

AI for Public Health and Population Health

Study Guide

by Richard J. Young, Ph.D.

richard@deepneuro.ai

Chapter 22 of "Healthcare Analytics and AI: Building Systems That Actually Work"

The essentials: how AI sees communities instead of patients, and why incentives decide whether it helps.

One-sentence thesis. Population health AI answers “what is happening to this community, and why?”, a question you cannot reach by scaling up an individual risk model, because the data, the incentives, and the interventions are all different.

PART 1: Two Alerts, Two Paradigms

Individual Versus Population

Two alerts fire at 7:42 a.m., and they belong to different worlds. An individual readmission model flags Mrs. Patterson at 34 percent risk; a care manager calls her, reconciles meds, and books a follow-up, preventing perhaps one readmission for about \$85, one patient at a time. At the same moment, a population platform on an all-payer database finds 312 women overdue for mammography in four zip codes, sharing one barrier: transportation. The fix is not 312 phone calls; it is a mobile van in a church parking lot on three Saturdays plus multilingual texting.

ANALOGY

A flashlight versus a map. The flashlight (individual AI) lights up one patient at a time; the map (population AI) shows where risk is concentrated and what structural feature concentrates it there.

Both are AI and both matter, but the second needs multi-payer aggregation, geospatial analysis, and community-level intervention design that no individual readmission model ever had to develop.

Watch out. Scaling up an individual model does not produce population health. The questions, and therefore the data and the interventions, are categorically different.

Pre-Diagnostic Signals

Classical outbreak detection is reactive: a clinician sees a case, a lab confirms it, a report is filed, and days to weeks pass. Syndromic surveillance inverts this by monitoring pre-diagnostic signals in near real-time (ED chief complaints, pharmacy sales, absenteeism, search volume). Any one of them is noisy, but they become convincing when independent signals converge in the same place at the same time, the way smelling smoke, hearing a crackle, and feeling heat separately mean little but together mean fire before you see flames.

The canonical wins are concrete. **HealthMap** (2006) mines multilingual news into a live global outbreak map. **BlueDot** flagged the Wuhan pneumonia clus-

ter on December 31, 2019, **nine days** before the WHO notification, by combining multilingual NLP with airline-mobility data to forecast where cases would surface next. **Ensemble forecasting** (the CDC FluSight and COVID-19 Forecast Hub) beats any single model, the same ensemble principle behind random forests. Genomic surveillance (Pangolin, Nextstrain) detected Omicron within days in November 2021.

Watch out. These systems are only as strong as the surveillance feeding them, and that infrastructure is chronically underfunded between outbreaks. Forecasts also train on biased counts: reported cases are a biased sample and the true infection denominator is unknown.

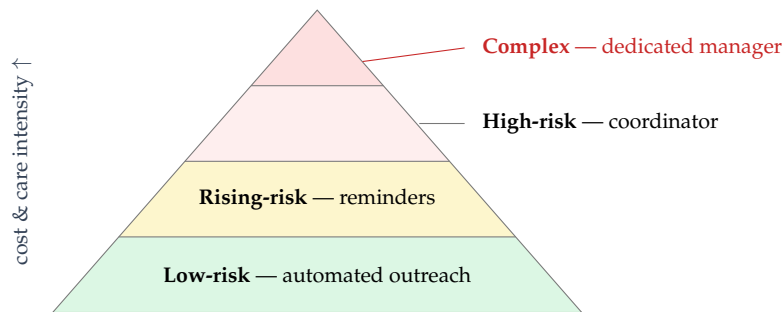
System	Data layer	ML method	Canonical win
HealthMap	Multilingual news, reports	NER + spatiotemporal	Continuous global outbreak map
BlueDot	News + airline mobility	NLP + network analysis	Wuhan cluster, 9 days before WHO
Pangolin / Nextstrain	Genome sequences	Phylogenetic ML + embeddings	Omicron within days, Nov 2021

PART 3: Population Risk Stratification

From Scores to Tiers

An individual model outputs a continuous score; a population system does something structurally different, sorting an entire attributed panel (hundreds of thousands of lives) into tiers, each getting a different care-management intensity. It is triage for a whole city rather than a waiting room: you cannot give every resident a case manager, so you sort them.

The Adjusted Clinical Groups (ACG) system, from Johns Hopkins, maps ICD-10 codes into roughly 100 morbidity categories with a resource utilization band. Modern approaches layer ML on top, predicting who will escalate a tier and using NLP to pull SDOH indicators absent from structured claims. The tiers translate directly into resource allocation, from automated outreach for the low-risk to a dedicated care manager for the complex.



The population risk pyramid. A whole attributed panel is sorted into tiers, each getting a different care-management intensity; cost and intensity rise toward the narrow apex, where the complex few live.

Tier	Criteria	Care intensity	Cost band (per year)
Low-risk	No chronic conditions	Automated outreach	under \$1,000
Rising-risk	1 to 2 chronic conditions	Reminders, check-ins	\$1,000 to \$10,000
High-risk	3+ chronic or 1+ inpatient	Pharmacist, coordinator	\$10,000 to \$50,000
Complex	High-risk plus SDOH or behavioral	Dedicated care manager	over \$50,000

The Denominator Problems

The cruelest failure mode is that the patient at greatest risk is often the least visible. Someone who has not seen a PCP in two years generates no recent codes, so an activity-based model scores them low, reading data absence as health presence and deprioritizing the person who most needs outreach.

KEY IDEA

Absence of data is not absence of risk. A model that scores the invisible patient as low-risk has mistaken silence for health.

A second denominator problem sits one level up: a

single payer never sees a whole community. One insurer's claims might show 11 percent diabetes prevalence in a zip code where the true all-payer figure is 19 percent. **All-payer claims databases (APCDs)** aggregate across payers (about 20 states operate them), but ERISA-protected self-insured employers need not participate, so most APCDs are incomplete by design.

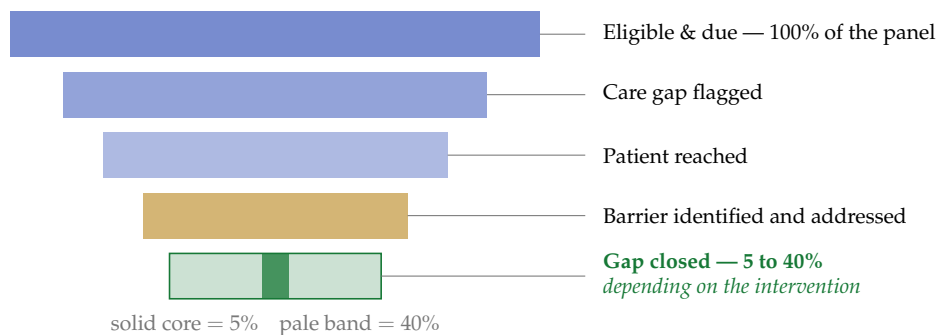
Watch out. Geospatial analysis by census tract, not patient panel, exposes patterns claims cannot: if 60 percent of highest-cost patients cluster in six contiguous tracts that all lack a grocery store, pharmacy, and clinic, the answer is community-level (mobile clinics, food access), not more care-manager calls.

From Reminder to Resolution

A care gap is an operational failure, not a clinical mystery: the system knew a mammogram or HbA1c was due and did not make it happen. HEDIS (about 90 NCQA measures) defines each gap by denominator, numerator, and exclusions. Finding the gap is the easy part; a leak you can spot with a flashlight is not the hard part, knowing whether it needs tape, a wrench, or a plumber is. So AI identifies gaps and, more usefully, predicts the **barrier** so the intervention matches it, from a text reminder up to a community health worker. Matching intervention to barrier is

what actually closes gaps: transportation assistance raises mammography closure by 20 to 40 percentage points over reminders alone, but only for the patients whose barrier is transport.

Watch out. HEDIS permits exclusions for unreachable, declining, or contraindicated patients, and an AI trained to maximize Star Ratings can learn that documenting an exclusion is functionally equivalent to closing a gap. Decide whether your objective is “close the gap” or “improve the measured rate”, because the difference is felt by the patients whose gaps are documented rather than closed.



The care-gap funnel. Attrition at every step means the leverage is not finding the gap but matching the intervention to the real barrier. Bar width is share of the eligible panel; only the two ends are quantified in the chapter — everyone eligible and due at the top, and a 5 to 40 percent closure rate at the bottom depending on which intervention is used (SDOH referrals themselves close only 20 to 40 percent of the time). The two middle widths are an illustrative taper, not measured values.

Intervention	Cost per contact	Closure rate	Best for
Text reminder	under \$1	5 to 12 percent	Forgetfulness, low salience
Care-manager call	\$30 to \$85	15 to 30 percent	Identifying the specific barrier
Transportation assistance	\$20 to \$50	+20 to 40 pp	Structural transport barrier
Community health worker	\$50 to \$150	25 to 40 percent	Multi-barrier, trust required

Referral Loops and Equity Dashboards

SDOH platforms (FindHelp, Unite Us, and the 211 network) match identified needs to community resources, but only 20 to 40 percent of SDOH referrals close, an **interoperability** failure: community organizations lack EHR access and have no standard way to report outcomes back. Equity dashboards stratify every metric by race, ethnicity, language, geography, and payer, turning a reassuring “78 percent screening rate” into six numbers that reveal two standard deviations of inequity. The distinction that matters is that

model fairness is not system fairness: a readmission model can be perfectly calibrated across racial groups while the system’s readmission rate for Black patients runs 50 percent higher, driven by access and social determinants no model predicts.

Watch out. Over 70 percent of U.S. hospitals have equity dashboards; fewer than **15 percent** fund action on them. A dashboard that accurately documents unchanging inequity year after year is a memorial, not a tool.

Monitoring Without Surveillance

Some of the best population signals never touch a patient at all. Environmental health AI fuses satellite imagery, air-quality monitors, and weather to forecast PM2.5, heat vulnerability, and wildfire smoke at census-tract resolution, enabling pre-emptive outreach (cooling centers, medication pre-positioning) before anyone reaches a clinic. Wastewater surveillance goes further: it tests sewage for pathogen genetic material, a population signal that leads clinical case counts by 4 to 10 days, is inherently anonymized, and is unaffected by testing access. It is a tide gauge for community infection, reading the water level for the whole harbor without identifying a single boat.

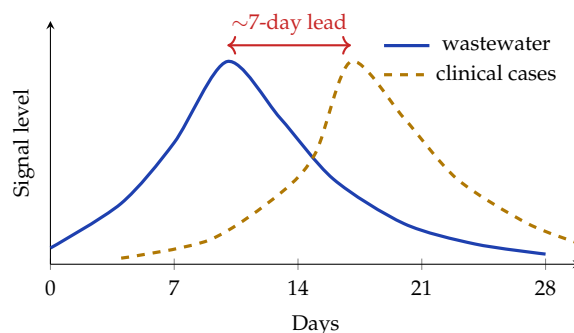
During Omicron, wastewater forecasts called the surge about a week ahead of case data, giving hospitals time to activate surge plans, and the infrastructure now extends to opioid metabolites, antimicrobial-resistance genes, and multiplexed respiratory panels. Because a sample from a plant serving 500,000 people traces to no individual, it is population monitoring without population surveillance, which is exactly what makes it acceptable to a community.

Watch out. The rate-limiting factor is alignment, not algorithms. Public health departments have the mandate but not the funding (CDC core funding fell **24 percent** from 2010 to 2024). Health systems have the strategic interest but a quarterly time horizon, so

revenue-cycle tools beat population health in capital allocation. Payers have the sophistication but optimize for Star Rating scores, not substance. Communities are modeled and monitored but rarely invested in. The pandemic-preparedness cycle (surge funding, then collapse, then rebuild from zero) is a political failure, not a technical one.

Should You Build This?

Build it when the intervention exists at the same level as the analysis: a mobile mammography van, a wastewater alert a health department will act on, a care-gap campaign a value-based organization is structurally committed to funding. Population health AI pays off where multi-payer data, geospatial context, and an accountable owner for community-level action already exist. **Think twice when** the money is emergency supplemental funding that expires with the crisis, the build-during-crisis, abandon-after-crisis cycle this chapter documents, or when the system will optimize the measured score over the substance, closing gaps on paper by documenting exclusions rather than delivering services. **Who bears the risk:** the communities being modeled, whose data generates the dashboards and the Star Rating gains, and who are being extracted from rather than served if their conditions never measurably improve; and the underfunded health departments left maintaining the pipelines when the grants end.



Wastewater leads the clinic. The sewage signal rises 4 to 10 days before clinical case counts, is inherently anonymized, and is unaffected by testing access; during Omicron it called the surge about a week early. Both curves are schematic and each is scaled to its own peak, so the vertical axis carries no units and the two series are not comparable in level — only in timing.

Quick Self-Test

1. Two 7:42 a.m. alerts, one individual and one population, demand different responses. What structural capability does the second require that the first never needed?
2. Explain the EHR denominator problem and why it systematically deprioritizes the patients who most need outreach.
3. One payer sees 11 percent diabetes prevalence where the true figure is 19 percent. What is the mechanism, what fixes it, and what political barrier keeps the fix incomplete?
4. How can an AI improve a HEDIS-measured screening rate without improving actual care, and what design choice invites this?
5. Give two reasons wastewater surveillance is valuable that a clinical case-count system cannot match.

If You Remember One Thing

Population health AI operates on a different data layer than clinical AI, and its bottleneck is almost never the algorithm. The models to detect outbreaks, map risk, and find care gaps largely exist and largely work. Whether they help a community depends on multi-payer data that is fragmented by design, on interventions matched to the real barrier, and on stakeholders whose incentives (funding cycles, quarterly horizons, Star Rating scores) rarely align with the community being measured. A dashboard that documents unchanging inequity is a memorial, not a tool.

Richard J. Young, Ph.D. · richard@deepneuro.ai · Senior AI Research Scientist, UnitedHealth Group · UNLV Lee Business School · ORCID 0000-0002-1109-7552