている。甲状腺機能低下症は、橋本病など中年女性に多く認められ、また高齢者では特別な症状がなくとも進行していく例も多い。スクリーニングとしての甲状腺ホルモン値測定は割と頻繁に行われ、偶発的に発見される例も多い。また、アミオダロンなどの薬物による副作用として発症することもあり、定期的なスクリーニングは今回の結果からも有効と考えられる。

間質性肺炎に関しては、病名や検査値異常、治療薬を含む初期の抽出条件による PPV が42.4%であった。その他のケースとなった例は、検査値のみに異常が認められ、画像所見が伴わない例が大半であった。疑わしいケースについては、ほとんどが検査値異常とともに画像所見

も認められるが、新規あるいは追加の治療を必要としない例であった。疑わしいケース、その他のケースのいずれにおいても、治療薬は間質性肺炎の治療以外の目的で使用されていた。

間質性肺炎は、多くが原因不明の特発性間質性肺炎であるが、薬剤性や放射線性の二次性間質性肺炎も少なくない。さらに、膠原病等の原疾患に合併する例も存在する。治療は薬物療法が基本的であり、ステロイドや免疫抑制剤を使用する。本研究では、疑わしいケースあるいはその他のケースとなった例には、膠原病等に対してステロイド等により治療しているが、間質性肺炎に対しては治療を要しない例や、間質性肺炎以外の疾患(主に悪性腫瘍)に対する治療目的でステロイドを使用している例が多く含まれていた。ステロイドは非常に多くの疾患に使用される薬剤であり、間質性肺炎の治療目的に投与されている例を区別するのは困難であった。

一方、ステロイドはパクリタキセルをはじめとした抗がん剤投与時にも用いられ、通常デキサメタゾン注射を使用する。実際、本研究でその他のケースとなった抗がん剤治療を受けた例で使用されたステロイド剤は、すべてデキサメタゾン注射であった。それに対し、間質性肺炎に対する治療目的で使用されたステロイドにはデキサメタゾン注射は含まれず、プレドニゾロンやヒドロコルチゾン等であった。そこで、デキサメタゾン注射を使用した例は除外することにより、抗がん剤治療例を抽出対象外とすることが可能であると考えられた。その結果、PPVは55.8%、疑わしいケースを真のケースとして扱うと74.7%となることが見込まれた。但し、デキサメタゾン注射が間質性肺炎に使用されないのは東北大学病院特有である可能性があり、保険適用があることから他施設では投与されていることは十分考えられる。このため、単純に本研究の結果を一般化し、他施設に適用することはできない可能性があり、本研究の研究限界として考慮する必要がある。

また、膠原病等に使用されるステロイドは間質性肺炎と同様に、プレドニゾロンやヒドロコルチゾン等であるため、薬剤の種類によって間質性肺炎と膠原病等を区別することはできず、膠原病等を抽出対象外とするのは困難であった。可能性として、病名開始日の前後で治療薬の増量があるかどうかなどを検討することにより、真のケースと疑わしいケースを区別することが有用かもしれない。

E. 結論

甲状腺機能低下症において、初期ルールでは PPV が低かった。機械学習において、FT4 低値がきわめて強く関連しており、抽出条件に追加すると、陽性的中度は顕著に改善した。他の要素によらず、FT4 低値のみを抽出条件とすることを提案する。

治療を要する間質性肺炎に対する初期の抽出条件では、真のケースは半数程度であった。治療薬等の抽出条件の調整により、PPV の上昇が見込まれた。

F. 健康危険情報：

(分担研究報告書では記入不要です)

G. 研究発表：

1. 論文発表

1. Tanigawa M, Kohama M, Nonaka T, Saito A, Tamiya A, Nomura H, Kataoka Y, Okauchi M, Tamiya T, Inoue R, Nakayama M, Suzuki T, Uyama Y, Yokoi H. Validity of identification algorithms combining diagnostic codes with other measures for acute ischemic stroke in MID-NET®. *Pharmacoepidemiology and Drug Safety*. 2022;31:524–533. DOI: 10.1002/pds.5423
2. Nakayama M, Inoue R. Electronic Phenotyping to Identify Patients with Heart Failure Using a National Clinical Information Database in Japan. *Studies in Health Technology and Informatics*. 2021; 281: 243-247. DOI: 10.3233/shti210157.
3. 中山雅晴, 永島里美, 大江和彦, 山下暁士, 白鳥義宗, 山下貴範, 中島直樹, 堤英樹, 東海林晋, 窪田成重, 厚生労働省標準規格 SS-MIX2 標準化ストレージのデータ品質改善の試み, *医療情報学* 2021 年 41 巻(Suppl), 975-977.
4. 中山雅晴, 宋翀, SS-MIX2 データを活用するための FHIR®ベース PHR の開発, *医療情報学* 2021 年 41 巻(Suppl), 370-371.

2. 学会発表

1. 厚生労働省標準規格 SS-MIX2 標準化ストレージのデータ品質改善の試み, 中山雅晴, 第 41 回医療情報学連合大会(第 22 回日本医療情報学会学術大会), 2021/11/21, 国内, 口演
2. SS-MIX2 データを活用するための FHIR®ベース PHR の開発, 中山雅晴, 第 41 回医療情報学連合大会(第 22 回日本医療情報学会学術大会), 2021/11/20, 国内, 口演
3. MID-NET を用いた動脈解離の検索精度に関する検討, 井上 隆輔, 中山 雅晴, 第 25 回日本医療情報学会春季学術大会, 2021/6/11, ポスター.
4. Electronic Phenotyping to Identify Patients with Heart Failure Using a National Clinical Information Database in Japan, Nakayama M, 31st Medical Informatics Europe Conference (MIE2021), 2021/5/31, 国外, Oral

H. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得
2. 実用新案登録
3. その他