

2019年度大学教育再生戦略推進費
「医療データ人材育成拠点形成事業」
申請書

【様式1】

事業の構想等

代表校名 (連携校名)	東京大学 (筑波大学、富山大学、自治医科大学) 計4大学
事業名 (全角20字以内)	医療リアルワールドデータ活用人材育成事業

1. 事業の構想 ※事業の全体像を示した資料(ポンチ絵A4横1枚)を末尾に添付すること。

(1) 全体構想

①事業の概要等

電子カルテシステムの普及やデータ収集基盤の社会的整備が進み、大規模な医療データの創出が進んでいるが、**この大規模データから知見を得て、国民の健康課題の解決に資することが必須である。**しかし、医療現場から創出される大規模医療リアルワールドデータ(RWD)を適切に解析するためには**高い専門性を持って意味あるデータを抽出、形成し、課題にマッチしたデータ分析を行える総合力と実践力を有する人材の育成が喫緊の課題**である。そこで本事業では、大規模医療RWDについて1) **データ特性・意義やバイアスを理解**し、2) データ標準化と変換及びクレンジングにより**解析可能な形式のデータベースに構築**し、3) 具体的な**医療課題解決と知見創成に必要なデータ処理技術を習得**し、4) それを自ら実践でき**指導者層にもなりうる人材**を、種々の医療生データを素材として実践的技術と知識を修得できる教育コースを設置し、**高度な医療人材の育成を推進**する。

②大学・研究科等の教育理念・使命(ミッション)・人材養成目的との関係

2015年に策定した「**東京大学ビジョン2020**」では、公共的な視点から主体的に行動し新たな価値創造に挑む「**知のプロフェッショナル**」**人材の育成を最重要ビジョンの一つ**として掲げている。また、**本医学系研究科**は、疾病の克服および健康の増進に寄与する最先端研究を推進するとともに、**医学系領域の各分野において卓越した学識と高度な独創的研究能力を有する国際的リーダーを養成することを目的**として掲げているところである。

その上で、本事業の課題でもある、医療現場から創出される大規模なデータの解析は世界的にも未だ発展途上の段階であり、さらなる学術的成長が求められている領域であるとともに、現実世界に対する成果が強く求められている領域である。全ての医療課題に臨床試験を計画し、エビデンスを創出することは現実的ではなく、現存するリアルワールドデータをどのように活用するかについて独創的なリーダーシップを発揮できる人材の育成は我が国だけにとどまらず世界的に望まれている。

こうした背景等を踏まえ、**本事業は、我が国の大規模な医療リアルワールドデータから新規知見を創出し、成果を世界へ発信できる「知のプロフェッショナル」人材の育成を目的としており、我が国の医学研究の更なる発展のために重要なミッションであることから、本教育プログラムが本学及び医学系研究科の理念とも十分に合致するものである。**

③新規性・独創性

本教育プログラムでは、保健医療領域特有の汚さや癖・歪みを持つ大規模医療データのハンドリングに関して即戦力と実践力を持ち、独り立ちできる人材を育成するため、**東京大学大学院医学系研究科の10年以上にわたる公衆衛生学専門職修士の人材育成での経験を医療現場における大規模データ解析人材育成に特化させ、課題解決に結びつけるための実践的な人材育成プログラムを新たに策定**する。本プログラムの**新規性**として、

①多様な医療の大規模生データ(10万件以上)を実際に扱う複数の研究機関のデータ解析現場に常駐し、**直接スタッフとともに大規模生データを解析する「データ実地研修科目」を必修**とする。

②医療データを扱う実務経験者も対象に、**匿名加工後データを医療機関外で扱う人材も養成**。

③大規模データ収集・クレンジング・解析を一人でを行う技術を習得した上で、役割分担できるチームを編成して課題解決するために、**データ分析の実務マネジメント力を養う科目を設定**する。

④本プログラムで利用する大規模な生データを対象に、**履修生と議論を行いながら実際に匿名加工を行う**。さらにこれらの**匿名加工した成果物は、今後の大規模医療データを使用する新たな教育リソースとして多種類にわたって公開し、後続の他の人材育成事業でも活用できる教育資源**とする。

⑤仕事に従事中の医療職者を対象にするため、「データ実地研修科目」を除き必修科目は**すべて新規に平日18時以降か土曜に開講**し、既存の大学院コースが提供する科目の読み替えによる利用は一部の選択科目を除き原則として行わないことが挙げられる。さらにプログラムでは、

①「**講義科目**」、**講義と実習をセット化した「実践科目**」、「**データ実地研修科目**」の**3区分**を用意し、一般履修コースとインテンシブコース両方でこの各区分を必修とする。講義科目と実践科目は1科目単位を90分8コマとコンパクトにし、さらに講義科目コマのうち少なくとも2コマは必ずケーススタディー形式をとることで、講義形式で学んだ知識を実際の課題にどう適用するかを**講師と受講者の双方向議論することで実学知識化**する。

②代表校の**工学系研究科で実績と定評のあるAI・機械学習コース、学際情報学府の生物統計情報学コースとも連携協力**して、履修生に背景知識に応じた**幅広い科目選択機会**を与える。

③履修生の達成度評価を、講義・実習・データ実地研修での出席率、議論、試験やレポート、といった従来型の評価方法に加えて、**1名30分の個別面接評価試験**(実データによる実技プレゼンと3名程度の試験委員による質疑応答)を実施し、**技術応用力、問題解決力、チームマネジメント力の視点から評価する**といった点が**特徴**であり、これまでにない**独創性を有するプログラム**である。

④達成目標・評価指標

本教育プログラムでは、**実践科目・知識科目・実地研修科目の合計9科目の教育カリキュラムを新たに開発**するとともに、以下の2つのコースを立ち上げ事業を展開する。

一般履修コース : 開設時期 2020年度4月、履修期間 2年
履修者数 2020年度:10名、2021年度:10名、2022年度:10名、2023年度:10名(合計40名)

インテンシブコース: 開設時期 2020年度4月、履修期間 1年以内
履修者数 2020年度:8名、2021年度:8名、2022年度:8名、2023年度:8名(合計32名)

個人ごとに**ペーパーによる試験で知識を評価し、1名30分の個別面接評価試験**(プレゼンと3名程度の試験委員による質疑応答)を実施し、両方の成績を総合的に評価して、成績をS(優れた習熟が得られた)、A(十分な習熟が得られた)、B(一定程度習熟できた)、C(やや習熟が足りない)、D(未習熟の部分が多かった)のように個人評価を行う。その上で、評価指標として、**A以上の割合およびB以上の割合、Dの割合を採用し、A以上が10%以上でかつB以上が80%以上かつD5%以下となることを目標**とする。

(2) 教育プログラム・コース → 【様式2】

2. 事業の実現可能性

(1) 運営体制

①実施体制

本教育プログラムは、代表校である東京大学内では医学部附属病院と医学系研究科が連携して実施し、**医学系研究科長が事業責任者として事業を総括**する。事業の運営に当たっては関係講座と事務による教職員によって**運営調整会議を構成**し、会議の長は医療データ管理・処理を専門とする大学院医療情報学分野教授（病院企画情報運営部長兼務）があたる。同会議のもとに**カリキュラム委員会、事業評価委員会、教材データ品質整備委員会及び運営事務局**を設置する。

1) カリキュラム委員会は詳細なカリキュラムの作成と必要に応じた途中改革、履修認定等の決定等の教学面の統括管理を行う。

2) 事業評価委員会は外部有識者3名以上および連携校からの教員を交えた構成により、事業としての進捗評価と達成度評価を行う。また自己評価および履修者からのアンケートによる評価を指揮する。

3) 教材データ品質整備委員会は、教材作成のガイドラインを整備し、実習や実地研修で使用されるデータの規模や質の妥当性および倫理的妥当性のチェックし、教材資料について科目間もしくは講師間の品質のばらつきを小さくなるように管理する役割を担う。

4) 運営事務局では①履修生の募集、②授業の実施支援、③遠隔講義環境の整備と運用、など全般的な事業運営の実務をマネジメントする。

以上により運営調整会議のもとで本プログラムの立ち上げ時から**機動性を有した事業の円滑な実施を図る**。また、プログラム開始前と各年度に連携校や協力機関・協力教員との**事業連携協議会を年間2回以上開催**し、事業の運営全般について積極的な連携を図るとともに、**拠点形成による事業継続へ向けた体制構築**を進めていく。

また、地方都市等からの履修のケースでは、受講生個人毎PCからのリアルタイム双方向遠隔講義履修環境を整備し、計算機環境とデータアクセス環境をクラウドコンピューティングにより提供し完全なリアルタイム遠隔実習の体制を構築する。

事業実施にあたり講義実習等を担当するとともに運営体制にも加わる機関等は以下のとおりである。

代表校：東京大学、**連携校**：筑波大学、富山大学、自治医科大学、**協力校**：国際医療福祉大学

②評価体制

プログラム開始後に継続的に事業内容を改善するため、各年度末に行われる連携校との事業連携協議会において、自己評価及び履修者からのアンケート、および各科目担当責任教員の科目評価を行い、当該年度の**事業内容の自己評価**を行う。特に履修者による科目評価と担当教員評価を5段階スケールで収集する。さらに**外部有識者を3名以上含む事業評価委員会を常設**し、評価基準表を作成してそれに基づき講義と実習を実際に視察評価し、また実施状況点検を行う。事業評価委員会は、前記の自己評価結果とこれらの評価結果を総合して**年に1度、総合客観評価を行うことで、運営調整会議は同評価に基づく継続的なプログラム改善・強化が図られ、PDCAサイクルによる事業展開が確立**される。

③連携体制（連携校との役割分担等）

事業の実施にあたっては**連携校が強みを持つ特定領域の授業を積極的に本プログラムに取り入れ**、参画するそれぞれの連携校の優れた教育を組み合わせることで、単独校ではなし得ない高度な教育プログラムの構築を行う。また、それぞれが有するリアルワールドデータを教育に用いることで、**バリエーションに富んだデータに触れる機会を履修生に提供**する。具体的には以下のとおり。

筑波大学：医療データの自然言語処理に関するアノテーションと形態素解析などの自然言語処理により意味のあるデータを抽出する手法を奈良先端科学技術大学院らとともに実施しているため、この領域の教育資源を電子カルテデータと連結して扱う手法について遠隔講義と実習を行う。また、ヘルスサービスリサーチ研究に資する介護保険レセプトデータの分析、ゲノム医療データ分析について遠隔講義を担当する。

富山大学：全国糖尿病データベース事業であるJ-DREAMSに関わってきた経験に加え、現在国民健康保険の特定健診データを用いてすでに地域の健康課題の可視化に取り組んでいる実績があり、これらの経験とデータを生かした遠隔講義と遠隔実習を行う。加えて循環器領域の臨床研究症例レジストリを用いて臨床研究でのデータ解析に関する遠隔講義と遠隔実習を行い、また北陸地域での教育拠点として履修生の募集を行う。

自治医科大学：全国の自治体で活躍する出身者の医院や病院から収集した診療データを有し、これを教育資源とした遠隔講義と実習を行う。また自治体から集積している健診・レセプトデータの資源を活用した実地研修に協力する。

協力機関として国際医療福祉大学（担当教員：石川ベンジャミン光一医学研究科教授）とも連携し、医師以外の医療職を目指す大学院生に履修機会を提供するとともに、同大学が有する九州地区との遠隔講義システム等を活用して、地方からの遠隔履修者への履修機会を拡大する。

その他、協力機関として、**(1)一般財団法人医療情報システム開発センター(MEDIS-DC)が参画し**、標準病名マスター、標準臨床検査マスタ、標準医薬品マスターHOTコードなど同財団がメンテナンスする標準化業務を実地で説明を受け業務の一部研修することに協力する。**(2)一般財団法人匿名加工医療情報公正利用促進機構(FAST-HDJ)が参画し**、医療機関からのデータと収集と名寄せ、および匿名加工処理について実地研修可能な環境を提供する。

④連携体制（自治体、企業等との連携体制や連携のメリット等）

協力教員として、**①医療経済研究機構** 満武巨裕研究副部長（AMED研究事業で、NDBの分析、自治体レセプト分析の実績）、**②国立国際医療研究センター情報基盤センター** 美代賢吾センター長（糖尿病症例DB構築事業J-DREAMS、セキュリティ管理）、**③国立保健医療科学院保健医療経済評価センター** 福田敬センター長、佐藤大介主任研究官（医療資源配置の全国調査データ分析）、**④FAST-HDJ** 山本隆一理事長（元東京大学情報学環准教授、次世代医療基盤法に基づく匿名加工事業）、**⑤ひかり綜合法律事務所** 板倉陽一郎弁護士（データ保護法、IT関連法、知的財産権法等）、**⑥東北大学大学院医学系研究科医学情報学分野** 中山雅晴教授（宮城地域医療ネットワークMMWINのデータ整備）、**⑦九州大学病院メディカルインフォメーションセンター** 中島直樹教授（MID-NET事業における標準化、JLAC10変換）、**⑧千葉大学予防医学センター** 藤田伸輔教授（予防医学関連エビデンス分析）、**⑨奈良先端科学技術大学** 荒牧英治准教授（医療自然言語処理の実践）、**⑩帝京大学医療情報システム研究センター** 澤智博教授の各氏がそれぞれ括弧内記載の担当として講義・実習の実施もしくは関連する科目での教材資料提供や助言をする協力体制が敷かれている。

また、ゲノム医療データベースシステム構築のケーススタディー担当として、実システムを開発・運用する富士通(株)関係部署が協力機関となり体制を整備している。多くの公的な症例レジストリーDB構築事業において医療機関側データ変換出力を担当する企業や、自治体からレセプトデータ分析によるアドバイス事業を請け負う企業とも連携し、履修生の多様な実体験ケースとなるよう体制を組んでいる。

(2) 継続・事業成果の普及に関する構想等

①継続に関する構想

本教育プログラムにおいて育成する人材は、近い将来次世代医療基盤法に基づいて医療機関外へ供される医療データから、課題解決のための知見を創出するために必須の人材であり、事業の継続によって国民の健康に有益となる医学的知見の持続的な創出体系を我が国に確立することが重要である。このため、事業立ち上げ後から連携校とともにプログラム継続に向けて拠点としての体制構築を行い、4年ないし5年後に**連合大学院修士課程に相当するようなバーチャルな修士課程の開講を目指す**。また、自己財源確保の手段として、本プログラムの講義および遠隔実習の一部を、企業在籍者が有償で受講できるように、**本教育プログラムおよびオンライン教材を東京大学TL0に知財登録して企業に配信させ、対価収入を得る仕組み**を動かすことを目指す。

②事業成果の普及に関する計画

本教育プログラムを知財として東京大学TL0に登録し、医療データ人材育成教育を提供する**民間企業や非営利団体に有償でライセンス**することで、履修生を増やす。また、特色でも記載したように、本プログラム実施のために作成された実習資源や実地研修で使用された大規模な生データを対象に、履修生が履修過程で匿名加工した成果物を、**今後の医療データ人材育成にも供するオープンでフリーの教育資源として公開**を目指す。これにより、後続の他の人材育成事業でも活用できる教育資源とすることができる。また、**リアルワールドデータを解析できる人材の価値・育成の重要性について各種研修会やセミナー、ホームページ等で広く情報発信**を行う。さらに本プログラムでの成果物として**教科書の出版を目指す**。

3. 実施計画

(1) 年度別の計画

2019年度	<ul style="list-style-type: none"> ①7月 学内運営チームを組織。 ②9月 キックオフシンポジウムの開催・第1回連携協議会開催。 ③9月 コア・カリキュラムの作成を開始。教材・データの作成開始 9月～2月：5回程度の公開講義（シンポジウム形式）を行うトライアルコースを開催し、これによりアンケート調査を行って、策定中のカリキュラム改良に反映する。 9月～2月：リアルタイム遠隔講義・実習環境を導入・整備 ④10月 実習用データの教育利用の倫理承認へ向けた準備の開始。 ⑤10月 コア・カリキュラムに沿った教育・実習プログラムの作成を開始。 ⑥11月 教育・実習プログラムの作成完了 ⑦12月 受講者の全国公募開始 ⑧3月 受講者の決定。第2回連携協議会開催。年度評価実施。 データ利用各倫理承認の完了 教材・データの作成完了
2020年度	<ul style="list-style-type: none"> ①4月 一般コース、インテンシブコース開始。 ②9月 第3回連携協議会開催(中間評価へ向けた体制の見直し、評価を実施) ③12月 受講者の全国公募開始。 ④3月 受講者の決定。 ⑤3月 第4回連絡協議会開催。拠点体制構築調整。年度評価実施。
2021年度	<ul style="list-style-type: none"> ①4月 通常コース、インテンシブコースの継続。 ②9月 第5回連携協議会開催。補助金以外の確保財源による事業継続の調整。教育プログラムのTL0への登録と、ライセンス先調整の開始。 ③12月 受講者の全国公募開始。 ④3月 受講者の決定。 ⑤3月 第6回連絡協議会開催。年度評価実施。
2022年度	<ul style="list-style-type: none"> ①4月 通常コース、インテンシブコースの継続。 ②10月 第7回連携協議会開催。拠点体制の確立。 ③12月 受講者の全国公募開始。 ④3月 受講者の決定。 ⑤3月 第8回連絡協議会開催。年度評価実施。
2023年度	<ul style="list-style-type: none"> ①4月 通常コース、インテンシブコースの継続。 ②10月 第9回連携協議会開催。財政支援完全終了後の財源確保体制の確立。 ③12月 受講者の全国公募開始。 ④3月 受講者の決定。 ⑤3月 第10回連絡協議会開催。事業総括及び事業継続体制の確認。

教育プログラム・コースの概要

大学名等	東京大学大学院医学系研究科
教育プログラム・コース名	医療リアルワールドデータ活用人材育成事業 一般履修コース
対象者	以下のいずれかであるもの 1) 医師、看護師、その他健康医療介護職の国家資格または日本医学会分科会等の医学医療系学会の認定資格等を有する者 2) 健康管理や医療における業務、それらのデータ分析業務経験を3年以上有する者で直近経験後3年以内の者
修業年限（期間）	2年
養成すべき人材像	1)保健医療現場から発生する多様なデータの特性、医学的意義、リアルデータ特有のバイアス（特定の条件に依存して収集される偏り）の原理や実態を理解し、2)データのコード等識別子と形式の標準化変換及びクレンジングにより、対象となるデータを課題解決に資する解析可能な形式のデータベースに再編成でき、3)具体的な医療課題解決と知見創成に必要な分析手順作成方法とAI・機械学習を含む新しい技術を含めたデータ処理技術を習得し、4)自ら手順を立てて必要ならチームを組んでデータ収集と分析を実践でき、 医療リアルワールドデータ基盤を構築でき、指導者層にもなりうる人材であり、すなわち「即戦力と実践力のある医療データハンドリング独り立ち人材」 である。
修了要件・履修方法	本教育プログラム・コースにおける講義科目を3科目36時間（90分×24コマ）以上、実践科目を8科目計96時間（90分×64コマ）以上、実地研修を2科目30時間以上の合計162時間以上を履修すること。上記を満たした者に修了証を発行する。
履修科目等	履修科目は「講義科目」、講義と実習をセット化した「実践科目」、「データ実地研修科目」の3区分に別れ、この各区分を必修とする。講義科目と実践科目は1科目単位を90分8コマ（1単位相当）とコンパクトにし、さらに講義科目は少なくとも2コマは必ずケーススタディー形式をとることで、講義形式で学んだ知識を実際の課題にどう適用するかを講師と受講者の双方向議論することで実学知識化する。 「データ実地研修科目」は1日3時間以上×5日間を1科目とし、データ解析またはデータ構築部門に常駐して、スタッフとともに課題解決型の実践的仕事を手伝うことで実践力と問題解決力を修得する科目で、3単位を算定する。 <必修> 知識科目3科目 ：①倫理・法制度概論（個人情報保護、匿名加工と次世代医療基盤法、研究倫理指針）、②医療データ管理（クラウド管理、データ・セキュリティ技術、暗号化、異なるデータ資源の統合化、リレーショナル・データベースとSQL言語、データベーススキーマ設計とデータ変換）、③医療リアルワールドデータ構築概論（NDB、SS-MIX2、MID-NET、J-DREAMS/J-CKD/J-ImpactレジストリDB、健診と生活圏データ、個人ゲノムデータ） 実践科目8科目 ①データ収集・コード化・標準化・クレンジング実践、②データ再編成と匿名加工実践、③SS-MIX2標準化ストレージ処理と二次利用データベース構築実践、④症例レジストリデータベース処理実践、⑤電子カルテ・自由記載データ自然言語処理実践、⑥臨床課題分析と実務マネジメント、⑦統計的データ解析・マイニング実践Ⅰ、⑧統計的データ解析・マイニング実践Ⅱ 実地研修2科目 ：①レセプト・特定健診データハンドリング、②臨床症例レジストリDBハンドリング、③SS-MIX2・MID-NETデータハンドリング、④医療資源調査データ分析、から2科目選択必修。 <選択科目>①Python入門、②Rプログラム入門、③AI・機械学習入門、④自然言語処理基礎、⑤臨床医学概論、⑥ゲノム解析入門、⑦医療画像のアノテーションと深層学習入門 などを新規開講および既存開講科目から選択可能とする。特にプログラミング言語であるPythonおよびRの入門的な授業を設ける他、医療に直接関わった経験のない受講者を対象とした臨床医学概論等を選択科目として設ける。

<p>教育内容の特色等 (新規性・獨創性)</p>	<p>講義と実習を一体化した実践科目、およびインターンシップのように実地環境で実地データをスタッフとともに処理することで実データの汚さを自分の手で手順を立てて整備し、与えられた課題を分析し、結果を出せる技能を身につけさせる。</p> <p>すべての実践科目において、実際のデータを扱う。用意されるデータ源として、(A)代表校由来の10万人分の非標準化、未整形、クレンジング前の電子カルテ生データで仮名化ID変換後にデータウェアハウスに格納されたデータ(検体検査、病名、処方、注射データ、手術データ)、(B)代表校と他2校がMID-NETに参画し作成した20万件規模のSS-MIX2標準化ストレージとレセプト由来の仮名化シャッフルされた標準化済みサンプルデータベース(MID-NETの全種別データ:検体検査、病名、処方、注射データ、レセプト)、(C)個人識別情報のない代表校由来の1000人分規模の退院時サマリまたは現病歴文章の未整形の自然言語処理を必要とする生テキストデータ、(D)代表校が症例レジストリ事業(J-DREAMS/J-CKD/J-Impactの各事業)にデータ提供した際の仮名化ID後の生症例登録データ、(E)自治体等から収集した多施設由来のレセプトと特定健診データ、の5種類を準備する。</p> <p>実践科目「①データ収集・コード化・標準化・クレンジング実践」と「②データ再編成と匿名加工実践」では主としてリソースA、「③SS-MIX2標準化ストレージ処理と二次利用データベース構築実践」ではリソースB、「④症例レジストリデータベース処理実践」ではリソースD、「⑤電子カルテ・自由記載データ自然言語処理実践」ではリソースC、「⑦統計的データ解析・マイニング実践Ⅰ」「⑧統計的データ解析・マイニング実践Ⅱ」ではリソースAを①②により履修生自身で整形したデータセットを使用し、一部でリソースEを使用する。各実践科目では、まず汚い生データを履修生自ら①②により整形データセットを作成させる。それを③により二次利用DBを作成し、④のレジストリの登録できる品質にさせる。この過程で作成したデータセットを⑦⑧で使用させることで、自身の前処理のどこが悪くどこが有用かを実際に体験させる。このように履修生が自分の実践科目で使うデータ・セットを自分で生データから処理過程を踏んで作成していくことが本実践科目の大きな特色である。なお、「⑥臨床課題分析と実務マネジメント」は直接データを使わないが、課題を実践に移すプロセスを教授する。</p>																																										
<p>指導体制</p>	<p>各科目責任教員を設定し、科目責任教員はカリキュラム委員会で教員間の調整協議を行う。科目責任教員は、科目の各コマを担当する担当教員とは別に、実践科目では2名以上の補助教員、講義科目では1名以上の補助教員を割当て、履修中の質問や逡巡に対応する。また実践科目では、履修生を履修開始前の調査書評価により、2ないし3名ずつにわけたチームを編成し、このチーム内で相互に議論と助言を出し合うことで自律的に実践できるように指導する。実践科目を遠隔で受ける履修生には、連携校で実践科目補助者を1名以上(履修生5名程度あたりに1名)配置して指導にあたる。データ実地研修では、研修担当教員のもとに、研修担当スタッフを1名割当て、実地指導を行う。</p>																																										
<p>教育プログラム・コース修了者のキャリアパス構想</p>	<ul style="list-style-type: none"> ・所属先医療機関(大学病院等):臨床業務に継続しながら将来的に医療リアルワールドデータの解析に参画し、プロジェクトにおいて指導的立場を担う。 ・大学院医学系研究科:医療リアルワールドデータ解析の研究・教育を行い、医療データ解析部門等で研究員、教員となる。また企業の研究所や匿名加工事業者への就職も視野に入る。 ・医療データ解析を行う企業等:企業内の解析プロジェクトの指導者となる。 																																										
<p>受入開始時期</p>	<p>2020年4月</p>																																										
<p>受入目標人数</p>	<table border="1"> <thead> <tr> <th>対象者</th> <th>2019年度</th> <th>2020年度</th> <th>2021年度</th> <th>2022年度</th> <th>2023年度</th> <th>計</th> </tr> </thead> <tbody> <tr> <td>医師</td> <td>0</td> <td>5</td> <td>5</td> <td>5</td> <td>5</td> <td>20</td> </tr> <tr> <td>看護師</td> <td>0</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>4</td> </tr> <tr> <td>メディカルスタッフ</td> <td>0</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>8</td> </tr> <tr> <td>その他社会人</td> <td>0</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>8</td> </tr> <tr> <td>計</td> <td>0</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>40</td> </tr> </tbody> </table>	対象者	2019年度	2020年度	2021年度	2022年度	2023年度	計	医師	0	5	5	5	5	20	看護師	0	1	1	1	1	4	メディカルスタッフ	0	2	2	2	2	8	その他社会人	0	2	2	2	2	8	計	0	10	10	10	10	40
対象者	2019年度	2020年度	2021年度	2022年度	2023年度	計																																					
医師	0	5	5	5	5	20																																					
看護師	0	1	1	1	1	4																																					
メディカルスタッフ	0	2	2	2	2	8																																					
その他社会人	0	2	2	2	2	8																																					
計	0	10	10	10	10	40																																					

教育プログラム・コースの概要

大学名等	東京大学大学院医学系研究科
教育プログラム・コース名	医療リアルワールドデータ活用人材育成事業インテンシブコース【インテンシブ】
対象者	以下のいずれかであるもの 1) 医師、看護師、その他健康医療介護職の国家資格または日本医学会分科会等の医学医療系学会の認定資格等を有する者 2) 健康管理や医療における業務、それらのデータ分析業務経験を3年以上有する者で直近経験後3年以内の者
修業年限(期間)	1年
養成すべき人材像	1) 保健医療現場から発生する多様なデータの特性、医学的意義、リアルデータ特有のバイアス(特定の条件に依存して収集される偏り)の原理や実態を理解し、2) データのコード等識別子と形式の標準化変換及びクレンジングにより、対象となるデータを課題解決に資する解析可能な形式のデータベースに再編成でき、3) AI・機械学習を含む新しい技術を含めたデータ処理技術を習得して、医療リアルワールドデータ収集と整形および分析を支援できる人材
修了要件・履修方法	本教育プログラム・コースで定める科目について、講義科目を3科目24時間(90分×16コマ)以上、実践科目を4科目計48時間(90分×32コマ)以上、実地研修を1科目15時間以上の合計87時間以上を履修すること。上記を満たした者に修了証を発行する。
履修科目等	履修科目は「講義科目」、講義と実習をセット化した「実践科目」、「データ実地研修科目」の3区分に別れ、この各区分を必修とする。講義科目と実践科目は1科目単位を90分8コマ(1単位相当)とコンパクトにし、さらに講義科目ではケーススタディー形式をとる1コマを必修とすることで、講義形式で学んだ知識を実際の課題にどう適用するかを講師と受講者の双方向議論することで実学知識化する。 「データ実地研修科目」は1日3時間以上×5日間を1科目とし、データ解析またはデータ構築部門に常駐して、スタッフとともに課題解決型の実践的仕事を手伝うことで実践力と問題解決力を修得する科目で、3単位を算定する。 <必修> 知識科目3科目から指定計16コマ ：①倫理・法制度概論より5コマ以上(個人情報保護、匿名加工と次世代医療基盤法、研究倫理指針)、②医療データ管理より5コマ以上(クラウド管理、データ・セキュリティー技術、暗号化、異なるデータ資源の統合化、リレーショナル・データベースとSQL言語、データベーススキーマ設計とデータ変換)、③医療リアルワールドデータ構築概論より5コマ以上(NDB、SS-MIX2、MID-NET、J-DREAMS/J-CKD/J-ImpactレジストリDB、健診と生活圏データ、個人ゲノムデータ) 実践科目4科目(一般履修コース8科目から指定4科目) ：①データ収集・コード化・標準化・クレンジング実践、②データ再編成と匿名加工実践、③SS-MIX2標準化ストレージ処理と二次利用データベース構築実践、④統計的データ解析・マイニング実践I 実地研修1科目 ：①レセプト・特定健診データハンドリング、②臨床症例レジストリDBハンドリング、③SS-MIX2・MID-NETデータハンドリング、④医療資源調査データ分析、から1科目選択必修。 <選択科目>①Python入門、②Rプログラム入門、③AI・機械学習入門、④自然言語処理基礎、⑤臨床医学概論、⑥ゲノム解析入門、⑦医療画像のアノテーションと深層学習入門などを新規開講および既存開講科目から選択可能とする。特にプログラミング言語であるPythonおよびRの入門的な授業を設ける他、医療に直接関わった経験のない受講者を対象とした臨床医学概論等を選択科目として設ける。

<p>教育内容の特色等 (新規性・獨創性)</p>	<p>講義と実習を一体化した実践科目、およびインターンシップのように実地環境で実地データをスタッフとともに処理することで実データの汚さを自分の手で手順を立てて整備し、与えられた課題を分析できる結果を出せる技能を身につけさせる。</p> <p>すべての実践科目において、実際のデータを扱う。用意されるデータ源は、(A)代表校由来の10万人分の非標準化、未整形、クレンジング前の電子カルテ生データで仮名化ID変換後にデータウェアハウスに格納されたデータ(検体検査、病名、処方、注射データ、手術データ)、(B)代表校と他2校がMID-NETに参画し作成した20万件規模のSS-MIX2標準化ストレージとレセプト由来の仮名化シャッフルされた標準化済みサンプルデータベース(MID-NETの全種別データ:検体検査、病名、処方、注射データ、レセプト)、(E)自治体等から収集した多施設由来のレセプトと特定健診データ、の5種類のリアルワールドデータを準備する。の3種類のリアルワールドデータを準備する。一般履修コースで使用するリソース(C)(D)は本コースでは使用しない。</p> <p>実践科目「①データ収集・コード化・標準化・クレンジング実践」と「②データ再編成と匿名加工実践」では主としてリソースA、「③SS-MIX2標準化ストレージ処理と二次利用データベース構築実践」ではリソースB、「④統計的データ解析・マイニング実践I」ではリソースAを①②により履修生自身で整形したデータセットを使用し、一部でリソースEを使用する。各実践科目では、まず汚い生データを履修生自ら①②により整形データセットを作成させる。それを③により二次利用DBを作成させる。この過程で作成したデータセットを④で使用させることで、自身の前処理のどこが悪くどこが有用かを実際に体験させる。このように履修生が自分の実践科目で使うデータ・セットを自分で生データから処理過程を踏んで作成していくことが本実践科目の大きな特色である。</p> <p>医学分野において、実践的かつ体系的にデータ解析技術および関連分野の知識を習得できるプログラムはこれまで限られており、1年という期間に集中してこれらを学ぶ本プログラムはこれまでにない実用性を有する。</p>																																										
<p>指導体制</p>	<p>各科目責任教員を設定し、科目責任教員はカリキュラム委員会で教員間の調整協議を行う。科目責任教員は、科目の各コマを担当する担当教員とは別に、実践科目では2名以上の補助教員、講義科目では1名以上の補助教員を割当て、履修中の質問や逡巡に対応する。また実践科目では、履修生を履修開始前の調査書評価により、2ないし3名ずつにわたったチームを編成し、このチーム内で相互に議論と助言を出し合うことで自律的に実践できるように指導する。実践科目を遠隔で受ける履修生には、連携校で実践科目補助者を1名以上(履修生5名程度あたりに1名)配置して指導にあたる。データ実地研修では、研修担当教員のもとに、研修担当スタッフを1名割当て、実地指導を行う。</p>																																										
<p>教育プログラム・コース修了者のキャリアパス構想</p>	<ul style="list-style-type: none"> ・所属先医療機関(大学病院等):臨床業務に継続しながら将来的に医療リアルワールドデータの解析に参画する。 ・大学院医学系研究科:医療リアルワールドデータ解析の研究・教育を行い、医療データ解析部門等で研究員、教員となる。また企業の研究所や匿名加工事業者への就職も視野に入る。 ・医療データ解析を行う企業等:企業内の解析プロジェクトの指導者となる。 																																										
<p>受入開始時期</p>	<p>2020年4月</p>																																										
<p>受入目標人数</p>	<table border="1"> <thead> <tr> <th>対象者</th> <th>2019年度</th> <th>2020年度</th> <th>2021年度</th> <th>2022年度</th> <th>2023年度</th> <th>計</th> </tr> </thead> <tbody> <tr> <td>医師</td> <td>0</td> <td>3</td> <td>3</td> <td>3</td> <td>3</td> <td>12</td> </tr> <tr> <td>看護師</td> <td>0</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>4</td> </tr> <tr> <td>メディカルスタッフ</td> <td>0</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>8</td> </tr> <tr> <td>その他社会人</td> <td>0</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>8</td> </tr> <tr> <td>計</td> <td>0</td> <td>8</td> <td>8</td> <td>8</td> <td>8</td> <td>32</td> </tr> </tbody> </table>	対象者	2019年度	2020年度	2021年度	2022年度	2023年度	計	医師	0	3	3	3	3	12	看護師	0	1	1	1	1	4	メディカルスタッフ	0	2	2	2	2	8	その他社会人	0	2	2	2	2	8	計	0	8	8	8	8	32
対象者	2019年度	2020年度	2021年度	2022年度	2023年度	計																																					
医師	0	3	3	3	3	12																																					
看護師	0	1	1	1	1	4																																					
メディカルスタッフ	0	2	2	2	2	8																																					
その他社会人	0	2	2	2	2	8																																					
計	0	8	8	8	8	32																																					

医療リアルワールドデータ活用人材育成事業

— 代表校：東京大学 連携校：筑波大学、富山大学、自治医科大学（協力機関：国際医療福祉大学） —

我が国の課題

医療現場から創出される大規模医療リアルワールドデータを適切に解析できる人材の不足

問題点

標準化されていない項目値、各種欠損値の存在、実臨床との間に存在するデータの歪みのために、そのままでは解析できない

本事業で育成する人材像

- 医療大規模データの特性・意義やバイアスを理解
- データ標準化・クレンジングによりデータベース再編成ができる
- 具体的な医療課題解決と知見創成に必要なデータ処理技術を習得
- 自らデータ基盤の構築ができ、指導者層になりうる

医療大規模データ活用人材育成の拠点構築

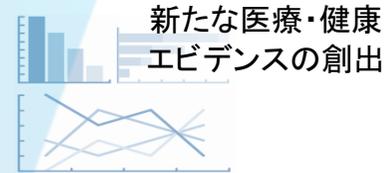
- ・大規模生データ実習と実地研修：各大学等が有する現場データを使用
- ・ケーススタディ：プロジェクトを指導的にマネジメントする能力を身につける
- ・領域専門知識の習得：医療情報学、生物統計学、臨床疫学
個人情報保護や倫理・法制度に関する専門知識

一般コース	集中コース
2年	1年
講義3科目	講義3科目
実践8科目	実践4科目
実地2科目	実地1科目



大規模医療データ解析の一連の工程を自ら技術を持って実践でき、**医療データ基盤を構築でき**
指導的に課題解決できる人材を育成

⑥患者の元へ還元



⑤課題解決を見据えた知見創出

④10万件を超える大規模なデータの適切な解析処理

①医療・保健の現場からデータ抽出

そのままでは解析できない生データ

	Hb	CRP
患者A	16.7	+++
患者B	8600	13.5
患者C	15.5	<0.3
患者D	5800	凝固 6.34
患者E	検査不能	0.54

②データの特性・意義偏りの把握

③データ標準化・整形・クレンジングによるデータベース構築

対象とする受講者 一般10名、集中8名×4年 計72名

医師・看護師を始めとする医療免許保有者
医療機関外で医療関連データ解析の経験を有する者

成果普及へ

教育内容について教科書を出版・オンライン教材として公開
大規模生データを適切に匿名加工処理し、教育リソースとして公開する