Epidemiology Information Synthesis for Focal Segmental Glomerulosclerosis (FSGS): An Innovative Approach Using Human-in-the-Loop AI

MSR182

Vatsal Chhaya, Shaurya Deep Bajwa*, Kapil Khambholja, Anoop Ambika
Catalyst Clinical Research, Wilmington, NC, USA

Presented at ISPOR Europe 2024: November 17-20, 2024; Barcelona, Spain

INTRODUCTION

- **Challenges in Rare Disease Data**: Collecting epidemiological data for rare diseases like Focal Segmental Glomerulosclerosis (FSGS) is hampered by limited and inconsistent sources.
- **Data Complexity**: Variability in case reports, studies, and registries complicates the synthesis of reliable epidemiological insights.
- **AI-powered Solution**: AI models can streamline and standardize data integration from diverse epidemiological sources.
- **Tool Development**: This poster presents the conceptual framework for an AI tool to synthesize epidemiological data for FSGS.
- **Future Impact**: The AI solution aims to enhance rare disease data analysis, improving decision-making and patient outcomes.

METHODS

A semi-automated, AI-powered evidence integration synthesis workflow for FSGS was developed by a multidisciplinary taskforce. This team included AI/ML experts, UI/UX designers, software engineers, senior health economics and outcomes research (HEOR) analysts, and consultant epidemiologists. The proposed roadmap comprises six key steps:

1. **Design Thinking Workshop**: Conducted to align the team on objectives and strategies.
2. **Input Data Retrieval**: Gathering relevant data from various sources.
3. **FSGS-specific Named Entity Recognition (NER) Corpus Development**: Creating a specialized corpus for accurate entity recognition.
4. **Training Dataset Annotation**: Utilizing Natural Language Processing (NLP) techniques such as entity linking and relationship extraction to annotate the dataset.
5. **Integration with a Knowledge Graph**: Connecting annotated data with a structured knowledge graph for enhanced insights.
6. **Product Testing**: Implementing a predefined User Acceptability Testing (UAT) plan, executed by technical and HEOR teams to ensure the product meets user requirements.

This structured approach ensured a comprehensive and efficient workflow for synthesizing evidence in FSGS research.

Training Dataset Annotation Using NLP Techniques

Training dataset annotation involves tagging and organizing data using NLP techniques to teach our AI models how to understand complex medical texts.

Techniques Used

Entity Linking

- **What's this?** Connecting identified entities to established databases or knowledge bases.
- **Why?** Ensures that terms like "glomerulosclerosis" are linked to their correct medical definitions and synonyms, enhancing the AI's contextual understanding.

Relationship Extraction

- **What's this?** Identifying and categorizing the relationships between entities.
- **Why?** Helps the AI understand how terms are connected, such as linking "treatment" with "outcome" or "symptoms" with "diagnosis."

Benefits

- **Improved Accuracy**: Trains AI to recognize and correctly interpret complex relationships and entities in FSGS data.
- **Enhanced Insights**: Provides clearer, actionable insights from medical texts, leading to better-informed research and clinical decisions.

Process Highlights

1. **Annotation Tool Setup**: Deploy NLP tools to assist with tagging and linking.
2. **Expert Annotations**: Medical professionals review and tag data with high precision.
3. **Model Training**: Use annotated data to train and refine AI models, ensuring they learn from high-quality examples.

Quality Metrics

- **High Precision** in FSGS Metric Identification
- **Faster Semantic Query Response** with Knowledge Graph
- **Strong Accuracy** in Literature Curation
- **Significant Time Savings** in Report Generation

Multidisciplinary Team

Multidisciplinary taskforce, including AI/ML experts, UI/UX designers, software engineers, senior HEOR analysts, and consultant epidemiologists.

Input Data Retrieval

- **Scientific Literature Databases**: Extract peer-reviewed articles from PubMed, Google Scholar, and Web of Science focused on FSGS research.
- **Clinical Trial Registries**: Gather data from ClinicalTrials.gov and WHO International Clinical Trials Registry Platform (ICTRP) for insights on ongoing and completed trials related to FSGS.
- **Health Organization Reports**: Incorporate guidelines and reports from the National Kidney Foundation and KDIGO to reflect best practices and consensus statements.
- **Patient Registries**: Leverage real-world data from patient registries to capture disease progression in FSGS.

FSGS-specific Named Entity Recognition (NER) Corpus Development

The FSGS NER Corpus is a specialized dataset designed to help AI systems accurately identify and understand key entities related to FSGS from vast medical texts.

Why is it essential?

Developing this corpus allows our AI models to:

- **Spot Critical Terms**: From disease names to treatment protocols and patient symptoms, ensuring every relevant entity is recognized.
- **Enhance Precision**: Improve the accuracy of data extraction and information retrieval specific to FSGS.
- **Facilitate Research**: Accelerate the discovery of patterns and relationships in medical literature, leading to better understanding and potential breakthroughs.

How does it work?

1. **Data Collection**: Gather a wide range of medical texts, research papers, and clinical reports on FSGS.
2. **Entity Annotation**: Experts mark and categorize entities such as disease names, drug names, and patient conditions.
3. **Corpus Creation**: Compile these annotated entities into a structured dataset for training AI models.

Integration with Knowledge Graph and User Acceptability Testing (UAT)

Enhanced Data Connectivity: Knowledge graph integration allows linking of diverse data points (clinical trials, patient registries, genetic studies) to uncover hidden patterns, improving evidence synthesis accuracy.

Streamlined Decision Support: By organizing data hierarchically, the knowledge graph aids users in quickly identifying relevant insights.

Personalized User Experience: UAT ensures that the interface is intuitive and tailored to clinicians, researchers, and stakeholders.

Improved Clinical Relevance: Testing the tool with real users helps refine features to prioritize actionable FSGS-related evidence.

Figure 1.

Proposed Flow Chart: AI-based Epidemiology Evidence Synthesis PoC

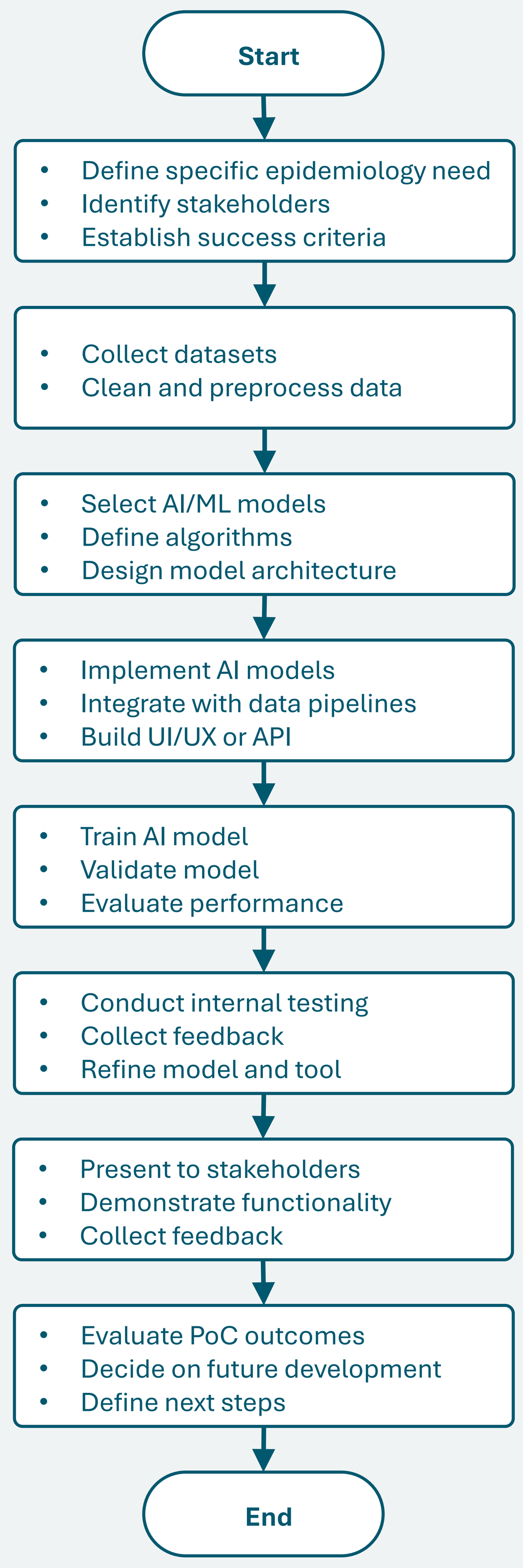

Figure 2.

FSGS PoC – Design Thinking Workshop Elements

Design Thinking Stage	Description	Application to FSGS Epidemiology Information Synthesis (PoC)
1. Empathize	Understand user needs	- Identify challenges in accessing and interpreting FSGS data. - Focus on integration issues and reporting inconsistencies.
2. Define	Narrow the problem scope	- Synthesize FSGS data from limited sources. - Assess feasibility using case reports and select registries.
3. Ideate	Generate potential solutions	- Brainstorm AI solutions for data integration. - Target essential features for standardization.
4. Prototype	Build a low-fidelity version	- Create a basic AI prototype for data synthesis. - Focus on core data extraction functionalities.
5. Test	Conduct initial user tests	- Evaluate the tool with a small group of epidemiologists. - Assess data integration effectiveness.
6. Refine	Gather feedback for improvement	- Use feedback to enhance the tool. - Validate its FSGS data synthesis capabilities.

CONCLUSION

- **Integrated Expertise:** This hybrid workflow synergizes nephrologists, epidemiologists, AI/ML scientists, and HEOR specialists to refine evidence synthesis for rare nephropathies like Focal Segmental Glomerulosclerosis (FSGS). This ensures rigorous, domain-specific insights essential for addressing FSGS's unique pathophysiology and clinical progression.
- **Real-world Relevance:** Utilizing robust training datasets informed by glomerular disease registries, patient-reported outcomes, and tailored ontologies pertinent to FSGS phenotypes anchors findings in real-world, patient-centered data reflective of variability in glomerular injury and progression.
- **Future Implications:** This methodology informs the advancement of novel epidemiological synthesis tools, optimizing their utility in guiding clinical decision-making and improving health outcomes in FSGS and other rare glomerulopathies.

CONTACT INFORMATION

Kapil Khambholja, Ph.D.
Executive Director, Head of Medical Writing and Product Strategy Lead
Catalyst Clinical Research
Phone: +91-77029 49998 | Email: kapil.khambholja@catalystcr.com
www.CatalystCR.com

SCAN HERE
TO LEARN MORE