Longitudinal encounter histories as text:
Using graph machine learning and natural language processing methods to locate patient voices within constituencies.

Jonas Bambi, Ernie Chang, Austin Hawkins-Seagram, Joseph Howie, Kenneth Moselle, Ashlin Richardson, Abraham Rudnick, Yudi Santoso

ABSTRACT

In this work, funded by the Michael Smith Foundation and Victoria Hospitals foundation, we tackle the challenge of locating the patient voice within a constituency – using health service encounter data. We employ tools that start with individual patient service encounters, assemble them into within-patient-over-time journeys typically criss-crossing the Island Health service terrain, and then identify clusters of persons who track to the same clusters of services as they navigate across that terrain.

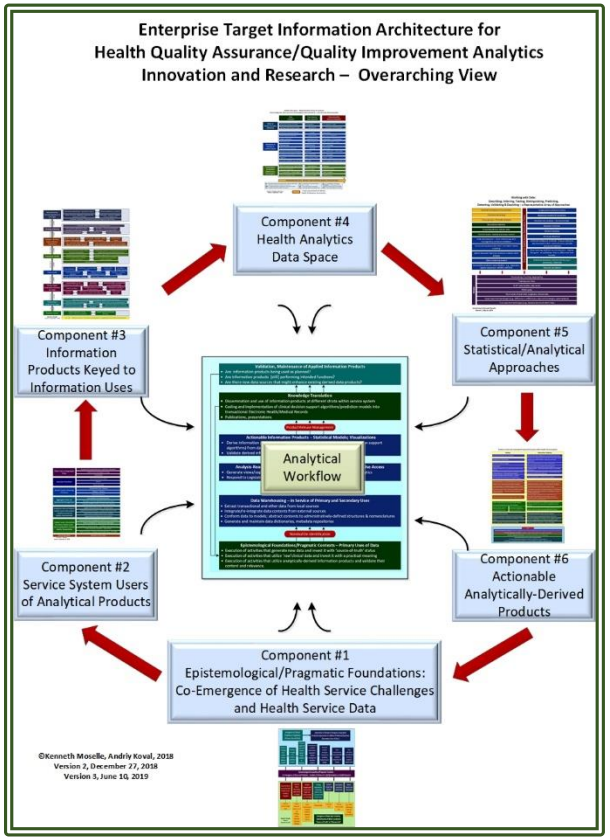
Pragmatically, the work is organized around a model of system dynamics in which service system pathways function as limiters/enablers for clinical pathways, which are carriers for clinical protocols keyed to focal problems. Service pathways (together with local capacity) are critical determinants of whether system potential to deliver "best practice" is actualized. Cohort-specific pathways are suited to cohort-comparison designs, which we use to operationalize constructs such as "equity".

More concretely, we present products of an analytical pipeline that treats each patient encounter with a service entity as a statement (literally). The location build within the Island Health cross-continuum deployment of Cerner provides the lexicon. When encounters are strung together (a journey), the product for any given patient is a narrative that conveys a different message than the sum of diagnoses and treatments associated with single service units. We use natural language processing/neural network tools and graph machine learning tools (community detection) applied to large arrays of person-level trajectories to make sense of these messages.

Specifically, we create two co-emergent cluster solutions: (a) service groupings accessed routinely by groups of patients, directly reflecting need and preference; and (b) groups of patients who track similarly to service groupings - their journeys are "saying the same thing". Results are expressed in terms of Island Health service entities, so this knowledge is already explicitly contextualized and pre-translated into a form that speaks directly to quality oversight personnel and decision-makers.

I. Why this analytical focus on "patient journeys"?

- Our analytical "modus operandi" – work from a Target Information Architecture (TIA) to determine how/where derived information products will do work in the context of service system clinical operations, clinical/operational quality oversight, demand estimation, planning, service system redesign and macro-level service system planning (see figure).
- Population-level strategies require targeting to translate from goals/objectives to tactics/plans. This targeting entails two components:
    - People – who will be beneficiaries of improved services.
    - Service Units – the last-mile connecting strategic intent to patient-level services.
- Counting instances of poor processes is essential for making a business case for change and for monitoring outcomes
- However, counts of processes do not tell us how we can improve upon them. For those insights, we need to be able to "see" the processes.
- Pressures bearing down on one program may be the "effect" of causes that are positioned in other locations. The connections between cause and effect are created conjointly by patients and organizational structures and functions. If we want to disentangle those contributing factors, we need to compare and contrast different cohorts interacting with the same service system.
- Quality Assurance/Quality Improvement incident review processes would benefit from more proactive methods that relate patient trajectories [journeys] to likely outcomes.
- Service system capacity shortfalls = waiting. However, need does not go away – it may be displaced. We need tools to relate waiting to altered Service Pathways.
- We should not rely on Emergency Department encounters [and NACRS/CTAS data] plus Acute Care admissions/readmissions [and Discharge Abstracts Database data] as a proxy for full service system utilization. ***We need to bring the full health service system into analytical focus.***
- Missing pieces – the dynamics embodied in patient journeys conveys important information. Repeating cycles that involving one or more services – often associated with deteriorating health status of the patient, conveys two important pieces of information:
    - Something needs to be added to the service mix to produce change.
    - That new service needs to be located somewhere within that reverberating cycle of functionally networked services.
- Right service, non-ideal functional location – services need to be functionally "proximal", e.g., directly accessible, for patients who have difficulty navigating the service system or sustaining engagement. We need tools to characterize the reliability and directness of connections.



II. Navigating the Health Space between One Person's Voice and Populations Speaking at Once

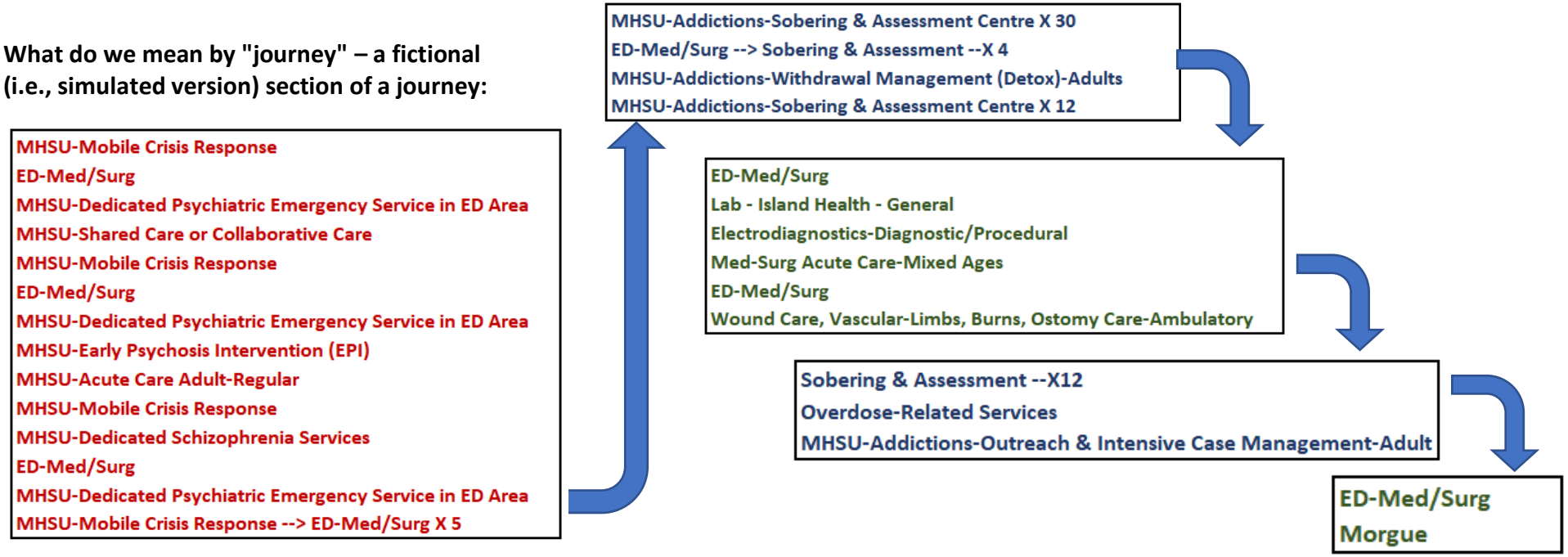
Constituencies Matter for Patient-Oriented Research

- Health service systems "think" and plan in terms of groups of people who have similar need/risk profiles.
- The service system must relate a "patient voice" to one or more constituencies if the results of patient-oriented research are going to span the gap that separates population-level strategies from tactics that speak to the needs of constituencies.
- There's strength in numbers – one patient/client journey can legitimately call upon the system ***not to do* X – *for anyone***. However, the service system needs to hear from groups of similar persons to put in place a program ***to do Y – for a constituency***.
- Service system change must be informed *and* motivated. If we do not have "strong" numbers we cannot create a compelling business case.

Journeys Matter for Patient-Oriented Research – Patients "vote" for service system structures & functions with their service encounter "feet".

- Quality improvement methods (e.g, statistical process control) are suited to ***operations*** within Service Unit – e.g., narrow the range of variation in practice performed by Provider X for Problem Y within Service Unit Z.
- Many of the issues for patients with complex/chronic conditions are reflections of service system ***interoperations*** (or lack thereof) – patient's "falling through the cracks"; or patients cycling through clusters of services until a critical threshold is reached and then Scarce Service Y is provided.
- Outcomes associated with care for complex/chronic problems may be related to emergent characteristics of service system interoperations – the "journey" – as well as what happens when a patient arrives at any particular service unit.

What do we mean by "journey" – a fictional (i.e., simulated version) section of a journey:



III. How we think about "the patient journey" – constituencies that track to similar pathways across the service terrain.

The Context

- ***People experience health challenges – one or more over their lifetime, and they can change.*** ICD-10-CM identifies roughly 69,000 health conditions, though they aggregate up in meaningful ways.
- ***Health problems over time*** are multiply/complexly determined by pathophysiology, health risk/promoting behaviour (proximal determinants), population-health determinants (e.g., vaccination rates within geographies), contextual factors (distal determinants)
- Multiple factors characterize ***health services***, e.g., diagnostic and treatment procedure, "best practices"; program mandates, ***clinical pathways –*** consisting of series of procedures associated with a problem or a routinely-experience set of emergent co-morbidities.
- Multiple factors characterize the ***manner in which health services are deployed*** – ***how the tools are used*** – e.g., supply of services and associated wait-times – which may be tightly or loosely coupled with need and risk; staffing for programs; care delivery models reflecting possibly multiple factors, e.g., supply of services, beliefs around what average or typical people prefer.

The Patient Service Journey

- The Patient Health Service Journey reflects the joint interplay of all of those factors. Importantly, the pathways reflect more than a unilateral effect of a system responding to problems and diagnoses.
- Transdiagnostic context – people navigating across the health service landscape display dynamic characteristics. For example, they have "mass", reflected in impacts when health problems hit the person. They may have "velocity", reflected in the speed at which a problem unfolds under the "gravitational pull" of etiologic factors at work. They may have "momentum" reflecting both "mass" and "velocity", which determines the "service intensity" that must be supplied in order to change trajectories or velocity or impacts. They may experience disparities associated with non-medical factors (e.g., race/ethnicity) or stigma associated with health conditions themselves – these may introduce a "coefficient of friction" into the service access equation.
- There may be one or more chapters to the patient journey, with different interwoven themes and characters. People are never "just X".
- This patient service journey is coupled to the service system journey as it seeks to provide best possible service to populations consisting of people whose journeys are all fundamentally different, but not completely different.

Outcomes

- In an ideal world, all services would be accessible to all persons in a timely fashion. In these cases, the patient journey would distil down to ***Clinical Pathways*** associated with problems. In this sense, Clinical Pathways are "ideal cases" or best practice reference standards for patient journeys.
- In the real world, the patient journey consists of ***Service Pathways*** that emerge as the person, conceived transdiagnostically, interacts with the reality of how the service system attempts to realize best practices under an unyielding set of constraints that makes that impossible.
- For a person experiencing problems over time, **Service Pathways function as the "carrier" for Clinical Pathways**. If we want to understand determinants of outcomes for persons with complex/chronic conditions, we must relate outcomes to those Service Pathways, because real outcomes for real people is not simply the sum of outcomes achieved in relationship to procedures located on clinical pathways.

IV. Methodological and Pragmatic Challenges

Methodology

- If we want to be able to relate outcomes to Service Pathways, we must be able to locate pathways and outcomes in the data.
- The ***Person X Service data space is high-dimensional***, e.g., 69,000 diagnoses, 70,000 procedures x 2450 Service Units (in Island Health) X thousands of medications, lab results X Events ***over time*** – because a person can experience the same service multiple times.
- The space in which Service Pathways "live" is sparsely populated with data – nobody experiences all problems or accesses all services.
- Every patient journey is likely to be unique in this sparse, high-dimensional space.
- However, ***patient journeys are not random sequences of events, and there are similarities between journeys.***
- ***Constituencies*** - how can we identify groups of people who traverse sectors of the service system terrain (sub-journeys) in similar ways?
- ***Cohort-specific patterns of service utilization (PSUs) –*** how can we identify groups of service that may or may not be parked under the same clinical/administrative governance structures but nevertheless interoperate on behalf of patients who share needs/risks.

The Nomenclature Challenge – patient journeys expressed in electronic health records using opaque terminology.

- The journey is characterized by a diverse array of events, e.g., interactions with services (i.e., "encounters"), procedures, prescriptions, etc.
- The base architecture for the "journey" within a health service system reflects the architecture of electronic health records (EHRs) – a long list of Service Units [a functional and physical *location build*]; patient encounters with Service Units; various other pieces of information (e.g., diagnoses; prescriptions, structured/unstructured clinical documentation) associated with those encounters.
- Service Units have names that are known by people within programs but not by others. E.g., "LILAC" (post-withdrawal stabilization unit) is called "LILAC" in the Island Health Implementation of Cerner because that is what people working in LILAC call their program. The program *must* be called "LILAC" in Cerner if clinicians are to put their documentation in the right service location. Who else knows what is "LILAC"?
- The foundational data structure that embodies the Patient Journey expresses that journey in terms that are not decipherable.

The Practical Challenge

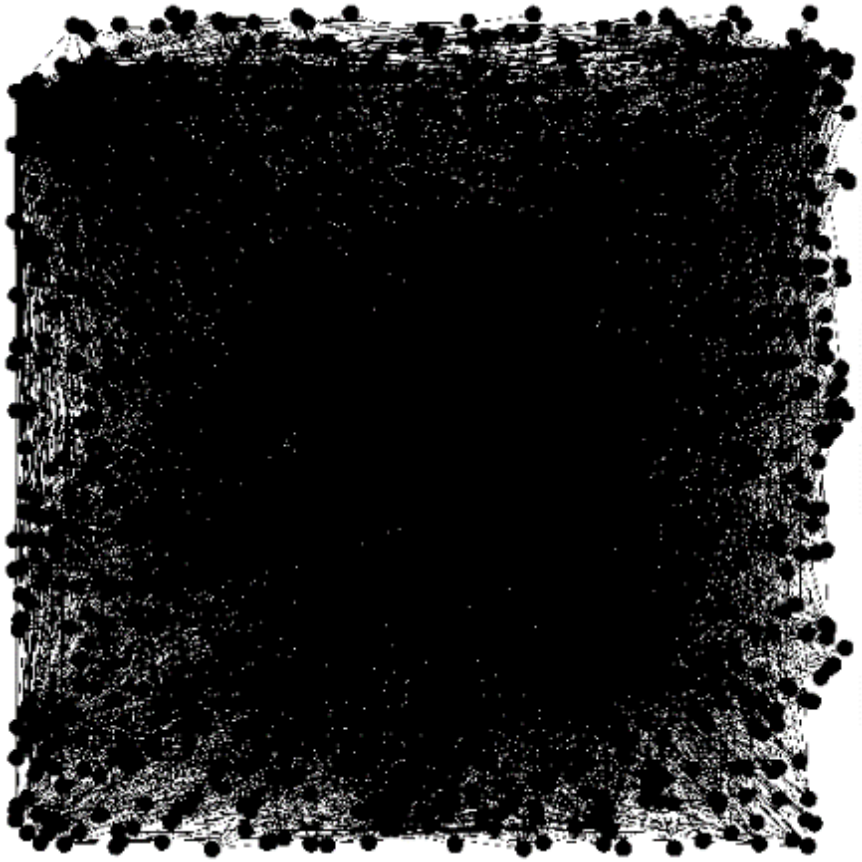
- By clustering people whose journeys are "saying the same thing" about the people (construed transdiagnostically) and the services they access, can we supply information that is sufficiently granular that it can inform ***translation of broad strategy into tactics*** – pointing to the space mid-way between population priorities or broadly-construed service models and individual service units.
- ***Measuring inequities entails measuring events that didn't happen***, e.g., in relationship to non-medical causes or stigmatizing conditions. Can we compare cohorts at the level of Service Pathways to identify disparities?
- Can we identify problematic service system dynamics in an actionable and useful way by focusing in on sections of service pathways – e.g., repeat visits to Emergency Department ***and*** select other services, e.g., for homeless persons.
- Can we identify service access obstacles associated with patient journey pathways that are "etched" into the service system terrain – "you can't get there from here" within this particular service system?
- Can we characterize at a macro-level some features of the service system topology – the overarching clinical "logic" build into the system through the interplay of different patterns of service utilization, e.g., ED-centric service system configurations *vs* an array of linear pathways between services that bypass th ED.

V. Methodology – Foundations
A transparent vocabulary to name Service Units in terms of Functions

- **The Problem** – 2450 Service Units with names like: 1471 Suites; KING; CVH-Emergency Hold; Male Beds; Female Beds; Swing Beds, etc.
  - For some you can make a good guess, e.g., "NRG – Emergency" is probably Nanaimo Regional Hospital, Emergency Department. But even then, there are 21 different Service Units (in the encounter data) with the Unit Name = NRG – Emergency.
  - When we model patient journeys using Service Units to express those journeys, we do not want to leave the reader guessing as to what they are seeing.
- **What is the Clinical Context Coding Scheme**
  - Six sets of codes – service intensity (2 sets), problem, service type, service location (physical), service age range.
  - Code sets applied to all 2450+ Service Units as they are architected in the Enterprise Data Warehouse (we do not "touch" the clinical information system with the codes).
  - This yields approximately 185 equivalence classes – i.e., unique configurations of the six codes. We term these "Service Classes".
- **Before Island Health Clinical Context Coding Scheme (CCCS):**
  - Pathway from *Unit_Name = "LILAC"* to Unit_Name = "*Swing Beds*".
  - What does this mean? Who would you ask? Do you have time to find out?
- **After Island Health CCCS**
  - Pathway from *Service_ Class = "MHSU_Post-Withdrawal Stabilization*" to *Service Class="MHSU-Addictions-Sobering & Assessment Centre"*
  - This means the person went to a Post-Withdrawal Stabilization Unit (LILAC], spent up to 30 days, relapsed, and then ended up in the Sobering Assessment Centre (in a "Swing Bed"), in all likelihood because they ended up homeless after completing their maximum 30 days in "LILAC".
  - This pathway is associated with a cohort or a "constituency" – people who traverses this pathway – people contending with a chronic relapsing substance use disorder.
  - This scheme reduces "spurious granularity" in the raw (prior to CCCS) encounter data. For example, 99 Family Care Homes show up as 99 Service Units (and possibly 99 Service Pathways) as opposed to one single function, like an acute care unit with multiple beds. . In theory the graph machine learning tools could show that these 99 Service Units are a single entity – but with the high-dimensionality of the data and the imprecision inherent in the modeling tools, this is not a realistic expectation – and humans know that the 99 Family Care Homes are relatively homogeneous with respect to CCCS characteristics.

VI. Methodological Challenge: Finding Patterns in Sparse High-Dimensional Encounter Data

Island Health Overdose Cohort – 38,089 "Statements" Configured as 1124 Journeys Through a Space of 21 (out of 185) Classes of Services



Is this one constituency?

Are there understandable Service Pathways "hiding in plain sight"?

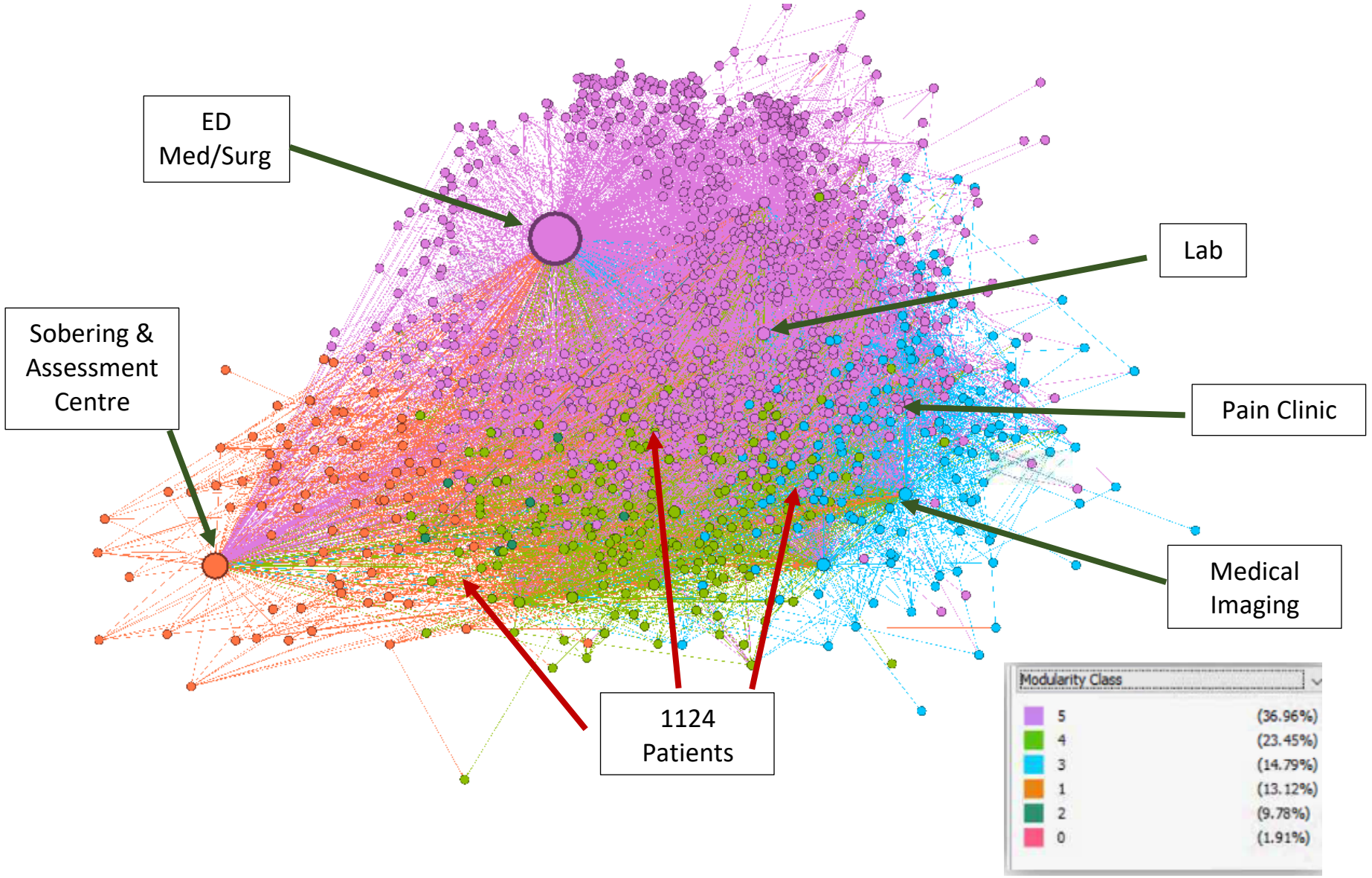
If so – what are they? What are the "take-away messages" for providers or planners?

Dynamics – how much "momentum" is associated with Service Pathways??

What will produce change? What will not produce change?

If the strategic priority is to change what is depicted in this figure, how do we translate that strategy into tactics – based on what we are seeing??

VII. Graph Machine Learning – Community Detection



VIII. Community Detection – Communities of Service and Co-Emergent Communities of Persons– Similar Groups of Patients and Functionally Interoperating Clusters of Services

6 Modularity Class Solution for Overdose Cohort- 2 of 6 Communities

171 Patient 'Gravitating' to this Community of Services;
Average 10.5 Encounters/Person

Service Class	# Encounters
MHSU- Addictions- Clinical Intake Adult	473
MHSU – Clinical Intake – Adult	415
MHSU – Addictions – Clinic-Adult Ambulatory	394
MHSU- Addictions – Withdrawal Management (Detox) Adults	388
MHSU-Addictions – Post-Withdrawal Stabilization – Residential – Adults	227
MHSU- Shared-Care or Collaborative Care	30

239 Patient 'Gravitating' to this Community of Services;
Average 8.4 Encounters/Person

Service Class	# Encounters
Medical Imaging – Diagnostic/Procedural-Ambulatory	822
Orthopedic-Ambulatory-High-Intensity	244
Home & Community Care Services	193
Med/Sur Ambulatory-Episodic, Intermittent, On-Going	135
Surgery- Surgical Day care – Same Day – Acute Care	120
Pain – Ambulatory	113
Electrodiagnostics – Diagnostic/Procedural-Ambulatory	110

We see two distinguishable clusters of services with two groups of patients that "gravitate" towards these clusters – two constituencies.

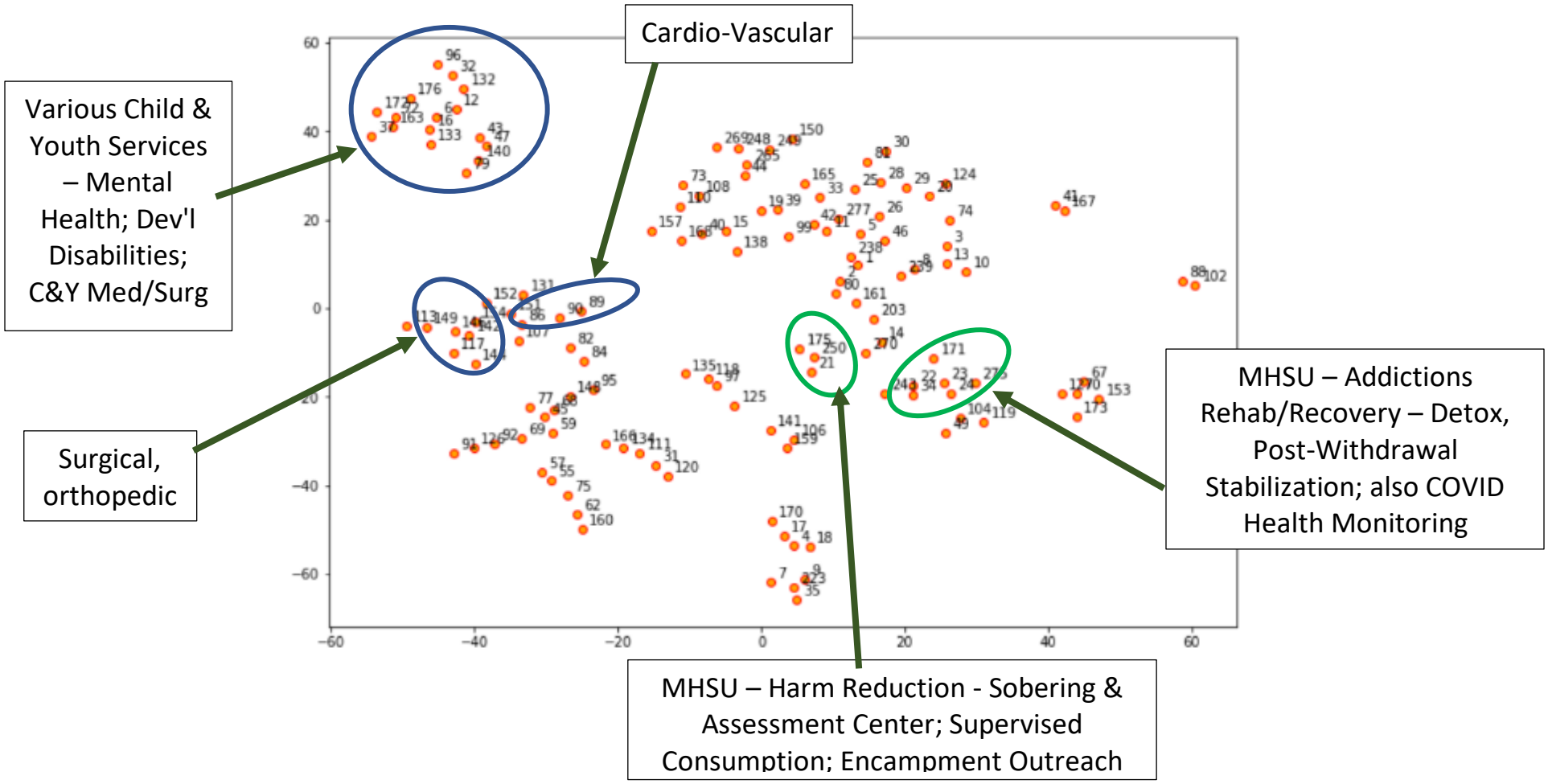
However, it is important to realize that patients may track to multiple Service Pathways entailing access to more than one cluster of services – a person may experience various physical health challenges and be left with difficult-to-manage pain. They may then "cross-over" to a different sector of the service terrain – addictions services.

By that same token, a person could engage with various services for persons with severe chronic psychiatric conditions such as the schizophrenia. They could also trace a service pathway through addictions services.

If they then develop cardiovascular illness (major cause of excess morbidity/mortality for persons with schizophrenia) they may then trace a pathway through the relevant medical/surgical services. However, maybe they will not access those services for persons with cardiovascular problems even if they need those services, because they may be limited in their capacity to negotiate and maintain access – see Bambi, et al., "medically hard-to-house"

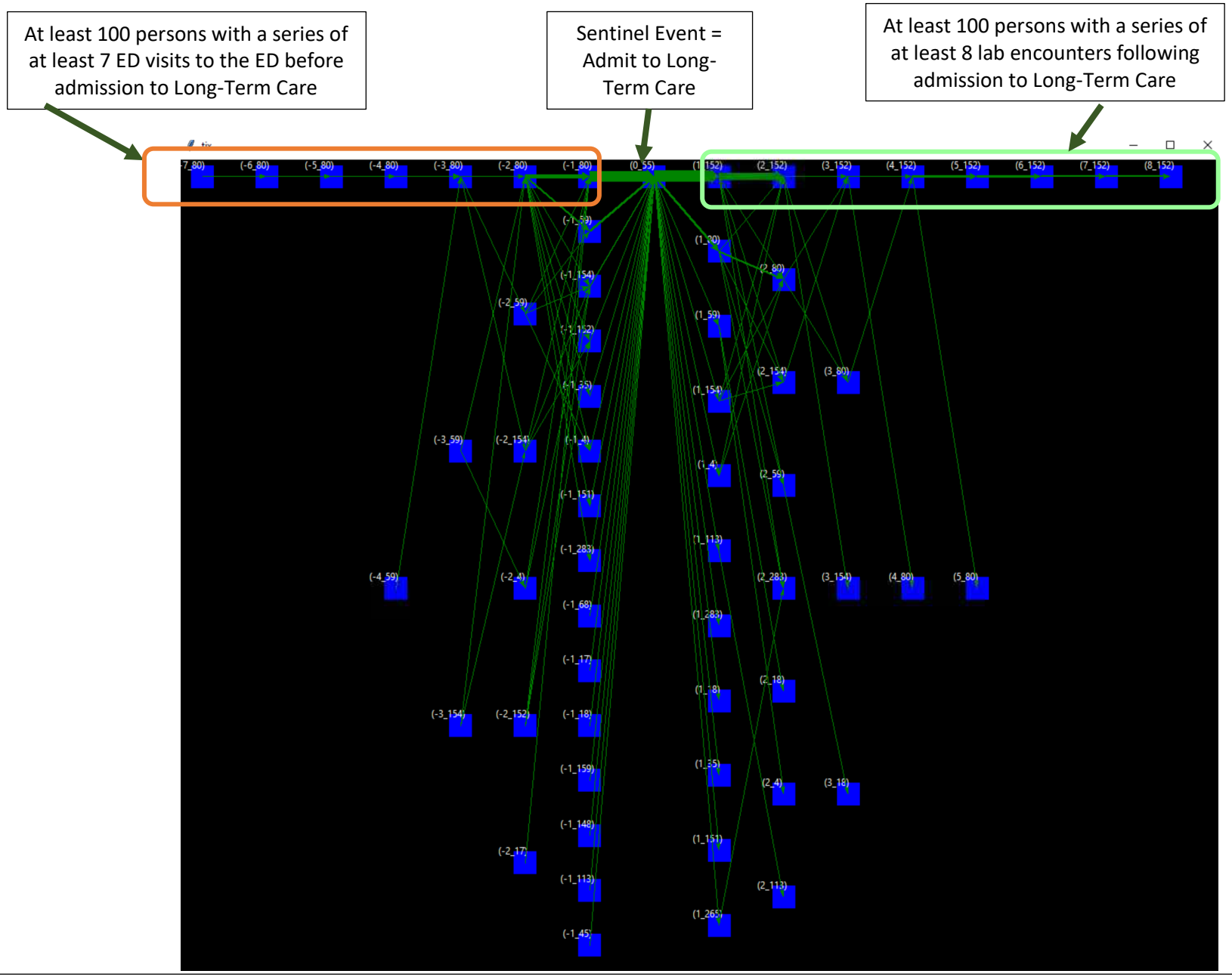
The complete patient journey is in effect the superposition of the patient's interactions with all of the communities generated *via* graph machine learning community detection algorithms applied to encounter data. That clinical wave function may then "collapse" when they access services for a particular problem in a particular location.

IX. "Reality Check" for these Graph Machine Learning Clustering Solutions – Natural Language Processing (NLP)



- Each service encounter is a "token" in longitudinal string of "words". These services are clustered by virtue of proximity to one another across multiple patient service encounter "narratives" – e.g. *via* word2vec algorithm.
- The result is clusters of services that are "connected" by virtue of Service Pathways traced by constituencies who share a history with those services, and presumably they share an interest and need.

X. Hourglass Tool – Sequences of Service Encounters Leading Up To/Following Sentinel Event



At least 100 persons for every encounter sequence in this figure. In the live version, you mouse over a box to see the service class associated with that box.

XII. Conclusions, Next Steps

A Health Service System Macroscope – to locate patterns "hiding in plain sight"

- The work reported made use of unsupervised machine learning tools applied to sparse high-dimensional sets of service encounter data to find patterns.
- The tools we have demonstrated are ***not*** operating within a ***supervised machine learning paradigm***, where we have labelled data generating by humans (e.g., "… extra-axial mass ... at the level of the posterior frontal and anterior parietal lobes") and we want to see if we can train a machine to duplicate this performance, or maybe do it faster.
- The graph machine learning and NLP tools entail ***unsupervised machine learning***. There are no "correct answers". What we are trying to find are patterns in datasets that are so complex, sparse and high-dimensional that humans cannot see the patterns.

Subject Matter Experts ("SMEs"); Service System Operations Experts ("SSOES")

- Because we are working within an unsupervised machine learning paradigm, there is no "fixed truth" to the patterns we are seeking. The tools will yield different clustering of services depending on how parameters are set. This requires a process of sifting through candidate aggregates of services – to determine which clusters "make sense".
    - Recall that with the machine learning tools we employ, communities of service and communities of person ***co-emerge***.
    - What services "go together" because they are keyed directly to problems shared by a cohort defined in terms of a pattern of service utilization? A cluster of services "makes sense" in relationship to a co-emerging cluster of people.
    - What services "don't belong" in that they reflect variable/chance features of persons associated with a pattern of service utilization?
- Various metrics (e.g., in-degrees vs out-degrees) and different algorithms or repeat runs with different starting points can help – to a point. The process of landing on working specifications for ***cohort-specific patterns of service utilization*** requires engagement of:
    - Clinical Subject Matter Experts (SMEs) – people with excellent knowledge of the relationship between problems and required services for a group of diagnoses or problems.
    - "Service System Operations Experts" (what we term "SSOEs") – people who know how components of a given service system interoperate – the pragmatics of service system operations. This may consist of "in this system" or "in any system" knowledge.
- SMEs and SSOEs also address the question: how do depictions of patient journeys and localized processes-of-concern point to need/opportunities for enhancements to service system structure and function.

Next Steps –

- Patient engagement in interpreting/validating cohort-specific patterns of service utilization.
- Proactive and automated case identification for quality assurance/quality improvement reviews – cases selected within cohorts defined in terms of the cross-continuum journeys. The sequence of service events for case-exemplars functions as a roadmap for reviewers.
- Prediction models – associating sentinel events with patterns of service utilization. These events may include person-level events – e.g., overdoses. They may also include predictions/prescriptions for service engagement based on patient similarity.