**Statistical Methods in Epidemiology**

**Lab 3**

**Interactions**

**Introduction and Rationale**

This practical builds upon the theoretical concepts introduced in Lecture 3 (“Interactions and Effect Modification”) and aims to apply them through Stata analyses using epidemiologic data.

While previous sessions focused on independent effects of exposures (e.g. risk ratios or rate ratios), this session introduces **interaction analysis**, exploring whether the effect of one exposure differs depending on the level of another.

Students will learn how to:

* distinguish between **additive** and **multiplicative** interaction models,
* calculate and interpret **Relative Excess Risk due to Interaction (RERI)** and **Synergy Index (S)**,
* and understand the implications of correct coding and reference category selection.

This lab emphasizes both the statistical and biological interpretation of interactions, as presented by Rothman, Knol, and VanderWeele, focusing on how combinations of exposures may produce effects greater (synergistic) or smaller (antagonistic) than expected from their individual contributions.

**Structure and Flow**

The structure mirrors the logical analytic process and theoretical framework of Lecture 3.

| **Section** | **Theoretical correspondence** | **Practical focus** |
| --- | --- | --- |
| **1.1 Data exploration** | Understanding exposures and outcome; identifying suitable variable coding for interaction analysis | Inspect and recode BMI and Mediterranean diet adherence |
| **1.2 Variable creation** | Concepts of joint exposure categories | Generate combined exposure variable (cat\_bmi\_nomd) |
| **1.3 Additive interaction** | Definition and estimation of RERI and S (Rothman 1986) | Estimate RERI and Synergy Index using Stata nlcom |
| **1.4 Multiplicative interaction** | The multiplicative model and deviation testing | Fit models with interaction term; interpret exp(b3) |
| **2.1 Classic interaction analysis** | Comparing additive vs. multiplicative models | Fit logistic regression including main effects and interaction term |
| **2.2–2.3 Incorrect and alternative codings** | Sensitivity to reference category and exposure coding | Demonstrate interpretational pitfalls and reversals |

| **Lecture content** | **Practical component** | **Conceptual goal** |
| --- | --- | --- |
| Slides 1–10: *Concept of interaction and effect modification* | 1.1 Data exploration and variable coding | Understand how combined exposures may modify disease risk |
| Slides 11–22: *Multiplicative model of effects* | 1.5 Multiplicative interaction analysis | Apply logistic regression models including interaction terms and interpret multiplicative effects (exp(b3)) |
| Slides 23–38: *Additive interaction and its indices (RERI, Synergy Index)* | 1.3 Additive Interaction Analysis, 1.4 Synergy Index Calculation | Estimate and interpret measures of additive interaction (RERI, S) |
| Slides 39–48: *Coding and interpretation issues (Knol & VanderWeele, 2012)* | 2.2 Exploring Incorrect Approaches, 2.3 Alternative Coding Analysis | Understand how coding direction and reference category affect interaction interpretation |
| Slides 49–60: *Comparing additive vs. multiplicative scales* | 2.1 Classic Interaction Analysis | Contrast interpretations from additive and multiplicative models |
| Slides 61–70: *Biological vs. statistical interaction and interpretation* | Discussion and interpretation throughout | Integrate statistical results with biological reasoning (synergy vs. antagonism) |

**Objectives**

The objectives of this practical are to introduce statistical methods for analyzing and interpreting interactions between risk factors in epidemiological studies, focusing on both additive and multiplicative interactions. Using a dataset examining hypertension in relation to BMI and Mediterranean diet adherence, students will learn to calculate and interpret key interaction measures (RERI, Synergy index) while understanding the importance of proper methodological approaches and potential pitfalls.

# The Dataset

**File:** hyper

**Description:** This dataset contains data from 328 individuals examining hypertension as the primary outcome of interest (represented by the binary variable 'hyper'). The study investigates the association between hypertension and two key factors: Body Mass Index (BMI) and adherence to Mediterranean diet (categorized in two groups), with particular focus on how these characteristics interact with each other.

**Part I. Additive Interaction (RERI and Synergy Index)**

***1.1. Initial Data Exploration***

**1.** Open and examine the dataset ("hyper").

. use hyper, clear

. desc

**2.** Inspect the variables using descriptive statistics.

. tab bmi

. tab nomd

**3.** Evaluate the coding of variables for interaction analysis.

***1.2. Variable Creation and Coding***

**1.** Generate a new combined variable (cat\_bmi\_nomd) with the following coding:

* 1: both factors absent (BMI = 0 & nomd = 0)
* 2: high BMI only (BMI = 1 & nomd = 0)
* 3: low Mediterranean diet adherence only (BMI = 0 & nomd = 1)
* 4: both factors present (BMI = 1 & nomd = 1)

. gen cat\_bmi\_nomd = 1 if bmi == 0 & nomd == 0 // Neither risk factor

. replace cat\_bmi\_nomd = 2 if bmi == 1 & nomd == 0 // High BMI only

. replace cat\_bmi\_nomd = 3 if bmi == 0 & nomd == 1 // Low adherence only

. replace cat\_bmi\_nomd = 4 if bmi == 1 & nomd == 1 // Both risk factors

**2.** Verify the correct coding of the new variable.

. tab bmi nomd

. tab cat\_bmi\_nomd

***1.3. Additive Interaction on the OR Scale (RERI)***

**1.** Fit the logistic regression model using the combined variable.

. logistic hyper i.cat\_bmi\_nomd

**2.** Compute Relative Excess Risk for Interaction (RERI) on the OR scale from the logistic odds ratios.

**3.** Calculate the Relative Excess Risk for Interaction (RERI) using nlcom command.

. nlcom m1\_RERI\_OR: exp(\_b[4.cat\_bmi\_nomd]) ─ exp(\_b[2.cat\_bmi\_nomd]) ─ exp(\_b[3.cat\_bmi\_nomd]) + 1

**4.** Interpret RERI results including clinical implications.

**Note:** The RERI derived from logistic regression reflects additive interaction on the odds ratio scale. It quantifies excess odds due to the joint exposure. Only under the rare-disease assumption (OR ≈ RR) can it be interpreted approximately as excess risk.

***1.4. Synergy Index Calculation (S)***

**1.** Calculate Synergy Index (S) from the previous output.

**2.** Calculate Synergy Index (S) using nlcom command.

. nlcom m1\_ln\_Syn: ln(exp(\_b[4.cat\_bmi\_nomd]) - 1) - ln(exp(\_b[2.cat\_bmi\_nomd]) + exp(\_b[3.cat\_bmi\_nomd]) - 2), post

. nlcom m1\_Syn\_index: exp(ln(exp(\_b[4.cat\_bmi\_nomd]) - 1) - ln(exp(\_b[2.cat\_bmi\_nomd]) + exp(\_b[3.cat\_bmi\_nomd]) - 2))

**3.** Generate 95% confidence intervals using delta method.

. scalar m1\_Syn\_index\_low95 = exp(\_b[m1\_ln\_Syn] - invnormal(0.975) \* \_se[m1\_ln\_Syn])

. scalar m1\_Syn\_index\_high95 = exp(\_b[m1\_ln\_Syn] + invnormal(0.975) \* \_se[m1\_ln\_Syn])

**4.** Create and examine the matrix of Synergy index results.

. mat define Model1\_Synergy\_index = (m1\_Syn\_index, m1\_Syn\_index\_low95, m1\_Syn\_index\_high95)

. mat rown Model1\_Synergy\_index = Syn\_index

. mat coln Model1\_Synergy\_index = S\_index S\_low95 S\_high95

. mat list Model1\_Synergy\_index

**5.** Interpret S results including clinical implications.

**Part II. Multiplicative Interaction (Product Term Approach)**

***2.1. Multiplicative Interaction Analysis***

**1.** Calculate the multiplicative interaction estimates using the previous output of logistic regression.

**2.** Calculate the multiplicative interaction estimates using nlcom command.

. nlcom m1\_ln\_mult\_int: \_b[4.cat\_bmi\_nomd] - \_b[2.cat\_bmi\_nomd] - \_b[3.cat\_bmi\_nomd], post

**3.** Generate 95% confidence intervals for multiplicative interactions.

. scalar m1\_mult\_interaction = exp(\_b[m1\_ln\_mult\_int])

. scalar m1\_mult\_interaction\_low95 = exp(\_b[m1\_ln\_mult\_int] - invnormal(0.975) \* \_se[m1\_ln\_mult\_int])

. scalar m1\_mult\_interaction\_high95 = exp(\_b[m1\_ln\_mult\_int] + invnormal(0.975) \* \_se[m1\_ln\_mult\_int])

**4.** Create and examine the matrix of multiplicative interaction results.

. mat define Model1\_mult\_interaction = (m1\_mult\_interaction, m1\_mult\_interaction\_low95, m1\_mult\_interaction\_high95)

. mat rown Model1\_mult\_interaction = mult\_interaction

. mat coln Model1\_mult\_interaction = mult\_interaction m\_int\_low95 m\_int\_high95

. mat list Model1\_mult\_interaction

**5.** Interpret results.

**6.** Compare the results of additive and multiplicative interaction analyses, discuss any differences in direction or significance between the two scales, and comment on which measure is more relevant for public health interpretation.

**Part III. Alternative Approaches**

*This section demonstrates that additive and multiplicative interactions can also be estimated using a single logistic model including both main effects and an interaction term.*

***3.1. Classic Interaction Analysis***

**1.** Generate interaction term (bmi\_nomd) between BMI and Mediterranean diet adherence.

. gen bmi\_nomd = bmi \* nomd

. label variable bmi\_nomd "Interaction: BMI × Low Mediterranean Diet Adherence"

**2.** Fit a logistic regression model including main effects and interaction term.

. logistic hyper bmi nomd bmi\_nomd

**3.** Calculate RERI and Synergy index as before using nlcom command in stata

. nlcom m2\_RERI: exp(\_b[bmi] + \_b[nomd] + \_b[bmi\_nomd]) - exp(\_b[bmi]) - exp(\_b[nomd]) + 1

. nlcom m2\_ln\_Syn: ln(exp(\_b[bmi] + \_b[nomd] + \_b[bmi\_nomd]) - 1) - ln(exp(\_b[bmi]) + exp(\_b[nomd]) - 2), post

**4.** Create and examine the matrix of results.

. scalar m2\_Syn\_index = exp(\_b[m2\_ln\_Syn])

. scalar m2\_Syn\_index\_low95 = exp(\_b[m2\_ln\_Syn] - invnormal(0.975) \* \_se[m2\_ln\_Syn])

. scalar m2\_Syn\_index\_high95 = exp(\_b[m2\_ln\_Syn] + invnormal(0.975) \* \_se[m2\_ln\_Syn])

. mat define Model2\_Synergy\_index = (m2\_Syn\_index, m2\_Syn\_index\_low95, m2\_Syn\_index\_high95)

. mat rown Model2\_Synergy\_index = Syn\_index

. mat coln Model2\_Synergy\_index = S\_index S\_low95 S\_high95

. mat list Model2\_Synergy\_index

**5.** Compare results with those from Part I.

**Part IV. Incorrect Coding and Reference Approaches**

*Proper coding of exposures is crucial for a meaningful interpretation of interaction on additive and multiplicative scales.*

*For the RERI and Synergy Index to be interpretable, it is strongly recommended that both variables are coded in the same risk direction (e.g., higher values corresponding to higher hypothesised risk). Although this is not mathematically required for estimation, inconsistent coding makes the resulting interaction terms difficult or impossible to interpret.*

***4.1 Wrong Reference (Incorrect Approach #1)***

Demonstrate the incorrect approach by changing reference category:

**1.** Set category 2 (bmi > 30kg/m2 and Mediterranean diet adherence) as baseline.

. char cat\_bmi\_nomd [omit] 2

**2.** Fit logistic regression (model 2) with incorrect reference.

. xi: logistic hyper i.cat\_bmi\_nomd

**3.** Calculate and interpret RERI and multiplicative interaction.

. nlcom m1\_err\_RERI: exp(\_b[\_Icat\_bmi\_n\_4]) - exp(\_b[\_Icat\_bmi\_n\_3]) - exp(\_b[\_Icat\_bmi\_n\_1]) + 1

. nlcom m1\_err\_ln\_mult\_int: \_b[\_Icat\_bmi\_n\_4] - \_b[\_Icat\_bmi\_n\_3] - \_b[\_Icat\_bmi\_n\_1], post

. scalar m1\_err\_mult\_interaction = exp(\_b[m1\_err\_ln\_mult\_int])

. scalar m1\_err\_mult\_interaction\_low95 = exp(\_b[m1\_err\_ln\_mult\_int] - invnormal(0.975) \* \_se[m1\_err\_ln\_mult\_int])

. scalar m1\_err\_mult\_interaction\_high95 = exp(\_b[m1\_err\_ln\_mult\_int] + invnormal(0.975) \* \_se[m1\_err\_ln\_mult\_int])

. mat define Model1\_err\_mult\_interaction = (m1\_err\_mult\_interaction, m1\_err\_mult\_interaction\_low95, m1\_err\_mult\_interaction\_high95)

. mat rown Model1\_err\_mult\_interaction = mult\_interaction

. mat coln Model1\_err\_mult\_interaction = mult\_interaction m\_int\_low95 m\_int\_high95

. mat list Model1\_err\_mult\_interaction

**4.** Show why results are problematic.

***4.2. Reversed Coding (Incorrect Approach #2)***

**1.** Demonstrate problems with reversed coding for Mediterranean diet:

* Generate err\_nomd (reverse coding of Mediterranean diet adherence)

. gen err\_nomd = 1 - nomd

* Create new interaction term (err\_bmi\_nomd)

. gen err\_nomd = 1 - nomd

. gen err\_bmi\_nomd = bmi \* err\_nomd

* Fit logistic regression with reversed coding

. logistic hyper bmi err\_nomd err\_bmi\_nomd

. logit hyper bmi err\_nomd err\_bmi\_nomd

. logit hyper bmi nomd bmi\_nomd

**2.** Show why interpretation becomes problematic with one risk and one protective factor.

**Part V. Summary and Key Takeaways**

This lab demonstrated how to assess interaction on additive and multiplicative scales using logistic regression.

Although no statistically significant interaction was detected, the exercise illustrates key methodological principles:

proper coding, correct reference category selection, and clear distinction between additive and multiplicative interpretation.