

의료기관 빅데이터를 활용한 임상개발전략지원 사례집



들어가는 글

국내외 의료빅데이터 활용 가이드라인 발표 이후 전세계적으로 실사용증거(Real World Evidence, RWE)를 생성하는 연구가 활발히 진행되고 있습니다. 국내에서도 치료제 품허가기관, 요양급여 심사기관, 보건의료기술 평가기관 등에서 관련 사업을 통해서 지원 방안을 추진하고 있습니다.

연세의료원 연구개발자문센터에서는 'SALT PLUS' 협력 프로그램을 통해 국내 최초로 의료빅데이터 분석 기반의 임상 자문을 제공하고 있습니다. 제약·바이오 기업이 보유한 치료 제품을 대상으로 의료기관의 연구인프라를 활용하여 산학병 간의 열린 협력을 지원하고 있습니다. 본 사례집에서는 지난 3년 간 프로그램 운영을 통해 축적한 경험을 토대로 연구성과 사례들을 공유하고자 합니다.

먼저, 환자 중심의 치료 제품 개발을 위해 필요한 전략 7가지를 사례별로 제시하였습니다. 또한, 시판 제품 혹은 질병 연구 관점에서 등록 연구(Registry Study)의 중요성을 언급하면서 데이터베이스(Database, DB) 구축 사례도 소개하였습니다. 최근 의료빅데이터는 연구개발 과정에서 환자와 연구자 간의 소통 채널로도 인식되고 있습니다. 이에 따라, 빅데이터를 의미 있는 정보로 생성하는 분석 기술은 반드시 필요합니다. 빅데이터 관련 기술 동향과 연구 사례는 마지막 장에서 확인하실 수 있습니다.

본 사례집은 실사용증거 생성을 위한 가이드북('RWE Guidebook', 연구개발자문센터, 2020)에서 소개한 내용을 사례 중심으로 구성하였습니다. 의료기관의 임상연구자와 제약·바이오 기업 R&D 전문가에게 많은 도움이 될 수 있기를 바랍니다.



목 차

Part I

환자 중심 임상개발전략

1. 목표 질병 특성 파악	05
2. 환자의 치료 여정 이해	06
3. 연구 대상자 강화 전략	08
4. 미충족 의료 수요 수집	09
5. 차별적 경쟁력 요소 반영	11
6. 임상 개념 입증 전략 수립	12
7. 시판 후 임상연구 확대	13

Part II

등록 데이터베이스 연구

등록 데이터베이스란	19
등록 데이터베이스 구축 현황	20

Part III

빅데이터 분석 기술

의료빅데이터 분석 기술의 연구배경	23
의료빅데이터 분석 모델과 솔루션	23

PART I

환자중심 개발전략

환자 중심의 임상개발전략 수립에 필요한
7가지 접근방법을 제시합니다



1. 목표 질병 특성 파악

Goal

목표 질병(Disease)에 대한 이해는 맞춤형 정밀의학 연구에 중요한 정보를 제공합니다. 하나의 진단명이라도 다양한 병인으로 인한 병태생리학적 특성이 다를 수 있습니다. 예를 들어, 제2형 당뇨병도 4가지 특성으로 구분됨을 시사한 논문이 최근에 발표되었습니다.¹⁾ 또한, 주 진단명 외에도 동반 질환을 확인하여 환자군 분포와 병용하는 약물 간의 상호작용 등도 이해해야 합니다. 이와 같이, 질병에 대한 자세한 정보는 맞춤형 치료 제품을 개발하는데 중요한 근거로 활용됩니다. 이러한 선행 조사를 통해 환자의 포괄적인 질환(Illness)도 이해함으로써 삶의 질에 영향을 주는 요소까지 파악할 수 있습니다.

Case

치료 제품 개발과정에서 병인이 충분히 밝혀지지 않는 질병에 대해 연구 대상 환자군을 선정하는 경우도 존재합니다. 다음 질병들은 EMR 빅데이터를 통해서 임상적 특성을 구분할 수 있습니다.

① 비알코올성지방간염(NASH)

해당 환자군은 간의 지방량, 섬유화 및 염증의 정도에 따라 경증에서 중증까지 구분됩니다. 또한 간경변증을 유발할 수 있는 위험 요인으로 당뇨와 비만 등이 있습니다. 따라서 진단 및 예후 인자를 고려할 때 다양한 환자 군의 분포가 존재합니다.

② 이식편대숙주병(GvHD)

이식편대숙주병(GvHD)은 동종 조혈모세포를 이식 받은 환자 대상 주로 발생하는 합병증입니다. 장기 이식 수술 시, 이식 거부 반응 예방을 목적으로 면역 억제제를 사용하기는 하나, 급성 및 만성 GvHD가 발생하면 표준 치료인 스테로이드가 사용됩니다. 이에 불응하는 경우에는 항암제 등 다른 치료제들로 대체됩니다.

EMR 빅데이터 분석 정보는 연구 기획 혹은 수행 단계에서 미처 확인되지 못한 질병의 특성을 이해하는데 도움되어 임상개발전략을 보완하는 중요한 역할을 합니다.

- + 진단 환자 규모 및 진료 협진
- + 진단 및 예후 인자들에 따른 환자군 분포
- + 표준 치료 종류 및 불응 대상 환자 정보



Check Point

- ✓ 연구 대상군 선정 시 질병의 특성을 자세히 파악하는 것이 필수 조건입니다.
- ✓ 환자의 삶의 질에 영향을 주는 요소를 이해하는 것도 필요합니다.

2. 환자의 치료 여정 이해

Goal

주 호소(Main Complaint)로 병원에 방문한 환자의 여정(Patient Journey)²⁾ 정보는 치료 제품 연구개발에 중요한 의사결정 판단 근거가 됩니다. 진단 이후 일차 치료의 종류(수술·시술 및 약물)를 파악하고, 이후 해당 치료의 지속 혹은 변경 과정을 확인할 수 있습니다. 예를 들어, 주 치료로 인해 질병 관해가 되면 이후 유지 요법 치료가 사용되지만, 기존 치료에 불응하는 경우에는 다른 치료법이 적용되기도 합니다. 이와 같이 EMR 빅데이터 분석 정보는 실제 진료 환자의 치료 과정을 자세히 대변할 수 있는 장점이 있습니다. 이러한 정보는 컨설팅 기업의 보고서 등에서는 쉽게 파악하기 어렵습니다. 따라서 의료기관의 빅데이터를 통해 분석한 환자의 치료 여정 정보는 치료 제품의 포지셔닝 전략 수립에 중요한 정보를 제공해 줍니다.

Case

EMR 빅데이터 분석을 통해 연구 목표 환자군의 환자 치료 여정 지도를 제시한 사례들을 소개합니다.

① 성인 말단비대증(Acromegaly)

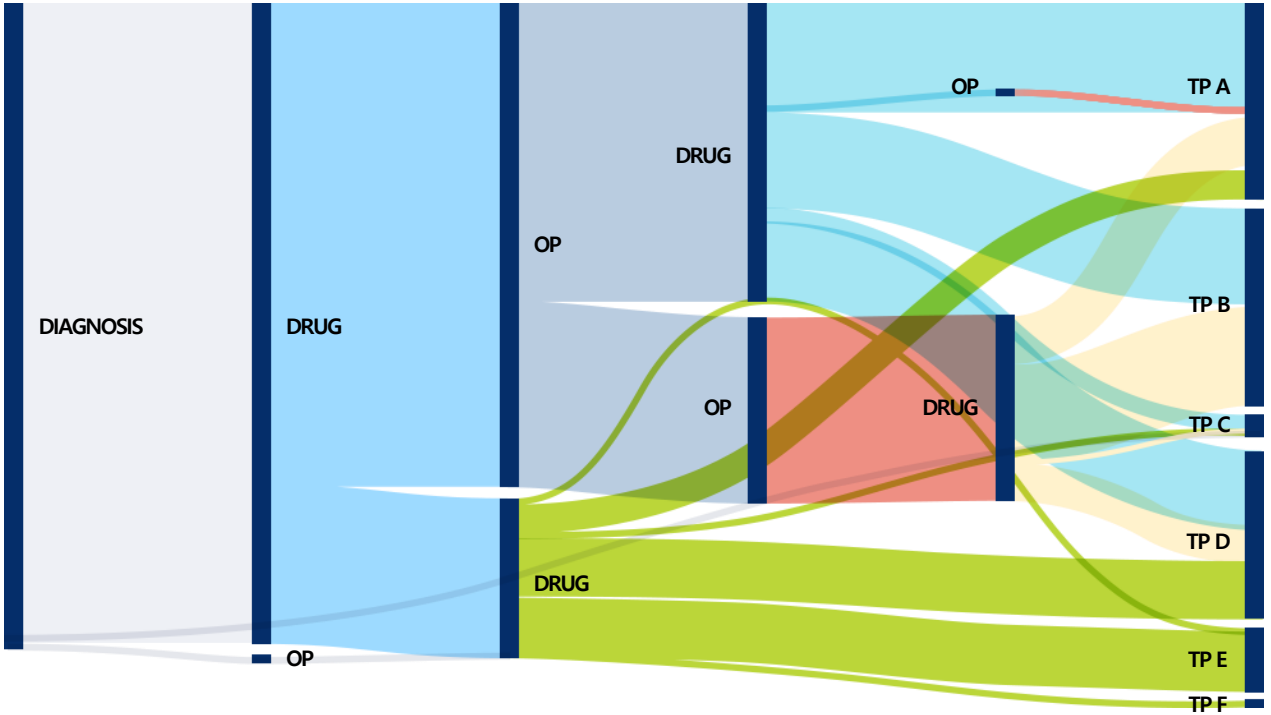
과도한 성장호르몬 분비가 원인으로 뇌하수체 선종을 제거하는 수술과 호르몬 분비를 억제하는 약물 치료 등이 존재합니다. 주로 종양 제거 수술 이후 방사선 치료법과 약물 보조요법으로 구분 될 수 있습니다. 약물 치료제 개발 시 이러한 수술 및 시술에 노출되는 환자군을 정확히 파악하는 것이 중요합니다.

- + 환자 여정에서 치료법의 변화 양상
- + 치료 과정에서 수술·시술법과 약물 요법 간의 분포
- + 표준 치료 불응성 환자의 대안 치료법 종류
- + 조직병리학적 검사를 통한 치료제 선별 기준 유무

② 난치성 뇌질환(Brain Disorder)

뇌 병변 수술 전에는 약물 치료법이 선호되지만, 기존 표준 약물 치료에 잘 반응하지 않거나, 치료 효과가 없는 경우 절제술을 선택하기도 합니다. 수술 과정 중 시행된 조직 생검으로는 병인 혹은 위험 요인을 파악할 수 있으며, 수술 후 치료법 선별 근거로도 사용할 수도 있습니다.

Figure 1. 난치성 희귀병 치료 여정 예시(Sankey Diagram)



* OP (Operation): 수술
* TP (Treatment Progress): 치료경과(완치, 약물유지, 재발 등)



Check Point

- ✓ 개발 대상 치료 제품의 포지셔닝 전략 수립을 위해서 의료빅데이터 분석을 통한 환자 여정 파악이 중요합니다.
- ✓ 임상 연구자의 자문은 치료 여정 지도에서 파악된 치료법 변화에 대한 임상적 근거를 보다 명확하게 제시해 줍니다.

3. 연구 대상자 강화 전략

Goal

정밀의료 임상시험을 위해서는 환자 층화 구조를 이해하고 연구 대상자를 강화하는 전략(Enrichment Strategy)³⁾ 중요합니다. 환자의 층화 구조에는 기저질환, 증상의 양상, 치료 반응 예측 및 질병 예후 인자가 속하며 최대한 동질(Homogeneous)한 환자군을 선별하는 것이 필요합니다. 과거에는 같은 적응증을 가진 환자 모두에게 동일 치료제가 처방되는 방식이었다면, 이제는 목표 환자 소그룹에 따라 치료를 구체화 하는 맞춤형 시대입니다. 새로운 치료기술 개발 과정에서 질병의 특성 및 치료 효과와 연관된 생체지표(Biomarker), 대리지표(Surrogate Endpoint) 및 치료적 임상 결과(Clinical Outcome)를 구체화 하여 층화 요인과 평가 지표 간의 상관관계를 이해해야 합니다. EMR 빅데이터 분석 정보는 이러한 층화 요인을 통해 맞춤형 치료제 개발에 필요한 정보를 제공해 줍니다.

Case

현재, 난소암(Ovary Cancer)과 폐암(Lung Cancer)의 특정 종양 자료를 수집하고 데이터베이스로 구축하고 있습니다. 본 연구들은 다국적 또는 국내 다기관 형태로 진행되는 과제로서 다음과 같은 정보를 수집하여 층화 요인으로 활용하고 있습니다.

- + 환자 기본 정보
- + 종양의 병기 및 조직병리학적 특성
- + 약물치료타겟 및 유전자 변이 분포
- + 질병 예후 및 치료 예측 인자
- + 용적 축소 수술, 방사선 보조치료, 선행/보조 화학요법
- + 치료적 임상 결과(무진행생존기간, 전체생존기간)

Figure 2. 질병의 특성과 관련된 데이터 수집

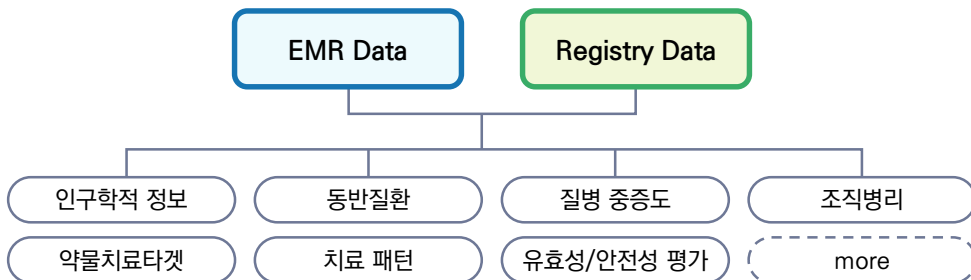


Figure 3. 비소세포폐암(NSCLC)에서 약물치료타겟인 EGFR mutation subtype에 따른 분포 예시



Check Point

- ✓ 임상개발 대상 목표 환자군을 정의하고 이를 선별할 수 있는 환자 층화 요인을 파악하는 강화 전략(Enrichment Strategy)이 중요합니다.
- ✓ 강화 전략의 목적은 임상시험 대상자의 이질성을 가능한 배제하여 동일 환자군에서 맞춤형 치료 제품의 효과를 확인하는 것입니다.

4. 미충족 의료 수요 수집

Goal

실제 진료 환경에서 쌓인 의료빅데이터와 임상 연구자의 경험 등을 통해 환자중심 미충족 의료 수요(Unmet Medical Needs)를 정확하게 이해하면 새로운 치료 제품 개발을 위한 정확한 임상 개발 전략을 수립할 수 있습니다.

당뇨병 환자의 경우, 다양한 계열의 약물들이 개발되었지만, 장기간 투여 이후에도 목표 혈당 수준에 도달하지 못하는 경우가 흔하며, 유병 기간 동안 발생한 합병증으로 인해 사망률이 높은 고위험군은 점차 늘고 있습니다. 이와 같은 경우는 목표 혈당 관리와 동시에 합병증을 예방하는 것이 미충족 의료 수요입니다. 이러한 수요를 개발 전략에 반영하는 것은 시장에서의 성공과 직결됩니다.

Case

당뇨병 치료제로 개발되어 심부전 관련 혜택을 입증한 약물이 있습니다. 심부전 동반 환자는 예후가 좋지 않아서 심부전 위험 조절 여부가 주요한 치료 목표입니다. 최근 당뇨병 치료제 중에서 SGLT-2 억제제 계열 내 다수 약물이 심부전 예방 효과를 입증하였습니다. 이런 근거들로 인해 2019년 유럽심장학회 가이드라인에 심질환을 동반한 제2형 당뇨병 환자의 경우, SGLT-2 억제제가 1차 치료 약물로 권고되었습니다. 이뿐만 아니라 SGLT-2 억제제는 신기능 보호 작용을 통해 심부전 환자군에서도 새로운 혜택을 보여줄 가능성이 있습니다. 따라서, 이 경우 다양한 임상 근거 생성을 위해 추가 연구기획 단계에서 다음과 같은 사항이 고려될 수 있습니다.

- + 내분비내과, 심장내과, 신장내과 등 다학제적 의견 수렴 필요
- + 생체지표(Biomarker), 대리지표(Surrogate Endpoint) 등 평가 지표 선정
- + 당뇨병 환자 외에도 비당뇨병 환자 대상 연구 모델 확장 가능성
- + 심부전과 급성 심부전 동반 환자 대상 연구 가능성
- + 복합제 개발을 고려한 약물 후보군

Check Point



- ✓ 진료환경에서 미충족 의료 수요를 이해하기 위해 임상 연구자의 다양한 의견을 체계적으로 수집하는 것이 중요합니다.
- ✓ 미충족 의료 수요는 개발, 규정, 급여, 진료 등 관련 이해관계자에 따라 바라보는 관점이 다를 수 있습니다.⁴⁾

5. 차별적 경쟁력 요소 반영

Goal

규제기관의 위험-혜택 평가를 통해 제품이 허가되더라도 시장에서 실패하는 주된 이유는 경쟁 치료제 대비 부족한 차별성입니다. 경쟁 요소를 파악하기 위해서는 시장이 원하는 가치를 파악하고 이를 입증하는 임상개발전략이 중요합니다. 또 다른 실패 이유는 환자가 원하는 혜택(예, Clinical Outcome)이 허가용 무작위 배정 임상시험(RCT)에 평가 요소로서 반영되지 않았기 때문입니다. 미국 FDA는 환자의 요구사항을 체계적으로 임상시험에 반영하여 평가하고자 4단계 연구방법론을 제시하였습니다.⁵⁾ 마지막으로, RCT와 같이 다소 실험적 환경에서 획득한 임상 결과가 실제 진료 환경에서 얻어진 유효성·안전성과 차이가 있을 수 있다는 점도 충분히 인지하여야 합니다.

Case

폐동맥 고혈압(Pulmonary Arterial Hypertension, PAH)은 흔히 전신고혈압과 혼동되지만 전혀 다른 질환입니다. 적절한 평가 요소는 전신 수축기혈압(SBP) 이 아닌 폐혈관 저항으로 인한 수축기 우심실압(Right Ventricle Systolic Pressure, RVSP)의 개선 여부입니다. 주된 사망 원인이 우심부전이기 때문입니다. 안전성 평가 측면에서도 국내 허가된 계열 약물 중에 RVSP 조절이 선택적이지 않아 SBP에 관여하여 어지러움 등 부작용이 발생하는 경우도 있습니다.

PAH 치료에서 1차 약물로는 경구용 엔도텔린 수용체 길항제(ERA)가 사용됩니다. ERA는 복용 편의성과 부작용 발현 비율이 낮아서 약 50% 이상 단독으로 처방됩니다. 따라서 임상개발 후보 포지셔닝은 ERA와 병용 가능한 2차 약물이 될 것이며 경쟁 제품 역시 2제 병용 약물이 될 것입니다. 이외에도 제품 경쟁력을 위해서는 제품의 효능 외에 부작용 및 편의성에서도 차별성이 있어야 합니다. 예를 들어 정맥으로 투여되는 3제 약물은 진료상 입원을 요구하나, 이를 원하지 않는 환자가 상당수 존재합니다. 따라서 PAH 대상 약물 개발 시 경구 제형은 그 자체로 충분한 장점이 있습니다.

- + 경쟁 치료제를 파악하고, 부작용과 복용 중단 사례를 참고하여 우위 전략 수립
- + 진료 특성상 병용 투여가 불편하거나, 입원을 요구하는 약물 개발은 가능한 회피

Check Point



- ✓ 차별적 경쟁 요소는 미충족 의료 수요 파악과 긴밀하게 연결됩니다.
- ✓ 의료빅데이터 분석 정보와 임상 연구자의 진료 경험은 차별적 경쟁 요소를 명확하게 파악하는데 최적 조건입니다.

6. 임상 개념 입증 전략 수립

Goal

임상 단계 개념 입증(Proof-of-Concept, PoC)⁶⁾은 동물모델에서 확인된 제품의 경쟁력을 초기 임상시험에서 증명하는 것입니다. 다만, 모델의 차이에 의해 동물에서 확인된 효능이 임상시험에서 입증되지 않는 경우도 있습니다. 또한 목표 환자군이 불명확하여 실패하거나, 경쟁 제품 대비 차별화 계획이 간과되어 후기 임상시험에서 고전하기도 합니다. 따라서, 개념 입증 전략 수립을 위해 기업 연구개발 전문가와 임상 연구자의 면밀한 논의를 통하여 다학제적 접근에 대한 검토가 선행되어야 합니다. 임상시험 대상 목표 환자군을 분석하거나 혹은 부족한 임상 근거를 보충하기 위해 추가적인 비임상시험이 계획되기도 합니다. 성공적인 임상개발은 이러한 집단 지성에 의해 실현될 수 있습니다.

Case

임상시험용 의료기기인 지혈패드를 인체에 사용 하기 위해 수술 대상 환자군을 고려하여 선정합니다. 수술 건수가 많아 대상자 모집이 어렵지 않으면서, 동시에 출혈 정도가 심하지 않은 장기의 이식술 혹은 절제술을 선택함으로써 실패 위험성을 낮추어야 합니다. 또한 개복술 대비 복강경 수술의 빈도가 증가함에 따라 수술법에 의한 차이들도 제품 연구개발 과정에 반영되어야 합니다. 예를 들면, 복강경 수술 시 절개구멍으로 지혈패드가 쉽게 통과되는지, 검자에 의해 조작이 편리한 제형인지 등의 고려가 필요합니다. 또한 지혈패드는 일차 지혈법 이후 지속되는 지혈에 주로 사용되기 때문에 봉합 등 복강경에서 불가능한 시술법도 인지하여 임상시험 설계 시 반영해야 합니다.

- + 초기 임상시험에서 개념 입증에 적합한 연구 대상자 선별 조건
- + 해당 환자가 진료 환경에서 노출되어 결과에 영향을 줄 수 있는 요소 파악



Check Point

- ✓ 진료 환경에서 차별적 경쟁력을 명확하게 사전 정의하고, 이후 비임상시험에서 최대한 확인하여서 임상 개념 입증 전략에 반영해야 합니다.
- ✓ 개발 회사와 임상 연구자 간의 집단지성을 통해 필요시 위험관리 방안을 수립하여 임상시험 설계에 반영해야 합니다.

7. 시판 후 임상연구 확대

Goal

시판 후 임상연구는 진료 현장에서 처방된 의약품 등 치료제에 대한 유효성, 안전성, 유용성 등을 확인할 수 있다는 점에서 그 중요성이 강조되고 있습니다. 장기 관찰을 통해 허가 전 임상시험에서 확인하지 못한 임상 결과(Clinical Outcome) 확인도 가능합니다. 다만, 장기간 대규모 연구 진행을 위해서는 사전에 타당성 연구가 필요할 수 있습니다. EMR 빅데이터를 활용한 후향적 임상연구가 이것에 대한 솔루션이 될 수 있습니다. 빅데이터 조회, 추출, 처리 및 분석 과정을 통해서 보다 정교한 임상연구를 수행할 수 있는 분석툴이 개발되고 있기 때문입니다(PART III 내용 참조). 임상 결과(Clinical Outcome) 뿐만 아니라, 연구 대상 코호트의 인구학적 특성 분포, 치료 여정, 약물 처방 패턴, 치료 효과 등을 확인할 수 있습니다.

또한, 후향적 임상연구 결과는 전향적 임상시험 주제 발굴에 아이디어를 제공하거나, 연구 설계에 중요한 정보가 되는 환자 특성 분포, 층화 요인, 표준치료 패턴 등을 제공 합니다. 최근 관심을 보이기 시작하는 하이브리드 형태의 전·후향적 관찰 연구에서는 후향적 데이터를 기반으로 연구 대상자를 미리 선별하여 전향적 연구를 진행함으로써 해당 환자의 기존 임상 데이터도 함께 활용할 수 있습니다.

Case

제 2형 당뇨병을 목표 질병으로 다양한 목적의 후향적 임상연구를 수행하였습니다.

① 비교 유효성(Comparative Effectiveness)

본 연구는 의료빅데이터를 활용하여 연구 목적에 적합한 대조군을 탐색한 연구입니다.

세브란스 병원 내 기록된 최근 10년 동안의 의무기록 데이터를 활용하여 동종 치료 약물 계열 내 전환 처방 환자 수를 파악하고, 전환 처방에 따른 당화혈색소(HbA1c) 수치 수집 가능성도 조사하여 대조군을 확보하고자 하였습니다.

전환 처방 시행 환자들의 ① 당뇨병 최초 진단 이후 전환 처방이 시행된 시점과 ② 전환 처방 전후 병용 처방된 약물의 조성 종류가 이 다양하게 분포되어 있는 점을 고려하여 적합한 연구대상자를 선정하였습니다. 하지만, 대조군 선정 과정에서 처방 패턴이 매우 다양하고, 평가 항목이 실제 수집 시점에서 존재하지 않는 경우도 있어 이에 따라 환자수가 크게 줄어드는 점을 인지해야 합니다.

본 연구는 간질환 치료제와 당뇨병 치료제의 병용에 따른 당화혈색소(HbA1c) 개선 효과를 분석한 연구입니다.

세브란스 병원 내 기록된 과거 8년 동안의 의무기록 데이터를 활용하여 간기능 개선제와 당뇨병 치료제 병용 환자군과 당뇨병 치료제 단독 복용 환자군 간 당화혈색소(HbA1c) 개선 효과를 평가하였습니다.

환자마다 간기능 개선제와 당뇨병 치료제를 동시 처방한 구간이 진단 이후 최초 병용 시점부터 다양하게 분포함을 확인하였고, 단독 복용 환자군은 유사한 병용 시점 및 투여 기간을 고려하여 선정하였습니다. 당뇨병 치료제 전체 코호트에서 구간 유의미한 당화혈색소 개선 효과는 보이지 못했습니다만, 특정 소그룹에서 유의미적 차이를 보이는 환자 층화 조건을 확인하였습니다.

Figure 4. 전환 처방 전후 HbA1c 변화(%) 비교 예시

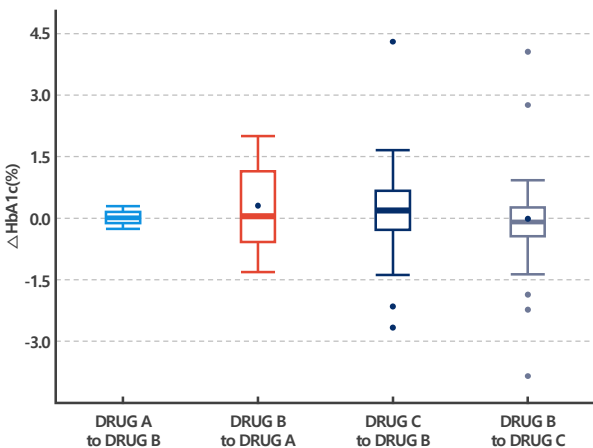


Figure 5. 성향점수매칭 전후 군 간 HbA1c 변화(%) 예시

	매칭 전			매칭 후		
	병용 처방군	단독 처방군	P-value	병용 처방군	단독 처방군	P-value
나이	51.1 ±11.5	54.5 ±10.7	0.0034	53.0 ±11.7	52.3 ±11.0	0.6798
성별			0.8853			0.6473
여자	38.4%	39.2%		40.0%	36.7%	
남자	61.6%	60.9%		60.0%	63.3%	
Baseline HbA1c(%)	7.85 ±1.79	7.44 ±1.48	0.0211	7.58 ±1.54	7.60 ±1.78	0.9361
Change in HbA1c(%)	0.27 ±0.93	0.51 ±1.43	0.0451	0.26 ±0.94	0.47 ±1.48	0.2737

Case

제 2형 당뇨병을 목표 질병으로 다양한 목적의 후향적 임상연구를 수행하였습니다(계속).

② 임상 결과(Clinical Outcome)

본 연구는 다기관 확대 연구의 가능성을 사전 파악하기 위해 실시한 시범 과제입니다.

세브란스 병원 내 기록된 최근 10년 동안의 의무기록데이터를 활용하여 약물 처방 환자 수를 파악하고, 당뇨병 치료제 복용에 따른 미세혈관 합병증 발생률을 조사하였습니다.

① 치료 기간 내 처방된 당뇨병 치료제의 종류와 ② 관심 약물의 처방 시점 및 기간이 다양하게 분포되어 있는 점을 고려하여 연구 대상자를 선정하였습니다. 또한, PDC(Proportion of Days Covered) 개념을 적용하여 관심 약물의 처방 기간이 전체 약물 처방 기간의 80% 이상 분포하고 있는 환자를 연구 대상으로 선정하였습니다. 본 연구에서는, 시범 단계에서 비교 약물 군 간 연구대상자의 기저 질환 분포가 유의미하게 상이한 점을 발견하여, 다기관 확대 연구의 필요성이 고려되었습니다.

본 연구는 스타틴 계열 별 신규 당뇨병 발생률을 분석한 연구입니다.

세브란스 병원 내 기록된 과거 6년 동안의 의무기록데이터를 활용하여 스타틴 계열 별 장기 복용에 따른 신규 당뇨병 발생(NODM) 비율을 비교 평가 하였습니다.

연구 대상자의 탈락을 최소화하기 위해 성향점수 역확률 가중치(IPTW) 방법론과 콕스 회귀분석을 사용하여 NODM 발생률을 분석하였습니다.

본 연구에서는 만성 신질환(Chronic Kidney Disease, CKD) 유병 여부에 따라 NODM 발생이 통계적으로 유의미하게 차이가 나는 것을 확인할 수 있었습니다. 유사한 결과가 최근에 발표된 연구 결과에서도 확인되었습니다.⁸⁾

Figure 6. Clinical Outcome 연구 결과 예시(Kaplan–Meier Curve)

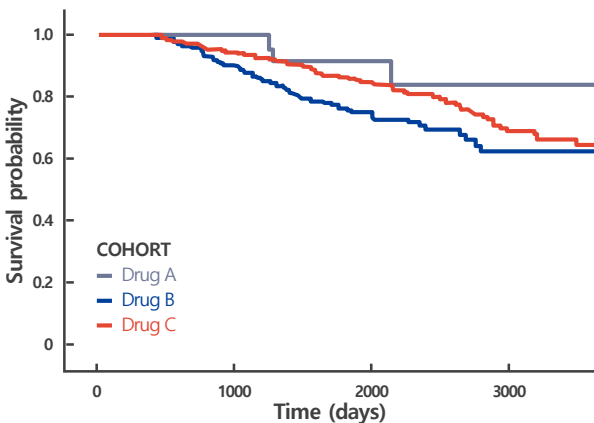
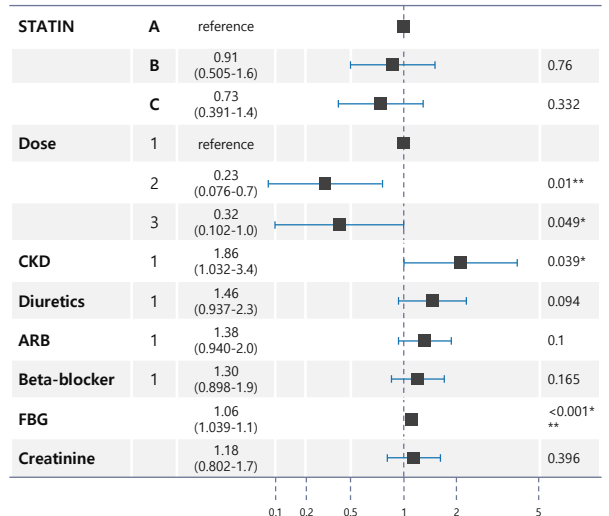


Figure 7. Clinical Outcome 연구 결과 예시(Hazard Ratio)



Case

이 외, **고혈압, 간세포성 암종 등** 목표 질병으로 다양한 목적의 후향적 임상연구를 수행하였습니다

③ 환자 탐색(Refined General Population)

본 연구는 본태성 고혈압 환자를 대상으로 관심 약물이 포함된 2제 병합 요법에 대해 효과를 보이는 환자군을 탐색하는 연구입니다.

세브란스 병원과 강남 세브란스 병원 내 기록된 과거 6년 동안의 의무기록데이터를 활용하여 환자군 별 병합 요법에 따른 유효성을 평가하였습니다.

아래 그림처럼 데이터 전처리 과정에서 수집된 혈압 측정 데이터의 4분면의 이질적인 분포가 확인됨에 따라 임상 연구자의 자문을 통해 데이터의 신뢰성을 검증하고, 관심 약물에 효과를 보이는 특정 환자군(여성, 나이 50세 이상)을 선정함으로써 시판 후 의약품에 대한 신규 포지셔닝 전략에 필요한 정보를 제공할 수 있었습니다.

Figure 8. 관심 약물 복용 시작 시점에서의 혈압 분포 예시

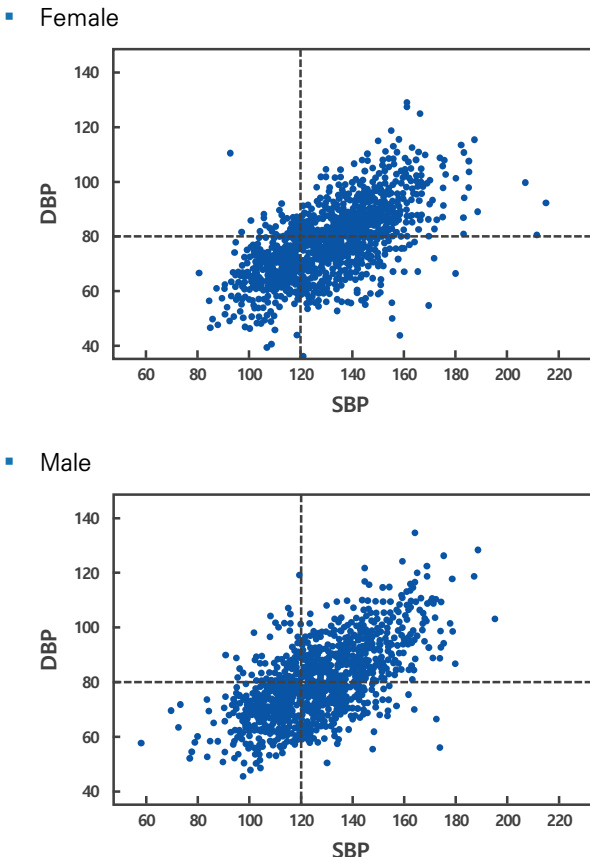
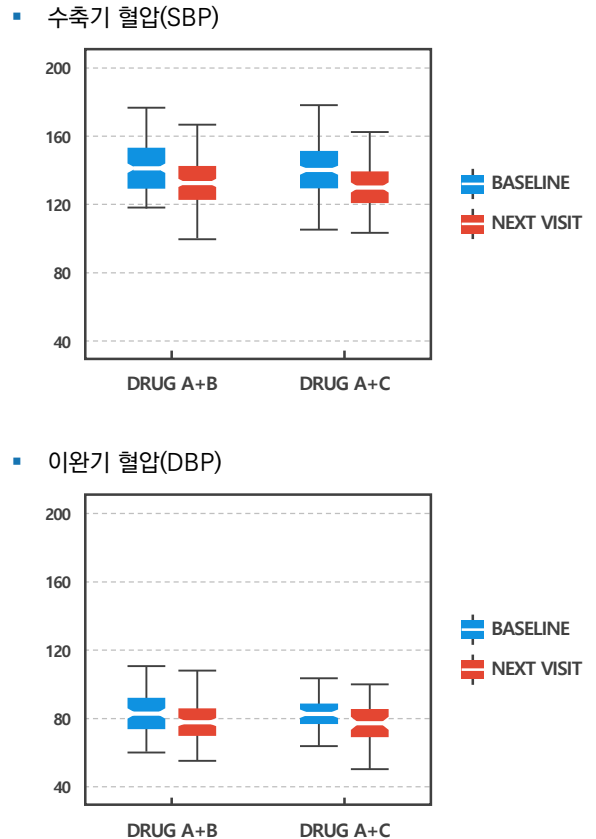


Figure 9. 관심 약물 처방 전후 혈압 변화 예시



Case

④ 요인 분석(Factor Analysis)

본 연구는 간이식술을 시행한 환자를 대상으로 HCC (Hepatocellular Carcinoma) 재발에 영향을 미치는 요인을 분석하는 연구입니다.



세브란스 병원 내 기록된 10년 동안의 의무기록데이터와 간이식 등록 데이터베이스(Registry DB)를 활용하여 간 이식 이후 HCC 재발에 영향을 미치는 요인을 평가하였습니다.

본 연구에서 간 이식술 판정 검사에서 AFP(α -fetoprotein) 수치가 높을수록 이식 후 1년 이내 조기 HCC 재발 위험도가 높아진다는 통계적 유의성을 확인하였습니다. 이를 통해 AFP가 조기 HCC 재발 요인이 될 수 있음을 후향적 연구를 통해서 재확인하였습니다.

⑤ 그 외 진행 중인 연구

연구개발자문센터는 현재 다음과 같은 연구를 진행하고 있습니다.



- 의료기기 대상 PMS 연구

의료기기를 대상으로 한 PMS(Post Market Surveillance) 연구로서 진료 및 수술 기록과 이상반응 데이터를 후향적으로 수집하여 안전성을 평가하고자 합니다.

- 등록 데이터베이스(Registry DB) 구축

종양(상피성 난소암, 비소세포폐암)과 만성질환(당뇨병, 이상지질혈증, 고혈압)에 대해 등록데이터베이스를 구축하고 있습니다. 등록데이터베이스 구축에 대한 내용은 Part II 에서 자세히 확인하실 수 있습니다.



Check Point

- ✓ 후향적 임상연구의 목적 및 설계에 따라 EMR 빅데이터에서 추출된 자료 항목에 대해 조작적 정의가 요구될 수 있습니다.
- ✓ 비교군 연구에서 인과관계에 영향을 주는 변수들을 통제하기 위해서 자료를 처리하고 분석하는 과정에서 ‘층화’, ‘매칭’, ‘보정’ 등 Bias 최소화 기법을 활용해야 합니다.
- ✓ 연구 목적에 맞는 코호트 정의, 추출된 분석용 데이터의 품질관리(Quality Control, QC)를 통해, 임상적 타당성을 확인하고 결과의 재현성을 검증하는 것이 중요합니다.

PART II

등록 데이터베이스 연구

레지스트리 기반 임상연구 시대를 대비하여
의료빅데이터로 데이터베이스를 구축하고 있습니다



등록 데이터베이스 구축 및 연구 사례

등록 데이터베이스(Registry DB)란

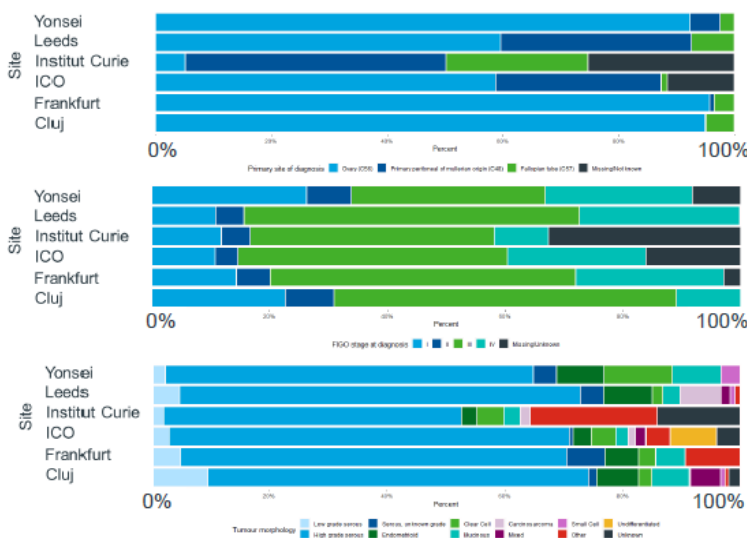
환자 중심의 치료제 개발을 목적으로 특정 질병 혹은 제품의 관심 데이터를 다양한 원천자료로부터 수집하여 일관된 형태로 보관한 자료 집합체입니다.⁷⁾

환자 1명 당 생성되는 자료는 약 90 여개 이상의 데이터 항목으로 구성됩니다. 관심 질병 기준으로 최소 수 백명 부터 최대 수 만명까지의 자료가 데이터베이스에 축적되며, 해당 자료에는 인구학적정보, 신체계측 정보와 같은 환자 기본 정보 부터 관심 질병의 동반 질환, 진단 검사, 병리 결과, 치료 및 반응, 임상 관찰, 임상 경과 추적 등이 포함됩니다.

데이터베이스에 구축된 질병 특성 및 진료 정보는 치료제 연구개발 과정에서 유용하게 활용됩니다. 목표하는 대상 환자군을 파악하고, 치료 과정에서 최적의 제품 포지셔닝 등에 도움을 주기 때문입니다. 이와 같이 등록 데이터베이스는 임상개발 전략수립에 중요하고 다차원적인 분석 정보를 제공할 수 있다는 장점이 있습니다.

최근에는 등록데이터베이스를 의사와 환자간의 연구용 소통 채널로도 인식하고 있습니다. 진료 과정을 통해 환자의 데이터가 남겨지고, 임상적 판단 자료들이 더해지면서 환자 중심의 연구개발이 가능해지기 때문입니다.

Figure 10. 등록 데이터베이스(Registry DB) 구축 연구 사례⁹⁾



Reference:
Sue Cheeseman, et al. (2020)

등록 데이터베이스 구축 현황

중양(상피성 난소암, 비소세포폐암)과 비중양(당뇨병, 이상지질혈증 등)으로 구분하여 등록 데이터베이스를 구축하고 있습니다.

중양분야

상피성 난소암의 경우 연구 목적에 따라 선별된 약 850명 코호트를 대상으로 약 90여개 변수의 데이터가 수집됩니다. 원천 자료에는 EMR 빅데이터 외에 암 등록 자료도 사용됩니다. 등록 데이터베이스는 환자, 치료, 임상관찰 정보 외에 총화, 예측, 예후 평가에 필요한 바이오마커, 조직학적 병리 결과, 생존 여부 등의 자료도 포함됩니다. 치료법에는 수술, 항암요법, 방사선치료가 해당되는데, 이 주제만으로도 약 40만 건 이상의 해당 자료가 DB에 저장됩니다.

비소세포폐암도 연구 목적에 따라 선별된 약 5,000명의 코호트를 대상으로, 유사한 형태로 자료가 수집됩니다.

비중양분야

당뇨병과 이상지질혈증의 경우에도 다양한 후향적 연구를 체계적으로 지원하기 위해 자료 수집 범위 및 항목을 선정하여 구축하고 있습니다. 만성질환은 동반질환 파악 및 주기적인 혈액학적 검사의 추적 관찰이 중요합니다. 이를 위해서 주로 EMR 빅데이터를 원천데이터로 사용하고 있습니다. 다만, 진료용 EMR 빅데이터를 2차적 목적인 연구용으로 가공해야 하므로 여러 주의를 요구합니다. 특히 연구용 진료 데이터 구축 시 자료입력 과정, 각종 서식지 간 연계성 등을 이해하기 위해서는 임상 의사의 자문이 반드시 병행되어야 합니다.



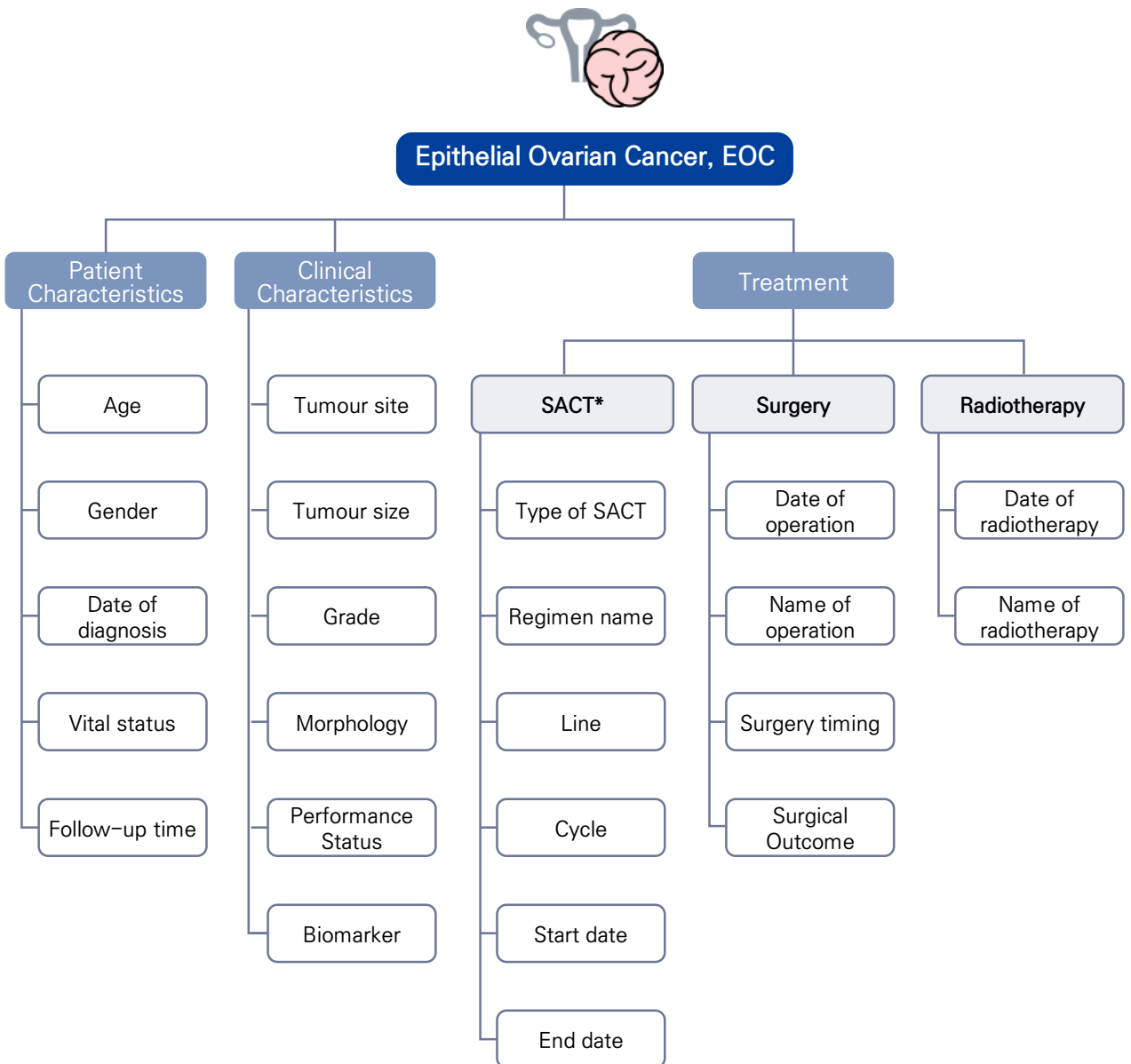
Check Point

- ✓ DB 구축 시, EMR 등 원천 자료의 속성 및 자료 생성 과정에 대한 이해와 숙련과정이 필수적입니다.
- ✓ DB 구축 과정에서는 정형화 자료 외에 반정형화, 비정형화 자료도 활용하는 경우가 있어 반드시 기획 과정에서부터 기술적 솔루션 검토가 병행되어야 합니다.
- ✓ 데이터 수집 및 처리 (Data Collection and Processing) 뿐만 아니라 데이터 관리(Data Management)에도 상당한 시간이 소요됩니다.



- ✓ 구축된 DB에 연동된 애널리틱스(Analytics) 전용 프로그램이 없다면, 분석용 데이터세트를 직접 만들어서 통계분석 프로그램(예, R 등)을 통해 분석할 수 있습니다.

Figure 11. 등록 데이터베이스 구축에 필요한 수집 항목 예시(상피성 난소암)



* SACT: Systemic Anti-Cancer Therapy

PART III

빅데이터 분석 기술

효율적인 RWE 생성 연구 수행을 위한
분석시스템 고도화 연구를 지속하고 있습니다



빅데이터 분석 기술 개발 및 연구 사례

의료빅데이터 분석 기술의 연구배경

최근 들어 신약재창출 후보 물질 탐구를 위한 데이터 간의 융합 활용이 관심을 받고 있습니다. 관심 물질의 화학구조, 세포 독성 및 효능 정보, 유전학 정보, 약물의 기전 연구 등의 기존 바이오데이터에 병원의 진료데이터를 연계하여 활용하려는 시도가 증가하고 있습니다. 즉, 연구 대상군에 대한 환자 기본 정보 외에도 진료 과정에서 수집되는 다양한 유효성 및 안전성 평가 자료가 분석에 적용됩니다.

이러한 추세는 환자가 원하는 제품을 발굴, 기획하는 단계에서 질병의 특성을 먼저 충분히 이해하고 특정 환자군의 미충족 의료 수요를 해결하는 관점에서 향후 타겟 물질이 발굴될 것임을 시사해주고 있습니다.

다시 말해, 의료빅데이터로부터 생성된 실사용 증거(RWE)를 통해 질병의 특성을 분석하고, 치료제 제품기획에 필요한 근거가 마련될 것입니다.

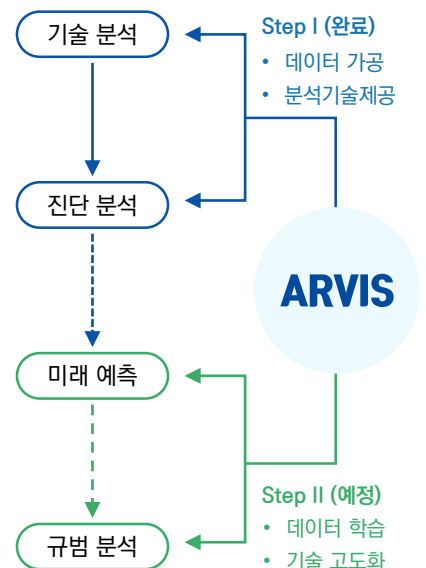
의료빅데이터 분석 모델과 솔루션

환자 진료 과정에서 생성된 빅데이터를 활용하여 유효성 및 안전성 평가 자료를 분석합니다. 이 과정에서 빅데이터 조회 및 분석에 특화된 알고리즘 개발이 필요합니다.

아비스(ARVIS, Augmented Research naVigation System)는 이러한 기능이 탑재된 조회·분석 시스템으로 환자 선별 기준에 따라 관심 코호트를 정밀하게 추출하고, 분석에 필요한 변수들을 원하는 데이터 세트 형태로 자유롭게 설정하여 시각화 정보 및 통계분석 결과를 제공합니다.

빅데이터 분석 모델은 크게 4단계로 구성됩니다. 기술(Descriptive), 진단(Diagnostic), 예측(Predictive), 그리고 규범(Prescriptive) 분석들로 구분됩니다.¹⁰⁾ 첫 두 단계 과정에서 요구되는 분석 기술을 아비스(ARVIS)를 통해서 완성했습니다. 향후 어떤 일들이 일어날지 예상하고 체계적으로 예측하기 위해서는 질 좋은 학습용 데이터세트가 필요합니다.

Figure 12. 빅데이터 분석 모델

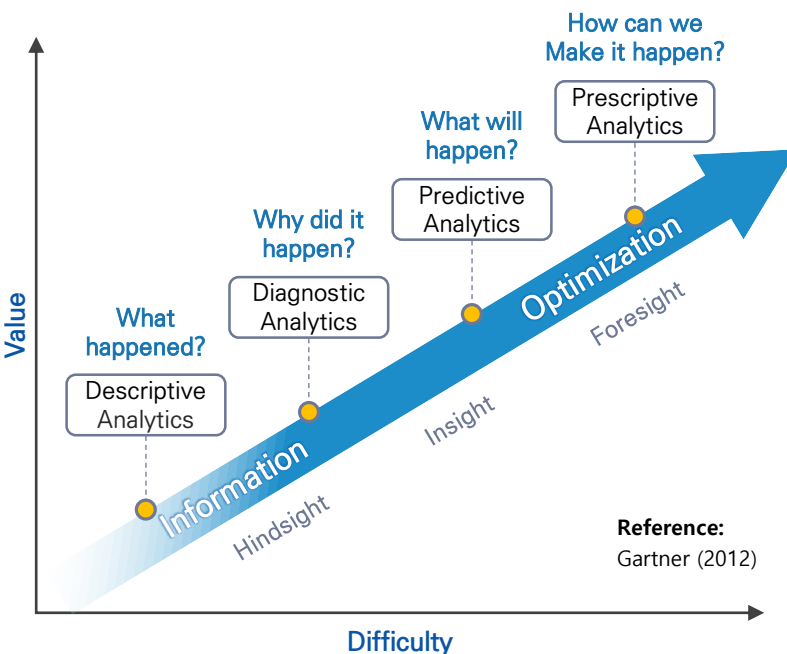


따라서, 아비스(ARVIS)는 다양한 임상조건에서 정밀 추출된 질 좋은 학습용 데이터를 통해 나머지 두 단계의 고도화 분석을 지원할 수 있습니다. 또한 필요 시 데이터 탐색 과정에서 유용한 자료를 생성하기 위한 데이터 전처리 가공도 가능합니다.

초기 두 단계의 분석 수행 과정을 통해서 쌓은 지식 및 경험은 다양한 임상 예측 모델을 개발하는데도 도움이 됩니다. 현재 연구개발자문센터는 'CDM기반의 지능형 진단·처방 조회 시스템 개발 및 실증' 정부 R&D 사업에 참여하고 있습니다. 이를 통해 대리지표 및 임상 결과를 분석하는 과정에서 다양한 예측 분석을 수행할 예정입니다.



Figure 13. Analytics Ascendancy Model



Check Point

- ✓ 의료빅데이터 분석 경험은 처음부터 순차적으로 진행되어야 합니다. 전향적인 예측모델을 연구하기 위해서 후향적으로 축적된 데이터의 속성 및 생성 과정을 이해해야 합니다.
- ✓ 인공지능 학습용 데이터의 질과 양은 예측 모델 적합성 평가에 매우 중요한 요소입니다. 다양한 조회 조건에 맞게 데이터들 정밀하게 추출하고 정제하는 기술이 핵심입니다.

참고 문헌

1. Ahlqvist, Emma, et al. "Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables." *The Lancet Diabetes & Endocrinology* 6.5 (2018): 361–369.
2. Infosys Ltd. *Empowering the patient journey in clinical trials*. (2021).
3. FDA. *Enrichment strategies for clinical trials to support approval of human drugs and biological products*. (2019).
4. Vreman, Rick A., et al. "Unmet medical need: an introduction to definitions and stakeholder perceptions." *Value in Health* 22.11 (2019): 1275–1282.
5. FDA. *CDER Patient-focused drug development*. (2020).
6. Cartwright, M. E., et al. "Proof of concept: a PhRMA position paper with recommendations for best practice." *Clinical Pharmacology & Therapeutics* 87.3 (2010): 278–285.
7. ARHQ. *Registries for evaluating patient outcomes: A User's Guide: 4th Edition*. (2020).
8. Wang, I-Kuan, et al. "Increased risk of new-onset type 2 diabetes in people with chronic kidney disease." *International Urology and Nephrology* 51.4 (2019): 707–712.
9. Sue Cheeseman, et al. *Real world retrospective study of patients with epithelial Ovarian cancer: an international collaboration, ESGO 2020*
10. Lim, Ee-Peng, Hsinchun Chen, and Guoqing Chen. "Business intelligence and analytics: Research directions." *ACM Transactions on Management Information Systems (TMIS)* 3.4 (2013): 1–10.

문의사항

본 사례집에 대해 의견이 있으시거나 혹은 연구개발 협력에 대한 관심이 있으신 경우
연구개발자문센터 홈페이지 내 문의하기 또는 이메일 주소로 연락주시길 바랍니다.



홈페이지

<http://alynd.yuhs.or.kr>

이메일

saltplus@yuhs.ac

전화

02-2228-0223(4)

연세의료원 연구개발자문센터

편집장 김동규
참여자 김동규, 도해원, 엄희상, 임수빈(가나다 순)
발행인 심재용
발행처 연세의료원 의과학연구처 연구개발자문센터
디자인 율공간

