公募企画
# 公募企画シンポジウム3
# 診療情報から「病態」を取り出そう！ Phenotyingのススメ
2017年11月21日(火) 14:15 〜 15:45 F会場 (10F 会議室1004-1005)

[2-F-2-PS3-2] 診療情報による Phenotypingの現状・限界

河添 悦昌[1], 香川 璃奈[2], 今井 健[3], 大江 和彦[2] （1.東京大学医学部附属病院 企画情報運営部, 2.東京大学大学院 医学系研究科 医療情報学分野, 3.東京大学大学院 医学系研究科 疾患生命工学センター）

診療の場で得られる表現型を効率よく収集するための情報源として電子カルテが注目されており、これに含まれる診療情報から計算機処理によって表現型を同定するタスクを e-Phenotypingと呼んでいる。ここで対象とする表現型は主に疾患であり、その重要な情報源は登録病名であるが、これは保険請求を目的とすることから真の疾患を反映しておらず、また、 DPCで登録される資源最投入病名も入院患者にしか登録されないという制限がある。そのため、病名や処方歴、検査値歴、診療録など複数種類の診療情報を組み合わせ、これらに含まれる幾つかの項目を変数として目的の疾患を有する症例を選択するアルゴリズムが必要となる。筆者らのグループはいくつかの e-Phenotypingアルゴリズムを開発し精度の評価を行ったが、その過程において、判断に必要な情報がそもそも記載されていない場合が多いことや、評価のための正解ラベルの定義の仕方によりアルゴリズムの精度が大きく変わるため、アルゴリズムそのものよりも、質の高い正解ラベルを付与するための方法論がむしろ重要であるという認識に至った。これに加え、現在の e-Phenotypingは異なる目的で記録された複数の断片的な診療情報から、なんとか目的の Phenotypeを取り出そうとしているに過ぎず、これを診療行為と比較すると、生命の設計図であるゲノムの異常が疾患の本態である場合に、症状・所見・血液検査などから疾患を類推しているのと似た状況のように見える。これを変えるためには、まず Phenotypeの根幹をなす情報モデルを規定した上で、それに基づき電子カルテのテンプレートが展開され、必要情報が簡易に入力されるような仕組みを導入し、表現型を明示的に登録する必要があると考えている。

診療情報による Phenotyping の現状・限界

河添 悦昌[*1]、香川 璃奈[*2]、今井 健[*3] 、大江 和彦[*12]
*1 東京大学医学部附属病院 企画情報運営部, *2 東京大学大学院 医学系研究科 医療情報学分野
*3 東京大学大学院 医学系研究科 疾患生命工学センター

Current status and limitations of EHR Phenotyping

Yoshimasa Kawazoe[*1] , Rina Kagawa[*2] , Takeshi Imai[*3], Kazuhiko Ohe[*12]
*1 Department of Healthcare Information Management, The University of Tokyo Hospital
*2 Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo
*3 Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo.

An electronic health record (EHR) is attracting attention as an information source to collect phenotypes obtained in clinical care, and the task of identifying the phenotypes from EHR by computer processing is referred to as EHR Phenotyping. The main target phenotypes here are diseases, and their easily available source is administrative codes, but it has not reflected the actual disease because it is aimed at the insurance claim. This is why it is necessary to develop an algorithm that combines several clinical information such as the administrative data, prescription history, laboratory test results, and medical records. Authors has developed some EHR Phenotyping algorithms and assessed their performances, and during the process, they became aware of the facts that information necessary for the determination itself has not been described in many cases and that the performance of the algorithm varies significantly depending on the ways to define correct labels, so the methodology to assign high-quality correct labels is rather important than the algorithm itself. In addition to that, the current EHR Phenotyping is no more than trying to manage to extract the target phenotype based on several fragmentary information recorded for different purposes. In order to change this, it is considered to be necessary to specify an information model that forms the basis of phenotype first, and then have a template of electronic medical record developed based on the model, introduce a system where necessary information can be input easily, and register the phenotypes explicitly.

Keywords: Electronic Health Record, Phenotyping, e-Phenotyping, Type 2 Diabetes

1. 電子カルテデータのゲノム研究への活用

遺伝学の問題は個々人の遺伝子の塩基配列の違いが身体の特徴に及ぼす影響、つまり遺伝型と表現型との対応を調査することであるが、診療の場で得られる表現型を効率よく収集するための情報源として電子カルテが注目されている。例として、ゲノムワイド関連解析（GWAS）は関心のある表現型を有する群と対照群を集め、群間における一塩基多型（SNPs）の発現頻度の違いから SNPs と表現型との関連を調査するものであるが、米国 NIH が主導する eMERGE プロジェクト [1) 2)] は、多施設の電子カルテデータを情報源としてより大規模な症例コホートを作成している。また、リバース GWAS とも呼ばれるフェノムワイド関連解析 [3)]（PheWAS）も同様の目的で行われ、こちらは多様な臨床的表現型を含む情報源を必要とすることから、電子カルテの利用を前提とした方法である。遺伝学の研究で着目された電子カルテからの表現型抽出であるが、通常の臨床試験や薬剤疫学研究にも同様のニーズが有る。例えば、既存薬の新たな効能を発見するドラッグ・リパーパシング [4)]は、電子カルテを用いて対象集団を同定することでより安価に迅速に施行可能であると考えられるし、後方視的な臨床研究も同様に迅速に繰り返して施行することで研究の仮説形成に役立つことが期待される。

2. EHR Phenotyping:カルテからの表現型抽出

電子カルテからの表現型の抽出を言い換えると、関心のある臨床的な特性を持つ患者集団をカルテ情報から同定することである。ここで、臨床的特性とは、性別／血液型／薬剤アレルギーなどの先天的特性、喫煙や飲酒などの生活習慣、親が糖尿病であるといった家族歴、本態性高血／2 型糖尿病といった疾患歴、降圧薬／血糖降下薬といった薬剤の服用歴、血圧／HbA1c などの検査値歴、各種検査レポートや診療録に記載される検査所見／観察所見のことを指す。医療情報の分野では、これら診療情報の組合わせによって決定される臨床的な特性を表現型と呼び、コンピュータ処理によって表現型を同定することを EHR Phenotyping と呼んでいる（以下、e-Phenotyping と記載）。特定の Phenotype をもつ患者集団の抽出は臨床研究に役立つし、ある患者が特定の Phenotype をもつかの判定は臨床における意思決定支援に有用と思われることから、e-Phenotyping は電子カルテデータを活用するための注目すべき技術として認知されている [5) 6)]。

e-Phenotyping で対象とする主な表現型は疾患であるため、これを同定するために利用可能な情報源としてレセプトデータが考えられるが、これに含まれる病名は保険請求を主目的とすることから真の病名を反映していないことがしばしばある。例えば、2 型糖尿病症例の抽出を目的とした場合、2 型糖尿病の病名が登録された症例であっても、検査を行うために病名が登録されただけで、実際には 2 型糖尿病を有さない場合や、逆に 2 型糖尿病の病名が登録されずとも 2 型糖尿病を有する場合がある。また、単に糖尿病とだけ登録され、1 型か 2 型かの区別がつかない場合も多い。このようなことは、社会制度の異なる諸外国においても同様の傾向にあり、心不全症例を同定するにあたり、診断コード（ICD-9 もしくは ICD-10 コード）のみを利用するだけでは不十分であったとのメタアナリシスの結果も報告されている [7)]。

このような事情から、レセプト病名や処方歴、検査値歴、診療録など複数種類の診療情報を組み合わせ、これらに含まれる幾つかの項目を変数として目的の疾患を有する症例を選択するアルゴリズムが必要となる。前述の eMERGE プロジ

ェクトでは 10 を越える医療研究機関が参加して相互利用可能なアルゴリズムを開発し、これまでに 60 を越えるルール形式のアルゴリズムが PheKB[8]というカタログに登録されている。eMERGE の初期に開発された 13 のアルゴリズムの性能評価では、ケース例の陽性的中率がいずれも 90%を越えるなど高い性能が報告されているが、評価の過程における議論から、アルゴリズムの開発と評価を繰り返して行うこと、ならびに複数施設での評価が単施設での評価よりも重要であることが延べられている[9]。

3. e-Phenotyping アルゴリズムの開発例

PheKB を始め海外で開発されたアルゴリズムは、本邦との診断基準の違いや、医療情報の種類の違い、診療録や検査レポートなどの自然言語文章に対する解析精度の違いなどから、本邦でそのアルゴリズムを利用する場合に、報告される精度を外挿することはできない。そのため、本邦でも同様のアルゴリズムを作成し精度を評価する必要がある。この際に、情報源として SS-MIX2 ストレージ [10] を用いることで、共通した仕様に基づくデータを対象にすることができるため、アルゴリズムの共有がより容易になると期待される。以下、筆者らのグループが行った、e-Phenotyping アルゴリズムの開発と検討例を紹介する[11) 12) 13)]。

3.1. ルール形式のアルゴリズム

PheKB で公開される Northwestern 大学が開発した 2 型糖尿病症例を抽出するルール形式のアルゴリズムもとに修正を加えたものを図1に示す。オリジナルのアルゴリズムでは、病名コードは ICD-9、処方薬コードは RxNorm、検体検査コードは LOINC で表現されており、これを本邦で利用可能な ICD-10 コード、薬価基準収載医薬品コード、JLAC-10 コードにそれぞれ対応付けた。血糖や HbA1c などの検査値の閾値は、本邦の糖尿病診療ガイドラインに基づく値に変更した。また、オリジナルのアルゴリズムでは薬剤の使用などにより一時的に血糖が上昇する二次性糖尿病症例の除外が考慮されていなかったため、これを除外するルールを追加した。115,039 人を対象としてルールを実行したところ 1,662 人の 2 型糖尿病症例が抽出され、この陽性的中率は 100%であった。この例のように 2 型糖尿病に関しては幾つかの修正を行うことで、オリジナルのアルゴリズムで報告される精度とほぼ同等のものが得られることを確認した[11)]。

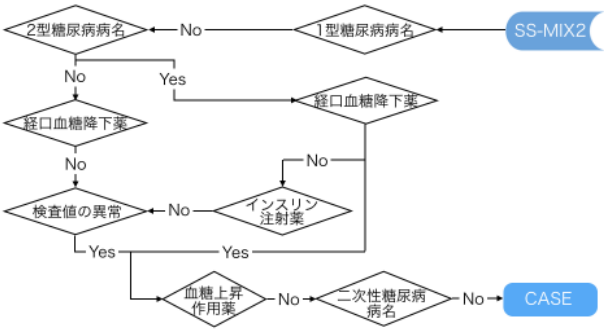

図1 2 型糖尿病を同定するアルゴリズムの修正版

3.2. ルール形式と機械学習を用いた方法の比較

e-Phenotyping アルゴリズムを開発する方法として、ルール形式の他に機械学習を用いる方法が考えられる。一般的に、ルール形式はアルゴリズムの解釈性が良い代わりに、ルールを作るために事前の知識が必要となり、機械学習を用いた方法は解釈性が犠牲になる代わりに、入力データと対応する正解ラベルさえ用意できれば、一定の手続きに従うことで精度の高い識別モデルを作ることができる。

筆者らのグループでは、高血圧症と 2 型糖尿病の有無を判別する e-Phenotyping アルゴリズムをルール形式と機械学習(Gradient Boosting Decision Tree)のそれぞれを使った方法で開発し比較検討した [12)]。入力変数は、病名登録の有無(糖尿病、高血圧症に関する ICD10 コード)、処方薬の有無(ACE/ARB 阻害剤、経口血糖降下薬、インスリン)、血液検査値(血糖、HbA1c)を使用し、2 型糖尿病、高血圧症の有無をそれぞれ判別するものである。650 名の症例に対して高血圧症を有するか否か、2 型糖尿病を有するか否かの正解付けを行ったデータセットを用意し評価した。機械学習を用いた高血圧症例の抽出結果は、感度が 67%、陽性的中率が 80%であり、2 型糖尿病は感度が 68%、陽性的中率が 90%であった。他方、ルール形式を用いた高血圧症例の抽出結果は、感度が 76%、陽性的中率が 76%であり、2 型糖尿病は感度が 47%、陽性的中率が 94%であった。つまり、高血圧症は機械学習を用いた場合に陽性的中率が高く、2 型糖尿病はルール形式を用いた場合に陽性的中率が高いという結果であった。機械学習とルール形式の違いを検討したところ、2 型糖尿病では単独で高い陽性的中率が得られる変数(経口血糖降下薬の投与有無)が存在していたことがルールに有利に働き、他方で高血圧症においては単独で高い陽性的中率を取る変数が存在しなかったことが、機械学習に有利に働いたと考えられた。このことから、識別に有効な変数があらかじめ知られている場合は、ルール形式で高い精度が得られると期待され、そのような変数が知られていない場合は機械学習を用いた方法を用いるのが良いと思われた。機械学習の手法によっては識別に有効な変数だけを残してモデルを構築する方法もあるため、入力変数を網羅的に用意した上で、機械学習モデル構築の過程で変数選択を行うといった方法の適用も考えられ、人手により選択された変数と、機械学習により選択された変数との比較も興味深い課題と思われる。

4. アルゴリズムの開発における問題点

4.1. 正解ラベル付けの必要性

前述の 3.1 ルール形式のアルゴリズムで述べた、2 型糖尿病症例を検出するアルゴリズムの評価には陽性的中率だけを用いたが、これはアルゴリズムが陽性と判定した症例のカルテを開き結果が正しいかどうかを判断すれば良いため、比較的簡便に行うことができる。しかしながら、アルゴリズムの詳細な性能を評価するためには陽性的中率だけでは十分ではなく、感度や特異度も評価に加える必要がある。極端な例では、確実に 2 型糖尿病といえる症例を1件だけでも見つければ陽性的中率は 100%となるが、一方で多くの 2 型糖尿病を見落とすことになるため、見落としの指標である感度も合わせて評価する必要があるためである。また、陽性的中率は有病率(データセットにおける 2 型糖尿病の割合)に影響をうけるため、同じアルゴリズムを有病率の異なる施設で利用するためには、有病率に影響を受けない感度と特異度がより適切な指標となるかもしれない。ここで、感度と特異度を測定するためには、対象とするデータセットの全症例に対して各カルテを調査し正解ラベル(2 型糖尿病、本態性高血圧といった抽出対象の表現型)をあらかじめ与える必要があり、これに極めて時間を要することがアルゴリズム開発の律速となる。

4.2. 明確な基準による正解ラベル付けの必要性

データセットの開発に時間を要することも問題ではあるが、専門的知識を持つ者がカルテを読めば、正解ラベルを正しく付けることができるのかという疑問が生じ、こちらはより本質的な問題と思われた。医療者であれば日常的に経験していることであるが、例えば、診療録に明示的に 2 型糖尿病と明示されていなくとも、それに準じた治療が行なわれていることから2型糖尿病が推察される場合や、紹介状に糖尿病性網膜症の記載があるが、それ以上の情報がないものなどがあり、このようなケースを 2 型糖尿病に含めるかどうかといった点が問題となる。諸外国を含むいくつものグループが e-Phenotyping アルゴリズムを開発しているが、正解ラベル付けの判断基準をどのように定義しているかの詳細な報告は調べる限りみられなかった。そこで、筆者らのグループは、1) 2 型糖尿病の記載があるもの、2) 経過から 2 型糖尿病と判断できるもの、3) その他の理由で 2 型糖尿病の可能性が高いと思われるもの、4) 型不明の糖尿病、5) 糖尿病の可能性が高いもの、6) 2 型以外の糖尿病、7) 糖尿病でないもの、の 7 分類を用意し、それぞれの判断基準を定義したフローチャートを作成した(図 2)。このチャートに基いて分類した症例のうち、前述の 1)2)3)に該当するものを 2 型糖尿病に分類するとした場合に、このうち診療録に2型糖尿病と明記されていたものは約57%に過ぎなかった。また、レセプト病名として ICD-10 の E11x(インスリン非依存性糖尿病)が登録されている症例のうち、2 型糖尿病に分類するとした患者は約半数であった [13]。この結果は、診療録の記載に対する文字列のマッチングやレセプト病名だけを用いて、機械的に 2 型糖尿病症例を抽出することの不十分さを改めて示すものであるが、それに加え、どの範囲を2型糖尿病に分類するかという閾値を変えることで、正解のラベル数が変化し、それに応じて e-Phenotyping アルゴリズムの精度も容易に変わることも示している。このことから、自然言語文章を含む電子カルテの記載から Gold Standard を判断しなければならない e-Phenotyping のようなタスクにおいては、アルゴリズムの開発そのものよりも、質の高い正解ラベルを付与するための方法論がむしろ重要であると考えられた。

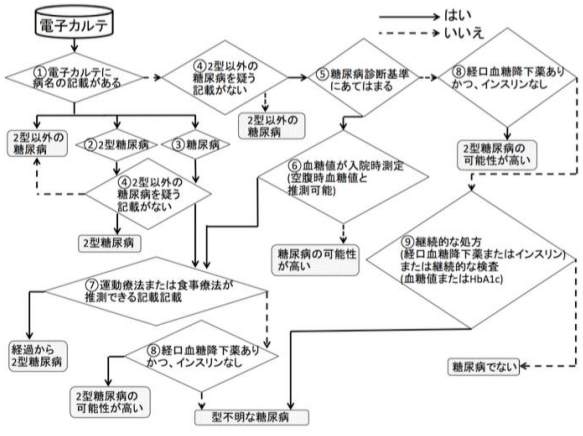

図 2 2 型糖尿病の正解ラベル付けのフローチャート

4.3. e-Phenotyping が困難である理由

前述の実験から、計算機処理で疾患を同定するにあたり具体的にどのような難しさがあるのかを検討するための手がかりも得られた。例えば、図 2 の⑦「運動療法または食事療法が推測できる記載」は、2 型糖尿病であればまず行われる生活習慣指導の有無から疾患の有無を判断しようとするものであるが、診療録に栄養指導を行ったことの記載はあるものの、2型糖尿病に対する指導であることが明示されないことから、他疾患に対する指導でないことと、糖尿病に関連する記載であることを文の内容から判断する必要がある。医療者であれば、栄養指導の対象となる疾患はいくつかに限定されることと、各疾患に対して行われるべき指導内容が想定されることから、2型糖尿病を対象としたものかどうかの判断は比較的容易であるが、そのような知識のない計算機処理で判定するためには、本質的にカルテに記載される情報が足りないと考えられた。また、⑨「経口血糖降下薬またはインスリンの継続的な処方」は、救急搬送や手術後の症例では一時的な血糖上昇に対して血糖を下げる薬剤が使われることがあるため、そのような症例を除外する目的のものである。医療者であれば一時的な血糖の上昇かどうかは、血糖上昇の前にどのような医療イベントが生じていたのかをみること、つまりカルテに記載された文脈から比較的明らかに判断できるが、計算機処理でこれを行うためには、血糖の上昇を伴う医療イベントの種類が何であるかということと、血糖上昇との前後関係の判断が必要となり、また一時的という表現を単純な期間の表現に置き換えられないことから、計算機処理で判断することが難しいタスクと考えられた。

5. e-Phenotyping の現状とこれからの展望

筆者らの経験した e-Phenotyping アルゴリズムの開発を通して、2 型糖尿病という診断基準が比較的明確な疾患であっても、診療録の記載に対する単純な文字列のマッチングやレセプト病名だけは、対象症例が十分に抽出できないことを改めて提示した。また、正解ラベル付けの詳細なフローチャートに基づく分類実験から、判断に必要な情報がそもそも診療録に記載されていない場合が多いことや、医療者であれば事前の知識とカルテの経過を追うことで容易に判断できることが、計算機処理によって行うことが難しいことを述べた。現状の e-Phenotyping は、異なる目的で記録された断片的な診療情報から、目的の Phenotype を取り出そうとしているに過ぎず、これを疾患の診断に例えると、生命の設計図であるゲノムの異常が疾患の本態である場合に、症状・所見・血液検査などから疾患を類推しているのと似た状況とも捉えることができる。現在のように過渡期における方法論として、日常的な診療の結果として発生する情報を流用することはやむを得ないと思われる一方、このような方法を続けることが e-Phenotyping 研究に求められ続けるとは考えにくい。それを変えるための一例として、Phenotype の根幹をなす情報モデルを規定した上で、それに基づく電子カルテのテンプレートが展開され、必要情報が簡易に入力できるとともに、その情報が日常診療の記録として反映される仕組みとして、電子カルテに入力する情報の主従を転換する方法などが考えられる。

6. 倫理的配慮

本研究の実施は東京大学大学院医学系研究科・医学部倫理委員会により承認を得ている(承認番号 10791)

7. 謝辞

本研究は MEXT 科研費 16K09161 と MEXT 科研費 16J05555 の助成を受けた。

参考文献

1) McCarty CA, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for

conducting genomic studies. BMC Med Genomics 2011;4:13.
2) Gottesman O, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013;15:761-71.
3) Denny JC, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010;26:1205-10.
4) Xu H, et al. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. J Am Med Inform Assoc. 2015 Jan;22(1):179-91.
5) Pathak J, et al. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J Am Med Inform Assoc. 2013 Dec;20(e2):e206-11.
6) Wei WQ, Denny JC. Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome Med. 2015 Apr 30;7(1):41.
7) McCormick N, et al. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. PLoS One. 2014 Aug 15;9(8):e104519.
8) Kirby JC, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc. 2016 Nov;23(6):1046-1052.
9) Newton KM, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc. 2013 Jun;20(e1):e147-54.
10) SS-MIX 普及推進コンソーシアム. [http://www.ss-mix.org/cons/ssmix2_about.html (cited 2017-Sep-14)].
11) 香川 璃奈 他. SS-MIX2 標準化ストレージを用いた 2 型糖尿病症例の EHR Phenotyping 手法の開発と評価. 医療情報学 35(Suppl.), pp.672-675, 2015.
12) 香川 璃奈 他. 高血圧の phenotyping 手法の開発および他疾患との比較検討. 医療情報学 36(Suppl.), pp.770-773, 2016.
13) 香川 璃奈 他. 特定患者集団の抽出 (Phenotyping) 手法確立に向けた技術的課題に関する考察. 人工知能学会 第 30 回全国大会 (北九州), 2C4-OS-21a-5:pp.1-4, 2016.