The consensus statement by Kahan and colleagues¹ addresses a critical methodological gap, extending the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidance for factorial trials.² Factorial trials are sample size–efficient designs requiring specific methodological considerations during design and analysis. A 2020 meta-analysis by Kahan et al³ found that reporting of these trials is often inadequate, hindering the validity of trials’ conclusions. From this point stems the need for dedicated guidance.

Compared with parallel-group studies, factorial trials allow for the concurrent evaluation of 2 or more interventions, effectively conducting 2 trials in 1.⁴ The meta-analysis by Kahan et al³ reported that only 22% of the analyzed trials provided a clear rationale for adopting this design. Based on a specific research question, the choice of a factorial design over a 3-group trial should be clearly motivated and grounded in a clinical and a statistical rationale.

The main advantage of a factorial design is in performing 2 or more comparisons in the most efficient way in terms of number of patients to be enrolled (sometimes called 2 trials in 1). Sample size calculation in a factorial design typically involves separate assessments based on target effect sizes for each treatment relative to their respective control. The final sample size is the larger of these, powered to detect the main effects of each intervention.⁵ By performing this analysis, it is possible to obtain evidence about efficacy from fewer patients than would be needed if the 2 treatments were individually tested in 2 separate trials or in a 3-group trial.⁶ The meta-analysis by Kahan et al³ reported that 42% of trials calculated sample sizes using a factorial design, while 28% lacked clarity whether calculation was based on a factorial of multigroup approach.

Most of the efficiency of the factorial design depends on the assumption of no positive or negative interaction between the treatments to be compared, although this assumption’s reasonability can vary across different contexts.⁷ Some trials could be designed to explore an interaction; in this case, the number of patients needed increases notably. Violations of the assumption of independency can affect the validity of conclusions about the effects of interventions. Proper evaluation of this assumption through an interaction test is essential, and it is recommended that multigroup analyses, which do not rely on the assumption of interaction, are also presented to complement the main factorial analyses. Interactions are usually assessed using a significance test, with $P > .05$ indicating no interaction. This approach has its pitfalls, and a preferable approach would be to present the size of the interaction term alongside its 95% CI. The challenge is whether an interaction test that is not significant truly indicates the absence of interaction or merely reflects inadequate statistical power. Factorial design can indeed reveal interactions, but when an interaction truly exists and the trial was designed under the assumption of no interaction, the planned sample size may not be enough to ensure adequate statistical power. Consequently, a high $P$ value, often interpreted as evidence of no interaction, can impart a deceptive sense of result robustness when it most likely reflects limited power. The meta-analysis by Kahan et al³ highlighted how no interaction was assumed in the sample size calculation for 82% of trials. Moreover, 37% of trials did not assess the presence of interaction for their primary outcome, while 44% of trials only reported the $P$ values and 18% of trials only reported a statement about the absence of interaction.⁸ A mere 12% of trials provided an estimate of the interaction size, while only 3% of trials reported CIs.³
The presence and nature of interaction are pivotal in determining the appropriateness of a factorial trial. If the direction of the effect of 1 intervention varies with different levels of another intervention (ie, a qualitative interaction), a factorial trial should be powered to detect this interaction when it is of primary interest. Conversely, for minor differences in the magnitude of the effect of the intervention (ie, a quantitative interaction), a factorial trial powered on main effects may be appropriate. Presenting intervention effects when a sizeable interaction is present requires caution. For qualitative interactions, main effects from the factorial analysis are likely to be misleading; the size of interaction with 95% CIs should be reported and results from the multigroup analysis should be included. For quantitative interactions, while the interaction and the cell means must still be presented, the main effects may nonetheless be a reasonable representation of intervention effects. In practice, when planning a factorial trial, the potential presence of interaction should be carefully considered. Even when an interaction is planned in the design, unrecognized interactions may still emerge, distorting results and interpretation. Such occurrence could make the primary results of the trial practically useless, with the unique and useful information coming from secondary, lower-powered analyses. Therefore, it is generally recommended to conduct factorial trials when no interaction is expected or when the outcome of interest is specifically the interaction between treatments.

Factorial trials often analyze secondary outcomes, demanding careful assessment of interactions for these outcomes. In the 2020 meta-analysis by Kahan et al, 62% of trials assessed the presence of interactions for secondary outcomes, with 41% revealing significant interactions for at least 1 secondary outcome. Interactions between secondary outcomes have their own implications, and ruling out an interaction for the primary outcomes does not guarantee the absence of interactions throughout the trial. Multiplicity adjustments are essential in this context, as assessing numerous interactions raises the risk of spurious statistical significance. Sensitivity analyses should be planned in advance for all outcomes, ensuring comprehensive evaluation while mitigating the risk of inflated type I errors.

The optimal analysis strategy for factorial trials hinges on whether treatments act independently or exhibit interactions. Factorial analyses are only appropriate when treatments act independently: in the presence of an interaction, the factorial approach is no longer suitable and multigroup analyses should be conducted. The choice of a factorial vs multigroup design should be explicitly stated in advance: basing the final analysis model on a preliminary test for interaction (the 2-stage approach) is discouraged, as it can introduce bias. The preliminary test has low power to detect interactions, leading to factorial analysis even for moderate interactions. Another generally discouraged practice is to include an interaction term in the statistical model for a factorial analysis: this approach can distort the interpretation of treatment effects, discarding 50% of the sample size from each comparison and leading to a substantial loss in power and precision. The meta-analysis by Kahan et al reported that 51% of trials assessed used a factorial analysis as their primary analysis approach, while 23% of trials used a multigroup analysis; however, only 26% of trials presented results for both analyses. Moreover, 18% of trials chose their primary analysis method of analysis on the basis of an interaction test, and 13% of trials included an interaction term in the model.

In summary, factorial trials offer a powerful means of efficiently evaluating multiple treatments within a single study. However, their methodological intricacies demand transparent reporting of rationale, assumptions, and analytical approaches. The extension of SPIRIT guidance for factorial randomized trial protocols represents a step toward enhancing the quality of evidence produced by such trials. Transparency in trial design, robust assessment of interactions, careful sample size determination, and comprehensive reporting practices are vital to harnessing the full potential of factorial trials and advancing evidence-based health care and policy decisions.