

・ 総括研究報告

厚生労働科学研究費補助金（政策科学総合研究事業（臨床研究等 ICT 基盤構築研究事業））
総括研究報告書

DPC とがん登録を軸としたデータベース拡充・連結に関する研究

研究代表者 東 尚弘 国立がん研究センターがん対策情報センターがん臨床情報部 部長

研究要旨

本研究は院内がん登録と、病理診断情報、放射線画像情報、臨床試験情報とのリンクによる臨床情報の拡充をめざし、その実施可能性の検討と応用事例の蓄積を測るものである。それぞれのデータに関して、まず初年度はデータ提供元において共通 ID で匿名化して収集することを基本とするために共通匿名化を自動化する匿名化ソフトウェアの開発と、いくつかの協力施設における手順の整理や円滑化に向けた検討を行っていく。病理診断情報、放射線画像情報については、合計3施設の協力が得られデータ収集を行った。病理情報についてはデータの抽出については2施設において円滑に抽出は可能であったが、必要情報が所見文の中に入っているためにそれを抽出するための機械学習ソフトと学習のためのインターフェースを開発中である。放射線画像診断については院内がん登録と DPC データにより、対象患者と検査日を特定し、それを施設内で院内 ID に戻して画像を取得の上再度匿名化する作業を自動化するソフトを開発した。ただし非常に時間がかかるため効率化の方法などを引き続き検討する。臨床試験についても対象を確定して審査を受けているところである。今年度の成果はデータを収集する前の準備に費やしたが、来年度からは実際のデータ解析を開始し、均てん化に資するデータベースの構築を目指す。

研究分担者氏名・所属機関名・職名

東 尚弘 国立がん研究センター
がん対策情報センター部長

中村健一 国立がん研究センター
中央病院臨床研究支援室室長

寺本典弘 四国がんセンター病理科医長

小林秀章 大隅鹿屋病院放射線科医長

合して解析を行うことで、研究、教育、診療支援に活用することを目的としている。院内がん登録は基本的な患者の台帳として機能するが、この突合により、病理診断、放射線診断情報が整理された状態で蓄積・活用が容易になる独創的な試みである。また、将来的には診断の自動化システムを開発する際の機械学習のための元データとなることも期待される。

また、情報の比較という意味では、非常に精度の高い臨床試験のデータを元に、院内がん登録などの検討を行うと共に、臨床試験参加集団の代表性の検討から、臨床試験の外的妥当性の検討を行うことを目的とする。

A. 研究目的

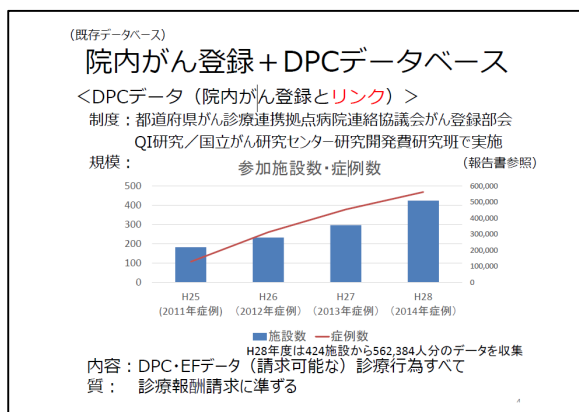
個別に存在している電子的なデータベースをリンクさせ、活用するための幅を広げることがアイデアとして存在するものの、わが国においてはその実施状況は限定的であり、そのリンクにおいても確率的にリンクすることが主となっており、もとの個人情報取扱主体において共通匿名化を行ってリンクする確実なリンクはなされていない。本研究は共通キーをもちいたリンク手法を用いて、既存の「院内がん登録 + DPC」データベースを軸として、病理診断、放射線診断情報、臨床試験の情報を突

B. 研究方法

上記の目的で、協力の得られる施設から病理診断、放射線画像の情報を院内がん登録 + DPC にリンク可能な形で抽出・集積するシステムや体制を構築すること、連携の可能な臨床試験との患者集団の重症度などの比較によ

る代表性(外的妥当性)の検討を行うこととの2点を行うこととした。

の病理診断、放射線画像は、院内がん登録 + DPCについては、匿名データが国立がん研究センターにがん診療連携拠点病院を中心とした数多くのデータが蓄積されていることから、協力施設においては、施設の許可の元でそれらを使用することとした。



次に、協力施設において、診療システム(放射線画像、病理画像)から抽出可能なデータ形式を同定し、対象症例を院内がん登録、DPCから、絞って、院内がん登録と突合可能な形で匿名化、提出用に暗号化するなどの一連の作業を決定、また可能な限り施設内での作業を自動化するために、ソフトウェアの開発を行った。特に病理データについては文章で所見が記載されていることから、その中から必要事項を効率的に抽出できるように、機械学習による自然言語処理の技術を活用しつつ、また人の確認や訂正作業を軽減できるようなデータ抽出のユーザーインターフェースを開発した。

の臨床試験情報との比較突合においては、臨床試験の結果を適用すべき標的母集団を院内がん登録によって構成することで、臨床試験参加患者と標的母集団患者との差異を検討するとともに、精度管理の厳しい臨床試験情報を使って院内がん登録の情報の検証を行い、正確性の向上に努めることを目標としている。初年度においては対象となる臨床試験を同定した。

(リンクデータ)

臨床試験データ

- JCOG (日本臨床腫瘍研究グループ) の臨床試験

臨床試験の課題：
 参加者は専門施設の元気な患者
 ⇒ 一般臨床応用できるのか (外的妥当性)

(解析計画)
 臨床試験データと院内がん登録データと、施設レベル、患者レベルで突合
 ・患者母集団と臨床試験参加者の比較
 年齢、性別、ステージ、予後など
 ⇒ 外的妥当性の検証
 ・院内がん登録の正確性確認

<http://jcog.jp/basic/clinicaltrial/>

C. 研究結果

病情報および画像情報の抽出システムの開発と実際のデータ収集1年目である本年度は、初期協力施設(病情報3施設、放射線画像2施設)において倫理審査・承認を得るとともに、以下の作業を行った。

(病情報)

・抽出作業の開発：

異なる病情報システムを使用している3施設(国際医療福祉大学、四国がんセンター、琉球大学)から、項目を一覧形式のカンマ区切りテキストファイルで抽出できることを確認した。そのデータを1箇所に集めて分析を行うことが可能なように、各施設で保存されている院内がん登録の対応表を用い、院内がん登録症例には共通した匿名IDを使い、院内がん登録症例以外については新しい匿名IDを自動的に付与する匿名化ソフトの開発を行った。倫理審査の終了した2施設で順次、抽出したファイルを使い当該プログラムの動作を確認し、匿名化した病情報を収集した。

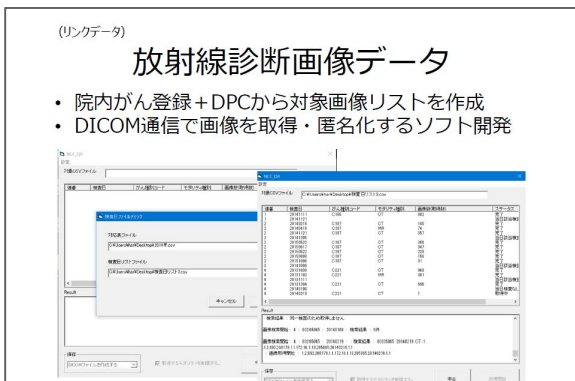
・解析：

腫瘍径、脈管侵襲、受容体の発現などの必要情報が病情報では病理所見欄の文調中に埋め込まれていることが確認されたため、機械学習による必要情報の抽出を行うためのプログラム開発を開始した。年度内に試作を完成させる。

(画像情報)

・抽出作業：対象患者のリストを読み込み、それらの患者に対して通常の臨床業務と同様にDICOMサーバと通信して画像を抽出するソフトを開発し、1施設で試行しつつ継続的な改善を行った。通常のPCでは非常に時間がかかることが判明し、別の機器を準備するなどの試行錯誤を行った。

・解析：腫瘍の含まれている臓器の画像を抽出するための処理ソフトを開発する業者を検討中である。



(臨床試験)

・日本臨床腫瘍研究グループ(Japan Clinical Oncology Group)の臨床試験情報について、データ二次利用委員会に申請している。想定している臨床試験はJCOG0212試験という、Stage II-III 直腸癌を対象とした第III相試験、JCOG0404と呼ばれる、I期の大腸がん患者への開腹手術と腹腔鏡下手術の非劣性試験である。現在、データ利用について審査中である。

JCOG0404

臨床試験の概要：

- ・ 大腸癌に対する腹腔鏡手術vs開腹手術
- ・ 30施設の多施設共同研究
- (対象) T3-4N0-2M0腫瘍径<8cm、20-75才(参加者) 1057例参加、IC承諾率 60%
- 年齢64才(中央)
- II期66%、III期34%、IV期0.03%

・ JCOG二次利用委員会で審査中

D. 考察

本年度は、データを本格的に収集するための基礎構築に主眼を置き、病情報、画像情報、臨床試験情報の3者における準備作業を行った。院内がん登録は、厚生労働省の指定するがん診療連携拠点病院においてその実施が指定要件であり「がん登録の推進に関する法律」(平成28年1月施行)においても、義務化された全国がん登録を支える基盤として位置付けられている。本研究が進むにつれて、この院内がん登録に放射線画像、病理診断情報、臨床試験データの突合による有効な活用モデルが示されるだけでなく、様々な効果が期待できる。

まず、がん対策基本法(平成18年6月23日成立)における重要課題である「がん医療の均てん化」推進のためには病理診断や画像診断などを含む専門家育成とともに、診療実態の把握も必要である。本研究で構築される放射線診断、病理データベースを活用、解析することで実態把握への道が開け、また、エビデンスに基づく人材育成にも貢献が可能であると考えられる。

さらに、第2期がん対策推進基本計画で対策が始まった希少がんにおいては一層貢献が期待される。希少がんは現状では症例経験の蓄積も困難であり、診断すら困難であることが問題視されている。本データベースにより希少がんの診断症例が蓄積され、現場の診断医が希少症例の放射線、病理診断画像に触れやすくなり、診断の効率や正確性の向上にも役立つ。

最後に、臨床試験の知見を正しく一般診療現場で適用するという現場の課題において本研究は一般患者と臨床試験参加患者の特徴の差異を定量的に検証するモデルが提供可能であり、これも医療の質の向上に寄与すると期待できる。

E. 結論

病理情報、画像情報、臨床試験との院内がん登録とのリンクデータの収集について、多様な現場においての試行を行い、一定の実績を得た。今後は対象を拡大するとともに、データ収集体制に関しても改善を重ねていく。

F. 健康危険情報 特になし

G. 研究発表

1. 論文発表 (東 尚弘)

1. Ogura K, Higashi T, Kawai A.
Statistics of bone sarcoma in Japan:
Report from the Bone and Soft Tissue
Tumor Registry in Japan. J Orthop Sci.
2017 Jan ; 22(1):133-143
2. Okuyama A, Nakamura F, Higashi T.
Prescription of Prophylactic Antiemetic
Drugs for Patients Receiving
Chemotherapy With Minimal and Low
Emetic Risk. JAMA Oncol. 2017 Mar
1;3(3):344-350
3. Maeda E, Higashi T, Hasegawa T,
Yokoya S, Mochizuki T, Ishii T, Ito J,
Kanzaki S, Shimatsu A, Takano K,
Tajima T, Tanaka H, Tanahashi Y,
Teramoto A, Nagai T, Hanew K,
Horikawa R, Yorifuji T, Wada N,
Tanaka T.Effects of financial support on
treatment of adolescents with growth
hormone deficiency: a retrospective
study in Japan. BMC Health Serv Res.
2016 Oct 21;16(1):602.

4. Takaoka M, Okuyama A, Mekata E,
Masuda M, Otani M, Higashide S, and
Higashi T, Staging discrepancies
between Hospital-Based Cancer
Registry and Diagnosis Procedure
Combination data Japanese Journal of
Clinical Oncology 2016 Aug;46(8):788-
91.

(寺本 典弘)

Roma AA, Teramoto N, et al. New
pattern-based personalized risk
stratification system for endocervical
adenocarcinoma with important clinical
implications and surgical outcome.
Gynecol Oncol. 2016 Apr;141(1):36-42

2. 学会発表

The 19th International Congress of
Cytology, 28 May–01 June, 2016
Pacifco Yokohama, Japan
Proffered paper O-045 QUALITY
ASSESSMENT OF CYTOLOGICAL
DIAGNOSIS OF LUNG CANCER
USING HOSPITAL CANCER
REGISTRY
Norihiro Teramoto

H . 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

