

Online Resource 4:

Receiver operating characteristics analysis of WEB-RADR reference set with shorter data collection

This is a *Drug Safety* online resource of the article 'An assessment of the utility of social media for broad-ranging statistical signal detection in pharmacovigilance: Results from the WEB-RADR project' by Ola Caster, Juergen Dietrich, Marie-Laure Kürzinger, Magnus Lerch, Simon Maskell, G. Niklas Norén, Stéphanie Tcherny-Lessenot, Benoit Vroman, Antoni Wisniewski, and John van Stekelenborg.

Corresponding author is Ola Caster (ola.caster@who-umc.org), Uppsala Monitoring Centre, Box 1051, 75140 Uppsala, Sweden.

As mentioned in the main article (see Section 2.4.1.1), the original intention was to perform the receiver operating characteristics (ROC) analysis of the WEB-RADR reference set in the same way as for the Harpaz reference [1], i.e. to collect data for positive controls only during the period before their index dates. The purpose of this is to ensure that positive controls are emerging safety signals rather than established adverse drug reactions (ADRs) at the point in time where they are evaluated.

The results from this design are shown in Fig. 1. The ROC curves from all data sources, including VigiBase, display an erratic and random pattern, and no data source offers any predictive ability.

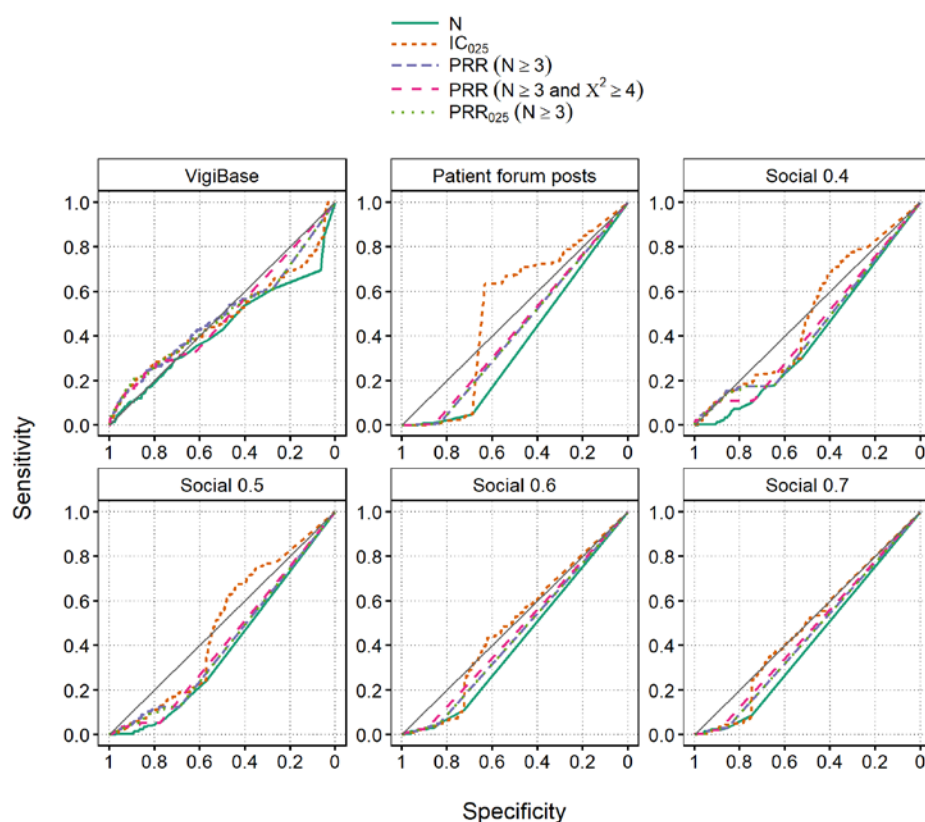


Fig. 1 ROC curves for the WEB-RADR reference set, using data from April 2012 up to the month prior to the index date (for positive controls) or up to March 2015 (for negative controls). 'Social 0.X' means Twitter/Facebook data with a post-level Indicator Score threshold of 0.X. The diagonal represents a random classifier

These results should be assessed in relation to those presented in Fig. 2 and Fig. 3 of the main article. The former shows an analysis of VigiBase only, using the same end of data collection as here, but without any restriction on the start of data collection. With that design there is predictive ability in VigiBase, and it was concluded that the reference as such is valid. Fig. 3, on the other hand, shows the results from a design with the same start of data collection as here (April 2012) but using the entire period up to March 2015 for all controls. Also with that design, VigiBase displays predictive ability.

The logical conclusion from these three analyses taken together is that the intended design (i.e. the one used here) suffers from too short data collection on the positive controls to generate any reliable results. Because we did not have access to further historical data from social media, the only viable option was to extend the data collection period for the positive controls, even though such a design might inflate predictive performance. The design used in the main article (with results in Fig. 3) included all available data for all controls, and yet no predictive ability whatsoever could be observed for social media. Hence, while this deviation from the intended design is regrettable, it does not threaten the main observation of no evidence for the value of statistical signal detection in the social media data sources considered here.

References

1. Harpaz R, Odgers D, Gaskin G, DuMouchel W, Winnenburger R, Bodenreider O et al. A time-indexed reference standard of adverse drug reactions. *Sci Data*. 2014;1:140043.