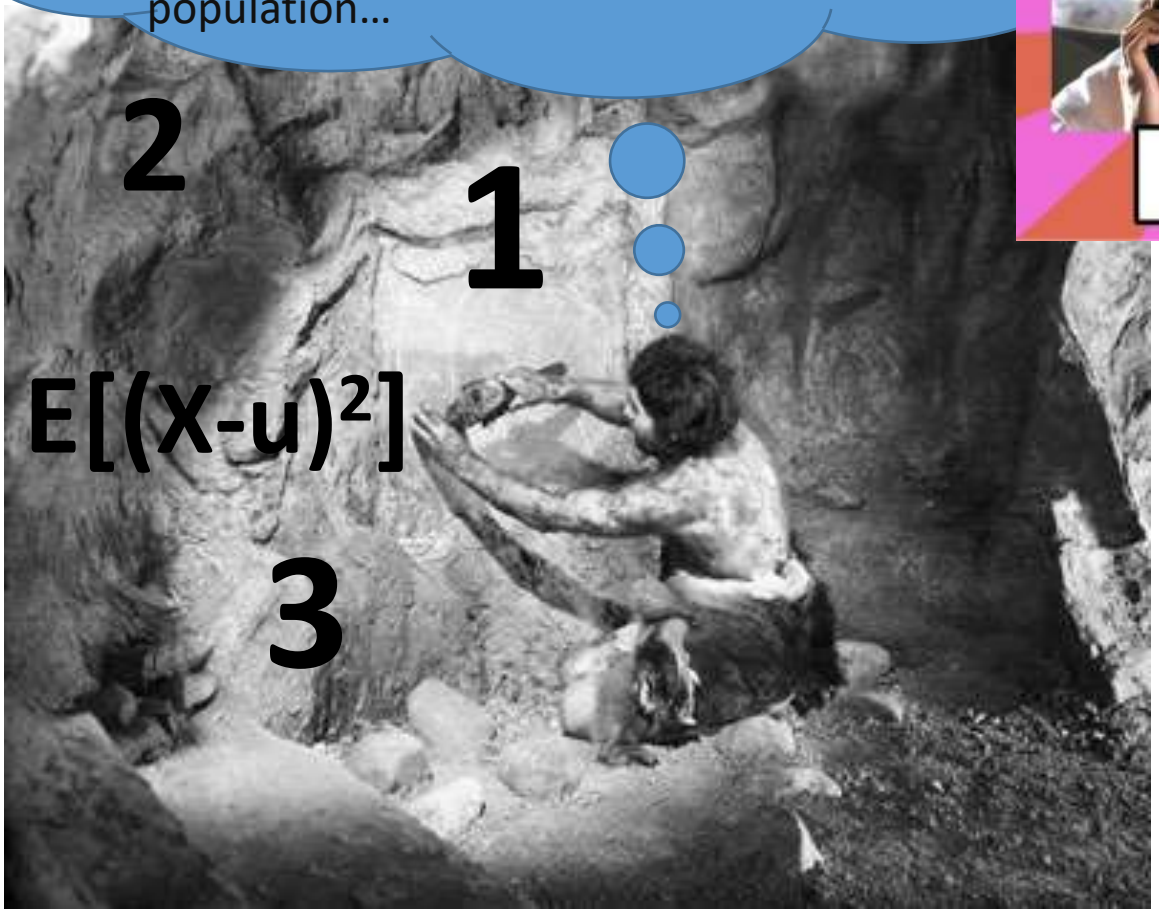
UVIC Psyc 513 Statistical Methods
Statistical Analysis of Administrative Health and Linked Longitudinal Encounter Data

Health Service System – Applied Clinical Research
Some Background, Context

Kenneth A. Moselle, PhD, R.Psych.
Director, Applied Clinical Research Unit
Island Health

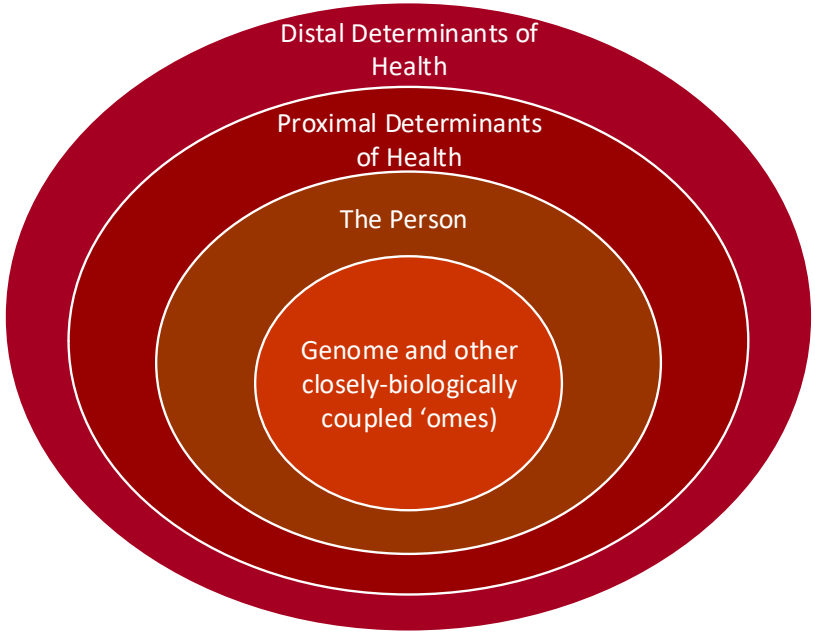
I call them numbers. I just invented them. You can add them, subtract them, multiply them, divide them, extract their integer roots, employ them to quantify scatter or estimate the true mean of the population…



Applied Clinical Research

- What is the problem addressed by the Applied Clinical Research Strategy (and tactics and tools)?
  - Research which is relevant to the providers/programs/systems that deliver healthcare
  - Designs, approaches that do not require years of additional research to render highly generic findings (e.g., products of RCTs) into a form that can be applied to all of those people for whom all other things are NOT equal!
  - How to get useful information out of the EHR (essentially, a clinical data repository and tool for viewing individual patient information) that can be applied to cohorts that interact with the systems of services that are users of the EHR (closed loop architecture)
- What are some of the key challenges –
  - Basic unit of data (encounters – reflecting persons interacting with services) – is essentially indecipherable in the form in which it shows up in the data warehouse
  - Most users of health data have only ever 'touched' administrative data and are not aware of the essentially ways in which those data are DEEPLY DIFFERENT from the data contents and structures that are accumulating (in real time) in the EHR.
- What is the Applied Clinical Research Unit (ACRU) in Island Health
  - Formally – as an entity within Island Health
  - Functionally – a mechanism for connecting data extracted from the EHR with research designs and statistical methodologies
  - Operationally – a key element in an interlocking network of partnerships
- ACRU – conceptual framework
  - From whence do research questions arise?
  - How are the answers put to work
  - Epistemological first principles – to provide clarity on how MUST certain types of questions be addressed, and what can you possibly do/not do with different types of data
  - Pragmatics – data access models (take-out-pizza model vs acquiring a cute little puppy dog model)
- Health Service System – a hierarchically-arranged set of regulatory mechanisms
- Data contents, data layering, data dependencies
- Key Concept – Target Information Architecture

The Person



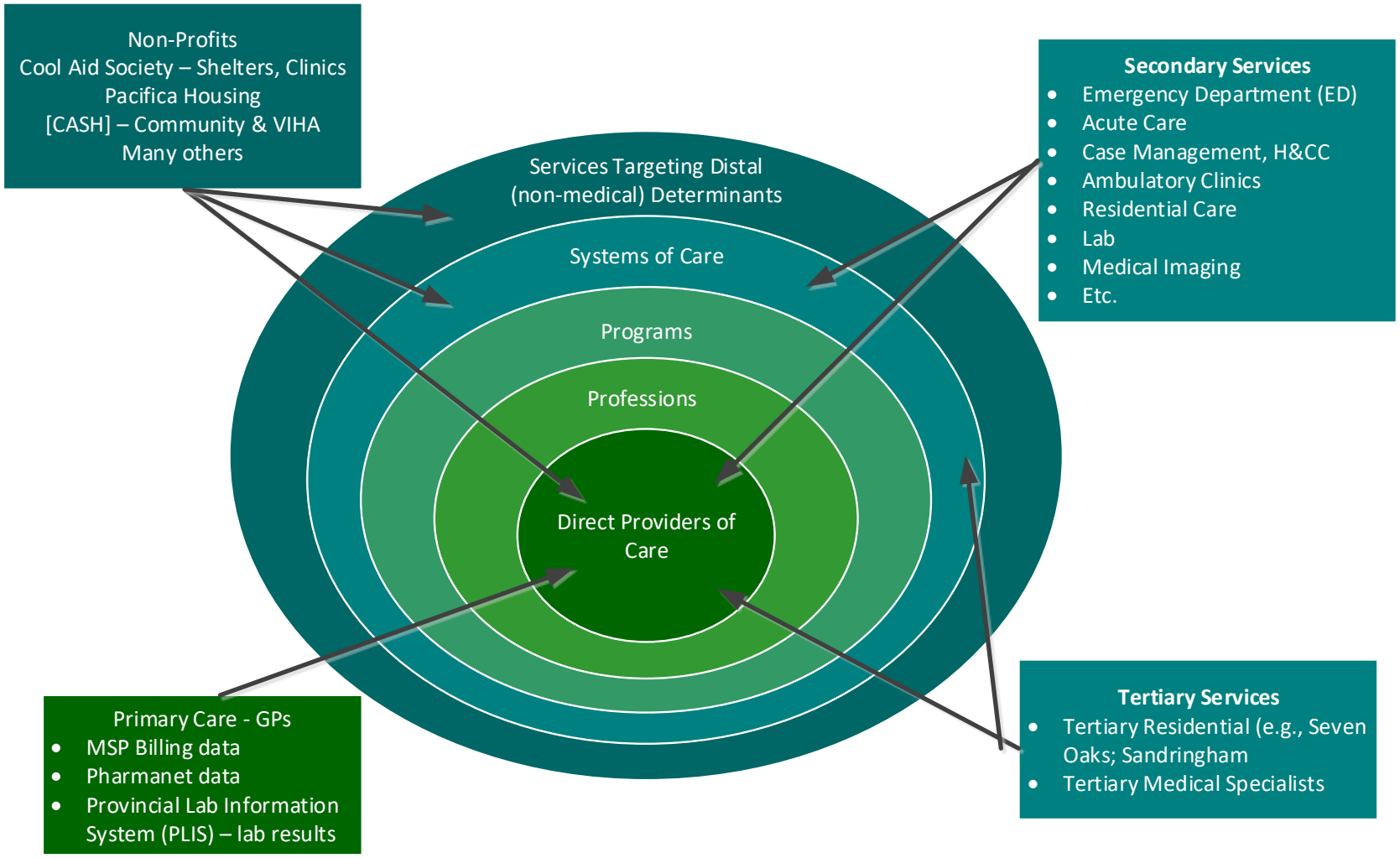
A collection of people who are similar with regard to point-in-time constellations of signs/symptoms (diagnoses) or symptoms-over-time (trajectories) and/or etiology and/or response to treatment = COHORT of interest. Ideally (for ACRU) that cohort will display some characteristic properties:

- Identifiable using data available within the health service system's data holdings
- Based on case definitions that are meaningful to the people who will be providing services (i.e., not an *ad hoc* grouping that exists ***purely*** as a statistically-derived entity)
- Relatively homogeneous with respect to service requirements and services provided
- Relatively homogeneous with respect to response to treatment
- Relatively homogeneous with respect to trajectory – or at least portions of trajectory

A not-parenthetical aside – so what does this have to do with research questions/design, or lets rephrase the question – Is "person-centred" more than just a feel-good term??

- The basic paradigm becomes Persons-with-Attributes, which we contrast with the paradigm Attributes-without-the-Person (e.g,. remove the effects of the person using RCT designs, in effect reducing the person to the Attribute).
- E.g., A study of distinctive features of inner speech in persons with schizophrenia – vs a study of persons with schizophrenia contending with a distinctive inner-speech profile, or
- A study of Type II diabetes and associated co-morbidities vs a study of persons with Type II diabetes and distinctive profiles (sub-types) characterized by emergent co-morbidities, or
- A study of relationships between phonological awareness and reading capacity by end of Grade 2 vs a comparative study of persons who cannot read by the end of Grade 2 due to primary [and enduring] deficits in phonological awareness vs less entrenched "getting-their-foot-in-the-reading-door" deficits in phonologically-based decoding secondary to a psycholinguistic-reading-guessing-game approach to reading, a manifestation of a more primary dysexecutive syndrome.
- With focal, bounded conditions – where the diagnosis and the episode are one and the same – an Attribute focus may work (e.g. if my attribute is ICD10 S62.602 – Fracture of unspecified phalanx of right middle finger).
- But with chronic disease, this approach is inherently limited – not necessarily in its validity, but in its usefulness. Consider for example a person with F60.81 plus F 60.2 [look them up!] or 296.4 (Bipolar manic spectrum) or E11.2 (Type 2 diabetes mellitus with kidney complications)

Providers

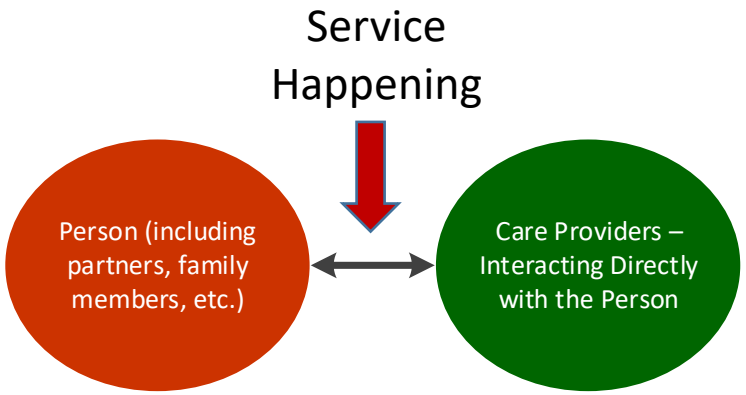

Non-Profits
Cool Aid Society – Shelters, Clinics
Pacifica Housing
[CASH] – Community & VIHA
Many others
Services Targeting Distal (non-medical) Determinants
Systems of Care
Programs
Professions
Direct Providers of Care
Secondary Services
Emergency Department (ED)
Acute Care
Case Management, H&CC
Ambulatory Clinics
Residential Care
Lab
Medical Imaging
Etc.
Primary Care - GPs
MSP Billing data
Pharmanet data
Provincial Lab Information System (PLIS) – lab results
Tertiary Services
Tertiary Residential (e.g., Seven Oaks; Sandringham
Tertiary Medical Specialists

Services enacted in the context of the interaction of Persons with Providers

- Note that I did not say “Services provided” – that may apply for surgical patients but it does not reflect what happens with people contending with chronic health concerns (including aging).
- Nevertheless, not everyone does everything, and not all problems are treated the same, so we have to grapple with the question of explaining what we mean by “Services” and “Service Delivery Models” (and outcomes associated with services….)
- Direct approach to defining “services” – ICD 9/10 Procedure Codes, e.g,. *DD010ZZ Beam Radiation of Stomach using Photons < 1MeV*

- Indirect – using programs or providers as proxies, e.g.,
  - If you were seen in the Cast Clinic then somebody probably did something involving a Cast
  - If you were seen by the Seniors Outreach Team then somebody did a comprehensive assessment (and the team probably also CAREFULLY screened the referral, so if they DID see a person then we can reasonably infer that the person seen was a frail elderly person, living at home but, in all likelihood, not receiving the level of care they require, or at risk for something bad if the level of care is not increased at some point in the foreseeable future)
  - If you were seen by a Case Manager in Mental Health & Substance Use, then they did at least one of the things that Case Managers do when they see their clients, e.g., ‘head to toe’ assessment of clinical/functional/ behavioural risk status; check on any medication-related issues; any housing-related issues, etc.
  - Using relationship-with-a-program as a proxy for ‘direct’ tracking and documentation of services/procedures is the method employed by the BC Ministry of Health Minimum Reporting Requirements (MRR) for Mental Health & Substance Use.

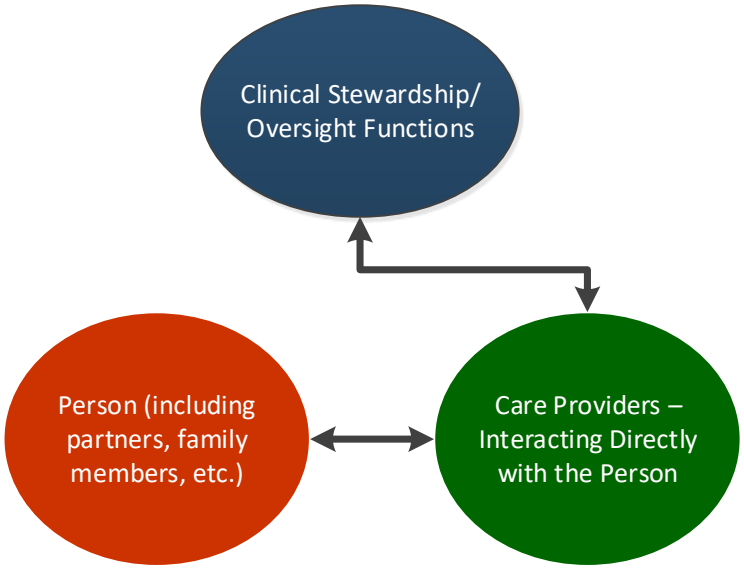
Foundation Layer - Person in Relationship with Providers – dynamic/real-time generation of events



Some characteristics of this relationship:

- Typically initiated by the Person – so this 'skews' the interaction in the direction of reactive care
- Information-dependent but not information-driven
- Two basic logics
  - Do Protocol X because of Y (X is indicated by Y) – De-Personalized Protocol-Based Care
  - Don't do Protocol X despite Y (X is contra-indicated by something(s) other than Y) – Personalized Care

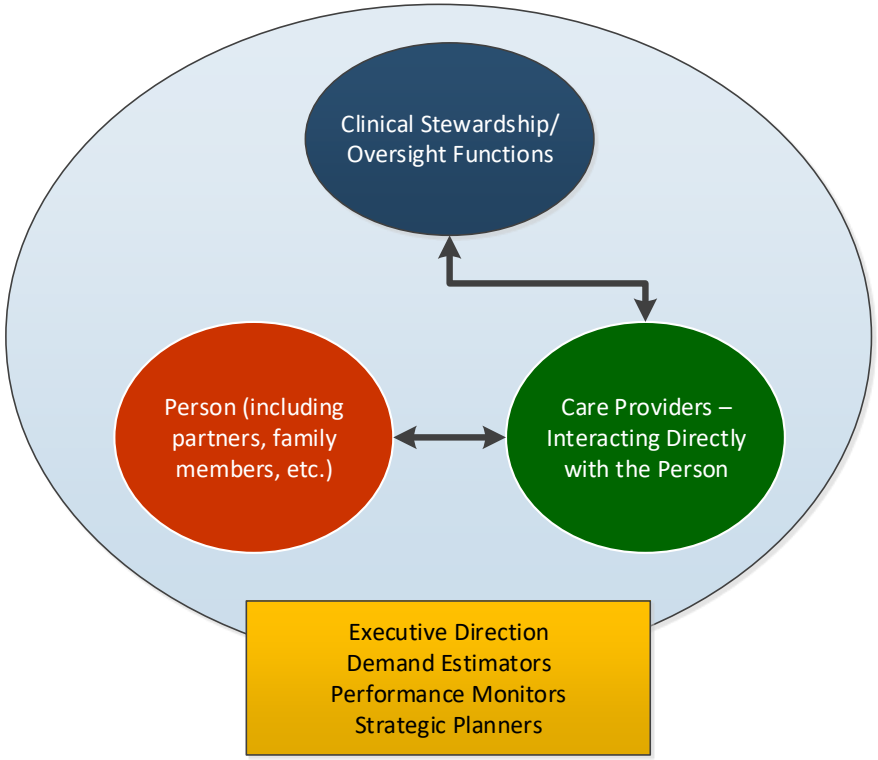
One layer ‘up’ – Clinical Oversight/Stewardship – information-product-based feedback, feedforward



Some characteristics of this relationship:

- Initiated by Provider or Clinical Steward
- Real-time functions – to support care of individual patients – working through intermediary of most responsible provider
- “Offline functions” - aggregate data views – e.g., is practice aligned with standards in those cases where the standards apply? And – to those standards engender expected outcomes (are the protocols ‘correct’)
- Two modes of action
  - Quality Assurance – Do X consistently when indicated by Y and not contra-indicated by anything else
  - Quality Improvement – Do W consistently when indicated by Y and not-contra-indicated by anything else – even though X was the preferred option at some point in the past

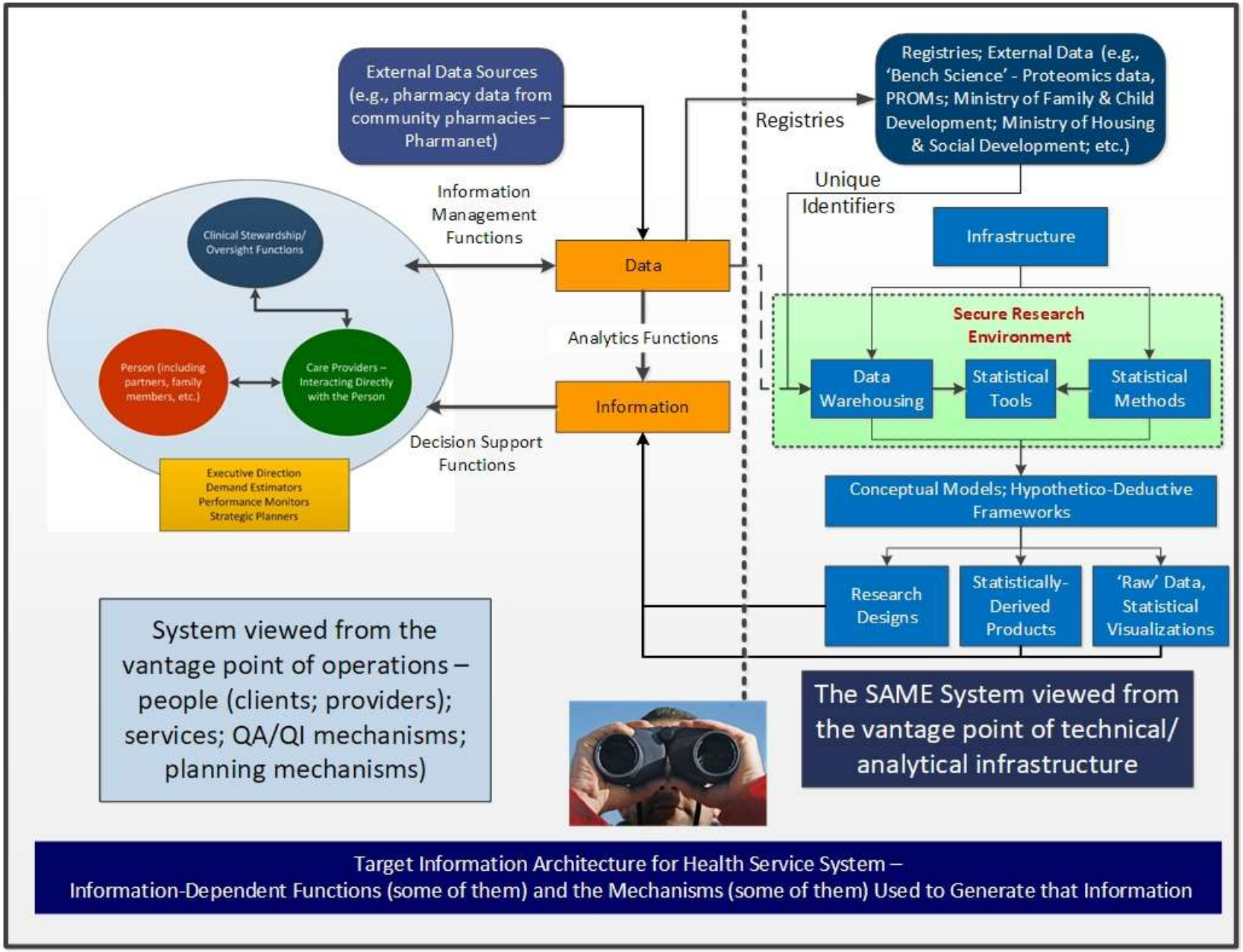
One further layer ‘up’ – Executive/Planning Functions



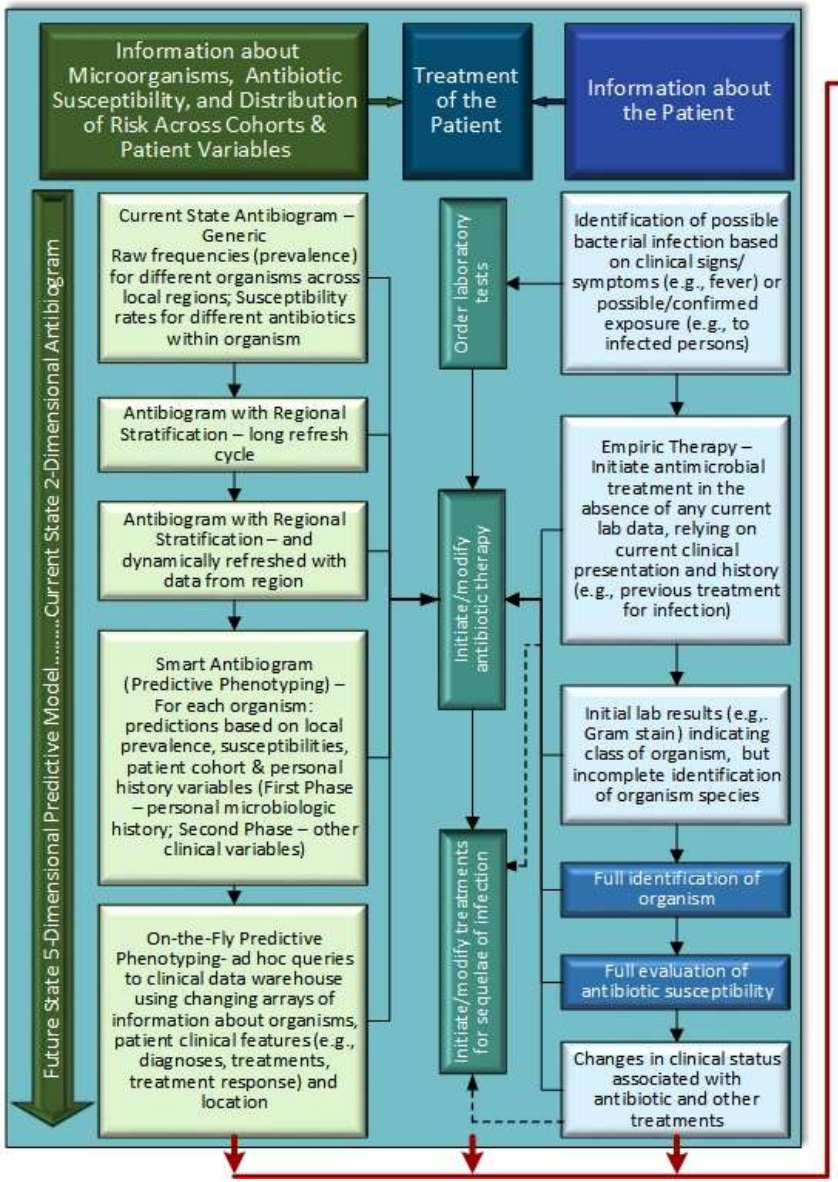
Some characteristics of this relationship:

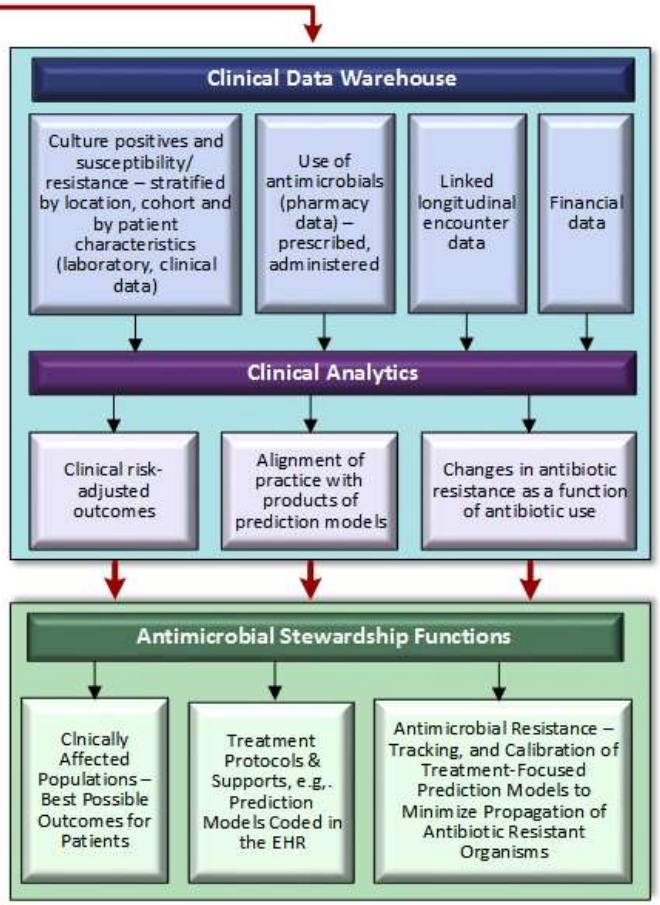
- Population-based – quality in relationship to at-risk or clinically impacted fractions of defined groups
- Information is only as good as underlying case definitions
- In the absence of direct measures of risk, need & outcomes – must live on a steady diet of proxies (which by definition contain unmeasured or uncontrolled sources of variance compared to direct measures)
- [At least] three modes of action
  - Set population priorities
  - Set funding levels, priorities
  - Promote service models (which will not be more granular than the population-views to which they refer – so very generic, e.g., “primary care integration”)

Information Enablers – Very Abstract Rendering of "Target Information Architecture" for Healthcare



Target Information Architecture – Real Example – Prediction Models in Service of Best Possible Treatment of Infections – and one 'layer' up – in service of antimicrobial stewardship agenda





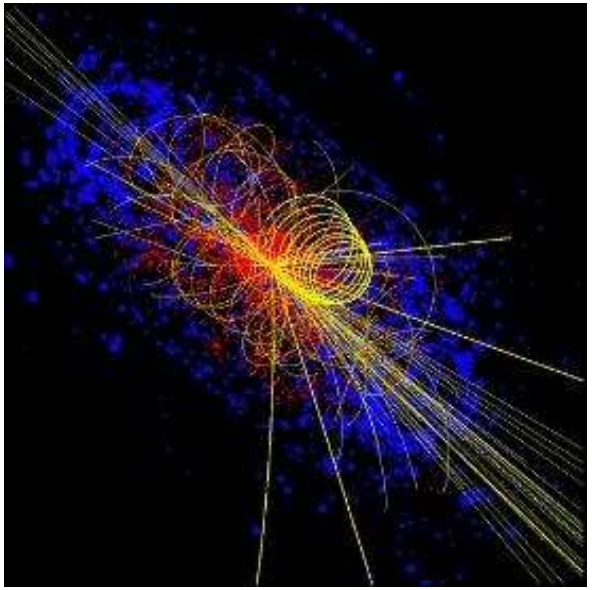
Smart Antibiogram Within the Context of Antimicrobial Stewardship

Island Health

June 1, 2015

People, Providers, Processes – Viewed Through Data Set Lenses

or



Traces Left by Transactions in the Data Warehouse

A Working Health Conceptual Data Model [don't worry – we'll unpack this over the course of several more slides]

Transactional Layer – the Sensory-Perceptual Layer for an Intelligent Health Service System
- Encounters – with Persons/ Professions/ Programs
- Procedures
- Prescriptions
- Lab results
- Expenditures

Attributes of Persons – sort of 'directly' measured - i.e., some significant element of inference or judgment on the part of providers
- Clinical Diagnoses (point-in-time)
- Functional, Behaviour Risk Profile
- 'Trajectories' at a visual-perceptual level (i.e., "eyeballing" the longitudinal record)

Attributes of Persons – Abstracted – i.e., Administrative Data sets
- Diagnoses – Most Responsible; Secondary, etc. (Discharge Abstracts Database)
- Procedures (Discharge Abstracts Database)
- Minimum Reporting Requirements – Home & Community Care, Residential Care (RAI)
- Minimum Reporting Requirements – Mental Health & Substance Use
- [These are keyed to some extent to transactions but the nomenclature and content scope is determined by users external to the service system – not to information requirements of patients/providers]
- Various provincial registries

Attributes of Persons – by analytical inference/validation
- Transactional Trajectories – aggregate views – common profiles of quantities of interaction with key services or programs, e.g,. Repeat users of ED
- Transactional Trajectories – longitudinal – e.g., characteristic patterns of interaction with a variety of services/programs over time
- Clinical Trajectories (e.g., characteristic patterns of emergent co-morbidities
- Common determinants profile (e.g,. Decision tree analysis)
- Common risk profile (quantified, e.g, as per prediction models)
- Group/cohort membership, profile similarity (e.g., cluster analysis)

Case Definitions; Case Costing, Episodes of Care
- Case definition based on single point-in-time diagnosis
- Case definition based on diagnostic profile over time
- Case definition based on transactions with the health service system
- Episodes – collection of services keyed to a case definition based on some combinations of all of the above
- Costing based on single encounters or single procedure (e.g., cost of an acute care day; cost of an MRI)
- Case costing based on an "episode of care" - e.g., all services associated with a person who fits a case definition
- Outcome-adjusted case costing – what does it cost to bias odds in favour of best possible outcomes for episodes of care associated with case definitions

Outcomes – Person Level; What is 'Best Practice'???
- Changes in clinical/functional/behaviour risk profile over time
- Transactional Trajectory-based – changes in slopes; inflection points
- Efficiency – wait times; time to access required services; outcomes associated with delays
- Cost-effectiveness – per procedure
- Cost to bias odds in favour of best possible outcomes
- Cost-effectiveness – per clinical condition [over time]

Quality – Practice in Relationship to Standards – in Relationship to Outcomes
- For a cohort associated with a protocol/service model – are services aligned with standards as embodied by clinical practice guidelines/protocols/service models?
- Practice in relationship to standards – stratified by person, by profession, by program
- Outcomes associated with practice that conforms/does not conform to standards – do the outcome data support the protocols?

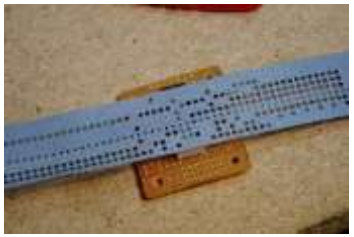
Quality – Practice in Relationship to Standards – in Relationship to Population Risks/Needs
- Incidence, prevalence – need, risk at a population level
- Population access to protocols that have a demonstrated relationship to better outcomes
- Proximal, distal determinants of service access
- Costs associated with population access to demonstrated better practices
- Costs associated with lack of access to better practices (pay now/pay later scenarios)

Evidence-Based Strategic Planning, Performance Monitoring
- Multivariate models that describe the operation of multiple components of the service system that are associated with service models that are associated with outcomes and costs
- Conjoint optimization of service system operation on the basis of population reach, cost and outcomes
- Strategic plans pre-validated via scenario testing using models
- Performance monitoring for strategic plan – based on indicators that have been validated against measures of service system operations and population reach/ outcomes

Point of Service – the Mission-Critical Data Source ⟷ Analytically-Derived Data Entities ⟷ Evidence-Based Quality Assurance, Quality Improvement, Strategic Planning

Events in spacetime are the real stuff of nature, unique and independent of any particular [*administrative data*] frame, whereas the [*administrative data*] frames are arbitrary human constructs that have no fundamental existence. There is no preferred or fundamental [*administrative data*] frame in Nature.

Foundation Layer – Health-Issue-Focused Transactions (and associated expenditures)



Think "VERY long paper tape with a bunch of events marked onto it – in a form that is deeply different from flattened, rectangular data files with one row of data per person and NO MISSING VALUES!!!!" In a sense – there is no missing data in a transactional record – because, unlike an administrative data set, there is no fixed structure [frame] from which something could be missing.

Transactional Layer – the Sensory-Perceptual Layer for an Intelligent Health Service System

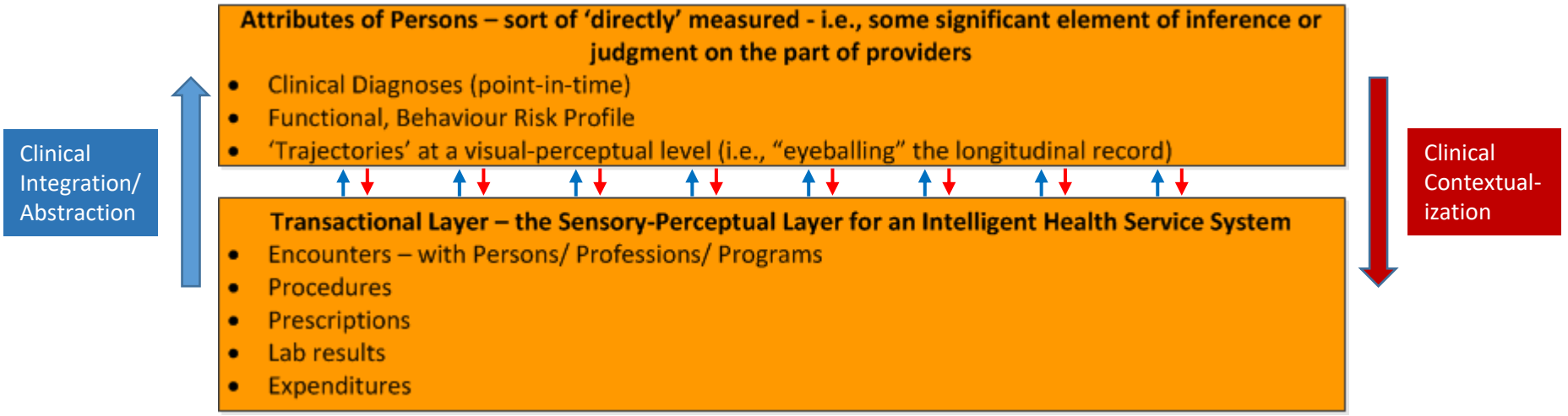
- Encounters – with Persons/ Professions/ Programs
- Procedures
- Prescriptions
- Lab results
- Expenditures
- Program expenditures linked to locations where patients/clients interact with providers

RE: Epistemological foundations of clinically actionable knowledge: do not forget the words of Heinz Werner (as in Werner & Kaplan – Symbol Formation – a classic text):

"*Out of nothing shall not come something*"

The transactional layer is ultimately the "something" out of which "clinically actionable knowledge" emerges.

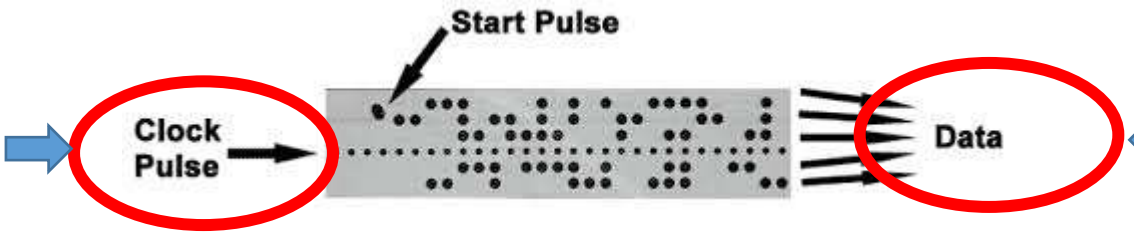
Clinical/Transactional Layers – Attributes of Persons (and Services) Layered onto Transactions



All health information that ties directly to people contending with health issues and accessing services arises from these transactions. The data arising from these transactions takes two forms:

1. The transactions themselves, e.g., "5 or more ED visits in a calendar year"
2. Information abstracted from these transactions and 'deposited' into ***administrative*** data sets (e.g., Discharge Abstracts Database) or ***registries.*** These are ***not*** clinical records and the source data will be transformed, in some fashion, in order to conform to the data structures and nomenclature standards of the administrative data set entities.

Repeat after me: "Within-person-over-TIME"



Transactional Data are inherently longitudinal

Data – by Abstraction – Administrative Data Sets

Attributes of Persons – Abstracted – i.e., Administrative Data sets

- Diagnoses – Most Responsible; Secondary, etc. (Discharge Abstracts Database)
- Procedures (Discharge Abstracts Database)
- Minimum Reporting Requirements – Home & Community Care, Residential Care (RAI)
- Minimum Reporting Requirements – Mental Health & Substance Use
- [These are keyed to some extent to transactions but the nomenclature and content scope is determined by users external to the service system – not to information requirements of patients/providers]
- Various provincial registries

Attributes of Persons – sort of 'directly' measured - i.e., some significant element of inference or judgment on the part of providers

- Clinical Diagnoses (point-in-time)
- Functional, Behaviour Risk Profile
- 'Trajectories' at a visual-perceptual level (i.e., "eyeballing" the longitudinal record)

Transactional Layer – the Sensory-Perceptual Layer for an Intelligent Health Service System

- Encounters – with Persons/ Professions/ Programs
- Procedures
- Prescriptions
- Lab results
- Expenditures
- Program expenditures linked to locations where patients/clients interact with providers

Note:

1. These are ***not*** clinical records and the source data will be transformed, in some fashion, in order to conform to the data structures and nomenclature standards of the administrative data set entities.
2. It is VERY MISLEADING to refer to transactional data coupled directly to the provision of care as "Administrative data" even if it is being used for non-clinical/administrative purposes.

Data – Analytically/Statistically Derived Entities – Cases, and that most difficult-to-operationalize construct: episodes of care

Case Definitions; Case Costing, Episodes of Care

- Case definition based on single point-in-time diagnosis
- Case definition based on diagnostic profile over time
- Case definition based on transactions with the health service system
- Episodes – collection of services keyed to a case definition based on some combinations of all of the above
- Costing based on single encounters or single procedure (e.g., cost of an acute care day; cost of an MRI)
- Case costing based on an "episode of care" - e.g., all services associated with a person who fits a case definition
- Outcome-adjusted case costing – what does it cost to bias odds in favour of best possible outcomes for episodes of care associated with case definitions

Attributes of Persons – by analytical inference/validation

- Transactional Trajectories – aggregate views – common profiles of quantities of interaction with key services or programs, e.g,. Repeat users of ED
- Transactional Trajectories – longitudinal – e.g., characteristic patterns of interaction with a variety of services/programs over time
- Clinical Trajectories (e.g., characteristic patterns of emergent co-morbidities
- Common determinants profile (e.g,. Decision tree analysis)
- Common risk profile (quantified, e.g, as per prediction models)
- Group/cohort membership, profile similarity (e.g., cluster analysis)

Attributes of Persons – Abstracted – i.e., Administrative Data sets

- Diagnoses – Most Responsible; Secondary, etc. (Discharge Abstracts Database)
- Procedures (Discharge Abstracts Database)
- Minimum Reporting Requirements – Home & Community Care, Residential Care (RAI)
- Minimum Reporting Requirements – Mental Health & Substance Use
- [These are keyed to some extent to transactions but the nomenclature and content scope is determined by users external to the service system – not to information requirements of patients/providers]
- Various provincial registries

Data – Analytically/Statistically Derived Entities – II – Outcomes; Outcome-Adjusted Case Costing (a methodologically lofty measurement objective)

Outcomes – Person Level; What is 'Best Practice'???

- Changes in clinical/functional/behaviour risk profile over time
- Transactional Trajectory-based – changes in slopes; inflection points
- Efficiency – wait times; time to access required services; outcomes associated with delays
- Cost-effectiveness – per procedure
- Cost to bias odds in favour of best possible outcomes
- Cost-effectiveness – per clinical condition [over time]

Case Definitions; Case Costing, Episodes of Care

- Case definition based on single point-in-time diagnosis
- Case definition based on diagnostic profile over time
- Case definition based on transactions with the health service system
- Episodes – collection of services keyed to a case definition based on some combinations of all of the above
- Costing based on single encounters or single procedure (e.g., cost of an acute care day; cost of an MRI)
- Case costing based on an "episode of care" - e.g., all services associated with a person who fits a case definition
- Outcome-adjusted case costing – what does it cost to bias odds in favour of best possible outcomes for episodes of care associated with case definitions

Attributes of Persons – by analytical inference/validation

- Transactional Trajectories – aggregate views – common profiles of quantities of interaction with key services or programs, e.g,. Repeat users of ED
- Transactional Trajectories – longitudinal – e.g., characteristic patterns of interaction with a variety of services/programs over time
- Clinical Trajectories (e.g., characteristic patterns of emergent co-morbidities
- Common determinants profile (e.g,. Decision tree analysis)
- Common risk profile (quantified, e.g, as per prediction models)
- Group/cohort membership, profile similarity (e.g., cluster analysis)

Data Roll-Ups – Practice in Relationship to Standards – in Relationship to Outcomes – at a Program Level (Quality); at a Population Level (Surveillance)

Quality – Practice in Relationship to Standards – in Relationship to Population Risks/Needs

- Incidence, prevalence – need, risk at a population level
- Population access to protocols that have a demonstrated relationship to better outcomes
- Proximal, distal determinants of service access
- Costs associated with population access to demonstrated better practices
- Costs associated with lack of access to better practices (pay now/pay later scenarios)

Quality – Practice in Relationship to Standards – in Relationship to Outcomes

- For a cohort associated with a protocol/service model – are services aligned with standards as embodied by clinical practice guidelines/protocols/service models?
- Practice in relationship to standards – stratified by person, by profession, by program
- Outcomes associated with practice that conforms/does not conform to standards – do the outcome data support the protocols?

Outcomes – Person Level; What is 'Best Practice'???

- Changes in clinical/functional/behaviour risk profile over time
- Transactional Trajectory-based – changes in slopes; inflection points
- Efficiency – wait times; time to access required services; outcomes associated with delays
- Cost-effectiveness – per procedure
- Cost to bias odds in favour of best possible outcomes
- Cost-effectiveness – per clinical condition [over time]

Case Definitions; Case Costing, Episodes of Care

- Case definition based on single point-in-time diagnosis
- Case definition based on diagnostic profile over time
- Case definition based on transactions with the health service system
- Episodes – collection of services keyed to a case definition based on some combinations of all of the above
- Costing based on single encounters or single procedure (e.g., cost of an acute care day; cost of an MRI)
- Case costing based on an "episode of care" - e.g., all services associated with a person who fits a case definition
- Outcome-adjusted case costing – what does it cost to bias odds in favour of best possible outcomes for episodes of care associated with case definitions

Roll-Ups – Operations Research-Type Models – Scenario-Testing; Pre-Validation of Strategies

Evidence-Based Strategic Planning, Performance Monitoring

- Multivariate models that describe the operation of multiple components of the service system that are associated with service models that are associated with outcomes and costs
- Conjoint optimization of service system operation on the basis of population reach, cost and outcomes
- Strategic plans pre-validated via scenario testing using models
- Performance monitoring for strategic plan – based on indicators that have been validated against measures of service system operations and population reach/ outcomes

Quality – Practice in Relationship to Standards – in Relationship to Population Risks/Needs

- Incidence, prevalence – need, risk at a population level
- Population access to protocols that have a demonstrated relationship to better outcomes
- Proximal, distal determinants of service access
- Costs associated with population access to demonstrated better practices
- Costs associated with lack of access to better practices (pay now/pay later scenarios)

Quality – Practice in Relationship to Standards – in Relationship to Outcomes

- For a cohort associated with a protocol/service model – are services aligned with standards as embodied by clinical practice guidelines/protocols/service models?
- Practice in relationship to standards – stratified by person, by profession, by program
- Outcomes associated with practice that conforms/does not conform to standards – do the outcome data support the protocols?

Outcomes – Person Level; What is 'Best Practice'???

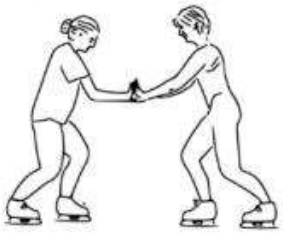
- Changes in clinical/functional/behaviour risk profile over time
- Transactional Trajectory-based – changes in slopes; inflection points
- Efficiency – wait times; time to access required services; outcomes associated with delays
- Cost-effectiveness – per procedure
- Cost to bias odds in favour of best possible outcomes
- Cost-effectiveness – per clinical condition [over time]

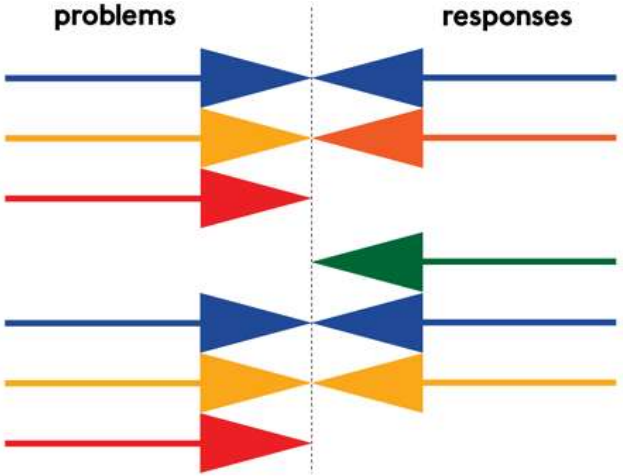
Epistemological Foundations

- What do we mean by “personalized healthcare”?
- You can/cannot get there from here – where MUST at least some portion of the data come from if the products derived statistically from the data can POSSIBLY serve the personalized healthcare agenda
- Data dependencies – e.g., how can you estimate the cost to bias odds in favour of a best-possible clinical outcome if you can’t measure a clinical outcome?
- Statistical/analytical dependencies - e.g.,
  - How can you measure change or relationships among variables using descriptive statistics?
  - What kind of designs do you require if you want to approximate the results of an RCT using ‘naturally-occurring’ experiments using data abstracted from the EHR??
- Information architecture/layering - is there more to optimized operation of the health service system than information infrastructure and products in service of personalized healthcare?
  - Sensory-perceptual level (primary projection area in Luria’s model)
  - Surveillance layer – the sensory-perceptual integration layer (secondary association area in Luria’s model)
  - Etc.
- The Law of Requisite Variety – foundational to a health service system that learns from experience and functions intelligently/ adaptively in response to [population] health challenges
  - What kind of data can be used to support the development of service models, strategies that will successfully make the [return] journey back to the point of service.
  - What happens when data that are NOT abstracted/extracted from the point-of-service transactional records are applied to the point of service (or what happens when attempts are made to ‘translate’ such contents back to the point of service?

Law of Requisite Variety

- W. Ross Ashby (1903-1972):
  If a system is to be stable, the number of states of its control mechanism must be greater than or equal to the number of states in the system being controlled.





insufficient (or incorrect) variety in responses to deal with variety of the problems

The Law of Requisite Variety

- In Cybernetics, there is basically one Law that relates to the governance of System X by information-based Control System Y - it is W.Ross Ashby's Law of Requisite Variety.

- It states, in effect, that if System X is to remain stable, and act in a predictable manner in the face of inputs from an environment, the number of states in Control System Y must be greater than or equal to the number of states in System X – the system being controlled.

- In slightly different terms: the Control System Y must operate on the basis of a MODEL that accurately describes all of the actions that System X can take in the face of all likely inputs.

- Or: there must be at least as many variations of control to be applied as there are ways for the system to get out of control.

Law of Requisite Variety – 'Penal Code' for a Layered Health Service System that aspires to systems-level intelligence

- Within Healthcare, the mechanisms that are involved in regulating and planning the operation of the components of the system, and the configuration of the system as a whole- MUST know what the clinicians and patients/ clients know about need, risk, interventions and outcomes. YOU ***CANNOT*** FINESSE AROUND THIS REQUIREMENT BY EMPLOYING DATA PROXIES, OR BY RELYING ENTIRELY ON ADMINISTRATIVE DATA ENTITIES.
- Should you try, the consequences may include:
  - Nothing – strategy, planning that does not produce change
  - Pathological inertia:
    - Perceiving the need to do nothing when there are really good reasons for doing something
    - Perceiving the need to do something when there are no clear reasons for departures from the status quo
  - Unanticipated (possibly unwelcome) changes in response to interventions into operations
  - Poorly comprehended changes (if your models do not reflect underlying operations, they cannot be used to understand the nature/extent of impacts of interventions on operations)
  - Poorly regulated processes of transformation
  - Inefficient use of resources

Okay – so now what's the big challenge?

Program Manager: "I would like to know if we can make a case for setting up a 'wet housing' program. And we would like to develop an evaluation strategy that shows the impact on levels/patterns of service utilization for the population that are candidates for such a program". So here is what I would like: I would like all of the encounters across the full range of Island Health services for this cohort of interest. All I need is the admission date and program name for each service accessed by members of this cohort. That shouldn't be that hard. I understand you have all of the encounter data in the warehouse so just please cut a slice that corresponds to this cohort of interest."

Applied Clinical Research Unit Person: "You want the older adults"

Program Manager: "Yes"

ACRU Person: "Do you mean 'older' as in "cut-off based on potentially years left given average life expectancy for cohort, or do you mean 'older' in terms of an absolute scale of ages based on life expectancy for the population at large???? "

Program Manager: "Okay, you can stop doing that now. I mean do not filter by age – ANYONE who accesses any of the following programs that SPECIFICALLY target persons with chronic/severe addictions should be included in the sample, regardless of age".

ACRU Person: "Okay….[a few hours later]….Okay, here is what your raw data will look like. What do you think" [[see next slide]]

Program Manager: "What planet do you come from. How am I supposed to make any sense of that. Do you really think that I know what the names are for all 1566 clinical services/programs and that I know what they do – who they serve, what services do they provide, where do they provide services – how am I supposed to know that. That's YOUR job. Get with it."

An illustration of the problem

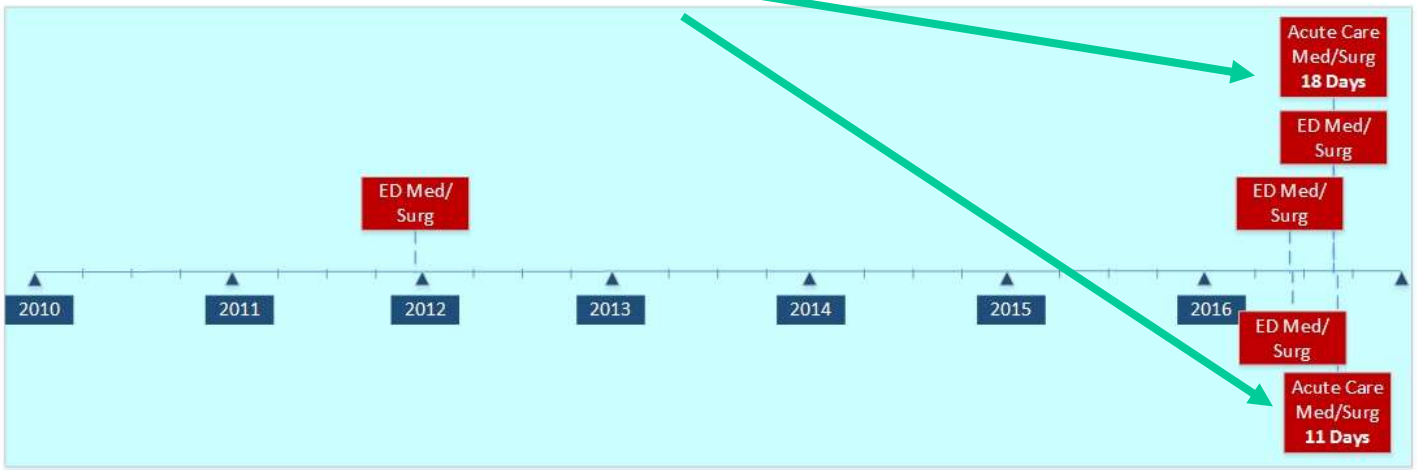
- Manager - Whoa! What are all of those unit names? How am I supposed to know what they are??
- ACRU – That's just a small sampling of your data file. The actual file contains 159,716 rows for your 4000 or so cohort. But there are only 1566 Unit Names (as of yesterday) that you need to translate into terms like "Post-Withdrawal Stabilization Unit" or "Sobering & Assessment Centre"
- Manager – forget it. Life is too short. Just give me the acute care plus ED encounters – I can make sense of that and use it as a proxy for total service system utilization.
- ACRU – I don't think you want to do that. Look at the graphic on the following slide.

Anonymized Identifier	Unit [i.e., Name of the Service in the EHR and in the Warehouse]	Admission Date
11242534732	1471 Suites	Days elapsed since Day 0 for a case
11242534732	1st Floor Dogwood Wing	Days elapsed since Day 0 for a case
11242534732	Adanac Street	Days elapsed since Day 0 for a case
11242534732	House 1	Days elapsed since Day 0 for a case
11242534732	CDH - Intensive Care Unit	Days elapsed since Day 0 for a case
11242534732	RJH - Post Op Openheart Surgery & Major Vascular Surgery	Days elapsed since Day 0 for a case
11242534732	Addiction Outpatient Services	Days elapsed since Day 0 for a case
11242534732	Substance Abuse	Days elapsed since Day 0 for a case
11242534732	Stabilization Unit	Days elapsed since Day 0 for a case
11242534732	Female Beds	Days elapsed since Day 0 for a case
11242534732	Intake	Days elapsed since Day 0 for a case
11242534732	RJH - Emergency	Days elapsed since Day 0 for a case
11242534732	Intensive Case Management Team - PES	Days elapsed since Day 0 for a case
11242534732	Grove	Days elapsed since Day 0 for a case
11242534732	RJH - 1 North West - Psychiatric Intensive Care Unit	Days elapsed since Day 0 for a case
11242534732	South 2 - Echo	Days elapsed since Day 0 for a case
11242534732	Diabetic Education Centre	Days elapsed since Day 0 for a case

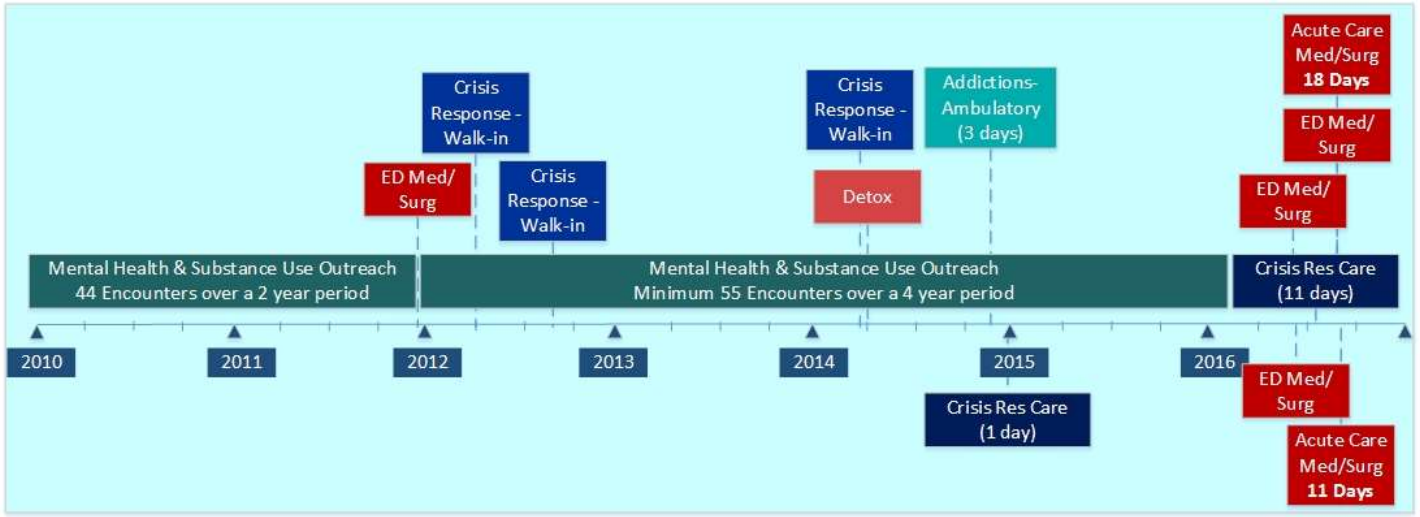
High Need/High Risk Substance User – two views

Note well – re: acute care encounters. Acute care for WHAT PROBLEMS? We ***really need*** the discharge abstracts data base data here (i.e., a key administrative data set) – that is we really need administrative data layered onto transactional encounter data.

The 'Usual' Acute Care plus ED Encounter lenses



All Secondary & Tertiary Encounters



The data dialogue continues

Manager – wait – that 'classic' Acute-Care-plus-ED-Encounters view is not just less encounters than the more complete picture. It is a fundamentally DIFFERENT picture.

ACRU – You got it. If you build a prediction model from the Acute-Care-plus-ED-Encounters data – it would say "low risk for problems, e.g., fentanyl OD". If you build a prediction model from the full body of encounter data, it says "Accident waiting to happen".

Manager – and if you are in the surveillance business, you will count this person (and people like this person) in fundamentally different ways.

ACRU – and if you are in the managing your budget business, you will get grossly different estimates of the amount of $$ the service system is using for this at-risk cohort. And if you are an epidemiologist in the burden-of-illness-severity business, your models will be very far off the mark – at least for this cohort.

ACRU – and for that group of 4000 persons, you have 41846 Sobering & Assessment Centre encounters which do not show up on the acute-care-plus-ED radar screen, AND that is not distributed evenly across the 4000 persons. E.g, you have some people with over 700 S&A Centre encounters over four years, and some with none. So guess what.

Manager – What?

ACRU – it is not a homogeneous entity. So not only do you NEED all of the encounters – you really want someone who can help you identify clusters of people who are relatively homogeneous with respect to patterns of service utilization.

Manager – can I do that with a reporting tools built around a basic set of descriptive statistics?

ACRU – sorry.

Manager – how about a slick new set of big data visualization tools that we can point at the encounter data and generate clusters, even if we don't understand the underlying statistical basis of cluster analysis, e.g., the concept of "distance from a cluster centre" does not speak to us, and we have no clue as to how we might go about validating the cluster solution or determining how many clusters we have. Oh yes - do you think it will be a problem if for each observation we could have a value on any of 1500++ entities, which I think you data people call "variables"???

ACRU – it will probably be a problem at least as large as a school of fish without sunglasses – and probably a much bigger problem – unless we can come up with a way of reducing those 1500 clinical service entities down to a far smaller number of service clusters that are relatively homogeneous with respect to target population served and services provided – and a few other characteristics – e.g, are 99 family care homes REALLY 99 different types of service?? I suggest you go home and google "curse of dimensionality" and then we can continue this conversation.

Learning objectives

- How the health service system is organized – what types of functions are performed at which layers in the system.
- How information is used by different parties performing different functions in the service system.
- The inescapable dependencies that relate essential information at higher 'layers' to information products derived from data at lower 'layers' – you CANNOT finesse around the challenges of working with transactional encounter data IF you want to leverage quality of care from data extracted from the transactional EHR;
- Why, historically, researchers and administrators have relied on administrative data or very coarse and simplified bodies of transactional data extracted largely from acute care plus Emergency Department locations.
- Exactly why it is that you don't want to do that.
- How to wrestle the longitudinal transactional data to the ground, and re-attach the administrative data to that transactional substrate – i.e., re-contextualize those administrative data in a way that responds to critical limitations in what you can do if you are working ONLY with transactional data.
- What kinds of statistical tools and research designs you can employ to generate clinically actionable products from those data that you have wrestled to the ground.
- Some sense of when all of this "working with transactional data" stuff is methodological overkill, i.e., when you can work quite meaningfully and effectively with the administrative data and let someone else worry about the transactional data.
- Enough experience derived over the course of two weeks to realize that (a) you can do this; and (b) you may be able to do something new and seriously interesting to a lot of people with these data.