Week 1 — Enrollment Overview: Weekly Report

Project: Clinical Trial Patient Recruitment & Adherence Monitoring
Date: September 15, 2025
Prepared by: Reuben Samuel / Group 9 Leader

Executive summary

Week 1 deliverables are complete. We designed and implemented the data model, ingested anonymized EDC CSVs, built the Enrollment Overview dashboard in Power BI, created required DAX measures, polished visuals (KPI row, funnel, velocity chart), and validated model behavior. The dashboard is saved and documented for handoff.

Data Set Acquisition

This week's focus was on building the foundational dataset required to support the development of the clinical trial analytics dashboards. Since access to real patient data is restricted due to privacy and compliance requirements, we generated a synthetic dataset that realistically simulates clinical trial operations within a US context. The goal was to ensure that the dataset not only mirrors the structure and complexity of actual Electronic Data Capture (EDC) exports but also provides enough breadth and depth to enable the key analyses outlined in the project plan.

To achieve this, we designed and implemented a synthetic data generation pipeline that produced multiple interrelated CSV files representing patients, trial recruitment statuses, site organizations, providers, encounters, adherence tracking, medications, conditions, and supporting operational data. The dataset was crafted to maintain relational consistency across tables (for example, ensuring patient IDs match across recruitment, encounters, and adherence records), while also embedding realistic trial scenarios such as screening failures, randomized enrollments, dropouts, and medication adherence trends.

This dataset serves as the starting point for the subsequent phases of the project. Here, we created synthetic US-based CSV datasets for this project using several libraries like os (to create directories and handle file paths), zipfile (to package all CSVs into a single compressed archive), random for randomization (to select genders, specialties, medications, etc.) We also relied on datetime and timedelta (to generate screening dates, enrollment dates, and encounter timelines). On the data side, we worked with pandas (to build, manipulate, and export structured datasets as CSV files) and numpy (to handle random number generation, ranges, and sampling).

Below are the first few lines of codes applied

```
import os
import random
import pandas as pd
from datetime import datetime, timedelta
import numpy as np
import zipfile

# ------------------
# Setup
# ------------------
random.seed(42)
np.random.seed(42)

output_dir = "synthetic_trial_dataset_final"
os.makedirs(output_dir, exist_ok=True)
```

```
def random_date(start, end):
    """Generate a random date between start and end."""
    return start + timedelta(days=random.randint(0, (end - start).days))
```

Completed tasks

Data preparation
- Imported all raw CSV exports into Power BI.
- Confirmed and set correct data types (IDs → text, dates → date, flags → whole number).
- Created Calendar (date) table and marked it as the model date table.
- Created core relationships:
  • patients ↔ recruitment
  • patients ↔ encounters
  • encounters ↔ adherence / observations / procedures
  • providers ↔ organizations
  • patients ↔ organizations

Modeling & DAX measures
- Implemented required measures:
  • Total Screened
  • Total Enrolled
  • Total Randomized
  • Screen Failure Rate
  • Enrollment Rate
  • Randomization Rate
  • Avg Days Screening → Enrollment
  • Avg Days Enrollment → Randomization

Visuals & UX
- KPI cards: screened, enrolled, randomized; plus failure / enrollment / randomization rates.
- Avg Days cards moved into KPI row and styled consistently (same size, padding, alignment).
- Funnel chart: recruitment funnel ordered logically (enrolled → randomized → dropped → screen_failed) by creating StageOrder and StageLabel (number + label) and using that as category so the funnel displays top → bottom in correct flow.
- Line chart: cumulative enrollment velocity with:
  • x-axis labeled (month) and tick interval adjusted for readability (every 2–3 months).
  • subtle horizontal gridlines for easier reading.
  • final cumulative value surfaced as a callout (data label on last point / constant line).
- Minor visual polish: reduced heavy shadows, standardized numeric font weight/size, increased label contrast, adjusted card icons to avoid clipping, added tooltips with underlying counts.

Interactivity
- Slicers added for date range, patient id, and site id so visuals respond to filters.

Validation & testing (completed)
- Validate numbers against raw CSV counts: cross-checked key KPIs (screened, enrolled, randomized, screen_failed) between raw CSVs and dashboard measures using direct row counts and quick DAX checks — all matched.
- Relationship & filter tests: verified filtering behavior for patient, site, and date:
  • Filtering by patient returns only that patient's records and updates KPIs and table visuals.
  • Filtering by site shows site-specific recruitment funnel and velocity correctly.
  • Date range slicer properly restricts line chart and cumulative totals.
- Data quality checks: confirmed Avg Days measures exclude nulls and negative diffs; handled unexpected statuses with fallback logic.

Save & documentation

- Power BI file saved as: Group9_Week1_EnrollmentOverview.pbix.
- Supporting artifacts saved/documented: data sources list, data model diagram (Model view screenshot), DAX measure list, visual style guide (font sizes, colors, padding settings).
- Weekly report (this document) prepared for submission.

Evidence & reproducibility (brief)

- Raw CSV counts and DAX validation queries are recorded in the project notes (sheet: validation_checks) within the project folder.
- DAX used for defensive Avg Days (example):

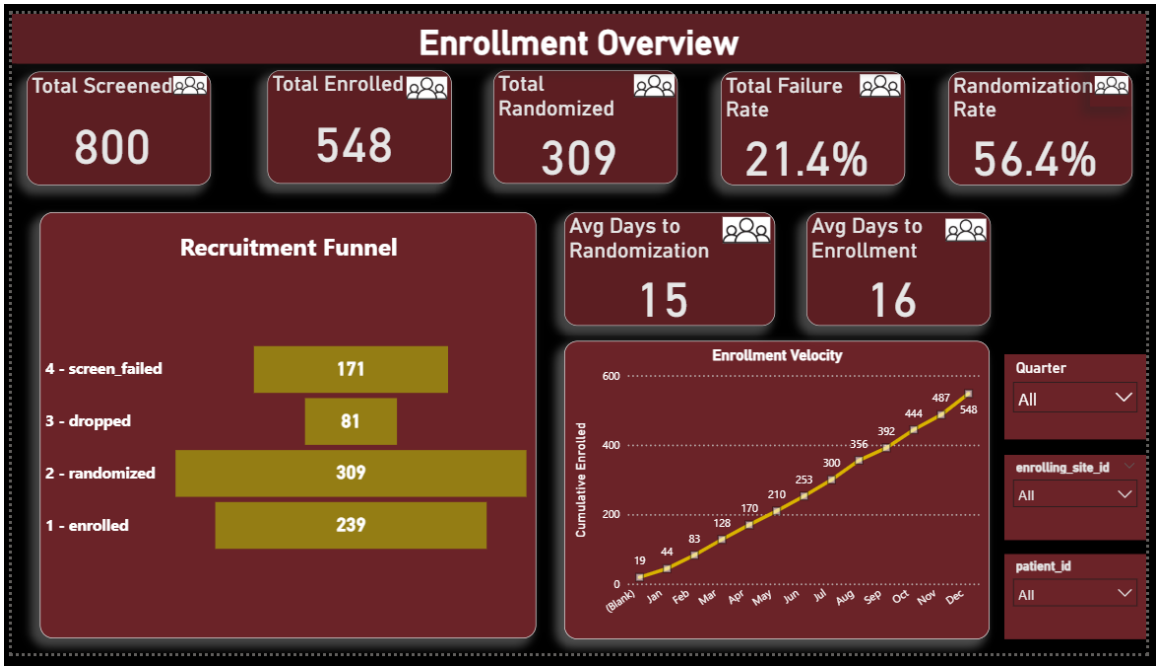
```
Avg Days to Randomization =
AVERAGEX(
  FILTER(
    Patients,
    NOT( ISBLANK(Patients[enrolled date]) ) &&
    NOT( ISBLANK(Patients[randomized date]) ) &&
    Patients[randomized date] >= Patients[enrolled date]
  ),
  DATEDIFF(Patients[enrolled date], Patients[randomized date], DAY)
)
```

- Funnel ordering implemented via StageOrder + StageLabel (created in Power Query) to ensure stable visual ordering.

Outstanding / next actions (hand-off to Week 2)

- Move into Week 2: model the recruitment funnel at site-level, compute site-level KPIs (screen failure rate, enrollment velocity per site), and build the Site Performance Leaderboard.
- Document any edge-case rows discovered during validation for follow-up (small list attached in validation_issues.csv).

Below is a screenshot of the dashboard



Conclusion

Week 1 is complete and validated. The Enrollment Overview dashboard is ready for stakeholder review and for Week 2 work (funnel modeling and site leaderboard).