

厚生労働科学研究費補助金（がん対策推進総合研究事業）  
（分担研究報告書）

全国がん登録と連携した臓器がん登録による大規模コホート研究の推進及び  
高質診療データベースの為のNCD長期予後入力システムの構築に関する研究

『臓器がん登録による』ビックデータ分析の課題  
エビデンスとしてのバイアスと価値、欧米の実情と実用化への課題

（研究分担者：水口 徹 札幌医科大学 消化器・総合、乳腺・内分泌外科学講座）

研究要旨

医療ビックデータの利活用の課題点を明らかにするため、米国・英国におけるがん登録システムの現状を明らかにした。米国においてはSEERプログラムとNPCRプログラムによって、ほぼ全人口をカバーし、データの利活用も行われている。英国においてはukiacrがデータを中央化し、登録・管理・教育からデータの還元までを行っている。ビックデータには、レセプトデータ、DPCデータ、ゲノムデータ、調剤データ、検査データ、病理データなどが含まれ、これらの質的側面・多様性・客観性を如何に評価・活用するかが課題である。国際的な視野から考えていくことが必要で、そのための国際教育化・産業活性化を通して、ビックデータの利活用が新たな医療情報産業として、本国から輸出するべくがん登録・利活用に本腰を入れるべきと考えられた。

A. 研究目的

現在行われている米国と英国の臓器がん登録について、登録システムとその運用方法を明らかにすることを目的とした。また、臓器がん登録情報を代表としたビックデータ解析・利用がもたらす利点と欠点を明らかにすることを目的とした。

B. 研究方法

米国に関してはSEERプログラムとNPCRプログラムを中心に解析した。英国ではukiacrを中心として解析した。

C. 研究結果

米国のがん登録

SEER (The Surveillance, Epidemiology and End Results) プログラムは1973年にThe National Cancer Institute (NCI) で開始された。米国人口の約28%のカバー率である。

NPCR (The National Program of Cancer Registries) は1992年にThe Centers for Disease Control and Prevention (CDC) によって開始され、45州において48の公的がん登録機関によって登録されており、米国人口の約96%のカバー率である。

NPCRは下部組織のデータ監査を行っておりCalifornia Cancer Registry (CCR) におけるQuality controlは98.7%の正確性であったと報告されている。

データ登録に際しては、検査部門や病理部門の情報が統合されて登録されており、登録者の教育も充実している。

これらの情報は、公表されており、特定の地域における高危険群の洗い出し、スクリーニングの普及、がんの原因究明などに

活用されている。

英国のがん登録情報

英国ではukiacrが、National Cancer Registry Ireland, Northern Ireland Cancer Registry, Public Health England, Scottish Cancer Registry, Welsh Cancer Intelligence and Surveillance Unitのデータを中央化し、登録・管理・教育からデータの還元までを行っている。

データ解析・教育コード分類化・管理保証などのチームから構成され、登録や出力のサポートなどの組織体制が整備されている。

ビックデータ活用

ビックデータは3V+ $\alpha$ とされ、Volume・Velocity・VarietyにVeracity・Visualization・Valueと定義されている。データの管理・分析・構造化・可視化によって、プロセスの最適化・効率化を図れるものである。

ビックデータには、レセプトデータ、DPCデータ、ゲノムデータ、調剤データ、検査データ、病理データなどが含まれるが、これらの質的側面・多様性・客観性を如何に評価・活用するかが課題である。

個人情報量としては、ペタバイト、エクサバイトの情報量となり、統合解析に加え能動的解析が必要になる。

D. 考察

米国のがん登録は、全米人口をカバーしており、データの利活用も含めて先進的と言える。一方で、州ごとに取り組み方に違いが見られ、統一した政策方針ではないことが問題点として挙げられた。

英国のがん登録は、組織が有機的に機能しており、がん登録のみならず、教育・普及なども含めて社会に還元できる体制が整備されている。

International Cancer Control Partnership (ICCP) 2012は、American Cancer Society (ACS), CDC, NCI, NCCN, UICC, WHOなどが参加した国際的ながん登録推進機構であるが、日本にはがん登録システムが存在しないとの情報が公開されている。ICCPでは、がん登録システムを発展国においてどのように導入するかマニュアルを公開し、データ悉皆性と信頼性を担保するための方策を提示している。がん登録後進国としては、先進諸国の登録システムやマニュアルを参考にデータ利活用も含めた包括的な体制整備が急務と考えられる。

ビッグデータの利活用には、データ処理過程におけるシステム自身の自己学習や能動的知識の蓄積ができる人工知能を応用したアプリケーションが必要であると考えられた。既にIBMではいくつかの診療支援モデルが開発されており、情報管理にベイズ理論を組み込んだ実地臨床における構造化整備が必要と考えられた。

ビッグデータ解析・利用の利点としては診療情報の共有化による効率化・均てん化、予防医学の発達に伴う、罹患率低下、病院機能の差別化と分担化、社会資源の有効活

用が考えられる。一方で、欠点としては特定病院への患者集中・中規模病院の経営難、特定の情報による人の選別、患者集中に伴うサービス低下・患者不満の増加、医師の診療科選択の制限などが考えられた。

今後の課題として、がん登録システムそのもののインフラ整備が必要で、現状の地域がん登録が印刷情報によってデータ登録が行われている現状は悲観すべきものがある。データ管理の安全性を確保した上で、統一したオンライン化が必要である。

また、今後はデータ利活用を国内のみならず、国際的な視野から考えていくことが必要で、そのための国際教育化・産業活性化が課題となる。結果として、ビッグデータの利活用が新たな医療情報産業として、本国から輸出するべくがん登録・利活用に本腰を入れるべきと考えられた。

#### E. 結論

米国・英国におけるがん登録システムを理解した。ICCPで指摘されるように、本国はシステムの構築がなされつつある発展過程にある。得られる情報量は莫大であるが、その利活用には国際的な視野で、人材教育を行いつつ、医療輸出国を目指して産業活性化を行っていく必要があるものと考えられた。