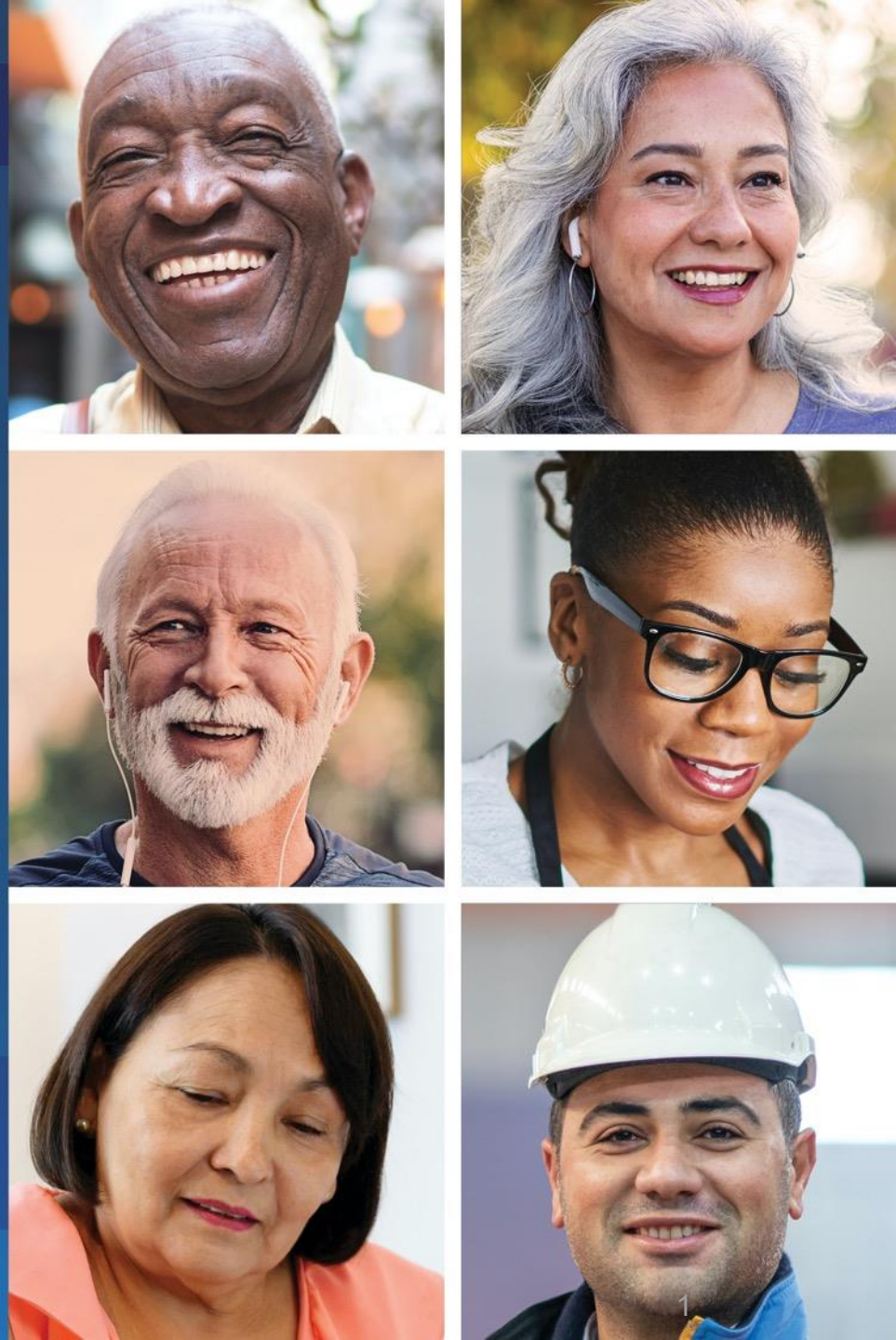
NCPCR Data Ecosystem Overview

Last updated 12/1/2025

National Institutes of Health
Community Engagement Alliance



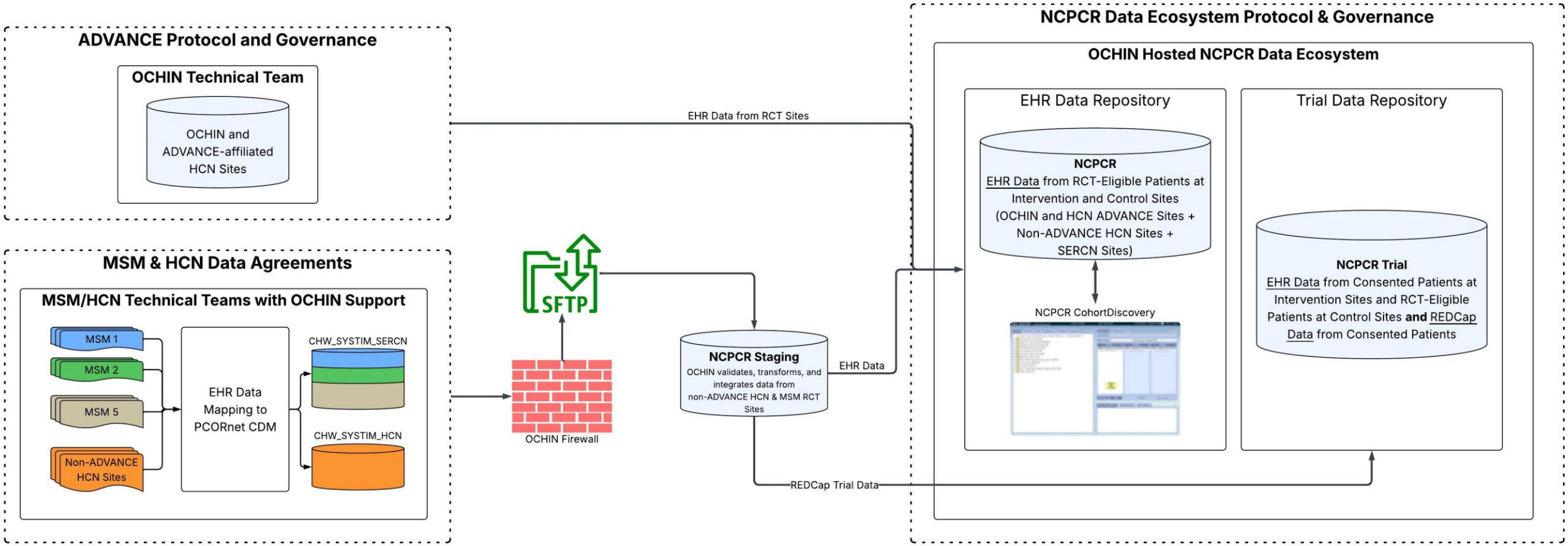
NCPCR Data Ecosystem Purpose

- The NCPCR Data Ecosystem (DE) was developed to support the Community Health Worker - Systematically Assessing and Addressing Social Determinants of Health (CHW-SYSTIM) randomized controlled trial and future primary care research across community health centers.

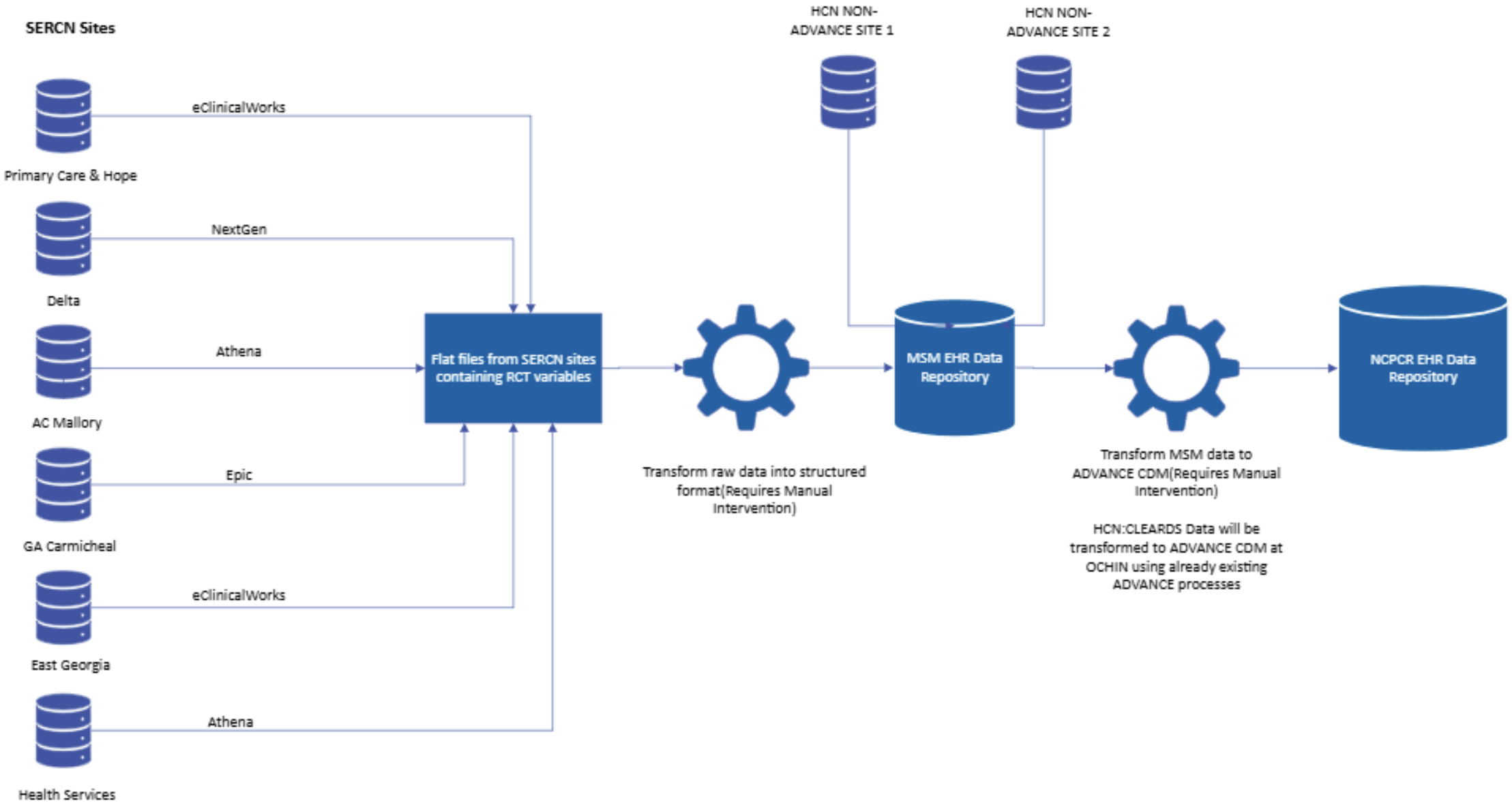
- The NCPCR DE integrates data from three major networks: **OCHIN**, **Health Choice Network (HCN)**, and **Morehouse School of Medicine/Southeast Regional Clinicians Network (MSM/SERCN)**

- Ultimately, the NCPCR DE creates a secure environment with standardized, research-ready data from these partner organizations.

NCPCR Data Ecosystem Data and Structure



Overview of MSM Data Transformation in NCPCR



NCPCR Data Ecosystem Data and Structure

EHR Data Repository

- RCT-relevant EHR data elements
- Eligible patients from the 26 RCT intervention and control health centers
  - Eligible patients: ≥18 years old, with a diabetes diagnosis and most recent A1c≥9, and / or a hypertension diagnosis and last systolic BP≥140 or last diastolic BP≥90

Trial Data Repository

- RCT-relevant EHR data for consented patients in intervention sites and eligible patients in control sites
- REDCap data from patients in intervention sites

Data Governance and Sharing

Governed by the NCPCR Data Ecosystem Governance Plan & Data Management and Sharing Plan

- Access to de-identified, aggregate data for prep-to-research queries
- Future potential: approved research project dataset extraction and release (detailed in Governance Plan)

EHR Data Repository

- De-identified aggregate data for prep-to-research
- Summary data (e.g., Table 1)

Trial Data Repository

- Two types of Trial Repository data that can be transferred to an external repository
- (1) Summary data (e.g., Table 1 type) for consented patients* at intervention sites and all patients at control sites**.
- (2) Second is HIPAA-deidentified data on consented patients at intervention sites.

**Patients that provide verbal consent to take part in CHW-SYSTIM, also give permission for future use of their trial specific EHR data and REDCap data, within specific parameters.*

***Patient data from control sites are covered under our waiver of consent and authorization*

NCPCR Data Ecosystem Hosting Environment

- The NCPCR Data Ecosystem is hosted at OCHIN and leverages existing infrastructure developed for ADVANCE, including:
  - PCORnet Common Data Model (CDM)
  - Software and technical controls for data security, privacy, and management
- **Hosting Environment**
  - Location: OCHIN's secure data center
  - Compliance: Meets organizational and regulatory standards for data protection and access control
  - Processing: All transactions and workloads run on secure servers

Design & Load of the NCPCR EHR Data Repository

- EHR data transformed into PCORnet CDM v7
- Site Integration Processes
  - ADVANCE Sites (OCHIN & HCN):
    - Governed by Data Transfer Agreements
    - Data extracted from ADVANCE Data Warehouse
    - Metadata includes standard terminologies, ontologies, and grouping systems
  - Non-ADVANCE Sites (HCN & MSM):
    - Integrated via separate processes
  - Uses proprietary ETL scripts to map raw EHR data to PCORnet CDM

NCPCR Cohort Discovery



The Cohort Discovery tool is fully operational with user guides and approval workflows integrated into governance plans.



New users must complete HIPAA and Responsible Conduct of Research training, sign agreements, and will be granted 60-day renewable access.



The tool supports preparatory research queries and returns de-identified, aggregate patient counts to ensure data privacy and organizational regulatory alignment.

Demonstration Queries

Key Considerations

- *The EHR Data Repository is limited to patients who were eligible for the RCT as of 9/30/2025 and ingested into the DE as part of Data Transfer 2 (10/31/2025).*
- *The following queries were executed before the completion of Data Transfer 2 (due 12/20/2025) and thus do not reflect the most up-to-date data for each site.*
- *As of the date these queries were run, there are ongoing data checks for the three sites that were delayed in sending data. These are not yet reflected in the queries presented here.*
- *Trial enrollment data is collected through REDCap and is not reflected in these EHR Data Repository queries. Please see the RCT updates for the enrollment data.*
- *These queries are sourced from the NCPCR EHR Data Repository which include eligible patients from the RCT intervention and control health centers.*

Number of Records and Number of Patients

In this example we demonstrate the number of records and the number of unique patients that appear in the “VITAL” table (where blood pressure measurements are stored

```sql
-- VITAL → ENCOUNTER → FACILITYID → rct_facility_crosswalk
SELECT
    'VITAL' AS TblName,
    cx.rct_healthsystem,
    cx.hs_intervention_yn,
    v.siteid,
    COUNT(*) AS RecordCount,
    COUNT(DISTINCT v.PATID) AS PatientCount
FROM NCPCR.dbo.VITAL AS v
LEFT JOIN NCPCR.dbo.ENCOUNTER AS e
       ON e.ENCOUNTERID = v.ENCOUNTERID
LEFT JOIN NCPCR_staging.dbo.rct_facility_crosswalk AS cx
       ON cx.FACILITYID = e.FACILITYID
WHERE cx.rct_healthsystem IS NOT NULL
GROUP BY cx.rct_healthsystem, cx.hs_intervention_yn, v.siteid
ORDER BY cx.hs_intervention_yn desc, v.siteid
```

TblName	rct_healthsystem	hs_intervention_yn	siteid	RecordCount	PatientCount
VITAL	COMMWELL HEALTH	1	HCN	96838	6194
VITAL	WEST HAWAII CHC/HICHC KUAKINI FAMILY	1	HCN	95461	5799
VITAL	GENESIS COMMUNITY HEALTH	1	HCN	13680	884
VITAL	NEOMED	1	HCN_PR	6033	1246
VITAL	LYNN COMMUNITY HEALTH CENTER	1	OCHIN	50598	6515
VITAL	EL CENTRO DE CORAZON	1	OCHIN	5416	1223
VITAL	AGAPE HEALTH SERVICES	1	OCHIN	41141	5591
VITAL	AXIS COMMUNITY HEALTH CENTER	1	OCHIN	27946	3928
VITAL	LIFELONG	1	OCHIN	83384	10905
VITAL	EAST GEORGIA HEALTH CENTER	1	SERCN	9406	4513
VITAL	HEALTH SERVICES	1	SERCN	2256	1879
VITAL	LA CASA FAMILY HEALTH CENTER	0	HCN	40851	2541
VITAL	HEALTH CARE CENTER FOR THE HOMELESS/ORANGE BLOSSOM	0	HCN	56792	3788
VITAL	MEDCENTRO	0	HCN_PR	20286	3476
VITAL	BAYWELL HEALTH	0	OCHIN	14243	2199
VITAL	KINTEGRA HEALTH	0	OCHIN	91189	15814
VITAL	FRANKLIN PRIMARY HEALTH CENTER	0	OCHIN	68342	12248
VITAL	MONTEREY COUNTY HEALTH DEPT	0	OCHIN	55499	6074
VITAL	AVENUE 360 HEALTH & WELLNESS	0	OCHIN	18576	3251
VITAL	ERIE FAMILY HEALTH CENTERS	0	OCHIN	93420	12300
VITAL	PEOPLE'S COMMUNITY CLINIC	0	OCHIN	21971	2173
VITAL	DELTA HEALTH CENTER	0	SERCN	4011	3082

The following queries were executed before the completion of Data Transfer 2 (due 12/20/2025) and thus do not reflect the most up-to-date data for each site; As of the date these queries were run, there are ongoing data checks for the three sites that were delayed in sending data. These are not yet reflected in the queries presented here.

Number of Records and Number of Patients

In this example we demonstrate the number of records and the number of unique patients that appear in the “LAB_RESULT_CM” table (where HbA1c is stored)

```sql
-- LAB_RESULT_CM → ENCOUNTER → FACILITYID → rct_facility_crosswalk
SELECT
    'LAB_RESULT_CM' AS TblName,
    cx.rct_healthsystem,
    cx.hs_intervention_yn,
    l.siteid,
    COUNT(*) AS RecordCount,
    COUNT(DISTINCT l.PATID) AS PatientCount
FROM NCPCR.dbo.LAB_RESULT_CM AS l
LEFT JOIN NCPCR.dbo.ENCOUNTER AS e
       ON e.ENCOUNTERID = l.ENCOUNTERID
LEFT JOIN NCPCR_staging.dbo.rct_facility_crosswalk AS cx
       ON cx.FACILITYID = e.FACILITYID
WHERE cx.rct_healthsystem IS NOT NULL
GROUP BY cx.rct_healthsystem, cx.hs_intervention_yn, l.siteid
ORDER BY cx.hs_intervention_yn desc, l.siteid
```

TblName	rct_healthsystem	hs_intervention_yn	siteid	RecordCount	PatientCount
LAB_RESULT_CM	COMMWELL HEALTH	1	HCN	11974	5029
LAB_RESULT_CM	WEST HAWAII CHC/HICHC KUAKINI FAMILY	1	HCN	7455	3533
LAB_RESULT_CM	GENESIS COMMUNITY HEALTH	1	HCN	1453	649
LAB_RESULT_CM	NEOMED	1	HCN_PR	1705	951
LAB_RESULT_CM	EL CENTRO DE CORAZON	1	OCHIN	2156	1045
LAB_RESULT_CM	LIFELONG	1	OCHIN	24006	7428
LAB_RESULT_CM	AXIS COMMUNITY HEALTH CENTER	1	OCHIN	10968	3390
LAB_RESULT_CM	AGAPE HEALTH SERVICES	1	OCHIN	9141	3853
LAB_RESULT_CM	LYNN COMMUNITY HEALTH CENTER	1	OCHIN	10754	4910
LAB_RESULT_CM	EAST GEORGIA HEALTH CENTER	1	SERCN	3882	2422
LAB_RESULT_CM	HEALTH SERVICES	1	SERCN	1738	1454
LAB_RESULT_CM	HEALTH CARE CENTER FOR THE HOMELESS/ORANGE BLOSSOM	0	HCN	5534	2644
LAB_RESULT_CM	LA CASA FAMILY HEALTH CENTER	0	HCN	4385	1886
LAB_RESULT_CM	MEDCENTRO	0	HCN_PR	2511	1669
LAB_RESULT_CM	ERIE FAMILY HEALTH CENTERS	0	OCHIN	26667	10904
LAB_RESULT_CM	PEOPLE'S COMMUNITY CLINIC	0	OCHIN	6762	1862
LAB_RESULT_CM	KINTEGRA HEALTH	0	OCHIN	29317	11895
LAB_RESULT_CM	MONTEREY COUNTY HEALTH DEPT	0	OCHIN	12370	4437
LAB_RESULT_CM	FRANKLIN PRIMARY HEALTH CENTER	0	OCHIN	14371	6887
LAB_RESULT_CM	AVENUE 360 HEALTH & WELLNESS	0	OCHIN	4567	2192
LAB_RESULT_CM	BAYWELL HEALTH	0	OCHIN	3034	1460
LAB_RESULT_CM	DELTA HEALTH CENTER	0	SERCN	3049	2481

The following queries were executed before the completion of Data Transfer 2 (due 12/20/2025) and thus do not reflect the most up-to-date data for each site; As of the date these queries were run, there are ongoing data checks for the three sites that were delayed in sending data. These are not yet reflected in the queries presented here.

Patients with 2+ Blood Pressure Readings

Number of patients with two or more blood pressure readings

```sql
--By Healthsystem
-- Patients with >=2 distinct encounters having BP values, grouped by RCT health system
SELECT
    sub.rct_healthsystem,
    sub.hs_intervention_yn,
    COUNT(*) AS patient_count
FROM (
    SELECT
        cx.rct_healthsystem,
        cx.hs_intervention_yn,
        v.PATID
    FROM NCPCR.dbo.VITAL AS v
    INNER JOIN NCPCR.dbo.ENCOUNTER AS e
        ON e.ENCOUNTERID = v.ENCOUNTERID
    INNER JOIN NCPCR_staging.dbo.rct_facility_crosswalk AS cx
        ON cx.FACILITYID = e.FACILITYID
    WHERE v.SYSTOLIC IS NOT NULL
       OR v.DIASTOLIC IS NOT NULL
    GROUP BY
        cx.rct_healthsystem,
        cx.hs_intervention_yn,
        v.PATID
    HAVING COUNT(DISTINCT v.ENCOUNTERID) >= 2
) AS sub
GROUP BY
    sub.rct_healthsystem,
    sub.hs_intervention_yn
ORDER BY
    sub.hs_intervention_yn desc, sub.rct_healthsystem;
```

rct_healthsystem	hs_intervention_yn	patient_count
AGAPE HEALTH SERVICES	1	4936
AXIS COMMUNITY HEALTH CENTER	1	3409
COMMWELL HEALTH	1	5585
EAST GEORGIA HEALTH CENTER	1	642
EL CENTRO DE CORAZON	1	1015
GENESIS COMMUNITY HEALTH	1	762
HEALTH SERVICES	1	377
LIFELONG	1	9534
LYNN COMMUNITY HEALTH CENTER	1	5733
NEOMED	1	1029
WEST HAWAII CHC/HICHC KUAKINI FAMILY	1	5265
AVENUE 360 HEALTH & WELLNESS	0	2799
BAYWELL HEALTH	0	1859
DELTA HEALTH CENTER	0	929
ERIE FAMILY HEALTH CENTERS	0	11146
FRANKLIN PRIMARY HEALTH CENTER	0	10162
HEALTH CARE CENTER FOR THE HOMELESS/ORANGE BLOSSOM	0	3235
KINTEGRA HEALTH	0	13944
LA CASA FAMILY HEALTH CENTER	0	2349
MEDCENTRO	0	3039
MONTEREY COUNTY HEALTH DEPT	0	5661
PEOPLE'S COMMUNITY CLINIC	0	2064

The following queries were executed before the completion of Data Transfer 2 (due 12/20/2025) and thus do not reflect the most up-to-date data for each site; As of the date these queries were run, there are ongoing data checks for the three sites that were delayed in sending data. These are not yet reflected in the queries presented here.

Patients with Blood Pressure ≥140/90, Stratified by Sex

Number of patients with elevated blood pressure (ever during study period)

```sql
SELECT
    cx.rct_healthsystem,
    cx.hs_intervention_yn,
    -- Patient counts
    COUNT(DISTINCT CASE WHEN p.SEX = 'M' THEN p.PATID END) AS male_patient_count,
    COUNT(DISTINCT CASE WHEN p.SEX = 'F' THEN p.PATID END) AS female_patient_count,
    -- Encounter counts
    COUNT(DISTINCT CASE WHEN p.SEX = 'M' THEN v.ENCOUNTERID END) AS male_encounter_count,
    COUNT(DISTINCT CASE WHEN p.SEX = 'F' THEN v.ENCOUNTERID END) AS female_encounter_count
FROM NCPCR.dbo.DEMOGRAPHIC AS p
INNER JOIN NCPCR.dbo.VITAL AS v
        ON v.PATID = p.PATID
INNER JOIN NCPCR.dbo.ENCOUNTER AS e
        ON e.ENCOUNTERID = v.ENCOUNTERID
INNER JOIN NCPCR_staging.dbo.rct_facility_crosswalk AS cx
        ON cx.FACILITYID = e.FACILITYID
WHERE (v.SYSTOLIC >= 140 OR v.DIASTOLIC >= 90)
  AND p.SEX IN ('F', 'M')
GROUP BY
    cx.rct_healthsystem, cx.hs_intervention_yn
ORDER BY
    cx.hs_intervention_yn DESC, cx.rct_healthsystem;
```

rct_healthsystem	hs_intervention_yn	male_patient_count	female_patient_count	male_encounter_count	female_encounter_count
AGAPE HEALTH SERVICES	1	2405	2904	7171	9015
AXIS COMMUNITY HEALTH CENTER	1	1540	2183	4416	6809
COMMWELL HEALTH	1	2546	2936	6288	7500
EAST GEORGIA HEALTH CENTER	1	1799	2206	2026	2522
EL CENTRO DE CORAZON	1	247	720	443	1560
GENESIS COMMUNITY HEALTH	1	323	508	868	1363
HEALTH SERVICES	1	591	691	631	751
LIFELONG	1	4500	5623	14405	18978
LYNN COMMUNITY HEALTH CENTER	1	2739	3092	8398	9023
NEOMED	1	443	670	893	1323
WEST HAWAII CHC/HICHC KUAKINI FAMILY	1	2747	2631	7202	6875
AVENUE 360 HEALTH & WELLNESS	0	1318	1711	3788	4851
BAYWELL HEALTH	0	895	1072	2385	2769
DELTA HEALTH CENTER	0	1103	1778	1392	2258
ERIE FAMILY HEALTH CENTERS	0	4552	6013	11343	16122
FRANKLIN PRIMARY HEALTH CENTER	0	5535	6150	14161	16165
HEALTH CARE CENTER FOR THE HOMELESS/ORANGE BLOSSOM	0	1712	1862	4654	5462
KINTEGRA HEALTH	0	6252	7372	13136	15653
LA CASA FAMILY HEALTH CENTER	0	1111	1191	2407	2672
MEDCENTRO	0	1350	1951	2659	3971
MONTEREY COUNTY HEALTH DEPT	0	2426	3176	7398	10404
PEOPLE'S COMMUNITY CLINIC	0	613	1269	1905	4808

The following queries were executed before the completion of Data Transfer 2 (due 12/20/2025) and thus do not reflect the most up-to-date data for each site; As of the date these queries were run, there are ongoing data checks for the three sites that were delayed in sending data. These are not yet reflected in the queries presented here.

Patients with HbA1c >9%, Stratified by Sex

Number of patients with elevated HbA1c (ever during study period)

```sql
SELECT
    cx.rct_healthsystem,
    cx.hs_intervention_yn,
    -- Patient counts
    COUNT(DISTINCT CASE WHEN p.SEX = 'M' THEN p.PATID END) AS male_patient_count,
    COUNT(DISTINCT CASE WHEN p.SEX = 'F' THEN p.PATID END) AS female_patient_count,
    -- Encounter counts
    COUNT(DISTINCT CASE WHEN p.SEX = 'M' THEN l.ENCOUNTERID END) AS male_encounter_count,
    COUNT(DISTINCT CASE WHEN p.SEX = 'F' THEN l.ENCOUNTERID END) AS female_encounter_count
FROM NCPCR.dbo.DEMOGRAPHIC AS p
INNER JOIN NCPCR.dbo.LAB_RESULT_CM AS l
        ON l.PATID = p.PATID
INNER JOIN NCPCR.dbo.ENCOUNTER AS e
        ON e.ENCOUNTERID = l.ENCOUNTERID
INNER JOIN NCPCR_staging.dbo.rct_facility_crosswalk AS cx
        ON cx.FACILITYID = e.FACILITYID
WHERE l.RESULT_NUM > 9
  AND p.SEX IN ('F', 'M')
GROUP BY
    cx.rct_healthsystem, cx.hs_intervention_yn
ORDER BY
    cx.hs_intervention_yn DESC, cx.rct_healthsystem;
```

rct_healthsystem	hs_intervention_yn	male_patient_count	female_patient_count	male_encounter_count	female_encounter_count
AGAPE HEALTH SERVICES	1	546	603	854	1011
AXIS COMMUNITY HEALTH CENTER	1	1085	1380	1822	2257
COMMWELL HEALTH	1	770	731	1317	1378
EAST GEORGIA HEALTH CENTER	1	556	536	620	623
EL CENTRO DE CORAZON	1	144	357	213	579
GENESIS COMMUNITY HEALTH	1	51	107	83	204
HEALTH SERVICES	1	430	604	530	756
LIFELONG	1	2878	3492	4497	5582
LYNN COMMUNITY HEALTH CENTER	1	1008	936	1557	1538
NEOMED	1	134	152	228	231
WEST HAWAII CHC/HICHC KUAKINI FAMILY	1	663	661	960	1004
AVENUE 360 HEALTH & WELLNESS	0	221	331	373	593
BAYWELL HEALTH	0	196	184	356	341
DELTA HEALTH CENTER	0	184	253	239	337
ERIE FAMILY HEALTH CENTERS	0	1635	1843	3082	3900
FRANKLIN PRIMARY HEALTH CENTER	0	917	852	1482	1501
HEALTH CARE CENTER FOR THE HOMELESS/ORANGE BLOSSOM	0	322	363	567	676
KINTEGRA HEALTH	0	1860	2016	3353	3799
LA CASA FAMILY HEALTH CENTER	0	306	302	508	531
MEDCENTRO	0	201	248	246	332
MONTEREY COUNTY HEALTH DEPT	0	984	1417	1649	2463
PEOPLE'S COMMUNITY CLINIC	0	524	1071	899	1883

The following queries were executed before the completion of Data Transfer 2 (due 12/20/2025) and thus do not reflect the most up-to-date data for each site; As of the date these queries were run, there are ongoing data checks for the three sites that were delayed in sending data. These are not yet reflected in the queries presented here.

Data Transfer 2: Progress Update

- Completed EHR data transfer from **25** participating HCs.
  - Data from the 3 outstanding SERCN sites have been received, transformed into the PCORnet CDM, and transferred to OCHIN for ingestion into the DE. Full transfer completion by 12/20.
- Completed REDCap data transfer from all 12 Intervention HCs.
- Data transformed to PCORnet CDM and ingested in NCPCR database.
- Study inclusion criteria applied for NCPCR database.

Data Transfer 2: Lessons Learned

- **Health center constraints:** Staffing shortages, competing priorities, and operational limitations affect their ability to meet project deadlines.
- **Need for flexibility:** To mitigate downstream delays, we need to proactively build in additional time, which was not included in previously approved timelines, and allow for greater flexibility in our planning to accommodate these external dependencies.
- **Extended preparation time required**: Health centers require additional time to prepare and transfer datasets, especially when juggling multiple demands.
- **Build more time into project workflow**: NCPCR team must allocate extra time and staffing levels for each transfer step to account for documentation needs, scheduling conflicts, and unforeseen delays.

Data Challenges and Resolutions

Step	New & Existing Challenges	Resolutions
SERCN CHC → MSM	- Non-standard file formats - Combined domains in raw files - Missing/null values - Duplicate records - EHR limitations (DX_Date, Encounter_ID)	- Standardized schema & formats - Split into domain-specific datasets - Applied PCORnet null flavors - Deduplicated using PAT_ID & date - Generated Encounter_IDs
MSM → OCHIN	- PCORnet CDM conformance errors - Missing Encounter_IDs in PRO, vitals, labs - Diagnosis linking issues - Minor QC errors (codes, formats)	- Applied PCORnet QC checks - Generated Encounter_IDs via patient ID + date proximity - Set DX_Date = admit_date - Corrected conformance issues
ADVANCE / HCN PR	- Inclusion/exclusion errors (e.g., NeoMed non-intervention sites) - Minor QC issues	- Filtered to intervention-only - Corrected PCORnet conformance issues
NCPCR DB	- Raw data included out-of-scope patients - Missing fields for inclusion logic	- Applied study inclusion criteria - Filtered and validated before ingestion

Potential Risks & Mitigation

- ***Delayed Extracts from CHCs:*** Some sites experienced vendor-related or staffing delays; sending early reminders and leadership escalations will help ensure timely delivery.
- ***Incomplete data extracts:*** Although clear extraction templates were provided, some sites still returned datasets missing key variables; follow-up validation and clarification will be emphasized.
- ***Requirement for multiple follow-ups:*** Frequent reminders were needed to confirm complete data extract.

Improvements for Final Transfer

- ***Early Data Readiness Checks:*** Request a sample extract early from each site and validate for format issues, missing fields, or delays before the full extract.
- ***Early reminders and identify gaps/ delays:*** Send reminders at least two weeks before extraction deadlines and track progress consistently
- ***Improve communication loops:*** Continue targeted email follow-ups and periodic check-ins with CHC data teams, summarizing progress and next steps for accountability.

Data Transfer 3: Timeline Considerations

- The team **requires** at least 8 weeks from the end of biomarker follow-up to complete the following tasks needed to complete Data Transfer #3 (final analysis dataset):
  1. Obtain data from SERCN health centers (3 weeks)
  2. Transformation of SERCN data to PCORnet CDM (2 weeks)
  3. Transfer to OCHIN and ingestion into the Data Ecosystem (2 weeks)
  4. Completion of documentation and deliverables (1 week)

Questions?