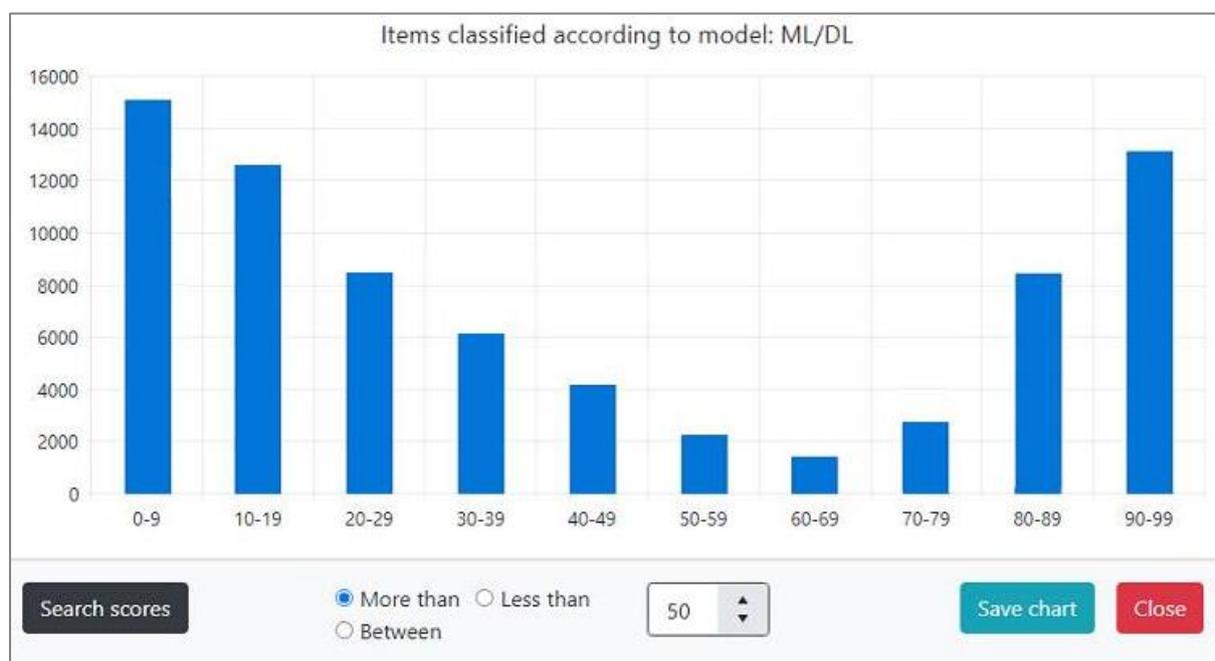


Classifiers in EPPI Reviewer

Background

Classifiers organise references, sorting them by the probability that they are relevant or not, according to existing data (i.e. an existing set of items already screened and thus coded with include or exclude codes).

Each item will be given a “probability of being relevant” score and can be presented in order - sorted on this score or banded into decile bands of probability (0-9% likely to be relevant, 1-10% likely to be relevant, through to 90-99% likely to be relevant).



If you have a dataset (references and coding, or references in batches according to your inclusion / relevant and exclusion / irrelevant criteria), you can create classifiers based on this information.

(This can be screening or coding done on items already in an existing review, whereby data from EPPI Reviewer’s screening system can thus be taken to build a classifier. You can amalgamate multiple exclusion or inclusion criteria to create two codes summarising whether an item is relevant or not.

Alternatively it can be done by giving codes to subsets of your dataset, according to whether the items are relevant or irrelevant, thus you can also work with data imported from outside EPPI Reviewer i.e. without screening or coding the items within ER.)

You can then apply the classifier to new items, automatically classifying them according to knowledge the system has learnt from your existing screening / coding.

There are also several pre-built classifiers available in EPPI Reviewer, designed to find, for example, Cochrane RCTs, original RCTs, or economic evaluations.

So, the models can classify items according to the probability they are an original Random Controlled Trial, a Systematic Review, etc.

The models are built from thousands of records and should be very accurate when used on items in the same subject area i.e. biomedical records of human studies, such as RCT records found on PubMed.

(Note that we are continually expanding and improving these in-built classifiers; for example, we have recently included specific models relating to Covid and Long Covid. We have incorporated current methods such as BERT models (Bidirectional Encoder Representations from Transformers), utilising a machine-learning technique developed for natural language processing by Google.)

Review home References Reports Search & Classify Collaborate

New Search Refresh List Delete Selected Combine Build Model **Classify**

Step 1: Select a Model

Long COVID binary model
COVID-19 map categories
Cochrane RCT
Economic Evaluation
Systematic Review
Original RCT
Show Custom Models per the current review
Show Custom Models for all reviews

Step 2: Choose method to apply

- Apply to all items in review
- Apply to items with this code
- Apply to items from this source

Run Model

Using an in-built classifier

Under the **Search & Classify** tab, click the **Classify** button (1).

Select the in-built model you wish to use (2), then which items you wish to apply the classifier to (3).

Finally click the **Run Model** button (4).

The screenshot shows the 'Search & Classify' interface. At the top, there are navigation tabs: 'Review home', 'References', 'Reports', 'Search & Classify' (highlighted), and 'Collaborate'. Below the tabs is a toolbar with buttons: 'New Search', 'Refresh List', 'Delete Selected', 'Combine', 'Build Model', and 'Classify' (circled in red with a '1' next to it). The main area is divided into two steps:

Step 1: Select a Model

- Long COVID binary model
- COVID-19 map categories
- Cochrane RCT** (circled in red with a '2' next to it)
- Economic Evaluation
- Systematic Review
- Original RCT
- Show Custom Models per the current review
- Show Custom Models for all reviews

Step 2: Choose method to apply

3 Apply to all items in review

Apply to items with this code

Apply to items from this source

At the bottom left, there is a green button labeled 'Run Model' (circled in red with a '4' next to it).

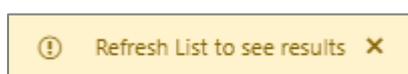
You will be asked to confirm the action, as it may take several minutes to process.

Please confirm

Are you sure you wish to run the selected model ?

Cancel
OK

You will then see a status message – “Refresh List to see results”.



Note that you can also apply the model to all items with a particular code, by selecting the relevant option, followed by the relevant codeset (in the code tree on the right).

The screenshot shows the 'Search & Classify' interface. At the top, there are tabs for 'Review home', 'References', 'Reports', 'Search & Classify', and 'Collaborate'. Below the tabs are buttons for 'New Search', 'Refresh List', 'Delete Selected', 'Combine', 'Build Model', and 'Classify'. The main area is divided into two steps:

- Step 1: Select a Model**: A list of models is shown, with 'Cochrane RCT' selected and highlighted in teal. Other models include 'Long COVID binary model', 'COVID-19 map categories', 'Economic Evaluation', 'Systematic Review', and 'Original RCT'. There are also options to 'Show Custom Models per the current review' and 'Show Custom Models for all reviews'. A green 'Run Model' button is at the bottom.
- Step 2: Choose method to apply**: Three radio button options are shown:
 - Apply to all items in review
 - Apply to items with this code **3a**
 - Apply to items from this source
 Below these options, a red oval highlights the text 'Current code: All items for HSRs'.

On the right side, there is a 'Code Tree' with a dropdown menu 'Assign search(es) to code'. The tree shows a hierarchy: 'SIGN Information Scientist Sift' > 'SIGN Methodology Checklist 1: Systematic Reviews and Meta-analyses' > 'Allocations' > 'Group 1' > 'Group 2' > 'All items for HSRs' **3b**. The 'All items for HSRs' item is highlighted in red.

Alternatively, you can apply the model to all items from a particular source (just select the relevant source from the drop-down list).

The screenshot shows the 'Search & Classify' interface. At the top, there are tabs for 'Review home', 'References', 'Reports', 'Search & Classify', and 'Collaborate'. Below the tabs are buttons for 'New Search', 'Refresh List', 'Delete Selected', 'Combine', 'Build Model', and 'Classify'. The main area is divided into two steps:

- Step 1: Select a Model**: A list of models is shown, with 'Cochrane RCT' selected and highlighted in teal. Other models include 'Long COVID binary model', 'COVID-19 map categories', 'Economic Evaluation', 'Systematic Review', and 'Original RCT'. There are also options to 'Show Custom Models per the current review' and 'Show Custom Models for all reviews'. A green 'Run Model' button is at the bottom.
- Step 2: Choose method to apply**: Three radio button options are shown:
 - Apply to all items in review
 - Apply to items with this code
 - Apply to items from this source **3a**
 Below these options, a dropdown menu is open, showing a list of sources:
 - EmBase RCTs December 2021.txt
 - Cochrane trials december 2021.txt
 - EmBase RCTs December 2021.txt** **3b**
 - Medline RCTs December 2021.txt
 - Manually Created Source
 The 'EmBase RCTs December 2021.txt' item is highlighted in a red oval.

Once the system has processed your reference (applied the model to your items and classified them accordingly), the results will be shown under your **Search & Classify** tab, along with your searches.

As mentioned, this can take several minutes. You may need to click the **Refresh List** button to see the classifier appear in your list.

Classifiers are indicated by the **True** link in the right-most "**Classifier**" column. To show a visualisation of the results, click the **true** link to the right of the name (in the **Classifier** column).

Review home References Reports Search & Classify Collaborate						
New Search Refresh List Delete Selected Combine ▼ Build Model Classify						
	No...	Name	Created By	Date	Hits	Classifier
<input type="checkbox"/>	14	<u>Items classified according to model: RCT</u>	Zak Ghouze	4 May 2022	41552	true
<input type="checkbox"/>	13	With at least one document uploaded.	Zak Ghouze	31 Aug 2021	1	false
<input type="checkbox"/>	12	"role" (in Abstract)	Zak Ghouze	12 Apr 2021	4275	false

The results are shown in probability bands (10 bands according to how likely an item is likely to be relevant or irrelevant, from 0-9% to 90-99%).

Note that you can search for particular score ranges or save the visualisation as a graphic (for reuse elsewhere).



You can display the probability score on the **References** tab. To show a list of the items in the classifier, click on the **Hits** number adjacent to the classifier Name.

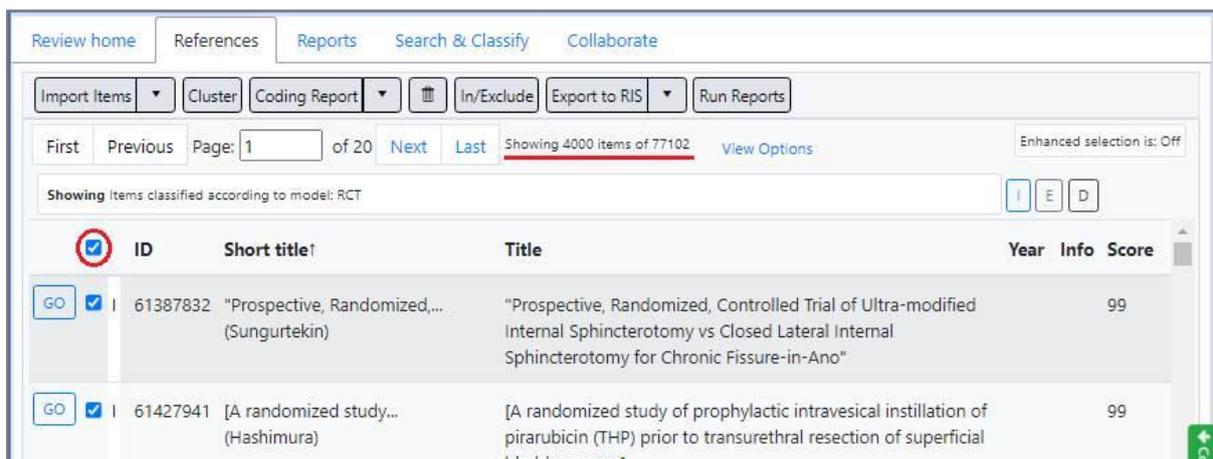
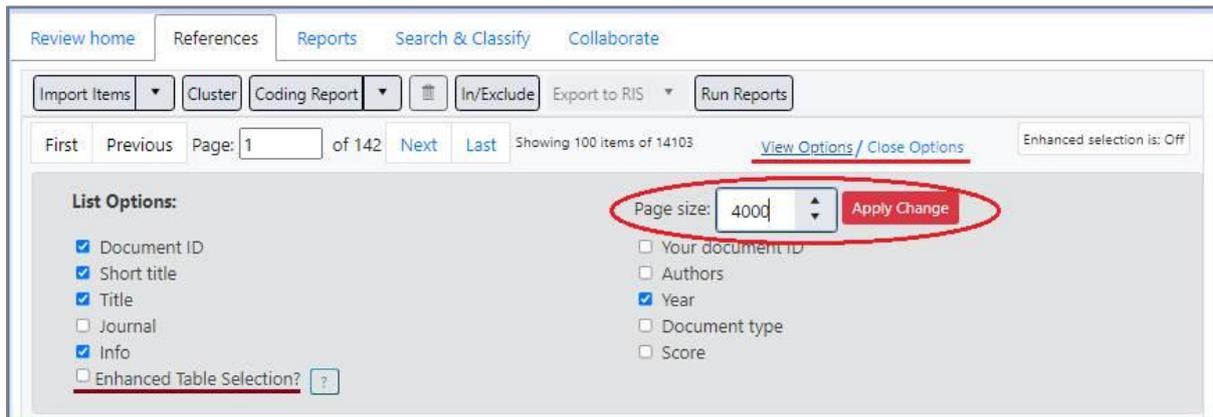
No...	Name	Created By	Date	Hits	Classifier
<input type="checkbox"/> 381	Search #377 scores between 40 and 49	Mark Engelbert		32377	false
<input type="checkbox"/> 380	Search #377 scores between 50 and 59	Mark Engelbert		18616	false
<input type="checkbox"/> 379	Search #377 scores between 60 and 69	Mark Engelbert		4405	false
<input type="checkbox"/> 378	Search #377 scores between 70 and 79	Mark Engelbert		170	false
<input type="checkbox"/> 377	Items classified according to model: 2017 T-A data (pub year 2017 only)	Mark Engelbert		66309	true
<input type="checkbox"/> 376	Not coded with: All screened (previous projects) 2021-03-31	Mark Engelbert		66317	false
<input type="checkbox"/> 375	Not coded with: All screened (previous projects) 2021-03-31	Zafeer Ravat		66356	false
<input type="checkbox"/> 374	Not coded with: All screened (previous projects) 2021-03-31	Zafeer Ravat		66361	false
<input type="checkbox"/> 372	Not coded with: All screened (previous projects) 2021-03-31 (TO RE-RUN)	Zafeer Ravat		66398	false

Alter the **View Options** to show **Score**, then you can sort by score (from lowest to highest or vice versa with each click).

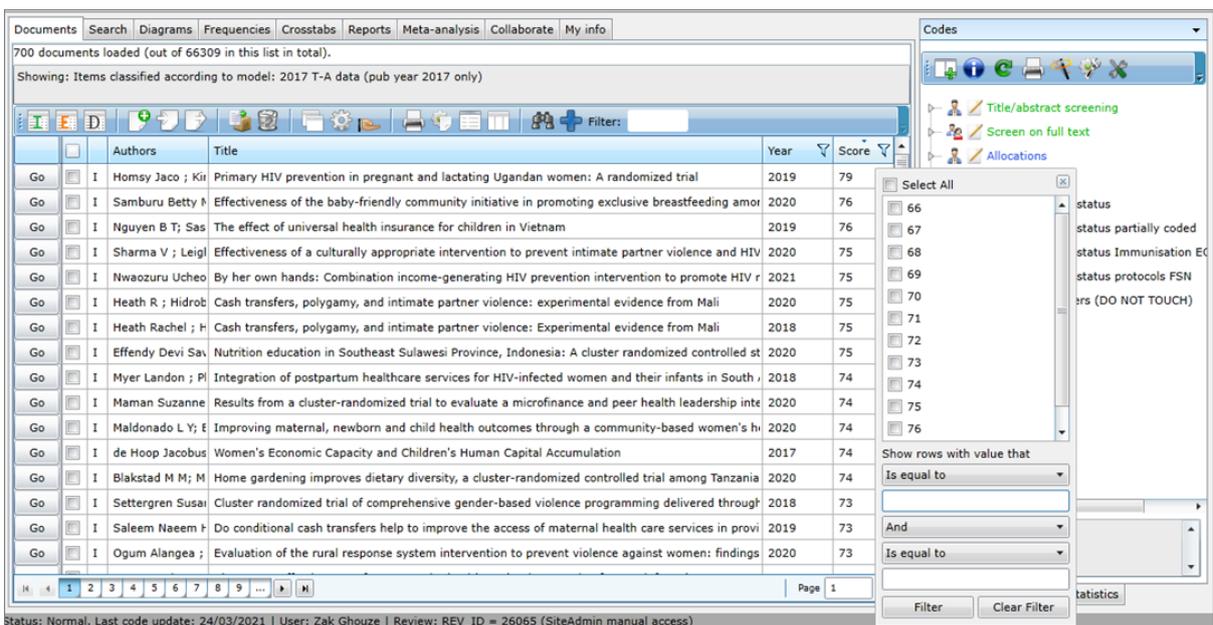
ID	Short title	Title	Year	Score
<input type="checkbox"/> I 56191257	Amstutz (2020)	Home-based oral self-testing for absent and declining individuals during a door-to-door HIV testing campaign in rural Lesotho (HOSENG): a cluster-randomised trial	2020	71
<input type="checkbox"/> I 56137425	Antwi (2020)	Primary school-based nutrition education intervention on nutrition knowledge, attitude and practices among school-age children in Ghana	2020	72

ID	Short title	Title	Year	Score
<input type="checkbox"/> I 56148203	Jain (2018)	Three Essays on Health and Aging	2018	60
<input type="checkbox"/> I 56193340	Jahan (2020)	Awareness Development and Usage of Mobile Health Technology Among Individuals With Hypertension in a Rural Community of Bangladesh: Randomized Controlled Trial	2020	60

(You can alter the number of items shown per page via the **View Options** link. A list of items can be selected in one operation via the checkbox at the top of the list.)

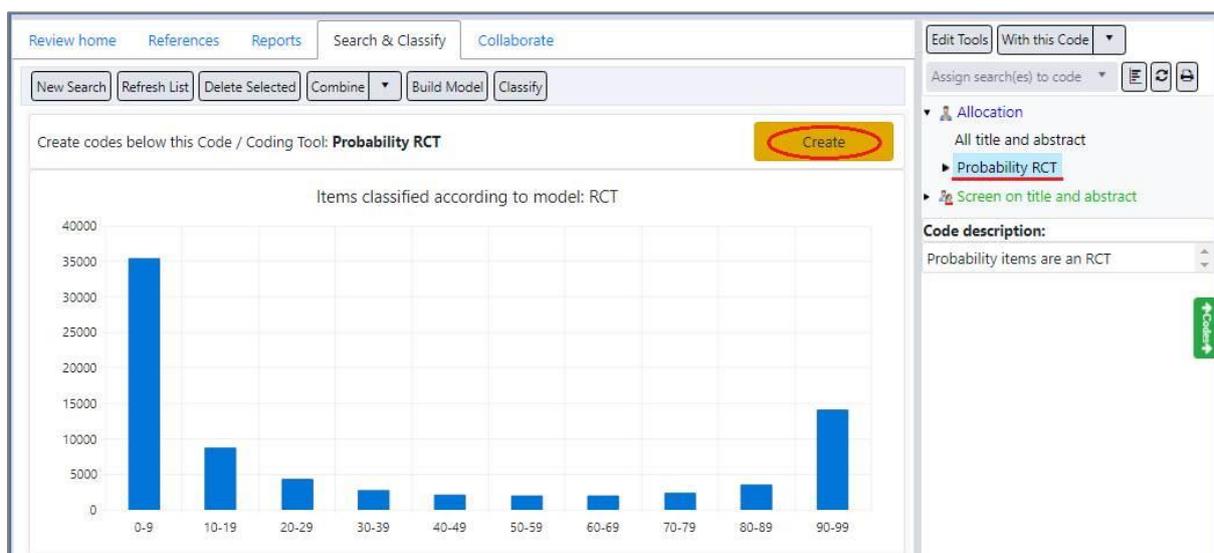


(You can also “filter” on score in ER 4, if you enable the column in your view options, so can get a specific subset of items by score from your classifier.)

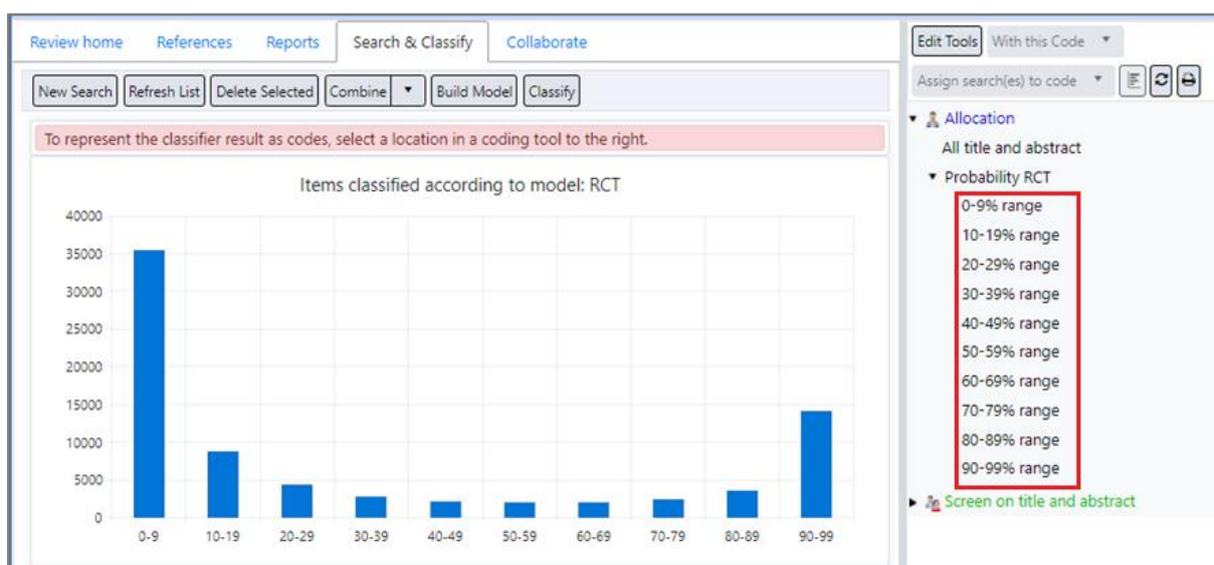


You can use the decile bands EPPI Reviewer produces to automatically create 10 codes within a codeset which are assigned to each item according to their band of probable relevance, useful for creating allocation codes or focusing on specific bands. (For example, you may wish to exclude all those items in the lower few bands en masse i.e. all items unlikely to be relevant / likely to be irrelevant, or you may wish to manually screen the items around the middle bands, where the system isn't sure if they will be included or excluded, or you may wish to work with only those items the system judges likely to be relevant i.e. those in the uppermost bands.)

If you select a code in the code tree on the right, you can then get the system to automatically make 10 child codes for the 10 bands by clicking the **Create** button. (You can easily create a code specifically for this purpose.)



The child codes will be applied to your items according to which band they are in.



So, for example, you can list and operate on the items in the 90-99% band by simply listing the items with the relevant code.

Showing 90-99% range.

ID	Short title†	Title	Year	Info
61387832	"Prospective, Randomized, ... (Sungurtekin)	"Prospective, Randomized, Controlled Trial of Ultra-modified Internal Sphincterotomy vs Closed Lateral Internal Sphincterotomy for Chronic Fissure-in-Ano"		
61427941	[A randomized study... (Hashimura)]	[A randomized study of prophylactic intravesical instillation of pirarubicin (THP) prior to transurethral resection of superficial bladder cancer]		
61350240	[Application of Alpha1-adrenergic... (Wang)]	[Application of Alpha1-adrenergic antagonist with extracorporeal shock wave lithotripsy for lower ureteral stone]		
61383198	[Approach to percutaneous... (Sedano-Portillo)]	[Approach to percutaneous nephrolithotomy. Comparison of the procedure in a one-shot versus the sequential with metal dilata]		

Allocation

- All title and abstract
- Probability RCT
 - 0-9% range
 - 10-19% range
 - 20-29% range
 - 30-39% range
 - 40-49% range
 - 50-59% range
 - 60-69% range
 - 70-79% range
 - 80-89% range
 - 90-99% range
- Screen on title and abstract

Code description:
FROM: Items classified according to model: RCT

(You can see the items' coding via a **Frequency Chart**. Clicking on the numbers brings up a list of those items.)

Frequencies and crosstabs

Rows: Probability RCT

Columns: Not set (only used for Crosstabs)

Filter: Not set (optional)

Get Frequencies

Get CrossTab

Current code: Probability RCT

Show results as: Table

Code	Count
0-9% range	35319
10-19% range	8769
20-29% range	4268
30-39% range	2704
40-49% range	2098
50-59% range	1966

Allocation

- All title and abstract
- Probability RCT
- Screen on title and abstract

Code description:
Probability items are an RCT

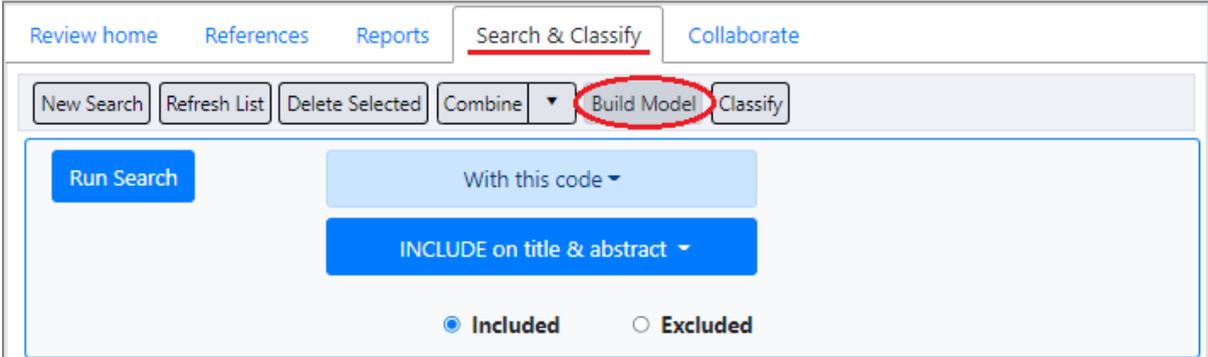
Creating Custom Classifiers

Further to the built-in classifiers described above, you can also create your own classifiers based on existing data.

You may have a batch of coded or screened items in the review that you want to use as the basis for a model. You may have sets of references elsewhere that you wish to import, coding each batch with an appropriate code for the model to work with. (You could also, for example, take references from another review, or combine your existing priority screening coding into two codes (*relevant* for items with an include code and *irrelevant* for items with an exclude code) and use that as the basis for a model.)

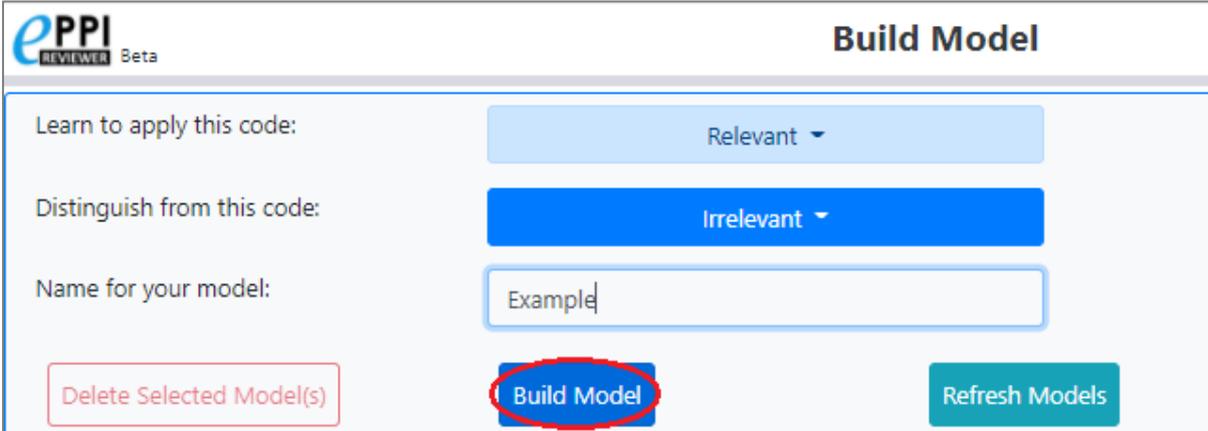
Your model will learn to classify items according to whether they are relevant or irrelevant, based on the dataset (refs and coding) you base the model on.

To create a new model, click the **Build Model** button (still under the **Search & Classify** tab).



The screenshot shows the 'Search & Classify' tab selected in the top navigation bar. Below the navigation bar, there is a row of buttons: 'New Search', 'Refresh List', 'Delete Selected', 'Combine', 'Build Model', and 'Classify'. The 'Build Model' button is circled in red. Below this row, there is a 'Run Search' button, a 'With this code' dropdown menu, and an 'INCLUDE on title & abstract' button. At the bottom, there are radio buttons for 'Included' (selected) and 'Excluded'.

Next, select the 2 codes which define the items which are relevant and those that are irrelevant. Give the model a name and click the next **Build Model** button.



The screenshot shows the 'Build Model' interface. At the top left is the 'ePPI REVIEWER Beta' logo. The title 'Build Model' is at the top right. Below the title, there are three rows of input fields: 'Learn to apply this code:' with a 'Relevant' dropdown, 'Distinguish from this code:' with an 'Irrelevant' dropdown, and 'Name for your model:' with a text input field containing 'Example'. At the bottom, there are three buttons: 'Delete Selected Model(s)', 'Build Model' (circled in red), and 'Refresh Models'.

As the process takes some time, you will see a confirmation that your request was submitted.



The system will work through the data and build its model. Whilst it is processing, you can use other functions within the software.

(If you wish to check if the model has been built, you can click the **Refresh Models** button. Whilst it is still building, the entry in the list of models will show “in progress” and the associated figures will all show as zero.)

Learn to apply this code: Close

Relevant ▾

Distinguish from this code: Irrelevant ▾

Name for your model:

Delete Selected Model(s)
Build Model
Refresh Models

<input type="checkbox"/>	ReviewId	Title	Att On	Att Not On	Accuracy	AUC	Precision	Recall
<input type="checkbox"/>	25830	Example (in progress...)	Relevant	Irrelevant	0	0	0	0

When the processing has completed, you will see the model listed with associated figures. (These figures are detailed later in this document.)

<input type="checkbox"/>	ReviewId	Title	Att On	Att Not On	Accuracy	AUC	Precision	Recall
<input type="checkbox"/>	25830	Example	Relevant	Irrelevant	0.907	0.94	0.306	0.807

Applying Custom Classifiers

To apply your newly built classifier, click the **Classify** button in the normal way.

[Review home](#)
[References](#)
[Reports](#)
Search & Classify
[Collaborate](#)

New Search
Refresh List
Delete Selected
Combine ▾
Build Model
Classify

Click the option to **Show Custom Models per the current review (1)** option, select the relevant model from the list below **(2)**, then select which items are to be classified **(3)**.

Review home
References
Reports
Search & Classify
Collaborate

New Search
Refresh List
Delete Selected
Combine ▼
Build Model
Classify

Step 1: Select a Model

Long COVID binary model

COVID-19 map categories

Cochrane RCT

Economic Evaluation

Systematic Review

Original RCT

1 Show Custom Models per the current review

Show Custom Models for all reviews

Step 2: Choose method to apply

3 Apply to all items in review

Apply to items with this code

Apply to items from this source

	ReviewID	ModelID	Title	Applies	Compar...	Precision	Recall	Rebuild
<input checked="" type="checkbox"/>	25830	2268	Example	Relevant	Irrelevant	0.306	0.807	Rebuild
<input type="checkbox"/>	25830	1202	Most used Exclude	INCLUDE on title & abstract	Most used excludes	0.256	0.636	Rebuild

You can also create models in other reviews and apply them in your current review. (You may have other data or existing models you wish to apply to items in a different project.)

Step 1: Select a Model

Long COVID binary model

COVID-19 map categories

Cochrane RCT

Economic Evaluation

Systematic Review

Original RCT

Show Custom Models per the current review

Show Custom Models for all reviews

(Equally, you can apply the model (or classify) items with a particular code, or items from a particular source – as with the built-in models.)

Step 2: Choose method to apply

- Apply to all items in review
- Apply to items with this code
- Apply to items from this source

Please select a code from the right.

When you have the parameters as you wish, click the **Run Model** button.

Review home
References
Reports
Search & Classify
Collaborate

New Search
Refresh List
Delete Selected
Combine
Build Model
Classify

Step 1: Select a Model

Long COVID binary model

COVID-19 map categories

Cochrane RCT

Economic Evaluation

Systematic Review

Original RCT

Show Custom Models per the current review

Show Custom Models for all reviews

Step 2: Choose method to apply

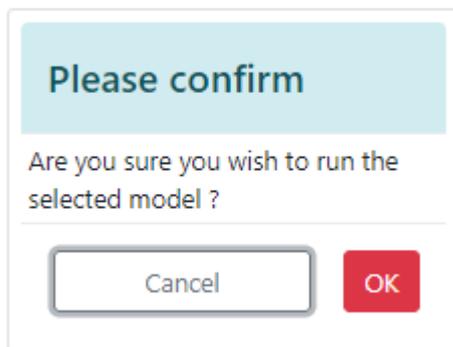
- Apply to all items in review
- Apply to items with this code
- Apply to items from this source

Current code: Unscreened 17-5

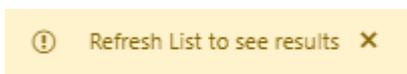
	ReviewID	ModelID	Title	Applies	Compared wi...	Precision	Recall	Rebuild
<input checked="" type="checkbox"/>	25830	2268	Example	Relevant	Irrelevant	0.306	0.807	Rebuild
<input type="checkbox"/>	25830	1202	Most used Exclude	INCLUDE on title & abstract	Most used excludes	0.256	0.636	Rebuild
<input type="checkbox"/>	25830	1195	Exclude year	INCLUDE on title & abstract	EXCLUDE - Year < 2000	0.842	0.842	Rebuild

Run Model

You will be asked to confirm the action, as it can take some time to complete.



You will then see a message, suggesting you refresh the Search list to check for the results coming in. (Again, you can continue to use other functions in EPPI Reviewer whilst the system runs the model.)



When the results are ready, you will see them listed, with the line showing **true** in the “Classifier” column. Click the **true** link to visualise their distribution according to probability of being relevant.

The chart can be saved as a graphic via the **Save chart** button.

	No...	Name	Created By	Date	Hits	Classifier
<input type="checkbox"/>	2076	Items classified according to model: Example	Zak Ghouze	17 May 2022	42225	true

Create codes below this Code / Coding Tool: **Unscreened 17-5** Create

Items classified according to model: Example

Score Range	Approximate Number of Items
0-9	31,000
10-19	6,000
20-29	3,000
30-39	2,000
40-49	1,500
50-59	1,000
60-69	800
70-79	600
80-89	400
90-99	300

Search scores

More than
 Less than
 Between

50

Save chart

Close

	No...	Name	Created By	Date	Hits	Classifier
<input type="checkbox"/>	2076	Items classified according to model: Example	Zak Ghouze	17 May 2022	42225	true

You can search for certain ranges of probabilities; the results will be listed in your searches. Enter the parameters and click the **Search scores** button.

Search scores

More than
 Less than
 Between

60

Save chart
Close

	No...	Name	Created By	Date	Hits	Classifier
<input type="checkbox"/>	2079	<u>Search #771394 scores less than 60</u>	Zak Ghouze	17 May 2022	41675	false

You can also list the items classified in order of probable relevance by clicking on the number of **Hits**, as described for built-in classifiers earlier.

	No...	Name	Created By	Date	Hits	Classifier
<input type="checkbox"/>	2076	Items classified according to model: Example	Zak Ghouze	17 May 2022	42225	true

Review home
References
Reports
Search & Classify
Collaborate

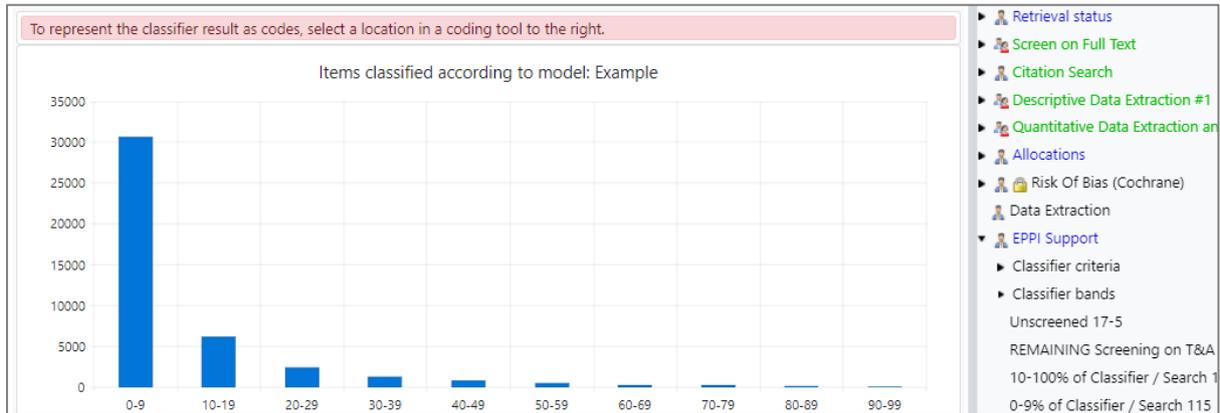
Import Items
Cluster
Coding Report
In/Exclude
Export to RIS
Run Reports

First
Previous
Page: 1 of 11
Next
Last
Showing 4000 Items of 42225
View Options

Showing Items classified according to model: Example
I
E
D

	ID	Short title	Title	Year	Score!
<input type="checkbox"/>	I	54977122 Ban, (2020)	Self-help groups, savings and social capital: Evidence from a field experiment in Cambodia	2020	98
<input type="checkbox"/>	I	54834975 Adoho (2014)	The impact of an adolescent girls employment program: The EPAG project in Liberia	2014	97
<input type="checkbox"/>	I	54978725 Naved, (2018)	A cluster randomized controlled trial to assess the impact of SAFE on spousal violence against women and girls in slums of Dhaka, Bangladesh	2018	96
<input type="checkbox"/>	I	54977125 McIntosh (2012)	The CLP's impact on women's empowerment	2012	95
<input type="checkbox"/>	I	54977119 Muhammad, (2012)	Women empowerment and microfinance: A case study of Pakistan.	2012	95
<input type="checkbox"/>	I	54834794 Buehren, (2017)	Adolescent Girls' Empowerment in Conflict-Affected Settings: Experimental Evidence from South Sudan	2017	95

Finally, the system can automatically create 10 codes based on the decile bands of probability, with each item being given one of the 10 codes accordingly.



Select the parent code where you want the 10 new codes to appear, then click the **Create** button.



The codes will be created and applied as requested.

Code	Count
0-9% range	30576
10-19% range	6109
20-29% range	2399
30-39% range	1304
40-49% range	794
50-59% range	493
60-69% range	247
70-79% range	147
80-89% range	77
90-99% range	37

The numbers

The Recall and Precision figures are built in the following way -:

- Upon receiving the training data, a proportion (10%) of randomly selected records is set aside for the purpose of calculating the “performance” figures - including precision and recall. (The algorithm will “fail” if it doesn’t receive enough records to train and then evaluate the result.)
- The remaining portion (the 90% majority of records) will then be used for training. Once the classification model is built, the initial “set aside” portion is used to evaluate performance. For this, precision and recall are calculated by setting the cut-off threshold at 0.5 (or “50%”).
- (NOTE: Rebuilding models, especially when there isn’t a great deal of training data, can produce slightly different precision and recall numbers, depending on what gets randomly selected for “evaluation”.)
- When you apply a model, you will hopefully find the majority of references that scored more than 50% are “true positives” i.e. have been classified correctly. You may also find a tail of true positives below the 50% score, generally accounting for a small minority of all true positives.
- Accuracy is usually calculated using precision / recall, and is produced from the same 10% records used for the other performance stats. (The specific formula used for the Accuracy score is documented here: https://scikit-learn.org/stable/modules/model_evaluation.html#accuracy-score)

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i)$$

You’ll find some background information here <https://paulvanderlaken.com/2019/08/16/roc-auc-precision-and-recall-visually-explained/> and formal definitions here https://en.wikipedia.org/wiki/Precision_and_recall. (Excuse the source, but it’s not a bad summary.)

We have a document on the Machine Learning available within ER Web available at https://eppi.ioe.ac.uk/CMS/Portals/35/machine_learning_in_eppi-reviewer_v_7_web_version.pdf.

(Other information on automation in ER can be found at <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3772>. Other information about the software can be found at <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3822>.)