

Automatic Identification of Rare Disease Epidemiology Studies

Jennifer John, Qian Zhu, PhD, Dr. Eric Sid



National Center
for Advancing
Translational Sciences

Rare disease epidemiology

- Prevalence/incidence estimates or case/family counts
- Informs research funding
- Suggests patterns in disease etiology
- Presented through epidemiology studies or case reports



NIH
National Center
for Advancing
Translational Sciences

Current process

GARD curator manually searches for epidemiology studies on PubMed (tedious and error-prone)

New process

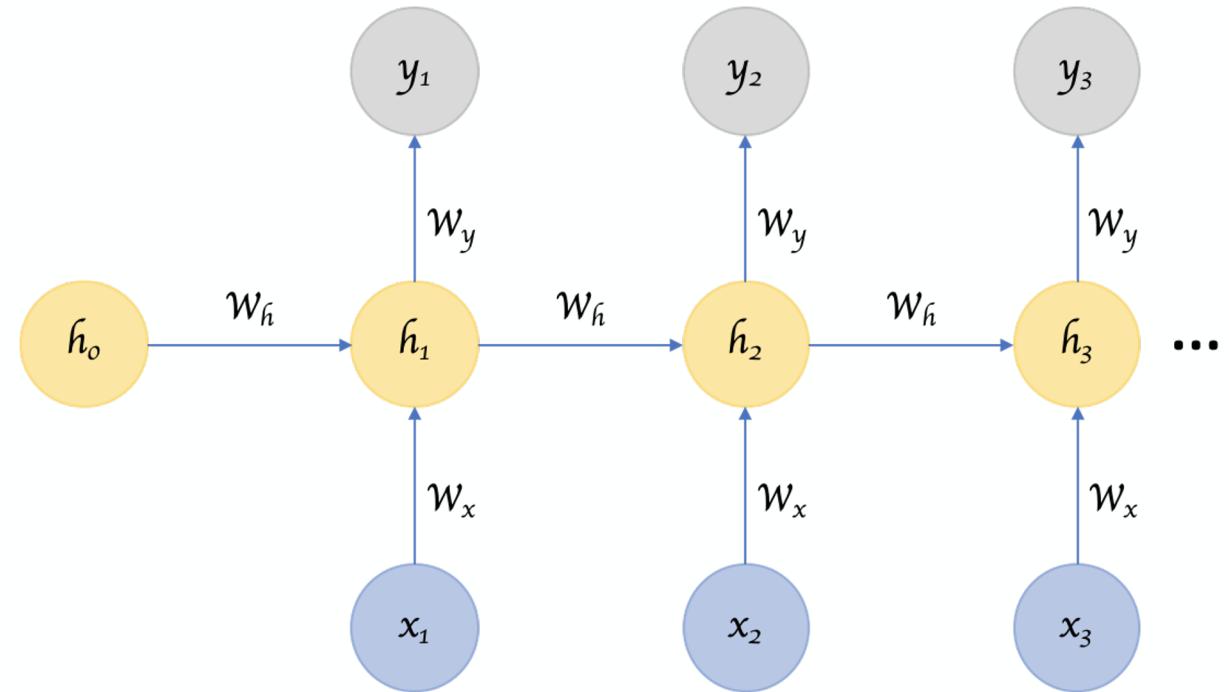
Epidemiology classification model alerts GARD as new epidemiology studies are published



National Center
for Advancing
Translational Sciences

Materials

- Recurrent neural network
- Positive dataset from Orphanet
- Negative dataset from PubMed



Results

	Holdout test set	Evaluation set
Precision	0.846	0.726
Recall	0.937	0.700
F1 score	0.886	0.701
AUC	0.967	0.751



National Center
for Advancing
Translational Sciences

> Arch Dis Child. 2003 Mar;88(3):263-4. doi: 10.1136/adc.88.3.263.

Birth prevalence of Prader-Willi syndrome in Australia

A Smith ¹, J Egan, G Ridley, E Haan, P Montgomery, K Williams, E Elliott

Affiliations + expand

PMID: 12598399 PMCID: [PMC1719461](#) DOI: [10.1136/adc.88.3.263](#)

Free PMC article

→ 99.85%

Abstract

This is the first population based study to estimate the birth prevalence of DNA proven Prader-Willi syndrome. Thirty infants were reported to the Australian Paediatric Surveillance Unit between 1998 and 2000, a prevalence of 4 per 100,000 live births or approximately 1/25,000 live births per annum.



National Center
for Advancing
Translational Sciences

> Am J Med Genet. 1999 Jul 16;85(2):183-4.
doi: 10.1002/(sici)1096-8628(19990716)85:2<183::aid-ajmg15>3.0.co;2-4.

Dandy-Walker malformation with postaxial polydactyly: further evidence for autosomal recessive inheritance

D P Cavalcanti¹, M A Salomão



3.93%

Affiliations + expand

PMID: 10406674 DOI: 10.1002/(sici)1096-8628(19990716)85:2<183::aid-ajmg15>3.0.co;2-4

Abstract

We describe an infant with Dandy-Walker malformation and tetramelic postaxial polydactyly type 1A. Parental consanguinity reinforces previous suggestions for autosomal recessive inheritance.



NIH
National Center
for Advancing
Translational Sciences

Next steps

- Submitting research paper to AMIA next week
- Apply information extraction
- Adapt to other publication types (case reports, natural hx, clinical trials)
- Combine with UX research



NIH ➤ National Center
for Advancing
Translational Sciences