

# Erratum to: Using text mining for study identification in systematic reviews: a systematic review of current approaches

Alison O'Mara-Eves<sup>1</sup>, James Thomas<sup>1\*</sup>, John McNaught<sup>2</sup>, Makoto Miwa<sup>3</sup> and Sophia Ananiadou<sup>2</sup>

## Erratum

Following publication of our article [1], it has come to our attention that two of the formulae in Table 1 were incorrect. The formulae for the measures of precision and burden have been corrected (Table 1). We are publishing this erratum to update these formulae to the following:

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Burden} = \frac{tp^T + tn^T + fp^T + tp^H + fp^H}{N}$$

Received: 6 March 2015 Accepted: 11 March 2015

Published online: 28 April 2015

## References

1. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Rev.* 2015;4:5.

\* Correspondence: j.thomas@ioe.ac.uk

<sup>1</sup>Evidence for Policy and Practice Information and Coordinating EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London, UK

<sup>2</sup>The National Centre for Text Mining and School of Computer Science, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK

<sup>3</sup>Toyota Technological Institute, 2-12-1 Hisakata, Tempaku-ku, Nagoya 468-8511, Japan

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



**Table 1 Definitions of performance measures reported in the studies**

Measure	#	Definition	Formula
<b>Recall (sensitivity)</b>	22	Proportion of correctly identified positives amongst all <i>real</i> positives	$\frac{TP}{TP+FN}$
<b>Precision</b>	18	Proportion of correctly identified positives amongst all positives.	$\frac{TP}{TP+FP}$
<b>F measure</b>	10	Combines precision and recall. Values of $\beta < 1.0$ indicate precision is more important than recall, whilst values of $\beta > 1.0$ indicate recall is more important than precision	$F_{\beta,k} = \frac{(\beta^2 + 1)TP_k}{(\beta^2 + 1)TP_k + FP_k + \beta^2 FN_k}$ Where $\beta$ is a value that specifies the relative importance of recall and precision.
<b>ROC (AUC)</b>	10	Area under the curve traced out by graphing the true positive rate against the false positive rate. 1.0 is a perfect score and 0.50 is equivalent to a random ordering	
<b>Accuracy</b>	8	Proportion of agreements to total number of documents.	$\frac{TP+TN}{TP+FP+FN+TN}$
<b>Work saved over sampling</b>	8	The percentage of papers that the reviewers do not have to read because they have been screened out by the classifier	WSS at 95% recall = $\frac{TN+FN}{N-0.05}$
<b>Time</b>	7	Time taken to screen (usually in minutes)	
<b>Burden</b>	4	The fraction of the total number of items that a human must screen (active learning)	$Burden = \frac{tp^r + tn^r + fp^r + tp^u + fp^u}{N}$
<b>Yield</b>	3	The fraction of items that are identified by a given screening approach (active learning)	$Yield = \frac{tp^r + tp^u}{tp^r + tp^u + fn^u}$
<b>Utility</b>	5	Relative measure of burden and yield that takes into account reviewer preferences for weighting these two concepts (active learning)	$\frac{\beta \cdot yield + (1 - burden)}{\beta + 1}$ Where $\beta$ is the user-defined weight
<b>Baseline inclusion rate</b>	2	The proportion of includes in a random sample of items before prioritisation or classification takes place. The number to be screened is determined using a power calculation	$\frac{n_i}{n_t}$ Where $n_i$ = number of items included in the random sample; $n_t$ = total number of items in the random sample
<b>Performance (efficiency)<sup>a</sup></b>	2	Number of relevant items selected divided by the time spent screening, where relevant items were those marked as included by two or more people	$\frac{\text{Selected, relevant items}}{\text{Time}}$
<b>Specificity</b>	2	The proportion of correctly identified negatives (excludes) out of the total number of negatives	$\frac{TN}{TN+FP}$
<b>True positives</b>	2	The number of correctly identified positives (includes)	TP
<b>False negatives</b>	1	The number of incorrectly identified negatives (excludes)	FN
<b>Coverage</b>	1	The ratio of positives in the data pool that are annotated during active learning	$\frac{TP^L}{TP^L + FN^L + TP^U + FN^U}$ Where $L$ refers to labelled items and $U$ refers to unlabelled items
<b>Unit cost</b>	1	Expected time to label an item multiplied by the unit cost of the labeler (salary per unit of time), as calculated from their (known or estimated) salary	$time_{\text{expected}} \times \text{cost}_{\text{unit}}$
<b>Classification error</b>	1	Proportion of disagreements to total number of documents	100 % – accuracy %
<b>Error</b>	1	Total number of falsely classified items divided by the total number of items	$\frac{\sum (FP+FN)}{\sum (TP+FP+FN+TN)}$
<b>Absolute screening reduction</b>	1	Number of items excluded by the classifier that do not need to be manually screened	TN + FN
<b>Prioritised inclusion rate</b>	1	The proportion of includes out of the total number screened, after prioritisation or classification takes place	$\frac{n_{ip}}{n_{tp}}$ Where $n_{ip}$ = number of items included in prioritised sample; $n_{tp}$ = total number of items in the prioritised sample