BUILDING AN ONTOLOGY OF PUBLIC HEALTH INTERVENTIONS AND TOOLS FOR EVIDENCE-BASED PUBLIC HEALTH
Introduction
Tools for evidence-based public health (EBPH)
Evidence-based Public Health

- PH decisions made using better knowledge and information are better
  - i.e. lead to better outcomes

- Better knowledge and information:
  - relevant to task at hand
  - in a usable format
  - sufficient amount
  - high quality
PopHR: Population Health Record

- Web-based software platform for intelligent analysis and visualization of population health information
- Integrates heterogeneous data from multiple sources to calculate up-to-date indicators:
  - health determinants
  - diseases and conditions
  - health system performance
- Indicators are contextualized by public health knowledge
- Addresses common limitations of existing web portals
  - out-of-date indicators
  - low geographical resolution
  - poor support for simultaneous analysis of multiple indicators
PopHR as an EBPH tool

Knowledge in epi & PH: Public Health Ontology

Intervention evidence: HealthEvidence© ⇔ OPHI ⇔ Interventions KB

Applicability & Transferability:
How well does this evidence fit my population?

Population health information: Ontology-based Health Indicators
Public Health Ontology (PHOnt)

- Based on the Australian Classification of PH Activity *Jorm, Gruszin, and Churches (2009)*

- Definitions of concepts in following categories:
  - determinants of health and risk factors
  - health issues (diseases, functioning, well-being)
  - populations and their characteristics
  - PH resources and settings

- Formal encoding of causal epi knowledge
Ontology-based health indicators

- Systematized using multiaxial taxonomy in Health Indicators Ontology
- Defined properties that have implications for analysis and visualization (e.g. measurement units)
- Linked to health conditions and determinants of health via PHOnt
Ontology of PH Interventions (OPHI)

Intervention evidence

Knowledge in epi & PH + Intervention evidence → Population health information
Barriers to using evidence about PH interventions in practice

- Poorly organized evidence
  - Ongoing efforts (HealthEvidence.org, CDC DB, BEEM project) have limitations
  - Need a conceptual framework and a universal vocabulary to consistently describe interventions and evidence about their effectiveness

- Varying quality of research and reporting (difficult to assess)

- Remaining knowledge gaps

- Lack of tools providing access to the evidence within health information systems
Advantages of ontology

- Supports multi-axial classifications
  - interventions can be described by many properties
- Relationships among entities have clear semantic (unlike keywords and tags)
  - e.g. interventions to *prevent COPD* vs. interventions for improving outcomes in *people with COPD*
- Machine-readable
  - automatic tools can be developed
- Relationships encoded as logical axioms to allow automated reasoning
  - e.g. use knowledge about risk factors to search for upstream interventions
Is there an existing ontology?

- Extensive classifications exist for clinical, but not for public health interventions
- ICHI: an attempt to extend clinical interventions taxonomy to public health
  - in practice, not many attributes of clinical interventions are relevant to PH, and the other way around
- Australian classification of public health
  - broad scope extending beyond interventions
  - includes comprehensive lists of methods, settings and resources relevant to interventions
  - lacks important properties to fully describe interventions
Search for existing frameworks

- 3 independent reviewers
- Literature search:
  - PubMed
  - Scopus
  - Google Scholar
  - Cochrane Library
- Gray literature search (misc. documents):
  - PHAC
  - INSPQ
  - WHO: PHE
  - NPHP

Keyword search:
“public health” AND “interventions” AND “classification”/“taxonomy”/“ontology”

995

Title and abstract review:
removed ineligible
246

Full-text review:
agreement by all 3 reviewers
14
Synthesis and term alignment

- high-frequency concepts
- synonyms
- classification dimensions
- properties of interventions

OPHI conceptual framework
Core concepts

**Intervention**
- Refers to an organized activity that intends to improve population health
- Not limited to activities performed by PH agencies
- Instances are specific programs deployed within specific populations and geo-spatial context

**Evidence**
- Refers to a study establishing an effect of an intervention on a specific health outcome
- Unless there is evidence of effect, knowledge about possible interventions is of little relevance to practitioners
## Properties of interventions

<table>
<thead>
<tr>
<th>Property (preferred term)</th>
<th>Range</th>
<th>Cardinality</th>
<th>Definition / description</th>
<th>Synonyms (in other frameworks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>intent</td>
<td>PH Function(?)</td>
<td>1</td>
<td>High-level purpose category (e.g. promote-prevent-protect)</td>
<td>purpose, strategy</td>
</tr>
<tr>
<td>distal target</td>
<td>Health Issue</td>
<td>1+</td>
<td>Existing health issue being addressed by an intervention, an ultimate goal of an intervention (as opposed to specific immediate outcomes of an intervention).</td>
<td>problem, issue, health priority</td>
</tr>
<tr>
<td>recipient</td>
<td>Population</td>
<td>1</td>
<td>The population, to which an intervention applies</td>
<td>target, target population</td>
</tr>
<tr>
<td>actor</td>
<td>Organization</td>
<td>1+</td>
<td>Entity (organization) that makes an intervention possible in some capacity (Who does it and in what role?)</td>
<td></td>
</tr>
<tr>
<td>setting</td>
<td>PH Setting</td>
<td>1(?)</td>
<td>Institutional and social environments, partnerships, and locations in which public health activity occurs</td>
<td>context</td>
</tr>
<tr>
<td>resources</td>
<td>PH Resource</td>
<td>1+</td>
<td>The means available for the operation of health systems, including human resources, facilities, equipment and supplies, financial funds and knowledge</td>
<td>resources &amp; infrastructure, means</td>
</tr>
<tr>
<td>time frame</td>
<td></td>
<td></td>
<td>Start, end, duration</td>
<td></td>
</tr>
<tr>
<td>geography</td>
<td></td>
<td></td>
<td>Geo location</td>
<td></td>
</tr>
<tr>
<td>evidence</td>
<td>Evidence</td>
<td>0+</td>
<td>Evidence of effectiveness for particular target issue(s)</td>
<td></td>
</tr>
</tbody>
</table>
Properties of evidence

- Using Ontology of Clinical Research (OCRe) to represent study design & other properties
- Additional attributes of evidence are being considered based on the input from knowledge translation experts

<table>
<thead>
<tr>
<th>Property (preferred term)</th>
<th>Range</th>
<th>Cardinality</th>
<th>Definition / description</th>
</tr>
</thead>
<tbody>
<tr>
<td>intervention</td>
<td>Intervention</td>
<td>1</td>
<td>Which intervention has been evaluated</td>
</tr>
<tr>
<td>effect</td>
<td>Health Issue, Health Determinant</td>
<td>1</td>
<td>What was demonstrated to change as a result of the intervention? (measurable outcome)</td>
</tr>
<tr>
<td>effect size</td>
<td>{strong, moderate, weak}</td>
<td>1</td>
<td>The magnitude/significance of observed changes</td>
</tr>
<tr>
<td>effect direction</td>
<td>{positive, negative}</td>
<td>1</td>
<td>Although we may assume that an intervention would have an intended effect on the outcome, it is better to specify explicitly how an intervention affected the outcome</td>
</tr>
<tr>
<td>type</td>
<td>EvidenceType</td>
<td>1</td>
<td>What's the type/quality of evidence? (e.g. individual peer reviewed study vs. systematic review)</td>
</tr>
</tbody>
</table>
Instances of evidence from literature encoded in terms of OPHI
- Essential to include original studies and not just systematic reviews
- Each study can produce multiple instances of evidence – for each measured outcome

**Phase I:** Manually encoded small set of studies (smoking cessation)

**Phase II:** Use NLP to populate PH Interventions Knowledgebase
Transferability assessment
For an intervention with proven effectiveness, can I expect similar results in my target region?

- Assess the similarity between study population/setting & target population/setting
  - Define phenotype (characteristics to be considered)
  - Choose similarity metrics (conceptual/semantic similarity, Euclidian distance, ...)
  - Possibly pool results across different characteristic

- The better the match between study population/setting and the target, the higher transferability
Visualization

Type 2 Diabetes Mellitus Interventions

Quality

- High quality
- Good fit
- Modest effect

Match

- Good fit
- Large effect
- Poor quality

Individual studies
Putting it all together in PopHR

Causal knowledge

Region profile

Interventions tab

Matching Algorithms
- euclidian distance
- semantic similarity

3D ranking:
- match
- effect size
- quality

Extend/narrow search for conditions & RFs

HealthEvidence©

Reviews
+ annotations
+ original studies

Intervention KB

Reviews+

Original Studies

Population characteristics

PopHR 1.5

PopHR 2.0
Extras
PHOнт taxonomy examples: Determinants of Health
PHOnt taxonomy examples:
PH functions, settings, and resources
Consensus knowledge exists regarding risk factors and downstream conditions.

Links established at the population level.
Causal knowledge: encoding challenges

- Non-deterministic causation
- Causal links established from the analysis of populations don’t always hold for all individuals
- Ontology languages, like OWL, don’t deal well with uncertainty

Need to express uncertainty vs. Rigid formal logic
From causal diagram to ontology

- **Nodes:**
  - Diseases and conditions
  - Health determinants
  - Health-related events and procedures (e.g. Amputation)

- **Arrows** = probabilistic causal links:
  - `has_effect_on` (positive/negative)
  - generally transitive
  - allow versatile DL queries

- **Implications of *consensus knowledge***
  - functional form not fully known or agreed upon
  - limitations in measurement
  - assumption of independence
PHOnt Future work

- Causal knowledge in PHOnt is incomplete
  - current scope: PHAC Chronic Disease framework
  - built in collaboration with experts from INSPQ and DSP
  - manual knowledge extraction process is not scalable

Use statistical NLP for automatic extraction of causal findings from published literature
Current PopHR Ontology Suite

Domain ontologies

- **Disease Ontology** *(DOID)*
  - Taxonomy of diseases

- **PopHR - Geography**
  - Geographic units
  - Spatial relations
  - Geographic locations (instances)

- **Public Health Ontology** *(PHont)*
  - Health issues
  - Health determinants
  - PH interventions
  - Demographics
  - Causal pathways

Upper ontologies

- **BFO***
- **SIO***

compatible with

Application ontology

- **Health Indicators**
  - Health indicator taxonomy
  - Epidemiological concepts
  - Data specification
  - Statistical methods
  - Temporal units and relations
  - …

* External resources
Indicator encoding example
Indicators at the PopHR front-end