higher (≥) than 80% (2) (8) (23). However, a further specific restriction applies to the sensitivity of the proposed *in vitro* test method inasmuch as only two *in vivo* Category 2 chemicals, *1-decanol* and *di-n-propyl disulphide*, may be misclassified as No Category by more than one participating laboratory. The specificity should be equal or higher (≥) than 70% (2) (8) (23). There is no further restriction with regard to the specificity of the proposed *in vitro* test method, *i.e.* any participating laboratory may misclassify any *in vivo* No Category chemical as long as the final specificity of the test method is within the acceptable range. The overall accuracy should be equal or higher (≥) than 75% (2) (8) (23). Although the sensitivity of the VRM calculated for the 20 Reference Chemicals listed in Table 1 is equal to 90%, the defined minimum sensitivity value required for any similar or modified test method to be considered valid is set at 80% since both *1-decanol* (a borderline chemical) and *di-n-propyl disulphide* (a false negative of the VRM) are known to be non-irritant in humans (31) (32) (33), although being identified as irritants in the rabbit test. Since RhE models are based on cells of human origin, they may predict these chemicals as non-irritant (UN GHS No Category).

Table 2: Required predictive values for sensitivity, specificity and overall accuracy for any similar or modified test method to be considered valid.

Sensitivity	Specificity	Overall Accuracy
≥ 80%	≥ 70%	≥ 75%

Study Acceptance Criteria

13. It is possible that one or several tests pertaining to one or more test chemicals does/do not meet the test acceptance criteria for the test and control chemicals or is/are not acceptable for other reasons. To complement missing data, for each test chemical a maximum number of two additional tests is admissible ("retesting"). More precisely, since in case of retesting also PC and NC have to be concurrently tested, a maximum number of two additional runs may be conducted for each test chemical.

14. It is conceivable that even after retesting, the minimum number of three valid runs required for each tested chemical is not obtained for every Reference Chemical in every participating laboratory, leading to an incomplete data matrix. In such cases the following three criteria should all be met in order to consider the datasets acceptable:

1. All 20 Reference Chemicals should have at least one complete run sequence.

2. In each of at least three participating laboratories, a minimum of 85% of the run sequences need to be complete (for 20 chemicals; *i.e.* 3 invalid run sequences are allowed in a single laboratory).

3. A minimum of 90% of all possible run sequences from at least three laboratories need to be complete (for 20 chemicals tested in 3 laboratories; *i.e.* 6 invalid run sequences are allowed in total).