**INTRODUCTION**

Oncology data collected for individual patient care or registries use dictionaries to report disease diagnosis, biomarkers, procedures and disease progression, for example NAACCR Dictionary and CAP Checklists. However, these collections are not suitable for use in standardized vocabularies and network research because of:

1. Lack of nomenclature (table 1)
2. Lack of context (table 2)
3. Inconsistent post-coordination (table 3)
4. Mixture of transactional with observational content (table 4)
5. They are not integrated or linked to each other.

**METHODS**

1. We undertook an effort to develop a common syntax and consistent set of concepts.
2. We used the semantic content of both NAACCR and CAP.
3. We converted combinations of schema/protocol, variables and values to create fully pre-coordinated concepts. Only presence and absence attributes were left post-coordinated.
4. We consolidated the concepts across NAACCR and CAP.
5. We categorized the resulting concepts into a set of concept classes.

**RESULTS**

- 2666 concepts are available in Athena. The concept classes within vocabulary are shown in figure 1.