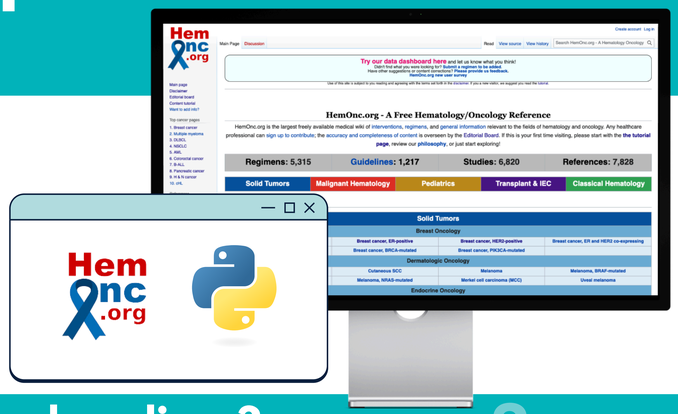
Interface for description and analysis of systemic oncology protocols

Python modules to integrate extended HemOnc data model with OMOP data

Georgina Kennedy,[1,2,3] Travis Zack,[4,] Ivy Cerelia Valerie,[1.] Michael Gurley,[5] Jeremy L Warner [6]



AIM

Systemic anti-cancer treatment (SACT) regimens are complex and multi-dimensional — defined by their patient populations, components, dosages, periodicity, and duration. In real world use, they are often modified through dose reduction, early termination and / or scheduling changes to address side-effects, treatment-response, resources availability, or patient preference. There is limited evidence for the effect that these modifications have in practice, as the modifications are in and of themselves also multi-factorial, and therefore outside the bounds of what can be practically studied in clinical trials.

To meaningfully assess the effect of dose modifications, one must be able to identify the full baseline intended treatment, and thus produce an accurate 'treatment delta' that goes beyond the basic measure of relative dose intensity (RDI). These full treatment specifications are available in the HemOnc data model, however are not included in the vocabulary version that is loaded into the OMOP vocabularies.

Looking for ways to explore the full HemOnc data model including dosing baselines?

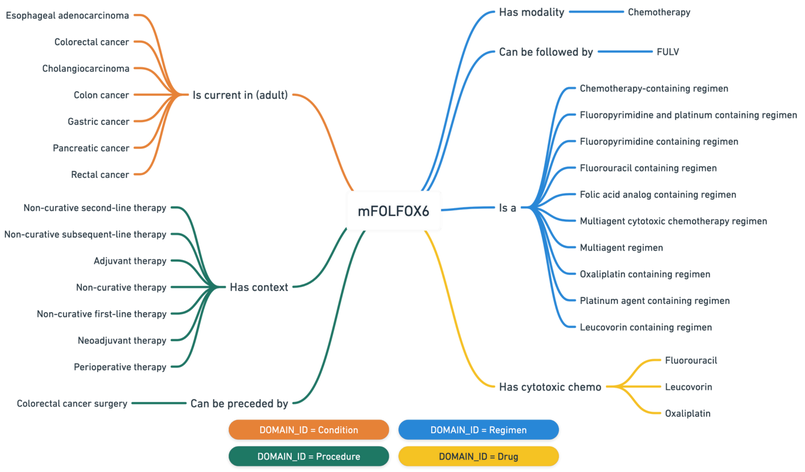


Fig 1. Example existing HemOnc OMOP relationships showing selected relationships centred around the mFOLFOX6 regimen in condition, drug, procedure and regimen domains.

METHODS

The HemOnc ontology consists of controlled terms and relationship metadata to describe the evolving practice of anti-cancer treatment and supporting evidence. Inputs are limited to oncology regimens backed by at least phase II clinical data and published in peer-reviewed medical journals. Each entry includes the complete description the regimen as specified in the relevant clinical trial, including not only pharmaceutical agents and their links to base regimen, therapeutic contexts, and source evidence, but also the specification of many complex relative and absolute scheduling factors. These schedules frequently include irregular and/or multipart signetur (sig) specifications at both the cycle and component level.

Although regimen properties such as inheritance / classification are easily expressed in the typical OMOP triples, extending even a single variant (of 18) for the regimen in Fig 1 into a fully specified dose prescription in a minimally useful from at the regimen part, cycle, and component levels requires 80+ additional relationships. Expanded to the entire HemOnc vocabulary of more than 2,000 regimens, 5,000 variants and 15,000 [regimen part - component] pairs is not a meaningfully workable solution, as the breadth and count of variables becomes unwieldy. It is also not a good fit for reassembling the reference baselines required for analysis. We have therefore chosen instead to use the wide format of the HemOnc Sig table specification while retaining tight integration with the OMOP vocabulary and clinical tables.

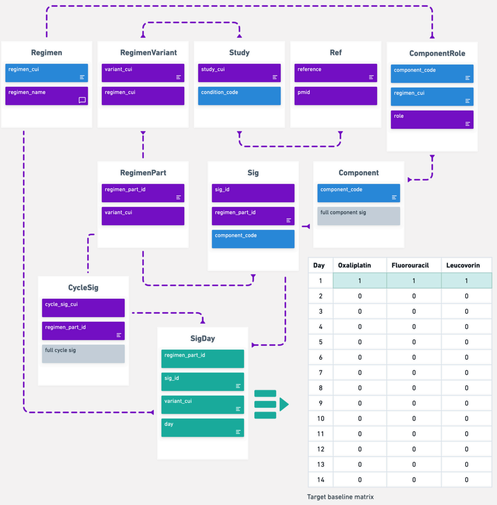
By following the existing OMOP Alchemy object relational model (ORM) definition, the extended interface can sit alongside and be jointly queried with the clinical data - fulfilling the specific analytic use-case for SACT baseline derivation, without being 'shoe-horned' into an artificially restrictive format, nor affecting typical CDM users who do not require this detail. This sets the groundwork for a family of such extensions to support other high complexity use-cases in a modular and repeatable fashion.

Richly integrated extension between full reference model and clinical data

Exemplars can be found at the github repository (QR code below). In particular the reader is referred to the directory notebooks.

This introduces and describes the following use-cases: 1.Importing HemOnc source tables into the target data model, including mapping to concept IDs and importing of required vocabularies into the sqlite database, as well as spaCy matcher rules that have been used to extract details from the semi-structured sig strings; 2.Code for the automatic generation of entity relationship diagrams used to provide orientation to the data model. 3.Some useful example queries that can provide insights from the extended HemOnc reference that is not currently available in the OMOP vocabularies. Specific examples provided for navigating the study ↔ variant ↔ condition relationships, temporal availability of regimens and component types, as well as the rolling out of the full drug delivery schedule and dosing to serve as the baseline comparator when seeking dose variation information. 4.Some example queries for the handling of extended functionality to support episode-driven complexity of CDM, as required to assemble systemic anti-cancer therapy treatment episodes from source data.

Note that this functionality assumes that the user has already assembled their source drug exposure records into a sensible hierarchy of treatment episodes with appropriate inferences made regarding links to the disease being treated (where available). This assumption is in itself an advanced and complex implementation; however work is underway to extend these interfaces so these steps can also be completed in a repeatable and sharable fashion.



Day	Oxaliplatin	Fluorouracil	Leucovorin
1	1	1	1
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0

Fig 2. Illustrative subset of HemOnc relationships to define interface – blue fields integrate directly (via concept_code) to OMOP vocabularies. SigDay extension provided to support production of target base cycle matrices.

```python
# specify target complex joins for SACT episodes by joining diagnostic episodes with earliest chemo
# delivery date in a linked treatment episode to identify start of systemic treatment

systemic_therapy_with_dx = (
    sa.join(
        dx_subquery,
        systemic_therapy_start,
        sa.and_(
            systemic_therapy_start.c.episode_parent_id==dx_subquery.c.dx_episode_id,
            systemic_therapy_start.c.person_id==dx_subquery.c.person_id
        ),
        isouter=True
    )
)

# we then assemble the above complex queries into target mapping classes, which draw from multiple
# sub-queries and / or joins as required

class Systemic_Therapy_Episode(Base):
    __table__ = systemic_therapy_with_dx
    episode_id = systemic_therapy_start.c.sact_episode_id
    person_id = systemic_therapy_start.c.person_id
    sact_start = so.column_property(systemic_therapy_start.c.sact_start)

    dx_ep_id = dx_subquery.c.dx_episode_id

    dx_object: so.Mapped[Optional['Episode']] = so.relationship(foreign_keys=[dx_ep_id])
    person_object: so.Mapped['Person_Episodes'] = so.relationship(back_populates="sact_episodes", foreign_keys=[person_id])
    episode_object: so.Mapped['Episode'] = so.relationship(foreign_keys=[episode_id])
    sact_events: AssociationProxyList['Episode_Event']] = association_proxy("episode_object", "events")

    @property
    def episode_agents(self):
        return list(set([s.event_polymorphic.drug_label for s in self.sact_events if s.event_polymorphic.polymorphic_label=='drug_exposure']))


# this means that the end user can perform much simpler queries, such as

import omop_alchemy as oa
engine = oa.oa_config.engine

with so.Session(engine) as session:
    person_object = session.query(Person_Episodes).filter(Person_Episodes.person_id == 1234).first()

# this will return a person object, for person_id=1234, through which all linked overarching dx episodes may be accessed directly as
# person_object.condition_episodes, as well as any systemic therapy episodes as person_object.sact_episodes

# if only the dx episodes that are linked to systemic therapy episodes are desired, one can instead access [sact.dx_object for sact in
# person_object.sact_episodes]
```

1 Faculty of Medicine & Health, UNSW Sydney,
2 Ingham Institute of Applied Medical Research, Liverpool, Sydney,
3 Maridulu Budyari Gumal (SPHERE) Cancer Clinical Academic Group, Australia,
4 University of California, San Francisco, CA, USA,
5 Northwestern University, Chicago, IL, United States,
6 Brown University and Lifespan Cancer Institute, Providence, RI, USA.

HEMONC ALCHEMY

OMOP ALCHEMY

Proudly supported by Cancer Institute NSW

Contribution to this work from JLW was funded by NCI U24 CA265879. Contribution to this work from GK was funded by CINSW Capacity Building Grant 2021/CBG003