Institutional Name Confusion on Citation Indexes: The Example of the Names of Turkish Hospitals

Zehra Taşkın & Umut Al
{ztaskin, umutal}@hacettepe.edu.tr
Outline

- Research evaluation
- Data accuracy and consistency
- Findings
  - Types of affiliation information mistakes
  - Effects of name confusion
  - Unification techniques
- Conclusion
Research Evaluation

- Evaluation of scientific products (articles, patents, etc.)
- Aims of research evaluation
  - Deciding fund distribution
  - Allocation of scarce sources
  - Using for academic appointment and staff assignment
  - Analyzing the impact of scientific outputs
  - Observation on science policy applications
- Importance of accurate data
Data Accuracy and Consistency

- Data accuracy is key to providing the best results.
- Data used in bibliometric studies should be accurate and consistent.
- Some mistakes create insoluble results.
- Unstandardized addresses are very big problem for research evaluation studies.
Methodology

- Data source *Web of Science*
- Data set cover 1928-2009
- Authored by the scholars affiliated with Turkish institutions
- Different forms of Turkish addresses (e.g., “Turkey,” “Turkiye”, “Turkei”)
- There is no document type distinction
- 198,687 Turkey-addressed publications
- Data cleaning and unification process
Unification Process

- Address information for authors found in C1 and RP fields of Web of Science
- A new column named “institution” has been created to write unified addresses for each institution by using Excel
- “DR ZEKAL TAHIR BURAK WOMEN HOSP, ANKARA, TURKEY; ZUBEYDE HANIM MATERN HOSP, ANKARA, TURKEY”, => “ZEKAI TAHIR BURAK TRH; ZUBEYDE HANIM TRH”
Aim of the Study

- Identifying institutional name confusions in citation indexes
- Turkish Training and Research Hospitals
- Specifying the most productive hospitals and mistakes in their names
- Presenting the effect of name confusion
- Displaying some collaboration maps for the effects
Findings
Types of Affiliation Information

- **Wrong spelling**
  - e.g. Dr Saaaaami Ulus Childrens Hosp (Dr. Sami Ulus TRH); Sisil Etfal Hosp (Sisli Etfal TRH)

**Prothrombin G20210A mutation in Turkish children**

**Author(s):** Gurgey, A (Gurgey, A); Unal, S (Unal, S); Okur, H (Okur, H); Duru

**Source:** PEDIATRIC HEMATOLOGY AND ONCOLOGY Volume: 22 Issue: 5

**Times Cited:** 8 (from Web of Science)

**Cited References:** 7 [view related records]  

**Abstract:** The prothrombin G20210A mutation has been described as the prothrombin gene variant namely prothrombin C20291G has also been found to thrombin variants have been searched in two different groups. Group 1: A and 2004 and screened for prothrombin G20210A mutation. Twenty-four The mean age of the patients was 6.1 years (median 5 years, range: 4 m. Fifteen of 24 children, (62.5%) had arterial thrombosis. most of whom (53.200 thrombotic patients and in 200 healthy subjects. alone of the thrombolytic arterial thrombosis as the cerebral infarct is the most prominent type of the C20291G variant is not an important risk factor for the population studied.

**Accession Number:** WOS:000229812400007

**Document Type:** Article

**Language:** English

**Author Keywords:** children with thrombosis; prothrombin C20291G; prothrombin G20210A

**KeyWords Plus:** METHYLENETETRAHYDROFOLATE REDUCTASE, RISK

**Reprint Address:** Gurgey, A (reprint author), Hacettepe Univ, Fac Med, Dep Med, Dept

**Document Type:** Article

**Language:** English

**KeyWords Plus:** BENIGN NEONATAL HEMANGIOMATOSIS

**Reprint Address:** Altunay, IK (reprint author), Barbaros Mah Yoyal Paza Sok, Yuusakent Sit E-16, Istanbul, Turkey

**Address:**
1. Sisil Etfal Hosp, Dept Dermatol, Istanbul, Turkey
2. Sisl Etfal Hosp, Dept Pediat, Istanbul, Turkey
Types of Affiliation Information

Mistakes

- Abbreviation mistakes
  - e.g. TYIH Hosp (Türkiye Yüksek İhtisas TRH);
    Dr AY Oncol Training & Res Hosp (Dr. Abdurrahman Yurtaslan TRH)
Types of Affiliation Information

Mistakes

- Translation mistakes
  - e.g. Ankara Postgrad Training Hosp, Higher Specializat Hosp, Turkey High Special Hosp, High Specializat Hosp, Adv Specialist Hosp etc. (Turkiye Yuksek Ihtisas TRH)
<table>
<thead>
<tr>
<th>Hospital</th>
<th>Pub.</th>
<th>Mistakes</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankara Numune TRH</td>
<td>2,325</td>
<td>437</td>
<td>18.7</td>
</tr>
<tr>
<td>Türkiye Yüksek İhtisas TRH</td>
<td>1,070</td>
<td>415</td>
<td>38.7</td>
</tr>
<tr>
<td>Ankara TRH</td>
<td>1,023</td>
<td>354</td>
<td>33.0</td>
</tr>
<tr>
<td>Şişli Etfal TRH</td>
<td>821</td>
<td>50</td>
<td>6.9</td>
</tr>
<tr>
<td>İzmir Atatürk TRH</td>
<td>648</td>
<td>493</td>
<td>76.0</td>
</tr>
<tr>
<td>Haydarpaşa Numune TRH</td>
<td>643</td>
<td>100</td>
<td>15.5</td>
</tr>
<tr>
<td>Dışkapı Yıldırım Beyazıt TRH</td>
<td>538</td>
<td>314</td>
<td>58.3</td>
</tr>
<tr>
<td>Dr. Siyami Ersek TRH</td>
<td>498</td>
<td>54</td>
<td>10.8</td>
</tr>
<tr>
<td>Dr. Abdurrahman Yurtaslan TRH</td>
<td>463</td>
<td>84</td>
<td>18.1</td>
</tr>
<tr>
<td>Dr. Sami Ulus TRH</td>
<td>444</td>
<td>10</td>
<td>2.2</td>
</tr>
</tbody>
</table>
Effects of Name Confusion

- Performance evaluation
- Governmental supports and academic studies
- As a Turkish case => National Academic License for Electronic Resources
- Inappropriate and different bibliometric results
Collaboration Map with Inaccurate Affiliation Information
Collaboration Map with Unified Hospital Names
Unification Techniques

- Clustering
  - Based on cleaning, sorting, clustering, checking and updating stages
  - Incorrect addresses are determined by using measure of similarity

- Finite state
  - Based on finite sets of states and its relations to each other
  - Transducers; Nooj - Xerox
Conclusions

- Data inconsistency is the main problem for evaluating by using citation databases
- Mistakes in institution names due to spelling, translation or indexing errors
- Non-standardized addresses can reduce institutional visibility
- With the non-standardized addresses, bibliometric studies can produce unreliable results
Suggestions

- Before the evaluation process, all existing institutional affiliation information must be unified
- There are some techniques to make unification automatically
- The main solution to the confusion about institutional names is to assign unique numbers to institutions
- There are some responsibilities for authors, editors, librarians, indexers and decision-makers
Institutional Name Confusion on Citation Indexes: The Example of the Names of Turkish Hospitals

Zehra Taşkıın & Umut Al
{ztaskin, umutal}@hacettepe.edu.tr