Project Background:
Rare Disease Algorithm Overview

MAT-US-2107797 v2.0 - P
Expiration Date: 10/03/2025

Opportunity to further patient diagnosis using EHR data

EHR OPPORTUNITIES

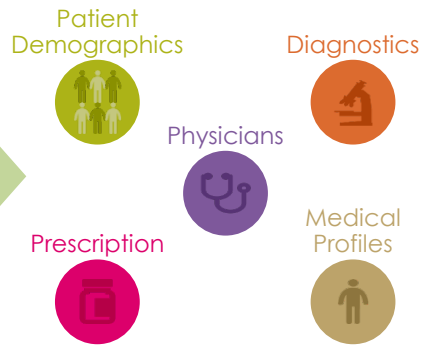
LSD patients are not common, as a consequence they may be missed.

- EHR platforms frequently contain the information needed for physicians to identify rare disease patients
- EHR screening program opportunity
  - Extend reach
  - Early intervention through retrospective searches or Point of Care

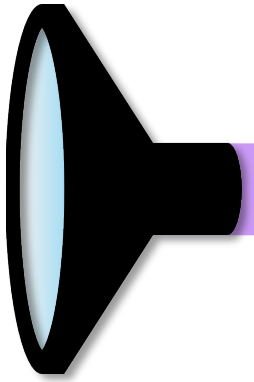
HCP EDUCATION

Educate HCP & Institution on the utilization of EHR algorithm

EHR CONTENT



1 ALGORITHM IMPLEMENTATION



Disease specific algorithm is applied to the EHR

2 IDENTIFICATION OF POTENTIAL PATIENT LIST

EHR patients ranked by likelihood. HCP reviews charts

3 SCREENING/ TESTING

HCP offers patients diagnostic testing

4 DIAGNOSIS

Patient is successfully diagnosed

Activity

Advanced Analytics and/or Site Activity → HCP Activity → HCP & Lab Activity → HCP Activity

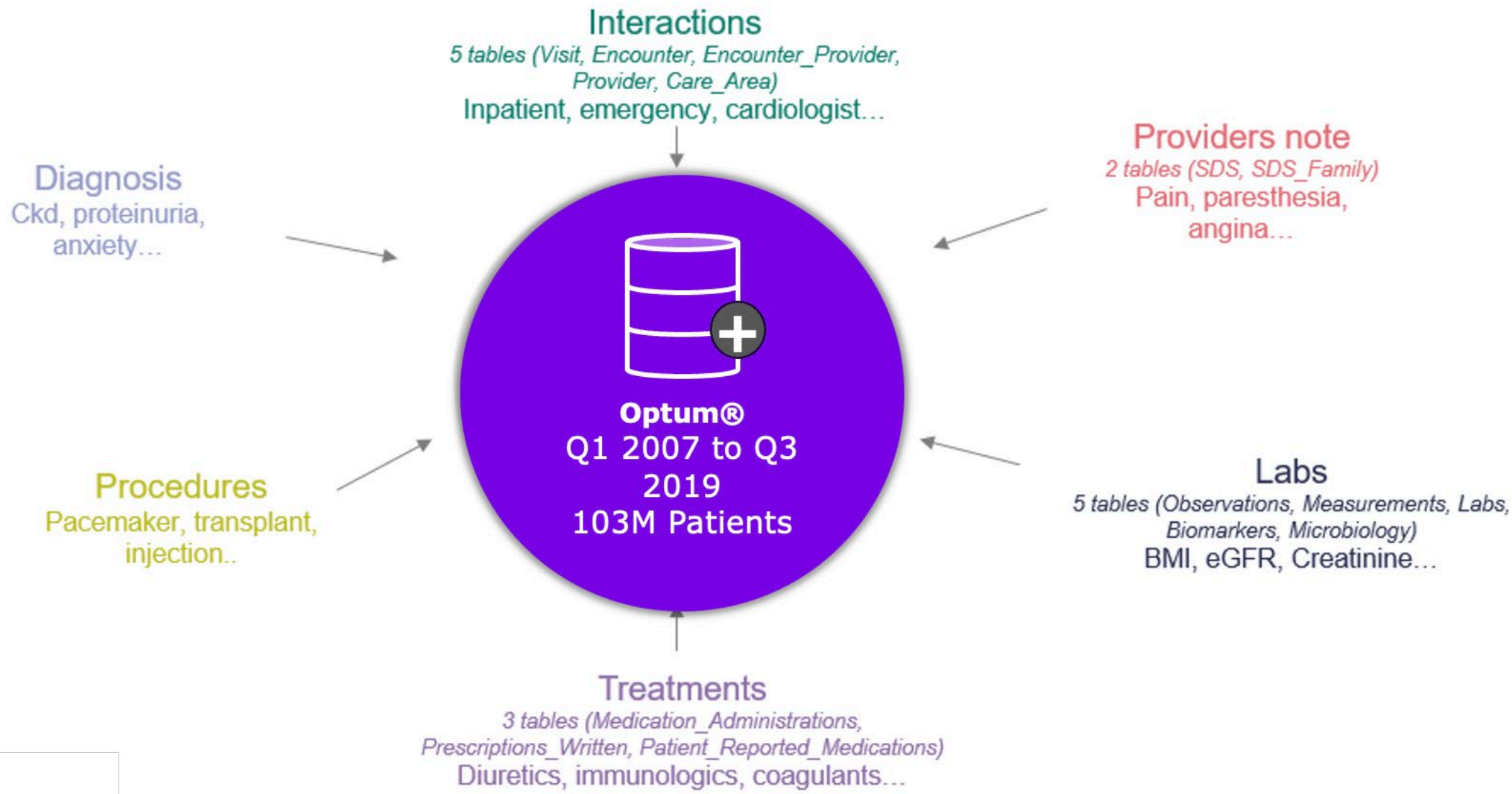
Continuous Observation & Learning
There will be no identifiable patient data obtained during this project

Developing the Rare Disease Algorithm

A behind-the-scenes look

sanofi

What's in the De-identified Database?



There will be no identifiable patient data obtained during this project

Algorithm (Model) Development – 3 Stages

1 DEFINE

What is a Gaucher Patient?

	Number of patients	Step	Description
Inclusion criteria	N = 829	Diagnosed or Treated Gaucher	*At least 2 diagnosis of ICD10 E75.22* <u>OR</u> *At least 1 medication by NDC & HCPC codes*
	- 14 patients		Exclusion: - Death prior to index (-2 patients) - Missing age or gender (-3 patients) - Misdiagnosis (-9 patients)
Exclusion criteria	N = 815	Representative Gaucher	

715 → TRAINING

100 → TESTING

Gaucher patients

N=715

N=656

At least 1 year of coverage

TRAINING DATASET

Random control patients

N=328,000 (ratio 1:500)

Gaucher patients

N=656

2 TRAIN

How does a typical Gaucher patient present longitudinally within EHR Data?

LITERATURE FEATURES	• Symptom from literature • Enriched with SDS, labs, procedures
DATA DRIVEN FEATURES	• Features the model picked up as differentiators of Gaucher vs controls
DEMOGRAPHICS	• Region, race, gender, age
HEALTHCARE INTERACTIONS	• Provider specialty • Visits • Encounters

▲ Age model ▲ Prevalence model

3 TEST

Based on features identified during the training, how well do they identify GD patients?

Random Gaucher patients

N=100

TESTING DATASET

Patients

N=1,000,000 (ratio 1:10,000)

Gaucher patients

N=100

MAT-US-2107797 v2.0 - P
Expiration Date: 10/03/2025

2 TRAIN

How does a typical Gaucher patient present longitudinally within EHR Data?

Literature Features	• Symptom from literature • Enriched with SDS, labs, procedures
Data driven Features	• Features the model picked up as differentiators of Gaucher vs controls
Demographics	• Region, race, gender, age
Healthcare Interactions	• Provider specialty • Visits • Encounters

Nervous system
- Ataxia
- Extrapyramidal disorder
- Bradykinesia
- Cranial nerve disorders
- Developmental retardation
- Dysphagia
- Gaze palsy
- Hearing impairment
- Hemiplegia/paresis
- Hydrocephalus
- Hypotonia
- Laryngeal spasms
- Myoclonic seizure
- Nerve root compression
- Oculomotor apraxia
- Optic kinetic nystagmus
- Parkinsonism
- Strabismus
- Tonic-clonic seizure
- Tremor

Development
- Delayed puberty
- Delayed skeletal maturation
- Developmental regression
- Growth delay
- Short stature

Bones
- Arthralgia
- Joint stiffness (e.g. arthrogryposis)
- Avascular necrosis
- Delayed skeletal maturation
- Erlenmeyer flask deformity
- Joint dislocation
- Kyphosis
- Osteoarthritis
- Osteopenia
- Osteolysis
- Osteoporosis
- Spinal deformation
- Bone pain
- Pathological fracture

Heart
- Aortic and mitral valve calcification

Hepatic
- Cirrhosis
- Hepatic fibrosis
- Portal hypertension
- Hepatitis
- Elevated CRP
- Elevated ferritin

Renal
- Hematuria
- Proteinuria

Respiratory
- Pulmonary fibrosis
- Pulmonary hypertension
- Respiratory failure
- Interstitial pulmonary abnormalities

Blood
- Pancytopenia
- Thrombocytopenia
- Gingival bleeding

Organomegaly
- Hepatomegaly
- Splenomegaly
- Ventriculomegaly

Ophthalmologic
- Corneal deposit & opacity
- Retinopathy (non-diabetic)

Malignancy
- Multiple myeloma
- Liver neoplasm
- Non-Hodgkin lymphoma
- Malignant melanoma
- Pancreatic cancer

Psychiatric
- Dementia (non senile)
- Depression

Perinatal
- Hydrops fetalis
- Ichthyosis

Anemia
- Anemia

Immunology
- Polyclonal gammopathy
- Monoclonal gammopathy

General signs
- Abdominal pain
- Fatigue
- Fever

The Two Models – How do they differ?

2 TRAIN

How does a typical Gaucher patient present longitudinally within EHR Data?

LITERATURE FEATURES	• Symptom from literature • Enriched with SDS, labs, procedures
DATA DRIVEN FEATURES	• Features the model picked up as differentiators of Gaucher vs controls
DEMOGRAPHICS	• Region, race, gender, age
HEALTHCARE INTERACTIONS	• Provider specialty • Visits • Encounters

Age Model

- Symptoms are derived with the age of first occurrence for each patient
- Favors features that have different ages of onset between the Gaucher and control patients
- Favors earlier onset
- Favors younger patients

Prevalence Model

- Symptoms are flagged with presence/absence
- Model favors features that have different prevalence between the Gaucher and control patients
- Favors accumulation of comorbidities
- Favors older patients

MAT-US-2107797 v2.0 - P
Expiration Date: 10/03/2025

Testing the Models

2 TEST

Based on features identified during the training,
how well do they identify GD patients?

Random Gaucher patients

N=100

TESTING DATASET

Patients

N=1,000,000 (ratio 1:10,000)

Gaucher patients

N=100

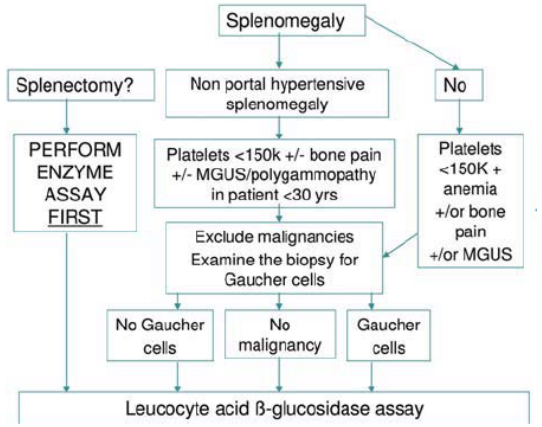
MAT-US-2107797 v2.0 - P
Expiration Date: 10/03/2025

Machine Learning models vs Each Other vs Rare Diagnostic Algorithm

❖ **Choice of model(s) and threshold should be assessed for each database and context of the application**

- There is flexibility in the application of the model that is <u>situation dependent</u>. We chose a threshold of **≥0.95 (95% probability)**.
- Consider what types of patients are *most likely* undiagnosed
- Testing capacities

	Age model	**Prevalence model**	**Rule-based filter (Cappellini Protocol)****
Based on threshold	• **1,073 patients with score ≥ 0.95** • **26 GD known patients (2%)**	• **1,793 patients with score ≥ 0.95** • **25 known GD patients (1%)**	• **20,743 patients have the information required by the rule-based filter** • **28 known GD patients (0.1%)**
How many to test to get the same number (28 GD pts) as identified from the Rule-Based Filter	• **1,204 patients to get to…** • **28 known GD patients (2%)** (≈17 times less than the rule-based filter).	• **2,862 patients to get to…** • **28 known GD patients (10%)** (≈7 times less than the rule-based filter).	
Testing as many as the Rule-Based Filter identified (20,743)	• **The top 20,743 patients** • **47 known GD patients (0.2%)**	• **The top 20,743 patients** • **47 known GD patients found (0.2%)**	



Both models identify more known GD patients in the dataset,
in addition to having a better known-GD-to-controls ratio, than the rule-based filter

**Adapted for RWD from published diagnostic algorithm/Cappellini Protocol (Mistry et al, 2011)

MAT-US-2107797 v2.0 - P
Expiration Date: 10/03/2025

Together or Separately?

❖ **Choice of model(s) and threshold should be assessed for each database and context of the application**

- There is flexibility in the application of the model that is situation dependent. We chose a threshold of **≥0.95 (95% probability)**.
- Consider what types of patients are *most likely* undiagnosed
- Testing capacities

Age model "OR" **Prevalence model**

- **2467 patients highly-ranked by either model**
- **34 known GD patients (1.4%)**

Age model "AND" **Prevalence model**

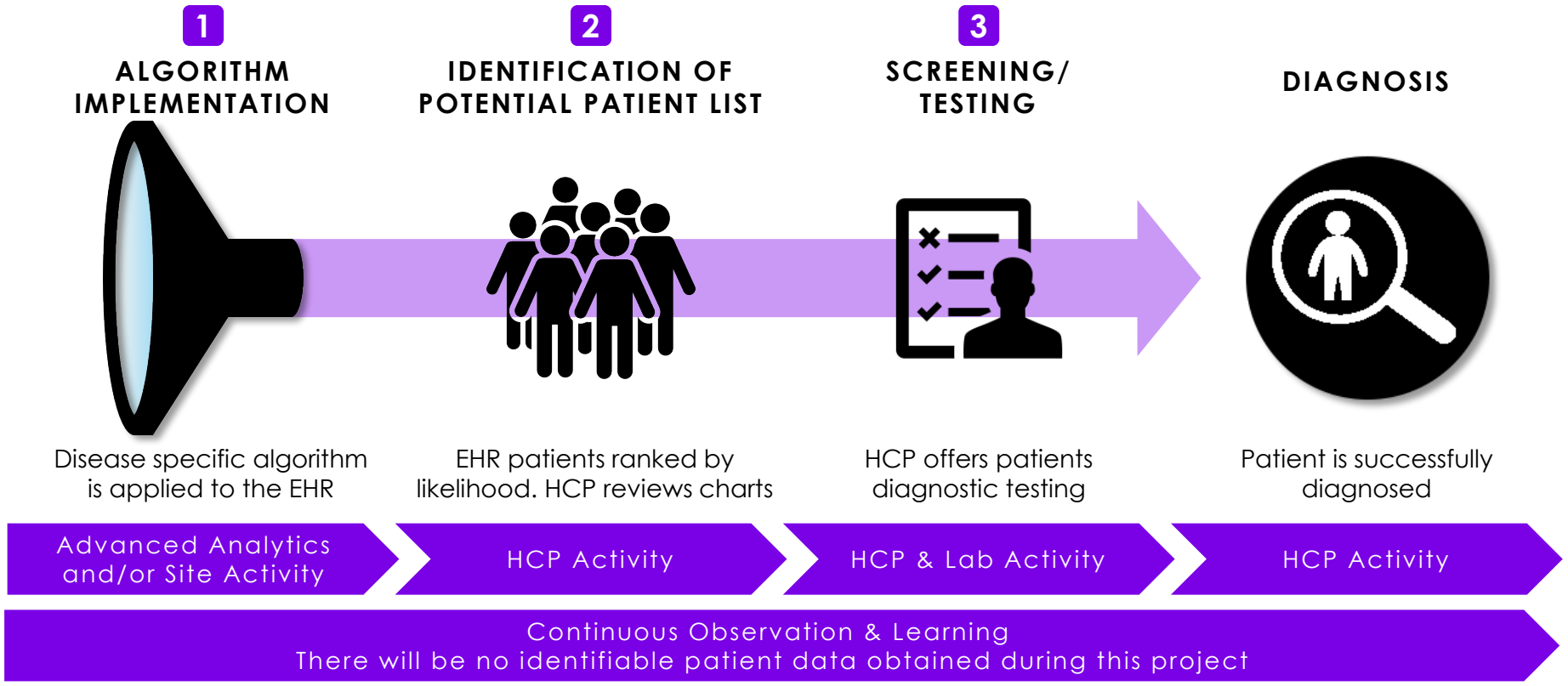
- **319 patients highly-ranked by both models**
- **17 known GD patients (5.3%)**

MAT-US-2107797 v2.0 - P
Expiration Date: 10/03/2025

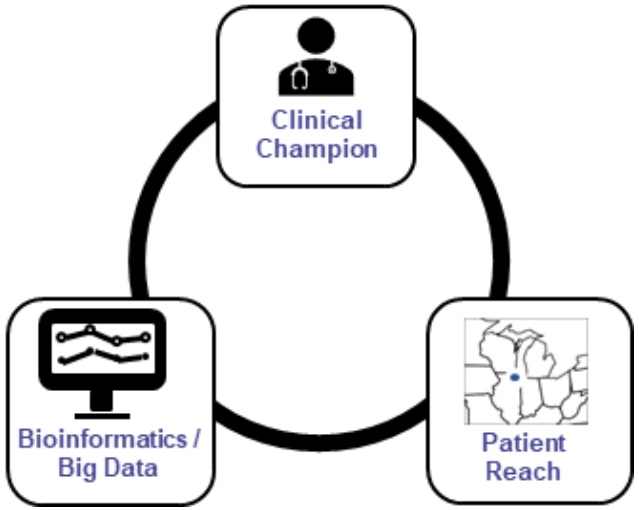
What's Next?

Implementation and Evaluation of a Rare Disease Algorithm to Identify Persons at Risk of Gaucher Disease Using Data from Electronic Health Records (EHRs) in the United States

Clinical Trial Design



Study Objectives



Primary

- Estimate the positive predictive value (PPV), or diagnostic yield, of the Rare Disease Algorithm (RDA) for Gaucher Disease (GD) (i.e., the proportion of persons with GD out of all persons highly ranked by the RDA)
- Identify the magnitude of "unrecognized" GD (i.e., previously undiagnosed persons highly ranked by RDA who subsequently test positive for GD, out of all persons with GD)

Secondary

- Describe characteristics (e.g., demographics, alternative and concomitant diagnoses, treatments, and patterns of provider visits) of all persons highly ranked by the RDA, by subgroups:
  - Persons previously diagnosed with GD
  - Persons previously undiagnosed with GD who subsequently test positive for GD
  - Persons previously undiagnosed with GD who subsequently test negative for GD

Exploratory

- Describe characteristics (e.g., demographics, alternative and concomitant diagnoses, treatments, and patterns of provider visits) of all persons highly ranked by the RDA, by subgroups:) of persons highly ranked by the RDA who are negative in confirmatory testing for GD but positive for ASMD
- Determine the diagnostic yield for ASMD among patients highly ranked by the RDA
- Explore the relationship of ASMD biomarker lysosphingomyelin in the diagnostic process for ASMD
- Describe the time and resources on a per-role basis (e.g., physician, study coordinator, bioinformatics) needed for the following:
  - Identifying potential predictors (demographics, diagnoses, treatments, lab values and visits) of those persons highly ranked by the RDA not previously diagnosed with GD who are found through confirmatory testing to be GD-positive (if sample size allows)
  - Describe IDN characteristics including demographics, academic vs. non-academic, EHR type, physician specialties, etc.

Next Steps:
Deploying RDA in a Compliant Manner

- Share algorithm with US Health Systems
- Access via secured URL or API
- Hands off approach
- Provide patient readout that fosters transparency with implementing the algorithm
- Gather feedback on implementation and patients identified with Gaucher
- Information **will not be shared** with Commercial colleagues

MAT-US-2107797 v2.0 - P
Expiration Date: 10/03/2025

Questions

Clinicaltrial.gov

Model Development

Patrick Pavlick
C: 908.842.4390
E: patrick.pavlick@sanofi.com