**AMSTERDAM EFFICIENT DEDUPLICATION (AED) METHOD**

**MANUAL**

|  |  |
| --- | --- |
| René Otten, Linda Schoonmade, Ralph de Vries  Universiteitsbibliotheek, department Research Support  Vrije Universiteit Amsterdam |  |

Contents

[INTRODUCTION 3](#_Toc34729823)

[1. PubMed 3](#_Toc34729824)

[1.1 Collecting IDs from PubMed (during search) 3](#_Toc34729825)

[1.2 Collecting IDs from EndNote (afterwards) 4](#_Toc34729826)

[1.3 Removing old PubMed records in future update 4](#_Toc34729827)

[2. Embase.com 4](#_Toc34729828)

[2.1 Collecting IDs from Embase.com (during search) 4](#_Toc34729829)

[2.2 Collecting IDs from EndNote (afterwards) 5](#_Toc34729830)

[2.3 Removing PubMed records 5](#_Toc34729831)

[2.4 Removing old Embase records in future update 5](#_Toc34729832)

[3. CINAHL (EBSCO) 6](#_Toc34729833)

[3.1 Collecting IDs from CINAHL (during search) 6](#_Toc34729834)

[3.2 Collecting IDs from EndNote (afterwards) 6](#_Toc34729835)

[3.3 Removing PubMed records 6](#_Toc34729836)

[3.4 Removing old CINAHL records in future update 7](#_Toc34729837)

[3.5 Other EBSCO databases 7](#_Toc34729838)

[4. APA PsycInfo (EBSCO) 7](#_Toc34729839)

[4.1 Collecting IDs from PsycInfo (during search) 7](#_Toc34729840)

[4.2 Collecting IDs from EndNote (afterwards) 8](#_Toc34729841)

[4.3 Removing PubMed records 8](#_Toc34729842)

[4.4 Removing old PsycInfo records in future update 8](#_Toc34729843)

[4.5 Other EBSCO databases 8](#_Toc34729844)

[5. Cochrane Library (Wiley) 9](#_Toc34729845)

[5.1 Collecting IDs from Cochrane (during search) 9](#_Toc34729846)

[5.2 Collecting IDs from EndNote (afterwards) 9](#_Toc34729847)

[5.3 Removing PubMed / Embase records 9](#_Toc34729848)

[5.4 Removing old Cochrane records in future update 10](#_Toc34729849)

[6. Scopus 10](#_Toc34729850)

[6.1 Collecting IDs from Scopus (during search) 10](#_Toc34729851)

[6.2 Collecting IDs from EndNote (afterwards) 11](#_Toc34729852)

[6.3 Removing PubMed records 11](#_Toc34729853)

[6.4 Removing old Scopus records in future update 11](#_Toc34729854)

[7. Web of Science 12](#_Toc34729855)

[7.1 Collecting IDs from Web of Science (during search) 12](#_Toc34729856)

[7.2 Collecting IDs from EndNote (afterwards) 12](#_Toc34729857)

[7.3 Removing PubMed records 12](#_Toc34729858)

[7.4 Removing old Web of Science records in future update 13](#_Toc34729859)

[DOI of all versions 13](#_Toc34729860)

[Cite this document as 13](#_Toc34729861)

[Appendix: handout 13](#_Toc34729862)

# INTRODUCTION

When searching extensively for literature, especially for systematic reviews, it is common and usually required to search multiple databases. Depending on the specific topic, this can vary from 2 to more than 10 databases. The results will therefore contain many duplicates. Deduplication is necessary to prevent the same article from being screened, read and cited twice or more. But deduplication can be a time-consuming process and is prone to errors.  
In this document we describe our method of deduplication:

*AMSTERDAM EFFICIENT DEDUPLICATION (AED) METHOD*

This method is 100% reliable.  
It can also be used with updates from older searches. The latter requires a small investment during the initial search, but delivers a substantial time saving when updating.

In this method it is explained per database / host which steps should be taken.

1. Collecting the accession numbers (if possible) during the search. Purpose: save for a later update and in the case of PubMed for duplicating between PubMed and other databases
2. Collecting the accession numbers from EndNote because the old ones were not collected at the initial search or because they cannot easily be retrieved from the database
3. Deduplicate PubMed records in another database
4. Deduplicate old records in a database upon update

# 1. PubMed

## 1.1 Collecting IDs from PubMed (during search)

* Download the PMIDs list in PubMed > Save > Format: **PMID**
* Edit this list in for example Notepad ++ or Word:

1. For deduplicating in a PubMed future update

* Convert line break to <space>[[1]](#endnote-1)
* For example, save as: *PMIDs for PubMed.txt*

1. For deduplication in other databases

* Select all [Ctrl-A] > Search/Replace [Ctrl-H]: replace <space> with <space>OR<space>  
  Remove (if necessary) the last OR
* For example, save as: PMIDs for other databases.txt

## 1.2 Collecting IDs from EndNote (afterwards)

This assumes that you have imported all old records from PubMed into EndNote.

* Use the Output Style "[Accession Numbers](https://zenodo.org/record/3582928/files/Accession%20Numbers.ens)"
* Select all titles [Ctrl-A] > right mouse click > Copy Formatted
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Remove (if necessary) the last OR
* You can use this version for deduplicating in PubMed future update

## 1.3 Removing old PubMed records in future update

* After performing your PubMed search
* Cut/paste the PMID **1.1**-list(without OR) into PubMed and perform the search
* Subtract from the new search
* #newsearch NOT #oldsearch

# 2. Embase.com

## 2.1 Collecting IDs from Embase.com (during search)

* Export the Embase identification numbers (PUI) via format: "CSV Fields by Column" >  
  "Specify fields to be exported" > Select: **PUI**
* Click Open > the file will be opened in Excel
* Remove the first row (containing "PUI")
* Select all [Ctrl-A] and copy [Ctrl-C]
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Select all [Ctrl-A] > Search/Replace [Ctrl-H]: replace <space> with **<space>OR<space>**
* Remove (if necessary) the last OR
* You can use this version for deduplicating in Embase future update

## 2.2 Collecting IDs from EndNote (afterwards)

This assumes that you have imported all old records from Embase into EndNote.

* Copy the field “Notes” with Embase identification number (PUI) to the field “Accession Number” (> Tools > Change/Move/Copy Fields> Move Fields)
* Use the Output Style “[Accession Numbers](https://zenodo.org/record/3582928/files/Accession%20Numbers.ens)”
* Select all titles [Ctrl-A] > right mouse click > Copy Formatted
* Separate or sort this output so that you have the Embase identification numbers (PUIs; starting with L) separately in one block / column
* Select this block/column
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>OR<space>1
* Remove (if necessary) the last OR
* You can use this version for deduplicating in Embase future update

## 2.3 Removing PubMed records

* After performing your Embase Search
* Cut/paste PMID **1.**2-list (with OR) in Embase between () with output **:ui**and perform the search
* Subtract from the Embase search
* Or directly:

#x NOT (30198739 OR 30183539 OR…):ui

## 2.4 Removing old Embase records in future update

* After performing your Embase Search
* Cut/paste Embase **2.1 or 2.2-** list (with OR) in Embase between () with output **:id**and perform the search
* Subtract from the new search
* Or directly:

#x NOT (L30198739 OR L30183539 OR…):id

# 3. CINAHL (EBSCO)

## 3.1 Collecting IDs from CINAHL (during search)

IRRELEVANT! Too cumbersome because of the maximum of 50 items each time for the function "Add to folder"

## 3.2 Collecting IDs from EndNote (afterwards)

This assumes that you have imported all old records from CINAHL into EndNote.

* Use the Output Style “[Accession Numbers](https://zenodo.org/record/3582928/files/Accession%20Numbers.ens)”
* Select all titles [Ctrl-A] > right mouse click > Copy Formatted
* Use for example Excel to separate or sort this output so that you have the CINAHL Accession Numbers separately in one block in one column
* Select the column with the accession numbers
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Remove (if necessary) the last OR
* You can use this version for deduplicating in CINAHL future update

## Removing PubMed records

* After performing your CINAHL Search
* Go to Notepad ++ or Word and paste the PMID 1.2 list (with OR) (as plain text [Ctrl-V]
* Convert line break to <space>1
* Select all [Ctrl-A]> Search / Replace [Ctrl-H]: replace <space> OR <space> with <space> OR <space> NLM. Add NLM at the beginning
* Cut/paste the edited list in CINAHL between () with precursor **PM**and perform the search
* Subtract the set with PMID list from the CINAHL search
* Or directly:   
  Sx NOT PM (NLM30198739 OR NLM30183539 OR…)

## 3.4 Removing old CINAHL records in future update

* Cut/paste CINAHL **3.2**-list ( with OR ) in CINAHL between () with precursor **AN**and perform the search
* Subtract from the new search
* Or directly:   
  Sx NOT AN (30198739 OR 30183539 OR…)

## 3.5 Other EBSCO databases

NB: For other EBSCO databases (i.e. not CINAHL/PsycInfo) , such as ERIC and SPORTDiscus, the above mentioned steps 2 and 4 can be used. The deduplication with PubMed is not possible as ERIC and SPORTDiscus contain no PMID’s.

# 4. APA PsycInfo (EBSCO)

## 4.1 Collecting IDs from PsycInfo (during search)

IRRELEVANT! Too cumbersome because of the maximum of 50 items each time for the function "Add to folder"

## 4.2 Collecting IDs from EndNote (afterwards)

This assumes that you have imported all old records from PsycInfo into EndNote.

* Use the Output Style “[Accession Numbers](https://zenodo.org/record/3582928/files/Accession%20Numbers.ens)”
* Select all titles [Ctrl-A] > right mouse click > Copy Formatted
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Remove (if necessary) the last OR
* You can use this version for deduplicating in PsycInfo future update

## 4.3 Removing PubMed records

* After performing your PsycInfo Search
* Cut/paste PMID **1.2**-list with OR in PsycInfo between () with precursor **PM**and perform the search
* Subtract the set with PMID list from the PsycInfo search
* Or directly:  
  Sx NOT PM (30198739 OR 30183539 OR…)

## 4.4 Removing old PsycInfo records in future update

* After performing your PsycInfo Search
* Cut/paste PsycInfo **4.2**-list with OR in PsycInfo between () with precursor **AN**and perform the search
* Subtract from the new search
* Or directly:   
  Sx NOT AN (30198739 OR 30183539 OR…)

## 4.5 Other EBSCO databases

NB: For other EBSCO databases (i.e. not CINAHL/PsycInfo) , such as ERIC and SPORTDiscus, the above mentioned steps 2 and 4 can be used. The deduplication with PubMed is not possible as ERIC and SPORTDiscus contain no PMID’s.

# 5. Cochrane Library (Wiley)

## 5.1 Collecting IDs from Cochrane (during search)

IT IS NOT POSSIBLE TO DOWNLOAD ACCESSION NUMBERS ONLY !

## 5.2 Collecting IDs from EndNote (afterwards)

This assumes that you have imported all old records from Cochrane into EndNote.

* Use the Output Style “[Accession Numbers](https://zenodo.org/record/3582928/files/Accession%20Numbers.ens)”
* Select all titles [Ctrl-A] > right mouse click > Copy Formatted
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Remove (if necessary) the last OR
* You can use this version for deduplicating in Cochrane future update

## 5.3 Removing PubMed / Embase records

THIS DOES NOT WORK WELL WITH LARGE SETS OF NUMBERS FROM PUBMED / EMBASE

**PMID list procedure:**

* Edit the PMID list in, for example, Notepad ++ or Word
* In the PubMed **1.2**-list, place **OR<space>** before the first number and **<space>OR** after the last number
* Select all [Ctrl-A], > Ctrl-H: Search: **OR<space>** Replace with: **OR (“PubMed<space>**
* Search: **<space>OR** and replace with: **"):an<space>OR**
* Remove the first and the last OR in the block
* Perform the search
* Subtract from the Cochrane search
* Or directly:   
  #x NOT (("PubMed 30403654"):an OR ("PubMed 30322793"):an OR …)

**Embase list procedure:**

* Edit the Embase **2.1 or 2.2-**list in, for example, Notepad ++ or Word
* In the Embase list, place **OR<space**> before the first number and **<space>OR** after the last number
* Select all [Ctrl-A] > Search/Replace [Ctrl-H] OR<space> by **OR ("Embase<space>**
* Replace <space>OR with **"):an<space>OR**
* Remove the first and the last OR in the block
* Perform the search
* Subtract from your Cochrane Search
* Or directly:   
  #x NOT (("Embase 624095871"):an OR ("Embase 615604185"):an OR …)

## 5.4 Removing old Cochrane records in future update

* After performing your Cochrane search
* Cut/paste Cochrane **5.2**-list (with OR) in Cochrane between () and perform the search
* Subtract from the new search
* Or directly:   
  #x NOT (CN-01577879 OR …)

# 6. Scopus

## 6.1 Collecting IDs from Scopus (during search)

* Select: All > Export | method of export: CSV > Select: **EID** (leave other boxes empty!)
* Open the CSV file in Excel and delete the top line
* Separate this output so that you have the Scopus EIDs (in the format *2-s2.0-77954009273*) separately in one column
* Select this column
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Select all [Ctrl-A] > Search/Replace [Ctrl-H]: replace <space> with **<space>OR<space>**
* You can use this version for deduplicating in Scopus future update

NB. With more than 2000 items split into parts of max. 2000 items!

## 6.2 Collecting IDs from EndNote (afterwards)

This assumes that you have imported all old records from Scopus into EndNote.

* Copy the field “URL” to the field “Accession Number” (go to: Tools > Change/Move/Copy Fields> Move Fields)
* Use the Output Style “[Accession Numbers](https://zenodo.org/record/3582928/files/Accession%20Numbers.ens)”
* Select all titles > right mouse click > Copy Formatted
* Paste the file in Excel to separate this output so that you have the Scopus EIDs (in the format 2-s2.0-77954009273) separately in one column
* Select this column
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Select all [Ctrl-A] > Search/Replace [Ctrl-H]: replace <space> with: **<space>OR<space>**
* You can use this version for deduplicating in Scopus future update

## 6.3 Removing PubMed records

* After performing your Scopus Search
* Cut/paste PMID **1.2**-list with OR in Scopus between () with precursor **PMID**and perform the search
* Subtract from the Scopus search  
  NB. Only works with combination of set numbers (#1 AND NOT #2)!
* PMID(30198739 OR 30183539 OR …)

## 6.4 Removing old Scopus records in future update

* After performing your Scopus Search
* Cut/paste Scopus **6.1 or 6.**2-list (with OR) in Scopus between () with precursor **EID**and perform the search
* Subtract from the new search  
  NB. Only works with combination of set numbers (#1 AND NOT #2)!
* EID( 2-s2.0-85064602041 OR 2-s2.0-85064120601 OR … )

# 7. Web of Science

## 7.1 Collecting IDs from Web of Science (during search)

* Export to Other File Formats > Full Record > Tab Delimited (Win)
* Open the file in Excel
* Select the column UT with the accession numbers
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Select all [Ctrl-A] > Search/Replace [Ctrl-H]: replace <space> with **<space>OR<space>**
* You can use this version for deduplication in Web of Science future update

## 7.2 Collecting IDs from EndNote (afterwards)

This assumes that you have imported all old records from Web of Science into EndNote.

* Use the Output Style “[Accession Numbers](https://zenodo.org/record/3582928/files/Accession%20Numbers.ens)”
* Select all titles > right mouse click > Copy Formatted
* Go to Notepad ++ or Word and paste as plain text [Ctrl-V]
* Convert line break to <space>1
* Select all [Ctrl-A] > Search/Replace [Ctrl-H]: replace <space> with **<space>OR<space>**
* You can use this version for deduplication in Web of Science future update

## 7.3 Removing PubMed records

* After performing your Web of Science Search
* Cut/paste PMID **1.2**-list with OR in Web of Science between () with precursor **PMID=**and perform the search
* Subtract from the WoS search
* Or directly:  
  #x NOT PMID=(30198739 OR 30183539 OR…)

## 7.4 Removing old Web of Science records in future update

* After performing your Web of Science Search
* Cut/paste Web of Science **7.1 or 7.2**-list with OR in Web of Science between () with precursor **UT=**and perform the search
* Subtract from the new search
* Or directly:

#x NOT UT=(WOS:000436466400018 OR WOS:00043646654321 OR …)

# DOI of all versions

10.5281/zenodo.3582927

# Cite this document as

René Otten, Ralph de Vries, & Linda Schoonmade. (2019, December 12). Amsterdam Efficient Deduplication (AED) method - manual. Zenodo.org . http://doi.org/10.5281/zenodo.3582927

# Appendix: handout

A handout can be found at

<http://doi.org/10.5281/zenodo.3582927>

March 2020

1. Convert line break to <space>

   In Notepad++: Go to Edit > Blank operations > EOL to Space  
   In Word: ctrl-H: Find: ^p Replace with: <space> [↑](#endnote-ref-1)