Exploring Variables Through CIPHER Data Dictionaries

Michael R. Murray, MS[1], Jacqueline Honerlaw, RN, MPH[1], Yuk-Lam Ho, MPH[1], Edward Zielinski, BS[1], Tiffany Sim, MPH[1], David Heise, MS[2], Keith Connatser, BS[2], Kelly Cho, MPH, PhD[1,3]

[1]VA Boston Healthcare System, Boston, MA; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]Mass General Brigham, Harvard Medical School, Boston, MA

CIPHER
CENTRALIZED INTERACTIVE PHENOMICS RESOURCE

VA U.S. Department of Veterans Affairs

OAK RIDGE National Laboratory

Visit CIPHER

Introduction

- Phenotype libraries facilitate the reuse of definitions derived from electronic health records (EHR) data.
- The Centralized Interactive Phenomics Resource (CIPHER) is a platform developed by the US Department of Veterans Affairs (VA) and Oak Ridge National Laboratory (ORNL), accessible at https://phenomics.va.ornl.gov, which contains a searchable library of computable phenotype definitions.[1]
- Phenotypes may include both structured, clinical data and external data, such as patient provided survey responses seen in the American Heart Association's "Life's Simple Seven."(Figure 2)
- CIPHER has developed a data dictionary interface to enable browsing of such external variable descriptions and referencing of variables in a phenotype definition.

Development

- CIPHER piloted the development of the data dictionary feature with surveys from the Million Veteran Program (MVP).[2]
- Phenotypes use both EHR and derived, non-clinical data to define phenotypes for conditions in the MVP cohort.
- MVP researchers desired to search for existing variables that could be used for phenotype development and cite the use of the variable directly in the phenotype definition.
- For each data dictionary, there is a minimum set of fields our standard captures: **author**, **source URL**, **creation date**, **version**. (Figure 1)
- For each variable, we captured: **variable name**, **description**, **data type**, **value description**, plus additional information such as counts of respondents by value.
- The additional data within each variable are entered as key-value pairs. The specific fields are not fixed, giving each variable's detail view the flexibility to show the information pertinent to its data dictionary.
- We improved our Knowledgebase Search to allow users to search for data dictionaries and variables by name.
- Applicable filters are available when browsing search results. (Figure 3)

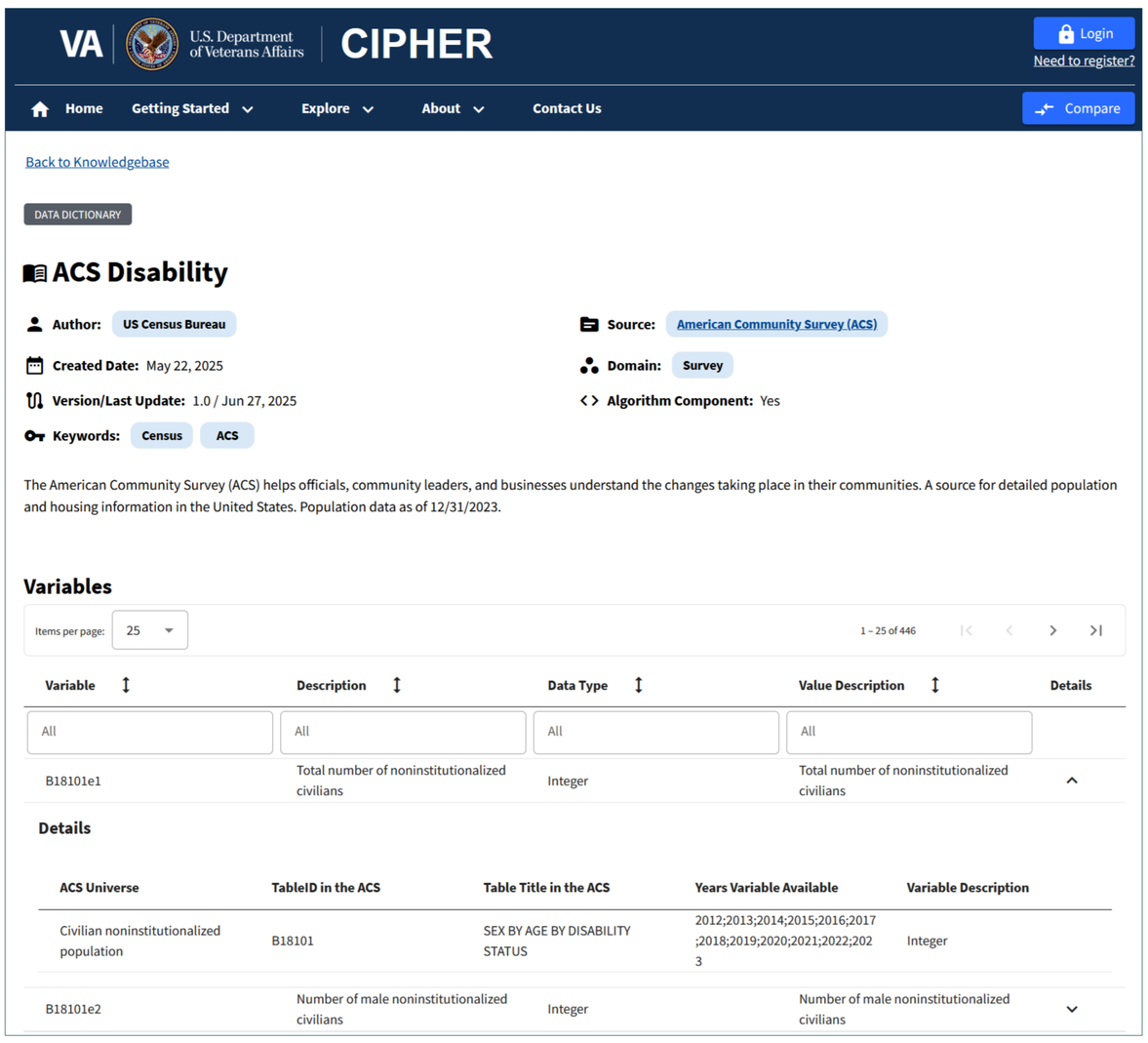

Figure 1. Data dictionary from the American Community Survey (ACS) on disability, currently showing the expanded detail view for variable B18101e1

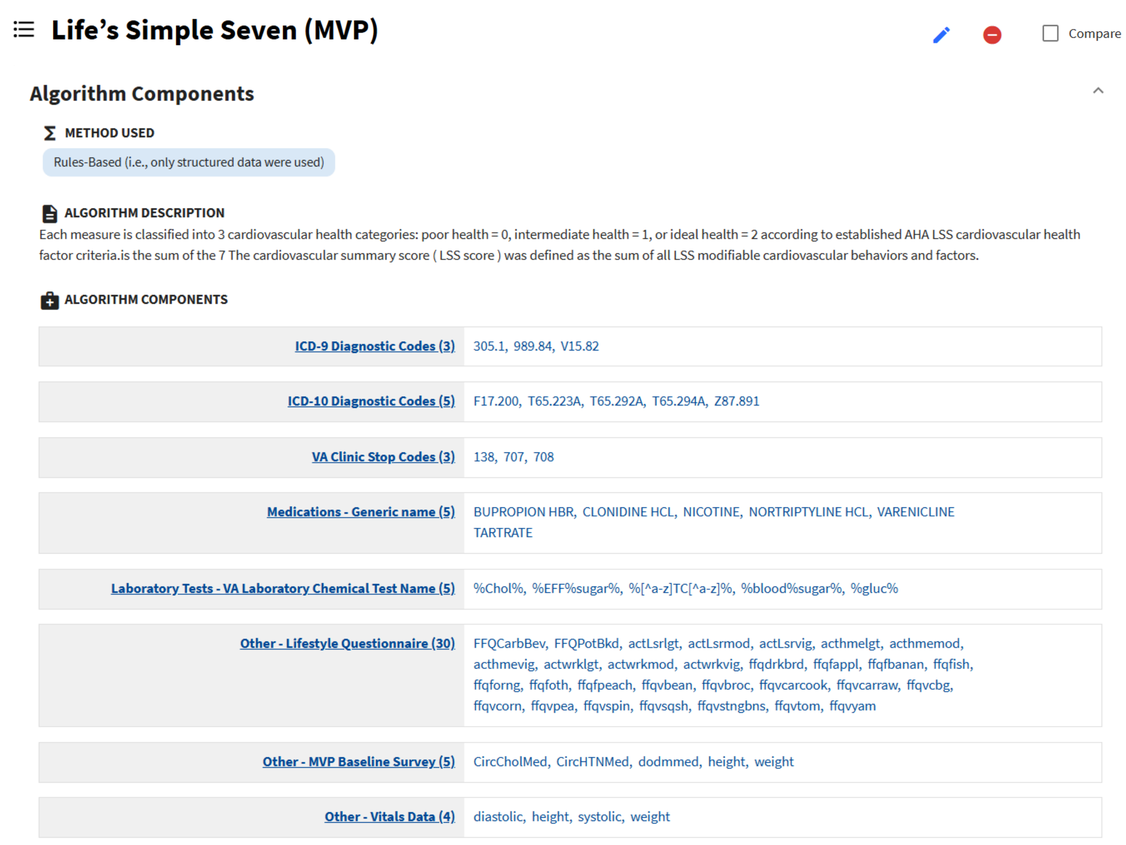

Figure 2. An example of a phenotype using survey responses.[3]

Discussion

- The data dictionary feature serves as a method for finding variables to use for phenotype development and referencing variables used in phenotype definitions.
- With this new functionality, we can catalog phenotypes using surveys and external data sources in a clear manner, providing prospective users of these phenotypes with necessary information.
- Additionally, CIPHER users have the potential to explore variable metadata and usage in phenotyping, driving the use of such data in EHR research.
- Planned improvements in the coming months include integration of additional data dictionaries for publicly available datasets and linkage from a phenotype definition to a data dictionary.

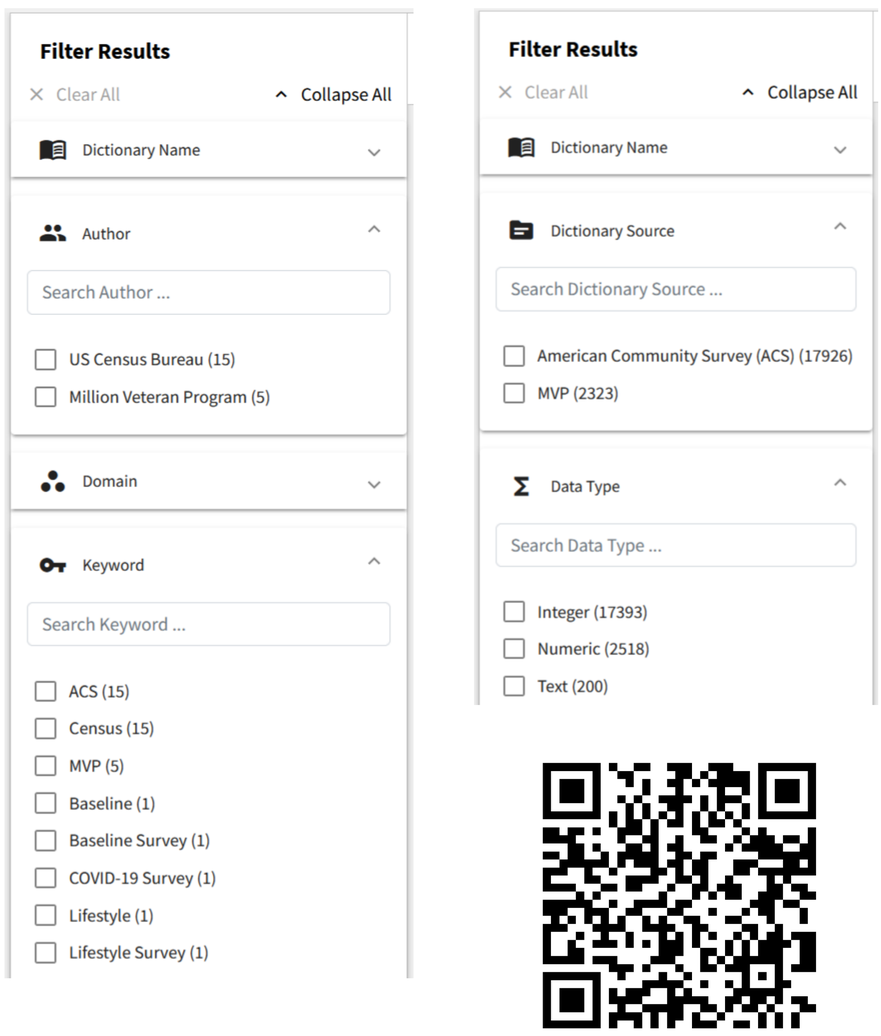

Figure 3. The filters available while searching for data dictionaries (*left*). The filters available while searching for specific variables (*right*). A QR code to view the data dictionary search page (*bottom-right*)

References

1. Honerlaw J, Ho YL, Fontin F, et al. Centralized Interactive Phenomics Resource: an integrated online phenomics knowledgebase for health data users. J Am Med Inform Assoc. Published online March 13, 2024. doi:10.1093/jamia/ocae042
2. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol. 2016;70:214-223. doi:10.1016/j.jclinepi.2015.09.016
3. Centralized Interactive Phenomics Resource (CIPHER). Accessed March 7th, 2025. https://phenomics.va.ornl.gov/web/cipher/phenotype-viewer/details?uqid=d38db2ef319e4321dd47453cf39b5b9a