



# APPLYING ONOMASTICS TO SCIENTOMETRICS

2015-01-19

Tania Vichnevskaja,  
French National Institute for Health



# Cultural bias and diaspora through publications

2

- Applying onomastics to approach cultural bias in research
- Analysis of International Cancer Research publications through co-authorship and inter-citation

# Two approaches combined: onomastics & bibliometrics

3

- Onomastics is the study of the origin, history, and use of proper names
- Bibliometrics is a statistical and structural analysis of written publications:
  - Quantitatives: number of publications, number of citations, ranking of the top 1% or 10% of cited publication.
  - Structural: Configuration of co-authorship, and inter-citation on different levels: Authorship/University/Country

# YOU are ALL onomasticians !

4

Active participants at the IREG conference:

- Prof. **Gero Federkeil**, CHE (Coordinator of Multi-Ranking),
- Prof. **Nian Cai Liu**, Jiao Tong University in Shanghai
- Prof. **Seeram Ramakrishna**, The National University of Singapore,
- Prof. **Santo Fortunato**, Aalto University,
- Prof. **Karin Stana Kleinschek**, The University of Maribor,
- Prof. **Henryk Ratajczak**, member of Czech Academy of Sciences,
- Prof. **Edvard Kobal**, Slovenian Science Foundation,
- **Roberta Sinatra**, PhD, Northeastern University,
- **Tania Vichnevskaja**, French National Institute for Health (INSERM),
- Prof. **Andrée Sursock**, Senior Adviser at EUA,
- Prof. **Øivind Andersen**, The University of Oslo.

What can YOU tell from the names?

# Data in Cancer Research

5

- China : The Fudan University Cancer Hospital
- USA: The Dana-Farber Cancer Institute
- France: French Comprehensive Cancer Centers (FCCC)
- Japan : National Cancer Center
- Poland : All applicable institutes
- Slovenia : All applicable institutes

# Thomson Web of Science

6

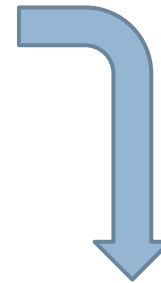
- For this study, we used Thomson Web of Science database
  - Web of Science provides access to the world's leading citation databases. Multidisciplinary content covers over 12,000 of the highest impact journals.
  - Thomson Web of Science contains information on scientists names, affiliation and citation (possibility to qualify Diaspora)
  
- In the corpus
  - 15k articles, 68k authors;
  - cited: 32k articles 168k authors;
  - 17 million author citing-author cited occurrences

# Name origin recognition : does it work?

Matrix with affiliation (row) and onomastic class (column)

7

		ie. Polish name								
		French	Brittish	Chinese	Polish	German	Japanese	Italian	...	
	US	2358	7072	2406	430	2384	680	1243		
	France	9375	318	88	93	370	42	607		
	China	13	35	5690	1	19	76	30		
ie. Poland affiliation >	Poland	23	16	3	5038	174	5	13		
	Great Brit	191	1786	57	52	145	28	115		
	Japan	9	10	26	7	7	3739	10		
	Germany	121	102	22	92	1819	9	60		
	Italy	83	4	1	11	31	3	2026		
	...									



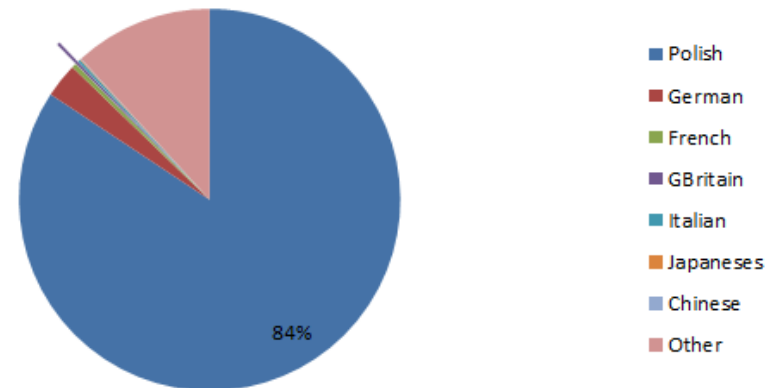
The USA was considered as a 'melting pot' of all origins, not having an onomastic class of its own.

Name origin recognition precision varies.

We can establish the strong relationship between affiliation and 'onomastic classes' for China, Japan, Poland, Slovenia.

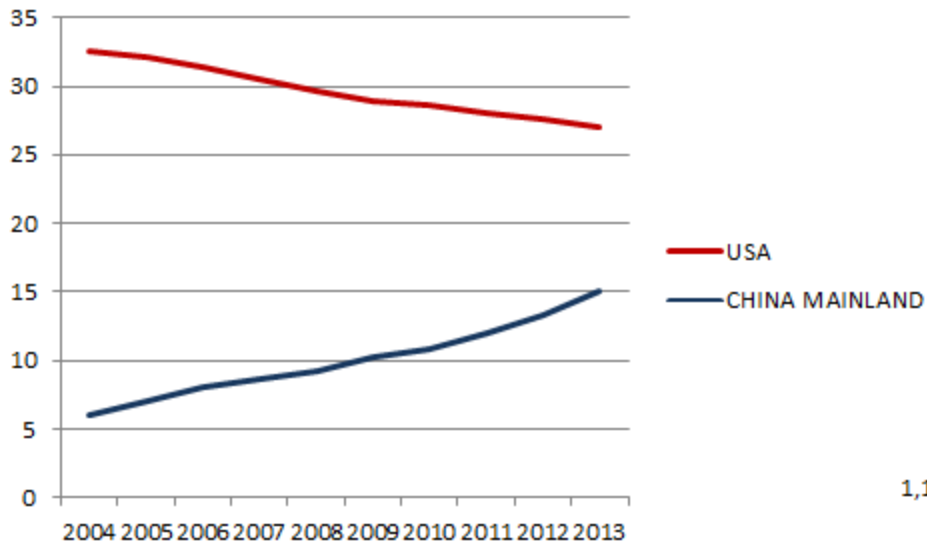
Ex: **84% of scientists affiliated to Poland have Polish names.**

**Scientists affiliated in Poland**  
Breakdown by onomastic class (Polish name, etc.)  
Source: Thomson WoS



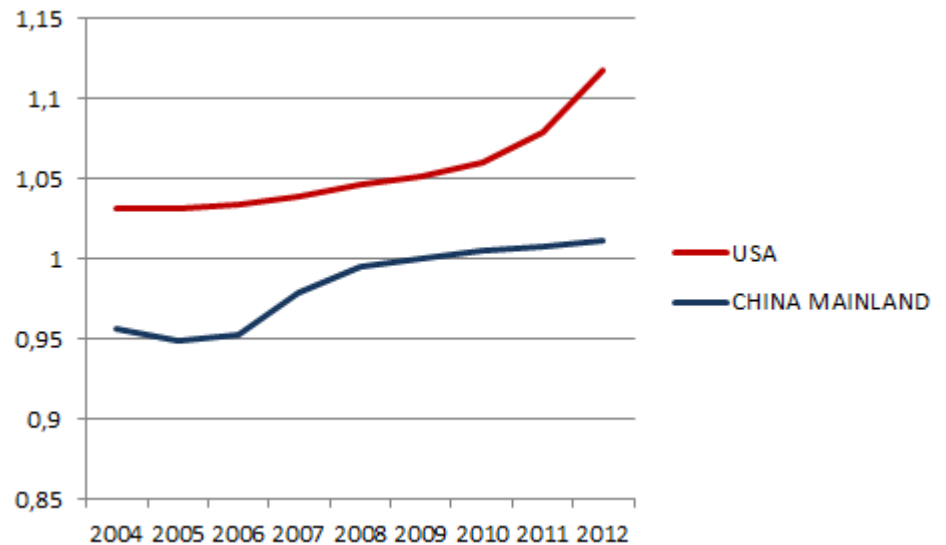
# Percentage of publications and citations

8



Percentage of Cancer Research publications in the World

Percentage of Cancer Research Documents cited relative to world publications



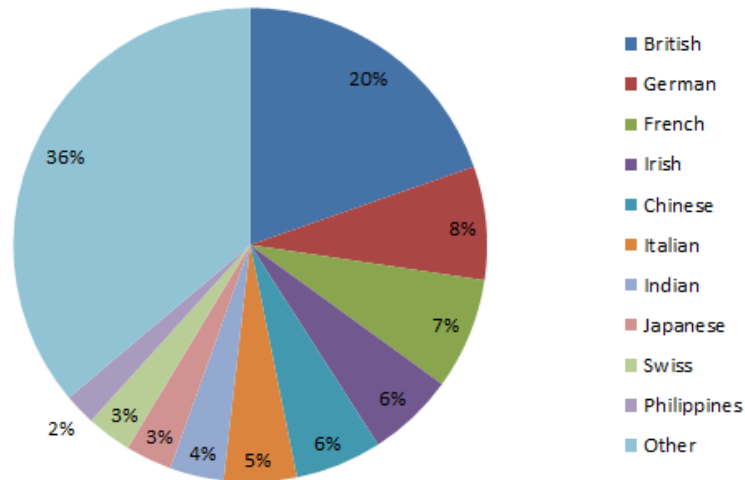


# Each corpus has a different profile

## Breakdown by co-authors onomastic classes

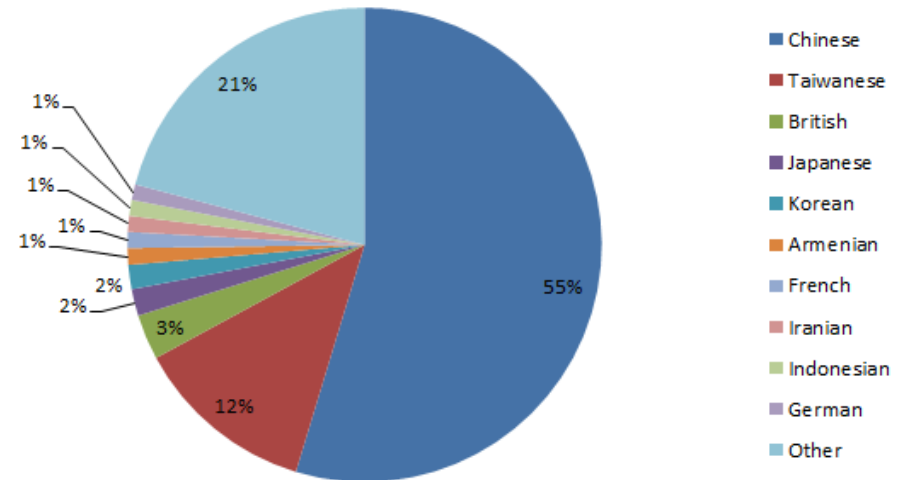
### Cancer Research US Danafarber

Co-Authors onomastics  
Source: Thomson



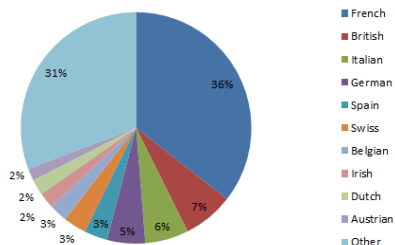
### Cancer Research China Fudan University

Co-Authors onomastics  
Source: Thomson WoS



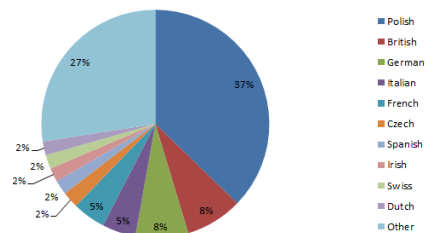
### Cancer Research France CLCC

Co-Authors onomastics  
Source: Thomson



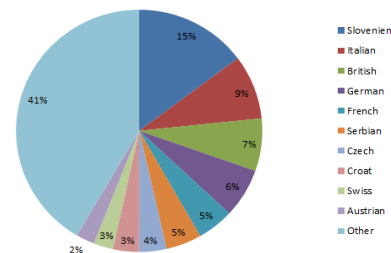
### Cancer Research Poland

Co-Authors onomastics  
Source: Thomson WoS



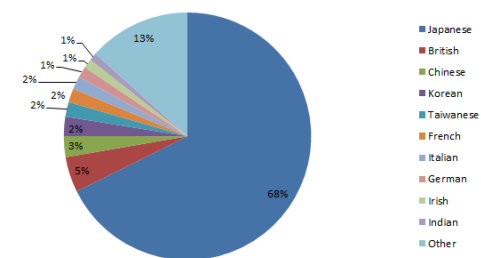
### Cancer Research Slovenia

Co-Authors onomastics  
Source: Thomson WoS



### Cancer Research Japan

Co-Authors onomastics  
Source: Thomson WoS 5



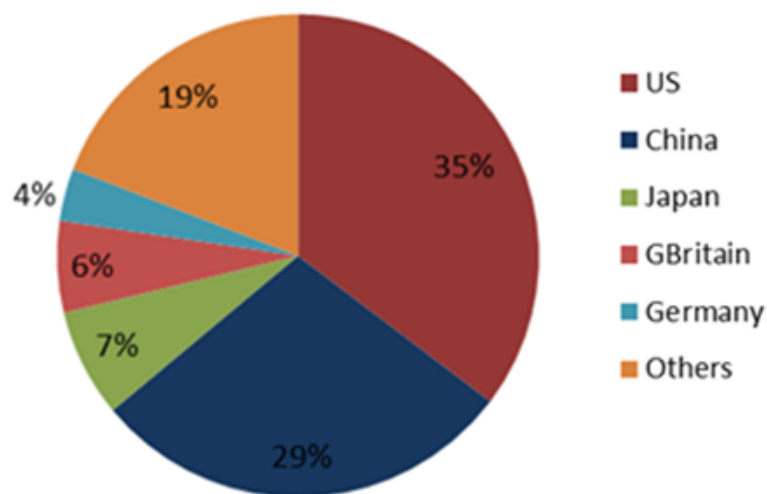
# Chinese scientists in Cancer Research Fudan, China

## Who do they cite?

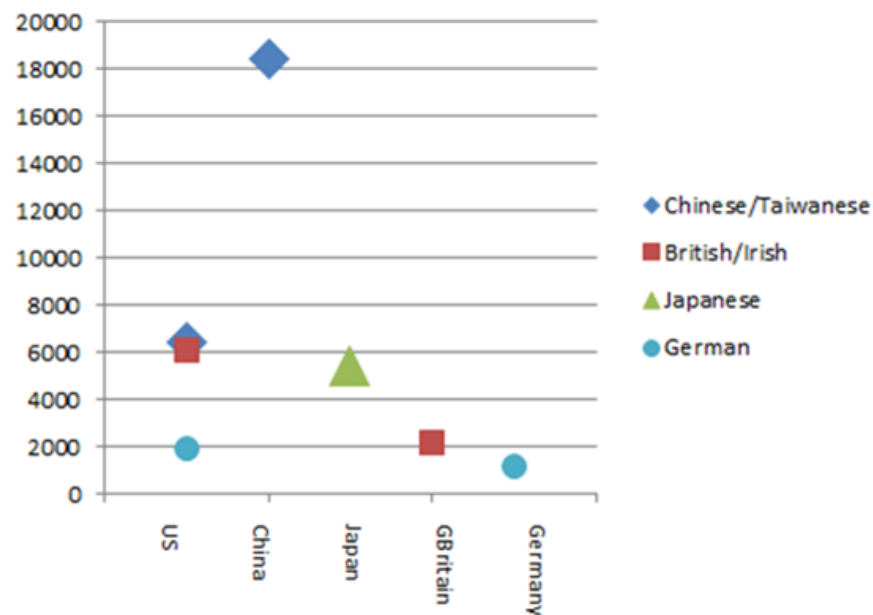
10

Authors in **Chinese** corpus, Affiliated to **China**, With a **Chinese or Taiwanese** names

Percentage of cited authors by affiliated country



Most frequently cited names by affiliated country



Most cited countries of affiliation:

- 1) **US**
- 2) China
- 3) Japan

Most cited onomastic class:

- 1) **Chinese (Chinese Scientists in the US)**
- 2) British
- 3) Japanese

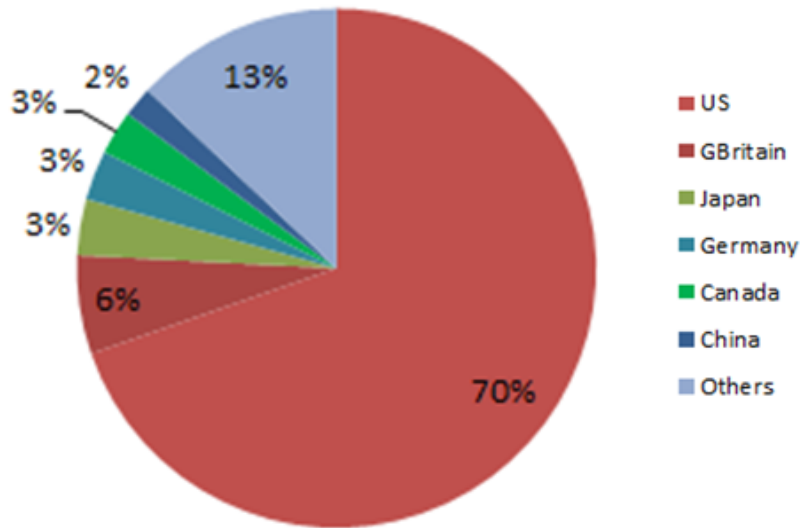
# Chinese Scientists in Cancer Research Dana-Faber, US

## Who do they cite?

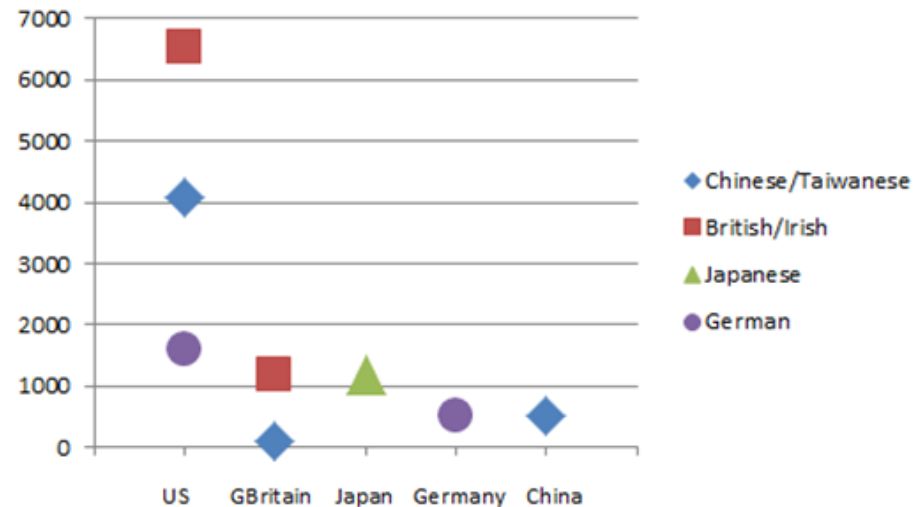
11

Scientists in **US** corpus, Affiliated to **US**, With a **Chinese or Taiwanese** names

Percentage of cited authors by affiliated country



Most frequently cited names by affiliated country



Most cited countries of affiliation:

- 1) **US**
- 2) **Great-Britain**
- 3) **Japan**

Most cited onomastic class:

- 1) **British**
- 2) **Chinese**
- 3) **Japanese**

# Diaspora a key 'hidden' dimension

12

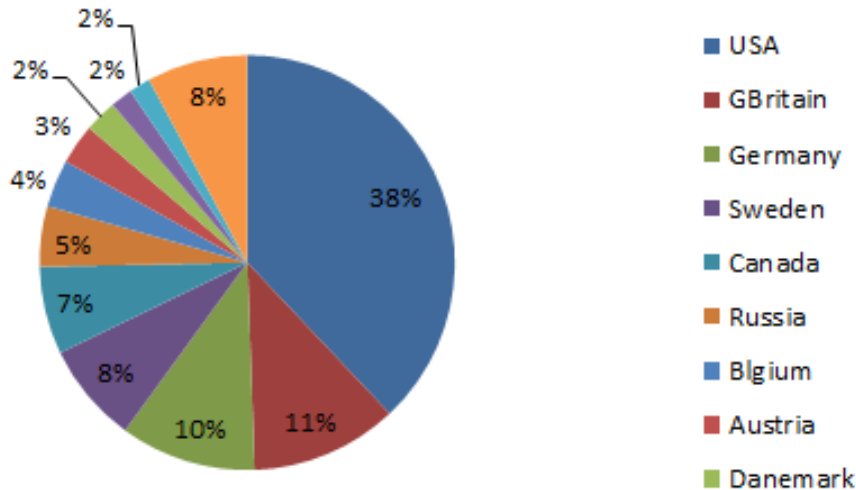
- **Ranking by Country & Affiliation** would entirely differ from **Ranking by Onomastic class** (ie. including Diasporas)
- Onomastics sheds light on the strong relationship that exists between China and its Diaspora, especially between the institutes of excellence in China and US
- **So China: Brain Drain or Brain Gain ?**

# Cancer Research in Poland and Slovenia

## Examining the 'brain drain'

13

### The Polish "brain drain"

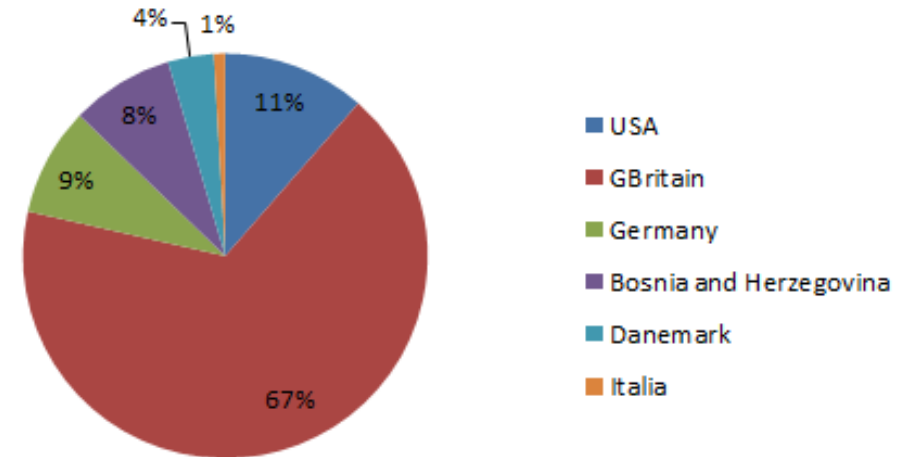


In the Polish Corpus, we look at co-authors with Polish names, affiliated abroad.

Top countries:

1. **US,**
2. **Great-Britain,**
3. **Germany.**

### The Slovenien "brain drain"



In the Slovenian Corpus, we look at co-authors with Slovenian names, affiliated abroad.

Top countries:

1. **Great-Britain,**
2. **US,**
3. **Germany.**

# Conclusion

14

- Applied onomastics can help view differences in
  - ▣ Co-authorship and citing patterns in various countries
  - ▣ Diasporas citing patterns (ex. Chinese in the US vs. non-Chinese in the US vs. Chinese in China)
  - ▣ ‘Brain drain’ and migration patterns
- The above has an potential impact on ranking, and should be further analysed
- More questions raised than answered!

# Thank you!

15

- Tatiana Vichnevskaja,
  - Contact: [tatiana.vichnevskaja@inserm.fr](mailto:tatiana.vichnevskaja@inserm.fr)
  
- About NamSor software and API
  - Contact: [elian.carsenat@namsor.com](mailto:elian.carsenat@namsor.com)
  - <http://namsor.com/>