

医療情報データベース解析における バイアスへの対応に関する研究

研究分担者 池田俊也（国際医療福祉大学薬学部教授）

研究要旨：国内外において、医療情報データベースを利用した医薬品の安全性評価が進められている、しかし、これらのデータベースを用いた分析は主に遡及的研究となるため、さまざまなバイアスが問題となる。そこで本研究では、糖尿病と発癌に関する分析を例として、群間の追跡期間の違いや属性の偏りから生じるバイアスに対応する方法について検討を行った。日本医療データセンター(JMDC)の保有するレセプトデータベースを用いて、Ⅱ型糖尿病患者の発がん率が、非糖尿病患者の発がん率と異なるかどうかを検証した。糖尿病患者では診断日、非糖尿病患者では台帳またはレセプト情報収集開始日を観察開始とした場合、発癌のなかった糖尿病患者で早期に打ち切りとなるケースが増え、結果にバイアスが生じることが予想される。また、年齢・性別・健康保険組合といった属性の偏りも懸念されることから、糖尿病患者1人につき同様の属性を持った非糖尿病患者を無作為に抽出し、Control Cohort とした。その結果、年齢・性別・健康保険組合・観察期間の偏りを補正してもⅡ型糖尿病患者と非糖尿病患者では無癌生存率に差があると考えられた。

研究協力者

佐藤敏彦 青山学院大学

A. 研究目的

医療情報が含まれるデータベースを利用した薬剤疫学研究は近年活発に行われており、米国や欧州の医薬品規制当局が医薬品の安全対策を講じる際の根拠ともなっている。最近では、糖尿病薬のピオグリタゾンの膀胱癌のリスクへの懸念からフランス、ドイツ等で同薬の新規処方を禁じる措置が取られ、その後、米国FDA (Food and Drug Administration、食品医薬品局)、欧州EMA (European Medicines Agency、欧州医薬

品庁)、日本では入手可能なデータの評価から治療中の膀胱癌患者への投与を避ける等の勧告が行われているが、これはフランスの医療情報データベース SNIRAM (System national interregimes de l'assurance maladie) のデータを用いたコホート研究である CNAMTS 試験において長期使用患者で膀胱癌発生リスクが高まるとの結果が示されたことをきっかけとしたものである。同研究ではSNIRAMの40～79歳の糖尿病患者1,491,060例の2006～2009年のデータを用いたリスク評価が行われている。

また、電子的な医療情報を医薬品などの安全性評価に役立てるためのシステムを構築する組

組織的な動きも米国、欧州等で見られており、米国 FDA では、レセプトや電子カルテ等の医療データベースのネットワークを利用して新たな医薬品などの市販後監視システムを構築する

Sentinel Initiative を 2008 年から開始している。欧州 EMA では、2006 年から医薬品の安全性とリスクベネフィットに関する質の高い市販後の薬剤疫学研究等を促進することによって市販後の医薬品のモニタリングを強化しようとするプロジェクト（ENCePP）を開始している。

我が国においても医療情報データベースの安全性評価への活用に向けた検討が進められており、厚生労働省及び PMDA で現在構築が進められている医療情報データベースは、10 の協力医療機関の保有するデータを安全対策のための医薬品安全性評価に利用しようとするものである。

これらのデータベースを用いた分析は主に遡及的研究となるため、さまざまなバイアスが問題となる。そこで本研究では、糖尿病と発癌に関する分析を例として、群間の属性の偏りや追跡期間の違いから生じるバイアスに対応する方法について検討を行った。

B. 研究方法

1. データベース

本研究では、日本医療データセンター(JMDC)の保有するレセプトデータベースを用いて、II 型糖尿病患者の発がん率が、非糖尿病患者の発がん率と異なるかどうかを検証した。今回使用したデータは 8 つの健康保険組合より提供を受け、少なくとも二年以上追跡できるデータであり、対象者（保険加入者）総数は 779,761 人であった。

2. 分析対象と除外対象

各対象期間内に診断名として'E10-E14(糖尿病)'を有する患者（疑い除く）を糖尿病患者とした。糖尿病患者の場合には、対象期間内の診断名が'E10-E14(糖尿病)'である最古の診療開始年月を観察開始年月とした。'E10（インスリン依存性糖尿病<IDDMM>）'をI型糖尿病とし、I型糖尿病診断の有無に関わらずE11-E14の診断がなされた患者をII型糖尿病とした。

観察期間内に'E10-E14(糖尿病)'の診断が一度もない加入者を非糖尿病患者とした。観察開始は各健保の台帳及びレセの収集開始年月とした。

年齢については、糖尿病患者、非糖尿病患者ともに観察開始時点の年齢とし、40代と50代を対象として解析を行った。

除外対象は以下の通りとした。

- (1) 糖尿病診療開始前から癌に罹患している場合
- (2) レセプト収集前から癌に罹患している場合
- (3) 癌と糖尿病を同時に診断された場合
- (4) 糖尿病の診療開始日が観察期間より前にある場合
- (5) 観察開始後 6 ヶ月間に発癌した場合

3. 追跡期間の違い等によるバイアスへの対応

糖尿病患者では診断日、非糖尿病患者では台帳またはレセプト情報収集開始日を観察開始とした場合、発癌のなかった糖尿病患者で早期に打ち切りとなるケースが増え、結果にバイアスが生じることが予想される。また、年齢・性別・健康保険組合の偏りによるバイアスも懸念される。

そのため、糖尿病患者 1 人につき同様の属性を持った非糖尿病患者を無作為に抽出し、Control Cohort とした。ここでは、性別・年齢

(40 歳以上 45 歳未満、45 歳以上 50 歳未満、50 歳以上 55 歳未満、55 歳以上 60 歳未満)・健康保険組合が一致するものを同一属性とした。なお、同じ属性をもった非糖尿病患者については、重複した抽出はされないものとした。

4. エンドポイント

全固形癌についての発症をエンドポイントとした。発癌しなかった場合にはレセ収集期間終了または資格喪失（死亡、退職など）にて観察打ち切りとした。固形癌の定義としては、下図のがんを非固形癌とし、それ以外の全悪性腫瘍を固形癌とした。

C81	ホジキン病
C82	ろく瀧>胞性[結節性]非ホジキンリンパ腫
C83	びまん性非ホジキンリンパ腫
C84	末梢性及び皮膚T細胞リンパ腫
C85	非ホジキンリンパ腫のその他及び詳細不明の型
C88	悪性免疫増殖性疾患
C90	多発性骨髄腫及び悪性形質細胞性新生物
C91	リンパ性白血病
C92	骨髄性白血病
C95	細胞型不明の白血病
C96	リンパ組織、造血組織及び関連組織のその他及び詳細不明の悪性新生物

Ⅱ型糖尿病患者では、診断日から発癌までの期間（ヶ月）を無癌期間とし、発癌しなかった場合には診断日から観察打ち切りまでの期間（ヶ月）を無癌期間とした。

非糖尿病患者では、台帳またはレセプト情報収集開始日から発癌までの期間（ヶ月）を無癌期間とし、発癌しなかった場合には台帳またはレセプト情報収集開始日から観察打ち切りまでの期間（ヶ月）を無癌期間とした

Control Cohort では、対となるⅡ型糖尿病患者における発癌がなかった場合の最長の観察期間（診断日から健保ごとのレセ収集終了期間：以下最長観察期間）の中で発癌があった場合には、レセ収集開始年月から発癌までの期間（ヶ月）を無癌期間とし、最長観察期間内で発癌しなかった場合には、個々のレセ収集開始年月から打ち切りまでの期間と最長観察期間のうち短いほうを無癌期間とした。

5. 解析方法

無癌率の推定値は上記で算出した個々の期間から、SAS 9.2 Lifetest Procedure を用いて Kaplan-Meier 法により分析集団ごとに算出した。

検定に関しては、Ⅱ型糖尿病患者と非糖尿病患者、Ⅱ型糖尿病患者と Control Cohort について Log-rank 検定を実施した（有意水準：両側 5%）。

さらに、Control Cohort の無作為抽出を 1,000 回繰り返し、各時点において無癌率及び信頼区間は正規分布するものとして群ごとに平均値を求め、その値を時点ごとの無癌率、信頼区間とする。検定については試行した結果のうち有意とならなかった割合を p 値とし、βエラーについて検討した。

（倫理面への配慮）連結不可能匿名化情報を用いており、臨床研究ならびに疫学研究に関する倫理指針の対象外であるが、データについてはパスワードをかけた CD-ROM で提供を受け、鍵のかかるロッカーに保管するなど、データの管理に万全の配慮を行った。

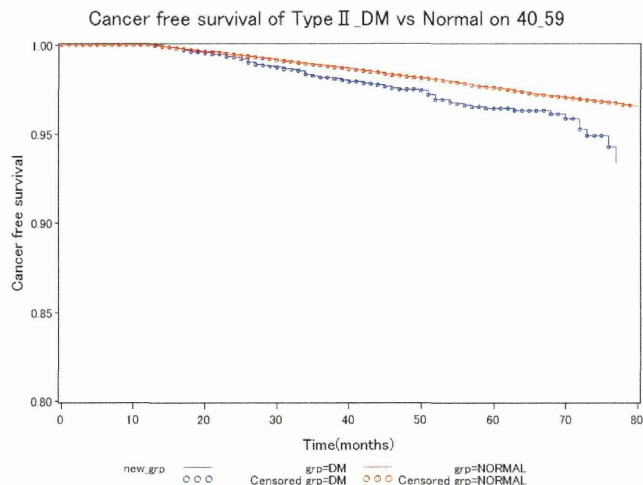
C. 研究結果

1. 対象患者の例数

Ⅱ型糖尿病患者の例数は 9,565 例、非糖尿病患者の例数は 129,180 例であった。

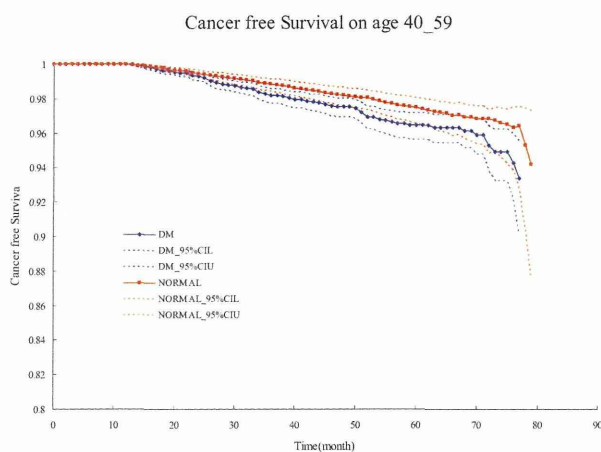
2. 両群の無癌率の比較（バイアスに対する対応なし）

以下の図の通り、Ⅱ型糖尿病患者の方が癌発症率が高く、log-rank test にて有意差が認められた ($p < 0.0001$)。



3. 両群の無癌率の比較 (Control Cohort を用いて属性と観察期間の偏りを補正)

結果は以下の図の通りⅡ型糖尿病患者の方が癌発症率が高い傾向が認められ、 $\alpha = 0.05$ とした時に有意な差が得られなかった確率は 23.8% であった。



D. 考察

糖尿病とがん発症の関連については、欧米の報告を中心としたメタアナリシスにより、関連

性がほぼ確立している。また、我が国においても、約 10 万人の男女を対象に 11 年間の追跡調査を行った国立がんセンターのコホート研究では、全部位平均で男性では 1.27 倍、女性では 1.21 倍との報告がなされている。

今回の解析結果では、40 歳以上 60 歳未満のⅡ型糖尿病患者と非糖尿病患者では、無癌生存率に有意な差が認められた。

Control Cohort を用いて追跡期間の整合を考慮した分析では、Ⅱ型糖尿病患者と Control Cohort では、 $\alpha = 0.05$ とした時に有意な差が得られなかった確率は 23.8%、即ち統計的検出力は約 80%であり、年齢・性別・健康保険組合・観察期間の偏りを補正してもⅡ型糖尿病患者と非糖尿病患者では無癌生存率に差があると考えられた。

E. 結論

データベースを用いた薬剤疫学研究の進展に伴い、バイアスへの対応についての研究が求められる。今回は追跡期間の差について調整する方法を試みたが、その他のバイアスへの対応も含め、今後、さらに適正な比較方法について検討する必要があると思われる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況

なし

III. 研究成果の刊行に関する一覧表

研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書 籍 名	出版社名	出版地	出版年	ページ

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Hori K, Yamakawa K, Yoshida N, Ohnishi K, Kawakami J.	Detection of fluoroquinolone-induced tendon disorders using a hospital database in Japan.	Pharmacoeconomics and Drug Safety	21	886-889	2012
若田好史、中島直樹、野原康伸	電子クリニカルパスにおけるオールバリエーション解析	医療情報学	32	62-65	2012
中島直樹、若田好史、野原康伸、井上創造、小妻幸男、副島秀久、田中雅夫	アウトカム志向型電子パスと生体センサを用いた探索的なクリティカルインディケータ抽出	第16回日本医療情報学会春季学術大会抄録集		84-85	2012

IV. 參考資料

第18回 日本薬剤疫学会学術総会 一般口演

医療情報データベースを用いたヘパリン起因性
血小板減少（HIT）検出アルゴリズムの構築

平成24年11月10日

○花谷 忠昭^{1),2)}、佐井 君江¹⁾、頭金 正博²⁾、瀬川 勝智¹⁾、
木村 通男³⁾、堀 雄史⁴⁾、川上 純一⁴⁾、斎藤 嘉朗¹⁾

- 1) 国立医薬品食品衛生研究所 医薬安全科学部
- 2) 名古屋市立大学薬学研究科 医薬品安全性評価学分野
- 3) 浜松医科大学医学部附属病院 医療情報部
- 4) 浜松医科大学医学部附属病院 薬剤部

日本のセンチネル・プロジェクトについて

厚生労働省に設置された「医薬品の安全対策等における医療関係データベースの活用方策に関する懇談会」がとりまとめた提言（平成22年8月）

- 医薬品等の安全対策の向上に関わる薬剤疫学研究等を実施するために利活用できるデータベースを構築する

⇒ まずは5年間で1,000万人規模のデータベースを目指す（平成27年目標）

⇒ 期待される成果として、

- ・ 医薬品等による副作用、安全上のリスクの抽出
- ・ リスクの定量的評価
- ・ ベネフィット/リスク バランスの改善のための企画・実施とその評価
- ・ 疫学研究を支援する情報基盤や研究者の育成

日本のセンチネル・プロジェクトについて

厚生労働省に設置された「医薬品の安全対策等における医療関係データベースの活用方策に関する懇談会」がとりまとめた提言（平成22年8月）

- 医薬品等の安全対策の向上に関わる薬剤疫学研究等を実施するために活用できるデータベースを構築する

⇒ まずは5年間で1,000万人規模のデータベースを目指す（平成27年目標）

⇒ 期待される成果として、

- ・ 医薬品等による副作用、安全上のリスクの抽出
- ・ リスクの定量的評価
- ・ ベネフィット/リスク バランスの改善のための企画・実施とその評価
- ・ 疫学研究を支援する情報基盤や研究者の育成

データベース構築前に
薬剤疫学的評価方法を
確立しておく必要がある

本研究の目的・内容について

● 目的

- ・ 浜松医科大学病院の医療情報データベース（D☆D）を用いて、日本のセンチネル・プロジェクトの先導的なパイロット研究を行う

⇒ 副作用のデータベース検索式の確立 と リスク因子の探索

● 対象副作用の選定

- ・ 厚生労働省の「重篤副作用疾患別対応マニュアル」に掲載され、年間の自発報告数が100件以上の副作用

⇒ **HIT（ヘパリン起因性血小板減少症）**を選定

⇒ よりリスクが高いとされる「未分画ヘパリン」を対象医薬品に選定

● 研究内容

- ・ 医療情報データベースからのHIT検出アルゴリズムの構築
- ・ HIT発症に影響を及ぼすと考えられているリスク因子の定量的評価

HIT について

(厚生労働省「重篤副作用疾患別対応マニュアル」より抜粋)

- ヘパリンによる軽度の血小板凝集作用の結果、血小板減少が引き起こされることが考えられている「Type I 型」と、一過性に出現するヘパリン依存性自己抗体が血小板を活性化するために血症板減少を引き起こす「Type II 型」に分類される
- 重篤な合併症を引き起こすのは「Type II 型」（以降、II 型をHITと呼ぶ）
- HITの約26～50%に血栓塞栓症を伴うため、血栓塞栓症状が見られる（HITT）
- 発症リスクは低分子ヘパリンよりも未分画ヘパリンの方が10倍高い
- HITの発症率は内科領域において約0～3.5%、外科領域において2.7～5.0%
- 過去100日以内にヘパリンを使用している場合、再使用后24時間以内に発症する。通常の初回治療時には使用開始5～14日後に血小板減少等が認められる

HIT 検出アルゴリズム

(1) 対象患者の選択（組入れ基準）

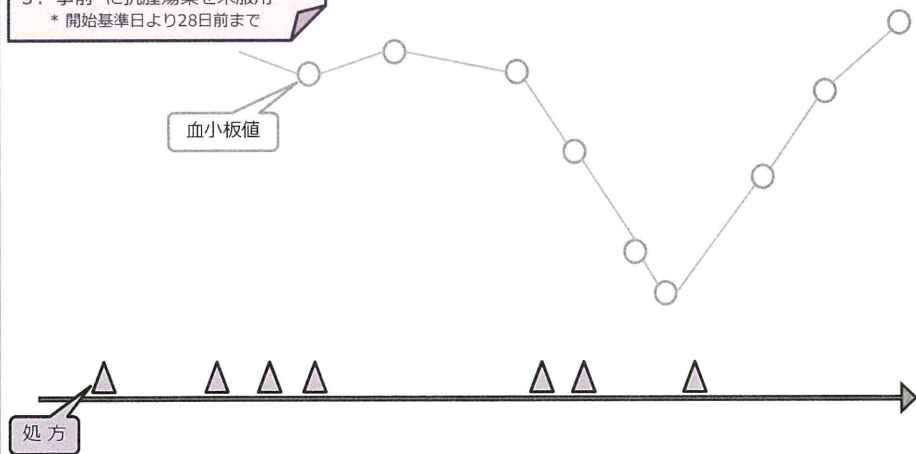
(2) 検査値によるHIT疑い症例の判定

(3) 病名情報による類似疾患との鑑別

HIT 検出アルゴリズム

(1) 対象患者の選択

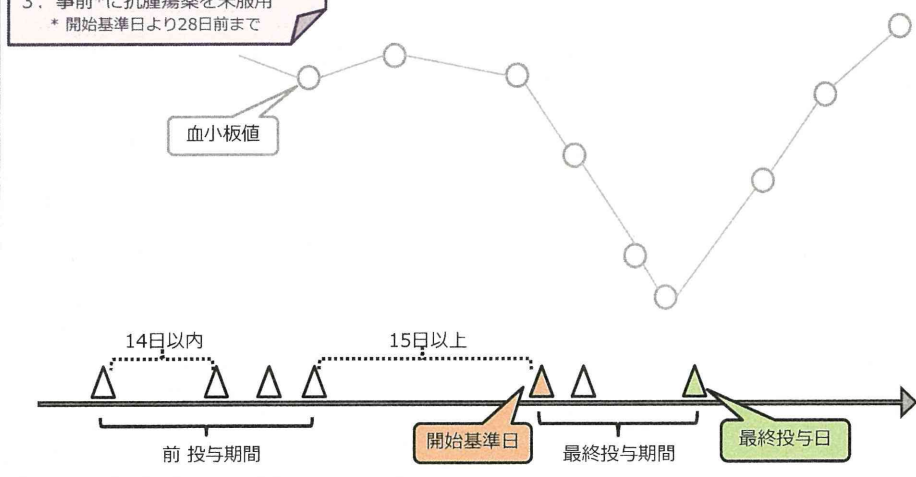
1. 対象薬の処方あり
2. 開始基準日前*後で血小板数の検査値データあり
* 14日前までの期間
3. 事前*に抗腫瘍薬を未服用
* 開始基準日より28日前まで



HIT 検出アルゴリズム

(1) 対象患者の選択

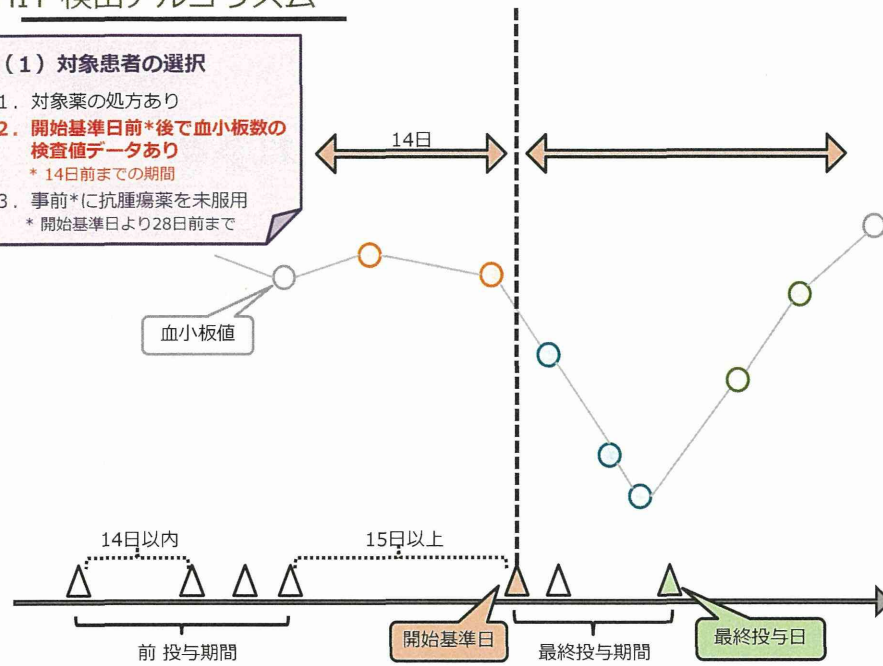
1. 対象薬の処方あり
2. 開始基準日前*後で血小板数の検査値データあり
* 14日前までの期間
3. 事前*に抗腫瘍薬を未服用
* 開始基準日より28日前まで



HIT 検出アルゴリズム

(1) 対象患者の選択

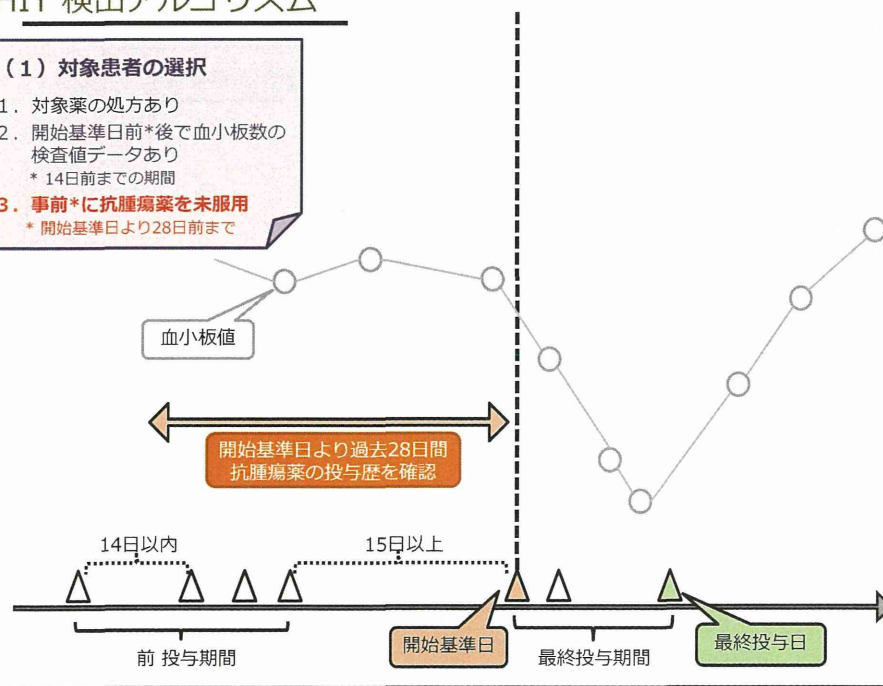
- 対象薬の処方あり
- 開始基準日前*後で血小板数の検査値データあり
* 14日前までの期間
- 事前*に抗腫瘍薬を未服用
* 開始基準日より28日前まで



HIT 検出アルゴリズム

(1) 対象患者の選択

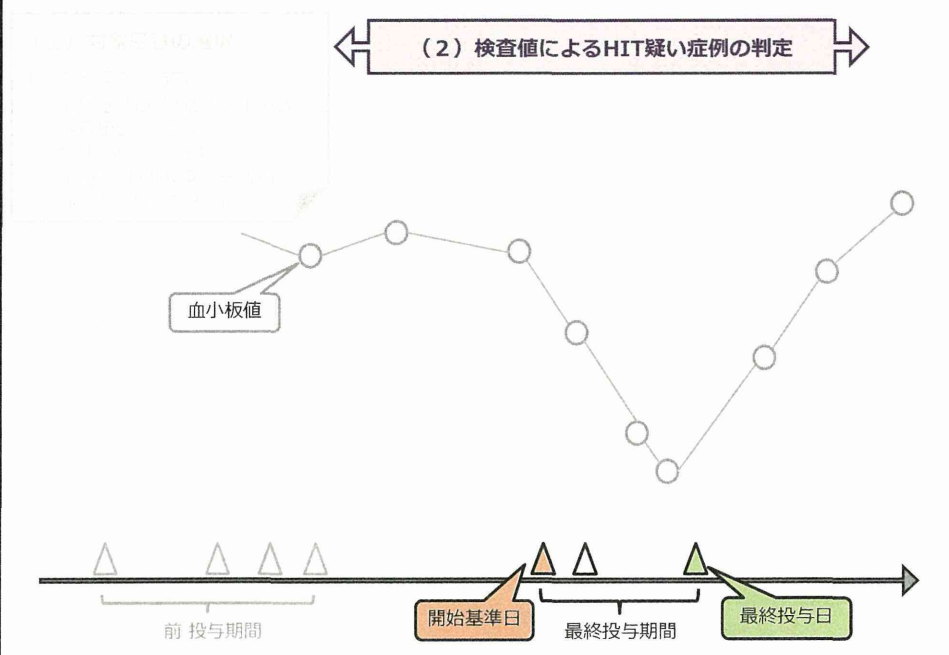
- 対象薬の処方あり
- 開始基準日前*後で血小板数の検査値データあり
* 14日前までの期間
- 事前*に抗腫瘍薬を未服用
* 開始基準日より28日前まで



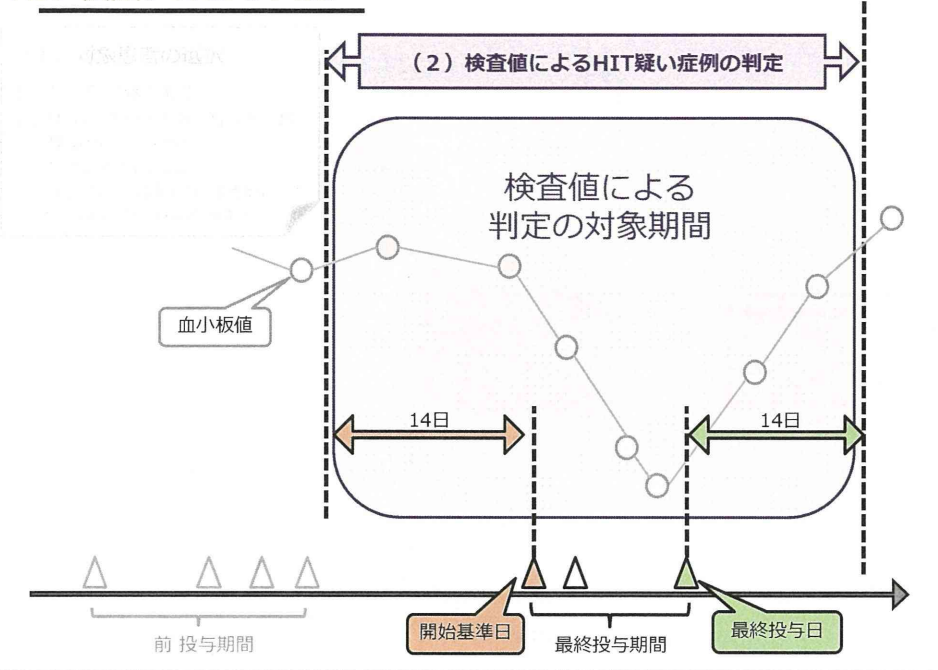
HIT 検出アルゴリズム

- (1) 対象患者の選択（組入れ基準）
- (2) 検査値によるHIT疑い症例の判定
- (3) 病名情報による類似疾患との鑑別

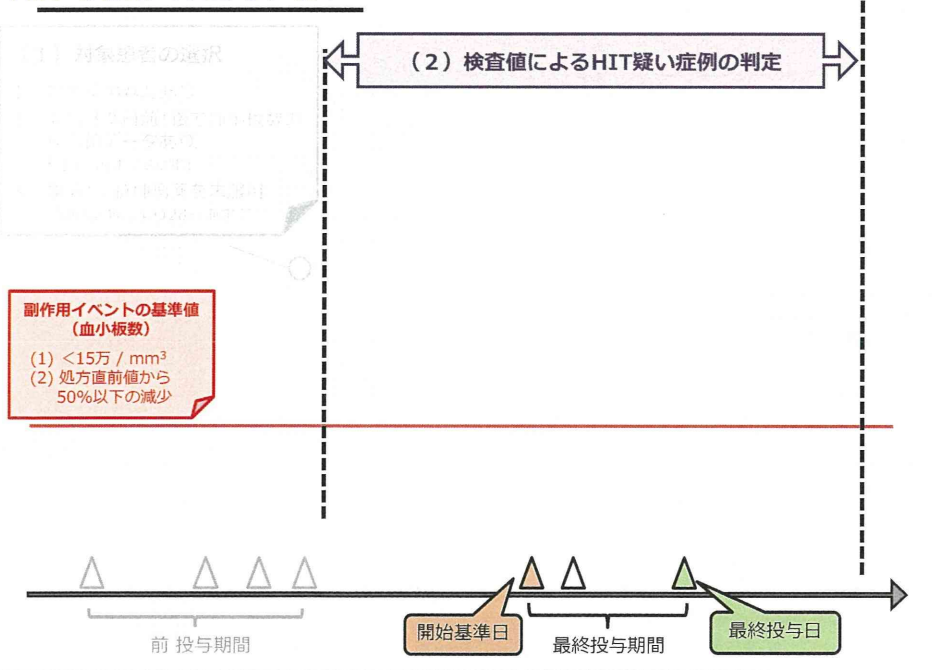
HIT 検出アルゴリズム



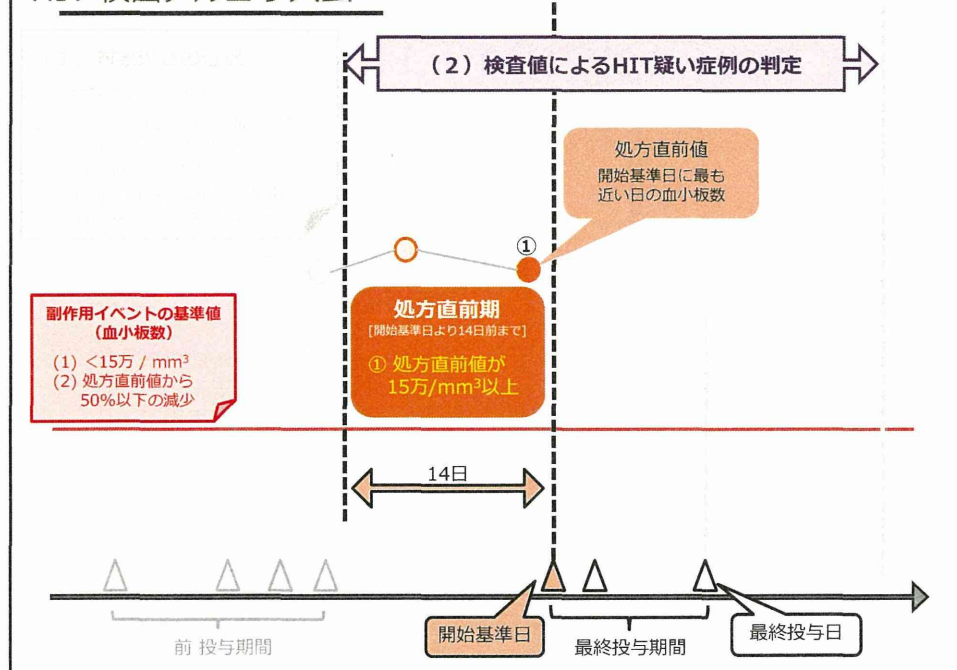
HIT 検出アルゴリズム



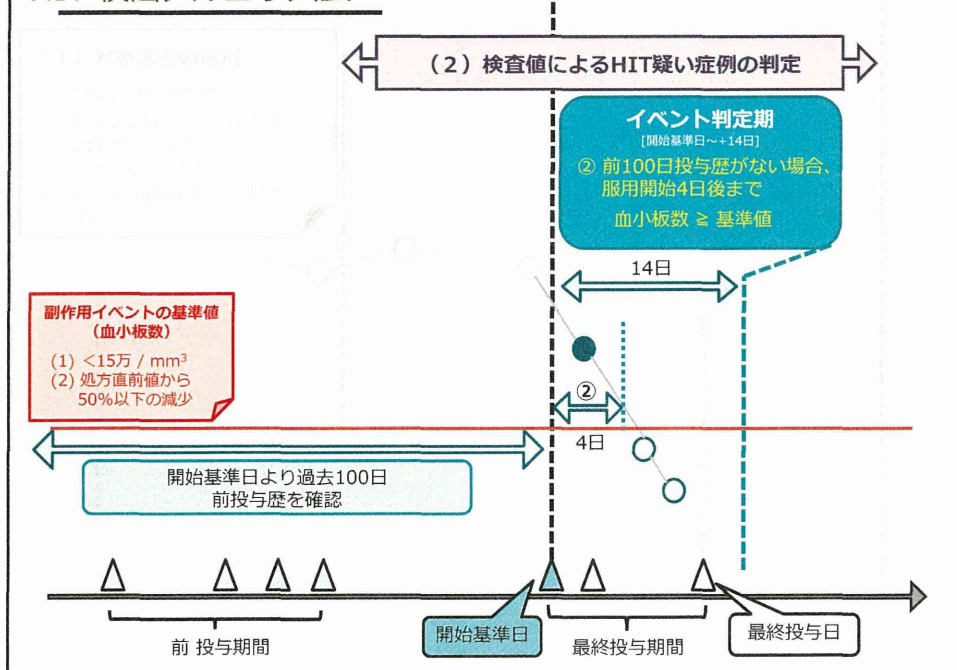
HIT 検出アルゴリズム



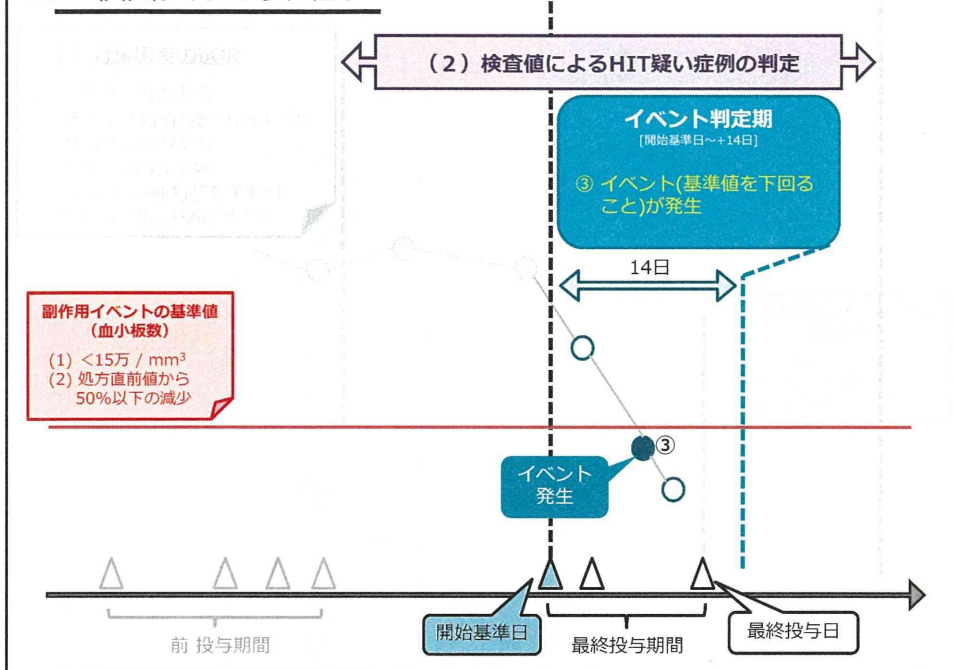
HIT 検出アルゴリズム



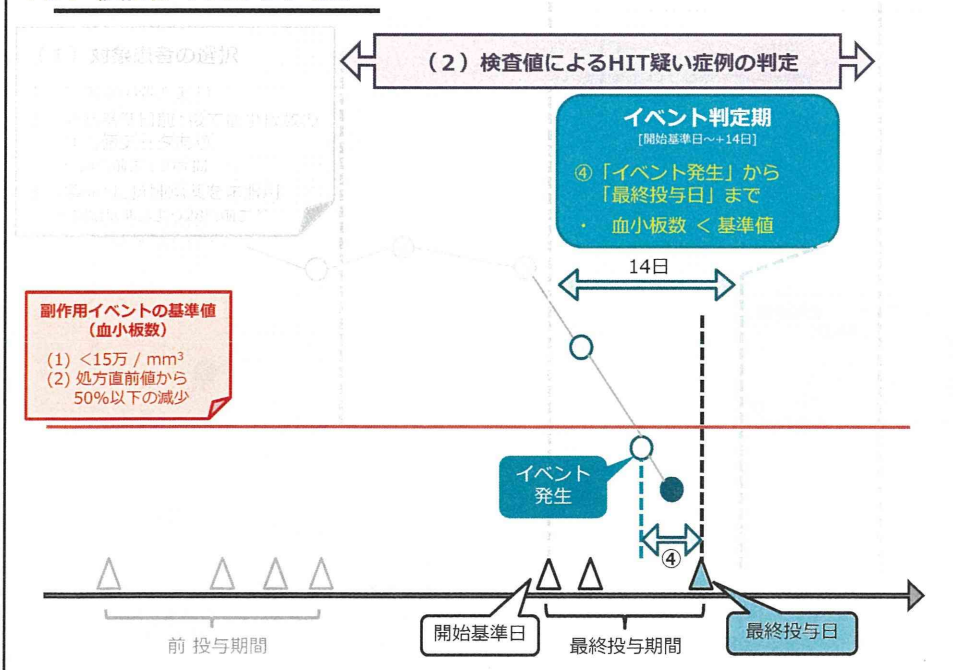
HIT 検出アルゴリズム



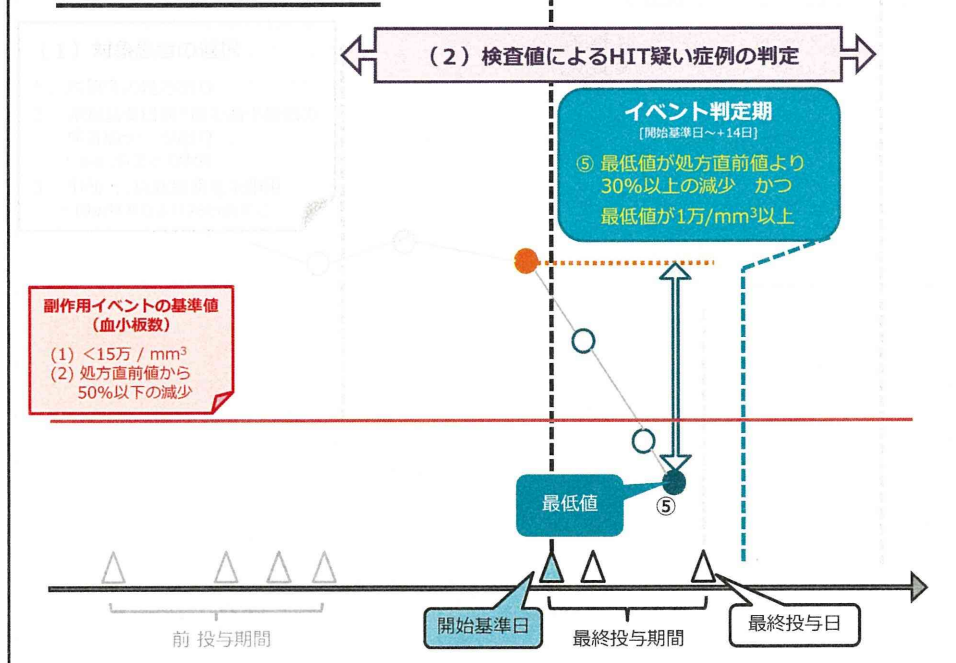
HIT 検出アルゴリズム



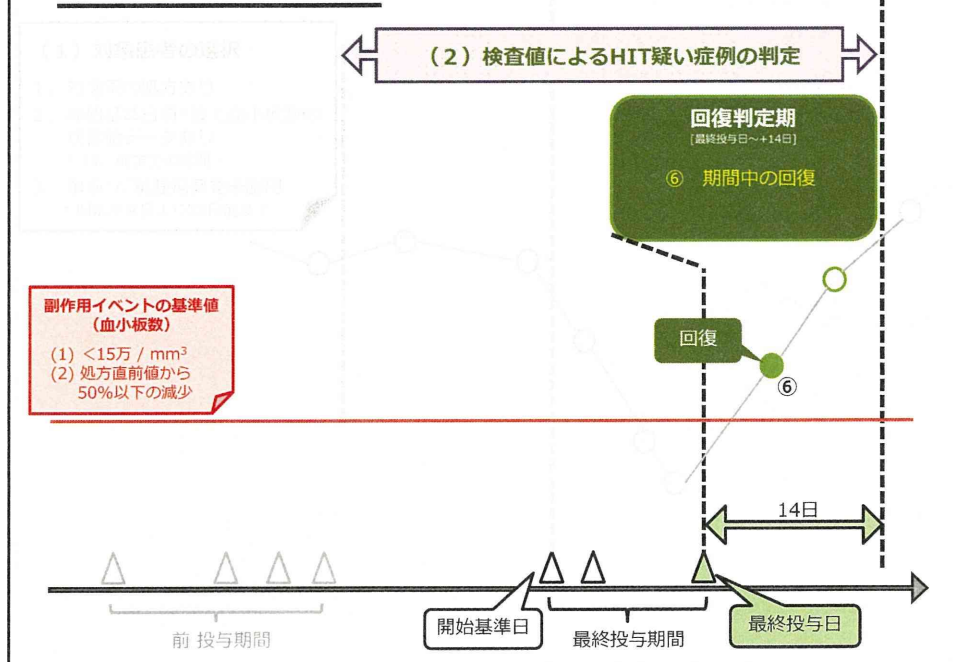
HIT 検出アルゴリズム



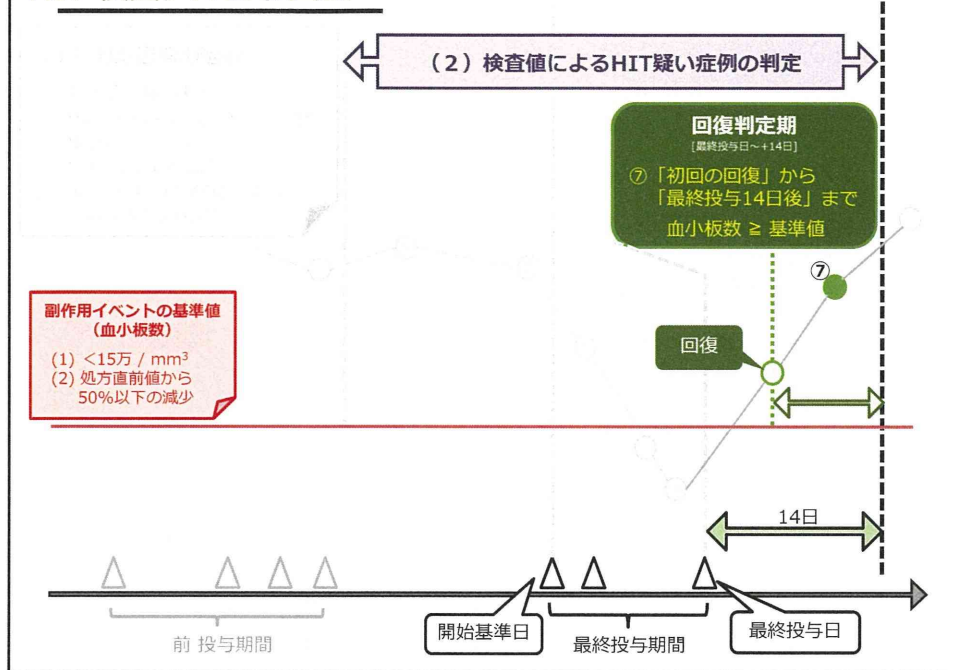
HIT 検出アルゴリズム



HIT 検出アルゴリズム



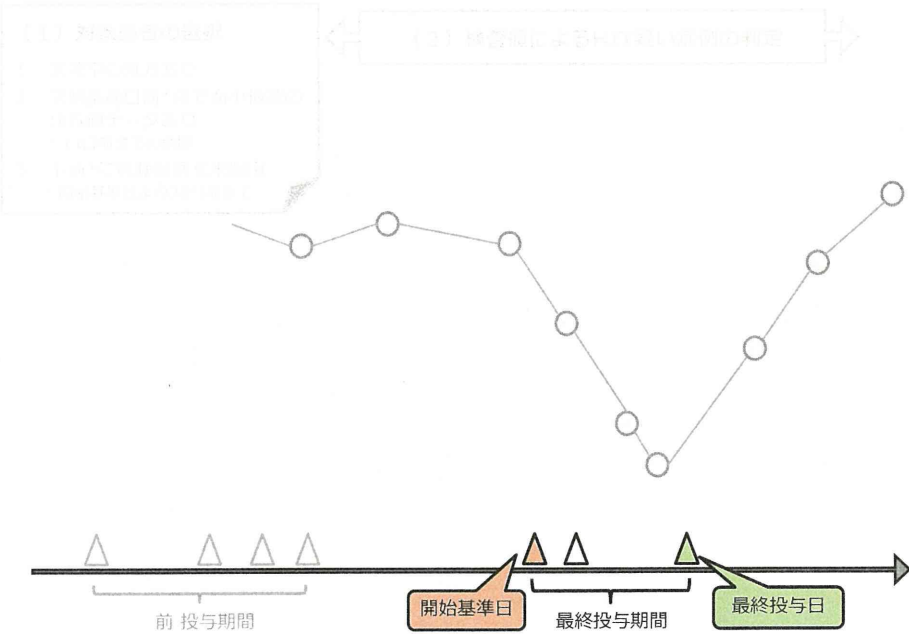
HIT 検出アルゴリズム



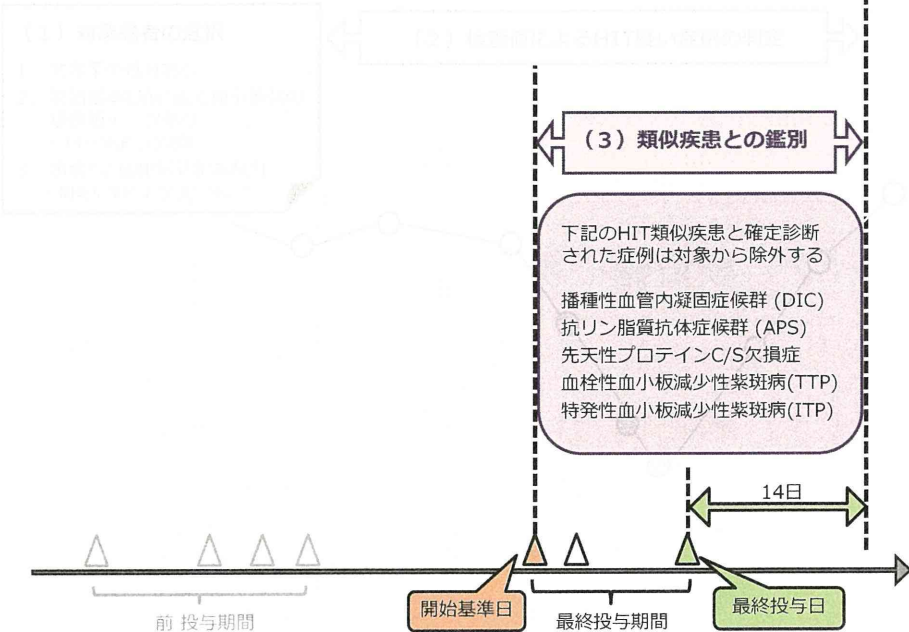
HIT 検出アルゴリズム

- (1) 対象患者の選択（組入れ基準）
- (2) 検査値によるHIT疑い症例の判定
- (3) 病名情報による類似疾患との鑑別

HIT 検出アルゴリズム



HIT 検出アルゴリズム



HIT検出アルゴリズムの適用結果

- 対象期間：2008年4月から2012年3月末まで（4年間）
- 対象薬剤：未分画ヘパリン

(1) 対象患者

2,875

- ① 対象薬の処方あり
- ② 投与前（開始基準日14日前以内）及び投与後の血小板値あり
- ③ 投与前28日の間に抗腫瘍薬の投与なし



(2) 検査値による抽出患者



(3) 検査値+病名による抽出患者

HIT検出アルゴリズムの適用結果

- 対象期間：2008年4月から2012年3月末まで（4年間）
- 対象薬剤：未分画ヘパリン

(1) 対象患者

2,875



(2) 検査値による抽出患者

58 (2.0%)



(3) 検査値+病名による抽出患者

- ① 開始基準日14日前～開始基準日 【検査値 $\geq$ 15万/mm³】
- ② 開始基準日～後4日目# 【検査値>基準値*】
- ③ 開始基準日～後14日目 【イベント発生】
- ④ イベント発生日～最終投与日 【検査値<基準値*】
- ⑤ イベント発生日～ 【検査値の最低値 $\leq$ 処方直前値 $\times$ 30%
かつ $\geq$ 1万/mm³】
- ⑥ 最終投与日～後14日目 【回復を確認】
- ⑦ 初回の回復日～最終投与後14日目 【検査値>基準値*】

* 基準値；(1)15万/mm³ 又は (2)処方直前値の50%

前100日に投与歴がない場合