

The expected positive or negative impact of each feature was rated across four domains:

1. *Patient safety*: Is the inclusion of the feature in software likely to improve, worsen or have no effect on patient safety?
2. *Quality of care*: Is the inclusion of the feature in software likely to support, detract from or have no effect on the provision of good quality clinical care?
3. *Useful to the clinician*: Is the inclusion of the feature in software likely to enhance, disrupt or have no effect on clinician work practices?
4. *Useful to the patient*: Is the inclusion of the feature in software likely to be useful to the patient, impact adversely on the patient, or have no direct impact on the patient. Examples include improving the communication between prescribers and consumers, assisting consumer decision-making about their management, or meeting expectations about the way their health information is used.

An 11-point rating scale was used, similar to the 15-point scale used by Bell et al.[11] The scale ranged from -5 (largest negative impact) to +5 (largest positive impact), with -4 to -5 being for effects that were “clearly negative” and +4 to +5 for effects that were “clearly positive”. Scores of -1 to -3 and +1 to +3 indicated some negative or positive effect, and a zero value indicated no effect. It was also possible to choose “don’t know” if the person felt that they had insufficient knowledge or experience to rate that feature. It was to be assumed that the feature had been implemented properly in the software and worked well.

The median score and the interquartile range (IQR) were calculated for each of the four domains for each feature. “Agreement” between panel members was defined as an item with scores where the IQR was ≤ 2 and there were no outliers (defined as those scores that were more than $1.5 \times$ IQR from the 25th or 75th percentile) or the range of all 12 scores was ≤ 2 . “Disagreement” included all items not meeting the criteria for “agreement” ie. those where the IQR was >2 or there were outliers (scores $> 1.5 \times$ IQR from the 25th or 75th percentile).