Variability in the Rabbit Skin Irritation Assay

J.P. Rooney[1], N. Choksi[1], P. Ceger[1], A. Daniel[1], J. Truax[1], D.G. Allen[1], N. Kleinstreuer[2]

[1]ILS, RTP, NC, United States; [2]NIH/NIEHS/DNTP/NICEATM, RTP, NC, United States

The in vivo rabbit test is the benchmark against which new approach methodologies for skin irritation are usually compared. In vitro methods are accepted as partial replacements for the in vivo test because they can distinguish irritant from non-irritant chemicals (as defined by the rabbit test). However, because none can classify substances as moderate and mild irritants, there are currently no in vitro methods that are considered full replacements for the in vivo assay. A limiting factor in identifying a complete replacement could be the variability inherent to the subjective scoring of erythema and edema responses in the rabbit test. This is particularly relevant for mild and moderate irritants, where interindividual differences in scoring are most likely to occur. To better characterize the reproducibility of the in vivo assay, we assessed variability in animal study results from substances tested multiple times. We compiled and curated 2624 test records, representing 990 unique mono-constituent substances, each tested at least twice. Methodological deviations from guidelines were noted, and multiple data sets with differing tolerances for such deviations were created. Where possible, primary dermal irritation indices were estimated from the available data and used to classify chemicals according to the U.S. Environmental Protection Agency (EPA) skin irritation classification criteria. Globally Harmonized System (GHS) hazard classifications were also extracted from study reports when available. Conditional probabilities were used to evaluate the reproducibility of the in vivo method in identification of EPA or GHS hazard categories. Chemicals classified as moderate irritants at least once were classified as mild irritants or non-irritants at least 40% of the time when tested repeatedly. Variability was greatest between mild and moderate irritants, which both had less than a 50% likelihood of being replicated. Increased reproducibility for the EPA and GHS systems was observed when a binary categorization was compared between corrosives/moderate irritants and mild/non-irritants. This analysis indicates that variability present in the rabbit skin irritation test should be considered when evaluating the performance of nonanimal alternative methods as potential replacements. This project was funded in whole or in part with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C.