How to compare Ovid MEDLINE & PubMed

Wolters Kluwer Health - Learning Research & Practice

November 2019
Index

- Content equivalence and timing differences
- 1. What causes the perception that “PubMed retrieves more records”?
- 2. What causes the perception that “PubMed retrieves more records”?
- Range searching – how to compare between Ovid MEDLINE and PubMed
- References to understand differences between Ovid MEDLINE and PubMed
- Ovid recommendations when comparing Ovid MEDLINE to PubMed
Content equivalence and timing differences

- There are no content differences:
  - Ovid MEDLINE is 100% equivalent in content to PubMed:
    - Ovid MEDLINE (R) ALL 1946 to November 19, 2019: `docz.dz` 30,418,369 references
    - PubMed November 19, 2019: `all[sb]` 30,346,249 references

- Timing differences:
  - Ovid MEDLINE is updated daily with a 1-day delay and 2 days on Monday compared to PubMed
  - Corrections on Ovid are loaded monthly, on the 2nd Wednesday of the month. When corrections are added to the front segment (medc), the corrected records are removed from the back-files

- Daily and weekly deduplication in Ovid MEDLINE:
  - Daily dedup: between Publisher/Ahead of Print, and In-Process/Non-Indexed
  - Weekly dedup: between In-Process/Non-indexed and MEDLINE

- The MeSH reload ensures PubMed (December) and Ovid (January) are “reset” annually.

- PubMed and Ovid MEDLINE will always show daily variations

Search examples executed on 19 November 2019
1. What causes the perception that “PubMed retrieves more records”?

PubMed auto-expands, e.g.: *(kidney disease) AND (orange juice)* results in 19 records. However, the actual search executed by PubMed is much wider than you might expect or want it to be:

((("kidney diseases"[MeSH Terms] OR ("kidney"[All Fields] AND "diseases"[All Fields])) OR "kidney diseases"[All Fields]) OR ("kidney"[All Fields] AND "disease"[All Fields])) OR "kidney disease"[All Fields]) AND ((("citrus sinensis"[MeSH Terms] OR ("citrus"[All Fields] AND "sinensis"[All Fields])) OR "citrus sinensis"[All Fields]) OR "orange"[All Fields]) OR "oranges"[All Fields]) AND ((("juice"[All Fields] OR "juice's"[All Fields]) OR "juiced"[All Fields]) OR "juices"[All Fields]) OR "juicing"[All Fields])

Ovid is precise. *kidney disease AND orange juice* results in 0 records. There are no articles with those concepts in MEDLINE text fields!

- Ovid’s default is phrase searching with the ADJ (adjacency) as the default operator between words:
  - ADJ = next to each other, in that specific order: kidney next to disease, and orange next to juice
- Ovid does not automatically include and explode MeSH terms
- Ovid does not have a hierarchy for subheadings, and does not auto-include sub-sub-headings
- Ovid’s default search is “.mp.”. Only relevant text fields in MEDLINE are searched: title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms

The equivalent search on Ovid MEDLINE to match PubMed’s would be:

*(exp kidney diseases/ or (kidney and disease?).mp.) and (exp citrus sinensis/ or (citrus and sinensis).mp. or citrus sinensis.mp. or (orange? and juice*3).mp.)*

- This search retrieves 22 records, 3 more then PubMed. Missed by PubMed are pmids: **17484478, 21858427, 25103218** (19Nov19)
2. What causes the perception that “PubMed retrieves more records”?

While the input is identical, and most fields exist on both platforms, indexation may differ!

- Phrase versus word indexed fields
  - e.g. AU (author) and JN (journals name) fields are **phrase** indexed on Ovid
  - **smith.au.** searches for that author without initials, and only retrieves 26 records.
  - **smith *.au.** Finds all authors with last name smith, and retrieves 237,721 records
  - smith[Author] on PubMed retrieves 233,750 results

- Not all fields are equal:
  - When fields are named identically, indexation is usually the same/similar, e.g. **Title or Abstract**.
    - **orange.ti,ab.** on Ovid: 28,724 results. On PubMed: "orange"[Title/Abstract] 28,980 results
    - Again PubMed is auto-expanding as it includes; citation's title, collection title, abstract, other abstract and keywords: https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.TitleAbstract_TIAB On Ovid: orange.ab,kf,oa,ti,cl. 28888 results
  - When field names differ, they can be indexed differently
    - “Date of Publication” (DP) on PubMed does not match “Publication Year” (YR) on Ovid
      - DP on Pubmed includes epub dates. YR on Ovid is the actual publication year
      - PubMed: "2015"[Date - Publication] 1,252,578 results Ovid: 2015.yr. 1,091,543 results
      - To correct for this, include the Electronic Publication Date: 2015.yr or 2015*.ep. 1,259,586 results
Another way to compare the content of PubMed with the content in Ovid MEDLINE is by applying a date range search. Again, it is important to make sure that date fields are matching between platforms. Below is a table with search examples that allow to make comparisons between the 2 platforms.

**Important Note:** Counts between the platforms will vary from day to day due to timing differences in database updates and deduplication as records get indexed changing status to: publisher – in process – MEDLINE or PubMed-Not-Medline.

More info on range searching can be found here: [https://wkhealth.force.com/ovidsupport/s/article/Limit-by-date-range-in-Ovid](https://wkhealth.force.com/ovidsupport/s/article/Limit-by-date-range-in-Ovid)

<table>
<thead>
<tr>
<th>PubMed Search on 19 November 2019, 2PM CET</th>
<th>#results</th>
<th>Ovid MEDLINE(R) ALL &lt;1946 to November 15, 2019&gt;</th>
<th>#results</th>
<th>PubMed vs Ovid</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&quot;1950&quot;[Date - Publication] : &quot;2015&quot;[Date - Publication])</td>
<td>25,288,744</td>
<td>1950:2015.(yr). or 19500101:20151231.(ep).</td>
<td>25,303,944</td>
<td>15,200</td>
</tr>
<tr>
<td>(&quot;1950&quot;[Date - MeSH] : &quot;2015&quot;[Date - MeSH])</td>
<td>24,624,946</td>
<td>19500101:20151231.(da). [DA - MeSH date]</td>
<td>24,630,222</td>
<td>5,276</td>
</tr>
<tr>
<td>(&quot;1950&quot;[Date - Entrez] : &quot;2015&quot;[Date - Entrez])</td>
<td>25,252,091</td>
<td>19500101:20151231.(ez). [EZ - Entrez date]</td>
<td>25,249,437</td>
<td>-2,654</td>
</tr>
<tr>
<td>(&quot;1950&quot;[Date - Create] : &quot;2015&quot;[Date - Create])</td>
<td>25,332,221</td>
<td>19500101:20151231.(dt). [DT - Create date]</td>
<td>25,344,736</td>
<td>12,515</td>
</tr>
<tr>
<td>(&quot;1950&quot;[Date - Completion] : &quot;2015&quot;[Date - Completion])</td>
<td>24,464,939</td>
<td>19500101:20151231.(ed). [ED - entry date]</td>
<td>24,474,891</td>
<td>9,952</td>
</tr>
</tbody>
</table>
References to understand differences between Ovid MEDLINE and PubMed

- Understanding the input data which is equal for PubMed and Ovid MEDLINE:

- Understanding fields and indexation of fields
  - Ovid Fields: http://ospguides.ovid.com/OSPguides/medline.htm

- Understanding search behaviour:

Search examples executed on 19 November 2019
Ovid recommendations when comparing Ovid MEDLINE to PubMed

- When comparing results in Ovid MEDLINE with PubMed always select **Ovid MEDLINE All <1946 – Present>**

<table>
<thead>
<tr>
<th>Ovid segment</th>
<th>Description</th>
<th>#Ovid docz.dz</th>
<th>PubMed Search</th>
<th>#PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td>medall</td>
<td>PubMed equivalent, all Ovid segments included</td>
<td>30,418,369</td>
<td>all[sb]</td>
<td>30,346,029</td>
</tr>
<tr>
<td>medp</td>
<td>Publisher/Ahead of Print</td>
<td>381,562</td>
<td>publisher[sb] ***</td>
<td>379,913</td>
</tr>
<tr>
<td>prew</td>
<td>In-Process &amp; Non-Indexed</td>
<td>3,688,674</td>
<td>inprocess[sb] OR pubmednotmedline[sb]</td>
<td>3,637,376</td>
</tr>
<tr>
<td>omes</td>
<td>Fully indexed MEDLINE records</td>
<td>26,347,143</td>
<td>medline[sb]</td>
<td>26,324,160</td>
</tr>
</tbody>
</table>

* To switch database this command line syntax in Advanced Search makes it quick: ..c <segmentname>

** For a full description of all segments see the Database guide: [http://ospguides.ovid.com/OSPguides/medline.htm](http://ospguides.ovid.com/OSPguides/medline.htm)

*** publisher[sb] includes "pmcbook" records. On Ovid nb$.bk retrieves all pmcbook records

More info can be found in this Ovid knowledgebase item: [OVID’S MEDLINE COMPARED TO PUBMED](http://ospguides.ovid.com/OSPguides/medline.htm)

- Use **Create Date** (DT) for comparisons when limiting result sets in time, and/or when repeating searches later
  - Create Date never changes, regardless of the record status (Publisher, In-Process, MEDLINE).
  - Create Date is commonly used by researchers to rerun searches from the past and identify new records (e.g. for Systematic Reviews).

- When comparing with PubMed, bear in mind:
  - Timing differences
  - Differences in indexation
  - Differences in search algorithm
  - Ovid does not assume, “what you search is what you get”