A Large Scale Experiment In N-Version Programming by John C. Knight, Nancy G. Leveson and Lois D. StJean, 1985.

Abstract
XH: Experience report
Studies reported here of <N-version programming> supported by <comparing the outputs of redundant programs written by 27 students> generate a number of findings concerning <how well N-version programming ensures reliability>, including <the fact that programmer errors are not statistically independent>. They indicate that <a geometric decrease in system error (as would be guaranteed by statistically independent programmer errors)> is not met by <the N-version programming approach that we studied>.

Why XH?
The authors tried out a supposedly promising approach and found that it didn’t work as well as anticipated.

Question - [Evaluation analysis of an instance]
How much does N-version programming improve reliability if we take a single specification, give it to multiple students at two universities who are forced to use the same tooling?

This is, of course, part of a bigger question (Method/means of development) asking about how we can produce more reliable software.

Results - [Report]
The authors tried it out and found that reliability didn’t go down as much as would be the case if programmer errors were uncorrelated. They report that error correlation actually spanned universities, which was a little surprising.

Validation – [Experience]
They actually tried the approach out and analyzed their empirical data statistically.