

平成27年3月18日	資料2-2
第24回レセプト情報等の 提供に関する有識者会議	

レセプト情報等オンサイトリサーチ センター(仮称)利用模擬申出

模擬申出一覧 運用目的

No.	申出者	研究名称 ※目的	抽出項目	公表その他
1	小林廉毅 (東大)	<p>「レセプト情報等オンサイトリサーチセンター利用普及に向けたセンター機能と性能の実証的評価に関する研究」</p> <p>当該システムの機能で保健医療にかかわる研究者が実際に行う研究が滞りなく実施できるか、システムの性能は十分か、研究を実施する上で研究者にはどのような技能・技術が必要か・またそれは利用普及の障壁となりうるか、今後どういった点を改善すべきかが明らかでないため、高度な負荷をかけることを予定している。</p>	医科、歯科、調剤、DPC、特定健診、保健指導の全データ	<ul style="list-style-type: none"> ・有識者会議報告 ・学術論文発表なし <p>(高度化室厚労科研(予定)「レセプト情報・特定健診等情報データベースの利活用の推進に関する研究」における報告)</p> <ul style="list-style-type: none"> ・利用者数26名(博士課程10名、修士課程1、研究員1を含む)
2	黒田知宏 (京大)	<p>利用割り当て時間や、センターの稼働を効率的にするために、</p> <p>一般利用に向けた技術的・運用上の問題点を明らかにすることを目的としている。</p>	医科、歯科、調剤、DPC、特定健診、保健指導の全データ	<ul style="list-style-type: none"> ・有識者会議報告 ・学術論文発表なし <p>(高度化室厚労科研(予定)「レセプト情報・特定健診等情報データベースの利活用の推進に関する研究」における報告)</p> <ul style="list-style-type: none"> ・利用者25名(博士課程4(+1)名、研究員2名、技術補佐員1名を含む)

模擬申出一覧 研究目的

No.	申出者	研究名称 ※目的	抽出項目	セキュリティ要件等その他
3	小林廉毅 (東大)	「後発医薬品の普及状況および関連要因に関する研究」 ※慢性疾病で通院する患者について後発医薬品と先発医薬品の使用割合の変化や関連要因を明らかにする。	2009年4月～2014年3月 医科、調剤、DPC 傷病名に別表1のいずれかを含むもの IR,RE,HO,KO,SY,SB,IY,等	7名(研究員1名、博士課程2名)(No1と重複有り)
4	大江和彦 (東大)	「レセプト情報等オンサイトリサーチセンターのレセプト情報等を用いた脳血管疾患の実態に関する研究」 ※脳血管疾患患者の薬物、リハビリテーション実施状況と医療費の検証。提供を進めている基本データセットを用いた研究成果と比較検証し精度管理等の課題を明らかにし利活用向上に貢献する。	2009～2014年度。 330項目指定 (基本データセット抽出条件に準拠)	2名(No1と重複有り)
5	康永秀生 (東大)	「周術期口腔機能管理による術後肺炎発症予防の効果」 ※DPCレセプトと歯科レセプトを用いて周術期口腔機能管理が術後肺炎発症を予防する効果について分析する。	2012年4月～2015年3月 歯科、DPC 指定Kコード手術を実施したもの IR,RE,HO,BU,SB,KK,SK,HH,GT,等	7名(No1と重複有り)
6	黒田知宏 (京大)	「高齢患者における不適切処方薬の処方実態に関する研究」 ※65歳以上の不適切処方の回数、投与量、併用薬を調査し、有害事象との関連、医療経済評価を行う。	2009年4月～2015年3月 医科、医科入院外、調剤 RE,IR,SY,BU,SB,IY,SI,CD,RE,YK,IY等	17名(No2と重複有り)

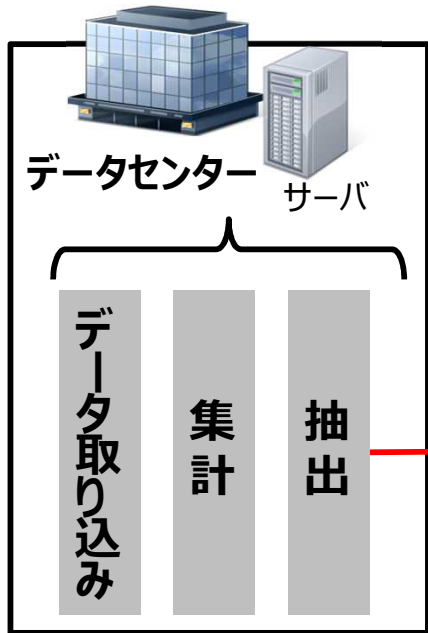
運用目的の研究について

背景：オンサイトリサーチセンターの現況把握の必要性

- 今回、オンサイトリサーチセンターが東京大学ならびに京都大学に開設され、オンサイトリサーチセンターから既存のデータセンターに直接アクセスし、データの抽出、集計、分析等を実施することが可能となった。
- しかし、これまでの「データ取り込み」「抽出」「集計」と異なる作業エリアを用いても、データセンターに「利用者による分析」という負荷がどの程度同時にかけられると機能不全を生じやすいか実証的評価を行っていない。
- オンサイトリサーチセンターには新たに統計ソフト(SAS, R)が導入され、幅広い分析が実施できるようになったが、これらソフトを利用した場合の動作等についても不明な部分がある。
- オンサイトリサーチセンターは両機関に2ブース、合計4ブースしか設置されていない。データセンターの機能把握をした上で、利用者に長い待機を強いたり、限られた端末を効率よく割り当てられなかったりすることなく、実運用を行う必要がある。
- 本研究は、限定的なデータを利用して特定の課題を実証研究するという従来型の研究目的ではなく、オンサイトリサーチセンターという新たなNDB利活用の運用システム全般について、現況把握を行うという目的において行うものである。
- また、本研究は東大・京大両機関において合同で行い、双方から現況把握を行うとともに、両機関の連携のもとでシステムの機能と性能を十分に把握し、オンサイトリサーチセンターを最大限に活用できる方法を提供側、ならびに利用者側の双方に提示することを目指すものである。

オンサイトリサーチセンターと従来のNDB運用の違い

従来のNDB運用



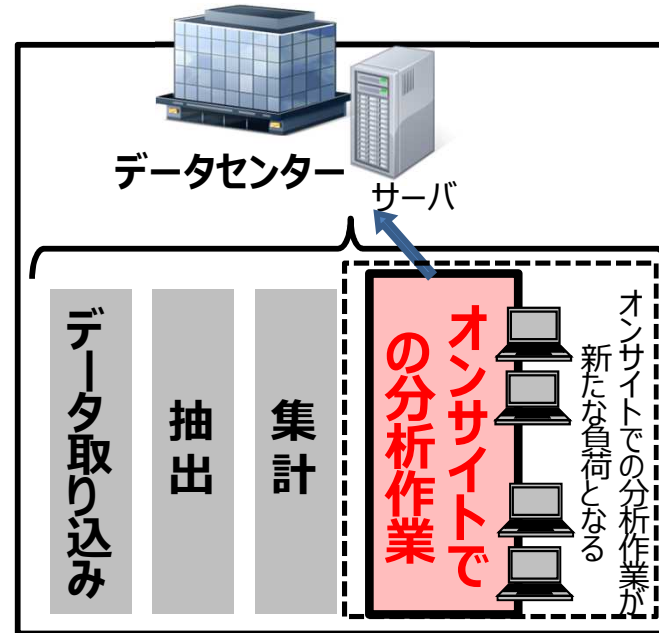
【利点】

- データセンターへの作業負荷の概要がこれまでの運用すでに蓄積されており、提供依頼等への作業時間の見積もりが容易。

【欠点】

- セキュリティ環境を確保出来る者しか利用することができない。

オンサイト開設後



【利点】

- セキュリティ環境を確保出来なかったデータ利用希望者も活用の道が開かれる。
- 新たな利用環境での利用者の増加も見込まれる。

【欠点】

- データセンター側にオンサイトでの分析作業がどの程度の影響を及ぼすのかわからない。

作業区分別の考え得る課題と運用目的の研究での作業評価項目例

作業区分	考え得る課題	具体的な作業内容
データハンドリング	多数の条件(例:「傷病名に『糖尿病』がある」または「インスリン利用歴がある」または「HbA1cが測定されている」または「食事指導を受けている」または「SU剤を服用している」または…または…)で絞り込んで抽出するような複雑な作業が集中した場合、数十億件のレセプトをその都度閲覧する作業を経なければならないため、負荷がかかりすぎて動かなくなる恐れがある。	データセンターに格納されているデータから、研究者が分析に必要なデータを、どの程度の所要時間、CPU占有時間等々で達成できるかを評価する。
データ解析	患者個人を単位としたパネル作成等の高度な分析がシステム上可能であったとしても、根本的なデータの組み直しを一斉に行う事で、データセンターへの多大な負荷となる可能性がある。	患者単位で時系列で評価可能なパネルを作成し、それに対する統計処理がどの程度可能かを評価する。 ⁴

運用目的の研究の内容、方法(統計手法等)

作業区分		集計表作成向けの現況把握 (クロスセクショナルデザイン)	患者単位・時系列分析向けの現況把握 (コホート・パネルデザイン)
データハンドリング	データの抽出	<ul style="list-style-type: none"> 現行システムで内科・歯科・調剤・DPCデータ、また特定健診データに対し年齢、性別、病名、使用薬剤、実施処置・手術等を検索キーとして対象となる患者を全て抽出可能か検証する。 	<ul style="list-style-type: none"> 現行システムで内科・歯科・調剤・DPC・特定健診データに対し年齢、性別、病名、使用薬剤、実施処置・手術、検診受診後の受診有無・異常値所見の有無等を検索キーとして対象を抽出し、その対象についてその前後の全レセプト情報を抽出し、適切なイベント粒度、順序で整列可能か検証する。
	特徴量抽出 データにみられる特徴的な偏りや変化を評価する数字である。	<ul style="list-style-type: none"> 現行システムで内科・歯科・調剤・DPCデータ、また特定健診に対し、患者単位の粒度で、年齢、性別、病名、受診した医療機関二次医療圏情報、受診した医療機関種別、医療費、特定の医療行為の実施時期(開始・終了)・実施量(実施日数・量)、転帰、健診異常値の有無等の特徴量の集計が可能か検証する(欠損、外れ値等評価)。 	<ul style="list-style-type: none"> 現行システムで、イベント粒度毎に、年齢、性別、病名、居住地市町村、受診した医療機関の所属二次医療圏、受診した医療機関種別、医療費(レセプト区分別)、特定の医療行為の実施時期(開始・終了)・実施量(実施日数・量)、転帰、異常値所見等の特徴量の集計が可能か検証する。
データの解析		<ul style="list-style-type: none"> 年齢層・性・2次医療圏、医療機関種別の医療費(入院・外来・歯科・薬剤) 年齢層・性・2次医療圏別の特定健診の受診率・異常所見割合 (可能な範囲で)上記の間の地域相関 年齢層・性・2次医療圏、病院種別の医療行為・薬剤利用などの比較(特に悪性新生物と糖尿病・循環器系疾患について) Charlson comorbidity index や、外部のデータから機械学習によって作成されたリスクスコア算出手法がデータに適応出来るか などの解析処理が可能かどうかについて、検証を行う。	<ul style="list-style-type: none"> 医療費予測や、制度変更の効果検証に関する分析 特定の入院処置などを受けたあとの入院・外来・薬剤医療費の動向 年齢や性・2次医療圏、病院種別比較を実施 特定健診受診の有無ないし健診異常の有無と入院・外来・薬剤医療費の動向 特定健診データと医療費データとの連結のテスト パネルデータを用いた解析 固定効果モデルやMCMCを用いたベイズモデリングなどを用いた、制度変更の影響 などの解析処理が可能かどうかについて、検証を行う。

運用目的の研究で網羅的に情報を扱う点について

これまでのNDBデータ提供における「必要最小限のデータ提供」 「探索的研究の禁止」の考え方との整合性

- 本研究は特定の目的を有した個別の研究として位置づけられるものではなく、今後オンサイトリサーチセンターに寄せられる様々な目的の研究に対し、最適な環境を提供するための現況把握として位置づけられるものである。
- 本研究によって得られた知見は報告書のみ公表を予定し、また、そこでの検討事項は、各種作業の実行過程で、作業実行の可否、あるいは実行可能時の所要時間やCPU占有率等が明示されるのみとし、検証に用いた解析データは持出さない。
- 以上の理由により、従来の審査基準を逸脱する申出内容を含んでいるものの、本研究における目的を最大限達成するために、レセプトに記載されているに関する全ての情報について、可能な限り長期間分のデータ利用を申出するものである。

運用目的の研究の成果物の公表イメージ

【報告書の内容 : オンサイトリサーチセンターの提供側／利用者側への参考資料】

- 成果物は、有識者会議への報告、東京大学、京都大学も含めたレセプト情報等オンサイトリサーチセンター連絡協議会、ならびに(予定)「レセプト情報等データベースの利活用の推進に関する研究」へ報告書とする予定である。
- いわゆる学術領域の知見として、論文等の媒体でデータそのものを公表する事を行わない。
- 提供側／利用者側に有用と思われる事項として、主に以下について検証を行いつつ、効率的にオンサイトリサーチセンターを運用するための、その他の課題等についても、可能な限り明示する。

・【データハンドリングの検証: 所要時間の把握】

オンサイトリサーチセンターにおけるデータ抽出、ならびにデータ集計作業の標準想定時間、ならびに負荷過剰となる作業について、具体例を報告する。

・【データ解析の検証: 当該システム下で実施可能な統計手法の確認】

オンサイトリサーチセンターには統計ソフト(SAS, R)が導入されているが、これらのソフトを用いた分析を行う際、主に研究内容に焦点をあて、どのような課題が適切に遂行可能で、どのような課題が遂行困難であるか、等を具体的に紹介・報告する。

・【データ解析の検証: 所要時間の把握】

オンサイトリサーチセンターにおける各種解析作業にどの程度の時間を要するかを具体的に把握する。ここで得られた知見は、試行期間終了後には、承諾された申出者がオンサイトリサーチセンターを利用する際の、アクセス権限の時期管理を行う際に活用される。

運用目的の研究報告例

作業例	実行可否	所要時間	CPU占有時間	その他
傷病名をキーとしたレセプト抽出	可能	1年分データを対象とした場合 ○○時間 5年分データを対象とした場合 △△時間	1年分データを対象とした場合 ●●時間 5年分データを対象とした場合 ▲▲時間	<ul style="list-style-type: none"> 傷病名单独であれば、100件までandで重複させても、所要時間に変化はみられなかったが、orで集計すると所要時間はデータ数に比例して長くなる傾向がみられた。 「疑い病名」を除外して集計する場合は、約1.5倍の所要時間を要する。
傷病名・診療行為・医薬品をキーとしたレセプト抽出	可能	1年分データを対象とした場合 ○○○時間 5年分データを対象とした場合 △△△時間	1年分データを対象とした場合 ●●●時間 5年分データを対象とした場合 ▲▲▲時間	<ul style="list-style-type: none"> 3つの項目それぞれにおいて、複数の事例でor検索をかけると、更に所要時間が延びる傾向がみられた。 端末1台でこの処理を行う場合は毎回比較的近い結果が得られたが、端末4第で同時にこの処理を行う場合は、所要時間に大きな差がみられた。
...	• ...