Tapping into the Power of Veterinary EMR Data

Tracy L. Webb, DVM, PhD

August 13th, 2024

CSU-CU Summit

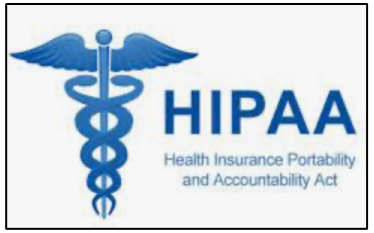
COLORADO STATE UNIVERSITY | COLLEGE OF VETERINARY MEDICINE AND BIOMEDICAL SCIENCES

Disclosures

- No financial conflicts of interest
- Stated opinions (and errors) are my own

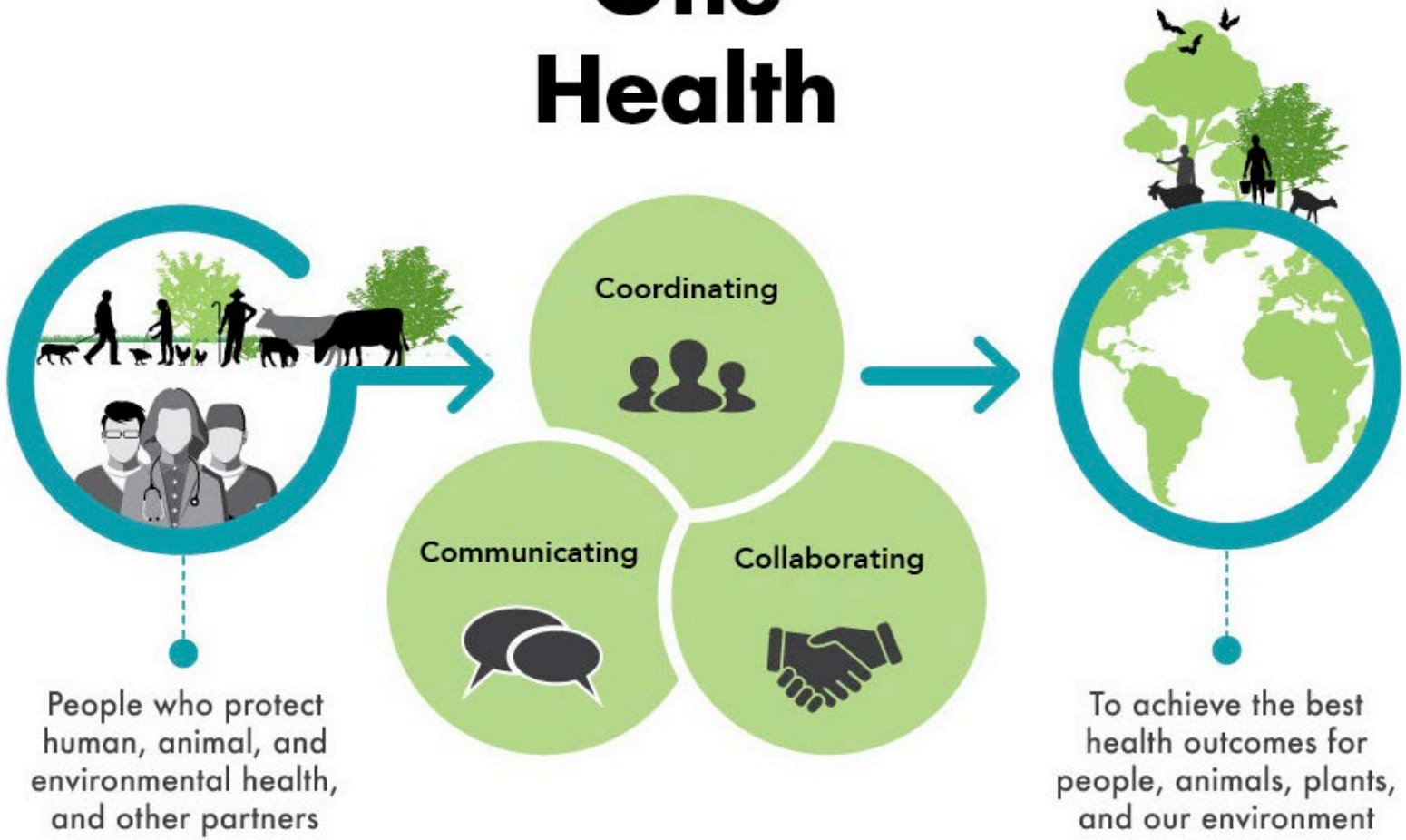
Veterinary Medicine

- Large amount of medical data:
  - ➢Approximately 270 million veterinary visits for dogs and 230 million for cats
    - Approximately 90 million dogs and 96 million cats in the United States
    - Average 3 visits/dog/year and 2.4 visits/cat/year
  - ➢Approximately 1 billion human physician office visits per year
    - Average 3.21/person/year

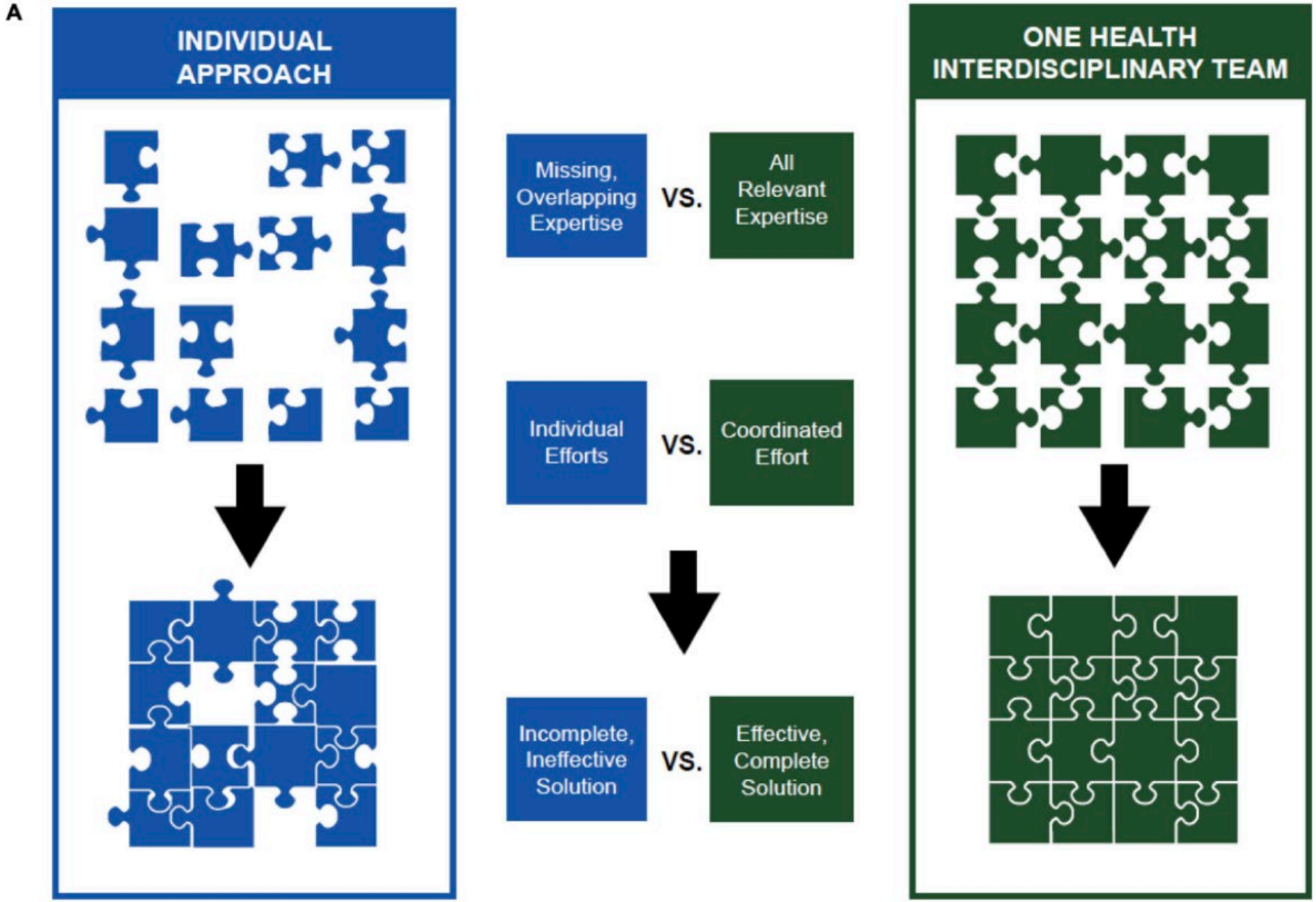


- Shared diseases, environments, exposures, and treatments

One Health



When we protect **one**, we help protect **all**.



PERSPECTIVE article

Front. Microbiol., 08 October 2021

Sec. Infectious Agents and Disease

Volume 12 - 2021 | https://doi.org/10.3389/fmicb.2021.718546

This article is part of the Research Topic

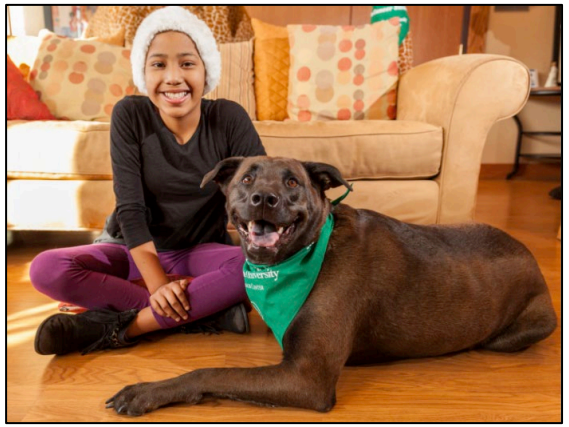
Needs and Potential Application of One Health Approach in the Control of Vector-borne and Zoonotic Infectious Disease

View all 14 Articles ›

Parallel Pandemics Illustrate the Need for One Health Solutions

Claire Tucker[1], Anna Fagre[1], George Wittemyer[2], Tracy Webb[3], Edward Okoth Abworo[4] and Sue VandeWoude[1*]

One Health Approach In Action



- Osteosarcoma:
  - People: 800-1000 cases/year
  - Dogs: >25,000 new cases diagnosed every year
  - https://cvmbs.source.colostate.edu/clinical-trial-for-pediatric-bone-cancer-underscores-one-cure-philosophy-at-csu/

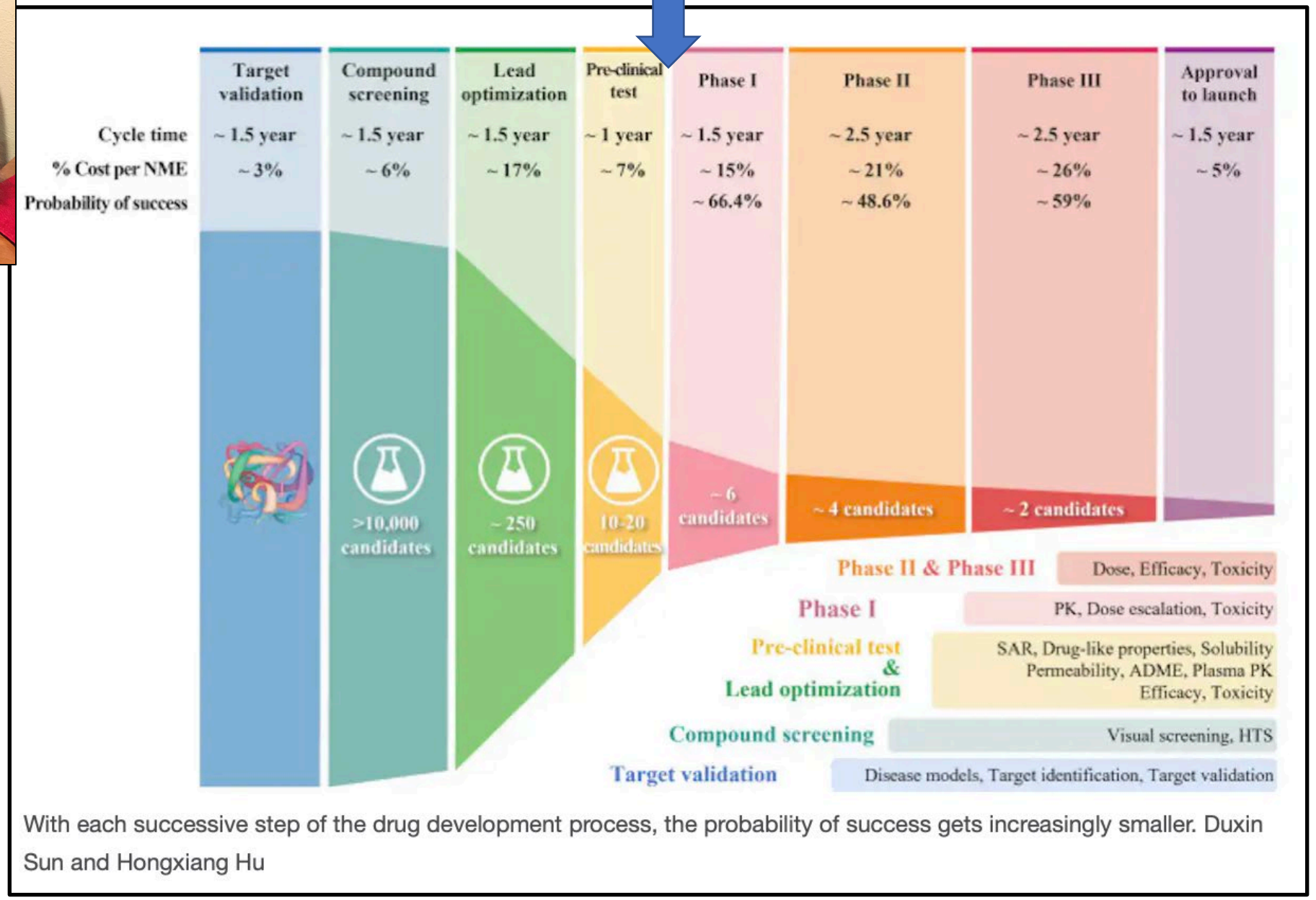
- 2022 - 60 Minutes “Comparative Oncology: Trial cancer treatments in dogs could also lead to breakthroughs for humans.”

- 2021 National Academies of Science Workshop “The Role of Companion Animals as Sentinels for Predicting Environmental Exposure Effects on Aging and Cancer Susceptibility in Humans”

COHA
Clinical and Translational Science Award
One Health Alliance

One Health Institute

C. WAYNE MCILWRAITH
TRANSLATIONAL MEDICINE INSTITUTE



With each successive step of the drug development process, the probability of success gets increasingly smaller. Duxin Sun and Hongxiang Hu

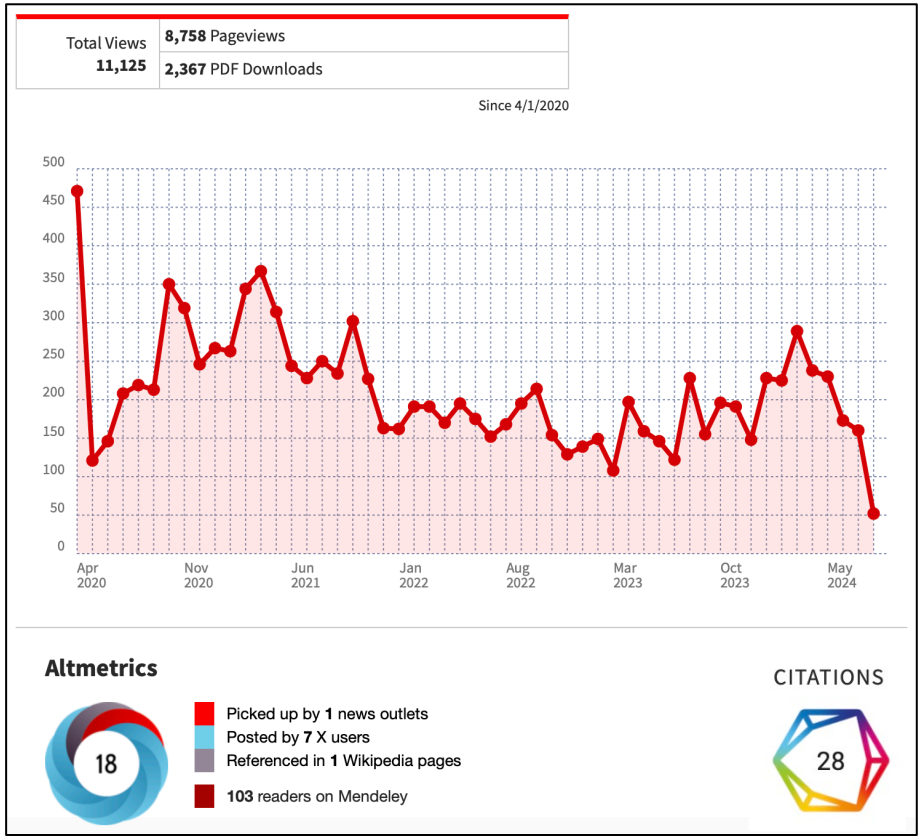
https://theconversation.com/90-of-drugs-fail-clinical-trials-heres-one-way-researchers-can-select-better-drug-candidates-174152

JOURNAL ARTICLE

Veterinary informatics: forging the future between veterinary medicine, human medicine, and One Health initiatives—a joint paper by the Association for Veterinary Informatics (AVI) and the CTSA One Health Alliance (COHA)

Jonathan L Lustgarten ✉, Ashley Zehnder, Wayde Shipman, Elizabeth Gancher, Tracy L Webb

JAMIA Open, Volume 3, Issue 2, July 2020, Pages 306–317,
https://doi.org/10.1093/jamiaopen/ooaa005

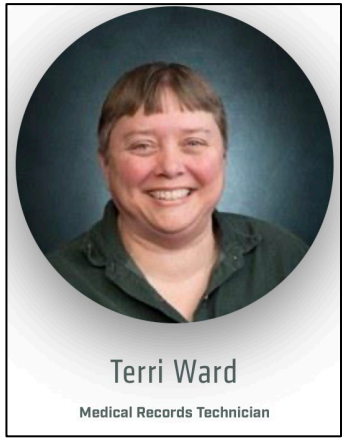


Challenges:

- ✓ Limited sources of training data
- ✓ Siloed data
- ✓ Inconsistent data formats
  - Lack insurance driver for medical coding in veterinary medicine

CSU VTH Coding

- 246,473 Encoded records (manual)
- 189,808 Unencoded Records



Created by Adam Kiehl

Goals

1. Develop a new state-of-the-art automatic diagnosis coding tool

2. Compare new tool to current state-of-the-art models for this task

3. Demonstrate the use of human clinical LLMs in the veterinary space

DeepTag (2017) & VetTag (2019)

DeepTag: inferring diagnoses from veterinary clinical notes

Allen Nie[1], Ashley Zehnder[1], Rodney L. Page[2], Yuhui Zhang[3], Arturo Lopez Pineda[1], Manuel A. Rivas[1], Carlos D. Bustamante[1,4] and James Zou[1,4]

VetTag: improving automated veterinary diagnosis coding via large-scale language modeling

Yuhui Zhang[1], Allen Nie[2], Ashley Zehnder[2], Rodney L. Page[3] and James Zou[2,4]

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Created by Adam Kiehl

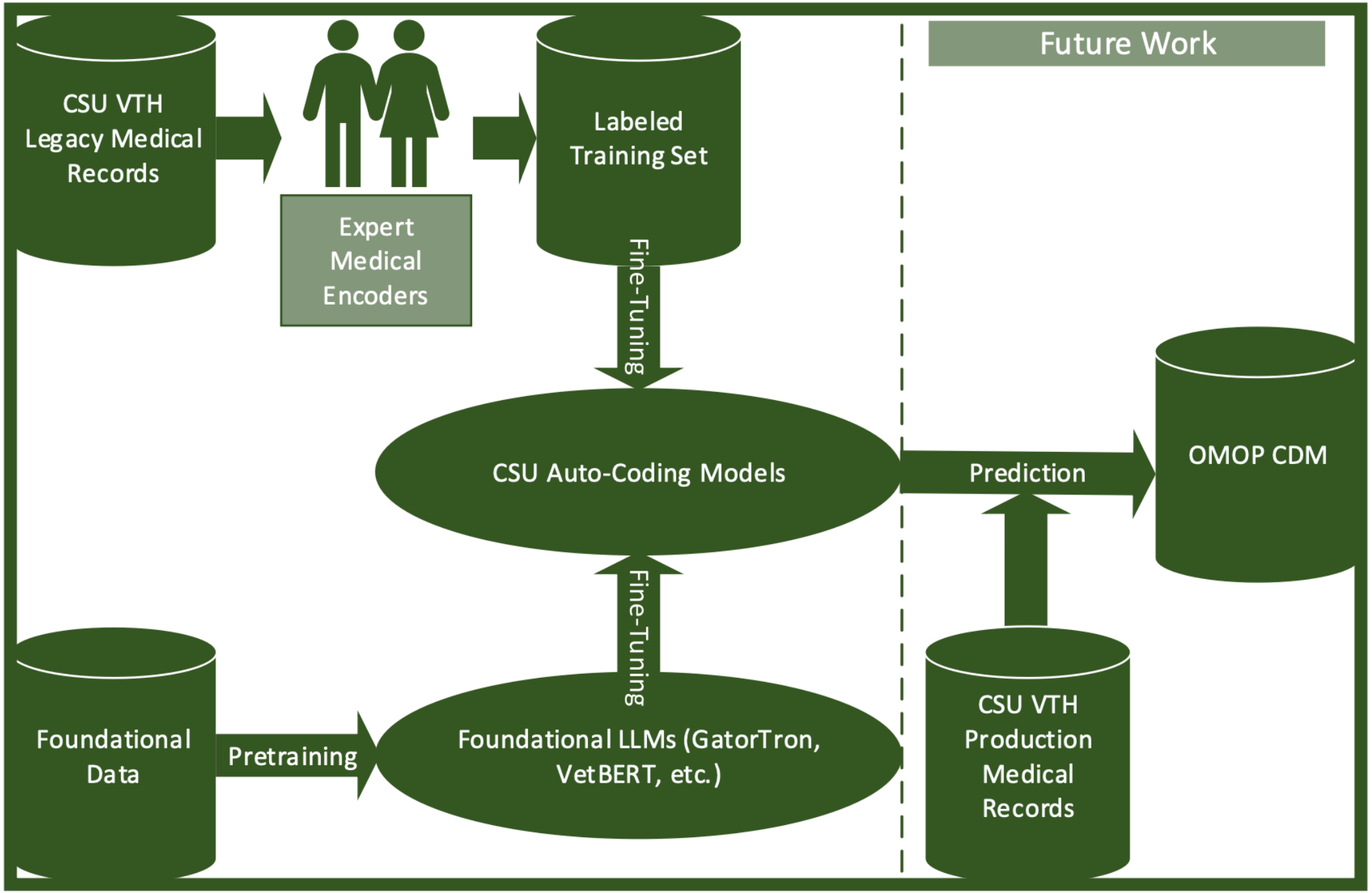
AI Tool Development at CSU

Data Science Research Institute

1. Foundational Models
   a) Large language models (typically publicly available)
   b) General or domain knowledge
2. Append Layers to Model
   a) Increases complexity
   b) Based on downstream task
3. Fine-Tune Models

Adam Kiehl
Health Data Scientist

Created by Adam Kiehl



Created by Adam Kiehl

Fine-Tuning Results

Model	F1	Precision	Recall	Exact Match
GatorTron	**74.9**±0.29	**80.8**±1.14	**71.8**±1.00	**51.6**±0.29

GatorTron

- 2021
- Partnership between University of Florida, NVIDIA
- University of Florida Health (2.9M human medical notes), MIMIC-III, PubMed, Wikipedia (91B words)

Current Findings (paper in progress):

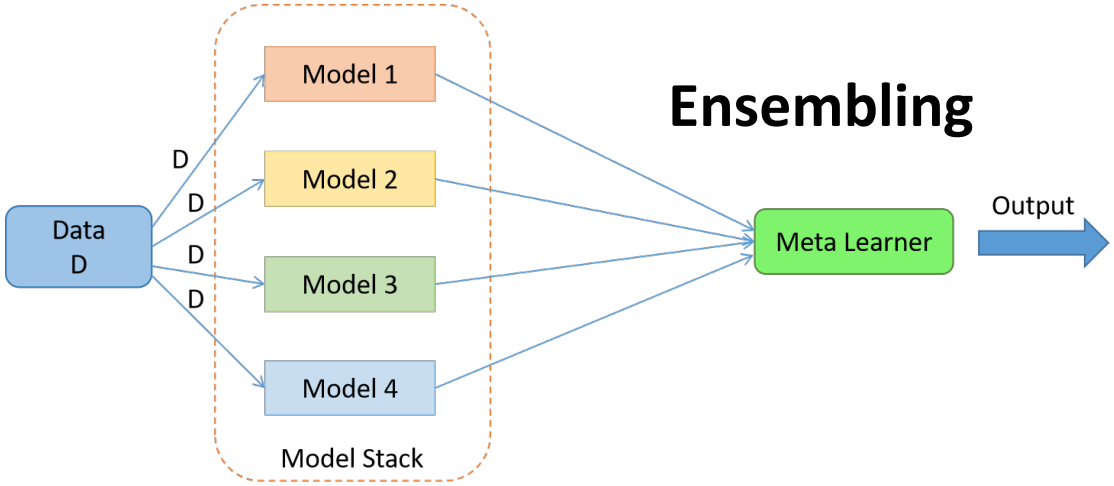
1. Foundational LLMs
   - Pretraining matters - can't fine-tune from scratch
   - Larger foundational models matter, to a point
2. Source of Pretraining Data
   - Human clinical data useful for veterinary tasks
   - May not be sensitive to content of pretraining data
3. Training Hyperparameters
   - Task-specific fine-tuning necessary
   - Fine-tune as much as possible without inducing forgetting
   - Volume of fine-tuning data matters, to a point

Next Steps:

Explainability

[84923006.0] continued from previous invoice for ardec septicemia clostridial enteritis . continued visit into new invoice for ardec . calf x1 ##30 presented to csu on 2 / 26 / 10 with a 24 hour history of diarrhea , lethargy and weakness . she had been down in the pasture but was able to stand on presentation . on presentation at the csu vth , she had very watery diarrhea with milk clots . she was

Ensembling



Textual Complexity

DIAGNOSES:
1. Lymph nodes, tonsil, spleen, nasal mucosa, kidney, bone marrow, lung: Lymphoma or lymphoid leukemia.
2. Nasal cavity: Rhinitis, necrotizing and ulcerative, marked.
3. Liver: Hydropic degeneration, marked.

Generalizability

UC DAVIS
UNIVERSITY OF CALIFORNIA

Tufts
UNIVERSITY

CSU Foundational Model (VTH-BERT) based on veterinary language

Created by Adam Kiehl

CU-CSU Data Linkage Pilot Project:

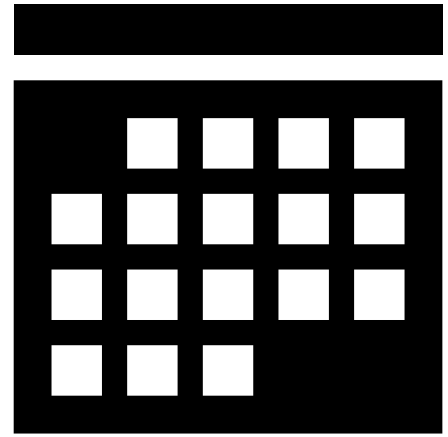
Goal: determine the feasibility of linking veterinary and human health records between UCDenver-Anschutz (UCD-AMC) and CSU-VHS for translational and comparative studies to improve patient outcomes.

- Determine overlap between people/households who receive healthcare locally and also bring their companion animals for treatment at the VTH.

- If the overlap is *significant*, determine how to undertake future studies that may look for similarities in diagnosis, therapies, etc., between people and their pets.

Process:

- Multiple years
- Many conversations
- Two IRBs
- Data Use/Sharing Agreement



Methods:

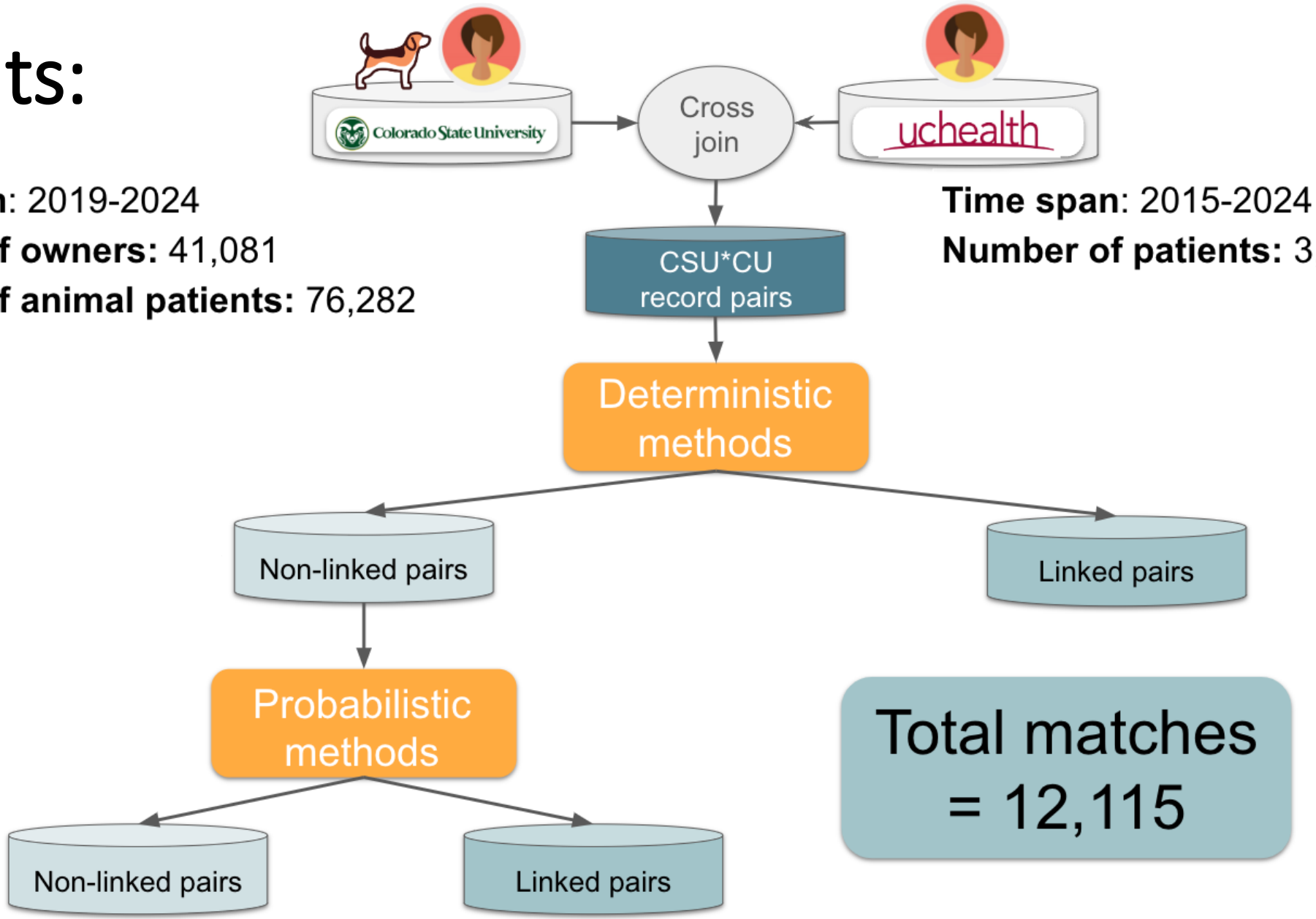
1. CSU-VTH dataset:
    - Extracted non-medical Personally Identifying Information (PII) from certain veterinary medical record (EMR) fields about animal owners
    - Securely delivered by CSU to Health Data Compass (HDC)
    - Uploaded through a Compass Eureka virtual machine (VM) gateway

2. Non-medical PII from CSU-VTH client records matched to patient records in the HDC HIPAA-compliant research data warehouse (RDW).

3. De-identified aggregate count was returned to the PI to share with CSU-VTH to determine the go/no-go status of a larger registry project

Results:

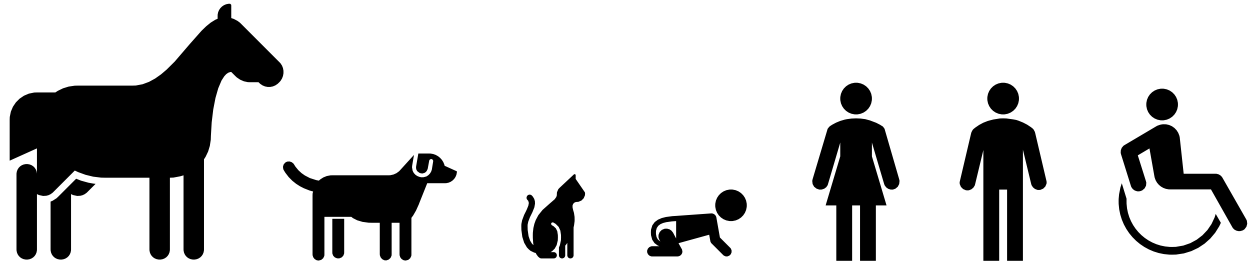
Time span: 2019-2024
Number of owners: 41,081
Number of animal patients: 76,282



Time span: 2015-2024
Number of patients: 3,282,860

Created by Katie Mullen

Next Steps:



- Create pet-patient registry to evaluate shared health questions

  example: leveraging companion animal and human EHR linkage to explore sex differences in household metabolic syndrome (MetS) co-occurrence

- Incorporate Ontologies

- Consider adding question about pet ownership to human medical records

Next Steps:

- Dissemination of research:
  - 11 abstracts/presentations
  - 7 manuscripts in progress
- Workforce development:

COHA Translational Research Fellowships

Nadia Saklou - COHA fellow

NADIA SAKLOU, DVM

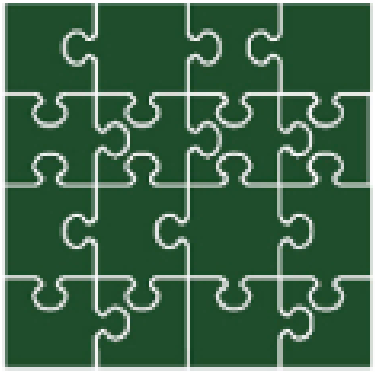
Residency trained in Large Animal Internal Medicine, current PhD student at Virginia-Maryland College of Veterinary Medicine

Veterinary Informatics and One Health Applications

Mentor Team at Colorado State University and the University of Colorado Anschutz Medical Campus:

- Tracy Webb, DVM, PhD
- Sue VandeWoude, DVM, DACLAM
- Melissa Haendel, PhD

Team Acknowledgments



- **CSU Data Science Research Institute**
  - Mayla Boguslav
  - David Kott
  - Michael Kirby
- **CSU CVMBS IT Services**
  - Paul Norvelle
  - Sally Winden
  - Wenjing Jiang
  - Adam Kiehl
  - Joe Strecker
- **CSU VTH**
  - Terri Ward
  - Tracy Webb
  - Nadia Saklou
  - Tara Marmulak
- **University Partners**
  - Chris Brandt – UC Davis
  - Manlik Kwong - Tufts
  - Allison Zwingenberger – UC Davis
  - Michael Kahn – UC Denver
  - Ian Brooks – UC Denver
  - Tislab
    - Melissa Haendel – UNC
    - Katie Mullen – UNC
    - Sabrina Toro -- UNC
- **Student Team**
  - Dawson Eliasen
  - Carson Eliasen
  - James Rudd
  - Maria Samson
  - Chris Reekie
  - Madelyn Reynolds
  - Marissa Ellingson
  - Krishnavamsi Gujju
  - Erik Gary