ISSN: 2755-0192

Journal of Psychiatry Research Reviews & Reports

SCIENTIFIC Research and Community

Review Article

Open Access

Relations- One Way out of the Replication Crisis

Rainer Maderthaner

Department of Cognition, Emotion and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria

Importance: Reproducibility, or the long-term validity of findings, is a precondition for the credibility of scientific results in several scientific disciplines. If different experts were asked how much published work in their field is reproducible, more than fifty percent of researchers in chemistry, physics, biology, medicine, and others said we have a replication crisis. This means that the scientific credibility of many disciplines in the eyes of the public is at risk, with significant consequences for the reputation and funding of science.

Challenges: It is therefore necessary to tackle the causes of the replication crisis, such as Questionable Research Practices (QRP), publication pressure, and weaknesses in the planning and statistical analysis of studies. The latter is the subject of this article, in which it is emphasised that many hypotheses do not correspond in their complexity to the phenomena studied, either in terms of the possible influencing variables or in terms of the measures of association.

Measures: It is suggested that the hypotheses should be more differentiated, take greater account of the presumed effect structure, and the variety of logical relationships in the empirical phenomena. This article uses several examples to show the extent to which more precise hypotheses have an impact on the accuracy of statistically reliable results. One computer program that can be used in the next time for these purposes is Relation Analysis (RELAN), which allows logical analyses, statistical tests, explorations and simulations of relations between variables.

Conclusion: In future, it will be necessary to adapt scientific hypotheses in the biological, human and social sciences more closely to the complexity and the structure of empirical phenomena.

***Corresponding author**
Rainer Maderthaner, Department of Cognition, Emotion and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria.

Received: November 20, 2024; **Accepted:** November 25, 2024; **Published:** December 04, 2024

Keywords: Logical-Statistical Analysis, Hypothesis Testing, Exploration, Modeling, Propositional Logic, Software

Introduction

Many scientific disciplines with a statistical focus (e.g. psychology, medicine, biology, economics, environmental sciences) have been discussing the causes of the so-called replication crisis for about twenty years [1-6]. Many reasons for the replication crisis can be found in the methodological literature [5-7]. I will mention only one of the most prominent factors, namely the Questionable Research Practices (QRP). One particular issue of them is the Null Ritual, whereby in statistical analysis we only focus on the null hypothesis, instead of postulating an alternative hypothesis, which moreover would also permit the calculation of statistical power [5]. This statistic is of significant value as it offers an estimation of the probability of the hypothesis being valid within the population, based on the sample data. Furthermore, Gigerenzer should be referenced, who discovered that approximately forty percent of professors and lecturers exhibited the "replication delusion," which refers to the belief that the probability of the alternative hypothesis is 1 - p (instead of 1 - β) [5]. Another questionable research practice is P-hacking, which involves increasing the sample size until a statistically significant result is obtained, and also HARKing, which stands for "hypothesis formulation after the results are known". A common error is also the repeated testing of different hypotheses on the same set of data (e.g. search in correlation matrices) without correcting the alpha level error. Furthermore, sampling errors, experimental influences and different operationalisation can also give rise to replication problems [7]. All of these reasons can compromise the validity of a study. Ten more possibly causes for the replication crisis are discussed elsewhere [6].

In addition to these certainly important reasons for valid scientific results, it has hardly been discussed that in many cases the empirical phenomena may be more complex than the hypotheses they are designed to explain [8]. Hypotheses should map as sharply as possible the regularities (structure and dynamic) in empirical phenomena. If the hypotheses are too simple, then the regularities in the data are only blurred because they are only represented in part of the data. The expectation that regularities will become increasingly apparent because random variations 'average out' with a larger sample (basic principle of statistical 'truth-finding') is only justified if there are no other superimposing regularities in the data in addition to the regularities sought. The more sharply hypotheses are formulated in terms of regularities, both the sought-after regularities and their overlaps (confounds), the greater the chance of identifying unadulterated structures and dynamics in the data.

Objectives

In my book sufficiently justified, I conclude that we don't ask ourselves enough questions about whether we are correctly mapping empirical phenomena onto hypotheses [6]. Therefore, I postulate the following claims

- Hypotheses about empirical phenomena need to be better adapted to empirical complexity.
- Multiple causes, moderators and mediators need to be

considered in empirical theories.

- Logical connections ("and", "or", "if-then", ...) must also be included in empirical hypotheses.

In order to take these postulations into account in statistical hypotheses, we need to refer to the mathematical concept of relations.

Materials and Method

Relations and Logical Functions

In mathematics the basic definition of a binary 'relation' R of two sets A and B is any subset (set of ordered pairs) of A cross B (Cartesian product) [9]. Relations are the most elementary and exhaustive relationships between sets. If we apply this definition to variables, then every combination of (ordered) variable values – and every set of these combinations – is a (elementary) relation. If the variables are only two-valued (1 = true, 0 = not true), the relations between the variables can be characterised by propositional functions (e.g. AND, OR, IF-THEN). In this way, a great many logical hypotheses (max. 10306) can be unambiguously formalised (Figure 1).

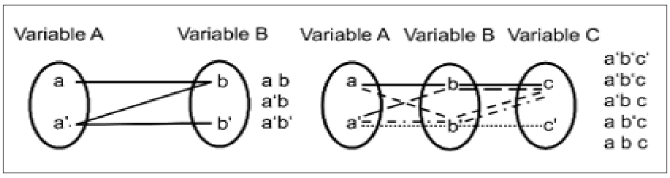

Figure 1: In the Diagram on the Left, the Three Lines Symbolise Three (Elementary) Relations between the Variables A and B and the Three Elementary Relations that can be Summarised in the Logical Formula IF A THEN B (Binary Relation). In the Diagram on the Right, the Five (Elementary) Relations can be Clearly Characterised by the Formula IF (A OR B) THEN C.

The statistical approach of the Relan Analysis (RELAN) is similar to the Prediction Analysis, as it is also based on Boolean algebra [10-13]. The logical functions that may be employed in the hypotheses are as follows: AND (∧), XOR exclusive (v̲), OR inclusive (∨), IF - THEN (→), ONLY IF - THEN (↔), NAND not AND (|). To illustrate, consider the following example from the field of chemistry, expressed in logical terms: IF (2 H2 AND O2 AND Ignition) THEN H2O (Explosion of oxyhydrogen gas with water as a result). It is very important to note that the use of IF - THEN (→) always expresses the idea that other triggers for effects are also taken into consideration, i.e. that multicausality is assumed, whereas the use of ONLY IF - THEN (↔) only asserts mono causality.

With RELAN a logical association analysis for variable pairs can be performed for up to one hundred variables. As will be demonstrated subsequently, an understanding of the logical functions is sufficient for those wishing to work effectively with the RELAN. The RELAN program is currently configured to analyse, test, explore and simulate hypotheses (relations) between up to ten two-valued variables. One special advantage of this approach is that, in lieu of correlations, directed implications (←, →) are employed to delineate variable associations.

Effect Structure: Multiple Causes – Moderators – Mediators

In order to correctly capture the empirical reality in our hypotheses, we should also assign variables to their empirical functional role or their effect structure [6,14]. In reality, variables are not merely causes or effects; they can also be conditions (moderators) of other variables or they can establish chains of effects (mediators) between variables [15] (Figure 2).

We can even apply this classification to many everyday phenomena, such as the movement of a car (see later), recipes for cooking, or building a house.

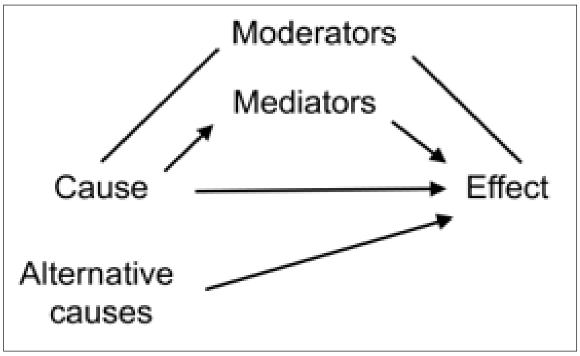

Figure 2: The Graphical Representation of the Hypotheses Improves a Useful Overview of the given Variable Relationships, but the Exact Structure of the Relationships between Variables still needs to be Specified by Logical Formulae.

Statistical Procedure

Here's a quick look at the statistical procedures of Relation Analysis (that the program does) and that are easily understood – especially compared to other complex statistical methods. As a simple example, consider the relation A implies B. On the basis of the number of variables, the truth values of the relation for all combinations of variables are first determined (T), then all corresponding frequencies are assigned (F), the expected frequencies (E) are calculated, and the significance testing is performed by using the approximate binomial test (z-statistic): $Z = (F – E) / SD$; $SD = \sqrt{E \cdot (1 - E/N)}$ [16]. This significance test is carried out in the same way for both the individual variable combinations (elementary relations) and for the proposed hypothetical relation (A → B) (Table 1).

Combination	A	B	T Truth Values	F Frequencies	E Expected F.	Z	P
1	0	0	1	15	10.435	1.607	0.054
2	0	1	1	15	19.565	-1.361	0.087
3	1	0	0	1	5.565	-2.064	0.020
4	1	1	1	15	10.435	1.607	0.054
				N = 46	46.000		

Table 1: If the p-Values of the Binomial Test Z-Statistics are Less Than the Specified Significance Level, the Logical Association is considered Significant; in the Example, only the Third Variable Combination is Significantly (a = .05) Understaffed. Since this Combination of Variables Values Represents Exactly the Negation of the Implication Function (T = 0), it is Negative to the Same Extent as the Implication Relation is Positive: Z = 2.064, P = .020, R+% = 98, Power =.66. R+% Indicates How Many Cases Agree with the Hypothetical Relation, and the Power is the Measure of the Probability of the Validity of the Result.

Results and Simulation

Example 1: Testing

In order to demonstrate the precision of a RELAN evaluation, data are constructed from an everyday example [6]. Let's assume that someone who is not familiar with driving observes the expression of the factors that seem to be important, such as the MOTOR, the GEAR, the COUPLING, the BRAKE and whether the car is moving (MOVE). Three technical observations could be the following (1 = yes, 0 = no):

- MOTOR (0), GEAR (0), COUPLING (0), BRAKE (0), MOVE (1) (car rolls downhill)
- MOTOR (1), GEAR (1), COUPLING (1), BRAKE (0), MOVE (1) (car is moving self-powered)
- MOTOR (0), GEAR (1), COUPLING (0), BRAKE (0), MOVE (1) (car is towed away)

Other technically possible combinations of the variable values are: 00001, 11101, 01001, 01110, 10010, 10110, 11000, 11001, 11010. We assume that the combinations of variables identified have been observed twice in the same way, giving a sample of N = 20.

The person's first statistical attempt may be to use conventional statistics, such as calculating correlations between all the variables, performing factor analysis and using logistic regression analysis. Unfortunately, the results are of little use: The only significant Phi coefficient shows the incompatibility between driving and braking: r_ϕ (MOVE, BRAKE) = -.67 ($p < .01$), the Principal Component Analysis (Kaiser criterion) shows a factor with trivial factor loadings (MOVE = -.892, BRAKE = .778), and the regression of MOVE by the predictors MOTOR, GEAR, COUPLING, BRAKE is insignificant. An alternative approach would be to cognitively process the past experiences with the moving or stationary car and try to specify a plausible logical hypothesis: "The engine is on and the first gear is engaged, then the power is transferred to the coupling and then the car is moving, but only if the brake is not applied".

((MOTOR (cause) ∧ GEAR (moderator)) → COUPLING (mediator) → MOVE (effect)) ∧ (MOVE | BRAKE Moderator)

If this hypothesis is tested with RELAN, we obtain the following very satisfactory statistics:
Z = 2.491, P = .0064, R+% = 100, Power = .80, N = 20

Complex relationships cannot be captured by correlative statistical methods, but Relation Analysis as a logical-statistical method provides a concise result. Working scientifically in this way means having a clear idea of the structure of effects and the dynamics in the variable system. But by better matching hypotheses to empirical phenomena, they can be elucidated more reliably, and their successful replication made more certain.

Example 2: Simulation

To talk about the next notable feature of Relation Analysis, I first need to define the concept of "perfect relations". Perfect relations are relationally optimal combinations of variable values in a hypothetical sample, that means maximal power and maximal effect size. These have two purposes

- To compare an empirical relation with its theoretical maximum (perfect relation).
- To check the appropriateness of sample sizes for hypothesized relations. Therefore, the simulation of relations provides information on how many cases must be present to be able to find a certain regularity.

To illustrate a simulation, a hypothesis with seven variables should be perfectly generated in a sample as small as possible. The variables come from psychology, where the many factors influencing social relationships (SOC) between people have been studied [17]. Six of these will be used here: Contact frequency (CON), Attractiveness (ATT), Similarity (SIM), Competence (COM), Reciprocity (REC), Promoting self-esteem (SEL). As there is (as yet) no detailed hypothesis in psychology about the exact structure of the effects of these factors, a possible explanation is proposed and simulated here: Social relationships can increase through two processes simultaneously: Social relationship increases,

- Only if a person is similar to me or attractive and the frequency of contact with that person is high, and
- Only if a person is competent, exchanging goods or services with me or the person enhances my self-esteem (the ONLY IF-THEN symbol ↔ indicates that these two explanations are the only ones):

Hypothesis: (SOC ↔ ((SIM ∨ ATT) ∧ CON)) ∧ (SOC ↔ (COM ∨ REC ∨ SEL))

This simulation could be perfectly realised in a sample of twenty-six cases, where all variable combinations that fit the relation are occupied by exactly one case:

Z = 5.101, P = .0001, R+% = 100, Power > .99, N = 26

This means that, in principle, no more than twenty-six cases would be needed to test these joint hypotheses (if the cases were representative).

Example 3: Comparing

The following example shows how RELAN is able to compare different hypotheses. In the original data, the program reveal four significant implications between the subsequent symptoms of depression: Agonising experience (AGO), Thought disturbance (THO), Broodiness (BRO) and not wanting to get up (in the morning) (NOT) [18]. These four implications were combined into a single hypothesis (Figure 3), which was statistically tested and confirmed:

Z = 2.990, P = .00160, R+% = 77, Power = .75, N = 150 (4 implications)

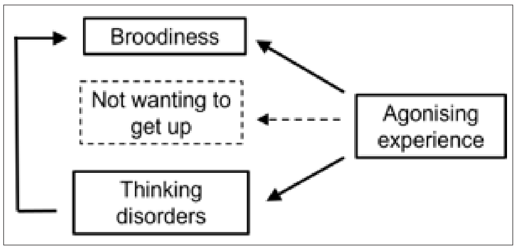

Figure 3: Four Symptoms of Depression are Implicitly Related and Combined in the First Integrating Hypothesis, the Second Hypothesis (Thick Lines) with only Three Variables seems more Valuable than the First.

As the statistical power of a combined hypothesis is desired to be somewhat greater, a further integrative hypothesis is formed with three variables, and this hypothesis has actually better values in the overall context:

Z = 3.890, P < 0.0001, R+% = 86, Power = .94, N = 150 (3 implications)

Since data configurations often leave room for different interpretations, the decision between competing hypotheses should always be based on the overall theoretical context.

Example 4: Exploration
The final ability of RELAN we are covering is 'exploration': The following example illustrates how a target variable (environmental behaviour) is influenced by seven causes and two mediators. The data comes from the interdisciplinary project 'Quality of Life and Environmental Behaviour: Consensus and Conflict in the Everyday Life of a Cultural Landscape', in the context of which a study was conducted on attitudes towards landscapes in Austria [19]. The questionnaire designed for the study was used in a telephone survey of the population (aged 15 to 80 years) in the Austrian federal state of Styria. A total of 247 variables from a sample of 401 people were subjected to a wide range of qualitative and quantitative analyses. For the explorative analysis of the psychological data the number of variables has been reduced to 98 Variables which were then median dichotomised.

Of particular interest was the prediction of environmental behaviour (Figure 4). It was not surprising that there was no direct link between environmental awareness and environmental behaviour, as many people say they are environmentally aware, but this does not translate into behaviour. The most direct influences on environmental behaviour are information on environmental topics, interest in environmental topics, and the impression that the surroundings are beautiful. Beautifulness is in turn influenced by landscape diversity, housing satisfaction, environmental satisfaction and, surprisingly, by environmental awareness, which in turn depends on political involvement. It was also unexpected that residential satisfaction and the social value of helping others influenced landscape connectednes

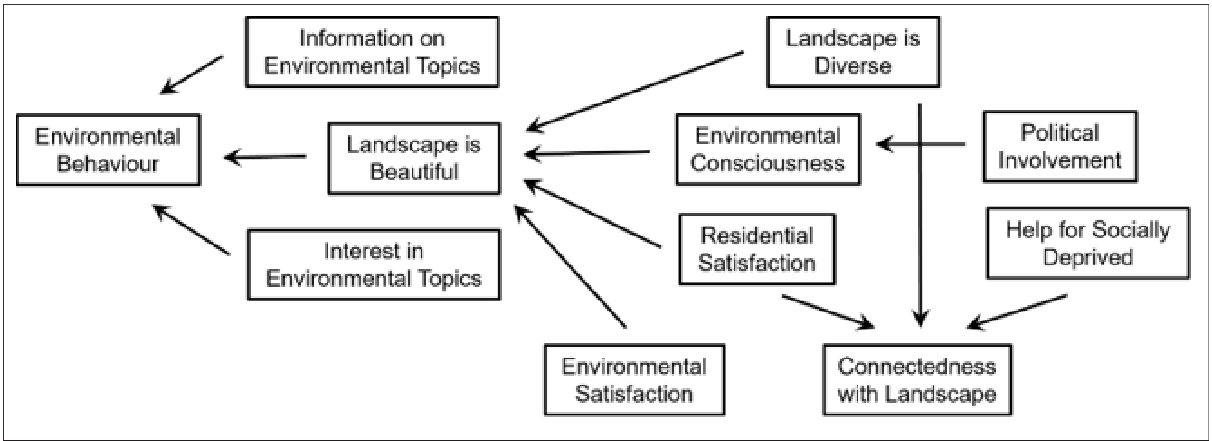

Figure 4: This Impact Scheme involves a Complex Effect Structure between Eleven Variables, including Six Causal Variables (not explained by others), Two Mediating Variables, and Two Effect Variables (explained by others). The Arrows indicate that the Variables at the Starting Point of the Implications are Significantly ($\alpha = .01$) related to the Variables at the End Point, and all have a (Effect) Statistic of $p(Y/X) \geq .80$ (PPV; Positive Predictive Value): Given the Variable X, the Variable Y Occurs with a Probability Greater than/Equal to 0.8.

This last example is intended to show how complex some variable relationships can be. It is important to recognize that such exploratory findings are merely preliminary hypotheses that require further verification through additional research or cross-validation.

Discussion
While the use of binary (two-valued) variables in RELAN may be perceived as a limitation, it also offers the benefit of allowing for the inclusion of up to ten variables in a single hypothesis, as well as the testing of six distinct logical functions between variables. Because of this increase in complexity, it is possible to describe effect structures more precisely through multiple causes, effects and moderators. There are probably only a few empirical laws that are so complex that ten determinants are not enough to capture the main structure of the effects.

To address the replication problem in biomedical research, Montgomery presents examples of logical errors in studies and suggests that at least a basic understanding of propositional logic and syllogistic deduction should be present in research, although he considers predicate logic and fuzzy logic to be even more helpful [20]. Predicate logic is actually very well suited as a methodology for representing scientific knowledge, but it requires considerably more basic mathematical knowledge than theory construction with RELAN [21,22]. The extension of the relationship analysis in the direction of fuzzy logic is a relatively straightforward process and is planned for implementation in subsequent versions of the software.

Statisticians might point out as a shortcoming of the method that no error terms are included in the structure of the hypotheses as 'deterministic models'. However, this can be contradicted by the fact that the program offers the option of checking each relation for potential error or interference relations using the program's extraction option and calculating these out of the hypothesis. This correction method takes into account not only the error variances, but also any complex error functions.

From the point of view of Bayesian statistics, RELAN's implication analysis must be criticized for not taking into account the effects caused by antecedent variables when testing implicative relationships between variables. If this is considered as a serious drawback, additional evaluations of the identified variable networks must be carried out using Bayesian statistical methods (e.g. "Bayesian Analysis Toolkit", "BAYESIA", "JASP").

Conclusion

The evaluation examples in this paper show that

- Relational Analysis can provide explanations for data configurations that are not revealed by conventional statistical methods,
- The simulation option offers the possibility of constructing data that are perfectly adapted to the hypothesis in question,
- Different competitive hypotheses can be compared statistically, and
- Extensive variable networks of significant variable associations can be identified.

In conclusion, the Relation Analysis (RELAN)

- Enables a much better adaptation of the statistical analysis to the complexity of the empirical phenomena than conventional statistical methods,
- Increases, despite of smaller samples, the chances of obtaining significant results and greater statistical power, and
- Improves theoretical structuring and scientific communication by formalizing the hypotheses and by including their presumed causal relationship (impact schema).

Funding: None

Conflicts of Interest: None

References

1. Baker M (2016) Is there a reproducibility crisis. Nature 533: 452-454.
2. Open Science Collaboration (2015) Estimating the reproducibility of psychological. Science. Available at: http://dx.doi.org/10.1126/science.aac4716.
3. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2: e1004085.
4. Brachem J, Frank M, Kvetnaya T, Schramm LFF, Volz L (2022) Replikationskrise, phacking und Open Science. Psychologische Rundschau 73: 1-17.
5. Gigerenzer G (2018) Statistical rituals: The replication delusion and how we got there. Advances in Methods and Practices in Psychological Science 1: 198-218.
6. Maderthaner R (2022) Relationsanalyse (RELAN)- Aussagenlogische, statistische und kausale Analyse von Daten. Heidelberg: Springer Spektrum. Available at: https://psycharchives.org/en/item/ed85b7e4-3b68-4ddc-be05-8fa50e8cba85.
7. Fletcher SC (2021) The role of replication in psychological science. European Journal for Philosophy of Science. Available at: https://link.springer.com/article/10.1007/s13194-020-00329-2.
8. Sanbonmatsu DM, Johnston WA (2019) Redefining Science: The impact of complexity on theory development in social and behavioral research. Perspectives on Psychological Science 14: 672-690.
9. Lehmann I, Schulz W. Mengen (2016) Relationen – Funktionen. Heidelberg: Springer Spektrum.
10. Hildebrand DK, Laing JD, Rosenthal H (1976) Prediction analysis of cross classifications, New York.
11. Eye A (1991) Prädiktionsanalyse. Vorhersagen mit kategorialen Variablen. Weinheim: Psychologie Verlags Union.
12. Denis-Papin M, Faure R, Kaufmann A, Malgrange Y (1974) Theorie und Praxis der Booleschen Algebra. Braunschweig: Vieweg.
13. Whitesitt JE (1973) Boolesche Algebra und ihre Anwendungen. Logik und Grundlagen der Mathematik. Braunschweig: Vieweg.
14. Nowak S (1976) Understanding and prediction. Dordrecht: Reidel.
15. Corraini P, Olsen M, Perdersen L, Dekkers OM, Vandenbroucke JP (2017) Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. Clinical Epidemiology 9: 331-338.
16. Freund PA (2010) Einstichproben-Tests. In: Holling H, Schmitz B (Eds). Handbuch Statistik, Methoden und Evaluation. Göttingen: Hogrefe pp: 382-390.
17. Maderthaner R (2021) Psychologie. Third edition. Wien: facultas, UTB basics.
18. Bortz J, Lienert GA, Boehnke K (2008) Verteilungsfreie Methoden der Biostatistik, 3. Aufl. Springer, Berlin.
19. Buchinger E, Burkart R, Fink M, Götzenbrucker G, Halper C, et al. (1999) Lebensqualität und Umwelthandeln: Konsens und Konflikt im Alltag einer Kulurlandschaft. Seibersdorf: Austrian Research Centers Seibersdorf (ARCS). Projektbericht (Band 2) pp: 231-283.
20. Montgomery EB (2024) Reproducibility in Biomedical Research (Second Edition), New York: Academic Press. Available at: https://doi.org/10.1016/B978-0-443-13829-4.00027-.
21. Bossel H, Hornung BR, Müller-Reißmann KF (1989) Wissensdynamik mit DEDUC. Grundlagen und Methoden dynamischer Wissensverarbeitung: Wirkungsanalyse, Folgenabschätzung und Konsequenzenbewertung. Braunschweig: Vieweg.
22. Relation Analysis (RELAN) A new Approach to Hypothesis Testing and Theory Construction (in preparation).