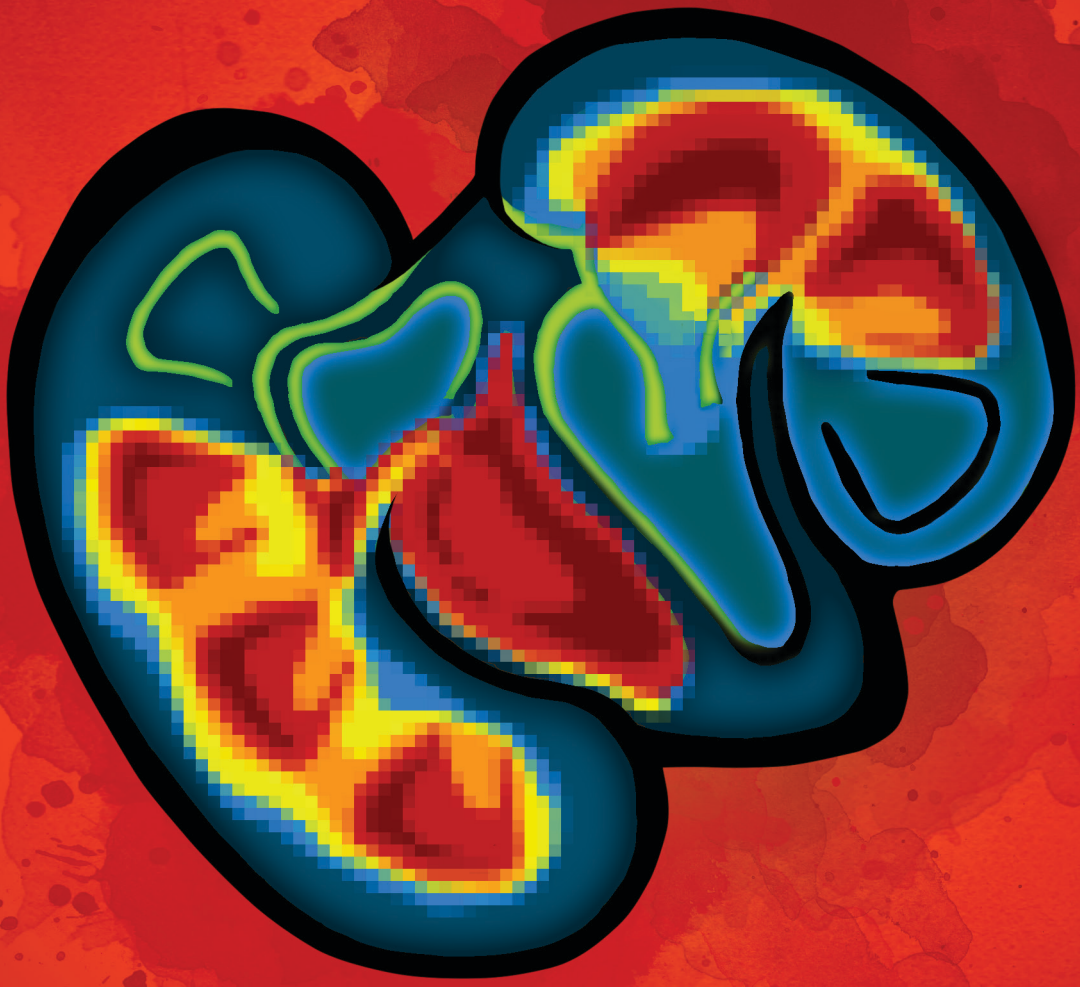


FLUIDAL CONNECTION

Part I

METHODOLOGICAL CONCEPTS

CHAPTER 3

On the Understanding of Statistical Interaction for Clinical Investigators

Milos Brankovic, Isabella Kardys, Ewout W. Steyerberg, Stanley Lemeshow, Maja Markovic, Dimitris Rizopoulos, Eric Boersma

Submitted

ABSTRACT

Despite testing for statistical interactions is usually stated as the secondary study objectives, it is not uncommon that these results lead to changing of treatment protocols or even modify the public health policies. For this reason, statistical interactions are studied frequently in clinical studies, but recent reviews have indicated that their proper assessment and reporting remains challenging for the clinical investigators. This article provides an overview of the challenges associated with the statistical interaction analysis to help the clinical investigators finding the best strategy to properly obtain and critically evaluate its presence in statistical models. Specifically, we discuss the importance of understanding the distinction between effect-measure modification and causal interaction, their qualitative and quantitative forms, the importance of a measurement scale on which interactions are tested, additive and multiplicative interaction measures, the relevance of multiple testing, and distinction between prespecified versus post-hoc analyses. Finally, we provide the recommendations that, if adhered to, could increase the clarity and the completeness of future studies. The understanding of the elements underlying statistical interaction analysis followed by its proper assessment and reporting may help in making the results more reliable, but also in facilitating clinical studies to use this type of analysis even more in the future.

INTRODUCTION

Many reasons motivate the study of statistical interaction of which the most fundamental are those to learn how to use an intervention most effectively, who would and who would not benefit (and who would benefit the most), or whether it would be harmful in specific subpopulations.[1] Although these reasons are usually stated as the secondary study objectives, if incorrectly performed statistical interaction analysis may cause false conclusions leading to unnecessary withholding of treatment, ineffective or even harmful treatment's effect.[2]

Despite the concept of statistical interaction is not new, it still poses a problem for the clinical investigators. In 2000, Assmann[3] et al. reviewed 50 randomized clinical trials (RCTs) in high-impact journals, and found that 70% of these trials performed interaction analysis but only 43% reported the test and 37% only a p-value. In 2006, Hernandez[4] et al. reported similar results after investigating published cardiovascular RCTs. In 2007, Wang[1] et al. evaluated 97 RCTs of which 61% used interaction analysis. Of those, 68% were unclear whether analyses were prespecified or post-hoc and only 27% reported an interaction test. Besides in RCTs, Knol[5] et al. found that vast majority of cohort and case-control studies also performed inappropriate interaction analysis. Finally in 2017, Wallach[6] et al. concluded that 61% of the RCTs the claimed the subgroup heterogeneity already in their abstracts (assuming these are the most credible) were, in fact, not supported by their results. For these reasons, previous reports tried to address this important topic.[1-3,7,8] These attempts, although informative, were directed for the most part to a narrow set of issues. For example, no discussion was performed for distinguishing different types of statistical interaction, or the importance of a measurement scale on which an interaction is tested. To date, a few reports[9,10] provide recommendations on some of these issues, but are intended mainly for an epidemiological audience.

In this paper, we summarize the evidence from the literature and provide the recommendations to assist the clinical investigators in selecting the best strategy to appropriately use, but also to critically evaluate, statistical interaction analyses as they might affect their decisions in clinical practice. In the following sections, we start by distinguishing different types and forms of statistical interaction; we then discuss how to properly analyze statistical interactions by the stratification or by an interaction modeling (i.e., inclusion of a cross-product term) and eventually how to report obtained results.

Types of statistical interaction

Statistical interaction can be classified as being either effect-measure modification or causal interaction. *Effect-measure modification* is present when the effect of one factor, exposure or intervention, on an outcome varies across the levels of another factor when no bias is present (Box 1).[11] Notably, the second factor does not need to affect the outcome for the effect-measure modification to be present, but only be related to another variable that does.[12] Some authors refer to this phenomenon also as an "effect heterogeneity".[13,14] Hence, the clinical motivation behind the effect-measure modification (or heterogeneity) analysis is to identify the subgroups of patients in whom a factor's effect differs based on patients' characteristics. If the effect of one factor is higher with higher levels of another factor an effect-measure modification is *positive*, whereas if this effect is lower an effect-measure modification is *negative*.

Causal interaction[15] is present when the combined effect of two factors on an outcome differs from their separate effects when no bias is present. (Box 1).[11] Unlike for effect-measure modification, both factors have to be causally related to an outcome in order for causal interaction to be present.[16] Despite it sounds theoretical, this distinction is important to be made especially if an intervention on the secondary factor is of interest.[17] For example, if an investigator would like to test whether cholesterol-lowering drug reduces the risk of myocardial infarction, and a positive interaction between the cholesterol treatment and hypertension is observed this would indicate that hypertension modifies the treatment's effect. Thus, targeting the subgroup of patients with hypertension would maximized the treatment's effect. However, if an investigator would also be interested in testing whether introducing secondary intervention (i.e., antihypertensive treatment) would further reduce the risk of myocardial infarction he/she should make sure that the secondary factor (i.e., hypertension) not only modifies the effect of the cholesterol treatment but is causally related to myocardial infarction. If so, causal interaction is present and a factorial design can be applied to confirm the hypothesis. Finally, a positive causal interaction indicates that the effect of two factors together is larger than the two factors considered separately, whereas a negative causal interaction indicates that this joint effect is smaller than these effects considered separately.

Effect-measure modification

- To assess if the effect of a factor (i.e., exposure or intervention) varies across levels of another factor when no bias is present
- Synonyms: effect modification, effect heterogeneity
- It can be positive or negative; qualitative or quantitative

Causal interaction

- To assess if the combined effect of two factors together (i.e., exposures or interventions) is different than their separate effects when no bias is present
- It can be positive or negative; qualitative or quantitative

BOX 1 Types and forms of statistical interaction. In a concrete analysis, the term "effect-measure" should be replaced with the name of exact measure that is used to estimate the effects in the statistical model. For example, if one would use the logistic regression model, a statistical interaction should be reported as the odds-ratio modification (or heterogeneity). Similarly, if Cox regression model is applied then hazard-ratio modification (or heterogeneity) would be more appropriate terminology. In this way, ambiguity about which effect is tested would be resolved.

Forms of statistical interaction

Statistical interaction can take either quantitative or qualitative form. The *quantitative* form (synonym[18]: "non-crossover") is the most common and is present when an effect of one factor has a different magnitude, but in the same direction, across strata of another factor (Figure 1: 1-4, 7, and 8).

The *qualitative* form (synonym[18]: "crossover") is present (1) if one factor does not have an effect on the outcome in one stratum, but does have effect in other stratum, of the second factor (Figure 1: 5a and 6a) or (2) if one factor has opposite effects depending on the strata of the second factor (Figure 1: 5b and 6b). Of note is that detection of qualitative interactions also depends on a study's selection criteria. For example, angiotensin-converting-enzyme inhibitors are beneficial in hypertensive patients, but are harmful in hypertensive patients due to reno-vascular disease.[19] If the latter group is excluded from the study due to selection criteria, an important qualitative interaction will be missed. This may lead to serious consequences if the study concludes that both groups of patients should be treated identically.

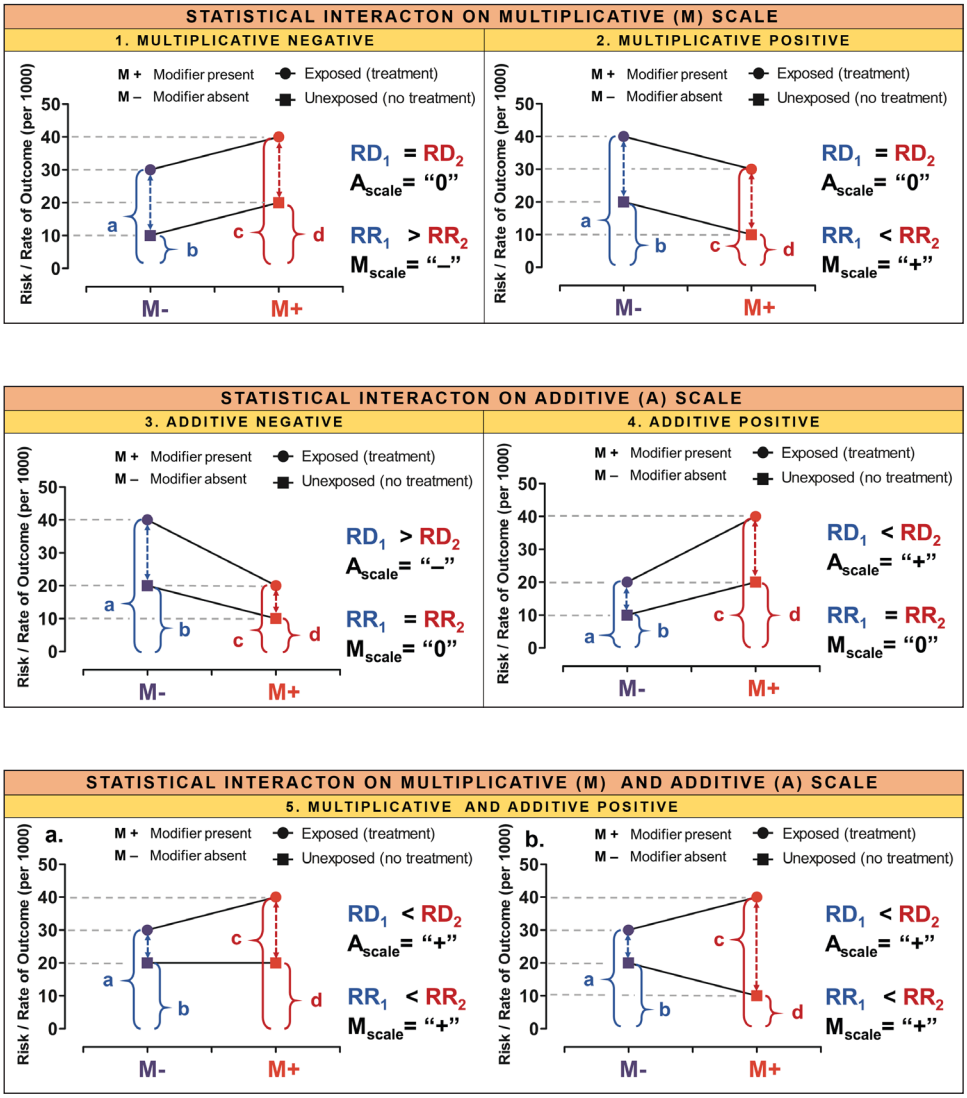
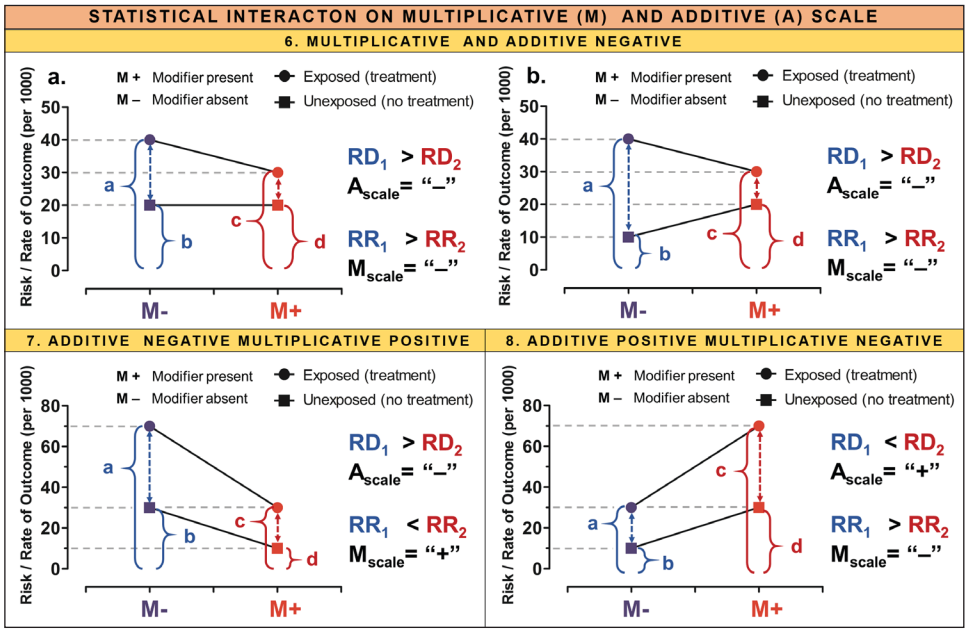


FIGURE 1 Potential scenarios that can be found when statistical interaction is detected by additive and multiplicative scales simultaneously. "a" denotes the effect in exposed (or treated) subgroup without modifier M; "b" denotes the effect in unexposed (or untreated) subgroup without modifier M; "c" denotes the effect in exposed (or treated) subgroup with modifier M; "d" the effect in unexposed (or untreated) subgroup with modifier M. RD_1 can be calculated as a – b; RR_1 can be calculated as a / b; RD_2 can be calculated as c – d; RR_2 can be calculated as c / d; numbers presented on Y-axes can be used to calculate RD_1, RD_2, RR_1, and RR_2. If there is departure on one of the two scales, eight possible scenarios can be observed: 1) no additive departure ($RD_1 = RD_2$), but negative multiplicative departure ($RR_1 > RR_2$); 2) no additive departure ($RD_1 = RD_2$), but positive multiplicative departure ($RR_1 < RR_2$); 3) no multiplicative departure ($RR_1 = RR_2$), but negative additive departure ($RD_1 > RD_2$); 4) no multiplicative departure ($RR_1 = RR_2$), but positive additive departure ($RD_1 < RD_2$); 5)



positive multiplicative and additive departures ($RD_1 < RD_2$ and $RR_1 < RR_2$) with two additional situations 5a) the effect is present only in one subgroup or 5b) the opposite effects are present in subgroups; 6) negative multiplicative and additive departures ($RD_1 > RD_2$ and $RR_1 > RR_2$) with two additional situations 6a) the effect is present only in one subgroup or 6b) the opposite effects are present in subgroups; 7) negative additive ($RD_1 > RD_2$) and positive multiplicative departures $RR_1 < RR_2$); 8) positive additive ($RD_1 < RD_2$) and negative $RR_1 > RR_2$) multiplicative departures.

ASSESSMENT OF STATISTICAL INTERACTION

As noted above, there are two ways to assess statistical interactions: (1) *stratification* (i.e., stratified or subgroup analysis) in which the effect of one factor is assessed within strata of another factor separately, (2) *interaction modeling* in which both factors are included into a statistical model together with their cross-product term ($F_1+F_2+F_1{}^*F_2$).

Before introducing their technical descriptions it is important to note that a statistical interaction is observed only if there is a departure from an underlying measurement scale on which a statistical model estimates effects. This means that a statistical interaction is scale-dependent. However, different statistical models estimate effects on different measurement scales. For example, standard linear regression coefficients estimate the sum of effects on an additive scale, whereas standard logistic regression and Cox regression exponentiated coefficients estimate the product of effects on a multiplicative scale such as risk ratio (RR), odds ratio (OR), or hazard ratio (HR)

scale. Importantly, additive and multiplicative scales do not always provide us with the same conclusion whether a statistical interaction is present or in which direction it operates. For this reason, both additive and multiplicative interaction measures are discussed below.

Additive interaction measures

A departure on an additive scale would mean that the combined effect of two factors is larger (in case of positive interaction) or smaller (in case of negative interaction) than the sum of their individual effects.[20]

For a binary outcome, e.g., death ("yes", "no"), and two binary factors, e.g., disease A and disease B ("yes", "no"), an additive interaction can be assessed using stratification and expressed as the absolute excess risk due to interaction (AERI) (Table 1: equation-1). For example, Weiner et al. studied the effects of chronic kidney disease (CKD) and cardiovascular disease (CVD) on the 10-year risk of the composite endpoint including cardiovascular and all-cause death.[21] Authors reported the absolute cumulative risk of 66% in individuals with both CKD and CVD, 34% in those with CKD but without CVD, 38% in those without CKD but with CVD, and 15% in those without CKD or CVD. The AERI is calculated as 66 + 15 – 34 – 38 = 9% which indicates a super-additive (i.e., positive) interaction because AERI >0 (detailed calculations are described in the supplemental text). This also indicates an absolute excess risk of 9% due to the interaction itself.

For a continuous outcome (e.g., blood pressure), and two categorical or continuous factors or their combination, an additive interaction can be assessed by including both factors together with their cross-product term into a linear regression model (Table 1: equation-2). In this case, β coefficient for the cross-product term would quantify the interaction on an additive scale.

When using continuous factors, a magnitude of statistical interaction will differ based on its unit-scale.[20] For example, if an investigator assesses whether a patient's age modifies the treatment's effect, the magnitude of the interaction between age and treatment will differ if age is expressed per 1-year, 5-year interval, or in some other units. Finally, a nice feature of regression models is that controlling for other covariates can easily be performed by including them into the model.

Multiplicative interaction measures

A departure on a multiplicative scale would mean that the combined effect of two factors is larger (in case of positive interaction) or smaller (in case of negative interaction) than the product of their individual effects.[20] Thus, the multiplicative scale corresponds to the ratios of effects rather than their difference as the additive scale does.

For a binary outcome and two binary factors, a multiplicative interaction can be assessed using stratification and expressed as the ratio of RRs (Table 2: equation-11). In the example above[21], the RRs of composite endpoint were 4.4 in individuals with both CKD and CVD, 2.3 in those with CKD but without CVD, 2.5 in those without CKD but with CVD as compared to those with neither, and 1.0 in those without CKD or CVD (supplemental text). Here, a multiplicative interaction is calculated as 4.4 / (2.5 * 2.3) = 0.8 which indicates a sub-multiplicative (i.e., negative) interaction between CKD and CVD because the ratio of RRs <1. This also indicates relative risk ratio due to interaction of -20%. However, the AERI indicated their super-additive interaction with absolute excess risk of 9%. Therefore, this example illustrates an aforementioned point that a measurement scale influences the presents and the direction of a statistical interaction.

For a binary outcome and two categorical or continuous factors or their combination, a multiplicative interaction can be assessed by including both factors together with their cross-product term into the logistic or Cox regression model (Table 2: equation-12 and equation-13). In the example above[21], OR or HR for the cross-product term would correspond to 0.8 indicating a sub-multiplicative interaction.

Additive versus Multiplicative scale

Figure 1 illustrates eight potential scenarios that can be found when statistical interaction is detected by additive and multiplicative scales simultaneously. In six of eight scenarios (Figure 1: 1-4, 7, and 8) these scales carry different information regarding statistical interaction. Therefore, it is not only possible, but even common to come to the different conclusions depending on the scale on which a statistical interaction is tested.

From the public health perspective, several authors have argued that under assumption that benefits, or costs, of certain factors are measured by excess, or reduction, in incident numbers (i.e., case-load per unit population), additive measures are more reliable than multiplicative measures to increase a net benefit by targeting the proper subpopulation.[13,22] The main reasoning behind was that if an excess effect

produced by each factor is nonadditive, a public health impact can only be predicted if the levels of all factor are known.[23,24]

Another important point is that both interaction measures can be considerably affected by falsely negative results, i.e., a type 2 error. This is because studies are usually only powered to show the significant differences in the total cohort and not in the subgroups.[3] In this context, obtaining significant p-values may be even more difficult when testing departure from additivity than from multiplicity of effects.

Taken together with previous reports,[9,16] we strongly advise the clinical investigators to report both additive and multiplicative interaction measures with corresponding 95% confidence interval (CI).

Additive interaction measures derived from multiplicative statistical models

Although statistical models such as logistic regression and Cox regression models operate on a multiplicative scale, additive interaction measures can still be calculated (Box 2). The following formulae apply for all ratio-measures (RR, OR, HR) equally.[16,25,26]

Relative Excess Risk due to Interaction (RERI)

The RERI (synonym: interaction contrast ratio [ICR]) is the difference between joint relative effect of two factors and their relative effects considered separately (Table 1: equations-3 and equations-4).[13] Although RERI is an additive interaction measure, it differs from the AERI because it operates with ratios instead of absolute risks. However, when only ratio-measures are given, the RERI can be used to determine additive interaction effect. For example, Jorgensen et al. reported that the 30-day risk of major adverse cardiovascular events (MACE) was associated with long-term use of β-blockers in patients with uncomplicated hypertension undergoing non-cardiac surgery.[27] They also found a super-multiplicative interaction between β-blocker use and diabetes. To quantify this interaction on an additive scale, we calculate the RERI using equation-3 as 2.20 –1.47 – 0.94 + 1.00 = 0.79 (supplemental text). The RERI indicated a super-additive interaction between β-blocker and diabetes (RERI >0). The 95%CI for RERI can be calculated using the delta method[28] or using the first percentile Bootstrap method which covers 95%CI better than the delta method[29] and is more suitable for continuous factors.[20] An interpretation of RERI may be sometimes less straightforward if additional covariates are included in the model because it varies across the levels defined by additional covariates.[30] The codes for calculating RERI with 95%CI are available in SAS[12,25,31], STATA[12], R[32,33], or using excel sheets.[9,20]

Relative Excess Risk due to Interaction (RERI)

- Use RR, OR, and HR
- RERI > 0, super-additive statistical interaction
- RERI < 0, sub-additive statistical interaction
- Depends on additional covariates adjustment

Attributable proportion due to interaction (AP)

- Use RR, OR, and HR
- Same direction as RERI
- Proportion of the outcome in double exposed group that is due to the interaction itself
- Depends on additional covariates adjustment

Modified AP*

- Use RR, OR, and HR
- Same direction as RERI
- Proportion of the joint effect of both exposures that is due to the interaction itself
- Does not depend on additional covariates adjustment

Synergy (S)-index

- Use RR, OR, and HR
- S-index > 1, super-additive statistical interaction
- S-index < 1, sub-additive statistical interaction
- Does not depend on additional covariates adjustment
- Interpretation is difficult if one or both factors are preventive

BOX 2 Additive interaction measures derived from the multiplicative (log-linear, logistic, Cox regression) models. RR, risk ratio; OR, odds ratio; HR, hazard ratio.

Attributable proportion due to interaction (AP)

The attributable proportion for the outcome, denoted here by AP, indicates the proportion of the outcome in double exposed group that is due to the interaction itself.[34] It is derived from RERI (Table 1: equation-5 and equation-6). Following the above example by Jorgensen[27], we calculate AP using equation-5 as 0.79 / 2.2 = 0.36 indicating that 36% of MACE in patients with diabetes and on β-blockers is due to the interaction itself. Similar to RERI, AP varies if additional covariates are included into the model. The codes for calculating AP with 95%CI are available in SAS[12,25,31], R[32], or using excel sheets.[9,20]

Alternatively, the attributable proportion for the effects, denote here by AP*, can be calculated which represents the proportion of the joint effect of both exposures that is due to the interaction itself (Table 1: equations-7 and equations-8).[34] In the same example[27], AP* can be calculated using equation-7 as 0.79 / (2.2 – 1) = 0.66 which indicates that 66% of joint effect of diabetes and β-blockers use is due to the interaction itself. Notably, AP* is independent of covariates adjustment.[34] The codes for calculating AP* with 95%CI are available in SAS[35], STATA[35], and R.[32,33]

TABLE 1 Additive measures of statistical interaction.

A. From additive statistical models:	**Eq. n.**
Absolute excess risk due to interaction (AERI) (using stratification)	
Formula: $AERI = R_{E+,M+} + R_{E-,M-} - R_{E+,M-} - R_{E-,M+}$ Description: E, the exposure (i.e., primary factor); M, a modifier (i.e., secondary factor); $R_{E+,M+}$, the risk in the patients who are exposed to both factors; $R_{E-,M-}$, the risk in the patients in whom both factors are absent; $R_{E+,M-}$, the risk in the patients who are exposed only to the primary factor; $R_{E-,M+}$, the risk in the patients who are exposed only to the secondary factor.	(1)
Linear regression model (using a cross-product term)	
Formula: $Y\ (continuos) = \beta_0 + \beta_1(E) + \beta_2(M) + \beta_3(ExM)$ Description: β_0, average Y in patients in whom both factors are absent (E–, M–); β_1, average difference in Y between the patients who are exposed only to the primary factor (E+, M–) and those in whom both factors are absent (E–, M–); β_2, average difference in Y between the patients who are exposed only to the secondary factor (E–, M+) and those in whom both factors are absent (E–, M–); $\beta_1+\beta_2+\beta_3$, average difference in Y between the patients in whom both factors are present (E+, M+) and those in whom both factors are absent (E–, M–); β_3, a coefficient for the cross-product term that represents an additive interaction measure.	(2)

continued

B. From multiplicative statistical models:	**Eq. n.**
Relative excess risk due to interaction (RERI)	
Formulae (can be used for RR, OR, HR equally): $RERI_{RR} = RR_{E+,M+} - RR_{E+,M-} - RR_{E-,M+} + 1$ (using stratification) $RERI_{OR} = OR_E \times OR_M \times OR_{ExM} - OR_E - OR_M + 1$ (using a cross-product term) Description: $OR_E \times OR_M \times OR_{ExM}$ equals to $OR_{E+,M+}$. Note: $OR_{E+,M+}$ is not provided in the output of the regression models using a cross-product term. The RERI is the difference between joint relative effect of two factors and their effects considered separately.	(3) (4)
Attributable proportion due to interaction (AP)	
Formulae (can be used for RR, OR, HR equally): $AP = RERI_{RR} / RR_{E+,M+}$ (using stratification) $AP = RERI_{OR} / (OR_E \times OR_M \times OR_{ExM})$ (using a cross-product term) Description: The AP is the proportion of the outcome in double exposed group that is due to the interaction itself.	(5) (6)
Modified attributable proportion due to interaction (AP*)	
Formulae (can be used for RR, OR, HR equally): $AP^* = RERI_{RR} / (RR_{E+,M+} - 1)$ (using stratification) $AP^* = RERI_{OR} / (OR_E \times OR_M \times OR_{ExM} - 1)$ (using a cross-product term) Description: The AP* represents the proportion of the effect of both exposures due to the interaction itself.	(7) (8)
Synergy (S)-index	
Formulae (can be used for RR, OR, HR equally): $S = (RR_{E+,M+} - 1) / [(RR_{E+,M-} - 1) + (RR_{E-,M+} - 1)]$ (using stratification) $S = (OR_E \times OR_M \times OR_{ExM} - 1) / [(OR_E - 1) + (OR_M - 1)]$ (using a cross-product term) Description: The S-index is the extent to which joint relative effect of two factors together exceed 1, and whether this exceeding is greater than the sum of relative effects of two factors separately exceed 1.	(9) (10)

Eg. n., equation number.

Synergy index

The S-index reflects the extent to which the joint relative effect of two factors together exceed 1, and whether this exceeding is greater than the sum of relative effects of two factors separately exceed 1 (Table 1: equation-9 and equation-10). For example, Andrews et al. studied the effect of an early resuscitation protocol on the in-hospital mortality in septic patients with hypotension.[36] They found that the use

of early resuscitation protocol increased the in-hospital mortality which was more pronounced in patients with Glasgov coma scale score (GCS) 13-15 than in those with score 3-12. The S-index can be calculated using equaiton-9 as (3.55 – 1) / (3.09 – 1 + 1.91 – 1) = 0.85 indicating a sub-additive interaction between the treatment protocol and worse GSC score because S-index <1 (supplemental text). Notably, the S-index is in independent of covariates adjustment.[30] However, the interpretation may be difficult if one of the factors are preventive rather than causative, i.e., when denominator of S-index is negative.[37] The codes for calculating S-index with 95%CI are available in SAS[12,25,31], R[33], or using excel sheets.[9,20]

TABLE 2 Multiplicative measures of statistical interaction.

Relative risk ratio due to interaction (using stratification)	**Eq. n.**
Formulae (can be used for RR, OR, HR equally): $RR_{E+,M+} / (RR_{E+,M-} \times RR_{E-,M+})$ Description: $RR_{E+,M+} / (RR_{E+,M-} \times RR_{E-,M+})$ equals to the relative risk of a product term in a regression model	(11)
Logistic regression model (using a cross-product term)	
Formula: $Ln[Pr_{Y=1}/(1 - Pr_{Y=1})] = \beta_0 + \beta_1(E) + \beta_2(M) + \beta_3(ExM)$ (exponentiation of both sides of equation to eliminate logarithm) $Pr_{Y=1} / (1 - Pr_{Y=1}) = e^{\beta 0} \times e^{\beta 1(E)} \times e^{\beta 2(M)} \times e^{\beta 3(ExM)}$ (this can also be rewritten as) $Odds = O_0 \times OR_E \times OR_M \times OR_{ExM}$ Description: O_0, odds of Y=1 (e.g., a patient dies) in patients in whom both factors are absent (E–, M–) i.e., this is a background risk because odds of outcome are determined by factors other than E and M; OR_E, odds ratio between the patients who are exposed only to the primary factor (E+, M–) and those in whom both factors are absent (E–, M–); OR_M, odds ratio between the patients who are exposed only to the secondary factor (E–, M+) and those in whom both factors are absent (E–, M–); $OR_E \times OR_M \times OR_{ExM}$, odds ratio between the patients who are exposed to both factors together (E+, M+) and those in whom both factors are absent (E–, M–); OR_{ExM}, OR for the cross-product term, that represents a multiplicative interaction measure.	(12)
Cox regression model (using a cross-product term)	
Formula: $Ln[H(t)] = \beta_0 + \beta_1(E) + \beta_2(M) + \beta_3(ExM)$ $H(t) = e^{\beta 0} \times e^{\beta 1(E)} \times e^{\beta 2(M)} \times e^{\beta 3(ExM)}$ (this can also be rewritten as) $H(t) = H_0(t) \times HR_E \times HR_M \times HR_{ExM}$ Description: The same description as for logistic model, but hazard ratio are used instead of odds ratio. HR_{E+xM+}, HR for the product term that represents a multiplicative interaction measure.	(13)

Eg. n., equation number.

Multiple testing

Multiple testing is common problem when testing statistical interactions because different data, hypotheses, and analyses are assessed simultaneously. Figure 2 outlines four steps that should be considered to reduce the probability of the false positive results, i.e., type 1 error. In hypothesis-generating studies, some authors suggest that no adjustments of the p-value are required.[38] In hypothesis-confirmatory studies, an adjustment for multiple testing should be done as these studies often lead to policy-making. To date, several methods exist to address multiple testing and are described elsewhere.[38] Finally, a multiple testing represents another reason why forming conclusions solely based on the p-value of an interaction test is unjustified.

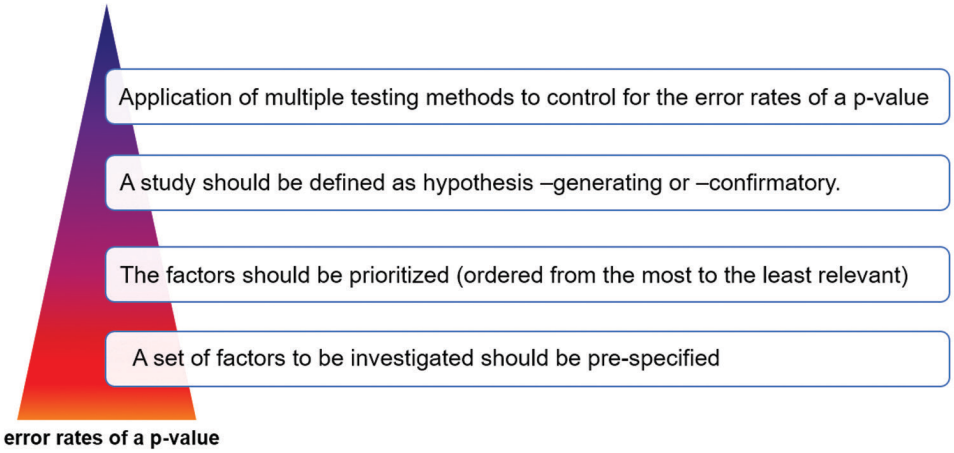


FIGURE 2 Four steps to be considered to reduce probability of having a significant interaction only as a result of chance findings.

Sample size calculation

Sample size calculation should be considered if an investigator is planning to analyze statistical interaction, and especially if an important subgroup analysis is expected to be performed. This helps defining the rule for stopping a trial in order for an adequate number of patients is recruited for each subgroup. For this purpose, a number of software programs[39] and excel sheets are available both for additive[40] and multiplicative[41] interaction measures and various study designs.[42]

REPORTING OF STATISTICAL INTERACTION

To make the results of an interaction analysis more reliable, an investigator should report all relevant information regarding the analysis which are discussed below.

The goal of statistical interaction analysis and set of confounders based on that goal (methods)

An important question that should be answered firstly is why statistical interaction is tested. For example, is the aim to find the subgroup of patients based on their baseline characteristics where the treatment has the greatest effect, or is intervening on those characteristics also considered? This is important to state because different set of confounders should be then chosen to control for the bias.

If effect-measure modification is investigated, only confounding of the primary factor on an outcome should be controlled for. In RCTs, this confounding of the treatment's effect is already addressed by randomization. Yet, one may still want to control for confounding in order to eliminate the possible imbalances between the subgroups that may occur despite the randomization.[17] However, if causal interaction is investigated, then confounding for the effects of both factors on an outcome must be controlled for.[17]

The origin of statistical interaction analysis (methods)

Based on the origin, statistical interaction can be classified as being either prespecified or post-hoc. The prespecified analysis[6] (synonyms: "a priory", "preplanned", "planned", "previously suggested") is considered if the analysis is specified before data are obtained. This specification includes: 1) factors that are considered for analysis, 2) outcomes that are considered for analysis, and 3) set of confounders. An investigator may also consider an attempt of corroboration, i.e., a subsequent study with the same analysis as reported previously (for the same strata, interventions, outcomes, and study population) as the prespecified analysis.[6]

The post-hoc analysis[6] (synonyms: "non-prespecified", "secondary", "explanatory", "preliminary") is considered in all other situations. Of note is that post-hoc analyses are usually data-driven and may be motivated with overall null findings.[43] In this case, one could aim to systematically assess all possible statistical interactions in order to reduce a chance of spurious results.[44] Nonetheless, the post-hoc analyses should be considered solely for exploratory purposes.

The results of statistical interaction analysis (results)

For effect-measure modification between two categorical factors, the results should include 1) effects per each stratum of both factors using a single reference category that should be a subgroup with the lowest risk, 2) effects of the primary factor in strata of the secondary factor, 3) effect per each multivariable adjusted models; 4) additive and multiplicative interaction measures with 95%CI; 5) the set of confounders for the primary factor–outcome relationship (template table 2). For causal interaction, the results should also include 6) effects of the secondary factor in strata of the primary factor; 7) the set of confounders for the secondary factor–outcome relationship (template table 2).

If one of the factors is continuous, a 2 x 2 table cannot be constructed and the results should be reported as in the template table 3. For easier interpretation, it is advisable to present the results using figures, which may also be helpful if more than two factors are tested. How these figures can be made in R is described elsewhere.[45] Alternatively, a continuous variable can be dichotomized and reported as in the template tables 1 and 2.

CONCLUSION

This article outlines the challenges associated with assessment and reporting of statistical interactions in clinical studies, as well as the recommendations that, if adhered to, could increase the clarity and the completeness of future studies. In the present article, we have discussed the importance of the distinction between effect-measure modification and causal interaction, their qualitative and quantitative forms, the importance of a measurement scale on which interactions are tested, additive and multiplicative interaction measures, the relevance of multiple testing as well as the origin of interaction analysis (i.e., whether is prespecified or post-hoc). In addition, we have summarized the information on publicly available SAS, STATA, and R codes, as well as the excel sheets, which can freely be used to calculate different interaction measures. Likewise, we have provided the templates to report obtained results. Altogether, we believe that this article will help in making the results of statistical interaction more reliable, and facilitate clinical studies to use this type of analysis even more in the future.

REFERENCES:

1. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. *The New England journal of medicine*. 2007;357(21):2189-2194.
2. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet (London, England)*. 2005;365(9454):176-186.
3. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet (London, England)*. 2000;355(9209):1064-1069.
4. Hernandez AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *American heart journal*. 2006;151(2):257-264.
5. Knol MJ, Egger M, Scott P, Geerlings MI, Vandenbroucke JP. When one depends on the other: reporting of interaction in case-control and cohort studies. *Epidemiology (Cambridge, Mass)*. 2009;20(2):161-166.
6. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JP. Evaluation of Evidence of Statistical Support and Corroboration of Subgroup Claims in Randomized Clinical Trials. *JAMA internal medicine*. 2017;177(4):554-560.
7. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21(19):2917-2930.
8. Lagakos SW. The challenge of subgroup analyses--reporting without distorting. *The New England journal of medicine*. 2006;354(16):1667-1669.
9. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *International journal of epidemiology*. 2012;41(2):514-520.
10. Boffetta P, Winn DM, Ioannidis JP, et al. Recommendations and proposed guidelines for assessing the cumulative evidence on joint effects of genes and environments on cancer occurrence in humans. *International journal of epidemiology*. 2012;41(3):686-704.
11. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology (Cambridge, Mass)*. 2009;20(6):863-871.
12. VanderWeele TJKMJ. A tutorial on interaction. *Epidemiologic methods*. 2014;3:33-72.
13. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology. 3rd ed*. Philadelphia: Lippincott Williams & Wilkins; 2008.
14. Szklo M, Nieto FJ. *Epidemiology. Beyond the Basics. 3rd ed*. Burlington: Jones & Barlett Learning; 2014.
15. Vander Weele TJ. Confounding and effect modification: distribution and measure. *Epidemiologic methods*. 2012;1(1):55-82.
16. VanderWeele T, Knol MJ. A tutorial on interaction. *Epidemiologic methods*. 2014;3:33-72.
17. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med*. 2011;154(10):680-683.
18. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*. 1985;41(2):361-372.
19. Brown MJ. Matching the right drug to the right patient in essential hypertension. *Heart (British Cardiac Society)*. 2001;86(1):113-120.
20. Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *International journal of epidemiology*. 2007;36(5):1111-1118.
21. Weiner DE, Tabatabai S, Tighiouart H, et al. Cardiovascular outcomes and all-cause mortality: exploring the interaction between CKD and cardiovascular disease. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2006;48(3):392-401.
22. Greenland S. Interactions in epidemiology: relevance, identification, and estimation. *Epidemiology (Cambridge, Mass)*. 2009;20(1):14-17.
23. Hoffmann K, Heidemann C, Weikert C, Schulze MB, Boeing H. Estimating the proportion of disease due to classes of sufficient causes. *American journal of epidemiology*. 2006;163(1):76-83.
24. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *American journal of*

epidemiology. 1980;112(4):467-470.

25. Li R, Chambless L. Test for additive interaction in proportional hazards models. *Annals of epidemiology*. 2007;17(3):227-236.
26. VanderWeele TJ. Causal interactions in the proportional hazards model. *Epidemiology (Cambridge, Mass)*. 2011;22(5):713-717.
27. Jorgensen ME, Hlatky MA, Kober L, et al. beta-Blocker-Associated Risks in Patients With Uncomplicated Hypertension Undergoing Noncardiac Surgery. *JAMA internal medicine*. 2015;175(12):1923-1931.
28. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology (Cambridge, Mass)*. 1992;3(5):452-456.
29. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. *Epidemiology (Cambridge, Mass)*. 1996;7(3):286-290.
30. Skrondal A. Interaction as departure from additivity in case-control studies: a cautionary note. *American journal of epidemiology*. 2003;158(3):251-258.
31. Lundberg M, Fredlund P, Hallqvist J, Diderichsen F. A SAS program calculating three measures of interaction with confidence intervals. *Epidemiology (Cambridge, Mass)*. 1996;7(6):655-656.
32. Mathur MB, VanderWeele TJ. R Function for Additive Interaction Measures. *Epidemiology (Cambridge, Mass)*. 2018;29(1):e5-e6.
33. Mark Stevenson with contributions from Telmo Nunes CH, Jonathon Marshall, Javier Sanchez, Ron Thornton, Jeno, Reiczigel JR-C, Paola Sebastiani, Peter Solymos, Kazuki Yoshida, Geoff Jones, Sarah Pirikahu, Simon, Firestone RK, Johann Popp and Mathew Jay. (2017). epiR: Tools for the Analysis of Epidemiological Data. R, https://CRAN.R-project.org/package=epiR pv-.
34. VanderWeele TJ. Reconsidering the denominator of the attributable proportion for interaction. *European journal of epidemiology*. 2013;28(10):779-784.
35. VanderWeele TJ, Tchetgen Tchetgen EJ. Attributing effects to interactions. *Epidemiology (Cambridge, Mass)*. 2014;25(5):711-722.
36. Andrews B, Semler MW, Muchemwa L, et al. Effect of an Early Resuscitation Protocol on In-hospital Mortality Among Adults With Sepsis and Hypotension: A Randomized Clinical Trial. *Jama*. 2017;318(13):1233-1240.
37. Knol MJ, VanderWeele TJ, Groenwold RH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. *European journal of epidemiology*. 2011;26(6):433-438.
38. Bender R, Lange S. Adjusting for multiple testing--when and how? *Journal of clinical epidemiology*. 2001;54(4):343-349.
39. http://biostats.usc.edu/Quanto.html.
40. VanderWeele TJ. Sample Size and Power Calculations for Additive Interactions. *Epidemiologic methods*. 2012;1(1):159-188.
41. Demidenko E. Sample size and optimal design for logistic regression with binary interaction. *Stat Med*. 2008;27(1):36-46.
42. Leon AC, Heo M. Sample Sizes Required to Detect Interactions between Two Binary Fixed-Effects in a Mixed-Effects Linear Regression Model. *Computational statistics & data analysis*. 2009;53(3):603-608.
43. Sun X, Briel M, Busse JW, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. *Bmj*. 2011;342:d1569.
44. Patel CJ, Chen R, Kodama K, Ioannidis JP, Butte AJ. Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. *Human genetics*. 2013;132(5):495-508.
45. Fox J. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*. 2003;8(15):1-9.

SUPPLEMENTARY INFORMATION

Example 1: Calculation of the absolute excess risk due to interaction (AERI) in the section entitled Additive interaction measures.

Weiner et al. studied the effects of chronic kidney disease (CKD) and cardiovascular disease (CVD) on the 10-year risk of the composite endpoint including cardiovascular and all-cause death.[21] Using numbers provided in the Tables 1 and 2 of their article, we can calculate absolute cumulative 10-year risk per each subgroup as shown in the table below:

Absolute risks		**Cardiovascular disease (CVD)***	
		No	Yes
Chronic kidney disease (CKD)*	No	15% (3053/20970)	38% (1344/3519)
	Yes	34% (565/1664)	66% (501/759)

*In parenthesis is shown the number of patients with event divided by the total number of patients in the corresponding subgroup.

To calculate AERI we will use an equation-1: AERI = $R_{CVD+,CKD+}$ + $R_{CVD-,CKD-}$ – $R_{CVD+,\ CKD-}$ – $R_{CVD-,CKD+}$ and calculate as 66% +15% - 38% - 34% = 9%. The AERI indicates a super-additive interaction between CKD and CVD because AERI > 0, but also shows an absolute excess risk of 9% due to the interaction itself.

Calculation of the ratio of RRs in the section entitled Multiplicative interaction measures.

In the same study by Weiner et al.[21] we can further calculate relative risk ratio due to interaction as shown in the table below. In the following table, relative risks are calculated by dividing the absolute risks per each subgroup with the risk in the subgroup of patients without CVD or CKD, i.e, subgroup with the lowest absolute risk.

Relative risks		**Cardiovascular disease (CVD)***	
		No	Yes
Chronic kidney disease (CKD)*	No	1.0 (15%/15%)	2.5 (38%/15%)
	Yes	2.3 (34%/15%)	4.4 (66%/15%)

*In parenthesis is shown relative risk which is calculated by dividing the absolute risk in the subgroup with the risk in the subgroup of patients without CVD or CKD, i.e, subgroup with the lowest absolute risk.

To calculate ratio of risk ratios we will use equation-11: $RR_{CVD+,CKD+}$ / ($RR_{CVD+,\ CKD-}$ x $RR_{CVD-,CKD+}$) and calculate as 4.4 / (2.5 x 2.3) = 0.8. This indicates a sub-

multiplicative interaction between CKD and CVD because the ratio of RRs < 1, but also shows relative risk ratio due to interaction of -20%.

Example 2. Calculation of the relative excess risk due to interaction (RERI) in the section entitled Relative Excess Risk due to Interaction (RERI).

Jorgensen et al. studied the effect of the long-term β-blockers use on the the 30-day risk of major adverse cardiovascular events (MACE) in patients with uncomplicated hypertension undergoing non-cardiac surgery.[27] Authors found a multiplicative interaction between the long-term use of β-blockers and diabetes on the 30-day risk of MACE. Using numbers provided in Figure 3 of their article, we can calculate both AERI and RERI of the aforementioned interaction as shown in the table below:

Absolute risks		**Diabetes (DM)***	
		No	Yes
β-blockers use*	No	0.85% (294/34691)	0.80% (48/5985)
	Yes	1.25% (164/13096)	1.87% (29/1548)

*In parenthesis is shown the number of patients with event divided by the total number of patients in the corresponding subgroup.

To calculate AERI we will use an equation-1: AERI = $R_{\beta\text{-blockers+,DM+}} + R_{\beta\text{-blockers–,DM–}} – R_{\beta\text{-blockers+, DM–}} – R_{\beta\text{-blockers–,DM+}}$ and calculate as 1.87% + 0.85% - 1.25% - 0.80% = 0.67%. The AERI indicates a super-additive interaction between the long-term use of β-blockers and diabetes with an absolute excess risk of 0.67% due to the interaction itself.

To calculate RERI we will use an equation-3: $RERI_{RR} = RR_{\beta\text{-blockers+,DM+}} – RR_{\beta\text{-blockers+, DM–}} – RR_{\beta\text{-blockers–,DM+}} + 1$. Furthermore, to obtain the relative risks we will divide the absolute risks per each subgroup by 0.85% which is the risk in patients who did not take β-blockers and did not have diabetes. Thus, the calculation is as follows $RERI_{RR}$ = 1.87% / 0.85% - 1.25% / 0.85% - 0.80% / 0.85% +1 = 0.79 indicating a super-additive interaction because $RERI_{RR} > 0$. Note that, although both AERI and RERI shows additivity of interaction, they are not the same (0.67≠0.79). This is because AERI operates with on a risk-difference scale and relative risk-difference scale.

Calculation of Attributable proportion due to interaction (AP) and modified AP* in the section entitled Attributable proportion due to interaction (AP).

In the same study by Jorgensen et al.[27] we can extend our investigation by calculating the attributable proportions of the outcome and of the joint effect that are due to the interaction itself. For former calculation, we will use an equation-5: AP =

$RERI_{RR}$ / $RR_{\beta\text{-blockers+,DM+}}$ and calculate as 0.79 / (1.87% / 0.85%) = 0.36 indicating that 36% of the 30-day risk of MACE is due to the interaction itself. For latter calculation, we will use an equation-7: AP* = $RERI_{RR}$ / ($RR_{\beta\text{-blockers+,DM+}}$ - 1) and calculate as 0.79 / (1.87% / 0.85% -1) = 0.66 indicating that 66% of the joint effect of β-blockers and diabetes is due to interaction itself.

Example 3. Calculation synergy index (S-index) in the section entitled Synergy index.

Andrews et al. studied the effect of an early resuscitation protocol on the in-hospital mortality in septic patients with hypotension.[36] Authors found a multiplicative interaction between the intervention and patients baseline Glasgov coma scale score (GCS). In their article, GSC score was tested as ordinal variable with three categories ≥13, 12-9, and 8-3. Considering that effect in the latter two categories were similar we dichotomized GSC score into ≥13 and 12-3. Using numbers provided in Figure 3 of their article, we can calculate S-index as shown in the table below:

Absolute risks		**GSC score***	
		≥ 13	3-12
Treatment*	usual care	22% (17 / 78)	68% (15 / 22)[a]
	early resuscitation protocol	42% (36 / 86)	78% (14 / 18)[b]

*In parenthesis is shown the number of patients who died divided by the total number of patients in the corresponding subgroup.
[a] These numbers are obtained after combining categories GSC score 3-8 and 12-9 into one category, GSC score 3-12. Thus, the number of patients treated with usual care who died is 10 + 5 = 15, and the total number of patients treated with usual care is 17 + 5 = 22.
[b] These numbers are obtained after combining categories GSC score 3-8 and 12-9 into one category, GSC score 3-12. Thus, the number of patients treated with the early resuscitation protocol who died is 4 + 10 = 14, and the total number of patients treated with the early resuscitation protocol is 7 + 11 = 18.

To calculate S-index we will use an equation-9: $S = (RR_{\text{protocol, GSC 3-12}} - 1) / [(RR_{\text{protocol, GSC ≥13}} - 1) + (RR_{\text{usual care, GSC 3-12}} - 1)]$. Furthermore, to obtain the relative risks we will divide the absolute risks per each subgroup by 22% which is the risk in patients who received usual care and had GSC score ≥13. Thus, the calculation is as follows S = (78% / 22%) / (42%/22% – 1 + 68% / 22% – 1) = 0.85 indicating a sub-additive interaction because S-index < 1.

TEMPLATE TABLE 1 Reporting the effect-measure modification analysis from multiplicative (logistic, Cox regression) models for two categorical factors.

	1st factor = 0			1st factor = 1			
	2nd factor = 0	2nd factor = 1		2nd factor = 0		2nd factor = 1	
Models adjustment	$n_{o.}/n_{pts.}$	$n_{o.}/n_{pts.}$		$n_{o.}/n_{pts.}$		$n_{o.}/n_{pts.}$	
Model 1	1 (reference)	RR(95%CI)	p-value	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 2	1 (reference)	RR(95%CI)	p-value	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 3	1 (reference)	RR(95%CI)	p-value	RR(95%CI)	p-value	RR(95%CI)	p-value

	1st factor = 1			
	2nd factor = 0		2nd factor = 1	
Models adjustment	$n_{o.}/n_{pts}$		$n_{o.}/n_{pts}$	
Model 1	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 2	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 3	RR(95%CI)	p-value	RR(95%CI)	p-value

	Effect modification: 1st factor x 2nd factor	
Models adjustment	**Additive measures**	**Multiplicative measures**
Model 1	RERI (95%CI) p-value	$RR_{1,1}$ / ($RR_{1,0}$ x $RR_{0,1}$) (95%CI) p-value
Model 2	RERI (95%CI) p-value	$RR_{1,1}$ / ($RR_{1,0}$ x $RR_{0,1}$) (95%CI) p-value
Model 3	RERI (95%CI) p-value	$RR_{1,1}$ / ($RR_{1,0}$ x $RR_{0,1}$) (95%CI) p-value

RR, risk ratio; 1st factor, the primary factor (i.e., exposure or intervention); 2nd factor, the secondary factor (i.e., exposure or intervention); RERI, relative excess risk due to interaction; 95%CI, 95% confidence interval; $n_{o.}/n_{pts.}$, number of outcomes / number of patients.
List of confounders for model 1, 2, and 3 should be noted in the footnote of the table. The RR can be replaced with odds ratio (OR) (logistic regression) or hazard ratio (HR) (Cox regression) depending on the model applied. Instead of RERI, other measures of additive effect modification can be used, such as attributable proportion (AP or modified AP), Synergy (S)-index, or their combination. The template provides example for three multivariable adjusted models, but if there are more than three models, additional rows can be added. The template provides example for 2x2 factors; if a factor has more than two subgroups additional columns can be added.*

TEMPLATE TABLE 2 Reporting the causal interaction analysis from multiplicative (logistic, Cox regression) models for two categorical factors.

	1st factor = 0			1st factor = 1		
	2nd factor = 0	2nd factor = 1		2nd factor = 0	2nd factor = 1	
Models adjustment	$n_{o.}/n_{pts.}$	$n_{o.}/n_{pts.}$		$n_{o.}/n_{pts.}$	$n_{o.}/n_{pts.}$	
Model 1	1 (reference)	RR(95%CI)	p-value	RR(95%CI) p-value	RR(95%CI)	p-value
Model 2	1 (reference)	RR(95%CI)	p-value	RR(95%CI) p-value	RR(95%CI)	p-value
Model 3	1 (reference)	RR(95%CI)	p-value	RR(95%CI) p-value	RR(95%CI)	p-value

1st factor = 1				
	2nd factor = 0		2nd factor = 1	
Models adjustment	$n_{o.}/n_{pts.}$		$n_{o.}/n_{pts.}$	
Model 1	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 2	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 3	RR(95%CI)	p-value	RR(95%CI)	p-value

2nd factor = 1				
	1st factor = 0		1st factor = 1	
Models adjustment	$n_{o.}/n_{pts.}$		$n_{o.}/n_{pts.}$	
Model 1	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 2	RR(95%CI)	p-value	RR(95%CI)	p-value
Model 3	RR(95%CI)	p-value	RR(95%CI)	p-value

Interaction: 1st factor x 2nd factor		
Models adjustment	**Additive measures**	**Multiplicative measures**
Model 1	RERI (95%CI) p-value	$RR_{1,1}$ / ($RR_{1,0}$ x $RR_{0,1}$) (95%CI) p-value
Model 2	RERI (95%CI) p-value	$RR_{1,1}$ / ($RR_{1,0}$ x $RR_{0,1}$) (95%CI) p-value
Model 2	RERI (95%CI) p-value	$RR_{1,1}$ / ($RR_{1,0}$ x $RR_{0,1}$) (95%CI) p-value

RR, risk ratio; 1st factor, the primary factor (i.e., exposure or intervention); 2nd factor, the secondary factor (i.e., exposure or intervention); RERI, relative excess risk due to interaction; 95%CI, 95% confidence interval; $n_{o.}/n_{pts.}$, number of outcomes / number of patients.

List of confounders for model 1, 2, and 3 should be noted in the footnote of the table. The RR can be replaced with odds ratio (OR) (logistic regression) or hazard ratio (HR) (Cox regression) depending on the model applied. Instead of RERI, other measures of additive effect modification can be used, such as attributable proportion (AP or modified AP), Synergy (S)-index, or their combination. The template provides example for three multivariable adjusted models, but if there are more than three models, additional rows can be added. The template provides example for 2x2 factors; if a factor has more than two subgroups additional columns can be added.*

TEMPLATE TABLE 3 Reporting the statistical interaction analysis from multiplicative (logistic, Cox regression) models if one or both factors are continuous.

Models adjustment	**1st factor**		**2nd factor**		**1st factor x 2nd factor**	
Model 1	OR(95%CI)	p-value	OR(95%CI)	p-value	OR(95%CI)	p-value
Model 2	OR(95%CI)	p-value	OR(95%CI)	p-value	OR(95%CI)	p-value
Model 3	OR(95%CI)	p-value	OR(95%CI)	p-value	OR(95%CI)	p-value

Models adjustment	**Statistical interaction: 1st factor x 2nd factor**
Model 1	RERI (95%CI) p-value
Model 2	RERI (95%CI) p-value
Model 3	RERI (95%CI) p-value

OR, odds ratio; 1st factor, the primary factor (i.e., exposure or intervention); 2nd factor, the secondary factor (i.e., exposure or intervention); RERI, relative excess risk due to interaction; 95%CI, 95% confidence interval; no. / npts., number of outcomes / number of patients.
List of confounders for model 1, 2, and 3 should be noted in the footnote of the table. The OR can be replaced with hazard ratio (HR) (Cox regression) depending on the model applied. Instead of RERI, other measures of additive effect modification can be used, such as attributable proportion (AP or modified AP), Synergy (S)-index, or their combination. The template provides example for three multivariable adjusted models, but if there are more than three models, additional rows can be added.*