

The true meaning of DICE: don't ignore chance effects

Mike Clarke

Centre for Public Health, Institute of Clinical Sciences, Queens University Belfast, Royal Victoria Hospital, Belfast BT12 6BJ, UK

Corresponding author: Mike Clarke. Email: m.clarke@qub.ac.uk

Introduction

Researchers often name their trials using an acronym. These include some from the past which might raise eyebrows now, such as ISIS¹ and the many attempts to incorporate 'cov' into the names of trials relating to COVID-19, including the RECOVERY trial.² However, one acronym with which all trialists, reviewers and users of research should be made more familiar is DICE – 'don't ignore chance effects'.

Over 20 years beginning in the early 1990s, a series of DICE studies have highlighted how chance could affect even a perfectly designed randomised trial, which had 100% adherence to the allocated interventions and no loss to follow-up, or a mathematically perfect meta-analysis. This should not be surprising, given the fundamental principle of randomised trials that, thinking about a typical 2-group individually randomised trial, participants are allocated to an intervention and a control group by chance.³ Thus, even if the intervention has absolutely no additional effect compared with a control, then, purely by chance, the groups could have different average outcomes. Whether this might have happened is tested using the statistical significance of the between-group difference. However, setting this to the traditional threshold of $p = 0.05$ will lead to 'statistically significant' differences with almost the same frequency as people rolling 11 with a pair of dice. The problem becomes even worse if multiple analyses are done and the one with the most striking difference, or lowest p -value, is elevated to become a key result of the trial.⁴ The DICE studies attempted to illustrate this for trials and reviews, and they serve as a cautionary tale for everyone involved in the conduct or use of controlled trials and meta-analyses.

DICE 1

In the early 1990s, as part of an exercise to teach doctors about clinical trials and systematic reviews, stroke doctors were asked to generate a series of simulated randomised trials of a therapy called 'Dice' which, when their results were combined in

meta-analyses, might have sufficient statistical power to detect a moderate treatment effect.⁵ The study, which became known as DICE 1, focused particularly on whether the combination of chance, biased decisions about including studies in a meta-analysis, inappropriate subgroup analysis and publication bias could lead to a conclusion that Dice therapy was beneficial and could save the lives of patients in specific circumstances, even though it should have no impact whatsoever.

Each participant on the course was given a red, green or white die and asked to write their name and the colour of the die on a data form. They then rolled their dice a specified number of times to represent the number of patients in the treatment group of a randomised trial, with each six recorded as a death on the form and all other numbers recorded as a survival. This was then repeated the same number of times for the control group. For each participant, this first trial was followed by a second of a different size. The trials varied in size from five in each group (total of 10) to a total of 200, with 100 rolls of the dice for the treatment group and 100 for the control group. None of the dice was biased and there was no reason other than chance for the 'trials' to produce different results. However, when the trials were analysed to test three hypotheses (see below), the danger that chance and bias can lead to misleading conclusions became clear.

The hypotheses were that the results would be meaningfully different for (a) trials of low quality (defined as incomplete or incorrect data forms) or specific different dice colours, (b) participants' second trials when comparing the results to those of their first trials and (c) when an adjustment for 'publication bias' was made by including in the meta-analysis 70% of the positive trials but only 40% of the null or negative trials.

A meta-analysis of the results of all 44 trials produced a statistically non-significant reduction in the odds of death of 11% (95% confidence interval (CI): 33% decrease to 11% increase, $p > 0.10$). However, if the low quality and red dice trials were excluded, the survival benefit rose to 22% (95% CI: 42% decrease

to 4% increase, $p=0.09$) and when publication bias was added in, the 'published' trials showed a decrease in the odds of death of 23% (95% CI: 43% decrease to 3% increase, $p=0.07$). Finally, when only the 'published' second trials were used, on the basis that the doctors were more familiar with the intervention after gaining experience with their first trials and therefore better able to administer it effectively, the intervention now showed a statistically significant decrease in the odds of death of 39% (95% CI: 60% decrease to 8% decrease, $p=0.02$). If this was true, the intervention would prevent 70 premature deaths for every 1000 stroke patients who received Dice therapy; but clearly, it is not a fair representation of what would happen with Dice therapy and the result is due to chance and bias.

DICE 2

Instead of using simulated data generated by rolling dice, DICE 2 used anonymised data on 580 patients who were in the control group of a randomised trial of a treatment for colorectal cancer, of whom 147 had died.⁶ Each patient was randomly coded to simulate allocation to a treatment or control group, and the resulting 100 'trials' were then analysed to compare the effects of treatment versus control on time to death. Furthermore, to highlight the possible dangers of multiple subgroup analyses, 50 subgroups were generated within each trial with patients being randomly categorised as a type A or type B person in each of these subgroups.

As expected, the overall analyses for most of the 100 trials yielded statistically non-significant differences, but four were conventionally statistically significant with p -values for the time to death analysis of less than 0.05. The most extreme had a p -value of 0.003 and showed an absolute reduction in four-year mortality of 40% (SD 15) for patients in the treatment group compared to those in the control group. Turning to the subgroup analyses for this trial, subgroup simulation 13 showed that the survival 'benefit' was present for patients in only one of the two subgroups, in whom the four-year mortality was reduced by 64% (SD 16, $p=0.00006$). If true, this would be a substantial benefit for this lucky subgroup of patients but, again, it is not a fair representation of the truth. It is due to chance and the biased decision to focus on the subgroup analysis within the most 'beneficial trial'.

DICE 3

The patient data used for DICE 2 were re-used in DICE 3,⁷ which explored how biases in meta-analysis

can, when combined with chance, lead to over-promising, but incorrect findings. The aim was to show the effects of chance on meta-analyses and how, if the results of a favourable trial prompted the decision to do a meta-analysis, this could have important implications for the interpretation of its results.

Using the anonymised data on 580 control patients from a randomised trial of a treatment for colorectal cancer, 100,000 randomised trials were simulated in which patients were randomly labelled as treatment or control. These trials were then combined into 10,000 meta-analyses, each containing 10 of the simulated trials. The main outcome was, once again, time to death.

As expected, approximately 5% of the 100,000 trials gave statistically significant results for the difference between treatment and control: 4897 (4.9%) at $2p < 0.05$. Furthermore, approximately 1% of the 10,000 meta-analyses were statistically significant at $2p < 0.01$: 123 (1.2%). However, also as expected, some of the results, all of which were due only to chance, were extreme. The most extreme result for a meta-analysis was a 20% reduction (99% CI: 0.70 to 0.91; $2p < 0.00002$) in the annual odds of dying in the treatment group compared with that in the control group, which would be an important benefit of a real treatment for patients with colorectal cancer.

Moving on to demonstrating the impact of initiating a meta-analysis because of particularly promising trial results, the simulations showed that if a meta-analysis contained at least one trial with a statistically significant result (at $2p < 0.05$) (as might be the case if an initial trial is regarded as hypothesis generating), there can be a striking increase in the likelihood of the meta-analysis being statistically significant (at $2p < 0.01$). For example, among the 473 meta-analyses in which the first trial in a batch of 10 was statistically significant (at $2p < 0.05$), the proportion of meta-analyses favouring treatment at $2p < 0.01$ rose to 3.8% (18 meta-analyses). Again, this is not a fair representation, but it is not unusual for promising results to lead to meta-analysis. As with DICE 1 and DICE 2, the result is due to chance and bias, which, in this case is the decision to do a meta-analysis because of the results of a trial that is then re-used in the meta-analysis.

Conclusions

The concept that chance can affect the results of trials and meta-analyses is fundamental to the use of statistical significance testing⁸ but too often seems forgotten once the statistical test has been done and the 'magic' p -value has been achieved. However, given the hundreds of thousands of trials and tens of

thousands of meta-analyses that might currently be in use by decision makers, including the vast number of reviews for COVID-19,⁹ people using the results of these research studies need to be cautious that any result that seems to have ruled out the effects of chance because of its low *p*-value or the distance of the confidence interval from the position of no difference, may still be due to chance and that things might be even worse if the result has been inflated by bias.

Everyone doing, analysing, reporting or reading the report of a trial or meta-analysis should consider: might the most likely cause of the result be bias, chance or a combination of the two? They should ask themselves many questions about why the result and conclusion might be wrong. Once they have exhausted those possibilities, then they might be right; but they should always remember DICE: don't ignore chance effects.

Declarations

Competing Interests: None declared.

Funding: None declared.

Ethics approval: Not applicable.

Guarantor: Mike Clarke.

Contributorship: Mike Clarke is the sole author of this paper.

Provenance: Invited article from the James Lind Library.

Acknowledgements: I am grateful for the opportunity to work on the DICE studies with Carl Counsell, Jim Slattery, Peter Sandercock and Jim Halsey. I thank Iain Chalmers for his comments on the initial draft of this article, for inclusion in the James Lind Library.

References

1. ISIS-2 (second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; 332: 349–360.
2. Glasziou PP and Tikkinen KAO. The RECOVERY trial platform: a milestone in the development and execution of treatment evaluation during an epidemic. *JLL Bulletin: Commentaries on the history of treatment evaluation* 2021. <https://www.jameslindlibrary.org/articles/the-recovery-trial-platform-a-milestone-in-the-development-and-execution-of-treatment-evaluation-during-an-epidemic/>.
3. The James Lind Library 2.2 The need to compare like-with-like in treatment comparisons. See <https://www.jameslindlibrary.org/essays/2-2-the-need-to-compare-like-with-like-in-treatment-comparisons/> and <https://www.jameslindlibrary.org/essays/2-5-bias-introduced-after-looking-at-study-results/>.
4. The James Lind Library 2.5 Bias introduced after looking at study results.
5. Counsell CE, Clarke MJ, Slattery J and Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *BMJ* 1994; 309: 1677–1681.
6. Clarke M and Halsey J. DICE 2: a further investigation of the effects of chance in life, death and subgroup analyses. *Int J Clin Pract* 1994; 55: 240–242.
7. Clarke M and Halsey J. Dicing with chance, life and death in systematic reviews and meta-analyses: D.I.C.E. 3, a simulation study. *J R Soc Med* 2014; 107: 116–119.
8. Matthews RAJ. The origins of the treatment of uncertainty in clinical medicine. Part 1: ancient roots, familiar disputes. *JLL Bulletin: Commentaries on the history of treatment evaluation* 2020. See <https://www.jameslindlibrary.org/articles/the-origins-of-the-treatment-of-uncertainty-in-clinical-medicine-part-1-ancient-roots-familiar-disputes/>.
9. Dotto L, Kinalska MA, Machado PS, Pereira GKR, Sarkis-Onofre R and Dos Santos MBF. The mass production of systematic reviews about COVID-19: an analysis of PROSPERO records. *J Evidence-Based Med* 2021; 14: 56–64.