

# Ian vs THE MACHINE

Using a PubMed Author Disambiguation  
Tool to Identify Publications

Mahria Lebow

Digital Projects Librarian, Kaiser Permanente Center for Health Research

PNC/MLA October 2016

# What I will touch on today:

What is a Researcher Networking System

PubMed Name Disambiguation Engine

Why Ian vs The Machine

# KAISER PERMANENTE. Researcher Profiles

**Keywords**

**Last Name**

**KP Region**

[More Search Options](#)

- Menu**
- [Find People](#)
  - [Find Everything](#)
  - [About This Site](#)
  - [Edit My Profile](#)
  - [Export RDF ?](#)
  - [Login to Profiles](#)

## Bobbi Jo Yarborough, PsyD

**Title** Assistant Investigator  
**Institution** Kaiser Permanente Northwest  
**Department** Center for Health Research - Northwest  
**Address** 3800 N. Interstate Avenue  
 Portland OR 97227  
**Email** [Bobbi.Jo.Yarborough@kpchs.org](mailto:Bobbi.Jo.Yarborough@kpchs.org)  
**Card Download** [Download vCard](#)



# What is a Researcher Networking System?

### Overview

Dr. Bobbi Jo Yarborough is a clinical psychologist and mental health services researcher. Her work focuses on improving care and outcomes and promoting recovery among individuals with serious mental illnesses, substance use problems, or both.

Dr. Yarborough joined the Center for Health Research in 2000. Her early work with Drs. Greg Clarke and Lynn DeBar focused on modifying evidence-based treatments developed for adults to manage depression and eating disorders so that they were acceptable and developmentally appropriate for adolescents. Together with Dr. Clarke she also modified a cognitive-behavioral therapy program for adults with subclinical and clinical depression for suitability as an online treatment. With Dr. DeBar she developed a primary care-based lifestyle intervention for overweight adolescent females and a cognitive behavioral treatment for binge eating in adolescent girls.

In 2005, Dr. Yarborough began collaborating with Dr. Carla Green on the STRIDE study, in which a weight loss program developed for adults in the general population was modified to meet the specific and unique needs of adults living with serious mental illnesses and/or taking antipsychotic medications. Materials from this study, including a facilitator guide, participant workbook, and food and fitness diary, are available for download.

Dr. Yarborough also played a key role on a project funded by the Agency for Healthcare Research and Quality that culminated in a definition of patient-centered services specific to people with serious mental illnesses, a theoretical model of a learning care delivery system, and an analytic framework to guide future comparative effectiveness research and patient-centered outcomes research.

Dr. Yarborough earned her bachelor's degree in psychology from the University of Oregon and her

### Yarborough's Networks

Click the "See All" links for more information and interactive visualizations!

### Concepts ?

- Life Style
  - Mental Health Services
  - Psychotic Disorders
  - Overweight
  - Mental Disorders
- 

### Co-Authors ?

- DeBar, Lynn
  - Perrin, Nancy
  - Green, Carla
  - Clarke, Gregory
  - Stevens, Victor
- 

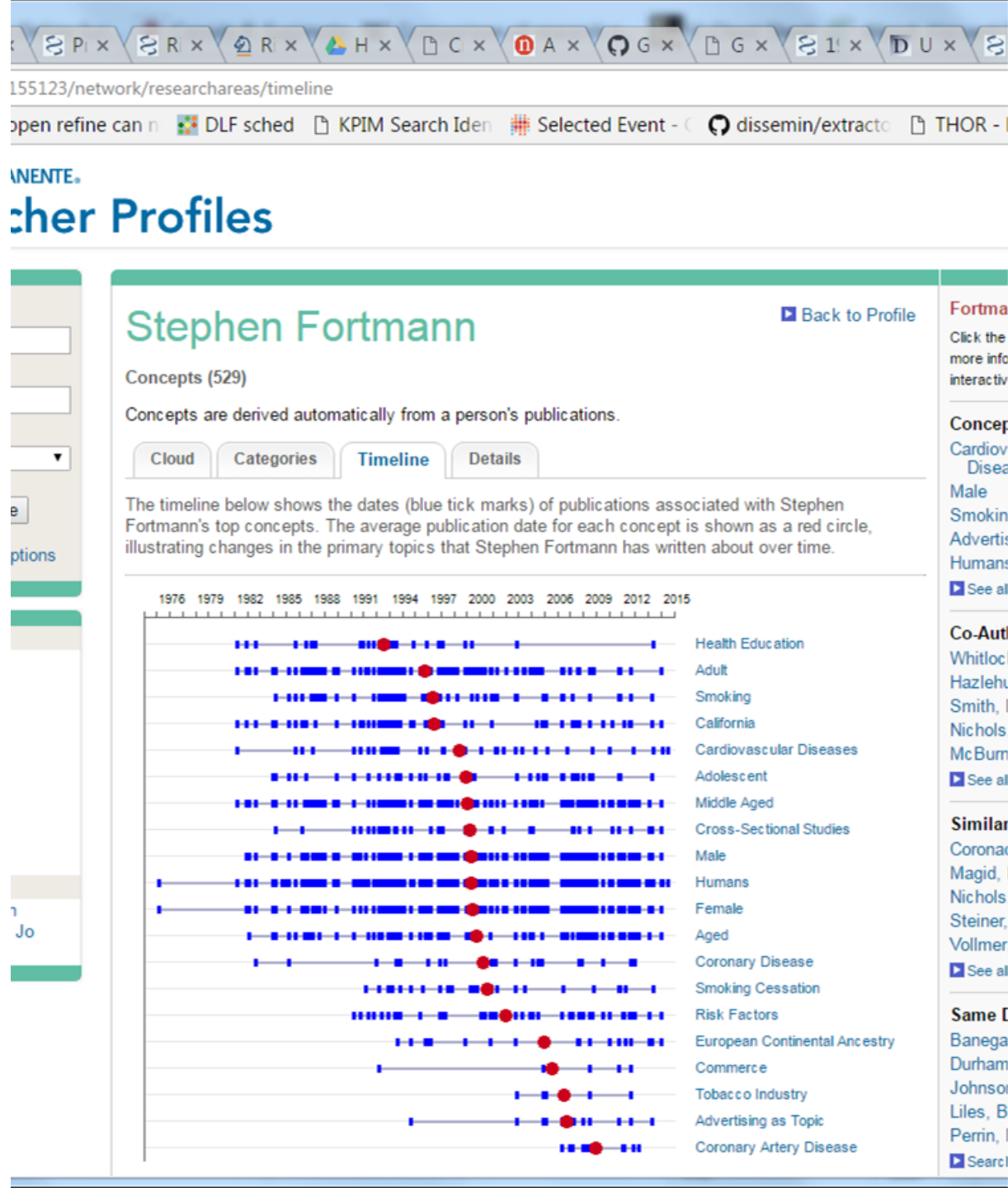
### Similar People ?

- Magid, David
  - Green, Carla
  - Steiner, John
  - Stevens, Victor
  - Vollmer, William
- 

### Same Department ?

- Clarke, Gregory
  - Naleway, Allison
  - Nichols, Gregory
  - Smith, Ning
  - Whitlock, Evelyn
-

# Concepts Derived From Research



# Concepts Derived From Research

## Researcher Profiles

**Keywords**

**Last Name**

**KP Region**

[More Search Options](#)

- Menu**
- [Find People](#)
  - [Find Everything](#)
  - [About This Site](#)
  - [Edit My Profile](#)
  - [Export RDF ?](#)
  - [Login to Profiles](#)
- History**
- Dickerson, John
  - Fortmann, Stephen
  - Health Care Reform
  - Health Education
- [See All \(4\) pages](#)

### Health Education

"Health Education" is a descriptor in the National Library of Medicine's controlled vocabulary thesaurus, MeSH (Medical Subject Headings). Descriptors are arranged in a hierarchical structure, which enables searching at various levels of specificity.

**MeSH information**

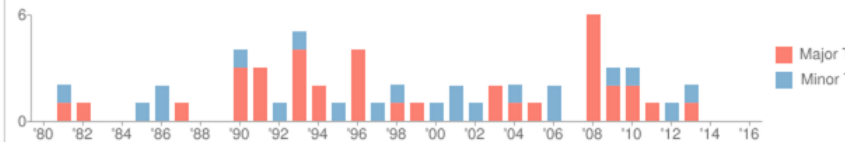
[Definition](#) | [Details](#) | [More General Concepts](#) | [Related Concepts](#) | [More Specific Concepts](#)

Education that increases the awareness and favorably influences the attitudes and knowledge relating to the improvement of health on a personal or community basis.

**publications**

[Timeline](#) | [Most Recent](#)

This graph shows the total number of publications written about "Health Education" by people in this website by year, and whether "Health Education" was a major or minor topic of these publications.



To see the data from this visualization as text, [click here](#).

**People ?**

- Coronado, Gloria
  - Fortmann, Stephen
  - Steiner, John
  - Goddard, Katrina
  - Binswanger, Ingrid
- See all people

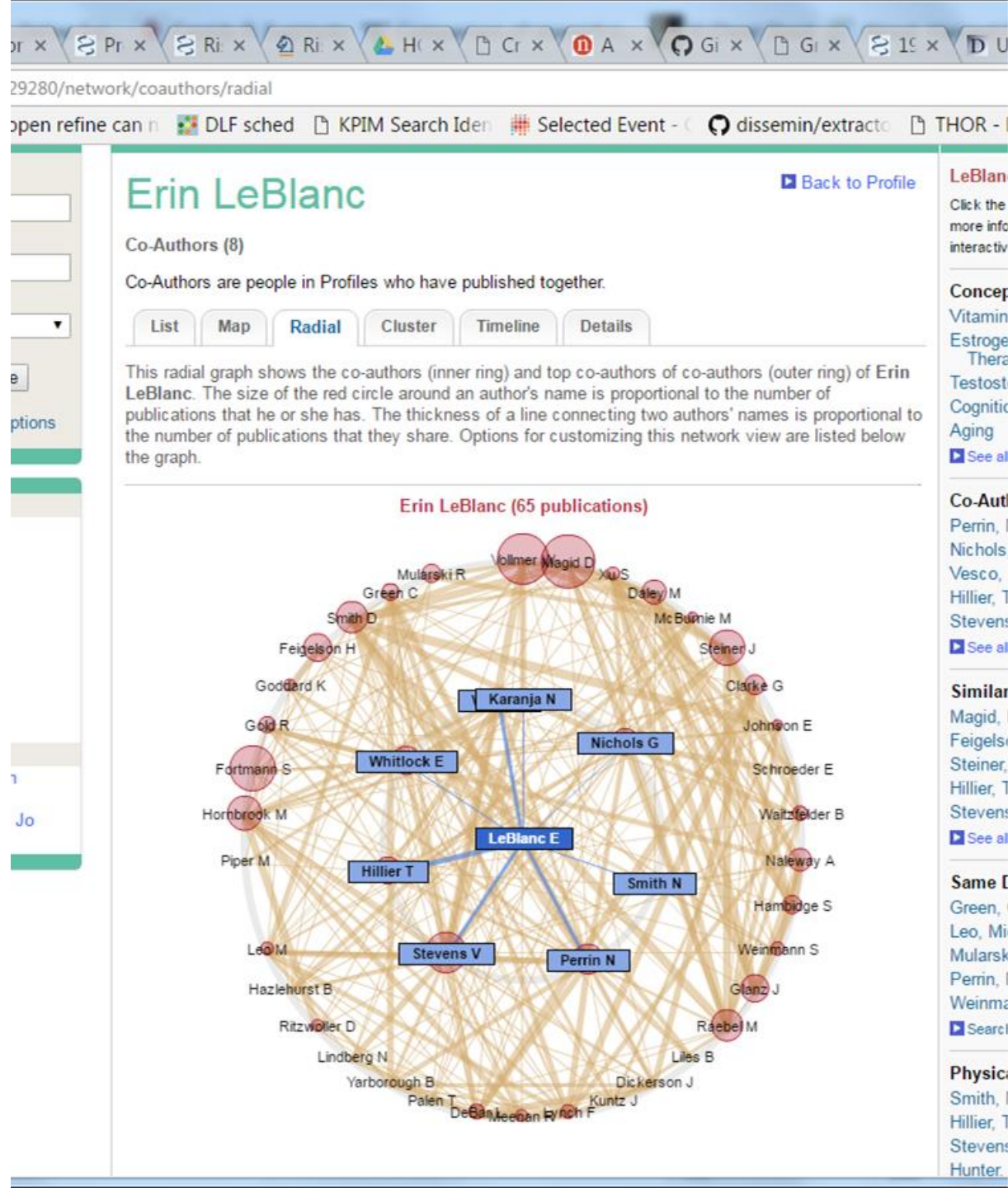
**Similar Concepts ?**

- Adult
- Cardiovascular Diseases
- Female
- Health Knowledge, Attitudes, Practice
- Hispanic Americans
- Humans
- Male
- Middle Aged
- Neoplasms
- Smoking

**Top Journals ?**

- Am J Epidemiol
- J Community Health
- JAMA
- Health Educ Res
- Am J Public Health
- Am J Prev Med
- Health Promot Pract
- Prev Med
- Circulation
- Arch Pediatr Adolesc Med

# Co-author Networks



# What is a name disambiguation engine?





# Who is Ian?

Disclaimer: Ian has seen this presentation and approved the content



# How does it work in Profiles RNS?

The screenshot displays a REST client interface for a POST request to `http://profiles.catalyst.harvard.edu/services/GetPMIDs/default.asp`. The request body is in raw XML format, containing fields for name, email, affiliations, and PMIDs. The response body is also in raw XML format, showing a list of PMIDs.

**Request Body (Raw XML):**

```
<FindPMIDs>
<Name>
<First>Mark</First>
<Middle>A</Middle>
<Last>Schmidt</Last>
<Suffix />
</Name>
<EmailList>
<email>.....@kp.org</email>
<email>.....@kpchr.org</email>
</EmailList>
<AffiliationList>
<Affiliation> %Kaiser Permanente% </Affiliation>
<Affiliation> %Permanente% </Affiliation>
<Affiliation> %Center for Health Research%Portland% </Affiliation>
<Affiliation> %Center for Health Research%Oregon% </Affiliation>
<Affiliation> %Center for Health Research%Hawaii% </Affiliation>
<Affiliation> %Center for Health Research%Northwest% </Affiliation>
<Affiliation> %Institute for Health Research%Denver% </Affiliation>
<Affiliation> %Institute for Health Research%Colorado% </Affiliation>
<Affiliation> %KP Research Affiliates Evidence-based Practice Center% </Affiliation>
<Affiliation> %@kp.org% </Affiliation>
<Affiliation> %kpchr.org% </Affiliation>
</AffiliationList>
<LocalDuplicateNames>1</LocalDuplicateNames>
<RequireFirstName>>false</RequireFirstName>
<MatchThreshold>0.98</MatchThreshold>
<PMIDAddList>
<PMID>26176570</PMID>
<PMID>26417034</PMID>
<PMID>26804385</PMID>
<PMID>27028626</PMID>
<PMID>27115485</PMID>
<PMID>27186944</PMID>
<PMID>27216163</PMID>
<PMID>26176570</PMID>
<PMID>25371489</PMID>
<PMID>25030940</PMID>
<PMID>24991025</PMID>
<PMID>24138765</PMID>
</PMIDAddList>
</FindPMIDs>
```

**Response Body (Raw XML):**

```
<PMIDList>
<PMID>26417034</PMID>
<PMID>26804385</PMID>
<PMID>27028626</PMID>
<PMID>27115485</PMID>
<PMID>27186944</PMID>
<PMID>27216163</PMID>
<PMID>26176570</PMID>
<PMID>25371489</PMID>
<PMID>25030940</PMID>
<PMID>24991025</PMID>
<PMID>24138765</PMID>
</PMIDList>
```

## **Mark M. Schmidt**

- Mark Schmidt
- Kaiser Permanente
- Pregnancy (SU)
- Risk factors (SU)
- Diabetes (SU)
- Hawaii/Epidemiology (SU)
- Hillier, T (co-author)
- Mullen, Judith (co-author)

## **Mark A. Schmidt**

- Mark Schmidt
- Kaiser Permanente
- Cohort studies (SU)
- Hep B (SU)
- Vaccines (SU)
- Vijayadeva, V (co-author)
- Trinacty, Connie (co-author)
- J Viral Hepat (journal)

# Methods

Date range: November 2015 through August 2016

Number of publications added by Ian to KPPL: 1592, 684 unique KP authors

Full name	Rank	# of pubs in PM with last name match	Mahria's difficulty rating
Steven Fortmann	Senior	306	Easiest
Erin LeBlanc	Senior	5957	Easy
Gregory Clarke	Senior	26426	Hard
Bobbi Jo Yarborough	Junior	210	Harder
Jessica Hunter	Junior	20896	Hardest
Mark Schmidt	Junior	55721	Hardest

# Results

Name	# of pubs added to KPPL (lan)	Profiles false hits (RNS)	Profiles misses (RNS)	KPPL misses (lan)	% missed (RNS as compared to lan)
Steven Fortmann	11	0	0 (1 epub)	0	0%
Erin LeBlanc	17	1	1 (1 epub)	0	6%
Gregory Clarke	5	0	0 (2 ebooks)	0	0%
Bobbi Jo Yarborough	8	1	5	0	62%
Jessica Hunter	2	0	1 (1 epub)	N/A?	100%
Mark Schmidt	7	0	6 (1 epub)	N/A?	100%

# Querying via Postman

The screenshot displays a Postman interface for a POST request to `http://profiles.catalyst.harvard.edu/services/GetPMIDs/default.asp`. The request body is an XML document with the following structure:

```
<FindPMIDs>
  <Name>
    <First>Mark</First>
    <Middle>A</Middle>
    <Last>Schmidt</Last>
    <Suffix />
  </Name>
  <EmailList>
    <email>....@kp.org</email>
    <email>....@kpchr.org</email>
  </EmailList>
  <AffiliationList>
    <Affiliation> %Kaiser Permanente% </Affiliation>
    <Affiliation> %Permanente% </Affiliation>
    <Affiliation> %Center for Health Research%Portland% </Affiliation>
    <Affiliation> %Center for Health Research%Oregon% </Affiliation>
    <Affiliation> %Center for Health Research%Hawaii% </Affiliation>
    <Affiliation> %Center for Health Research%Northwest% </Affiliation>
    <Affiliation> %Institute for Health Research%Denver% </Affiliation>
    <Affiliation> %Institute for Health Research%Colorado% </Affiliation>
    <Affiliation> %KP Research Affiliates Evidence-based Practice Center% </Affiliation>
    <Affiliation> %@kp.org% </Affiliation>
    <Affiliation> %kpchr.org% </Affiliation>
  </AffiliationList>
  <LocalDuplicateNames>1</LocalDuplicateNames>
  <RequireFirstName>>false</RequireFirstName>
  <MatchThreshold>0.98</MatchThreshold>
  <PMIDAddList>
    <PMID>26176570</PMID>
    <PMID>26417034</PMID>
  </PMIDAddList>
</FindPMIDs>
```

The response body is a JSON array of PMID objects:

```
{
  "PMIDList": [
    { "PMID": "26417034" },
    { "PMID": "26804385" },
    { "PMID": "27028626" },
    { "PMID": "27115485" },
    { "PMID": "27186944" },
    { "PMID": "27216163" },
    { "PMID": "26176570" },
    { "PMID": "25371489" },
    { "PMID": "25030940" },
    { "PMID": "24991025" },
    { "PMID": "24138765" }
  ]
}
```

# Results - Part 2

Name	# of <i>pubs</i> added to KPPL <i>(lan)</i>	# of pubs found <i>(Postman request)</i>	# of false hits <i>(Postm an reques t)</i>	# of misses <i>(Postman request)</i>	# of KPPL misses <i>(lan)</i>	% missed <i>(Postman as compared to lan)</i>
Bobbi Jo Yarborough	8	8	0	0	0	0%
Jessica Hunter	2	0	0	1 (1 epub)	0	100%
Mark Schmidt	7	9	0	0 (1 epub)	2*	0%

# Conclusion

?

# References

1. Weber, G. Profiles Research Networking Software PubMed Disambiguation Service. 01/20/2012.  
<http://profiles.catalyst.harvard.edu/Meetings/ProfilesRNSDisambiguation120120.pdf>
2. TORVIK, V. I., & SMALHEISER, N. R. (2009). Author Name Disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data*, 3(3), 11. (aside: this is the AUTHOR-ity paper)
3. Profiles Research Networking Software (RNS) Disambiguation Engine Documentation.  
[http://profiles.catalyst.harvard.edu/docs/ProfilesRNS\\_DisambiguationEngine.pdf](http://profiles.catalyst.harvard.edu/docs/ProfilesRNS_DisambiguationEngine.pdf).  
[Viewed 10/4/16/](#).
4. Torvik, V. I., Weeber, M., Swanson, D. R., & Smalheiser, N. R. (2005). *A probabilistic similarity metric for medline records: A model for author name disambiguation*. *Journal of the American Society for Information Science and Technology*, 56(2), 140-158. DOI: [10.1002/asi.20105](https://doi.org/10.1002/asi.20105)