

2023年度(令和5年度)
医療リアルワールドデータ活用人材育成事業
シラバス

最終更新:2023年6月8日

一般履修コース1年次・インテンシブコース 必修科目

科目名	医療データ管理概論		
日時	2023年5月6日(土曜)、2023年5月13日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	大江和彦(東京大学大学院医学系研究科社会医学専攻医療情報学分野 教授) 土井俊祐(千葉大学医学部附属病院 病院長企画室 特任講師) 岡田昌史(東京大学医学部附属病院企画情報運営部 特任講師)		
授業の目的と概要	【講義の目的】 医療データを取り扱うにあたり、クラウド管理などに関して学習する。さらに技術的な面で情報セキュリティ及び情報の完全性がどのように守られているか、不正侵入対策・暗号化技術も学習する。また医療データにおいては変更・記載履歴も重要であり、利用者認証・改ざん防止の技術についても学習する。後半では、異なるデータ資源の統合化について概論・各種標準規格を学習し、医療施設の部門内・部門間データベースについて実例を学ぶ。さらに、SQL言語を用いたリレーショナル・データベース操作について意義と方法を学び、ケーススタディを通じてこれからの実践につなげてゆく。 学習目標: 1.クラウド管理について、その基本特性、汎用化されているモデルについて説明できる。 2.医療データ管理における情報セキュリティ技術について、求められる要件、想定されるリスク要因と対策、具体的な技術について説明できる。 3.暗号化技術について、共通鍵暗号・公開鍵暗号とその基礎技術について説明できる。 4.異なるデータ資源の統合化の必要性と代表的な標準規格、医療機関内の部門内・部門間データベースの概要について説明できる。 5.小規模リレーショナル・データベースに対し、SQL言語を用いて初歩的な操作を行える。 【概要】 5月6日10:00-11:30 医療データ管理総論 担当:大江和彦 クラウド管理とオンプレミス管理の対比、クラウド管理の基本特性と各種商用クラウドの実例、クラウドでデータ管理を安全に行うために留意すべき点を学習する。 5月6日12:30-14:00 異なるデータ資源の統合化(1) 担当:大江和彦 大規模医療リアルワールドデータ活用にあたり、データ資源の統合化の必要性の総論を学習する。病院システムの構成について学習し、用いられる病院内データの概論を学習する。データの共通規格としてHL7, DICOM, SS-MIX2を取り上げる。 5月6日14:15-15:45 データ・セキュリティ技術・暗号化 担当:土井俊祐 医療情報管理の必要要件(機密性・完全性・可用性)を医療情報システムの安全管理に関するガイドラインに即して学習する。その基本となるネットワークアドレス変換技術、不正侵入・攻撃への対策となるファイアウォールとフィルタリング技術、利用者認証技術について学習する。暗号化技術の意義と具体的技術、例えば共通鍵暗号、公開鍵暗号(秘密鍵と公開鍵)の実例を学習する。改ざん防止・電子署名・認証の仕組み、ハッシュ値(一方向性関数)の特徴と使われ方を学習する。 5月6日16:00-17:30医療データ管理についてのケーススタディ① 担当:土井俊祐 医療リアルワールドデータを活用するにあたり、データサイエンティストとして望ましい対応方法について、セキュリティ技術や暗号化の観点からデータ管理に関する実践問題に挑戦する。設問は受講者間で議論し、グループごとに論点を整理する。 5月13日10:00-11:30異なるデータ資源の統合化(2) 担当:大江和彦		

	標準の統一に関わる組織として、ISO(国際標準化機構)、JIS(日本工業規格)を取り上げる。医療記録の標準コードとして、病名:ICD-10、MEDIS-DCが提供する病名マスター(ICD-10に対応)、薬品のコードとして医薬品HOTコードマスター、臨床検査項目の標準規格JLAC10などについて学習する。 5月13日12:30-14:00異なるデータ資源の統合化(3) 担当:岡田昌史 医療における多様な情報源から共通データモデルへデータを集約することを目的としたObservational Health Data Science and Informatics (OHDSI) コミュニティが管理するObservational Medical Outcomes Partnership (OMOP) 共通データモデルおよびその周辺の話題について学ぶ。 5月13日14:15-15:45リレーショナル・データベースとSQL言語 担当:土井俊祐 リレーショナル・データベースやSQL言語の基礎知識を学び、データ参照などに必要な基本的なコーディングを自身でできるように学習する。また、SQLクライアントソフトウェアの利用に関する画面デモ(CUI、Microsoft Accessなど)を通して、リレーショナル・データベース操作の意義と方法を学ぶ。 5月13日16:00-17:30医療データ管理についてのケーススタディ② 担当:土井俊祐 病院システムにおけるリレーショナル・データベースの活用について、自身が中規模病院であるデータを収集して利用する場合、または依頼を受けてデータを提供する場合に、どの部門にどのような規定をすべきか、収集する際に注意すべきことは何か、などを受講者間で議論する。
授業の方法	講義毎に異なる。適宜指示する。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 インテンシブコース: 5月6日10:00-11:30 医療データ管理総論 5月6日12:30-14:00 異なるデータ資源の統合化(1) 5月6日14:15-15:45 データ・セキュリティ技術・暗号化 5月6日16:00-17:30 医療データ管理についてのケーススタディ① 5月13日10:00-11:30 異なるデータ資源の統合化(2) *インテンシブコースは、講義科目である「倫理法制度概論」、「医療データ管理概論」、「医療リアルワールドデータ構築概論」の3つの科目の中から、指定された講義を含めて合計16コマ(1コマ90分)以上の受講が履修証明プログラム修了のために必要となるため注意すること。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	医療リアルワールドデータ構築概論		
日時	2023年5月20日(土曜)、2023年5月27日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	康永秀生(東京大学大学院医学系研究科公共健康医学専攻臨床疫学・経済学 教授) 岡田昌史(東京大学医学部附属病院企画情報運営部 特任講師) 松居宏樹(東京大学大学院医学系研究科公共健康医学専攻疫学保健学講座 助教) 平松達雄(国際医療福祉大学 未来研究支援センター 教授) 美代賢吾(国立国際医療研究センター 医療情報基盤センター センター長) 関根道和(富山大学学術研究部医学系疫学健康政策学講座 教授) 笹渕裕介(東京大学大学院医学系研究科 リアルワールドエビデンス講座 特任准教授) 佐藤大介(藤田医科大学大学院 医学研究科 病院経営学・管理学専攻 教授)		
授業の目的と概要	【講義の目的】 医療リアルワールドデータを課題解決のために活用することを目指し、現存する大規模医療データベースの構築背景、構造、利活用事例を学習する。個々のデータベースの事例から、医療データベースの構造的概観を把握し、新たな医療課題に対峙した際に、現存するデータベースをどのように応用するか、必要となればどのようなデータベースの構築が必要かを考え、具体的に実践できる知識を身につけることが本講義のねらいである。 学習目標: 1. 大規模医療リアルワールドデータにどのような種類のものが存在するのかについて説明できる。 2. 大規模医療リアルワールドデータがどのように利活用されているのかについて説明できる。 3. 特定の医療課題に対してどのデータベースを利用すれば解決に至ることができるか、どのようなデータベースを構築すれば良いかについて実践的な思考ができる。 【概要】 5月20日10:00-11:30大規模医療リアルワールドデータベース総論 担当:康永秀生 大規模医療データの種類について学習するとともに、医療機関ベース、保険者ベースなどのデータベースのそれぞれの特徴について学ぶ。また、データベース構築を行う際の目的設定やデータ標準化などの必要となる課題について概観を学ぶ。 5月20日12:30-14:00大規模医療リアルワールドデータベース各論1 担当:岡田昌史 電子レセプトデータは診療報酬請求や審査支払業務を目的として設計されており、データ分析に用いる際には留意すべき点多々存在する。電子レセプトデータの基本構造について記録様式、各レコードの詳細など、データ分析を行うにあたって必要となる知識および注意点について学ぶ。 5月20日14:15-15:45大規模医療リアルワールドデータベース各論2 担当:松居宏樹 我が国における代表的かつ重要な大規模医療リアルワールドデータベースとして、DPCデータベースの成り立ち、構造、利活用事例、また、レセプト情報・特定健診等情報データベース(NDB)の成り立ち、構造、利活用事例について学ぶ。 5月20日16:00-17:30医療課題に対するデータ利活用のケーススタディ① 担当:岡田昌史 レセプトデータを用いて実際にデータベース研究を実施することを想定して、研究対象集団を具体的に疾患コード等を用いて定義する演習を行う。 5月27日10:00-11:30大規模医療リアルワールドデータベース各論3 担当:平松達雄・美代賢吾 国内の大規模医療リアルワールドデータベースの具体例として、MID-NETおよびJ-DREAMSの成り立ち、構造、期待される利活用方法、また、各種レジストリデータについてその成り立ち、構造、利活用事例について学ぶ。		

	5月27日12:30-14:00大規模医療リアルワールドデータベース各論4 担当:関根道和 地域・行政のビッグデータについて学び、国保の特定健診データを用いた解析・データ可視化の試みについて実例を学ぶ。 5月27日14:15-15:45医療課題に対するデータベース利活用のケーススタディ② 担当:笹渕裕介 特定の医療課題について、リアルワールドデータを使ってどのように課題解決に結びつけるかを、具体的な事例を交えて理解を深める。 5月27日16:00-17:30大規模医療リアルワールドデータベース各論5 担当:佐藤大介 地域の医療体制に関する現状からリアルワールドデータ基盤の重要性について学び、急性期病院が有するDPCデータ等を活用した医療圏別・診療領域別の医療需要や、将来推計の実例を学ぶ。さらにデータ分析を活用した地域の医療体制の在り方に関する合意形成について、ディスカッション形式で理解を深める。
授業の方法	講義毎に異なる。適宜指示する。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 インテンシブコース: 5月20日10:00-11:30大規模医療リアルワールドデータベース総論 5月20日12:30-14:00大規模医療リアルワールドデータベース各論1 5月20日14:15-15:45大規模医療リアルワールドデータベース各論2 5月20日16:00-17:30個人情報に関するケーススタディ① 5月27日16:00-17:30大規模医療リアルワールドデータベース各論5 *インテンシブコースは、講義科目である「倫理法制度概論」、「医療データ管理概論」、「医療リアルワールドデータ構築概論」の3つの科目の中から、指定された講義を含めて合計16コマ(1コマ90分)以上の受講が履修証明プログラム修了のために必要となるため注意すること。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	倫理法制度概論		
日時	2023年7月22日(土曜)、2023年9月2日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	瀬尾雅子(東京大学医学部附属病院 弁護士) 板倉陽一郎(ひかり総合法律事務所 弁護士) 河添悦昌(東京大学大学院医学系研究科 医療AI開発学講座 特任准教授) 上竹勇三郎(東京大学大学院医学系研究科研究倫理支援室 准教授) 山本隆一(医療情報システム開発センター理事長) 川口英明(慶應義塾大学医学部 微生物学・免疫学教室 / ヒト生物学-微生物叢-量子計算研究センター(Bio2Q) 特任准教授 / 東京大学大学院医療情報学分野 客員研究員) 平松達雄(国際医療福祉大学 未来研究支援センター 教授) 香川璃奈(筑波大学医学医療系 臨床医学域 医療情報マネジメント学 講師)		
授業の目的と概要	【講義の目的】 医療関連情報を扱うに当たり、新しい知見の獲得という利点のみならず、情報漏えいをもたらす不利益にも配慮し、プライバシーを保護する必要性と法的整備について学習する。特に、昨今の社会的背景や個人情報保護法の要点を踏まえつつ、医療情報の利活用を促進するために制定された次世代医療基盤法の法的特徴について目を向けることが本講義のねらいである。 学習目標: 1. 医療関連情報を中心に医療に関連する法的整備について説明できる。 2. 個人情報保護法の概要と社会的背景を説明できる。 3. 次世代医療基盤法の概要と、匿名化を行う機関の運用について説明できる。 4. 人を対象とする医学系研究に関する倫理指針について説明できる。 【概要】 7月22日10:00-11:30医療と法 担当:瀬尾雅子 医療関連情報に関連した、診療録の法的位置づけや保存期間について学習するとともに、医師法・医療法・療養担当規則など、診療行為や病院・診療所の運営に関連した法的整備を学習し、医療を支える法的基盤について幅広く学習する。 7月22日12:30-14:00個人情報とプライバシー 担当:板倉陽一郎 プライバシーの権利や個人情報保護について、憲法上の根拠、法体系上の位置付けなどについて概説する。特にプライバシー侵害について、民事裁判例にあらわれた事例を交えながら、医療関連情報が不適当に扱われることで被る不利益について学習する。 7月22日14:15-15:45個人情報保護法概論 担当:板倉陽一郎 個人情報の取扱いに関する規制法の基盤となる個人情報保護法について、個人識別符号や要配慮個人情報といった、医療関連情報の取扱いについての規律を理解する上で重要となるキーワードを中心に学習する。また、一般データ保護規則(GDPR)の域外適用や米国の法制度などの海外の状況についても学習する。 7月22日16:00-17:30個人情報に関するケーススタディ① 担当:河添悦昌 データセット作成や提供、公開に関連した個人情報に関わる問題について、事例を交えて検討する。 9月2日10:00-11:30研究倫理指針概論 担当:上竹勇三郎 人を対象とする生命科学・医学系研究に関する倫理指針について、次世代医療基盤法との関わりを中心に学習する。また、医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律(薬機法)や臨床研究法等についても学習する。 9月2日12:30-14:00次世代医療基盤法とその周辺領域について 担当:山本隆一		

	次世代医療基盤法について、通知によるオプトアウト、匿名加工医療情報、第三者提供など、関連するキーワードを中心に、その法的体制について学習する。次世代医療基盤法で定められた認定匿名加工医療情報作成事業者について、事例を紹介しながら、その運用について説明する。また、医療データの取り扱いに関して次世代医療基盤法の周辺領域について学習する。 9月2日14:15-15:45公的ビッグデータの利用手続きと課題 担当：川口英明、平松達雄 NDB、MID-NET、制限公開ゲノムデータ、統計法に基づくデータなど、各種公的データの利活用について、データ利用にたどり着くための具体的な流れや注意点を、事例を交えながら学習する。 9月2日16:00-17:30個人情報に関するケーススタディ② 担当：香川璃奈 個人情報の利用について、事例を交えて議論を行う。
授業の方法	講義毎に異なる。適宜指示する。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース：全講義 インテンシブコース： 7月22日10:00-11:30医療と法 7月22日12:30-14:00個人情報とプライバシー 7月22日14:15-15:45個人情報保護法概論 7月22日16:00-17:30個人情報に関するケーススタディ① 9月2日12:30-14:00次世代医療基盤法とその周辺領域について *インテンシブコースは、講義科目である「倫理法制度概論」、「医療データ管理概論」、「医療リアルワールドデータ構築概論」の3つの科目の中から、指定された講義を含めて合計16コマ(1コマ90分)以上の受講が履修証明プログラム修了のために必要となるため注意すること。
教科書・参考書等	講義毎に異なる。下記に記載のない講義については適宜指示する。 7月22日10:00-11:30医療と法 担当：瀬尾雅子(特に予習や購入を指示するものではなく、より深い理解のための参考図書として提示するものである。) ①米村滋人『医事法講義』(日本評論社, 2016年) ②手嶋豊『医事法入門(第5版)』(有斐閣, 2018年) ③望月稔之, 並木洋, 小笠原一志『最新 医療関連法の基礎知識2022年版～これだけは知っておきたい医療実務101法』(医学通信社, 2022年) ④島崎謙治『日本の医療 制度と政策(増補改定版)』(東京大学出版会, 2020年) 7月22日12:30-14:00個人情報とプライバシー 担当：板倉陽一郎 参考文献： 小向太郎『情報法入門【第6版】』(NTT出版, 2022年) 曾我部真裕・林秀弥・栗田昌裕『情報法概説 第2版』(弘文堂, 2019年) 松尾剛行『最新判例にみるインターネット上のプライバシー・個人情報保護の理論と実務』(勁草書房, 2017年) 7月22日14:15-15:45個人情報保護法概論 担当：板倉陽一郎 参考文献： 日本の個人情報保護法について, 園部逸夫・藤原静雄編, 個人情報保護法制研究会著『個人情報保護法の解説<<第三次改訂版>>』(ぎょうせい, 2022年)(※2021年改正未対応) 田中浩之・薦大輔編著『60分でわかる! 改正個人情報保護法超入門』(技術評論社, 2022年) 宇賀克也『新・個人情報保護法の逐条解説』(有斐閣, 2021年) 一般データ保護規則(GDPR)について, 石井夏生利『EUデータ保護法』(勁草書房, 2020年) 米国の法制度(連邦法)について,

	クリス・フーフナグル著, 宮下紘・板倉陽一郎・河井理穂子・國見真理子・成原慧・前田恵美訳『アメリカプライバシー法』(勁草書房, 2018年) #いずれも, 辞書的に調べたい場合に備えるべき書籍であり, 全て購入する必要はないが, 実務での利用にも耐えうるものである。なお, 個人情報保護法は2020年, 2021年に改正されており, 2022年4月1日(2020年改正, 2021年改正のうち自治体部分以外)及び2023年4月1日(2021年改正のうち自治体部分)に施行された。書籍等を参照する際には両改正への対応状況に留意されたい。 9月2日12:30-14:00次世代医療基盤法とその周辺領域について 担当:山本隆一 (特に予習や購入を指示するものではなく、もし深く知りたい場合の参考図書としての提示) 参考文献: 宇賀克也『次世代医療基盤法の逐条解説』(有斐閣, 2019年)
--	---

科目名	データ収集・コード化・標準化・クレンジング実践		
日時	2023年6月24日(土曜)、2023年7月8日(土曜) 2023年7月15日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	大江和彦(東京大学大学院医学系研究科社会医学専攻医療情報学分野 教授) 岡田昌史(東京大学医学部附属病院企画情報運営部 特任講師)		
授業の目的と概要	【講義の目的】 医療データを用いて課題解決を目指す際には、データ解析手法の重要性もさることながら、解析に資するデータをいかに高品質なレベルで準備できるか、適切に目的に沿って利用可能なデータを準備できるかにかかっている。データ収集の段階からデータクレンジングの段階までの間で、どのようにその一連の処理を行い、高い品質のデータを得るかについて実践を行い、必要な知識、技術を身につけることが本講義・実習の狙いである。 学習目標: 1. 適切に解析に資するデータを作成するために、目的に沿ってデータ収集を計画することができる。 2. 電子診療録データから意図したデータを抽出することができる。 3. 目的に沿って適切にデータをコード化することができる。 4. 医療データにおける標準化コードの重要性を理解し、実データに対して標準化コードへの変換、評価を行うことができる。 5. 適切にデータクレンジングを計画し、実践後にデータに対して評価を行うことができる。 【概要】 6月24日10:00-11:30 データ分析におけるデータ概観の把握I 担当:岡田昌史 実際に大規模に収集された電子カルテ由来のデータベースを用いて、まず大きなデータのすべての項目を閲覧するために必要となる技術を学ぶ。 6月24日12:30-14:00 データ分析におけるデータ概観の把握II 担当:岡田昌史 データに付随して提供される仕様書と実際のデータから、使用されているコード体系を確認する。 6月24日14:15-15:45 データクレンジング実践I 担当:岡田昌史 具体的な研究課題を想定して、研究対象集団をデータベース上の情報を用いて定義する方法を検討する演習を行う。 6月24日16:00-17:30 データのクレンジング実践II 担当:岡田昌史 実際のデータベース上で研究対象集団を絞り込むために使われる変数の分布を確認して、検討した集団定義が機能するかを検討する。 7月8日10:00-11:30 データクレンジング実践III 担当:岡田昌史 研究のアウトカムとなる変数について、データベースから抽出するための定義を作成する。そして実際のデータベース上で関係する変数の分布を確認して、定義が機能するか検討する。 7月8日12:30-14:00 データクレンジング実践IV 担当:岡田昌史 実際のデータベース上で、これまでに作成・検討した研究対象集団の定義を適用して、研究対象者の絞り込みフローチャート(Patient Disposition)を作成する。 7月15日10:00-11:30 データのコード化 担当:大江和彦 医療データ解析において、入力されて保有されている時刻などの情報は時系列解析などにおいて重要となる一方で、期間などの量的な値に変換しないとそのままでは利用できないケースが多々存在する。入力されている内容の概観を適切に把握し、目的に応じて適切に日付や時刻データのコード		

	化を行う。さらに解析可能な形式への変換を行い、コード化を終えたデータの概観を可視化して正しくコード化されたかを評価する方法を学ぶ。 7月15日12:30-14:00 データの標準化 担当:大江和彦 複数施設のデータを統合し、解析するための重要な段階として医療データにおけるローカルコードを必要に応じて標準化コードへ変換する手法を学び、変換後に変換が適切であるかどうかを評価する手法を学ぶ。
授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 インテンシブコース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	データ再編成と匿名化実践		
日時	2023年9月9日(土曜)、2023年9月16日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	内村祐之(株式会社メディクト MEDICT, Inc.) 大坪徹也(東京大学医学部附属病院 国立大学病院データベースセンター 研究員) 平松達雄(国際医療福祉大学 未来研究支援センター 教授) 関倫久(東京大学医学部附属病院企画情報運営部 助教)		
授業の目的と概要	【講義の目的】 医療データベースの構築のためには必要なテーブルを定義し、電子診療録など二次利用を想定しないデータ配置のために直接の解析が困難となっているデータを、解析が容易になるような形式で再編成することが必要となる。また、個人情報を含む医療データを二次利用する場合において、個人情報保護とデータ有効活用の二つの命題を同時に満たすためには適切な手法で匿名化を行う必要がある。そのため、本講義及び実習では医療データベース構築のための適切なデータ再編成、データの匿名化について学び、実践的知識及び技術を身につけることを狙いとする。 学習目標: 1. リレーショナルデータベースの操作ができる。 2. 目的に沿ったテーブルを定義し、データを再編成することができる。 3. 匿名化の意義、手法を理解し、実践することができる。 【概要】 9月9日10:00-11:30リレーショナルデータベースの操作I 担当:内村祐之 大規模な医療データを操作するにあたり、正規化したデータ群をテーブル形式で保有するリレーショナルデータベースの理解、操作が必須である。リレーショナルデータベースの基本的な演算処理についてその概要を理解し、実際に操作する技術を身につける。 9月9日12:30-14:00リレーショナルデータベースの操作II 担当:内村祐之 大規模な医療データベースの構築・維持にあたり、リレーショナルデータベースを操作し、新たなデータベースやテーブルを定義するデータ定義や権限やトランザクションを扱うデータ制御が必要になる。これらに関して知識を身につけるとともに、実際に操作する技術を身につける。 9月9日14:15-15:45データ再編成I 担当:大坪徹也 二次利用を想定せずに蓄積されたデータベースから、解析のためにデータを再編成するにあたって、その一連の工程を学ぶ。まず最終的な目的を設定し、それに沿って適切にテーブルを定義し、データを再編成する工程を実践によって学ぶ。 9月9日16:00-17:30データ再編成II 担当:大坪徹也 特定の課題に対して再編成したデータを用い、その再編成工程が適切か、目的に応じた利用が可能かを実際に確認する。データテーブル定義が目的に沿っていない仕様となっていないかを検証し、必要があれば再編成を再度行い、最終的に解析利用に資するデータベースを構築する。 9月16日10:00-11:30匿名化実践I 担当:平松達雄 医療データにおいて匿名化は患者保護及びデータの利活用範囲の拡大に非常に重要な工程である。その意義及び具体的な手法(ソート、シャッフル、ハッシュ関数によるID置換、項目削除・レコード削除・セル削除、トップコーディングとボトムコーディング)について学び、実践を行う。 9月16日12:30-14:00匿名化実践II 担当:平松達雄 医療データベースの適切な利活用のために、匿名化の具体的な手法(一般化、スワッピング、マイクロアグリゲーション、ラウンディング、リサンプリング、ノイズ付加、擬似データ生成)について学び、実践を行う。		

	9月16日14:15-15:45匿名化実践III 担当: 関倫久 匿名化は再特定を100%防ぐものではなく、そのリスクを一般的に不可能と評価されるレベルに下げに過ぎない。その実際について、匿名化されたデータのk-匿名性の評価、参照リスクの評価に関して実践を行う。 9月16日16:00-17:30匿名化実践IV 担当:関倫久 匿名化における対応表の作成、匿名化したデータの連結について実践を行う。また、匿名化・リスク評価ツールを用いて匿名化の安全性の評価について実践を行う。
授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 インテンシブコース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	データの記述と図示		
日時	2023年9月30日(土曜)、2023年10月7日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	大原信(筑波大学医学医療系 教授 附属病院医療情報経営戦略部長) 香川璃奈(筑波大学医学医療系 臨床医学域 医療情報マネジメント学 講師) 讃岐勝(筑波大学医学医療系臨床医学域(医学数理情報学) 助教) 大江和彦(東京大学大学院医学系研究科社会医学専攻医療情報学分野 教授) 岡田昌史(東京大学医学部附属病院企画情報運営部 特任講師) 関倫久(東京大学医学部附属病院企画情報運営部 助教)		
授業の目的と概要	【講義の目的】 大規模医療情報データベースを用いた医学研究の報告においてよく利用される、データの記述統計や図示の方法を学び、教材データを自ら操作して図表の作成を行えるようになる。 学習目標: 1. データベース研究の報告においてよく利用される図表について実際の論文から学ぶ。 2. 図表作成のために利用されるソフトウェアパッケージの使用方法を知る。 3. 教材データを用いて記述統計表の作成、情報の可視化を自らの操作により行う。 【概要】 9月30日10:00-11:30 リアルワールドデータ研究の報告事例 担当:岡田昌史 リアルワールドデータからは、エビデンス創出を目的とした後ろ向きコホースタディを実施できるだけではなく、レセプトからは医療費の情報が得られ、投薬・注射オーダーからは治療パターンが読み取れる。これらの情報を記述的に集計して考察を加えた学術研究の実例を紹介する。 9月30日12:30-14:00 リアルワールドデータ研究でよく報告される図表 担当:岡田昌史 引き続き実際のリアルワールドデータを用いた研究事例から、図表の解釈と、図表を作成するために必要となる情報をデータベースから得る方法について説明する。 9月30日14:15-15:45 図表作成の具体的手法(1) 担当: 讃岐勝、香川璃奈 教材データを用いて研究報告でよく使われる表を作成する方法について具体的に演習形式で学ぶ。 9月30日16:00-17:30 図表作成の具体的手法(2) 担当: 讃岐勝、香川璃奈 教材データを用いて研究報告でよく使われる図を作成する方法について具体的に演習形式で学ぶ。 10月7日10:00-11:30 治療パターン記述の実践1 担当:岡田昌史、関倫久 グループワーク形式で、教材データを用いて治療パターンの記述を行う計画を立案する。 10月7日12:30-14:00治療パターン記述の実践2 担当: 岡田昌史、関倫久 グループワーク形式で、教材データを用いて治療パターンの記述を行うためのデータ抽出を実施する。 10月7日14:15-15:45治療パターン記述の実践3 担当: 岡田昌史、関倫久 グループワーク形式で、教材データを用いて治療パターンを記述した図表を作成する。 10月7日16:00-17:30治療パターン記述の実践4 担当: 岡田昌史、関倫久 グループワーク形式で、教材データを用いて作成した図表に対して考察を加え、発表する。		

授業の方法	演習部分、グループワーク部分については、実習用サーバにログインし、データ処理について実際に手を動かして学ぶ。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 インテンシブコース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	統計的データ解析・マイニング実践1		
日時	2023年10月28日(土曜)、2023年11月4日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	松山裕(東京大学大学院医学系研究科 公共健康医学専攻 教授) 大庭幸治(東京大学大学院情報学環 准教授) 小川光紀(東京大学大学院情報学環 特任講師) 上村鋼平(東京大学大学院情報学環 特任講師) 野村尚吾(東京大学大学院医学系研究科 生物統計情報学講座 特任講師)		
授業の目的と概要	【講義の目的】 医療データはそのままでは有益な知見につながることはできず、統計的に処理することで意味のある情報に変換することが肝要である。本講義のねらいは、医療データを医療情報に変換するための各種統計学的手法を基礎から学習し、自身の手で実際の解析を行えるようになることである。特に、医学統計や疫学研究で用いられる主な解析手法を中心に扱う。講義の前半では手法についての座学を行い、後半では実際に手を動かしてRによるプログラミングを行う。 学習目標: 1. 記述統計を用い、医療データを適切に集約できる。 2. 仮説検定の基礎的な考え方を学び、適切に検定を扱うことができる。 3. 一般化線形回帰モデルを扱い分析することができる。 4. 傾向スコアや欠損値補完など、観察研究で用いられる手法を適切に用いることができる。 【概要】 10月28日10:00-11:30記述統計 担当:小川光紀 平均・分散・標準偏差など、基礎的な統計量を学習し、データに合わせた適切な使用方法と、プレゼンテーション方法について学ぶ。また、名義尺度や順序尺度など、各種尺度についても学習し、適切な集約値の作成方法を学習する。 10月28日12:30-14:00確率分布 担当:小川光紀 正規分布を中心に、確率分布の概念やその使用方法について学習する。正規分布の他には、医学統計で頻繁に用いられる二項分布やポアソン分布についても学習し、プログラミングで確率分布を作成しながらその特性について正しく理解する。 10月28日14:15-15:45仮説検定① 担当:上村鋼平 帰無仮説や対立仮説、検定統計量などのキーワードを中心に、仮説検定を適切に用いるための考え方を学習する。代表的な仮説検定であるt検定を実行し、出力結果を正しく説明できるようになるのが目標である。 10月28日16:00-17:30仮説検定② 担当:上村鋼平 独立性の検定に用いられるカイ二乗検定や、3群以上からなるデータの母平均の差を検定する分散分析について学習する。また、実データに対して実行した各検定の出力結果を正しく説明できるようになるのが目標である。 11月4日10:00-11:30線形回帰モデル 担当:大庭幸治 統計モデルの中で最も代表的な線形回帰モデルの概要と、最小二乗法を用いたパラメータ推定の方法、使用に必要な注意点について学習する。また、実際にプログラミングを行い、出力結果を正しくプレゼンテーションする方法を学習する。 11月4日12:30-14:00一般化線形回帰モデル 担当:大庭幸治		

	医学統計において使用頻度の高い、二値の目的変数を対象とするロジスティック回帰モデルや、正の整数を対象とするポアソン回帰モデルの概要と、最尤推定を用いたパラメータ推定の方法を学ぶ。また、プログラミングを行い、出力結果を正しくプレゼンテーションする方法を学習する。 11月4日14:15-15:45傾向スコア 担当:野村尚吾 介入のできない観察研究に対し背景因子を調整する目的でしばしば用いられる、傾向スコア分析について学習する。特に、代表的な解析手法であるマッチングとInverse Probability Weighting (IPW)について、プログラミングを通じて適切な使用方法を学習する。 11月4日16:00-17:30欠損値の取り扱い 担当:上村鋼平 リアルワールドデータでは必ずと言ってよいほど存在する欠損値に関して、Missing at random (MAR)の考え方を中心に学習する。また、欠損値補完としては代表的な手法である、多重代入法について、プログラミングを通じて学習する。
授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 インテンシブコース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

一般履修コース2年次 必修科目

科目名	臨床課題分析と実務マネジメント		
日時	2023年4月1日(土曜)、2023年4月8日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	笹渕裕介(東京大学大学院医学系研究科 リアルワールドエビデンス講座 特任准教授) 山口聡子(東京大学大学院医学系研究科糖尿病・生活習慣病予防社会連携講座 特任准教授) 城大祐(東京大学大学院医学系研究科ヘルスサービスリサーチ講座 特任准教授) 松居宏樹(東京大学大学院医学系研究科公共健康医学専攻疫学保健学講座 助教) 康永秀生(東京大学大学院医学系研究科公共健康医学専攻臨床疫学・経済学 教授)		
授業の目的と概要	【講義の目的】 現場で生じる臨床課題を漠然と抱えているだけでは課題解決には不十分であり、分析のために検証可能な「問い」を立てる必要がある。本講義のねらいは、現場で生じる臨床課題を検討に適した形式に落とし込み、真に実務に結び付けるための具体的な方策を学習することである。実習形式でclinical questionをresearch questionに変換し、そのresearch questionの解決に向けたチーム編成・データ利用申請・倫理申請などの実務のシミュレーション・ケーススタディを行う。Research questionの設計と、チーム編成の意図、提案した課題解決手法、といった一連の流れに関する、受講生によるプレゼンテーションも適宜併用する。 学習目標: 1.clinical questionを分析可能なresearch questionに変換することができる。 2.research questionの分析に必要なチームを編成し、研究計画書作成・倫理審査などの実務手続きを進めることができる。 3.分析結果を学問的成果および実務改善に向けた方策に結び付け適切にプレゼンテーションができる 【概要】 4月1日10:00-11:30研究立案1 担当:笹渕裕介 臨床業務で想定されるclinical questionの抽出を行う。可能な限り体系的にclinical questionを抽出するのが好ましい。臨床業務に従事していない履修生については、医療・介護領域の一般的な問題点から派生したquestionの抽出でもよい。抽出したclinical questionを整理し、分類を行う。その際、より具体的な医療面のエンドポイント(有効性・安全性評価に使われる評価項目)を想定し、一定の基準を基にした切り口で分類することが好ましい。整理したclinical questionに対応するresearch questionに変換を行う。その際、PE(I)COといったフレームワークを用い、構造的に問いを分析する体系的なアプローチを採用する。 4月1日12:30-14:00研究立案2 担当:笹渕裕介 作成したresearch questionに関する文献精査を行い、その結果をフィードバックすることでresearch questionの精緻化を行う必要性について知る。立てたresearch questionが“FINER”かどうかのチェックを行う。 作成したresearch questionに対して、編成したチームではどのような解決手法が考えられるか検討を行う。その際、実務へのフィードバックを意識して具体的に解決手法を考案するのが好ましい。 4月1日14:15-15:45研究立案3 担当:笹渕裕介 Research questionの解決のために必要なチーム編成の検討を行う。その際、単施設か多施設か、利用可能なデータベース、同意取得の方法、研究計画の公表など課題解決に必要な舞台や発表方法、論文化する場合のオーナーシップについての検討を行う。研究デザイン時点での生物統計家へのコンサルテーションの必要性についても学習する。		

	4月1日16:00-17:30研究立案4 担当:笹渕裕介 倫理申請の研究計画書・同意書・オプトアウト文書例などを提示し、記載上の留意点について検討を行う。多施設共同研究の場合、分担施設でも無理なく承認される可能性が高いプロトコル記載になっているかについても留意が必要である。 4月8日10:00-11:30倫理申請・データ利用申請 担当:山口聡子、城大祐 倫理申請データ駆動型の研究に独特なNDB、DPCデータなどの利用申請の具体的手続き、申請書の作成を行う。DWHへのデータ利用申請例などもについても実践的に学習する。また、研究デザインおよび倫理申請における研究資金の取り扱い・利益相反管理について実践的に学習する。 4月8日12:30-14:00データマネジメント・データ管理 担当:松居宏樹 研究データを扱う上で、思考過程を明示しながら記録を残す具体的な方法、データの逸失や流出を防ぐ方法についておさらいする。論文化に向けたデータ管理の重要性についても学ぶ。 4月8日14:15-15:45統計解析・結果の公表1 担当:康永秀生 統計解析での留意点、および解析結果の公表について留意する点を学ぶ。必要に応じてグループワークでのプレゼンテーションも実習に併用する。 4月8日16:00-17:30統計解析・結果の公表2 担当:康永秀生 RWDを扱う際の論文化等における留意点について学習する。必要に応じてグループワークでのプレゼンテーションも実習に併用する。
授業の方法	講義毎に異なる。適宜指示する。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	自由記載データ自然言語処理実践		
日時	2023年6月3日(土曜)、2023年6月10日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	篠原恵美子(東京大学大学院医学系研究科 医療AI開発学講座 特任助教) 河添悦昌(東京大学大学院医学系研究科 医療AI開発学講座 特任准教授) 岡田昌史(東京大学医学部附属病院企画情報運営部 特任講師)		
授業の目的と概要	【講義の目的】 退院時サマリもしくは症例報告などの未整形の医療テキストを対象とした情報抽出を実践することで自然言語処理技術への理解を深める。 学習目標: 1. 自然言語処理に必要な前処理を行うことができる。 2. 自然言語処理の基本技術である形態素解析と係り受け解析の概要を理解する。 3. 自然言語を対象とする統計学的手法の概要を理解する。 4. 医療テキストを対象とする情報抽出課題を実践する。 【概要】 6月3日10:00-11:30 形態素解析と前処理1 担当:篠原恵美子 日本語テキストデータの解析の基本技術である形態素解析と係り受け解析の考え方と、そのために必要な前処理について解説する。 6月3日12:30-14:00 形態素解析と前処理2 担当:篠原恵美子 前講義で学習した前処理の具体的な実行方法を実習を通して学ぶ。 6月3日14:15-15:45 形態素解析と前処理3 担当:篠原恵美子 形態素解析器としてMeCabを用い、形態素解析の基本的な実行方法を学習する。 6月3日16:00-17:30 形態素解析と前処理4 担当:篠原恵美子 医療テキストでは医学専門用語や略語などが出現することが多い。このようなテキストに形態素解析を適用する際には、独自の形態素解析用辞書を用いることで解析精度を上げることができる。本講義では独自辞書の作成方法を学習する。 6月10日10:00-11:30 医療自然言語処理実践1 担当:河添悦昌 医療自然言語テキストに記載される情報の種類とこれを活用する必要性、ならびに課題について解説する。 6月10日12:30-14:00 医療自然言語処理実践2 担当:河添悦昌 単語の意味を扱うための方法の一つとしてシソーラスを使った考え方を解説する。また、別の方法として、機械学習によってテキストを扱うための基本となる単語の埋め込み表現と、この応用について解説する。 6月10日14:15-15:45 医療自然言語処理実践3 担当:河添悦昌 前講義までに行ったテキストの前処理、形態素解析、単語の埋め込み表現を用いて、設定した臨床的課題を実際のテキストを対象に実践する。 6月10日16:00-17:30 KH coderによるテキストマイニング 担当:岡田昌史 主として社会調査領域で活用されているGPLライセンスの計量テキスト分析ツールであるKH coderを用いて、フリーテキストデータに対して基本的なテキストマイニングを実施する方法を学ぶ。KH coderは事前にインストールしておくことが望ましい: https://khcoder.net/		

授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 インテンシブコース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	症例データベースレジストリ処理実践		
日時	2023年7月29日(土曜)、2023年8月5日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	大江和彦(東京大学大学院医学系研究科社会医学専攻医療情報学分野 教授) 岡田昌史(東京大学医学部附属病院企画情報運営部 特任講師)		
授業の目的と概要	【講義の目的】 症例レジストリのサンプルデータを用いて、設定されたテーマに沿ったデータベースの構築、データ操作を行うことで解析基盤を構築できる能力を身につけるとともに、クリニカルクエッションに沿ってデータ操作を行う能力を身につける。 学習目標: 1.症例レジストリ事業の対象となるテーマの特徴、レジストリ由来データの特徴・メリット/デメリットを知る。 2.症例レジストリデータ構築の手順を学び、具体的な構築構想ができる能力を身につける。 3.症例レジストリデータに対して設定されたクリニカルクエッションに沿った解析に向けたデータ操作を行うことができる。 4.レジストリデータを概観することで存在するデータが含む各種バイアスについて考察し、研究テーマの妥当性について評価できる能力を身につける。 【概要】 7月29日10:00-11:30 症例データベースレジストリ総論 担当:大江和彦 症例データベースレジストリについて、診療領域の特徴、レジストリ由来データの特徴・メリット/デメリットを学ぶ。特定のテーマについて、症例データベースレジストリ構築という手段が適するの可否かについて考察できる能力を身につける。 7月29日12:30-14:00レジストリ構築戦略策定プロセス 担当:大江和彦 症例データベースレジストリの構築工程について、収集情報の選定(リサーチクエッションの収集・選定)、データ収集方法の検討(紙ベース、電子カルテテンプレート、FileMaker、MCDRSなど)、データ固定化前のデータクリーニングなどを含めた具体的手順を学び、設定されたテーマについて症例データベースレジストリ構築の戦略を練ることを実践的に学ぶ。 7月29日14:15-15:45解析に向けたデータクリーニング1 担当:岡田昌史 実際に大規模に収集された電子カルテ由来のデータベースを用いて、具体的な研究課題を想定して、研究対象集団をデータベース上の情報を用いて定義する方法を検討する演習を行い、実際のデータベース上で研究対象集団を絞り込むために使われる変数の分布を確認して、検討した集団定義が機能するかを検討する。 7月29日16:00-17:30解析に向けたデータクリーニング2 担当:岡田昌史 研究の曝露とアウトカムとなる変数について、データベースから抽出するための定義を作成して、実際のデータベース上で関係する変数の分布を確認し、定義が機能するか検討する。 これまでに作成・検討した研究対象集団の定義を適用して、研究対象者の絞り込みフローチャート(Patient Disposition)を作成する。 8月5日10:00-11:30実践的データ加工 担当:岡田昌史 絞り込まれた研究対象者集団について、患者背景を一覧できる表(いわゆるTable 1)を作成する。 8月5日12:30-14:00データ概観評価 担当:岡田昌史 データベース内のデータから、曝露を示す変数、およびアウトカムとなる変数を作成して、その分布を把握する。		

	8月5日14:15-15:45実践的データ解析1 担当:岡田昌史 曝露とアウトカムとの関係について、層別解析や多変量解析を実施する。 8月5日16:00-17:30実践的データ解析2 担当:岡田昌史 これまでに得られた解析結果をもとに、得られた結果の内的、外的妥当性について議論を行い、結論をまとめる。
授業の方法	講義毎に異なる。適宜指示する。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

科目名	統計的データ解析・マイニング実践2		
日時	2023年12月2日(土曜)、2023年12月9日(土曜)	単位数	1
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	竹内由則(横浜市立大学 学術院(国際総合科学群) データサイエンス学部 データサイエンス学科 准教授) 野村尚吾(東京大学大学院医学系研究科 生物統計情報学講座 特任講師) 小川光紀(東京大学大学院情報学環 特任講師) 関倫久(東京大学医学部附属病院企画情報運営部 助教) 河添悦昌(東京大学大学院医学系研究科 医療AI開発学講座 特任准教授)		
授業の目的と概要	【講義の目的】 本講義のねらいは、多種多様なリアルワールドデータに対して適切なデザインを設計した上で、データに適したデータサイエンス手法を適切に選択し、実践できるようになることである。「統計的データ解析・マイニング実践1」で学習した内容以上に実践的で高度な解析ができるようになることを目指す。 学習目標: 1. リアルワールドデータ解析のためのデザインを適切に設計し、実践できる 2. 多重性の概念を適切に理解し、対処法を実践できる。 3. 機械学習の特性を理解し、代表的な手法を実践することができる。 【概要】 12月2日10:00-11:30リアルワールドデータ解析のためのデザイン設計 担当:竹内由則 リアルワールドデータを用いた研究において大きな役割を果たす統計的因果推論の基礎概念について学習する。さらに、研究デザインあるいは統計解析による交絡調整法について学習する。 12月2日12:30-14:00ケーススタディ1 担当:竹内由則 リアルワールドデータを用いた研究における、統計的因果推論の概念に基づく研究デザインおよび統計解析について、事例を交えて検討を行う。 12月2日14:15-15:45 多重性の概念とその対処法 担当:野村尚吾 リアルワールドデータを用いた統計解析においては、多数の検定が同時に発生するケースが多く、通常通りの有意水準で検定を行った場合、偽陽性が現れる確率が高くなるという多重性の問題が発生する。本講義では、FWERやFDRといった偽陽性指標を通じて、多重性の概念とその対処法について学習する。 12月2日16:00-17:30ケーススタディ2 担当:野村尚吾 多重性の概念とその対処法について、事例を交えて検討を行う。 12月9日10:00-11:30 予測モデル入門 担当:小川光紀 予測モデル構築に必要な基礎を学ぶ。より具体的には、教師あり学習や教師なし学習、over-fitting、などを中心に学ぶ。 12月9日12:30-14:00ランダムフォレスト・ブースティングモデル入門 担当:関倫久 決定木をベースにしたアンサンブル学習であるランダムフォレストとブースティングモデルの概要を学ぶ。Pythonを用いてプログラミングをし、予測結果について説明ができるようになることを目指す。 12月9日14:15-15:45ニューラルネットワーク入門 担当:河添悦昌 近年の機械学習的手法で最も注目を集めているディープラーニングの基礎となる技術である、ニューラルネットワークの概要を学ぶ。Pythonを用いてプログラミングをし、予測結果について説明ができるようになることを目指す。		

	12月9日16:00-17:30機械学習演習 担当:関倫久、河添悦昌 ランダムフォレスト、ブースティングモデル、ディープラーニングの各種手法をリアルワールドデータに対して実装し、その予測精度を比較することで、各種手法の特性について実習を通じて学習する。
授業の方法	講義毎に異なる。適宜指示する。
履修上の注意事項	履修証明プログラム修了のために以下の講義は必修となり、出席が必須である。 一般履修コース:全講義 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。
教科書・参考書等	講義毎に異なる。適宜指示する。

選択科目

科目名	UNIXコマンド入門		
日時	2023年6月17日(土曜日10:00-14:00)	単位数	なし
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	関倫久(東京大学医学部附属病院企画情報運営部 助教)		
授業の目的と概要	【講義の目的】大規模データを扱う際のサーバの操作、データの解析に利用するため、LinuxOSを操作できるように基本的なUNIXコマンドを学ぶ。 学習目標: 1. GUIとCUIの違いが理解できる。 2. UNIXコマンドを用いて簡単なデータ操作ができる。 【概要】 6月17日10:00-11:30 UNIXコマンド入門I 担当:関倫久 本講義はCUI操作未経験者を対象とした実習形式の講義である。UNIXコマンドにおける基本的な操作について学び、CUIにおけるディレクトリ操作などの基本的なコマンドを理解する。 6月17日12:30-14:00 UNIXコマンド入門II 担当:関倫久 大規模データ操作に必要となるUNIXコマンドについて学び、CUI操作によってテキストファイルのデータの把握、抽出などの基本的なコマンドを理解する。		
授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ実習形式の講義を予定している。		
履修上の注意事項	本科目はUNIX初学者の希望者のみを対象とした選択科目であり、本科目の履修によって履修証明プログラム修了のための単位は算定されないため注意すること。 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。		
教科書・参考書等	講義毎に異なる。適宜指示する。		

科目名	R入門		
日時	2023年5月10日、5月17日、5月24日、5月31日(水曜日18:00-19:30)、6月17日(土曜日14:15-17:30)	単位数	なし
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	岡田昌史(東京大学医学部附属病院企画情報運営部 特任講師)		
授業の目的と概要	【講義の目的】 近年データ解析に際して用いられる頻度の高いRプログラミングについて、主にプログラミング初学者やR未経験者を対象として、自身でRを用いてデータ処理ができる技術を習得することを狙う。 学習目標: 1. Rを用いて簡単なデータ操作ができる。 2. Rのデータ解析支援パッケージコレクションであるtidyverseを利用して基本的なデータ処理を行うことができる。 3. Rを用いて記述統計を用いたデータ集約や、基本的な統計検定、一般化線形回帰モデルの構築ができる。 【概要】 5月10日18:00-19:30 R入門I 担当:岡田昌史 本講義はRプログラミング未経験者を対象とした実習形式の講義である。まずRのインストールとR studioのインストールを行い、起動と終了の方法、Rによる基本的な演算を学ぶ。Rにおける基本的なデータ型とその操作について学び、簡単な計算をRを用いて実行できる能力を習得する。 5月17日18:00-19:30 R入門II 担当:岡田昌史 Rの基本的な操作を復習したのちに、パッケージのインストール方法、利用方法を学び、記述統計量の算出、外部からのデータ取り込み方法、グラフィックスの描画方法を学ぶ。 5月24日18:00-19:30 R基礎 担当:岡田昌史 tidyverseに含まれるdplyrをはじめとした各種データ解析用のパッケージを用いてデータの操作を行い、またggplot2を用いたデータの可視化を学ぶ。 5月31日18:00-19:30 Rによる統計解析入門 担当:岡田昌史 Rの各種データ解析支援パッケージを用いて、基本的な統計解析を実際に手を動かして学ぶ。また、一般化線形モデルの構築について実際にデータを用いて学ぶ。 6月17日14:15-15:45 R応用・質疑応答 担当:岡田昌史 Rの操作、各種統計解析手法について発展的事項を取り上げるとともに、これまでの実習について適宜質疑を受け付ける。 6月17日16:00-17:30 R応用・質疑応答 担当:岡田昌史 Rの操作、各種統計解析手法について発展的事項を取り上げるとともに、これまでの実習について適宜質疑を受け付ける。		
授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ実習形式の講義を予定している。		

履修上の注意事項	本科目はR初学者の希望者のみを対象とした選択科目であり、本科目の履修によって履修証明プログラム修了のための単位は算定されないため注意すること。 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。その場合、可能であれば事前にR studioのインストールを行い動作確認を行うのが望ましい。
教科書・参考書等	Rによる医療統計学 原書2版. 岡田昌史. 丸善出版.2017 (講義はこの教科書に沿って進行し、共通のサンプルデータを使用するが、特に予習や購入を指示するものではない。)

科目名	Python入門		
日時	2023年6月6日(火曜日)、6月13日(火曜日)、6月20日(火曜日)、6月27日(火曜日)いずれも18:00-19:30	単位数	なし
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	関倫久(東京大学医学部附属病院企画情報運営部 助教)		
授業の目的と概要	【講義の目的】 近年データ解析に際して用いられる頻度の高いプログラム言語であるPythonについて、主にプログラミング初学者やPython未経験者を対象として、自身でPythonを用いてデータ処理ができる技術を習得することを狙う。 学習目標: 1. Pythonの解析環境を自身で構築できる。 2. Pythonのデータ解析支援ライブラリであるPandasなどを利用して基本的なデータ処理を行うことができる。 3. PythonのNumpy、Pandas、scikit-learnなどを用いて簡単な機械学習モデルの構築を行うことができる。 【概要】 6月6日18:00-19:30 Python入門I 担当:関倫久 本講義はプログラミング未経験者を対象とした実習形式の講義である。Anaconda3環境下でJupyter lab環境を用いてPythonの基本的なデータ構造について学び、簡単なスクリプトを作成し実行できる能力を習得する。 6月13日18:00-19:30 Python入門II 担当:関倫久 本講義はプログラミング経験者かつPython未経験者を対象とした実習形式の講義である。Pythonの基本的なフロー制御、関数、データ構造について学習したのちに、小規模な公開データを操作し、Pandasを始めとしたデータ解析支援ライブラリに親しむ。 6月20日18:00-19:30 Python基礎 担当:関倫久 Pandas、scikit-learnをはじめとしたデータ解析に用いる各種ライブラリを用いてデータの操作を行い、実際の大規模データ解析を行う際に必要となるプログラムについてどのような手法が存在するのかを学び、自らデータ操作をできるようになることを目指す。 6月27日18:00-19:30 Pythonによる機械学習入門 担当:関倫久 各種ライブラリを用いて、機械学習モデル構築を実際に手を動かして学ぶ。scikit-learnを用いたモデル構築に取り組む際のデータ前処理、データ操作手法に関する知識の習得を目指し、簡単な予測モデル作成を行う。		
授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ実習形式の講義を予定している。		
履修上の注意事項	本科目はPython初学者の希望者のみを対象とした選択科目であり、本科目の履修によって履修証明プログラム修了のための単位は算定されないため注意すること。 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。		

教科書・参考書等	東京大学のデータサイエンティスト育成講座 Pythonで手を動かして学ぶデータ分析. 塚本邦尊, 山田典一, 大澤文孝(著), 中山浩太郎(監修), 松尾 豊(協力). マイナビ出版 独学プログラマー Python言語の基本から仕事のやり方まで. コーリー・アルソフ(著), 清水川 貴之(監訳). 日経BP
----------	--

科目名	自然言語処理入門		
日時	2023年9月7日、9月14日(木曜日18:00-19:30)	単位数	なし
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	篠原恵美子(東京大学大学院医学系研究科 医療AI開発学講座 特任助教)		
授業の目的と概要	【講義の目的】 医療分野におけるデータ解析では、構造化が不十分な自然言語で記載されたテキストデータへの対峙が余儀なくされるケースが多々ある。本講義では日本語のテキストデータを扱うための自然言語処理技術について概要を学ぶ。 学習目標: 1. 自然言語処理技術を用いたアプリケーションの例を挙げることができる 2. 形態素解析について概要を説明できる。 【概要】 9月7日18:00-19:30 自然言語処理入門 担当:篠原恵美子 自然言語処理技術の概要を事例の紹介を交えて学ぶ。 9月14日18:00-19:30 形態素解析入門 担当:篠原恵美子 代表的な形態素解析器であるMeCabを用いて自分で実際に形態素解析を行い、テキストデータの分析に親しむ。		
授業の方法	ノートパソコンを用い、データ処理について実際に手を動かして学ぶ実習形式の講義を予定している。		
履修上の注意事項	本科目は自然言語処理初学者の希望者のみを対象とした選択科目であり、本科目の履修によって履修証明プログラム修了のための単位は算定されないため注意すること。 本科目では実習においてPythonでのプログラミングを用いるため、Python未経験である場合には選択科目であるPython入門を受講することを強く推奨する。 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。 オンライン配信による受講も可能であるが、その場合教員による直接のサポートができない可能性があるため留意すること。		
教科書・参考書等	講義毎に異なる。適宜指示する。		

科目名	医用機械学習入門の手引き		
日時	2023年9月26日、10月3日(火曜日18:00-19:30)	単位数	なし
場所	代表校での対面による講義(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信を並行した授業を行う。対面による講義が開催不可能な場合はオンライン配信のみによる授業を行う。		
担当教員	河添悦昌(東京大学大学院医学系研究科 医療AI開発学講座 特任准教授) 関倫久(東京大学医学部附属病院企画情報運営部 助教)		
授業の目的と概要	医療分野における機械学習では、予測モデルの構築に際して適切なデータ収集、データ前処理、データ操作が必要になる。本講義では医療分野において機械学習モデル構築を行うに際して考慮が必要となる点について学び、実践する能力を身に付けることを狙う。 学習目標: 1. 医療分野における機械学習モデル構築におけるデータ準備の重要性を理解する。 2. 機械学習モデル構築の構築フロー、評価手法について説明できる。 3. 医療データを教師あり学習に資する形式へ変換して簡単な機械学習モデル構築を行うことができ、得られた結果の評価と解釈ができる。 【概要】 9月26日18:00-19:30 医用機械学習入門の手引きI 担当:河添悦昌・関倫久 機械学習モデルの構築時のフローを学び、医療データの準備に際してより質の高い学習データを得るための施策や適切な特徴抽出を行うためのデータ操作手法、必要となる評価手法について網羅的に学ぶ。 10月3日18:00-19:30 医用機械学習入門の手引きII 担当:河添悦昌・関倫久 ニューラルネットワークを用いた機械学習手法について、背景にある活性化関数・逆誤差伝搬法などの概念を理解する。実際のモデル構築を通じてその仕組みを学び、機械学習モデル作成の具体的な遂行方法について学ぶ。		
授業の方法	9月26日は座学形式、10月3日は実習形式での講義を予定している。実習形式の講義では、ノートパソコンを用い、データ処理について実際に手を動かして学ぶことを予定している。		
履修上の注意事項	本科目は医療分野の機械学習初学者の希望者のみを対象とした選択科目であり、本科目の履修によって履修証明プログラム修了のための単位は算定されないため注意すること。 本科目では実習においてPythonでのプログラミングを用いるため、Python未経験である場合には選択科目であるPython入門を受講することを強く推奨する。 実習に用いるノートパソコンは対面による講義が開催され教室内で使用する場合にのみ貸し出し可能(但し台数に制限あり)だが、学習の一環として履修生自身のノートパソコンの持ち込みによる受講が望ましい。		
教科書・参考書等	講義毎に異なる。適宜指示する。		

科目名	臨床医学概論		
日時	2023年度は講義動画を各自でオンデマンド視聴する方式とする。	単位数	なし
場所	オンライン配信のみによる授業を行う。		
担当教員	関倫久(東京大学医学部附属病院企画情報運営部 助教) 三宅加奈(東京大学医学部附属病院企画情報運営部 特任助教) 教材資料提供:康永秀生(東京大学大学院医学系研究科公共健康医学専攻臨床疫学・経済学 教授)		
授業の目的と概要	【講義の目的】 主に非医療者の履修生を対象とし、臨床医学に関する基礎知識を身に着けることを目的とする。臨床における主要かつ代表的な解剖・生理、診断・治療および病態生理について要点を学び、臨床的な医療課題の設計を理解できるようになることを目指す。また、医学用語を習得し、医学論文を読むための基礎力を養う。 学習目標: 1. 基礎的な医学用語を習得し、医学論文の理解を深める。 2. 臨床的な医療課題に対する具体的理解を深める。 【概要】(各90分間) 脳神経 循環器・呼吸器 消化器 腎・泌尿器、生殖器 感覚器、運動器 感染症、中毒、血液 内分泌・代謝、膠原病・アレルギー		
授業の方法	基本的に座学形式の講義である。		
履修上の注意事項	本科目は主に非医療者の履修生の希望者のみを対象とした選択科目であり、本科目の履修によって履修証明プログラム修了のための単位は算定されないため注意すること。 非医療者の履修生には本講義の履修を推奨する。なお、医療者である履修生にとっては既知の内容が大半となることに留意いただきたい。		
教科書・参考書等	講義毎に異なる。適宜指示する。		

一般履修コース1年次・インテンシブコース 実践課題研修 I

科目名	実践課題研修 I		
日時	課題研修成果発表会:2024年1月13日(土)・2024年1月20日(土) 発表会までに数か月をかけて課題に取り組む。	単位数	3
場所	課題設定および指導に関しては、担当教員と履修生で各自進めていく。 成果発表会は、代表校での対面開催(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信(外部公開を行う可能性がある)を並行して行う。対面での発表会が開催不可能な場合は、オンラインのみによる発表会を行う。		
担当教員	代表校および連携校の教員を中心に担当する。適宜指示する。		
授業の目的と概要	【講義の目的】 実践課題研修では、本事業の代表校・連携校・協力機関の複数名の教育担当者から提示された課題の中から履修生が取り組みたい課題を選択し、各履修生が課題に沿って文献等を参考に発表内容を準備し、教員及び他の履修生に対して発表を行い相互に議論を行う機会を通して医療データを巡る課題について実践的に思考する経験をする。なお、各課題は教育の範疇として行うものであり、本事業の範囲内では学術的な外部発表を想定した立て付けではないことに履修生および教員は留意する。 学習目標: 1. 複数講師より提案のあったテーマを自ら選択し、既存の利用可能な医療リアルワールドデータを用いてどのように分析が可能であるか検討する。 2. 設定されたテーマについて、テーマに対しての解決策を主体的に考えることができる。 3. 設定されたテーマに対して、適切な文献検討や分析などを行い、成果を分かりやすく適切に発表する手法を実践的に学ぶ。 【概要】 履修生に向けて実践課題研修の概要を提示する(2023年度前半を予定)。各履修生は、各自の興味・関心の方向に沿って複数提示されたテーマから一つの課題を選択し、取り組みを進める。課題研修成果発表会は2024年1月13日(土)・2024年1月20日(土)に予定する。 発表会は外部への公開形式とすることがある。		
授業の方法	上記の通りである。		
履修上の注意事項	履修証明プログラム修了のために本実習への参加は必修となり、成果発表会への出席が必須である(オンライン参加でも可)。 成果発表会当日は、自身の発表以外の発表時間にも参加し、他の発表を聴き議論に参加することを必須とする。 課題研修成果発表会(2023年度履修開始の一般履修コース履修生1年次・インテンシブコース履修生): 2024年1月13日(土)・2024年1月20日(土) 発表に使用したプレゼンテーション資料は別途提出すること。 なお本実習は教育の範疇として行うものであり、本事業の範囲内では学術的な外部発表を想定した立て付けではないことに履修生および教員は留意する。		

教科書・参考書等	必要である場合、担当教員より適宜指示する。
----------	-----------------------

一般履修コース2年次 実践課題研修Ⅱ

科目名	実践課題研修Ⅱ		
日時	課題研修成果発表会:2023年10月14日(土) 発表会までに数か月をかけて課題に取り組む。	単位数	3
場所	課題設定および指導に関しては、担当教員と履修生で各自進めていく。 成果発表会は、代表校での対面開催(東京大学 本郷キャンパス 医学部附属病院 南研究棟3階 鉄門臨床講堂)とオンライン配信(ただし外部公開は原則として行わない)を並行して行う。 対面での発表会が開催不可能な場合は、オンラインのみによる発表会を行う。		
担当教員	代表校および連携校の教員を中心に指導を行う。適宜指示する。		
実習の目的と概要	【実習の目的】 実際の医療データを取り扱うことで、一般履修コースで学んできた知識および技術を実際のデータ解析を経た成果物につなげるための実践的訓練とする。 なお、各課題は教育の範疇として行うものであり、本事業の範囲内では学術的な外部発表を想定した立て付けではないことに履修生および教員は留意する。 ただし、課題終了後も引き続き取り組み、将来的に学術的発表を検討するようなケースについては倫理申請などの手続きを含め、担当教員と個別に相談となる。 学習目標: 4.自らテーマを設定し、既存の利用可能な医療リアルワールドデータを用いてどのように分析が可能であるか検討する。設定したいテーマと利用可能なデータをすりあわせ、実際のリサーチクエスチョンを解決する手法を実践的に学ぶ。 5.本プログラム等で学習したR, SQLなどのコーディングを実践し解析を行うことで、コーディング技術のさらなる向上を図る。 6.本プログラムの範囲内では学術的な外部発表は行わないが、解析結果を分かりやすく適切に発表する手法を実践的に学ぶ。 【概要および授業の方法】 履修生に向けて課題研修の概要について提示を行う。詳細は該当履修生宛に周知する。 各履修生は、各自の興味・関心の方向に沿って仮のテーマを設定し、本事業事務局で希望テーマに応じた教員をティーチングアシスタントとして指名する。 利用するデータに関しては、仮テーマ設定時に希望を申し出ることが可能である。 ただし、履修生の所属施設等のデータ利用を希望する場合は適切に進めるための事前確認が必要となるため、事務局宛にも事前に相談することが望ましい。 研究計画書作成は、ティーチングアシスタントと相談し課題設定後約1か月をめどに行う。 各履修生とティーチングアシスタント間で個別に相談し(オンラインでも可)、数か月をかけて解析および発表会準備を進めていくこと。 課題研修成果発表会:2023年10月14日(土)に予定する。 外部への公開は原則として行わない。		
授業の方法	上記の通りである。		

履修上の注意事項	履修証明プログラム修了のために本実習への参加は必修となり、成果発表会への出席が必須である(オンライン参加でも可)。 成果発表会当日は、自身の発表以外の発表時間にも参加し、他の発表を聴き議論に参加することを必須とする。 課題研修成果発表会(2023年度履修開始の一般履修コース履修生2年次): 2023年10月14日(土) 発表に使用したプレゼンテーション資料は別途提出すること。ほか指定フォーマットでの要約の提出を別途指示する。 なお本実習は教育の範疇として行うものであり、本事業の範囲内では学術的な外部発表を想定した立て付けではないことに履修生および教員は留意する。 ただし、課題終了後も引き続き取り組み、将来的に学術的発表を検討するようなケースについては倫理申請などの手続きを含め、担当教員と個別に相談となる。
教科書・参考書等	必要である場合、担当教員より適宜指示する。