

As If *

Geoffroy de Clippel[†] Ryan Oprea[‡] Kareen Rozen[§]

July 2026

Abstract

We study the behavioral relationship between two core notions of rationality in economics: as-if rationality (behavior consistent with maximization of *some* utility function, the foundation of positive economics) and substantive rationality (maximization of *one's own* utility function, the foundation of normative economics). We *induce* textbook preferences over bundles for experimental subjects, and study to what degree subjects' choices reveal the preferences we've induced. We find that most subjects consistently behave as if they are rationally maximizing *some* utility function (indicating high as-if rationality), but only a minority consistently maximize the utility function they were assigned (indicating much lower substantive rationality). This suggests the internal consistency of as-if rationality often rests on a different behavioral foundation than true maximization, a finding with implications for applied welfare analysis.

Keywords: as-if rationality, substantive rationality, procedural rationality

JEL codes: C91, D91, G0

*First draft March 2024. Multiple seminar and conference audiences provided helpful comments. We are also grateful to Shachar Kariv for formative discussions. Since 9/2024 this research is supported by the National Science Foundation Award SES-2416935 and approved by Brown University IRB. It was previously supported by NSF Award SES-1949366 and approved by UC Santa Barbara IRB. AI was consulted for small coding improvements.

[†]de Clippel: Economics Department, Brown University, declippel@brown.edu.

[‡]Oprea: Haas School of Business, University of California, Berkeley, roprea@gmail.com.

[§]Rozen: Economics Department, Brown University, kareen.rozen@brown.edu.

1 Introduction

In this paper we study the relationship between two core notions of rationality in economics:

1. **Substantive rationality (SR)**: the claim that humans make choices that actually maximize *their own* objective functions (i.e., make choices that optimize over and therefore reveal their own tastes and preferences).
2. **As-if rationality (AIR)**: the claim that humans make choices *as if* they are maximizing *some* objective function (i.e., make internally consistent choices that obey the generalized axiom of revealed preferences or GARP).

Each notion is foundational to a separate branch of the field: AIR is the centerpiece of revealed preference theory and serves as a foundation of *positive economics*, while SR is the key assumption justifying traditional applied welfare analysis and serves as a foundation of *normative economics*. Understanding how the two notions are behaviorally related is important not only for understanding the nature of human rationality, but also for the practice of welfare analysis. However, to date, we know very little about this relationship. Our contribution is a novel experiment designed to gather some initial evidence.

Rationality and Welfare. Traditionally, welfare analysis has relied on the assumption that humans are substantively rational (i.e., make choices organized by their objectives), allowing economists to structurally infer preferences directly from choices to make welfare judgments. However, decades of behavioral research suggests that people fail substantive rationality in many settings: they make choices that are suboptimal or influenced by non-consequentialist factors, thereby undercutting the traditional identifying assumption of welfare analysis. An important line of research has attempted to develop a framework for conducting “behavioral welfare analysis” by taking account of the possibility that decision makers may not always make substantively rational maximizing choices that reveal their normative preferences. Central to this enterprise is the identification of a *welfare relevant domain* – a subset of choices

that are likely to be substantively rational and thus to reveal to the observer the normative objectives of the decision-maker (Bernheim and Rangel, 2009).

It is common in this literature to propose to apply behavioral welfare analysis precisely to choices that violate AIR (Bernheim and Rangel, 2007, 2009). By contrast, if choices comply with standard axioms of rationality and therefore satisfy AIR (i.e., are consistent with utility maximization), they naturally seem like particularly good candidates for welfare relevance. It is therefore natural to treat as-if rational choices as evidence of substantive rationality.

To what degree is this kind of inference justified? To answer this question we need to understand to what degree AIR serves as a diagnostic indicator of SR, but to date we have little evidence to inform an answer. Naturally, SR implies AIR: a decision-maker who maximizes her objective function makes choices that are consistent with utility maximization. But the reverse need not hold: decision-makers may use internally coherent choice procedures or heuristics that satisfy GARP, without maximizing the welfare-relevant objective function. If so, consistency-based criteria alone may be insufficient for identifying welfare-relevant choice.

Experimental Design. To study the relationship between AIR and SR, we conduct a novel experiment in which we *induce* simple, textbook objective functions and then incentivize subjects to reveal these preferences back to us in choice. In particular, we ask subject to choose bundles (x, y) from a sequence of budget lines, and reward their choice using one of several classic objective functions. By paying subjects proportional to $\min\{x, y\}$ we induce perfect complements (Leontief) preferences, while paying $(x+y)/2$ induces perfect substitutes and xy induces Cobb-Douglas preferences. Unlike traditional behavioral experiments (e.g., studying preferences for risk), these induced preferences allow us to directly assess substantive rationality, since failure to comply with the optimal demands implied by induced preferences are transparently money-losing mistakes. We include design features that allow us to assess robustness including direct measures of subjects' understanding of these payoff functions and fivefold manipulations of payoff levels.

This design allows us to directly assess and contrast AIR and SR in the same data and therefore answer our motivating question. Because we vary the budget line

across these tasks, we can naturally calculate Afriat’s CCEI score (a classic measure of GARP compliance), allowing us to assess as-if rationality: how closely subjects’ choices come to being rationalizable by *some* utility function. But by comparing subjects’ choices to the optimal demand functions induced in our design, we can *also* directly assess how substantively rational their choices are. Because we can assess both types of rationality in the same data, we can study the relationship between the two notions, and in particular, examine to what degree as-if rationality serves as a reliable indicator of substantive rationality. To our knowledge, this is a unique feature of our design, allowing us to answer our motivating question for the first time.

Empirical Findings. We find that subjects are highly as-if rational in our induced preference tasks, mirroring results from prior experiments attempting to measure natural preferences (e.g., for risk). Most of our subjects – about two-thirds – are *consistently* GARP compliant under all three induced preferences they face. However, the majority of these consistently as-if rational (AIR) subjects are not consistently substantively rational (SR).

Our empirical approach distinguishes between noisy implementation of optimal demands (e.g., frictions or optimization error) and the use of decision rules that are systematically different from the induced objective function. Allowing for noisy implementation of optimal demands, about 43% of consistently AIR subjects are also consistently substantively rational; under a stricter criterion, this fraction falls to 21%. We therefore find that AIR is not a reliable indicator of SR – a conclusion that is robust over a number of pre-registered metrics.

Most consistently AIR subjects exhibit choice patterns that substantially deviate from optimal demands under at least one induced preference. Nonetheless, these same subjects appear to follow internally consistent, as-if rational choice rules, often repeating distinct rules that maximize specific objective functions across the design. Moreover, these subjects exhibit several markers of economic rationality in our setting, including consistency across repetitions, compliance with the law of demand, homotheticity and symmetry. Borrowing terminology from Simon (1976a), we refer to AIR subjects who fail SR as *procedurally rational*. Because of the prevalence of such subjects in our data, we find that as-if rationality has limited diagnosticity for

substantive rationality.

Implications. The main conclusion we draw from these results is that as-if rationality and substantive rationality are only weakly related to one another behaviorally. This suggests there is a behavioral or cognitive tendency to be internally consistent (use GARP-compliant choice rules) even when failing to maximize, at least in some economically important settings (e.g., choices from budget sets). As a result, decision makers often mimic the properties of optimizing agents even when their choices are not well-described by their underlying preferences.

This finding is not only scientifically but also practically relevant: it suggests that as-if rationality may be an insufficient tool for identifying the welfare-relevant domain, at least under many theories of welfare. This conclusion does not require taking a position on whether a stable objective lies behind choice: even when we guarantee that one does—by inducing it ourselves—internal consistency fails to reveal it. In economically important contexts (such as choice under budget constraints), an independent drive towards consistency may generate as-if rationality even when behavior contains little information about underlying preferences. This points to the importance of developing alternative approaches for identifying welfare-relevant choice data when assessing welfare (Rees-Jones, 2024; Handel and Schwartzstein, 2018; Bernheim and Taubinsky, 2018; Bernheim, 2016).

In drawing out these implications, it is important to emphasize what we do *not* claim based on these data. While we document relatively low rates of substantive rationality and relatively high rates of as-if rationality in our setting, we expect there are settings where SR is high and other settings where AIR is low. Both types of rationality may vary substantially across environments, depending on task difficulty, framing, familiarity, availability of natural heuristics and opportunities for learning. The purpose of this paper is thus *not* to establish a universal measurement of either AIR or SR (a quixotic goal if ever there was one). Rather our contribution is to show that AIR need not indicate SR, because, at least in some important choice settings, humans exhibit independent tendencies towards GARP-like consistency even when they do not maximize their objectives. This finding calls into question the sufficiency of internal consistency for identifying the welfare-relevant domain in applied welfare

analysis.

Related Literature. Our work connects to several literatures, including a theoretical literature on revealed preferences that we discuss when describing the benchmark framework in Section 2. An important experimental literature measuring GARP compliance in the context of naturally-occurring preferences applies these theoretical tools. This literature employs similar budget-set experimental designs to measure as-if rationality in settings involving the expression of natural preferences in domains like risk (Choi et al., 2007, 2014; Halevy et al., 2018a; Dembo et al., 2021), time (Lanier et al., 2022), interpersonal distribution (Andreoni and Miller, 2002; Fisman et al., 2007; Zame et al., 2023) and ambiguity (Ahn et al., 2014). This literature has generally produced strong evidence of widespread as-if rationality in multiple domains, though – as in our data – it has also found heterogeneity in measures of GARP compliance.

Our paper also connects to a growing literature on behavioral welfare analysis, surveyed in de Clippel and Rozen (2024, Section 6). This literature can be divided into two broad approaches, and our findings are consistent with and relevant to each. One branch of the literature holds that there is a well-defined underlying objective and that the analyst’s task is to recover it from imperfect choice: by separating “revealed mistakes” from genuine preference (Kőszegi and Rabin, 2007, 2008), by distinguishing normative from positive preferences and modeling the wedge between them (Beshears et al., 2008), or by recovering preferences structurally from inconsistent data (Apesteguia and Ballester, 2015; Rubinstein and Salant, 2012; de Clippel and Rozen, 2023). For this literature, our findings suggest that choice consistency alone, à la AIR, is unlikely to be a reliable tool for identifying those objectives.

A second approach, more deferential to decision makers’ choices, declines to posit a hidden utility at all and instead assesses welfare on the basis of coherent choice: the choice-theoretic criterion of Bernheim and Rangel (2007, 2009), and more radically Sugden’s opportunity criterion and his critique of the “inner rational agent” (Sugden, 2004, 2018), which denies that there is any integrated set of latent preferences waiting to be uncovered. For this literature, internal consistency of welfare rankings is imposed by construction but GARP compliance in choice data à la AIR is

not necessarily taken as clear evidence of welfare relevance. In particular the literature emphasizes that if a decision maker fundamentally misunderstands her options (“characterization failure”), her choices may be perfectly AIR but nonetheless fail to reveal welfare-relevant information.¹ By contrast, the literature is more reluctant to disregard choices because of failures to optimize (“optimization failure”) for the simple reason that such failures are often difficult to empirically establish, and the analyst is therefore in danger of injecting contestable judgments about what the chooser ought to want (Bernheim, 2021; Bernheim and Taubinsky, 2018).

Our experiment produces a rare environment in which (unlike most natural choice contexts) true optimization failures are directly identifiable from choice.² Because of this we are able to reinforce and extend this literature’s suspicion of GARP consistency as a metric, by showing that even verifiable *optimization failures* can appear AIR-consistent. By the same token, because our AIR subjects choose coherently, their choices would be admitted to the welfare-relevant domain and deferred to by default in much of this literature,³ even though they often make choices that are in fact improvable.

Our results also tie in closely with the literature on procedural rationality: the idea that decision-makers often use coherent, orderly rules that don’t necessarily maximize preferences. This idea is particularly associated with Herbert Simon (Simon, 1955, 1976a) and has for decades been at the center of important research in cognitive

¹Bernheim and Taubinsky (2018, p. 408) give the example of a consumer who consistently mistakes apples for oranges: his choices satisfy the axioms even though he never gets what he prefers. As a result Bernheim and Taubinsky (2018, p. 408) caution that “a reduction in WARP/GARP violations is neither necessary nor sufficient for an improvement in the quality of decision making.”

²Nothing in our experiment calls into question the literature’s concern on this point. In standard field settings objectives are typically unobserved and must be imputed: as Bernheim and Taubinsky (2018, p. 400) put it, “the critical question is whether one can detect [optimization failure] using systematic evidence-based criteria without knowing the objective function,” and analyses of it “generally assume . . . that [the objective function] is known.” Our conclusions concerning optimization failure stem precisely from the fact that, unlike in typical choice settings, we *can* know the objective function because we have induced it. What is more, we can verify that subjects do understand their objective functions but are often unwilling or unable to make optimal use of them in choice.

³This is because our subjects do not seem to be suffering from characterization failure—they answer our incentivized comprehension questions about their objectives correctly and often apply the payoff-maximizing rule in some blocks. Their choices seem instead to be closer to optimization failures, which this literature is, in most contexts, justifiably wary of invoking as a basis for paternalistic intervention.

science (e.g. Payne et al., 1993). It has also been a topic of increasing interest in economics. Halevy and Mayraz (2024) uses a novel experimental design to show that subjects prefer to make decisions using procedural reasoning when allocating budgets between Arrow securities. Nielsen and Rehbeck (2022) show that subjects systematically prefer rules that comply with the axioms underlying expected utility theory, but fail to apply these rules in lottery choice. Arrieta and Nielsen (2024) use incentivized self-descriptions of strategies to show that subjects tend to use more rule-like decision procedures (procedures that are easier to describe and replicate) as decision problems grow more complex. The literature documents several systematic influences on the selection of choice heuristics, summarized for instance by DellaVigna (2009). One such influence is a tendency toward over-diversification, which is reflected in the choice patterns more frequently employed by our subjects.

The paper closest in spirit to ours is Ariely et al. (2003). They show demand behavior in simple monetary valuation tasks is internally coherent (with stable ordinal rankings over a fixed set of real objects, including chocolates and a cordless keyboard), and yet the valuations for each object are highly sensitive to initial random prices. Despite being downward sloping, the resulting demands should thus not be taken as rational in the conventional sense. As they point out, downward sloping demand is a necessary but not sufficient condition for rationality (Becker (1962) showed that even random choice will probabilistically yield downward-sloping demand). We directly establish as-if rationality over bundles (a stronger condition) and, by using an induced preference task, gather direct evidence on substantive rationality at the subject level.

Finally, there is a literature in economics regarding complexity (Oprea, 2020; Camara, 2024) and the role it plays in producing classic “behavioral” anomalies in risk and time preferences. This literature has produced evidence that anomalous behaviors that are typically interpreted as outgrowths of preferences, like small stakes risk aversion (Khaw et al., 2021), probability weighting (Enke and Graeber, 2023; Oprea, 2024), loss aversion (Oprea, 2024), and hyperbolic discounting (Enke et al., 2025), are to a great extent the consequence of systematic complexity-derived valuation errors. Overall, as Conlisk (1996, p. 676) notes, “*Biases, heuristics, and norms have been used in various economic models to explain otherwise puzzling behavior.*” Here we suggest, conversely, that behavior need not be puzzling for it to be generated by

systematic errors. Indeed, choice heuristics can appear deceptively rational.

2 Theoretical benchmark

We study the classic model of consumer choice, in which an agent makes choices from linear budget sets. For two goods, and income normalized to one, a budget is

$$B = \{(x, y) \in \mathbb{R}_+^2 \mid p_1x + p_2y \leq 1\}.$$

The workhorse model in this framework is **rationality**. A rational consumer has an increasing utility function $u : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ and picks the highest-utility bundle she can afford.⁴ This classic model abstracts from (or assumes away) any cognitive costs of finding the utility maximizing bundle: utility arises only from the bundle selected.⁵

The **data** comprises observed price-demand pairs $\{(p^k, d^k) \mid k = 1, \dots, K\}$, where $p^k = (p_1^k, p_2^k) \in \mathbb{R}_{++}^2$ and $d^k = (d_1^k, d_2^k) \in \mathbb{R}_+^2$. The data is **consistent with rationality** if there is an increasing utility function u such that for each k , the demand d^k maximizes u over the budget set $B^k = \{(x, y) \mid p_1^kx + p_2^ky \leq 1\}$. Since u is increasing, the optimal bundle is on the budget frontier.

As utility functions are not directly observed, rationality is assessed through revealed preferences, as first suggested by Samuelson (1938). The bundle d^k is revealed weakly preferred to any bundle at the frontier of B^k , and revealed strictly preferred to any bundle that is strictly within B^k . The **Generalized Axiom of Revealed Preference (GARP)** holds if there are no revealed-preference cycles (involving at least one strict preference) in the data.⁶ The choice data satisfies GARP if, and only if, it is

⁴Here, increasing means $u(x, y) > u(x', y')$ if $(x, y) > (x', y')$. But given Afriat (1967), the focus on testable implications means the analysis is unchanged for variants of this definition.

⁵Some may argue that using a simple heuristic when selecting bundles is rational because fully maximizing one's utility function is too costly. We are sympathetic with this viewpoint, and our results can be seen as giving credence to it, but this is not the prevalent use of the term rationality in consumer choice. Instead, we adopt the conventional definition used explicitly or implicitly in virtually all textbooks.

⁶With only two goods, we can restrict attention without loss to cycles of length two when checking GARP (Rose, 1958; Banerjee and Murphy, 2006); hence in this case, the absence of short cycles in the spirit of the weak axiom of revealed preference (WARP) is enough.

consistent with rationality (Afriat, 1967; Varian, 1982).

Consistency with rationality means that observed choices are indistinguishable from those of a rational agent. That is, it tests whether someone’s choices are *compatible* with rationality, but not whether they are truly rational. This is commonly called **as-if rationality**, and is justification enough for using rationality in descriptive models. When it comes to normative economics, however, one typically takes rationality at face value: the policymaker assumes that the revealed preference directly reflects what the decision maker *likes* (i.e., the decision maker’s *tastes*). This interpretation of choices relies on individuals actually maximizing their utility, or, in the terminology of Simon (1976a), being **substantively rational**.

Rather than focusing on whether GARP holds perfectly, the literature gradates the extent of as-if rationality in choices. Afriat (1973)’s **critical cost efficiency index (CCEI)** is the most commonly used yardstick.⁷ The CCEI measures the largest fraction of income that can be retained while avoiding revealed-preference cycles *à la* GARP. It is the supremum over all $\lambda \in [0, 1]$ such that GARP holds for the *modified* revealed preference, whereby the demanded bundle d^k in the k^{th} observation is revealed (strictly) preferred to any bundle (x, y) such that $p_1^k x + p_2^k y \leq (<) \lambda$. The condition becomes more demanding as λ gets bigger, with perfectly rational data achieving a CCEI of 1.

Using the CCEI and related metrics, high degrees of GARP consistency have been documented in the literature. But whether this is impressive depends on the combination of budget sets tested. To illustrate using an extreme example, if all budget sets are related by inclusion, then no matter what choices are made, GARP must hold. Bronars (1987) introduced a now common approach to assess power and benchmark results, by comparing the empirical distribution of CCEI scores to that obtained under simulated consumers, whose choices are drawn uniformly at random over the tested budget lines *à la* Becker (1962).

The CCEI and related metrics are *agnostic* about the subject’s true preference. As Varian (1990) observes, if the modeler has in mind a specific preference $\succeq$, then she

⁷Other proposed approaches include Houtman and Maks (1985), Varian (1990), Echenique et al. (2011), Dean and Martin (2016), Halevy et al. (2018b), and de Clippel and Rozen (2023).

could compute the budgetary adjustment needed *relative* to $\succeq$. If the demanded bundle is not at a point of tangency with the $\succeq$ -indifference curve passing through it, then the choice must be suboptimal. Given the specific preference $\succeq$, define the **preference-relative CCEI** as the supremum over $\lambda \leq 1$ such that for all observations k , the demanded bundle d^k is $\succeq$ -preferred to any bundle (x, y) with $p_1^k x + p_2^k y \leq \lambda$. This is the largest percentage of income that can be retained in all budget sets, while ensuring that no feasible choice is strictly preferred. As Varian (1990) hints, and Halevy et al. (2018b) shows more generally, the standard (preference-agnostic) CCEI is the *supremum* of the preference-relative CCEIs, taken over all continuous and increasing preferences $\succeq$ over bundles. Hence the two coincide for the preference that is actually being maximized (if such a preference exists).

3 Experimental Design

Our pre-registered experiment is divided into three blocks, each of which has 30 rounds.^{8,9} In each round, the subject selects a bundle (x, y) from a two-dimensional budget line. The budget lines vary in their x - and y -intercepts, and therefore in the relative prices and income represented. Using a graphical interface (shown in Figure 1), the subject makes a selection by clicking on the budget line. After clicking, they can adjust this choice as desired (either using the mouse again or the arrow keys). There is no ‘default’ bundle: until the subject clicks on the budget line, no selection appears on the screen. Subjects must then click on a button to finalize their selection before moving to the next round.

Each subject is assigned to one of two possible incentive conditions, *Higher* or *Lower*. The incentive condition changes only the dollar earnings associated with points achieved in a round, with the Higher treatment representing a fivefold increase. Other than this

⁸The pre-registration is available at Aspredicted #266054.

⁹An earlier version of this paper featured a slightly different experiment, detailed in Appendix C, that did not vary payoff levels or directly confirm understanding of payoff functions. That experiment, conducted in July 2023 included an additional block and a few other minor differences. In January 2026 we pre-registered and ran the new experiment when revising the paper to respond to questions and concerns from early readers and to substantially expand the size of our dataset. Nonetheless, the conclusions from the two experiments are very similar.

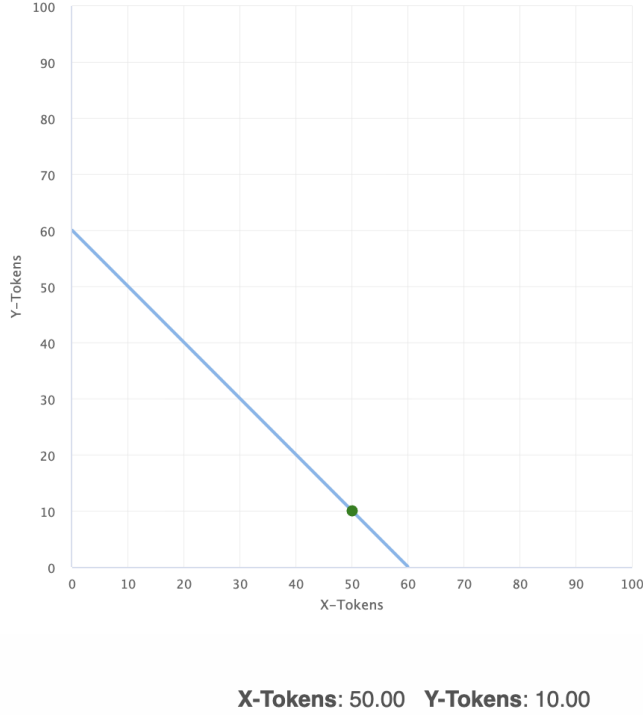


Figure 1: Screenshot from experimental software. The X -tokens and Y -tokens displayed at the bottom correspond to the green dot selected from the budget line in this example.

difference in incentive levels, all subjects face a survey with the same broad structure, and which is designed for a within-subject analysis of rationality. The experiment varies two treatment variables within subjects. First, across blocks we vary subjects' objective functions. Second, within each block, we vary the budget line across rounds. We discuss each of these variables and how they vary in the following two subsections.

3.1 Blocks and Preferences

In each block, we assign subjects a different payoff rule π (termed an 'earnings formula' in the survey). A payoff rule π maps a bundle (x, y) into *points*, which are translated into earnings. Doing this allows us to control and vary the subject's objective function, under the classic assumptions that subjects are consequentialist and monotone: caring only about outcomes, and preferring more money to less.

We chose three classic textbook preferences that are straightforward to describe:

- **Perfect Substitutes:** $\pi^S(x, y) = (x + y)/2$. Thus a subject's points are equal to the average of the x and y tokens in their bundle. To maximize π^S , one should spend the entire budget on the cheaper good, selecting the bundle at the largest intercept of the budget line (if the goods have equal prices, then any bundle on the budget line is equally good).
- **Perfect Complements:** $\pi^C(x, y) = \min\{x, y\}$. Thus a subject's points are equal to the smaller of the x and y tokens in their bundle. To maximize π^C , one should select the bundle on the diagonal ($x = y$).
- **Cobb-Douglas:** $\pi^{CD}(x, y) = xy$. Thus a subject's points are the product of x and y . To maximize π^{CD} , one chooses the bundle at the center of the budget line.

A benefit of using these classic preferences is that they generate starkly different demand patterns. A typical and visually appealing way to present the data from budget experiments plots the demand share of the X -dimension, given by $x/(x + y)$, against the log of the price ratio, p_1/p_2 . Figure 2 shows what these plots would look like for a perfectly rational subject (*homo economicus*), under each of our objective functions and using the budget lines we tested (described in the next subsection).¹⁰ As is clear from the Figure, the different demand patterns allow us to detect when subjects fail to properly change their behavior when induced payoff rules change across blocks.

Using induced-preference payoff rules removes scope for subjects to bring their own (naturally occurring) preferences into the task. These treatments are thus designed to allow us to clearly identify *mistakes* in preference maximization

3.2 Budget Lines

Within each block, we assigned subjects the same collection of 30 budget lines. Each of these budget lines can be described as a pair (x_{int}, y_{int}) , i.e., its x - and y -intercepts.¹¹

¹⁰Regarding the Substitutes panel of the Figure, all choices are of course optimal when $p_1 = p_2$.

¹¹Hence the price ratio p_1/p_2 is y_{int}/x_{int} .

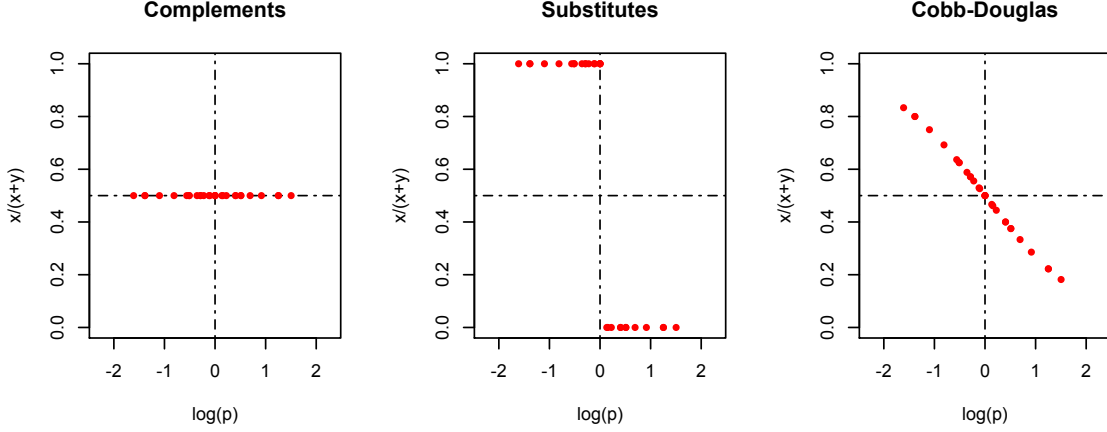


Figure 2: Optimal demand share $\frac{x}{x+y}$ plotted against $\log\left(\frac{p_1}{p_2}\right)$, for each treatment.

In choosing budget lines for the experiment we aimed to (i) include a wide range of budget lines that allow us to crisply assess subjects' demands, (ii) use the same set of budget lines in each block to facilitate clean comparisons of rationality across blocks and (iii) include some special diagnostic budget sets that can allow us to study to what degree subjects who fail to maximize induced preferences nonetheless reflect some regularity properties implied by the preferences we induce.

In order to satisfy goals (i) and (ii) we selected a wide range of budget lines using intercepts in the Cartesian product $\{60, 70, 80, 90, 100\} \times \{20, 40, 60, 80, 100\}$, and permute whether the first or second number serves as the x - or y -intercept resulting in 25 budget lines. To this set we add 5 budget lines that allow us to assess subjects' compliance with the following predicted properties of our induced preferences:

- **Consistency in repeat choices.** Our induced preferences all have a unique optimal demand (with the exception of Substitutes in the case of equal prices). Repeating budget lines sheds light on how consistent, or how noisy, subjects' choices are in two instances of an identical task. Five of our budget lines, given by the pairs of intercepts (20, 70), (60, 60), (60, 100), (80, 20), and (80, 60), appear twice over the course of the 30 rounds in a block. The set of repeated budget lines spans a wide range of slopes in our experiment, ranging from the 8th to the 96th percentile, and including the median slope (-1).

- **Symmetry.** Our induced preferences are all symmetric: that is, $\pi(x, y) = \pi(y, x)$. Budget sets that are symmetric mirrors of each other (but are not identical, due to unequal prices) offer a crisp test of compliance with symmetry. Our set of tested budgets include two such symmetric pairings: (60, 100) and its mirror (100, 60), plus (80, 100) and its mirror (100, 80).
- **Homotheticity.** Our induced preferences are all homothetic: that is, when income is scaled, the optimal demand is scaled proportionally. Homotheticity can be assessed using two groups of budget lines. One is the pair of budget sets (50, 30) and (100, 60), where income doubles. The other is the trio of budget sets (60, 60), (80, 80) and (100, 100), all with slope -1 .

The resulting full collection of 30 tested budget lines, which are faced by all subjects in all treatments, appears in Appendix A. The order in which the 30 budget lines appear in a block is randomized across subjects, and constant within subject.

3.3 Implementation

The dataset comprises 400 subjects, with 200 in each of the *Higher* and *Lower* incentive treatments. The survey was administered via Prolific in January 2026.¹² Screenshots of the experiment, including the instructions, are provided in Appendix OA.2.

After the instructions and before starting the main portion of the experiment, subjects confirm their understanding of the interface and our earnings formulas by taking two quizzes. The first quiz comprises four multiple-choice questions that check for an understanding of bundles and how to select them. The second quiz comprises three questions, each of which explains one of the three earnings formulas, and asks the subject to compute the formula for a simple example and enter their answer in a

¹²Initially, 403 subjects completed the survey, as Prolific continues to recruit if a subject ‘times out’. There was a recording error in the data of eight subjects, who were thus excluded from the sample. Five additional subjects were then collected through Prolific to attain the pre-registered total of 400 subjects, with 200 per treatment. All subjects were required to be fluent in English, reside in the United States, have an approval rate of at least 98% on Prolific, and have at least 100 but no more than 1,000 prior Prolific submissions.

text field.¹³ To ensure comprehension, each earnings formula is described using both commonplace and mathematical language. In each of the quizzes, subjects were not allowed to proceed without providing correct answers to all questions. To incentivize attention to the instructions, we awarded subjects an extra \$1 in (sure) earnings for each quiz that they completed entirely correctly on the first try.

A screen presenting both the earnings formula and noting how points are translated to monetary amounts is provided at the start of each treatment block. Specifically, in the *Higher* (*Lower*) incentives treatment, a point is worth 25 cents (5 cents, resp.) in all treatment blocks but Cobb-Douglas; in the latter, a square root is first applied to $\pi^{CD}(x, y)$ before multiplying by 25 cents (5 cents, resp.), to ensure greater comparability of payoff scale across treatments.¹⁴ In addition, the earning formula is constantly present for subjects’ reference on every decision screen. No feedback was provided after choice rounds. At the end of each treatment block, there is a screen asking subjects to estimate the likelihood with which their choices were close to optimal in that block. This unincentivized measure of “cognitive uncertainty” (Enke and Graeber, 2023) allows us to study whether imperfectly rational subjects are aware of their imperfection or are (alternatively) making errors that they are unaware of. Once all three blocks are complete, subjects take a three-question cognitive reflection test to measure susceptibility to intuitive over deliberate decision-making.¹⁵

Subjects received \$10 for completing the survey (and up to \$2 extra based on their quiz performance). On top of this, subjects had a twenty-percent chance of being selected for additional earnings, equal to their monetary earnings from one randomly selected round. The median completion time was just under 42 minutes.

¹³As we report in Section 4.6, 85% of subjects were able to compute all three payoff formulas correctly on their first attempt and 98.5% by their second. In Online Appendix B.1 we show that results are nearly identical if we drop subjects who made an error on their first attempt.

¹⁴Hence in all treatments, (z, z) gives $\$0.25z$ ($\$0.05z$) in the higher (lower) incentive condition.

¹⁵We use questions based on Toplak et al. (2014), which expands upon Frederick (2005).

4 Results

In Section 4.2, we use tools from revealed preference theory to establish that most subjects are consistently as-if rational (AIR) in our tasks, and yet most of these AIR subjects are not *substantively rational* (SR) in the sense that they do not reliably make choices that resemble their induced preferences. In Section 4.3 we argue that these AIR-but-not-SR subjects are *procedurally rational*, using choice rules that obey a number of signatures of rationality despite not being payoff maximizing. In Section 4.4, we show that because of the prevalence of procedural rationality in the data, AIR serves as a poor diagnostic for SR. In Section 4.5 we provide some evidence that procedural rationality stems from bounded rationality. Finally, in Section 4.6 we show that our results are not due to confusion about payoff functions and do not qualitatively change with stronger incentives. Because of this, we pool data from the Higher and Lower incentives treatments in most of our analysis.

4.1 Measuring Substantive and As-If Rationality

By calculating the CCEI index for each choice block, we can produce measures of both As-if and Substantive rationality in every choice block:

- **As-if rationality.** We follow common practice in the literature by calculating Afriat’s CCEI index, which is a measure of how much budget lines would have to be shifted to make a choice profile consistent with GARP, and benchmarking power using the Becker-Brønars approach. We classify a choice block as as-if rational if its CCEI index exceeds the 95th percentile CCEI from simulated random decision-makers (for our budget sets, we find this Becker-Brønars benchmark is 0.8795).
- **Strict substantive rationality.** To (strictly) measure the substantive rationality of the same choice block, we leverage the unique fact that we can observe subjects’ true objective function to calculate the CCEI score *relative to the true induced preferences*. This measure allows us to assess GARP compliance not relative to *any* objective function, but relative to the objective function relevant

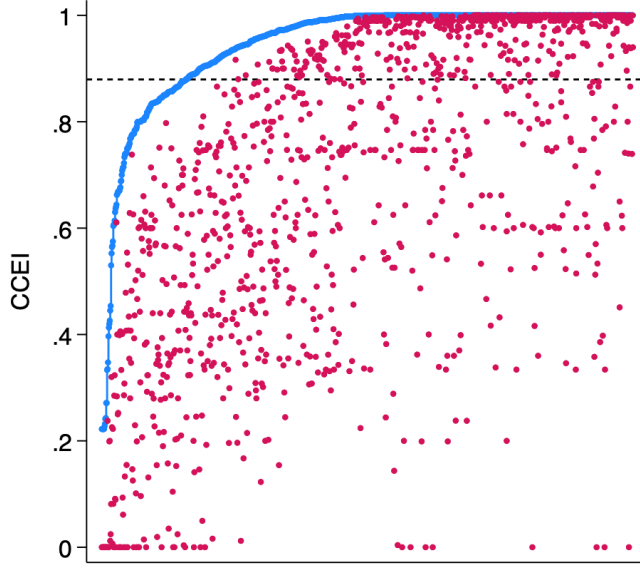


Figure 3: For each subject/block combination, the preference-agnostic CCEI and the preference-relative CCEI (for the induced objective in that block) plotted below it.

to a maximizing subject. As discussed in Section 2 the standard, preference-agnostic CCEI is the supremum of the preference-relative CCEIs, taken over all continuous and increasing preferences over bundles. The two thus coincide for the preference being maximized (if such a preference exists). We say a choice block is *strictly* substantively rational if its CCEI, *computed relative to the assigned objective function*, exceeds the 0.8795 benchmark above.

The top (blue) curve in Figure 3 connects, in increasing order, the preference-agnostic CCEI scores for each subject-block combination. The figure indicates high rates of as-if rationality, with about 84% of choice blocks exceeding the benchmark (the horizontal dashed line). As-if rationality rates of this magnitude are typical in the literature studying CCEI scores for natural preferences in the context of risk or time.¹⁶ Below each CCEI score, Figure 3 also shows (in red) the preference-relative CCEI score

¹⁶Indeed, as discussed in Appendix C, in our 2023 experiment we found very similar distributions of CCEI scores in (i) induced preference blocks and (ii) a (fourth) standard block eliciting naturally occurring risk preferences.

achieved for that same subject-block combination. Clearly this measure of strict substantive rationality is severely downward-shifted relative to the standard (preference agnostic) CCEI score, suggesting that choice blocks are far less substantively rational than they are as-if rational. Indeed, using the same 0.8795 benchmark, only half as many (41% of) choice blocks are strictly substantively rational as are as-if rational.

One of our key motivations for studying substantive rationality is to understand to what degree subjects reveal their objectives in choice. For this purpose, the preceding operationalization of substantive rationality may be too strict: subjects in our sample often broadly reveal their preferences but fail the preceding test simply because their choices are noisy implementations of optimal demand. To illustrate, Figure 4 plots the decisions from five sample subjects (one subject in each row) for each induced choice blocks, in the same manner as Figure 2. Subject 12 (top row) is a subject who exhibits strict substantive rationality in every choice block: their chosen bundles exactly match optimal the demands pictured in Figure 2, producing a preference-relative CCEI score of 1 in each block. By contrast, for both subjects 143 and 338, choices in the Complements block fail strict substantive rationality. There is, however, a marked difference between these subjects. Subject 143’s demand pattern is incorrect in Complements, and seemingly reveals the Cobb-Douglas preference instead. Subject 338, on the other hand, appears to noisily implement optimal demands in all three blocks. Looking at subject 338’s set of choices, an observer could easily identify which of our three induced preferences the subject was assigned in each block based on her choices, but the same is not true for subject 143 (one would mistakenly guess she was facing Cobb-Douglas in the Complements block).

Because our strict criterion for substantive rationality conflates failures-to-reveal with noisy implementation of choice rules, we will focus on a more forgiving operationalization:

- **Substantive rationality.** We classify a choice block as (noisily) substantively rational if it merely *reveals* the subject’s induced objective function in the sense above (this is a conservative choice in that the results to follow are strengthened if we focus instead on the stricter definition). Specifically, we run for each choice block an empirical horserace between three preferences (Complements, Substi-

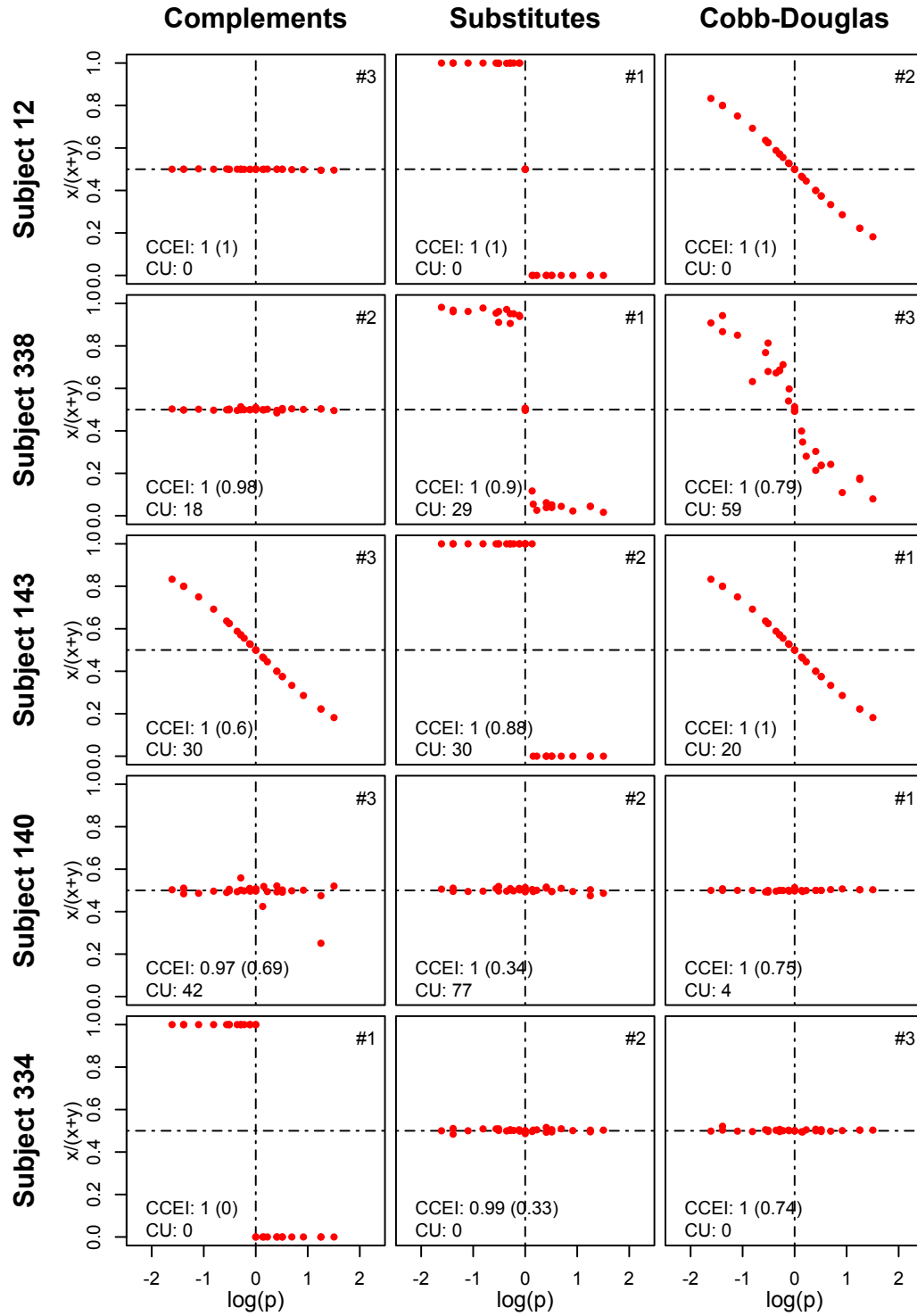


Figure 4: Choice data for five subjects who are as-if rational in all choice blocks. See Online Appendix OA.1 for similar plots for every subject in the dataset.

tutes, and Cobb-Douglas), examining whether demands *more closely* resembles optimal choices for the actual induced preference than optimal choices for the other two preferences we study in the design. For instance, we examine whether choices in a Complements block look more like optimal Complements choices than optimal Substitutes or Cobb-Douglas choices. If the correct induced preference is the winner, then the subject’s choice block reveals the preferences the subject was assigned in that block.

Determining whether a choice block is revealing allows us to distinguish between two conceptually distinct types of mistakes: mistakes in the implementation of the correct choice rule (e.g., due to noise or inconsistency), versus mistakes in the meta-decision of which choice rule to attempt to implement in the first place.

Of course, assessing which preference profile wins (i.e., is closest to the subject’s behavior) requires us to choose a measure of success. Our first and primary measure is the preference-relative CCEI score itself: we examine whether the preference-relative CCEI score in a block is *higher* when calculated relative to the *true* induced preferences in that block than it is when calculated relative to either of the other two preferences used in the design. To the degree this is true, the subject’s choices should be *more consistent* with the actual objective function than with other objective functions and therefore *reveal* those preferences. In the Appendix we consider additional metrics for robustness. These include Varian (1990)’s index (which aggregates budgetary adjustments by averaging instead of using a worst-case scenario), efficiency (the percent of optimal profit achieved on average), and average Euclidean distance (normalized by length of the budget frontier) between the demanded and optimal bundles. The results are almost identical regardless of which metric we use to assess substantive rationality. The robustness also extends when replacing the CCEI with Varian’s index (which, similarly to the CCEI, has a preference-agnostic extension) to determine as-if rationality in the first place.

4.2 The Rationality of Subjects

We now pose our main question: are subjects who are as-if rational actually *substantively rational*, or does as-if rationality rest on a different behavioral foundation? Following our pre-registration, we classify each of our subjects as follows:

- **As-If Rational (AIR):** We classify a subject as as-if rational if her choices are as-if rational in *each* of our main induced preference blocks. That is, if her (agnostic) CCEI score exceeds 0.8795 in every choice block.
- **Substantively Rational (SR):** We classify a subject as substantively rational if her choices correctly *reveal* induced preferences in each block (i.e., produce a higher CCEI score when computed relative to the block’s induced preferences than to the other induced preferences in the design).

Thus, here and throughout, we refer to a *block* as AIR or SR if choices in the block satisfy the criteria described above in Section 4.1. We refer to a *subject* as AIR or SR if they are *consistently* AIR or SR across all three blocks. In practice, we will be most interested in substantive types who are also classified as as-if rational and will sometimes simply refer to these subjects as “substantive.”

We find that a significant majority of subjects (about 67%, or two-thirds) can be classified as AIR subjects: their choices consistently come close to satisfying GARP under all three preference blocks. However, we find that only a minority (about 43%) of the AIR subjects are also consistently substantively rational, even using our highly forgiving definition of substantive rationality. About 57% of AIR subjects thus fail to use a choice rule that even noisily resembles optimal demands in at least one of the choice blocks. Note that this is based on a very permissive classification of substantive rationality: the rate of substantive rationality among AIR subjects falls to 21% using our stricter preference-relative CCEI criterion in all blocks.

Result 1 *Most subjects (about two-thirds) are consistently as-if rational, but the majority of these subjects are not consistently substantively rational, failing to even noisily maximize in at least one choice block.*

Using our conservative, noisy classification of SR, we can assemble a basic taxonomy of subjects. About 29% of all subjects are both AIR and substantively rational. Another 34% fail even AIR and can be classified as globally irrational. The most common “type” in the data, comprising 38% of subjects in our sample – are subjects who are simultaneously AIR and substantively irrational. As discussed in more detail in the next subsection, these subjects exhibit several signatures of rationality despite failing to consistently maximize. We therefore, following Simon (1976b) describe these subjects as ‘procedural rational.’ These subjects make decisions that are consistently internally coherent and predictable, arguably following a principled decision rule, but one that often fails to even noisily resemble maximizing behavior.

Once again, Figure 4 allows us to illustrate. Subjects shown in that figure are all as-if rational, with CCEI scores of 1 (or near 1) in every single induced choice block. However, as is clear from the figure, these subjects are nonetheless highly heterogeneous. To illustrate, we return to subjects 12 and 338: these are classified as substantively rational AIR subjects (the former, in the strictly substantive sense). Our procedure classifies the remaining subjects shown in Figure 4 as procedurally rational, consistently complying with AIR but not SR. Subject 143 makes near optimal choices in Substitutes and Cobb-Douglas but uses an internally consistent (and suboptimal) choice rule under Complements. Subject 140 chooses an optimal choice rule in only one choice block (Complements). Finally, subject 334, although always internally consistent and in near-compliance with GARP, *never* uses a substantively optimal choice rule.

As discussed above, Appendix B.1 shows our results are stable under different, pre-registered metrics for assessing preference revelation in a block.

Our main analysis is built around the CCEI given its prevalence in the literature. However, other indices for assessing AIR are available. As an exploratory analysis, we considered the robustness of results to redefining AIR based on Varian’s index (which instead averages over budgetary adjustments and is thus less sensitive to errors), and correspondingly using the preference-relative Varian’s index to assess preference revelation.¹⁷ Again, our conclusions are highly stable under this alternate measure:

¹⁷This requires adjusting the benchmark for rationality to the 95th percentile based on Varian’s index, so AIR may increase or decrease. Similarly, SR as measured by preference revelation may

about 73% of subjects would be classified as consistently as-if rational, with about 55% of these classified as procedural.

4.3 Procedural Rationality

The modal subject in our data is consistently as-if rational but not consistently substantively rational. As we will argue below, it is these subjects and their prevalence that, we will argue below, makes as-if rationality a poor diagnostic for preference revelation. These ‘procedurally rational’ subjects seem to have a motivation and capacity for internal consistency that exceeds their motivation and capacity to maximize their objectives.

4.3.1 Signatures of (procedural) rationality

Even though procedural types do not reliably maximize their objective functions, their choices nonetheless tend to comply with some key properties of the preferences we induce. By virtue of their GARP compliance, these procedural subjects tend to use decision rules that respect transitivity. Our experimental design allows us to show that these subjects also respect other rationality requirements given their induced preferences. In what follows, we compare the compliance with these properties across the three types of subjects in our experiment (non-AIR, substantively AIR and procedurally AIR), and as a benchmark, what we would expect from random decision-makers.

- **Stability across repetitions.** We repeat five of our budget lines within each choice block, allowing us to evaluate the extent of noisiness in those decisions. Most procedural subjects make fairly stable choices: when facing the same budget line twice within a block, their decisions are far more consistent than would be expected under random choice. Figure 5(a) shows CDFs of the Euclidean distances between repeat choices (normalized by length of the budget line) by our

increase or decrease, as the indices relative to each of the induced preferences change.

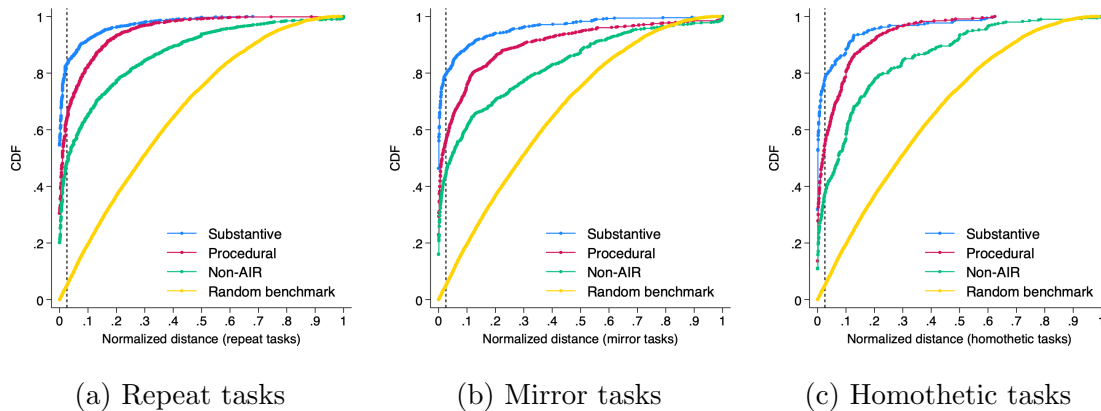


Figure 5: CDFs of normalized distance between choices, examining consistency across repeated budgets (a), symmetry of choices in mirrored budgets (b), and homotheticity of choices in scaled budgets (c), for substantively AIR, procedurally AIR, and non-AIR subjects. Distributions under simulated random choice are provided for reference.

rationality classification. Substantive types only slightly dominate procedural types in their degree of choice consistency.¹⁸

- **Symmetry.** The set of budgets tested in each block includes two pairs of budget lines that are mirror images of each other. If (x, y) is selected in one budget, then (y, x) should be selected in its mirrored counterpart (within the same induced-preference block). For each treatment, we compute the normalized Euclidean distance between (i) the choice (x, y) in one budget and (ii) the mirror $(\hat{y}, \hat{x})$ of the true choice $(\hat{x}, \hat{y})$ from its counterpart. Figure 5(b) plots CDFs of this measure for procedural types and the various benchmark groups. The results are strikingly similar to those from the left panel: procedural types show high degrees of symmetry, at levels only modestly below those of substantive types.
- **Homotheticity.** We included in our experimental design some budget lines in which the price ratio is held constant but income varies. In particular, if (x, y) is selected from the budget line given by intercepts $(100, 60)$, then within the same induced-preference treatment, $(x/2, y/2)$ should be selected in its smaller counterpart $(50, 30)$. We calculate the normalized Euclidean distance between

¹⁸Only 5% of random choices in a pair of repeated budget sets result in a normalized distance of 0.0258 (the vertical line in Figure 5), yet a high proportion of procedural subjects (about 60%) exhibit a choice discrepancy smaller than this benchmark. The analysis excludes repetitions of the $(60, 60)$ budget set within Substitutes, as there is no unique maximizer in that case.

the choice (x, y) in the larger budget and the bundle $(2\hat{x}, 2\hat{y})$, where $(\hat{x}, \hat{y})$ is the choice from the smaller budget. Figure 5(c) shows, again, strikingly similar results to those for consistency and symmetry. In Appendix B.2, we provide the corresponding figure for the $(60, 60)$, $(80, 80)$, and $(100, 100)$ budget lines; the same patterns hold, with even slightly stronger compliance with homotheticity when there is no price differential.

- **Law of demand.** Like the procedural subjects from Figure 4, the majority of procedural subjects obey the law of demand by demanding shares $x/(x + y)$ that are non-increasing in the price ratio of x to y . Among procedural subjects, the Spearman correlation between these quantities are *not* significantly positive at the 5% level in 97.4% of choice blocks. The correlations are significantly negative for 86.1% of these subjects in the Substitutes and Cobb-Douglas treatments.¹⁹ The strict law of demand seems to be quite ingrained in procedural subjects: 46.4% have a significantly negative Spearman correlation even in the Complements treatment, where it should be zero.

Result 2 *AIR subjects who fail to qualify as substantively rational nonetheless show signatures of rationality. Their choice behavior tends to be reasonably transitive, tends to obey the law of demand, tends to be consistent across repeated budget sets, and tends to reflect the symmetry and homotheticity inherent in our induced preferences.*

4.3.2 Heuristics

Clearly, procedurally AIR subjects maintain their apparent rationality by using choice rules with desirable properties (i.e. transitivity, consistency etc.) that are often unmoored from their objective functions. As is easily verified by inspection of the raw data in the Online Appendix, most procedurally rational subjects do this by *reusing* choice rules that are optimal in some other contexts, exporting them to contexts in

¹⁹This is much higher than the 58.7% that would be expected under random decision-making from our budget lines, but lower than the 97.8% displayed by substantively AIR subjects. Random choice has a tendency towards negative Spearman correlations, because the larger (smaller) is p_x relative to p_y , the greater is the relative proportion of the budget frontier lying below (above) the diagonal $x = y$; see Becker (1962).

which they do not belong. In other words, subjects use rules of thumbs or heuristics that they re-apply across multiple decision blocks, failing to sensitively respond to changes in their objectives.

To show this formally, we expand upon the preference-revelation methods we deployed when classifying substantive rationality. In particular, in each task we compare subjects’ choices to a large family of possible *archetype* demand patterns, and study, for each subject and each induced-preference block, to which archetype the subject’s choices are closest. Specifically, for each budget set in a choice block we calculate the normalized Euclidean distance between the subject’s choice and each archetype’s choice, and examine which archetype minimizes the average distance over that treatment. By examining how often this closest archetype is repeated across choice blocks, we characterize how often a subject re-uses a choice rule heuristically.²⁰

	<i>1</i>	<i>2</i>	<i>3</i>
Substantively AIR	0.9%	10.4%	88.7%
Procedurally AIR	17.2%	68.9%	13.9%
Non-AIR	20.9%	38.1%	41.0%

Table 1: Percent of subjects typed as using 1, 2 or 3 distinct choice archetypes in the induced preference treatments, by attributed rationality.

Table 1 summarizes the results by listing the number of distinct choice archetypes (1, 2 or 3) subjects use across the three induced-preference treatments. Unsurprisingly, we find that almost all subjects typed as substantively AIR use three distinct rules, sensitively changing their choice rule from objective function to objective function.²¹ Subject 12 in Figure 4 is such a subject. For procedurally AIR subjects, almost all

²⁰We deliberately selected a large set of potential archetypes (a decision that works against us finding evidence of repeated use of choice rules), including the optimal choice rules for Complements, Substitutes and Cobb-Douglas, random choice rules that are as-if rational, and three other deterministic choice patterns (inspired by some subjects’ actual choices): purchasing only X , purchasing only Y , and purchasing only the more expensive good. In order to identify haphazard behavior that is nonetheless as-if rational, we use 1,000 randomly generated choice patterns that fall in the top 5th percentile of the standard CCEI; we compute a subject’s average distance in a choice block to each of the 1,000 as-if-but-random choice patterns, and conservatively define the distance to the ‘random’ archetype as the minimal distance among these 1,000 numbers.

²¹The small number of substantive subjects classified as using fewer than 3 archetypes is due to our use of a slightly different classification procedure for this exercise: using a different, distance-based measure and based on a large set of archetypes instead of the three optimal choice rules.

(92%) of the demand patterns come closest to one of the three induced choice rules; but they are more prone to applying the *wrong* induced choice rule. By contrast to substantive subjects, the vast majority of procedurally rational subjects (about 81%) use only one (à la Subject 140 in Figure 4) or two (à la Subject 143) unique choice rules across tasks, repeating some choice rule at least once.²²

Result 3 *Most subjects who display procedural rationality, do so by using heuristics, re-using a choice rule in domains in which they are suboptimal.*

The percentages of subjects classified under each archetype, within each treatment and each rationality grouping, are provided in Appendix B.3.

4.4 As-If As a Diagnostic for Substantive Rationality

Given this gap between rates of SR and AIR, to what degree does as-if rationality serve as an accurate diagnostic sufficient statistic for substantive rationality, and therefore a reliable tool for identifying welfare-revealing (and therefore welfare-relevant) choices? To study this, Figure 6 plots kernel densities of the standard (preference agnostic) CCEI scores for choice blocks (i) that reveal induced preferences and that we therefore classified substantively rational and (ii) choice blocks that fail to reveal underlying induced preferences and that we therefore classify as substantively irrational. Unsurprisingly, CCEI scores for revealing, substantively-rational blocks are very high, with a strong mode at 1, and 90% rising above the 95th percentile random benchmark. However, our main finding is that CCEI scores are also very high for choice blocks that fail to reveal preferences: 74% of substantively *irrational* blocks, which fail to reveal preferences, can also be classified as as-if rational. Indeed, blocks that correctly reveal the induced preference are only 16 percentage points more likely to exceed the 95th percentile benchmark for rationality than blocks which do not. As our findings on procedural rationality suggest, subjects seem to have a drive to make

²²As discussed in DellaVigna (2009, p. 353), there is evidence that “*Individuals facing a complex choice may simplify it by diversifying excessively across the options.*” This tendency may relate to the particular choice of heuristics in our setting as well. The Complements and Cobb-Douglas patterns, for which choices are more centrally located in the feasible set, are commonly repeated heuristics. The Substitutes pattern, which conflicts with the tendency to diversify, is less used overall.

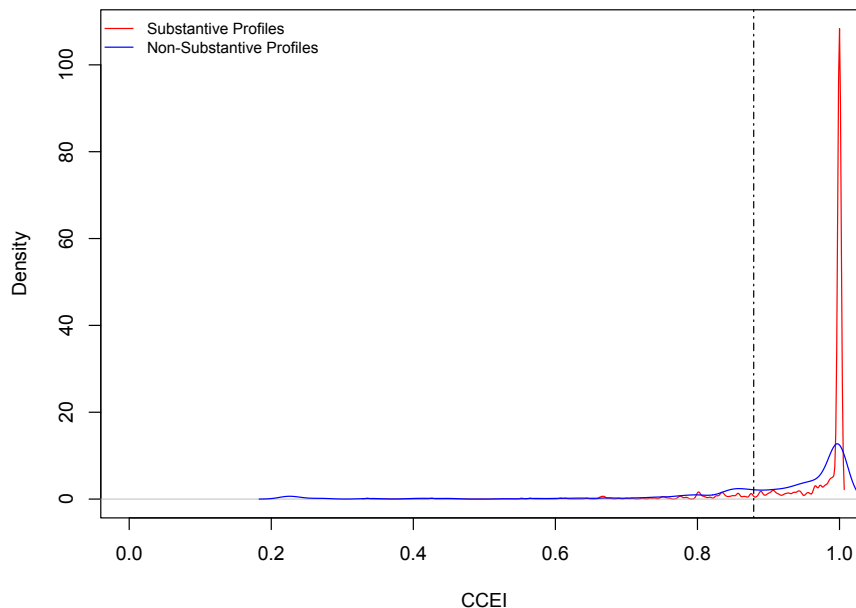


Figure 6: Kernel density estimates of the (agnostic) CCEI index for choice blocks that can (in red) and cannot (in blue) be classified as substantively rational.

internally-consistent, GARP-compliant decisions *even when failing to maximize their objectives*.

Result 4 *Most choice blocks that fail substantive rationality by failing to reveal underlying preferences (74%) are nonetheless highly internally consistent and therefore as-if rational.*

Because as-if rationality arises reliably on a basis distinct from maximization in our data, it serves as a poor diagnostic tool for detecting substantive rationality in a choice block, and therefore unreliable evidence that choice is welfare-revealing. This is easy to see using standard analysis from diagnostic test evaluation (Pepe, 2003). While as-if rationality is highly *sensitive*, generating a high true positive rate of 90%, it has extremely poor *specificity*, generating a very low true negative rate of only 26%. As a result the *balanced accuracy rate* of as-if rationality (the average of sensitivity and specificity) as a classifier of substantive rationality is only 58% – not much better

than a coin flip.²³

Result 5 *As-if rationality is a very weak tool for assessing substantive rationality, with an accuracy only marginally better than chance.*

4.5 Bounded Rationality

Our data provides some insight into why so many subjects deviate from substantive rationality. First, our data strongly suggests that procedurally rational subjects economize on effort relative to substantively rational subjects. To show this we examine *decision time* (the time spent formulating a choice), a widely used measure of effort in the experimental literature. Among those we classify as substantively rational, the median subject spent a total of 28.0 minutes formulating and submitting their choices in the main treatments. Relative to this, the median procedurally rational subject spent considerably less time – only 22.1 minutes. Thus subjects economize significantly on decision costs by using procedurally rational choice rules. The median non-AIR subject spends as much time as procedural subjects or even slightly more (22.8 minutes), and yet makes remarkably less consistent choices.²⁴

Second, there is significant evidence that many procedurally rational subjects are *aware* that their choices are not substantively rational. After each choice block we measure subject’s “cognitive certainty” (Enke and Graeber, 2023). Specifically, we ask subjects to estimate for us the percentage likelihood that over the course of the block, their chosen bundles were no more than 5 X-tokens, and no more than

²³Here we apply standard tools for evaluating diagnostic tests, where a test (here, as-if rationality) is used as a signal for some underlying condition (here, substantive rationality). *Sensitivity*, or the *true positive rate*, is the probability that the test is positive when the condition is present—in our setting, the fraction of substantively rational choices that as-if rationality correctly flags as rational (90%). *Specificity*, or the *true negative rate*, is the probability that the test is negative when the condition is genuinely absent—the fraction of substantively *irrational* choices that as-if rationality correctly declines to flag as rational (26%). A test can have high sensitivity yet low specificity, meaning it rarely misses true cases but also generates many false positives, so a positive result is weak evidence that the condition is present. *Balanced accuracy* is the simple average of sensitivity and specificity $((90\% + 26\%)/2 = 58\%)$; it summarizes overall classification performance while correcting for class imbalance, and a value near 50% indicates performance close to chance.

²⁴In a similar vein, procedural and non-AIR subjects look nearly identical in terms of their CRT scores; by contrast, the distribution of CRT score of substantive subjects is higher than that for non-substantive subjects (Wilcoxon rank sum test, p-value 0.0280).

5 Y-tokens, away from the payoff maximizing choices. On average, confidence is 76.4% for substantive types but significantly smaller for procedural types (65.7%, p -value < 0.001). This awareness of deviations from optimality opens the significant possibility that many subjects are avoiding maximization in a deliberate effort to economize on effort.

Putting these two pieces of evidence together, our data suggests that subjects are often consciously and maybe even deliberately using (and re-using) procedures in order to economize on the cognitive effort costs (i.e., the complexity) of substantively rational choice. In other words, subjects show indications of making consciously boundedly rational choices.

Result 6 *Procedural rationality requires significantly less effort than substantive rationality, and most procedurally rational subjects are aware that the resulting choices do not maximize their objective functions.*

4.6 Incentives and Comprehension

Finally, we include two additional findings that further aid in interpreting our results. First, as we discuss in Section 3, we included a free-form quiz testing subjects’ understanding of or attention to each of the three payoff functions used in the experiment (responses had to be typed into an empty field, *not* selected from a list of possibilities). The vast majority (85%) of subjects were able to answer all of these questions correctly on their first attempt, and virtually all (98.5%) do so by their second attempt – strongly suggesting that our results are not driven by confusion about or inattention to the payoff functions themselves. Indeed, when we drop subjects who failed to answer these questions on their first try, the results are nearly identical (see Appendix B.1 for details). Confusion about the nature of the payoff function is therefore not the driver of our results. Second, we randomly varied the size of the incentives in the experiment: the “higher incentives” condition featured payoffs that were five times larger than the “lower incentives” condition. As Figure 7 shows, this change in incentives did not improve decision making: if anything, AIR

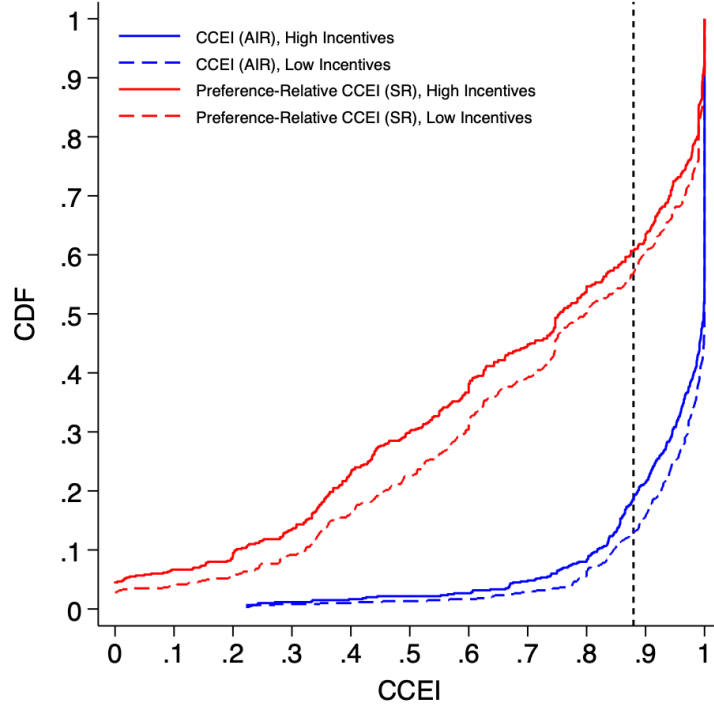


Figure 7: Distribution of CCEI and preference-relative CCEI for each choice block, by higher or lower incentive treatment.

and SR (as measured by the CCEI) are actually slightly *lower* at higher incentives.²⁵ This suggests that weak incentives are very unlikely to be driving our main results.

Result 7 *Procedural rationality remains more prevalent than substantive rationality when we (i) condition on subjects who demonstrated strong understanding of the payoff formulae or (ii) increase incentives by a factor of five.*

²⁵In terms of AIR, 63% of subjects in the higher incentive condition are AIR, versus 70% in the lower incentive condition. 26.5% of all subjects in the higher incentive condition are classified as substantive, versus 31% in the lower incentive condition. In terms of procedural rationality, 36.5% of all subjects are classified as procedural in the higher incentive treatment, versus 39% in the lower incentive treatment.

5 Discussion

We use novel induced preference experiments to study the relationship between as-if rationality (a tendency to make choices rationalizable by some utility function) and substantive rationality (a tendency to actually maximize welfare in choice). We find that these two core types of rationality are only weakly linked in our data. Although the vast majority of subjects satisfy as-if rationality, only a minority of these consistently maximize the induced objective functions we assign them. Instead, many subjects are procedurally rational, using (and re-using) internally consistent choice rules that mimic the properties of utility functions but that do not consistently optimize subjects' own welfare. As a result, as-if rationality (AIR) is a poor diagnostic for substantive rationality (SR), i.e., welfare revealing behavior.

Although our experiment was not designed to comprehensively identify *when* we should expect procedural rationality, our data gives us some clues. Results reported in Section 4.5 suggest that procedural subjects perform worse on cognitive reflection tasks and tend to make their decisions more quickly. This may suggest that procedural rationality is particularly likely to govern choice when decision makers are cognitively unsophisticated, distracted, inattentive or time constrained. Another clue is that subjects tend to use procedural heuristics more often when the optimal rule prescribes more extreme and less “interior” behaviors: the rate at which procedural subjects' choice patterns come closest to the optimal rule falls from 76.8% (Complements) to 58.9% (Cobb-Douglas) to 27.8% (Substitutes) as prescribed behavior approaches the corners of the budget line; see Table 5 in the Appendix. This suggests that caution and misleading hedging intuitions may be a driver of procedural rationality so that procedural rationality is more likely when optimality requires extreme behavior. Indeed, the heuristics procedural subjects reuse in the experiment are far more likely to be interior rules (Complements, 44.6%, or Cobb-Douglas, 49.2%) than the corner rule, Substitutes (3.1%). This reinforces the possibility that caution may be an important driver of procedural choice.

These findings may come as little surprise to researchers in the Samuelsonian tradition (Samuelson, 1938) who have long interpreted revealed preferences as positive and descriptive and emphasized that such preferences are not necessarily diagnostic of

welfare. Indeed, our findings provide a rare piece of direct evidence in favor of this caution. On the other hand, our results may be more surprising (and useful) to applied researchers tasked with inferring welfare from choice to inform policy. There, the rise of behavioral economics has generated a challenge to traditional welfare economics: how to identify choices that contain welfare-relevant information (a “welfare relevant domain”) in the face of the many heuristics, biases and errors documented in the literature. The growing “behavioral welfare economics” literature has explored and made use of criteria for identifying this domain. For this, our results provide a useful caution: apparent maximization of *some* utility function (AIR) is (at least in some economically important settings) an unreliable tool for identifying welfare relevant (SR) choices.

Because of this, our findings also reinforce the importance of this literature’s search for tools and criteria other than choice consistency to identify the welfare-relevant domain (see Rees-Jones (2024) for a recent synthesis). This literature searches for substantive criteria for when a choice is “suspect,” including characterization failure (Bernheim, 2016); *dominance*, since a dominated choice is a mistake whatever the chooser’s preferences, and a surprising share of real choices prove dominated (Bhargava et al., 2017); identifying assumptions about frames and defaults (Goldin and Reck, 2020)—building on the formal analysis of frame-dependent choice (Salant and Rubinstein, 2008)—or active-choice designs that strip their influence (Carroll et al., 2009); *process* data such as attention and response times (Caplin et al., 2011; Reutskaja et al., 2011; Rubinstein, 2007), or limited attention inferred from choice itself (Masatlioglu et al., 2012); a decision-maker’s own confidence in the quality of her decisions (Enke and Graeber, 2023); and benchmarks recovered from choices made under better conditions—full information, deliberation, or demonstrated competence (Allcott and Taubinsky, 2015; Bronnenberg et al., 2015; Handel and Kolstad, 2015; Ambuehl et al., 2022), an identification strategy reviewed by Handel and Schwartzstein (2018)—or estimated directly through a reduced-form “sufficient-statistics” approach (Mullainathan et al., 2012; Chetty, 2015).

References

- AFRIAT, S. (1967): “The Construction of Utility Functions from Expenditure Data,” *International Economic Review*, 8, 67–77.
- (1973): “On a system of inequalities in demand analysis: an extension of the classical method,” *International Economic Review*, 14, 460–472.
- AHN, D., S. CHO, D. GALE, AND S. KARIV (2014): “Estimating Ambiguity Aversion in a Portfolio Choice Experiment,” *Quantitative Economics*, 5, 195–223.
- ALLCOTT, H. AND D. TAUBINSKY (2015): “Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market,” *American Economic Review*, 105, 2501–2538.
- AMBUEHL, S., B. D. BERNHEIM, AND A. LUSARDI (2022): “Evaluating Deliberative Competence: A Simple Method with an Application to Financial Choice,” *American Economic Review*, 112, 3584–3626.
- ANDREONI, J. AND J. MILLER (2002): “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism,” *Econometrica*, 70, 737–753.
- APESTEGUIA, J. AND M. A. BALLESTER (2015): “A Measure of Rationality and Welfare,” *Journal of Political Economy*, 123, 1278–1310.
- ARIELY, D., G. LOEWENSTEIN, AND D. PRELEC (2003): ““Coherent Arbitrariness”: Stable Demand Curves Without Stable Preferences,” *The Quarterly Journal of Economics*, 118, 73–106.
- ARRIETA, G. R. AND K. NIELSEN (2024): “Procedural Decision-Making in the Face of Complexity,” *Unpublished manuscript*.
- BANERJEE, S. AND J. H. MURPHY (2006): “A simplified test for preference rationality of two-commodity Choice,” *Experimental Economics*, 9, 67–75.
- BECKER, G. (1962): “Irrational Behavior and Economic Theory,” *Journal of Political Economy*, 70, 1–13.

- BERNHEIM, B. D. (2016): “The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics,” *Journal of Benefit-Cost Analysis*, 7, 12–68.
- (2021): “In Defense of Behavioral Welfare Economics,” *Journal of Economic Methodology*, 28, 385–400.
- BERNHEIM, B. D. AND A. RANGEL (2007): “Toward Choice-Theoretic Foundations for Behavioral Welfare Economics,” *American Economic Review*, 97, 464–470.
- (2009): “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *The Quarterly Journal of Economics*, 124, 51–104.
- BERNHEIM, B. D. AND D. TAUBINSKY (2018): “Behavioral Public Economics,” in *Handbook of Behavioral Economics: Foundations and Applications*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Amsterdam: Elsevier, vol. 1, 381–516.
- BESHEARS, J., J. J. CHOI, D. LAIBSON, AND B. C. MADRIAN (2008): “How Are Preferences Revealed?” *Journal of Public Economics*, 92, 1787–1794.
- BHARGAVA, S., G. LOEWENSTEIN, AND J. SYDNOR (2017): “Choose to Lose: Health Plan Choices from a Menu with Dominated Options,” *The Quarterly Journal of Economics*, 132, 1319–1372.
- BRONNENBERG, B. J., J.-P. DUBÉ, M. GENTZKOW, AND J. M. SHAPIRO (2015): “Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium,” *The Quarterly Journal of Economics*, 130, 1669–1726.
- CAMARA, M. (2024): “Computationally Tractable Choice,” *Unpublished Manuscript*.
- CAPLIN, A., M. DEAN, AND D. MARTIN (2011): “Search and Satisficing,” *American Economic Review*, 101, 2899–2922.
- CARROLL, G. D., J. J. CHOI, D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2009): “Optimal Defaults and Active Decisions,” *The Quarterly Journal of Economics*, 124, 1639–1674.
- CHETTY, R. (2015): “Behavioral Economics and Public Policy: A Pragmatic Perspective,” *American Economic Review*, 105, 1–33.

- CHOI, S., R. FISMAN, D. GALE, AND S. KARIV (2007): “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” *American Economic Review*, 97, 1921–1938.
- CHOI, S., S. KARIV, W. MÜLLER, AND D. SILVERMAN (2014): “Who Is (More) Rational?” *American Economic Review*, 104, 1518–1550.
- CONLISK, J. (1996): “Why Bounded Rationality?” *Journal of Economic Literature*, 34, 669–700.
- DE CLIPPEL, G. AND K. ROZEN (2023): “Relaxed Optimization: How Close Is a Consumer to Satisfying First-Order Conditions?” *The Review of Economics and Statistics*, 105, 883–898.
- (2024): “Bounded Rationality in Choice Theory: A Survey,” *Journal of Economic Literature*, 62, 995–1039.
- DEAN, M. AND D. MARTIN (2016): “Measuring Rationality with the Minimum Cost of Revealed Preference Violations,” *Review of Economics and Statistics*, 98, 524–534.
- DELLAVIGNA, S. (2009): “Psychology and Economics: Evidence from the Field,” *Journal of Economic Literature*, 47.
- DEMBO, A., S. KARIV, M. POLISSON, AND J. K.-H. QUAH (2021): “Ever Since Allais,” *Journal of Political Economy*, *forthcoming*.
- ECHENIQUE, F., S. LEE, AND M. SHUM (2011): “The Money Pump as a Measure of Revealed Preference Violations,” *Journal of Political Economy*, 119, 1201–1223.
- ENKE, B. AND T. GRAEBER (2023): “Cognitive Uncertainty,” *The Quarterly Journal of Economics*, 138, 2021–2067.
- ENKE, B., T. GRAEBER, AND R. OPREA (2025): “Complexity and Hyperbolic Discounting,” *Journal of the European Economic Association*, 23, 1838–1867.
- FISMAN, R., S. KARIV, AND D. MARKOVITS (2007): “Individual Preferences for Giving,” *American Economic Review*, 97, 1858–1876.

- FREDERICK, S. (2005): “Cognitive Reflection and Decision Making,” *Journal of Economic Perspectives*, 19, 25–42.
- GOLDIN, J. AND D. RECK (2020): “Revealed-Preference Analysis with Framing Effects,” *Journal of Political Economy*, 128, 2759–2795.
- HALEVY, Y. AND G. MAYRAZ (2024): “Identifying Rule-Based Rationality,” *The Review of Economics and Statistics*, 106, 1369–1380.
- HALEVY, Y., D. PERSITZ, AND L. ZRILL (2018a): “Parametric Recoverability of Preferences,” *Journal of Political Economy*, 126.
- (2018b): “Parametric Recoverability of Preferences,” *Journal of Political Economy*, 126, 1558–1593.
- HANDEL, B. AND J. SCHWARTZSTEIN (2018): “Frictions or Mental Gaps: What’s Behind the Information We (Don’t) Use and When Do We Care?” *Journal of Economic Perspectives*, 32, 155–78.
- HANDEL, B. R. AND J. T. KOLSTAD (2015): “Health Insurance for “Humans”: Information Frictions, Plan Choice, and Consumer Welfare,” *American Economic Review*, 105, 2449–2500.
- HOUTMAN, M. AND J. MAK (1985): “Determining all Maximal Data Subsets Consistent with Revealed Preference,” *Kwantitatieve Methoden*, 19, 89–104.
- KHAW, M. W., Z. LI, AND M. WOODFORD (2021): “Cognitive imprecision and small-stakes risk aversion,” *Review of Economic Studies*, 88, 1979–2013.
- KŐSZEGI, B. AND M. RABIN (2007): “Mistakes in Choice-Based Welfare Analysis,” *American Economic Review*, 97, 477–481.
- (2008): “Revealed Mistakes and Revealed Preferences,” in *The Foundations of Positive and Normative Economics: A Handbook*, ed. by A. Caplin and A. Schotter, Oxford: Oxford University Press, 193–209, `VERIFY` page range and editor spelling against the printed volume.

- LANIER, J., B. MIAO, J. K.-H. QUAH, AND S. ZHONG (2022): “Intertemporal Consumption with Risk: A Revealed Preference Analysis,” *The Review of Economics and Statistics*, 1–45.
- MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): “Revealed Attention,” *American Economic Review*, 102, 2183–2205.
- MULLAINATHAN, S., J. SCHWARTZSTEIN, AND W. J. CONGDON (2012): “A Reduced-Form Approach to Behavioral Public Finance,” *Annual Review of Economics*, 4, 511–540.
- NIELSEN, K. AND J. REHBECK (2022): “When Choices Are Mistakes,” *American Economic Review*, 112, 2237–2268.
- OPREA, R. (2020): “What Makes A Rule Complex,” *American Economic Review*, 110, 3913–3951.
- (2024): “Decisions Under Risk are Decisions Under Complexity,” *American Economic Review*, 114, 3789–3811.
- PAYNE, J. W., J. R. BETTMAN, AND E. J. JOHNSON (1993): *The Adaptive Decision Maker*, Cambridge University Press.
- PEPE, M. S. (2003): *The Statistical Evaluation of Medical Tests for Classification and Prediction*, New York: Oxford University Press.
- REES-JONES, A. (2024): “Behavioral Incentive Compatibility and Empirically Informed Welfare Analysis: An Introductory Guide,” *Journal of Economic Perspectives*, 38, 155–174.
- REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): “Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study,” *American Economic Review*, 101, 900–926.
- ROSE, H. (1958): “Consistency of Preference: The Two-Commodity Case,” *The Review of Economic Studies*, 25, 124–125.

- RUBINSTEIN, A. (2007): “Instinctive and Cognitive Reasoning: A Study of Response Times,” *The Economic Journal*, 117, 1243–1259.
- RUBINSTEIN, A. AND Y. SALANT (2012): “Eliciting Welfare Preferences from Behavioural Data Sets,” *The Review of Economic Studies*, 79, 375–387.
- SALANT, Y. AND A. RUBINSTEIN (2008): “(A,f): Choice with Frames,” *The Review of Economic Studies*, 75, 1287–1296.
- SAMUELSON, P. A. (1938): “A Note on the Pure Theory of Consumer’s Behaviour,” *Economica*, 5, 61–71.
- SIMON, H. A. (1955): “A Behavioral Model of Rational Choice,” *The Quarterly Journal of Economics*, 69, 99–118.
- (1976a): “From Substantive to Procedural Rationality,” in *25 Years of Economic Theory*, 65–86.
- (1976b): “From Substantive to Procedural Rationality,” *25 Years of Economic Theory*, 65–86.
- SUGDEN, R. (2004): “The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences,” *American Economic Review*, 94, 1014–1033.
- (2018): *The Community of Advantage: A Behavioural Economist’s Defence of the Market*, Oxford: Oxford University Press.
- TOPLAK, M., R. F. WEST, AND K. E. STANOVICH (2014): “Assessing miserly information processing: An expansion of the Cognitive Reflection Test,” *Thinking & Reasoning*, 20, 147–168.
- VARIAN, H. (1982): “The Nonparametric Approach to Demand Analysis,” *Econometrica*, 50, 945–973.
- VARIAN, H. R. (1990): “Goodness-of-fit in optimizing models,” *Journal of Econometrics*, 46, 125–140.

ZAME, B., B. TUNGODDEN, E. SORENSEN, S. KARIV, AND A. CAPPELEN (2023):
“ Linking Social and Personal Preferences: Theory and Experiment,” *Journal of
Political Economy*, *forthcoming*.

Appendices

A Collection of budget sets tested

The last five budgets, denoted with a *-superscript in the table below, are those added to the initial set of 25 generated as described in Section 3.2.

Budget line #	x-Intercept	y-Intercept
1	60	20
2	40	60
3	60	60
4	80	60
5	60	100
6	20	70
7	70	40
8	60	70
9	70	80
10	100	70
11	80	20
12	40	80
13	80	60
14	80	80
15	100	80
16	20	90
17	90	40
18	60	90
19	90	80
20	100	90
21	100	20
22	40	100
23	100	60

24	80	100
25	100	100
26*	60	100
27*	20	70
28*	80	20
29*	50	30
30*	60	60

B Additional Details for Section 4

B.1 Number of revealing blocks and robustness checks

Among AIR subjects, the number of revealing blocks is highly stable across metrics, as seen in Table 3 below.²⁶

	CCEI	Varian	Earnings	Distance
3 revealing blocks	43.2%	44.4%	45.5%	42.1%
2 revealing blocks	33.5%	33.1%	32.0%	38.0%
1 revealing blocks	21.1%	20.3%	20.7%	18.4%
0 revealing blocks	2.3%	2.3%	1.9%	1.5%

Table 3: Number of revealing blocks among AIR subjects, under each metric.

The results are also similar when restricting attention to the approximately 85% of subjects who correctly complete – on their first attempt – the quiz checking for comprehension of earnings formulas. Conditioning on this sample, 69% are AIR

²⁶Though conceptually different, one can check that budgetary adjustments underlying the preference-relative CCEI and Varian’s index coincide with the achieved percentage of earnings when it comes to Perfect Complements and Perfect Substitutes. They differ for other preferences, such as Cobb-Douglas, and this can potentially impact the outcome of the three-horse races in all treatments. As seen from comparing the Varian and Earnings columns in Table 3 (both of which aggregate through averaging), no substantial differences materialize here.

subjects. Overall in this screened sample, 30.9% are substantively AIR, 30.9% are irrational, and 38.2% are procedurally AIR, with the number of revealing blocks again stable across metrics as seen in Table 4 below.

	CCEI	Varian	Earnings	Distance
3 revealing blocks	44.7%	45.5%	46.8%	44.3%
2 revealing blocks	33.2%	33.2%	31.9%	37.0%
1 revealing blocks	20.0%	19.2%	19.6%	17.5%
0 revealing blocks	2.1%	2.1%	1.7%	1.3%

Table 4: Number of revealing blocks, under each metric, for AIR subjects who answer the quiz checking comprehension of earnings formulas correctly on their first attempt.

Nearly all subjects (98.5%) complete this quiz (which requires typing each answer into a field, not selecting from a multiple choice list) correctly by the second try.

B.2 Additional homotheticity figure

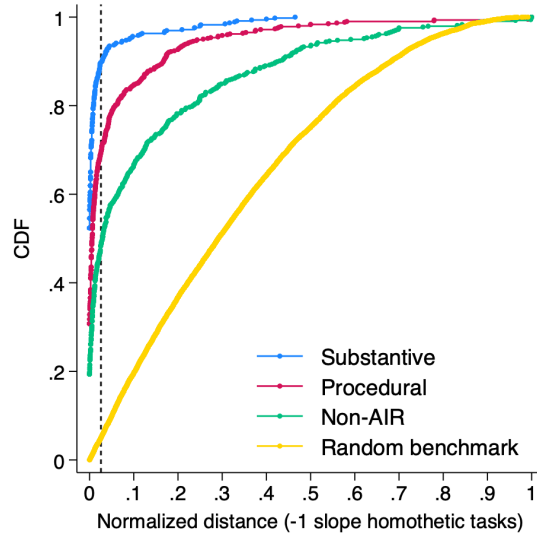


Figure 8: Counterpart to Figure 5(b) for the three budget sets with slope -1 .

B.3 Details on Heuristics

We create a family of choice archetypes that includes not only the ‘correct’ demand patterns, but also some mistaken ones inspired by some subjects’ scatterplots, and a large array of purely random patterns that are nonetheless as-if rational, to separate out haphazard behavior that only appears rational:

- (i) Equalizing the amount of each good that is purchased, as predicted by the Complements preference we induce (*‘comp’*);
- (ii) Purchasing only the cheaper good, or any bundle when prices are equal, as predicted by the Substitutes preference we induce (*‘sub’*);
- (iii) Purchasing the bundle at the midpoint of the budget line, as predicted by the Cobb-Douglas preference we induce (*‘c-d’*);
- (iv) Purchasing only good X (*‘just x ’*);
- (v) Purchasing only good Y (*‘just y ’*);
- (vi) The reverse of (ii): purchasing only the more expensive good, or any bundle when prices are equal (*‘rev-sub’*).²⁷
- (vii) 1,000 as-if-but-random choice patterns, each generated by selecting, for each budget line tested, one bundle uniformly at random (*‘rand-asif’*), and then confirming the entire pattern has CCEI exceeding the 95th-percentile threshold.

²⁷We do not include a reverse of Cobb-Douglas since the mirrored choices would lie strictly within the budget frontier, and not be available selections in our experiment. Reverse complements would simply be the complements pattern itself.

	Substantively AIR Subjects		
	<i>Complements</i>	<i>Substitutes</i>	<i>Cobb-Douglas</i>
<i>comp</i>	100%	0.9%	10.4%
<i>sub</i>	—	92.2%	—
<i>c-d</i>	—	3.5%	87.0%
<i>just x</i>	—	—	—
<i>just y</i>	—	0.9%	—
<i>rev-sub</i>	—	—	—
<i>rand-asif</i>	—	2.6%	2.6%
	Procedurally AIR Subjects		
	<i>Complements</i>	<i>Substitutes</i>	<i>Cobb-Douglas</i>
<i>comp</i>	76.8%	21.2%	29.1%
<i>sub</i>	1.3%	27.8%	2.7%
<i>c-d</i>	16.6%	42.4%	58.9%
<i>just x</i>	2.0%	1.3%	1.3%
<i>just y</i>	1.3%	2.0%	0.7%
<i>rev-sub</i>	—	—	—
<i>rand-asif</i>	2.0%	5.3%	7.3%
	Non-AIR Subjects		
	<i>Complements</i>	<i>Substitutes</i>	<i>Cobb-Douglas</i>
<i>comp</i>	50.0%	6.0%	10.5%
<i>sub</i>	0.8%	35.1%	1.5%
<i>c-d</i>	18.7%	34.3%	61.9%
<i>just x</i>	0.8%	—	—
<i>just y</i>	—	0.8%	—
<i>rev-sub</i>	9.0%	—	0.8%
<i>rand-asif</i>	20.9%	23.9%	25.4%

Table 5: Nearest choice archetype in each choice block in terms of average normalized Euclidean distance, by rationality grouping.

C Earlier Experiment

An earlier version of this paper featured a slightly different experiment. In January 2026 we pre-registered and ran the new experiment when revising the paper to respond to questions and concerns from early readers and to substantially expand the size of our dataset. The main differences in the design are:

- The sample size in the earlier experiment was about a third of the current size. The earlier experiment, which unlike the current version was not pre-registered, was conducted via Prolific in July 2023.
- There was only a single incentive condition in the earlier design, contrary to the two incentive conditions (higher/lower) here.
- The earlier design did not have a quiz to directly test comprehension of the payoff formulas.
- The earlier design was significantly longer: it had an additional block (on top of Complements, Substitutes, and Cobb Douglas). That 4th block tested natural risk preferences: if subjects pick a bundle (x, y) then the earnings rule gave x with probability $1/2$ and y with probability $1/2$. Both designs utilize the same collection of 30 budget sets in each block, with the order of blocks randomized.
- Like the current design, the earlier design included a quiz testing understanding of the interface, but the incentive for answering correctly was 50 cents smaller than here.
- The current design incorporates small refinements to the interface and instructions to improve subjects' understanding.

Using the same analysis, the two experiments reach very similar conclusions, and there is an even slightly higher degree of procedural rationality in the 2023 data (with about 62% of AIR subjects classified as procedural, in contrast to about 57% here).

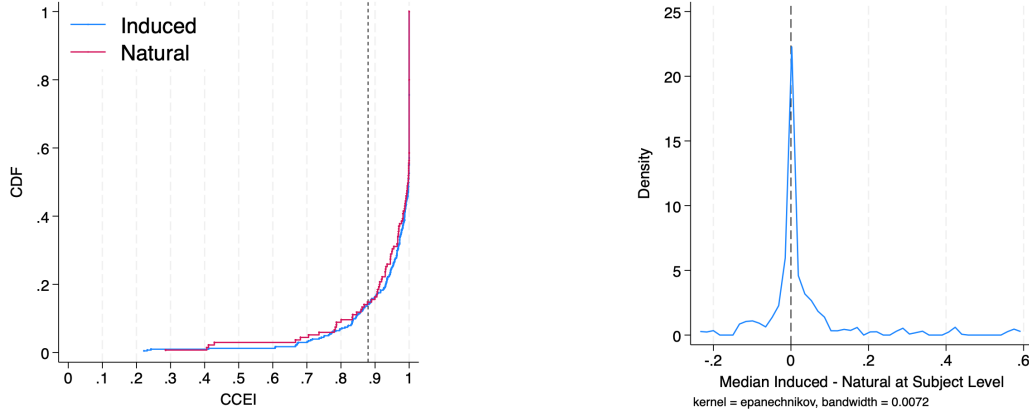


Figure 9: Distributions of CCEI scores in the left panel; and the difference between subject-wise median CCEI scores in natural vs. the induced preference blocks (all pooled together) in the right panel. *Notes: The left panel plots empirical CDFs; the right panel plots the kernel density.*

As discussed in Footnote 16, the earlier experiment can help compare measured as-if rationality in induced preference tasks with measured as-if rationality in standard natural-preference tasks. The left panel of Figure 9 plots empirical CDFs of preference-agnostic CCEI scores for subjects in the 2023 experiment, calculated at the subject-block level for the natural preference (Risk) block and the induced preference blocks. The distributions are very similar, though CCEI scores in induced treatments are slightly higher. At the subject level, the right hand panel compares each subject's CCEI from the natural-preference treatment with their median CCEI over the three induced-preference treatments, plotting a kernel density estimate of the distribution of subject-wise differences. These differences are tightly distributed around zero, suggesting that individual subjects are very similarly rational in each type of task.

Online Appendix to *As If*

de Clippel, Oprea and Rozen

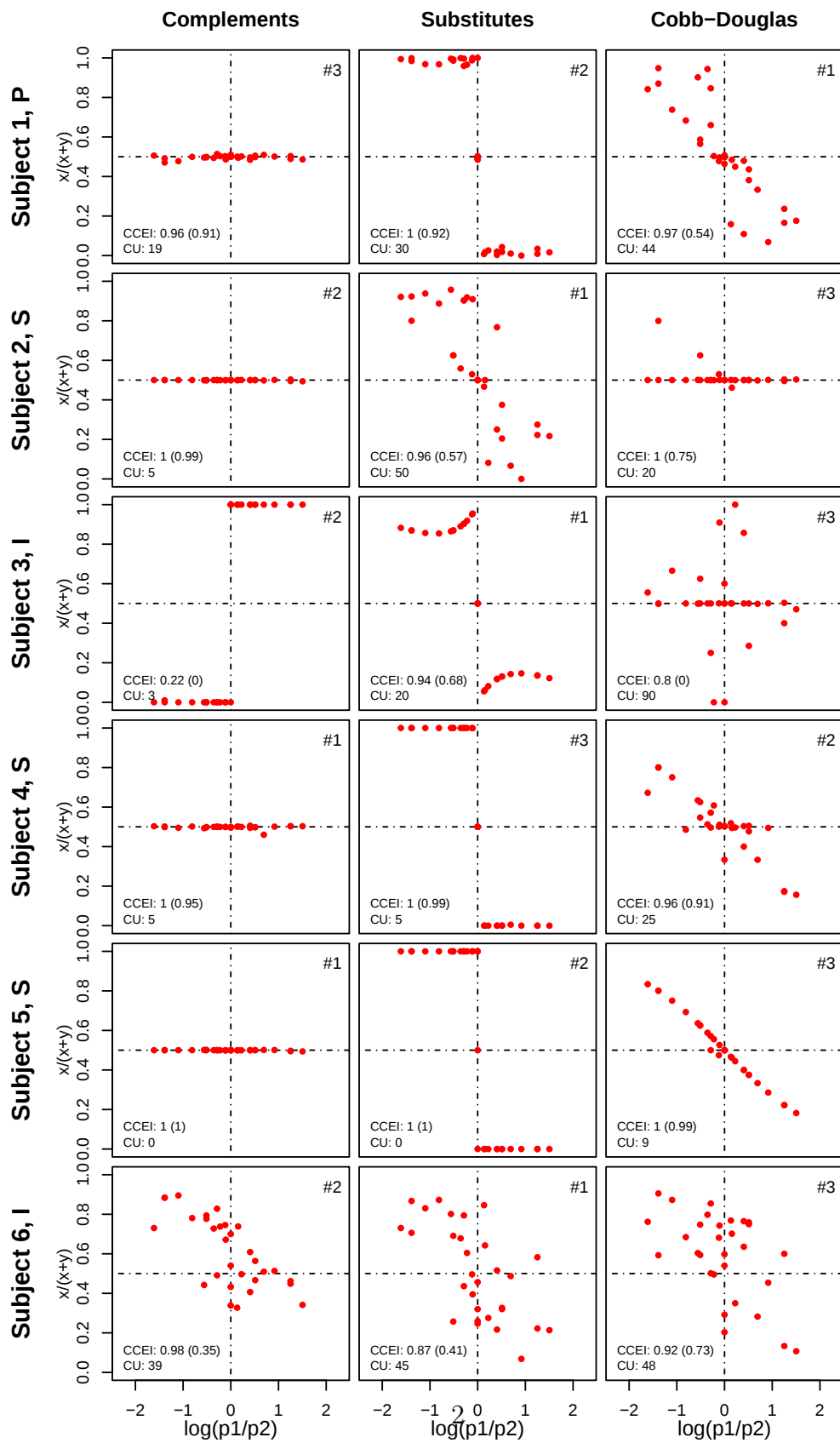
This Online Appendix contains:

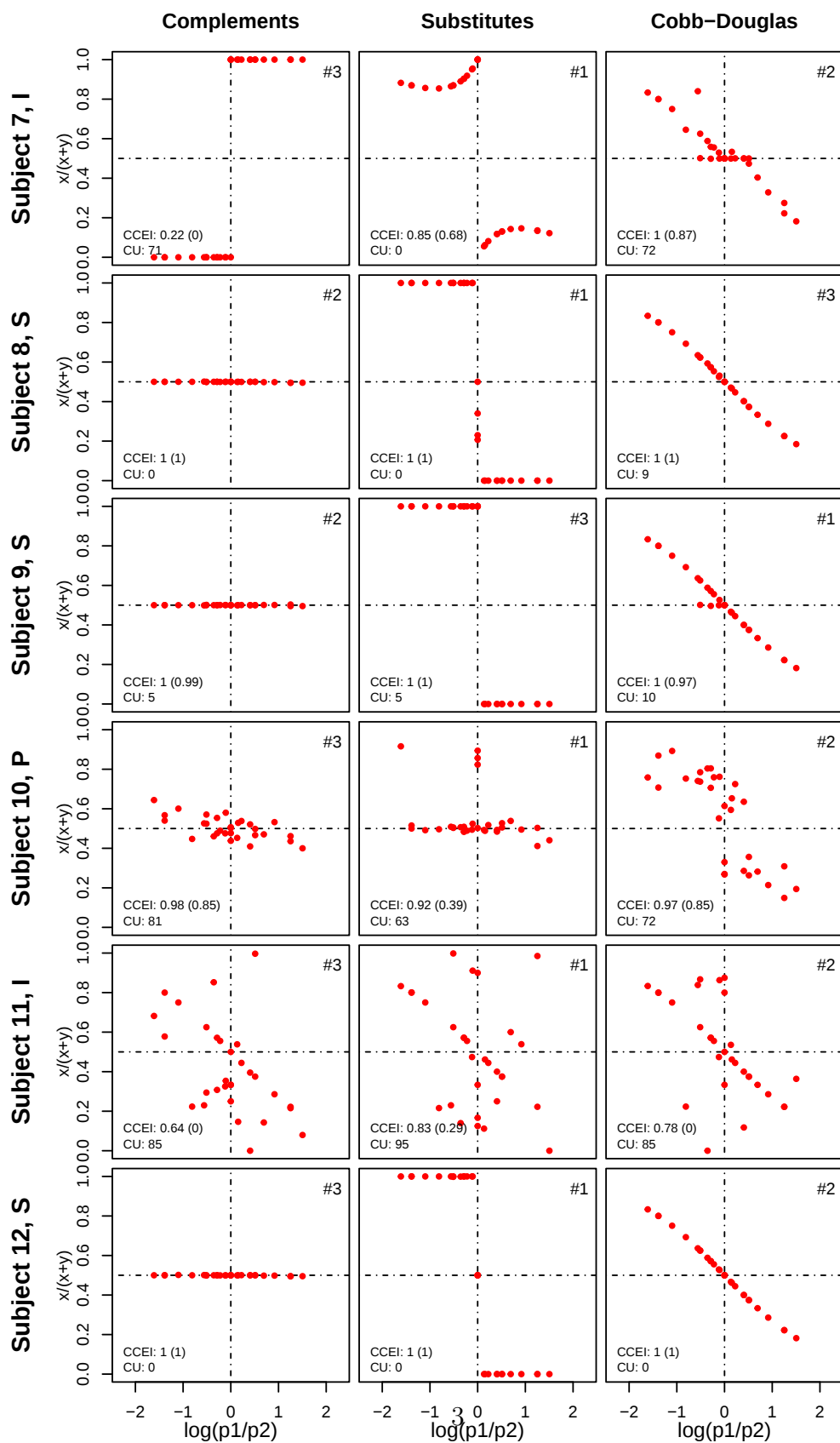
- Scatterplots with subject-level data (OA.1); and
- Screenshots of the experimental interface (OA.2).

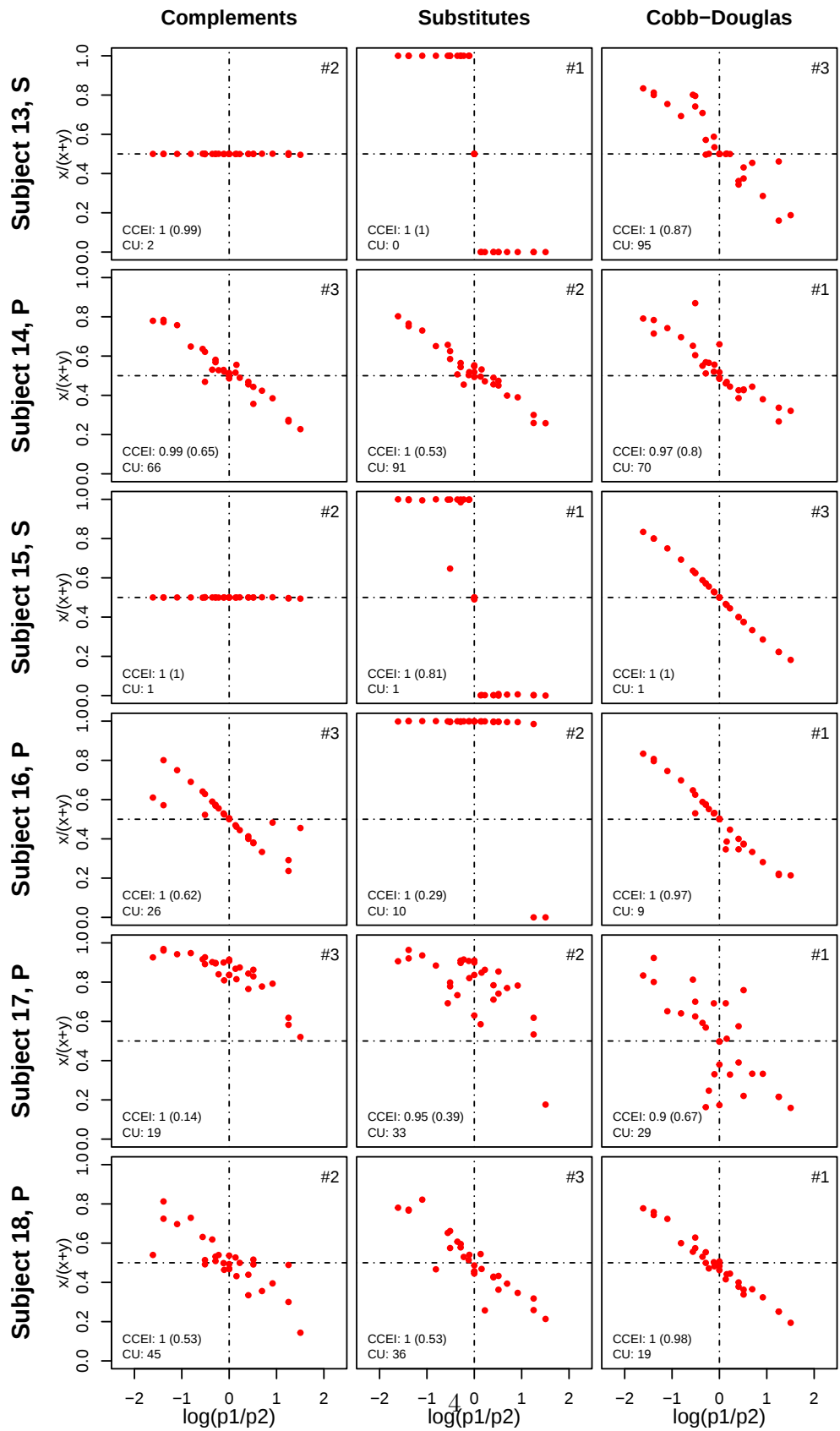
OA.1 Plots of demands

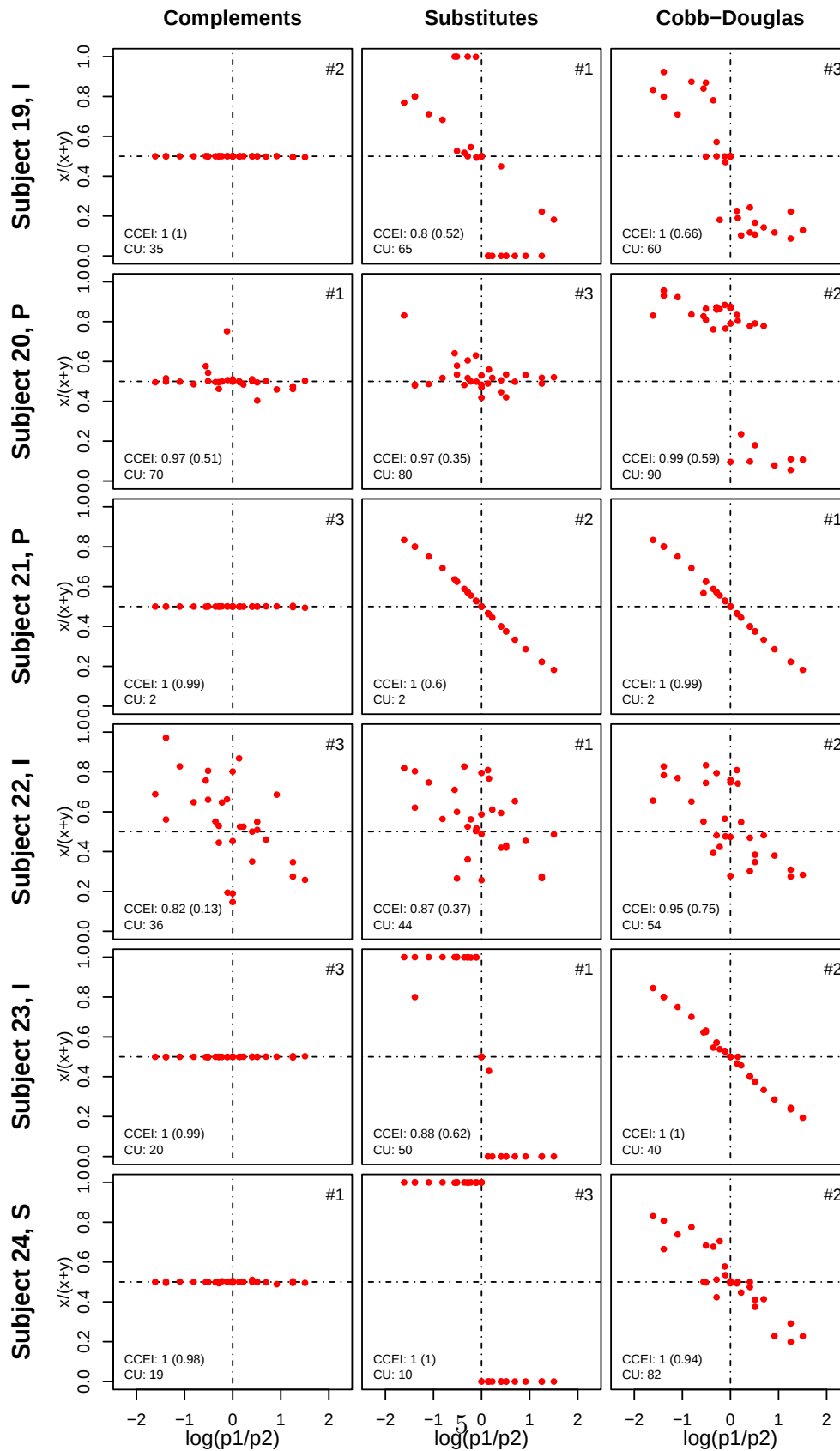
The following is raw data for the 400 subjects. There are four columns, labeled by the treatment. Each row is a subject, and the subject's designation (as Substantively AIR (S), Procedurally AIR (P), or Non-AIR/Irrational (I)) appears following the Subject ID in the row header.

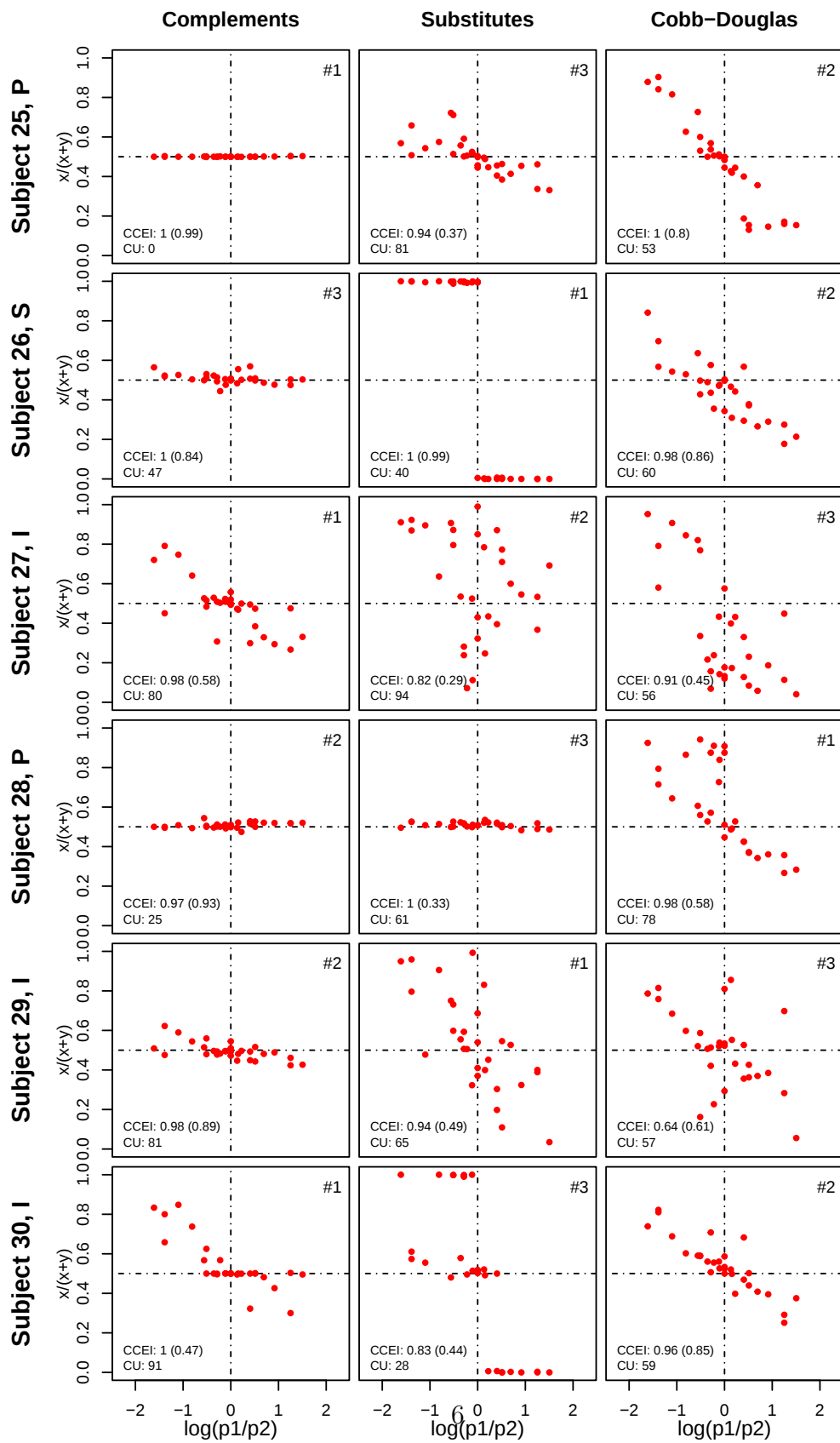
In addition to plotting the demand share $x/(x+y)$ against the log of the price ratio p_1/p_2 , each panel has the following information. In the bottom left corner there are two rows of numbers. In the top row, the first number (before the parentheses) is the standard, preference-agnostic CCEI score. The second (within parentheses) is the CCEI score calculated relative to the induced preferences for the task. Below these CCEI scores, the CU number reflects the subject's cognitive uncertainty, which was asked at the end of each treatment: it is 100 minus the certainty level expressed regarding being close (within 5 X-tokens and 5 Y-tokens) to the best choices in that treatment. In the upper right hand corner of each panel is a number (#1, #2, #3) identifying the order in the subject faced that treatment.

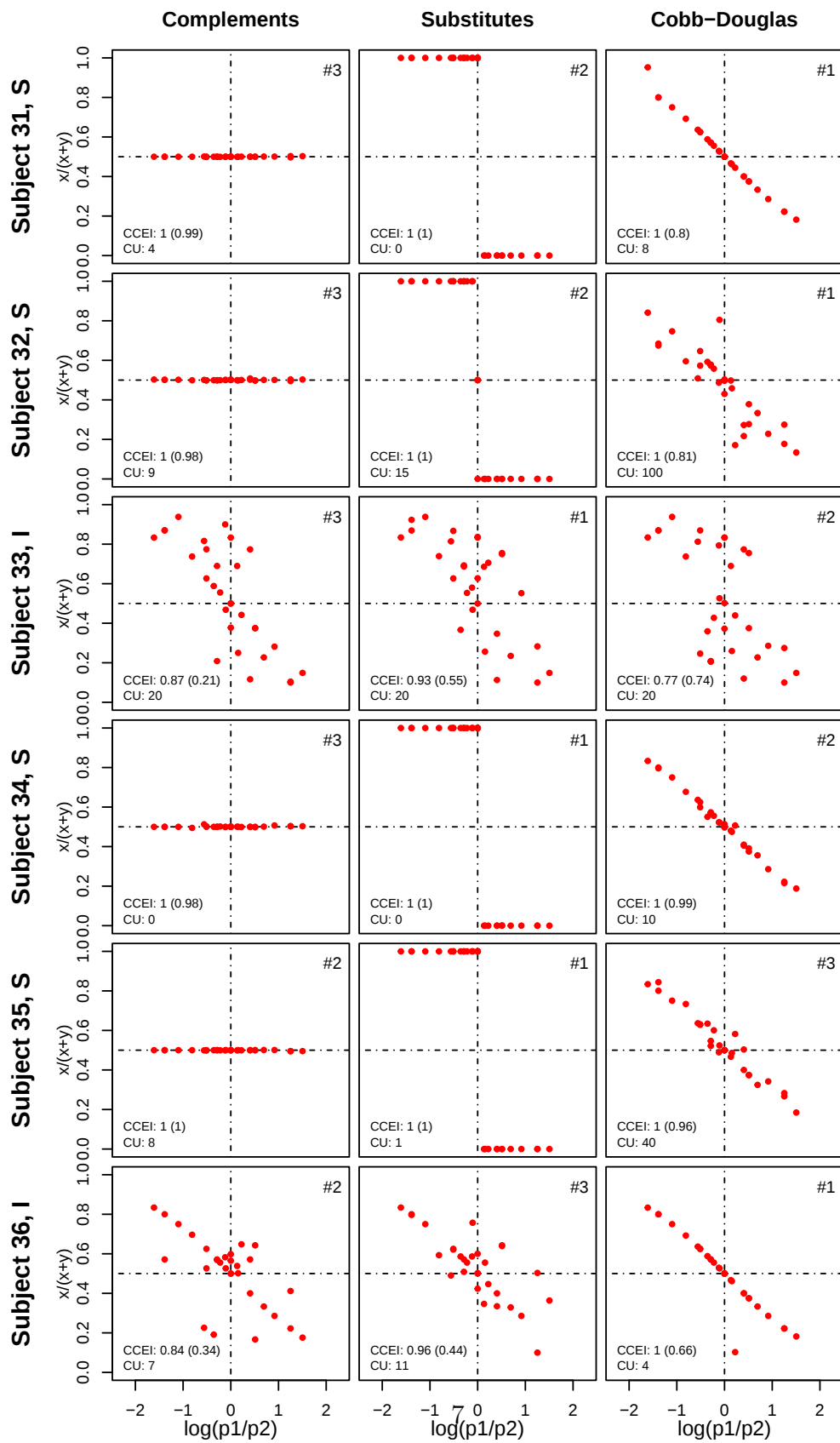


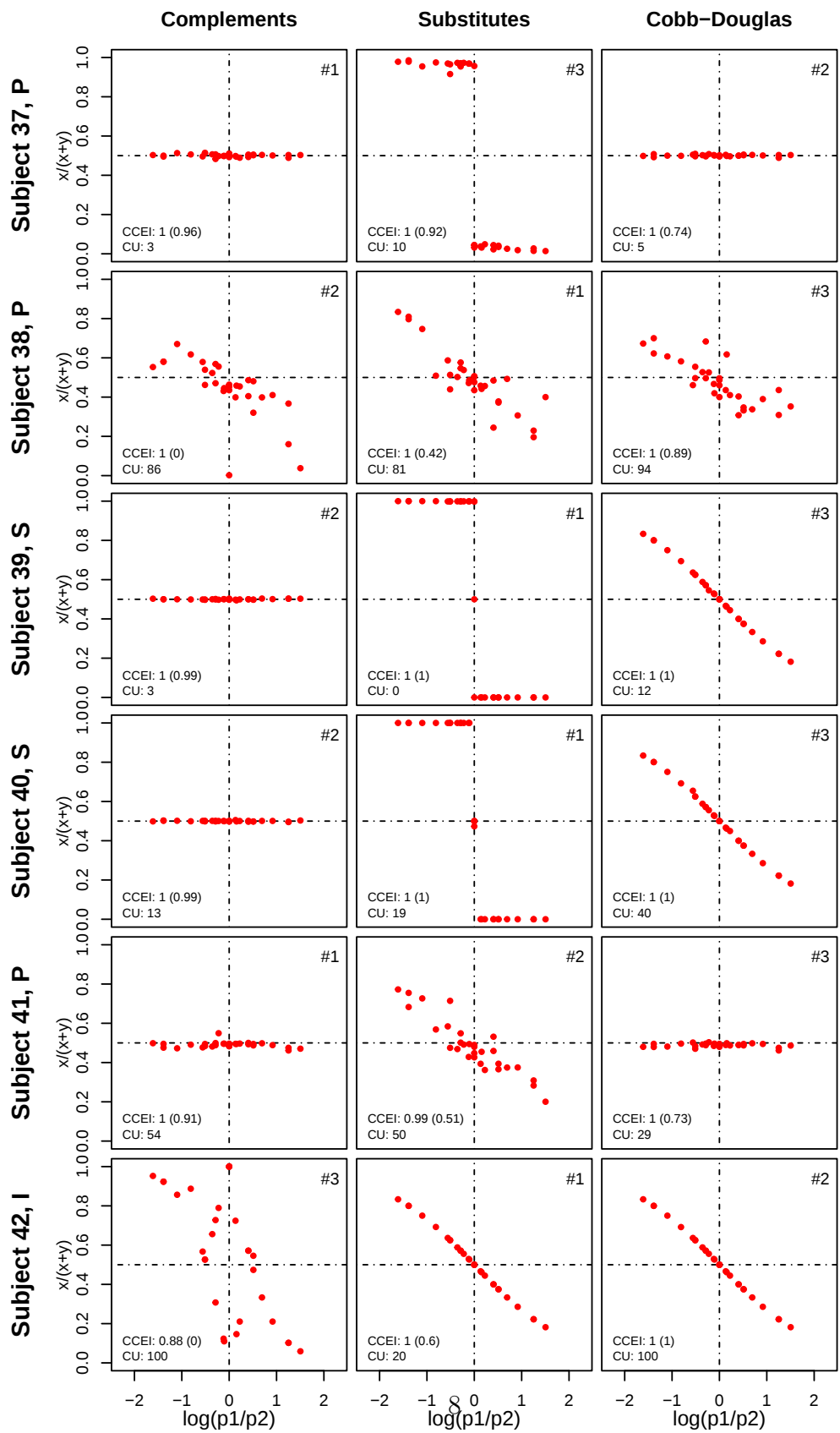


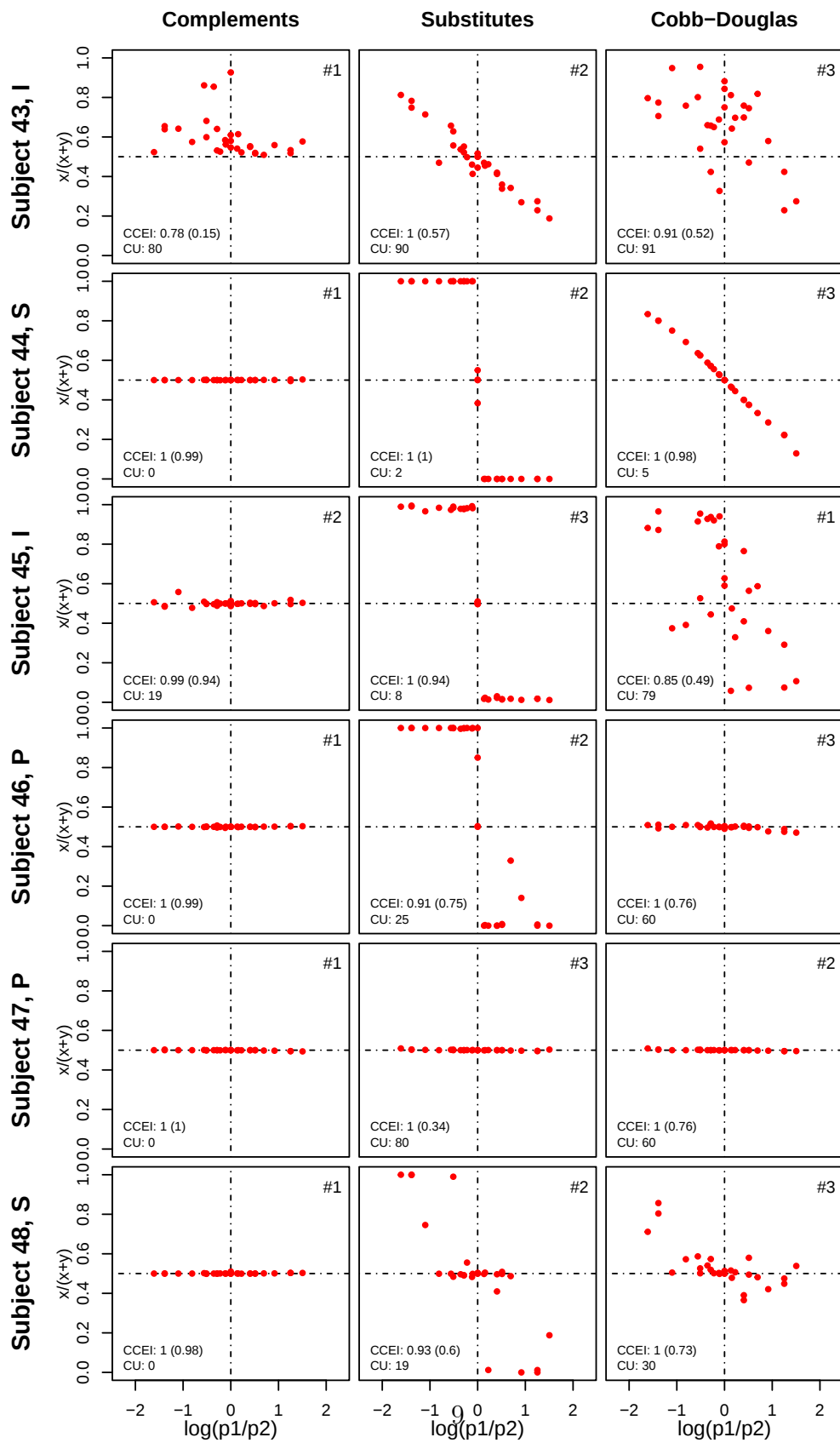


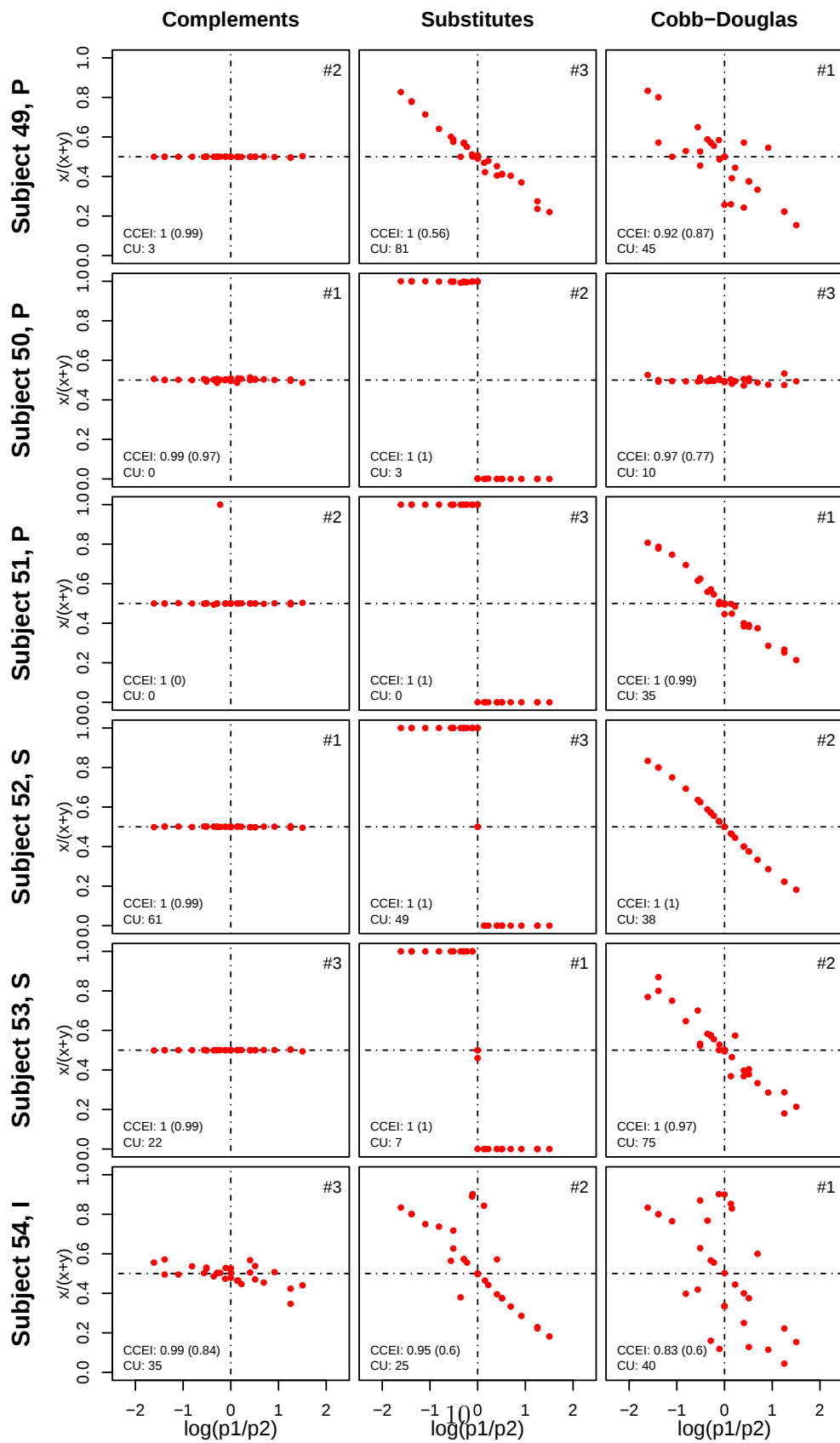


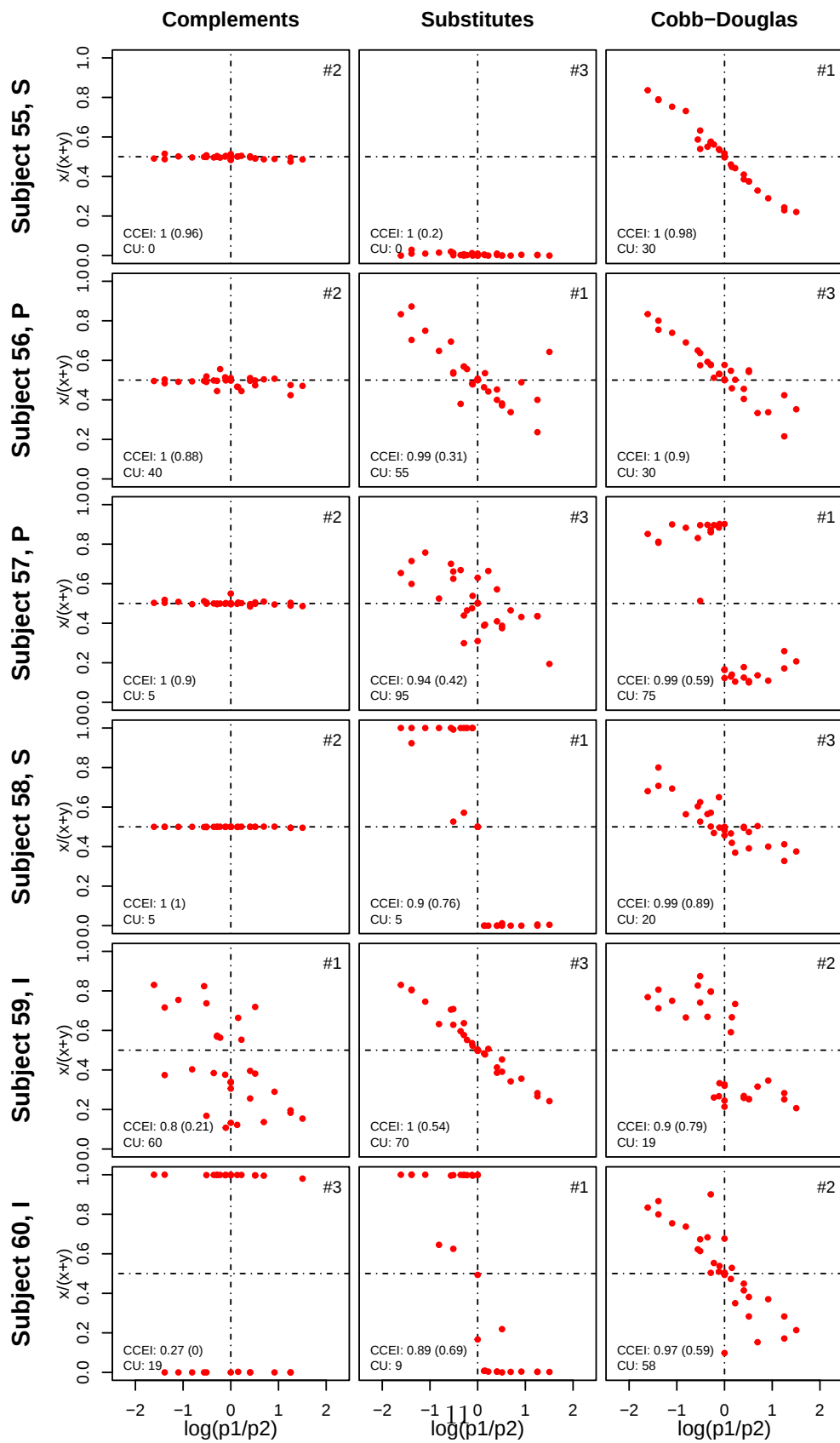


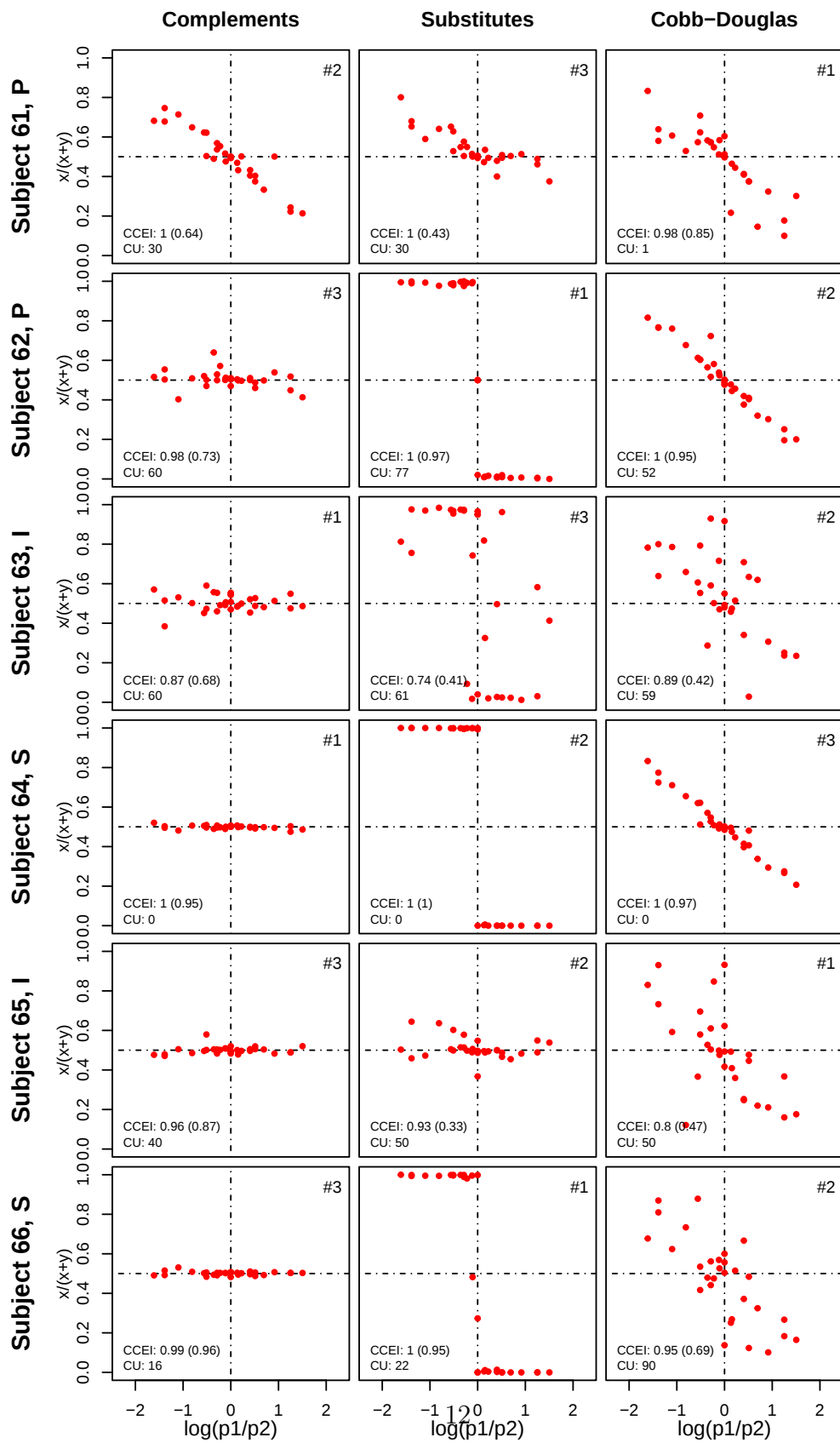


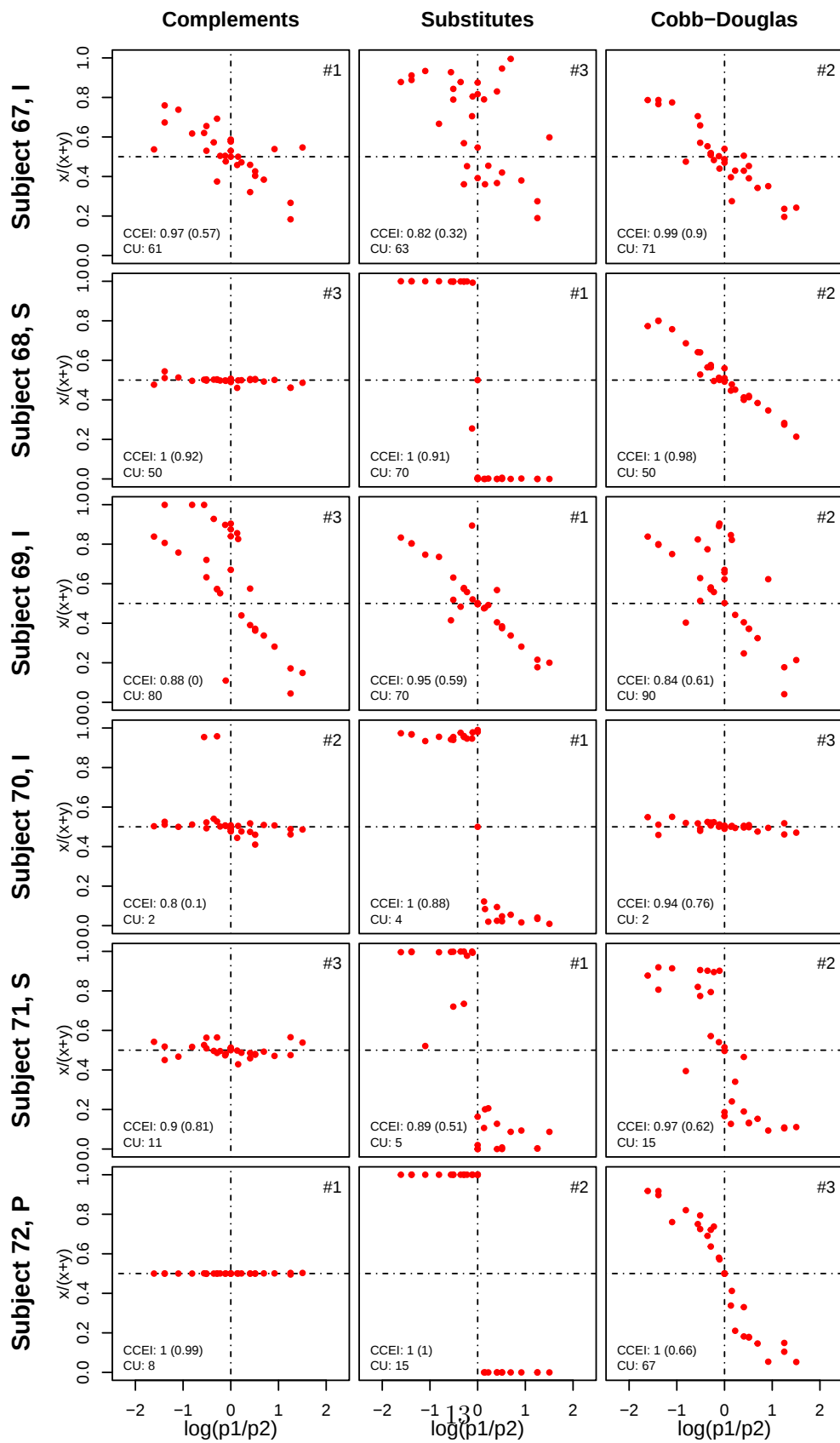


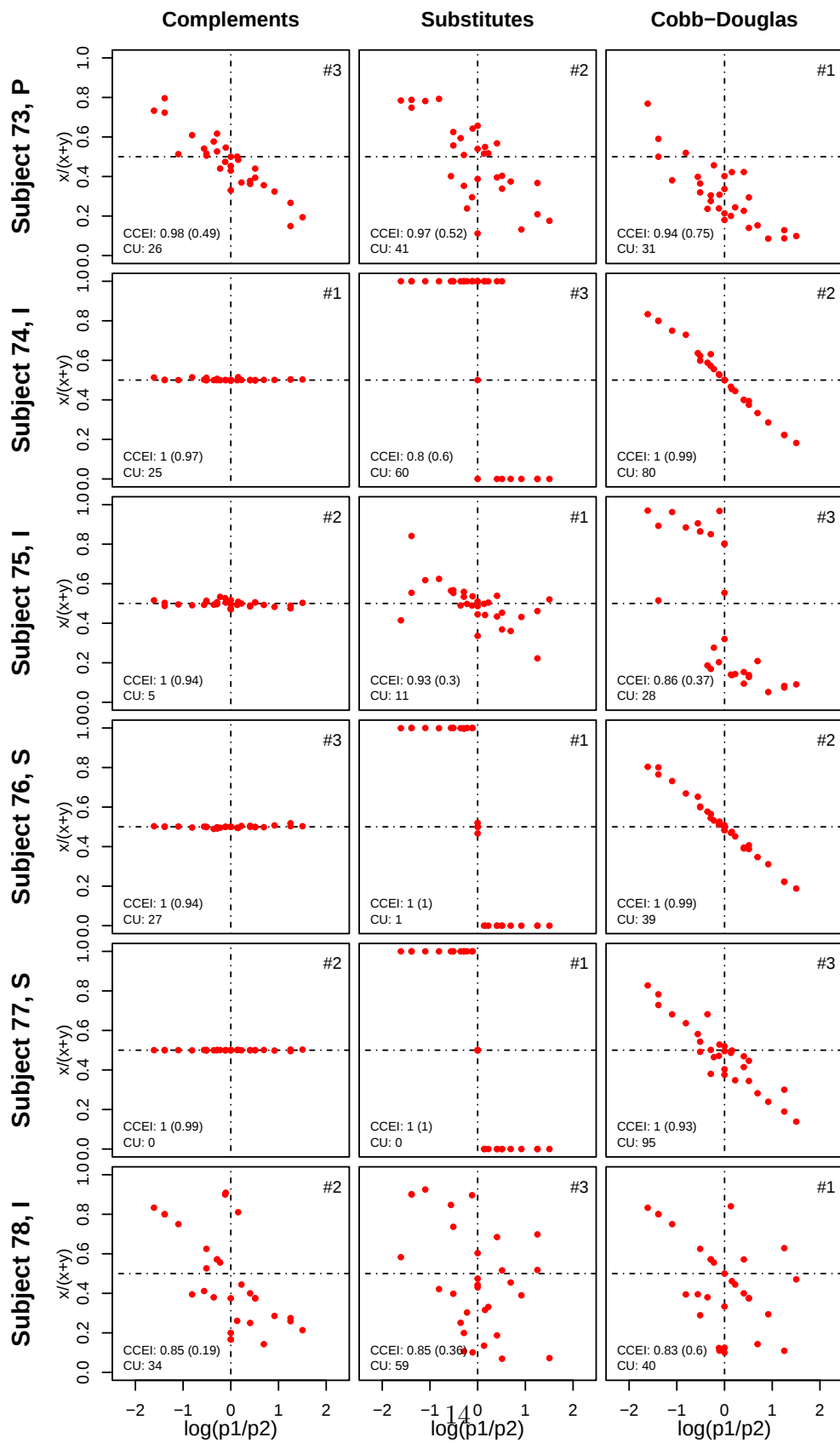


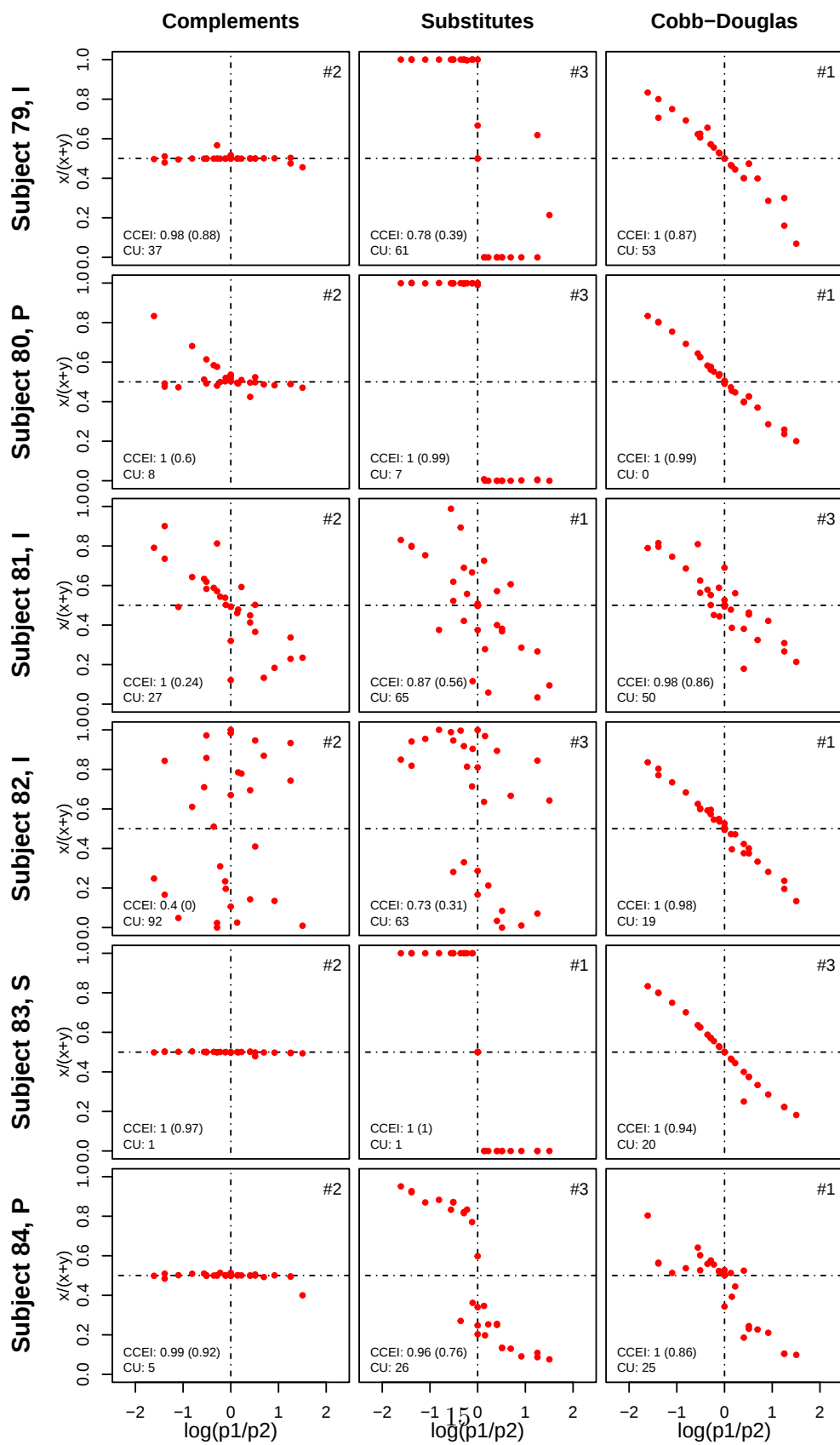


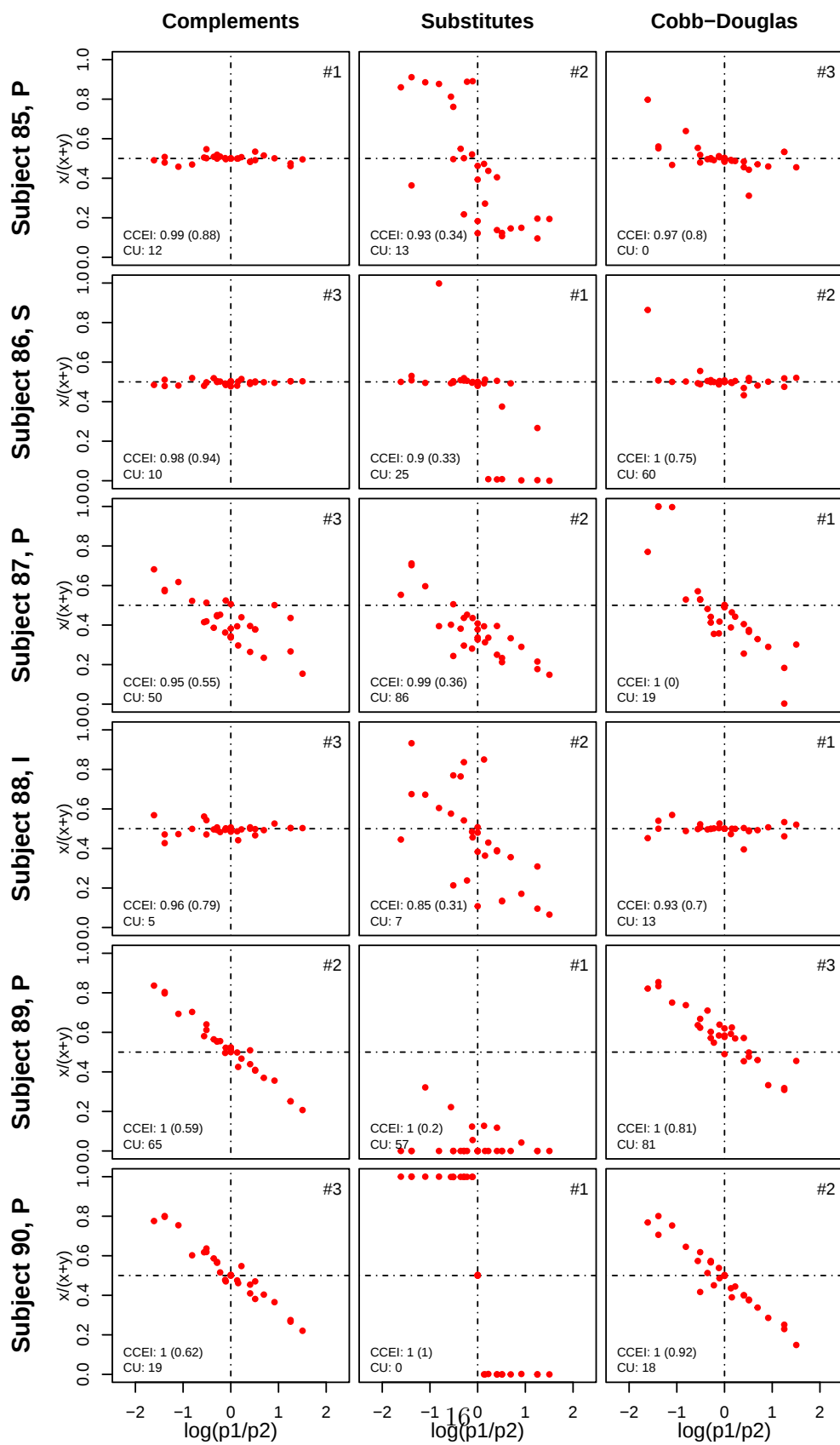


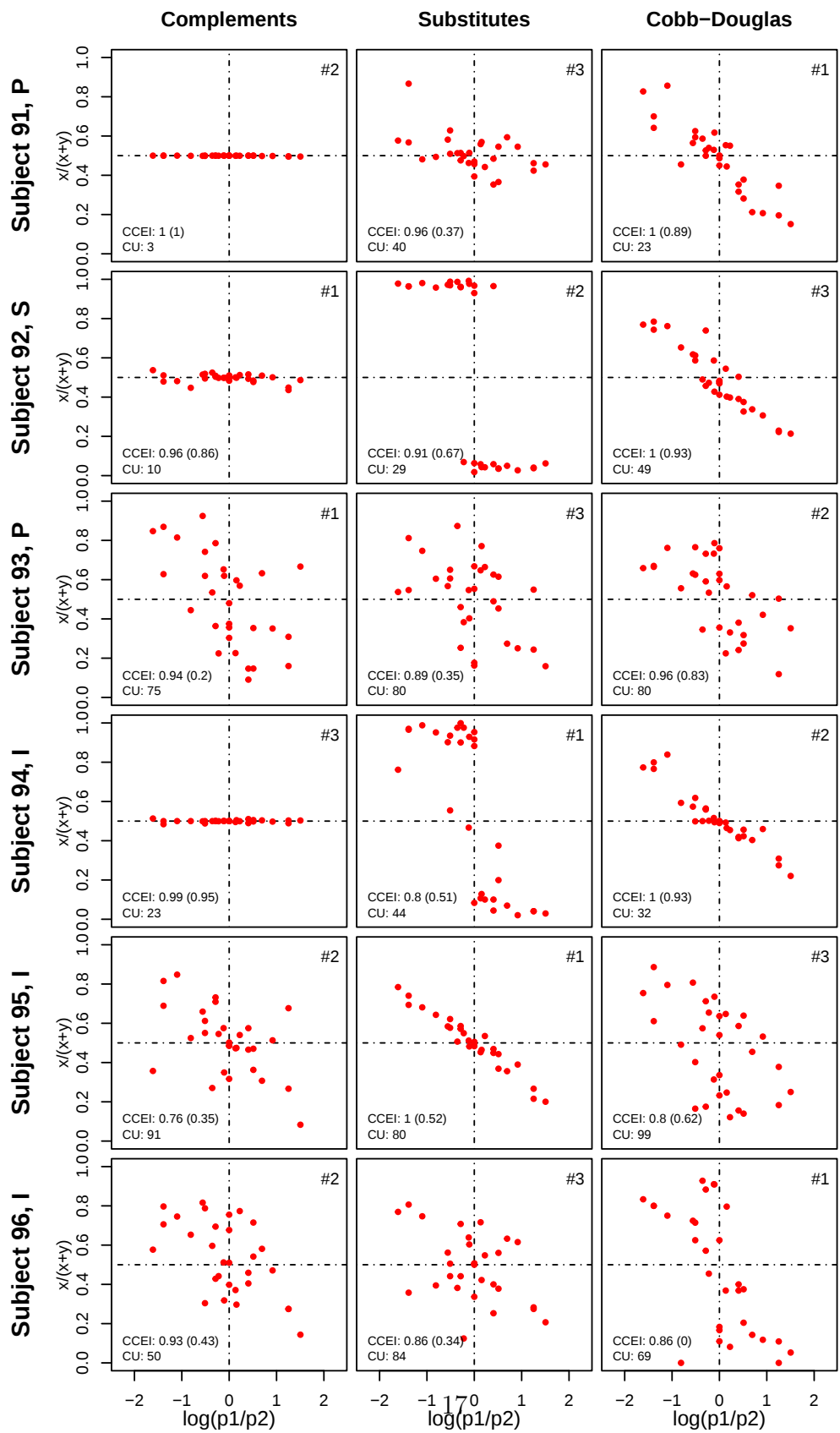


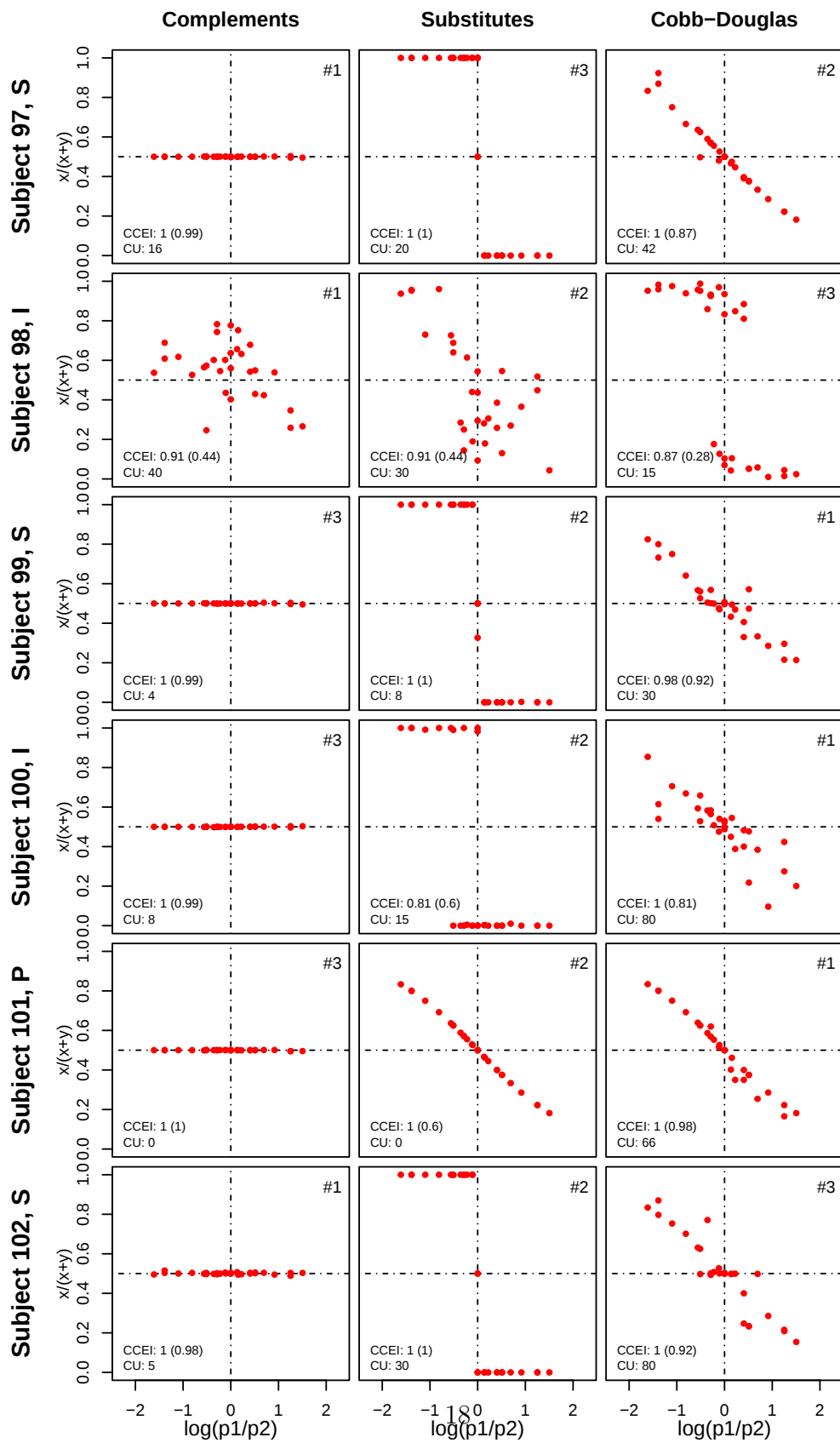


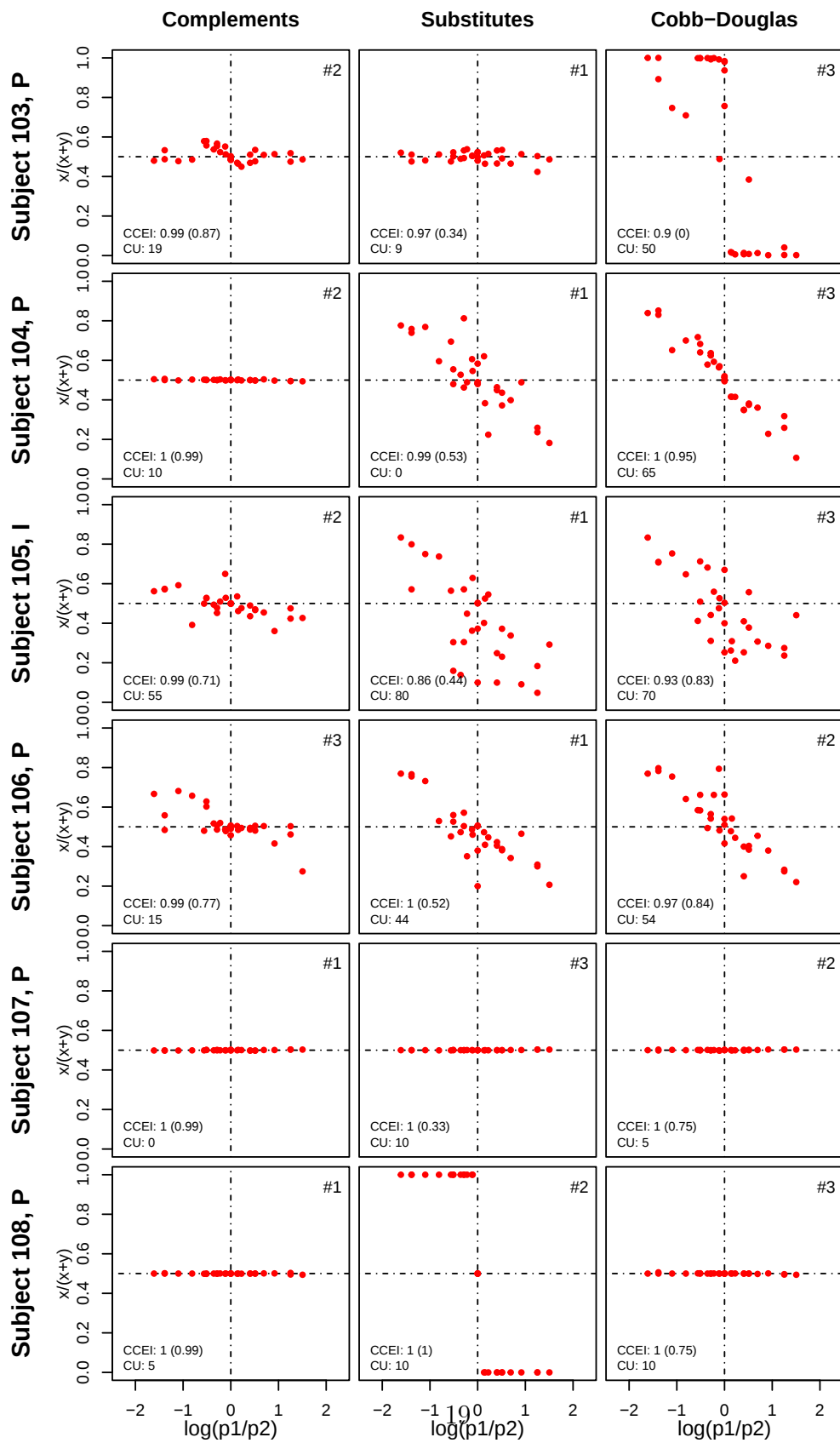


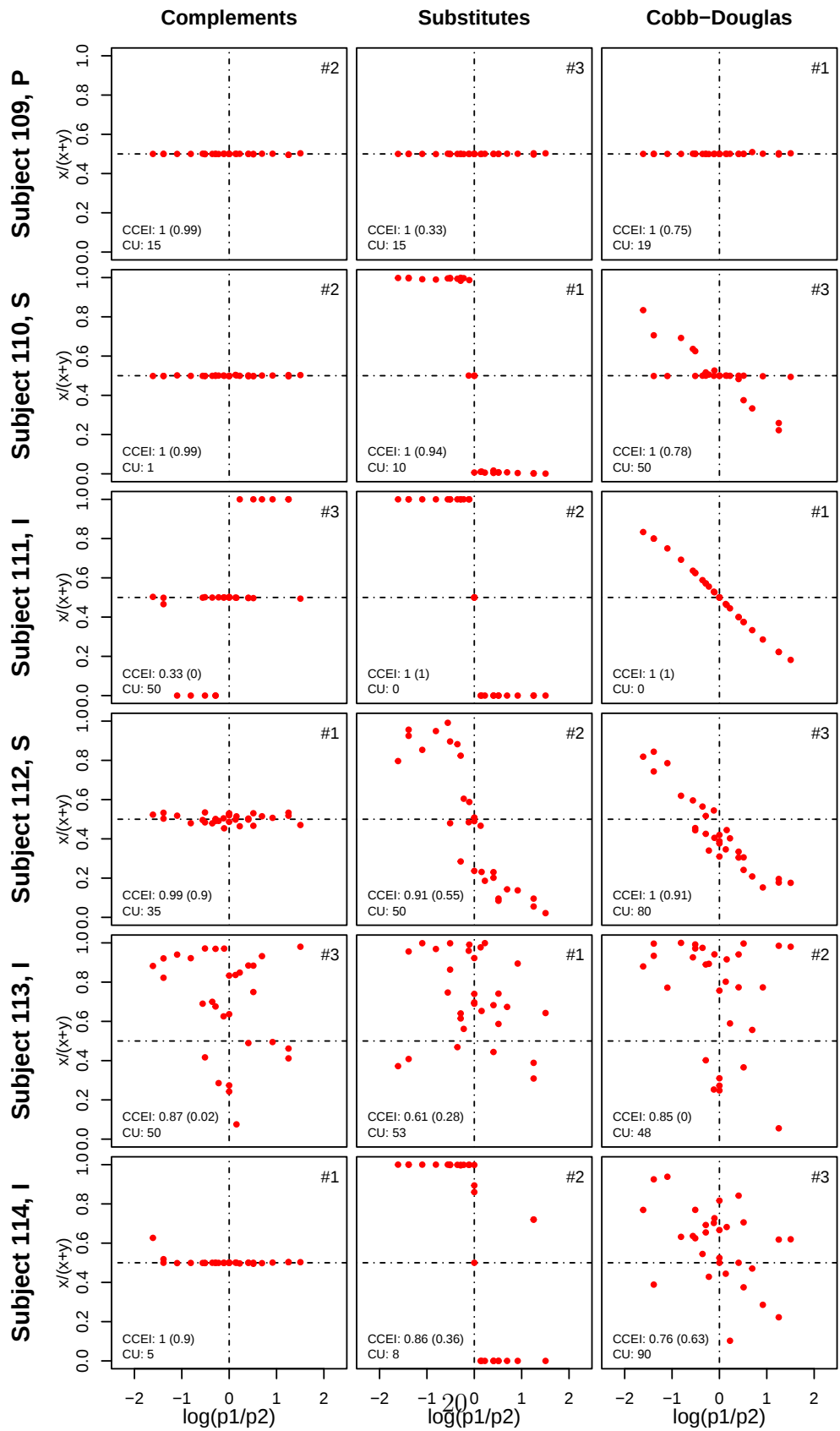


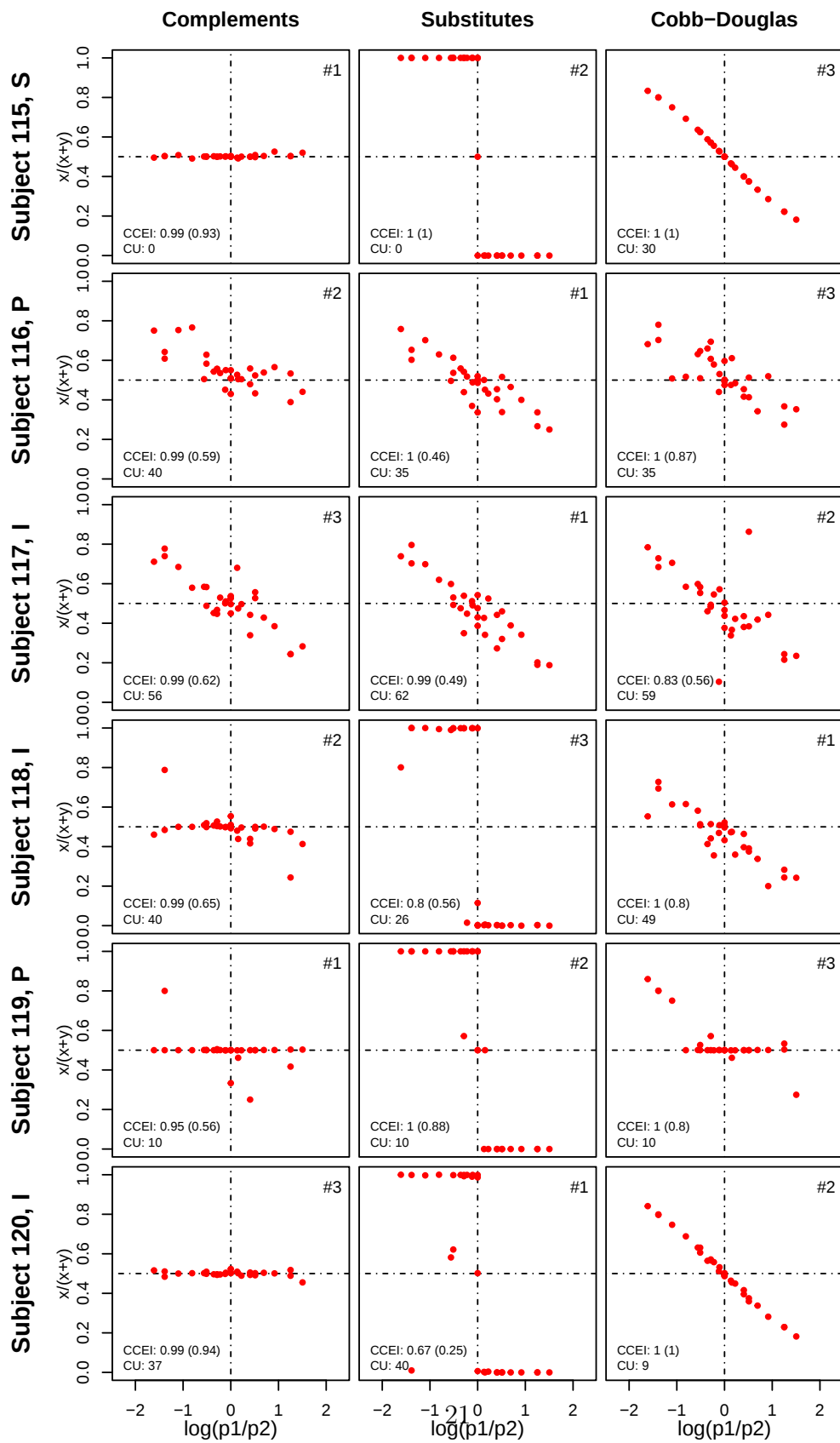


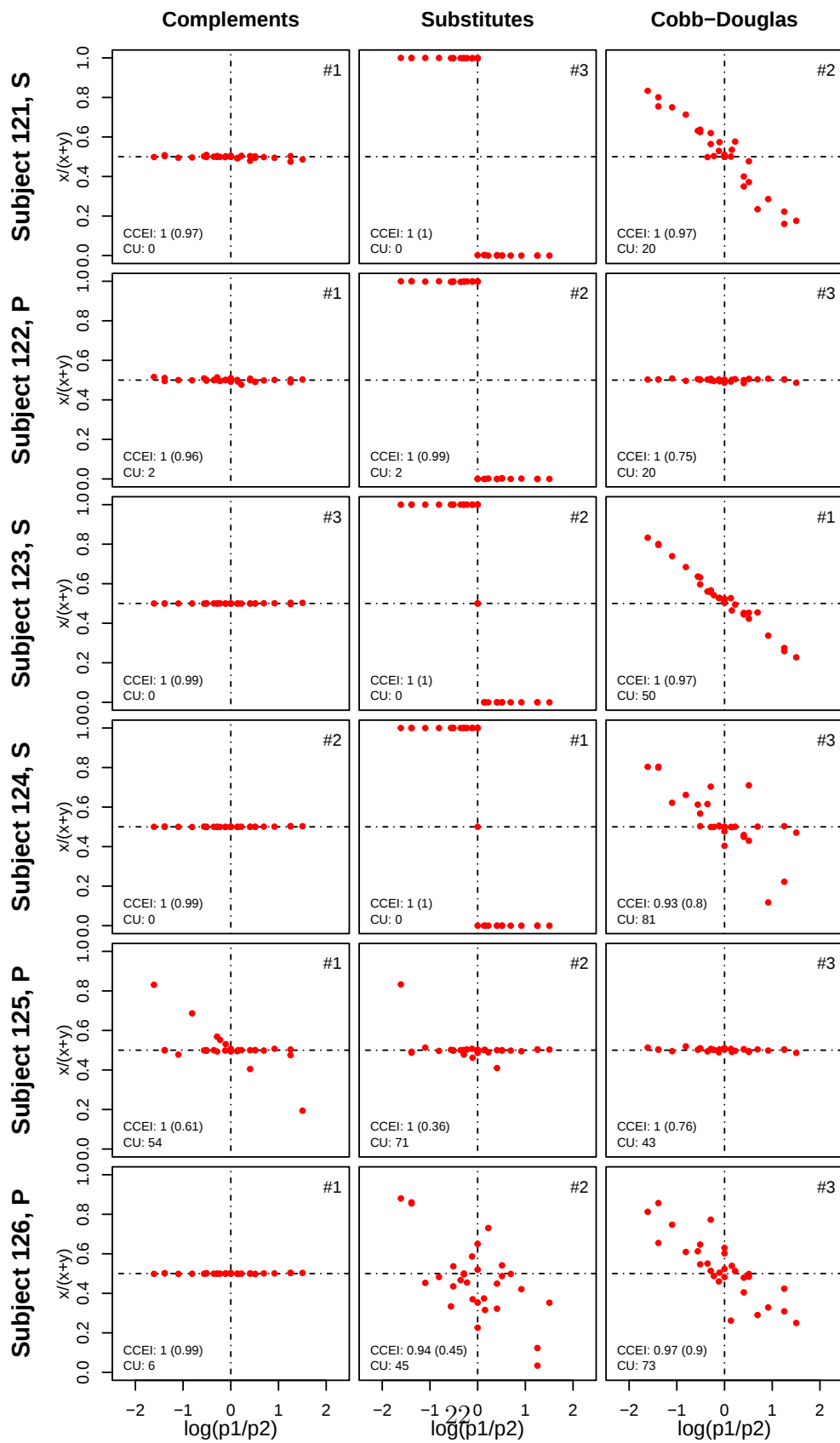


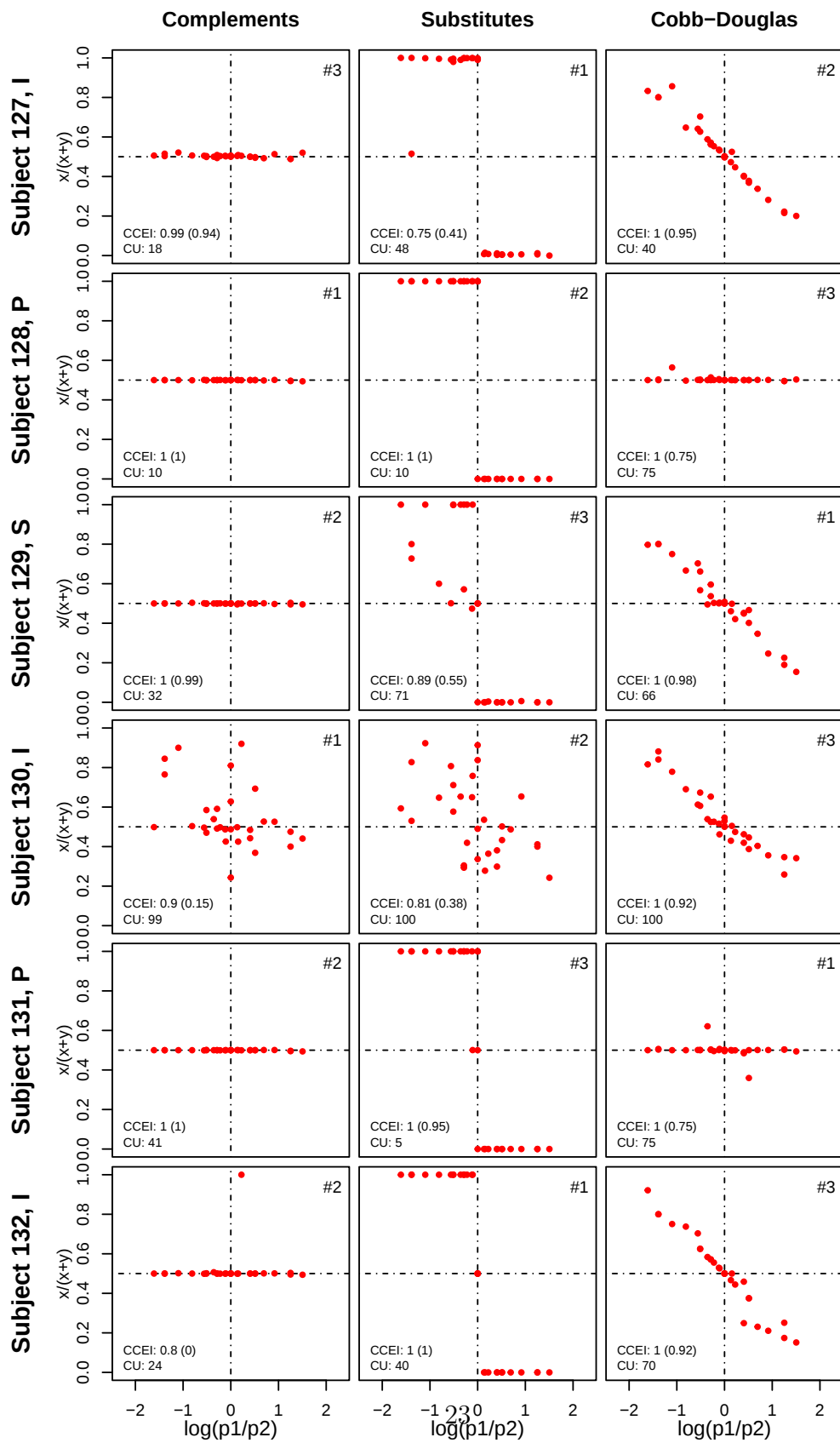


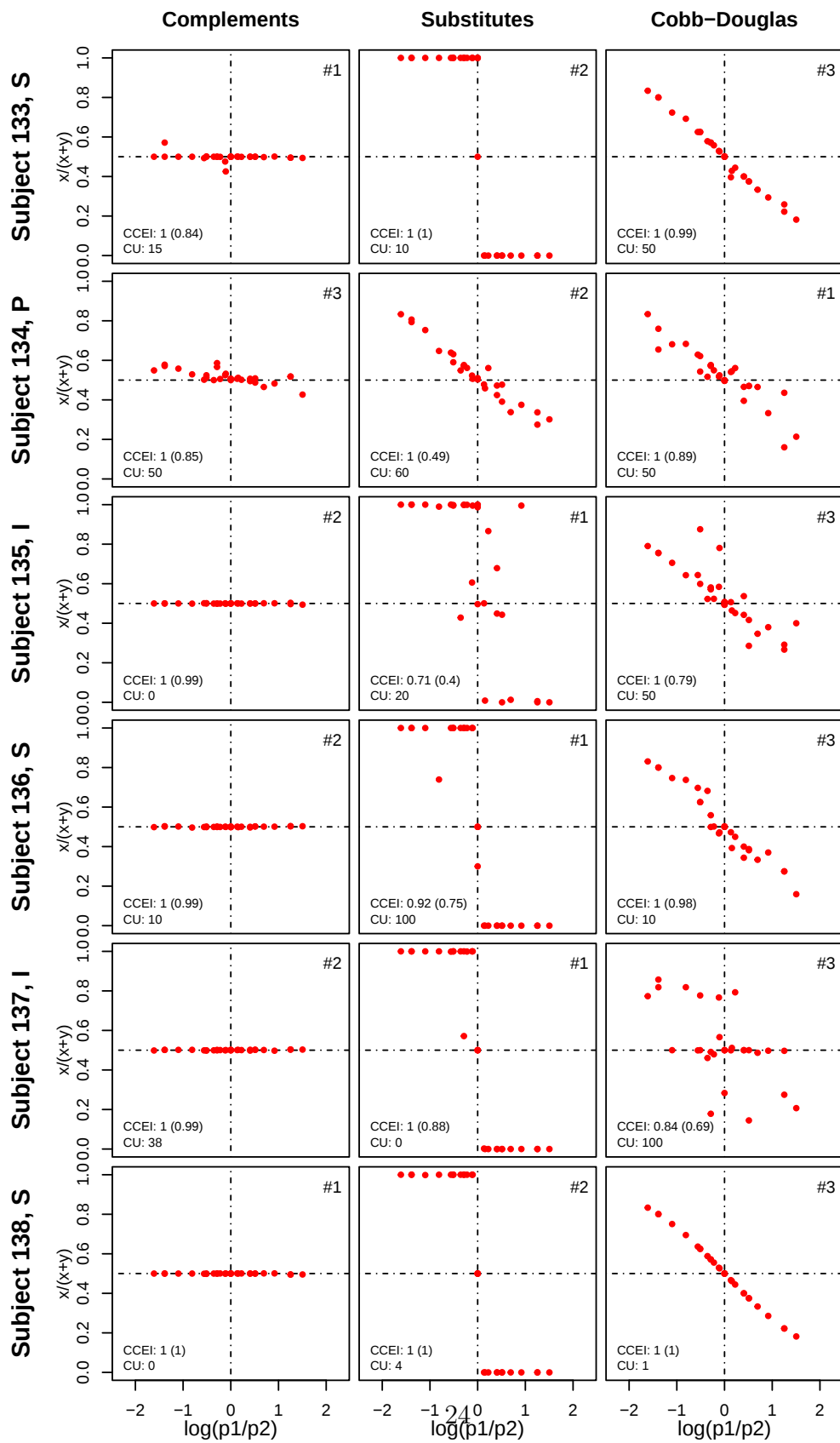


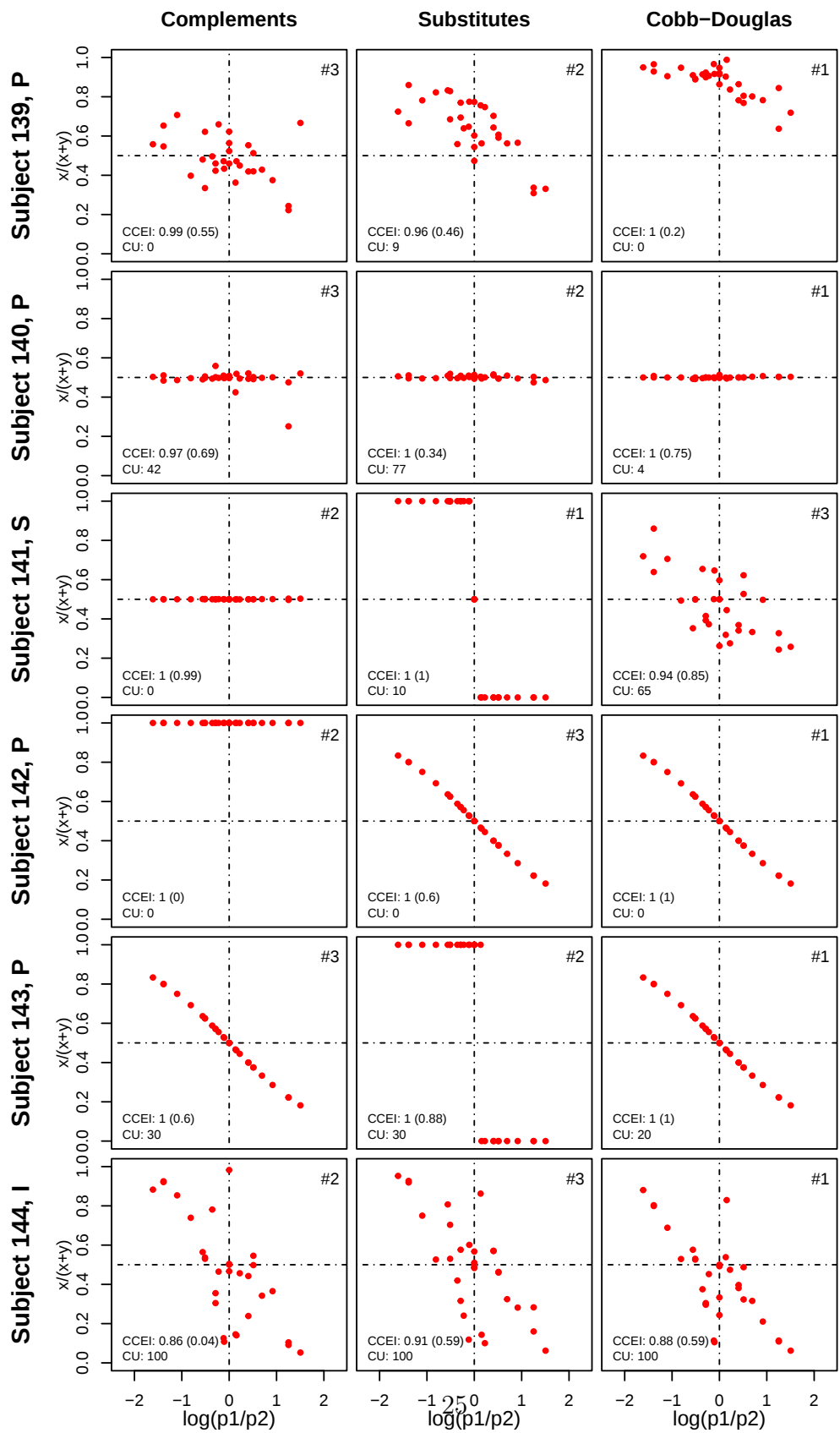


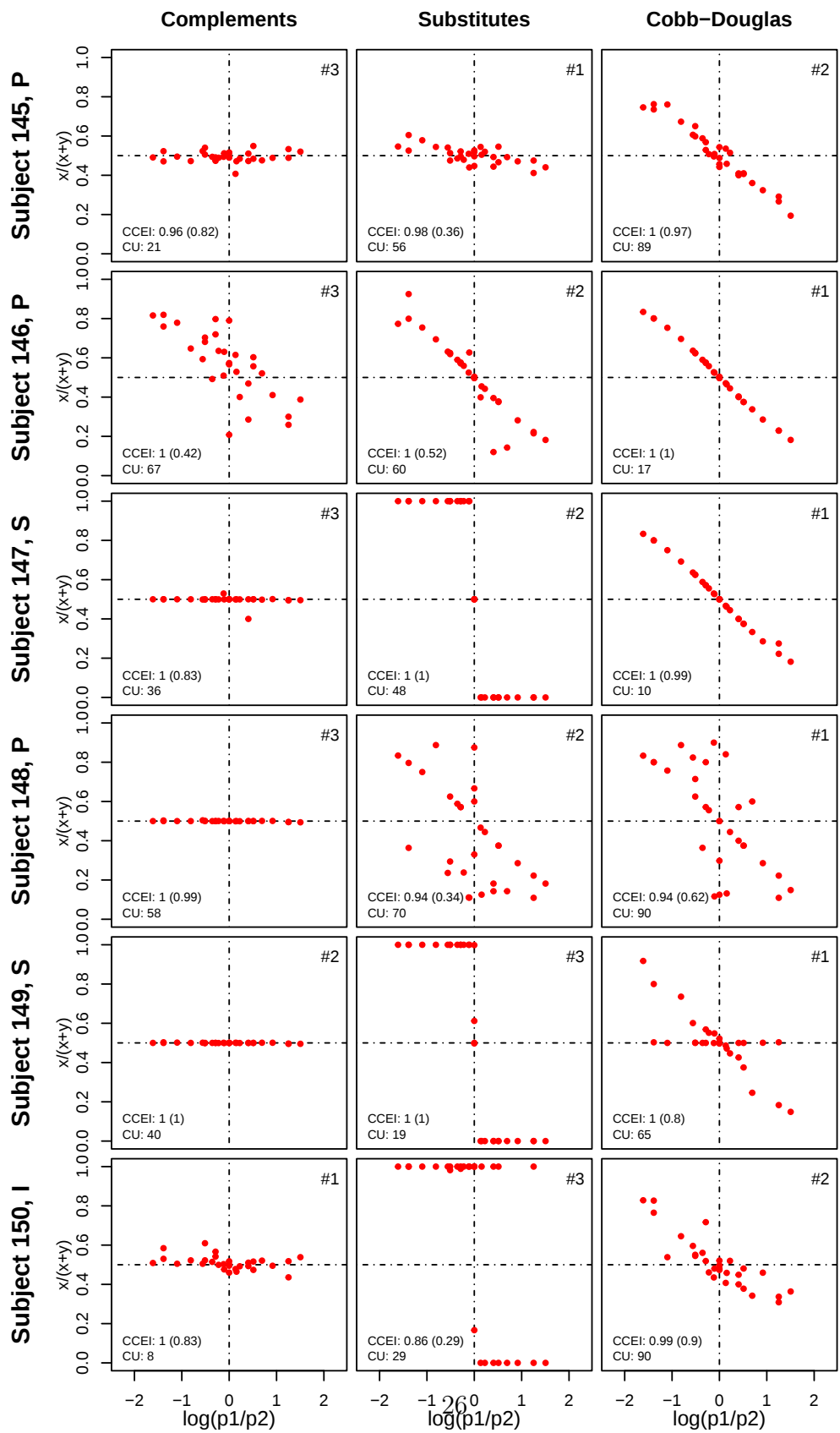


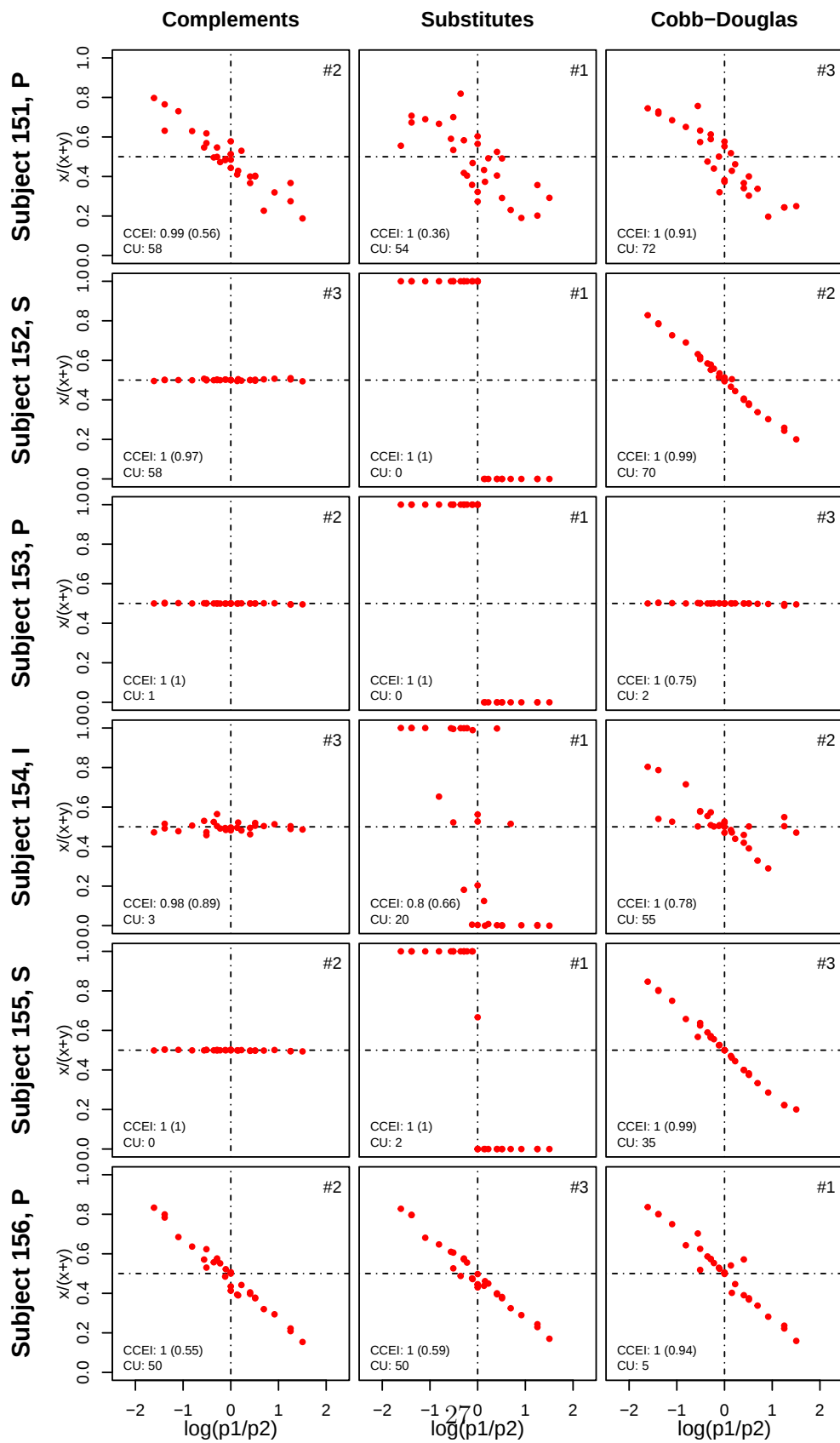


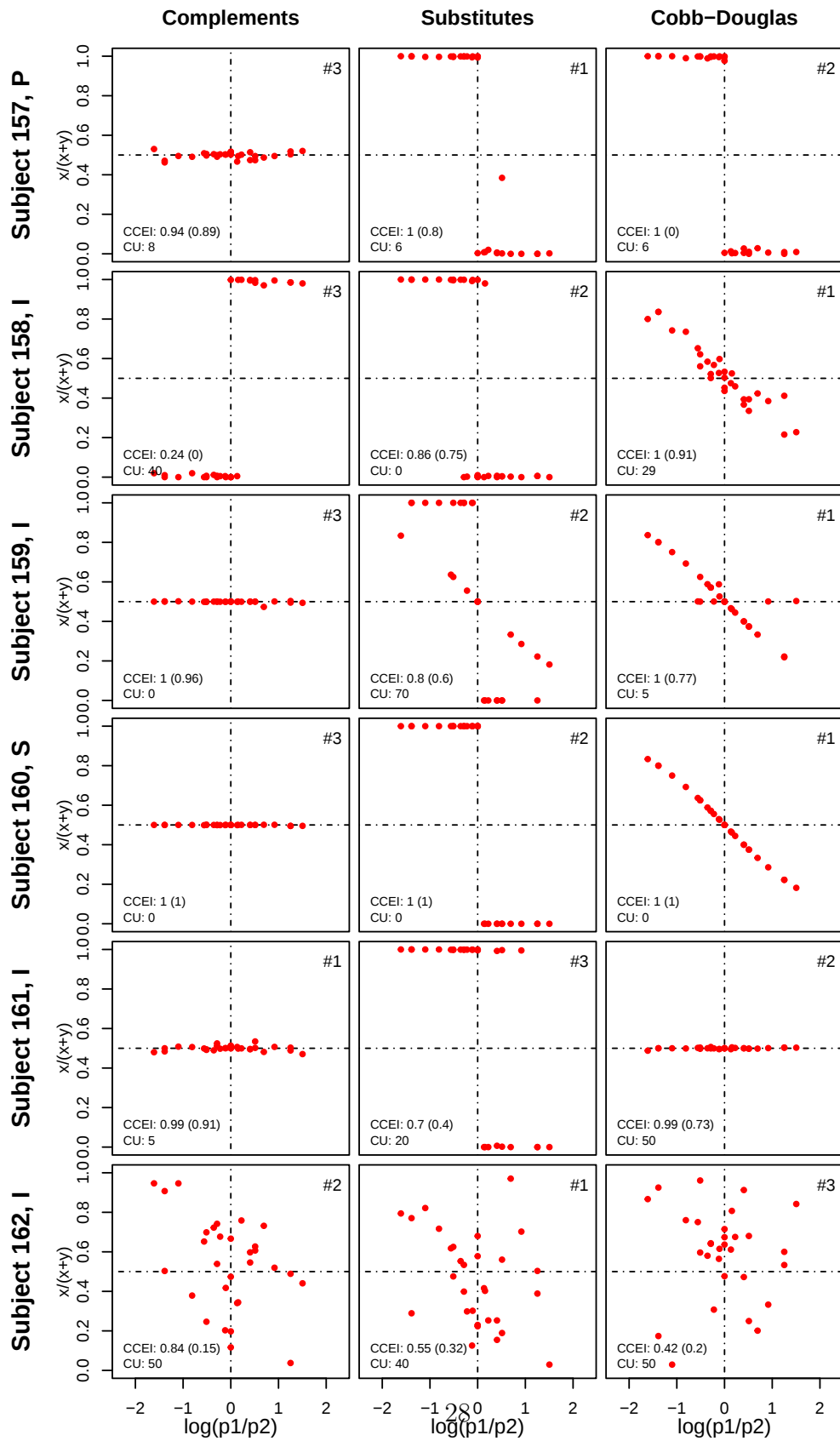


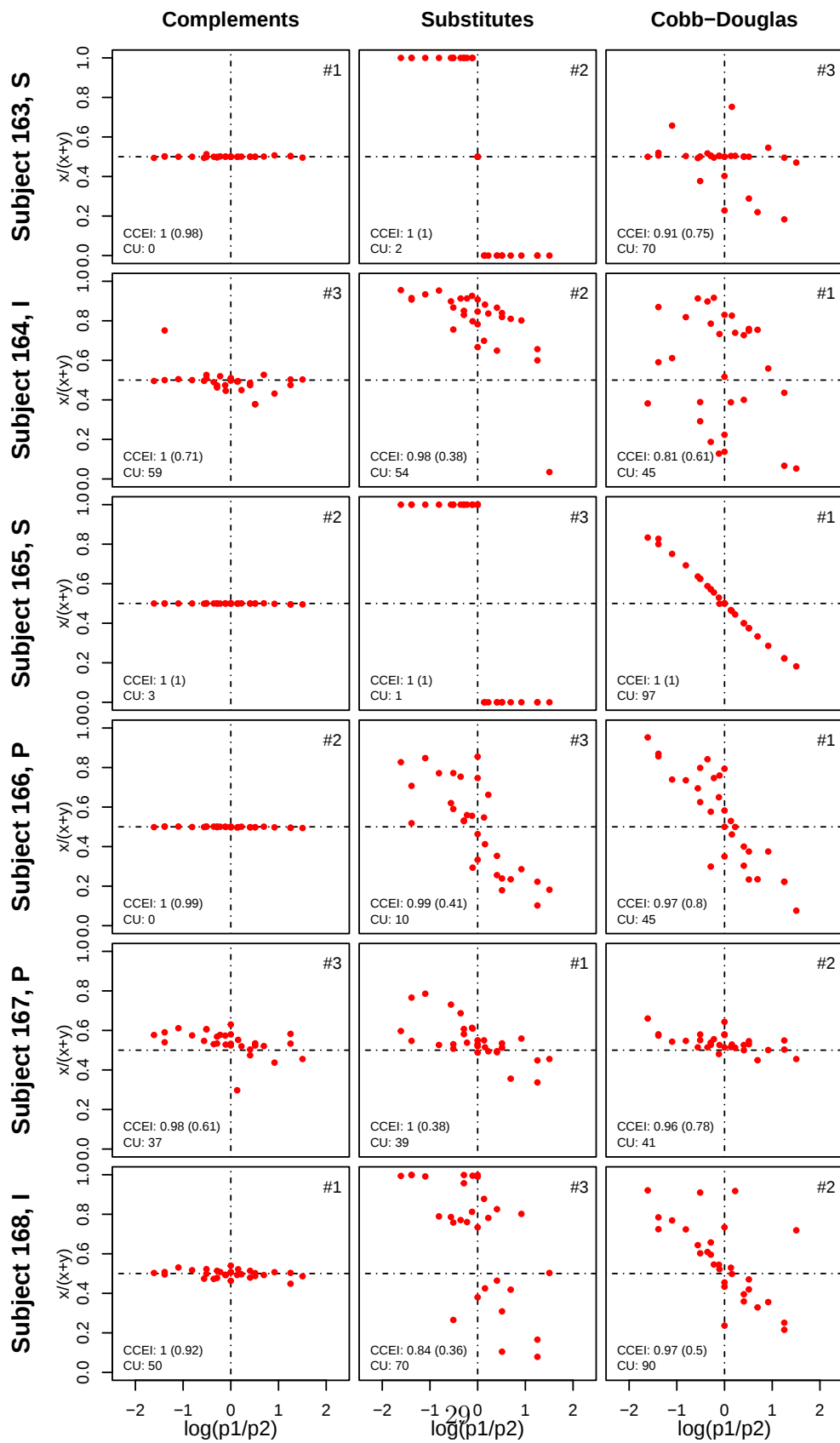


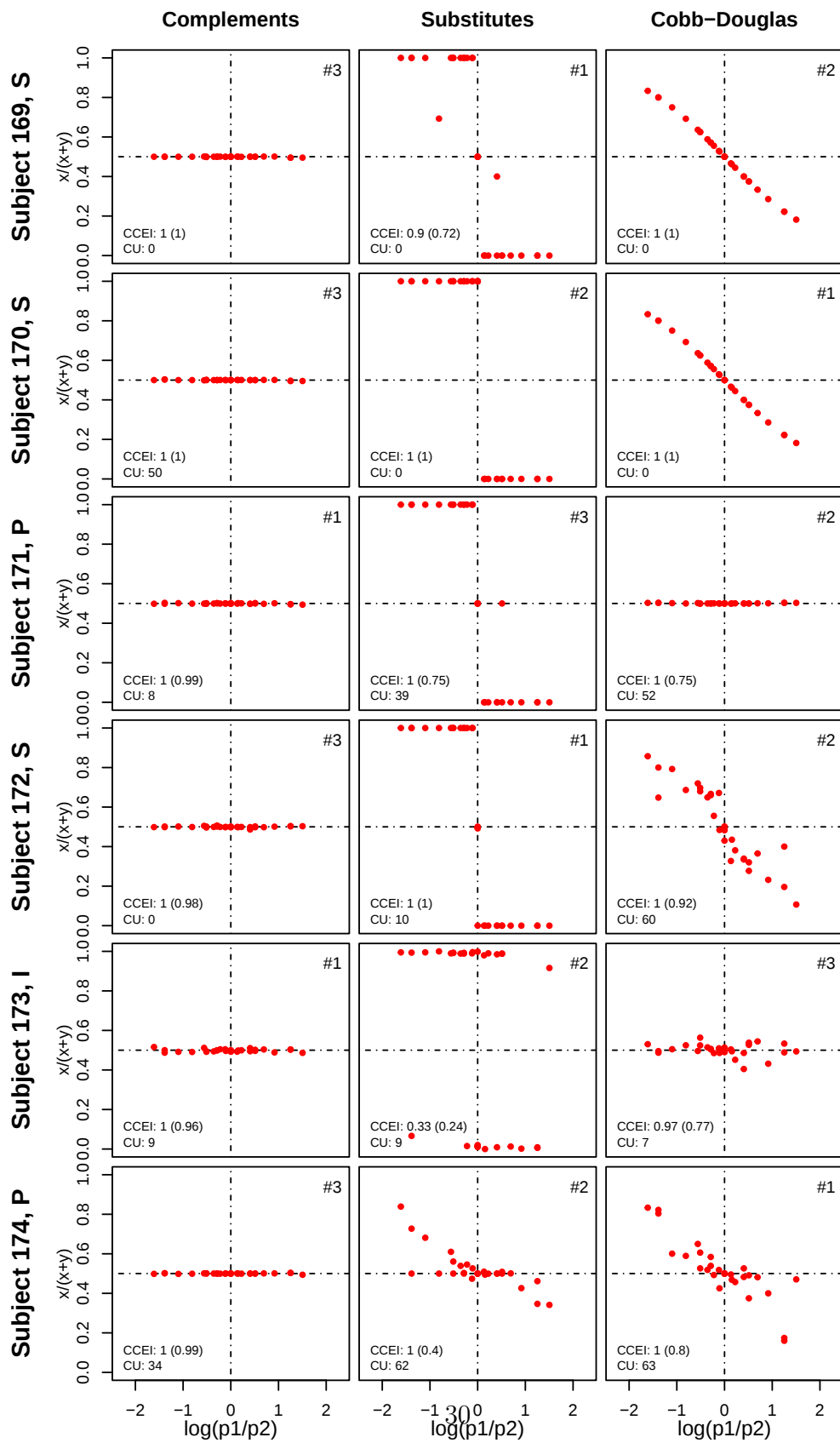


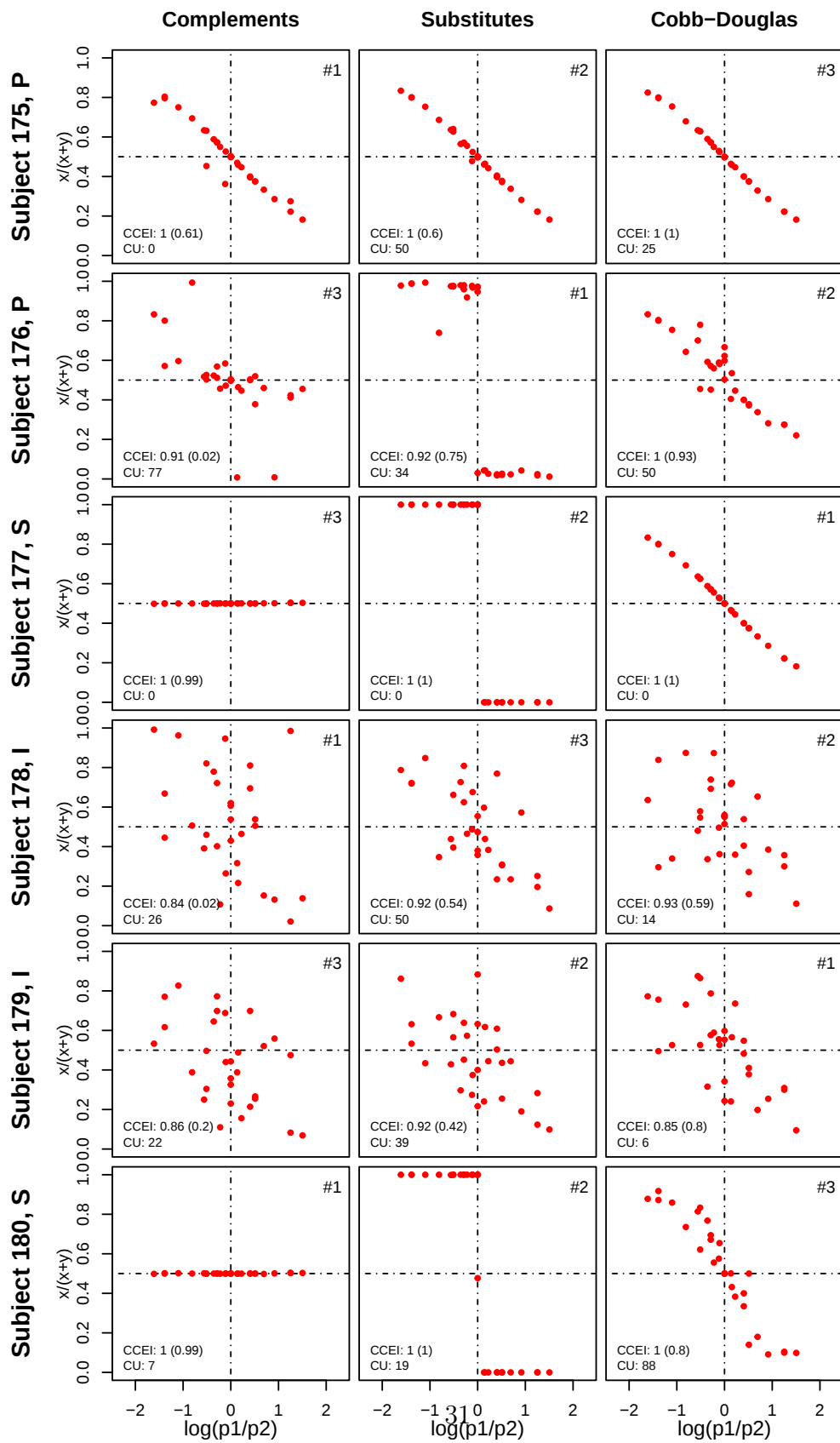


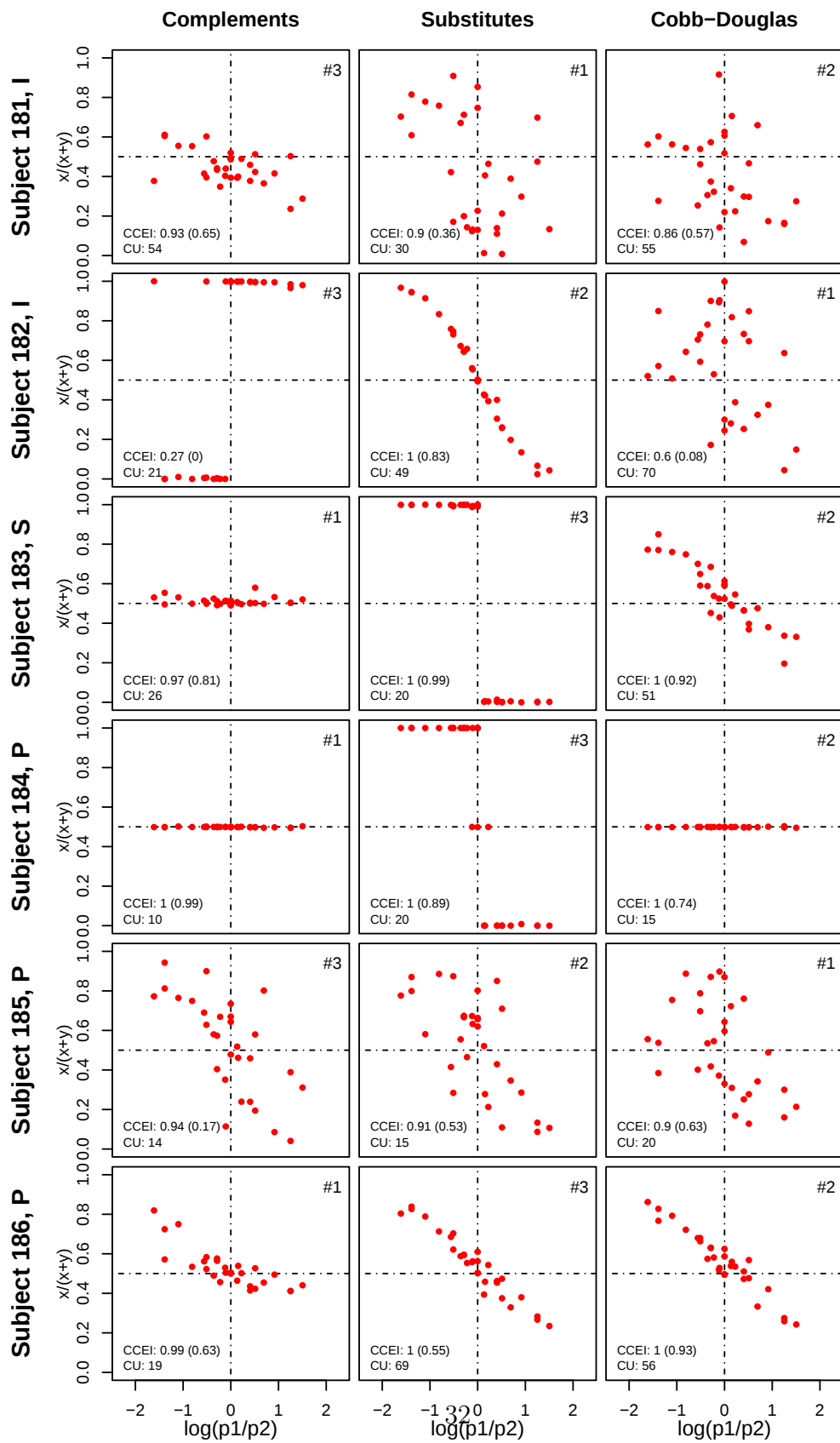


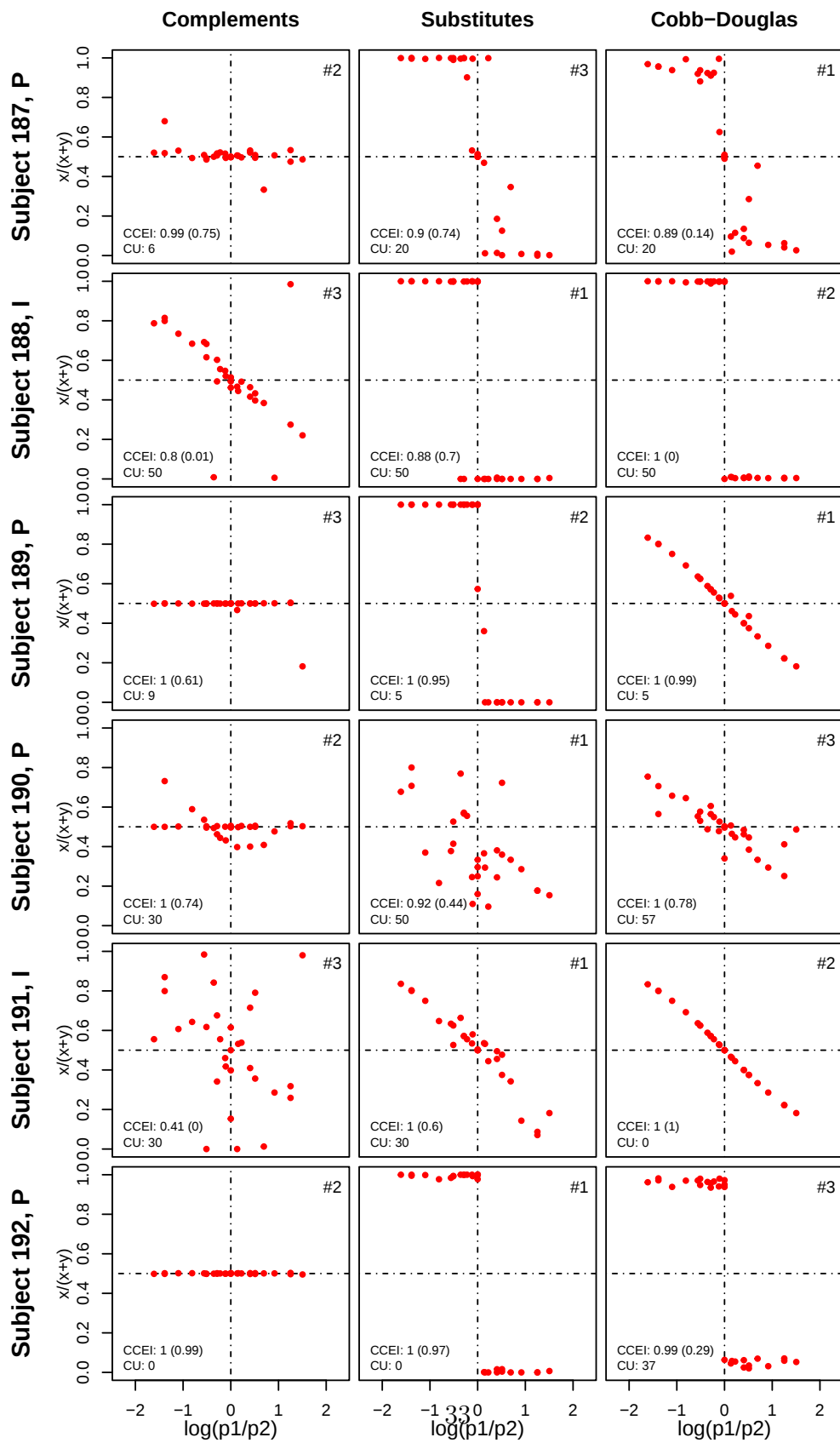


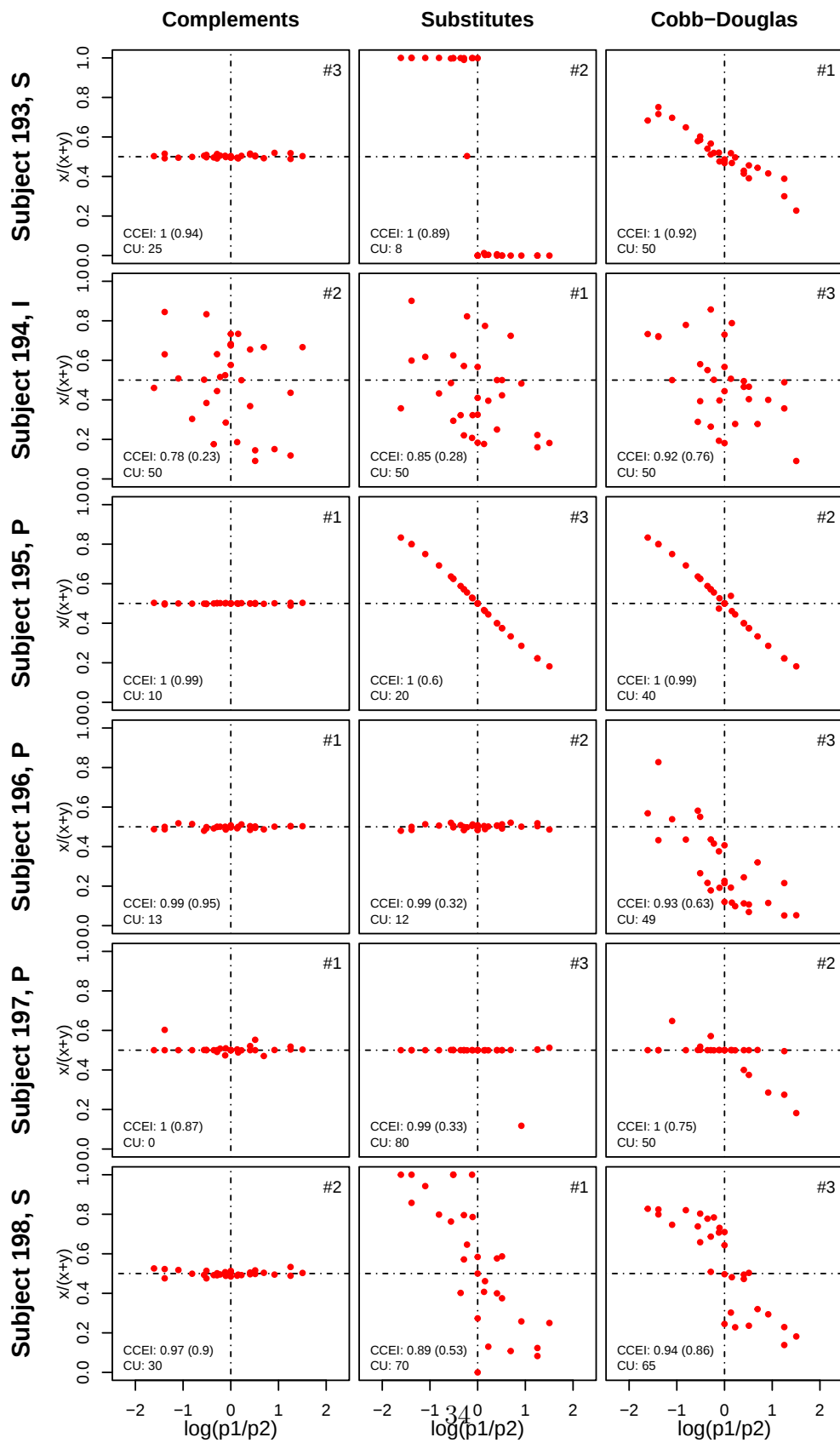


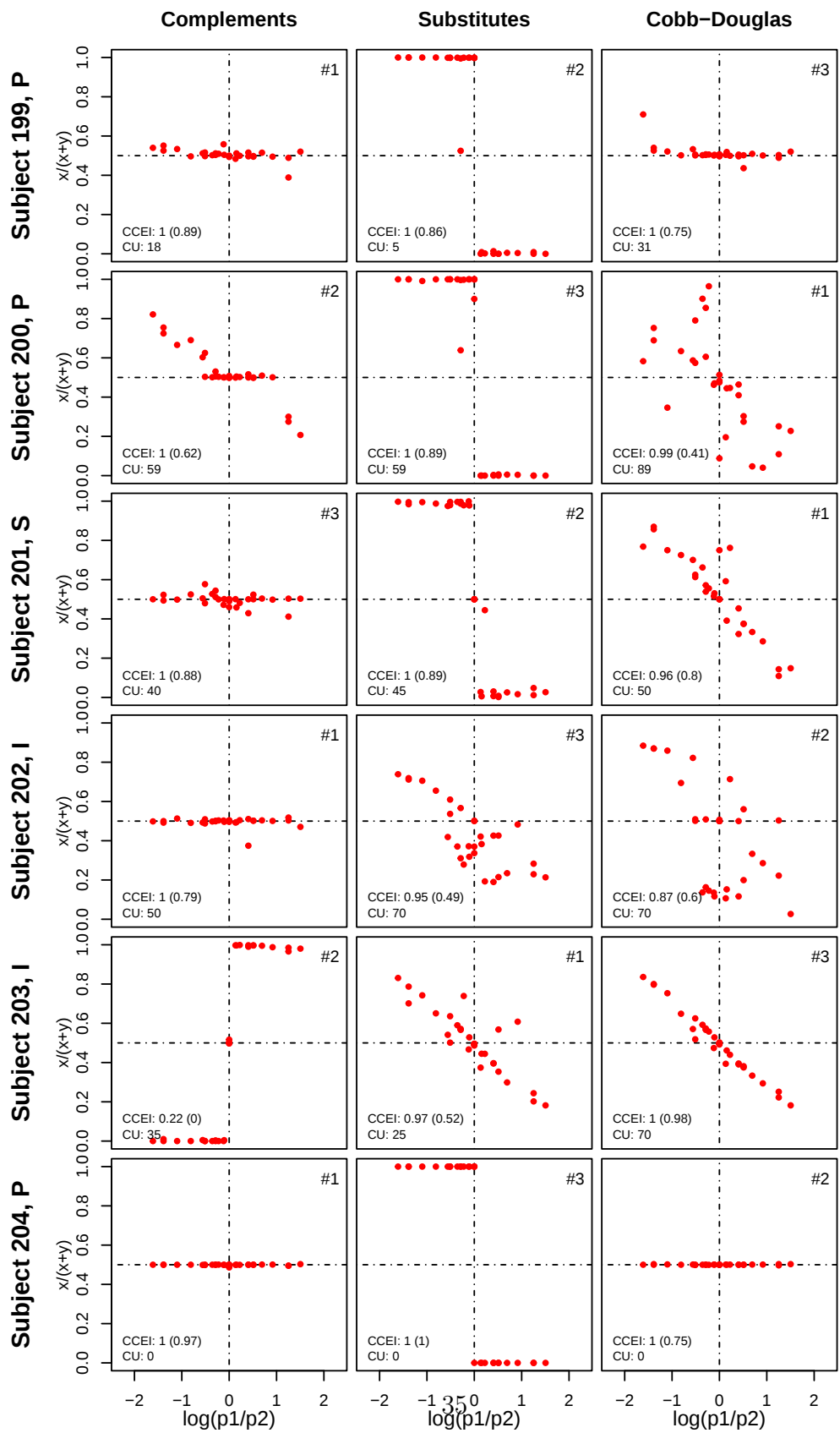


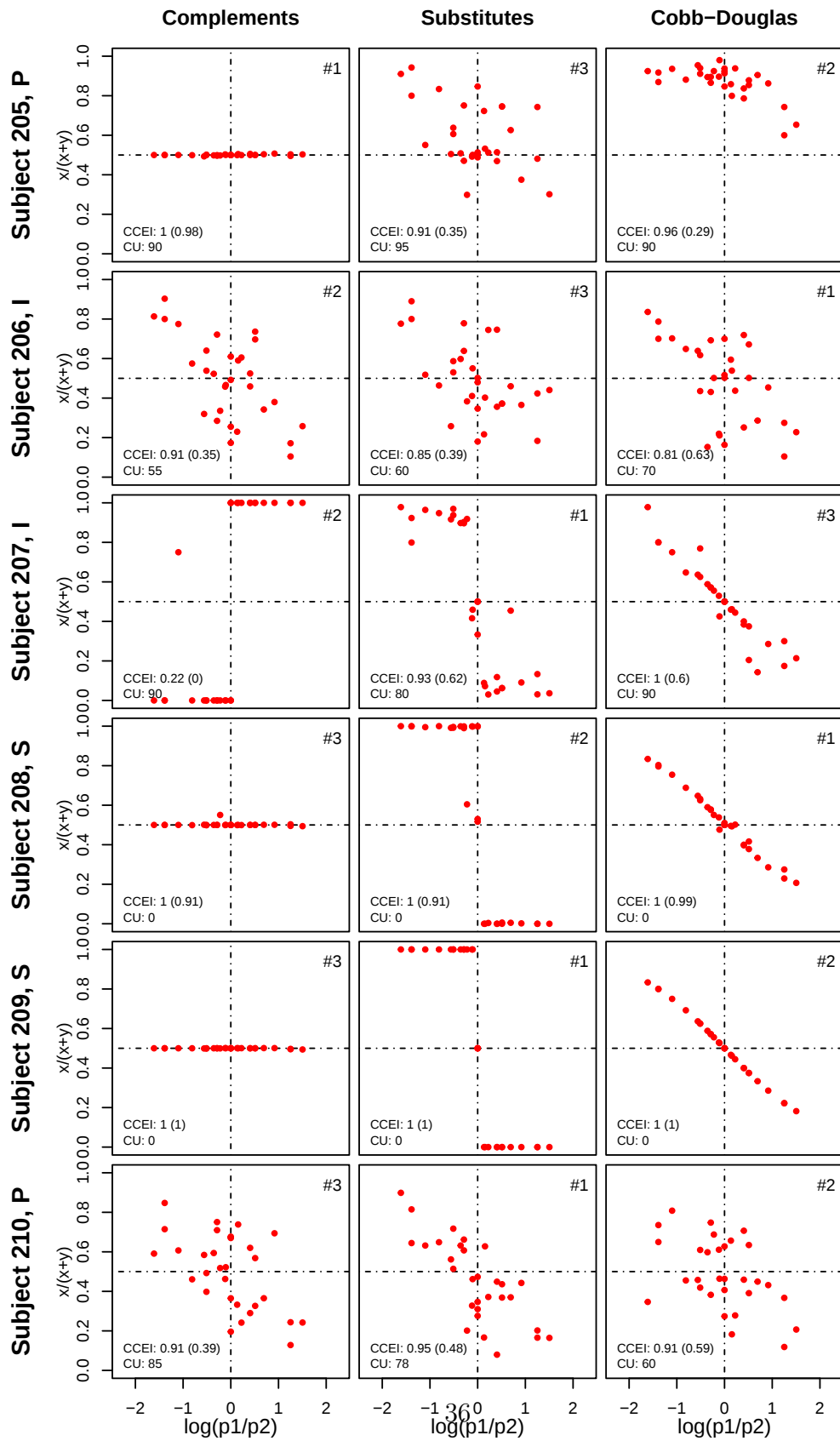


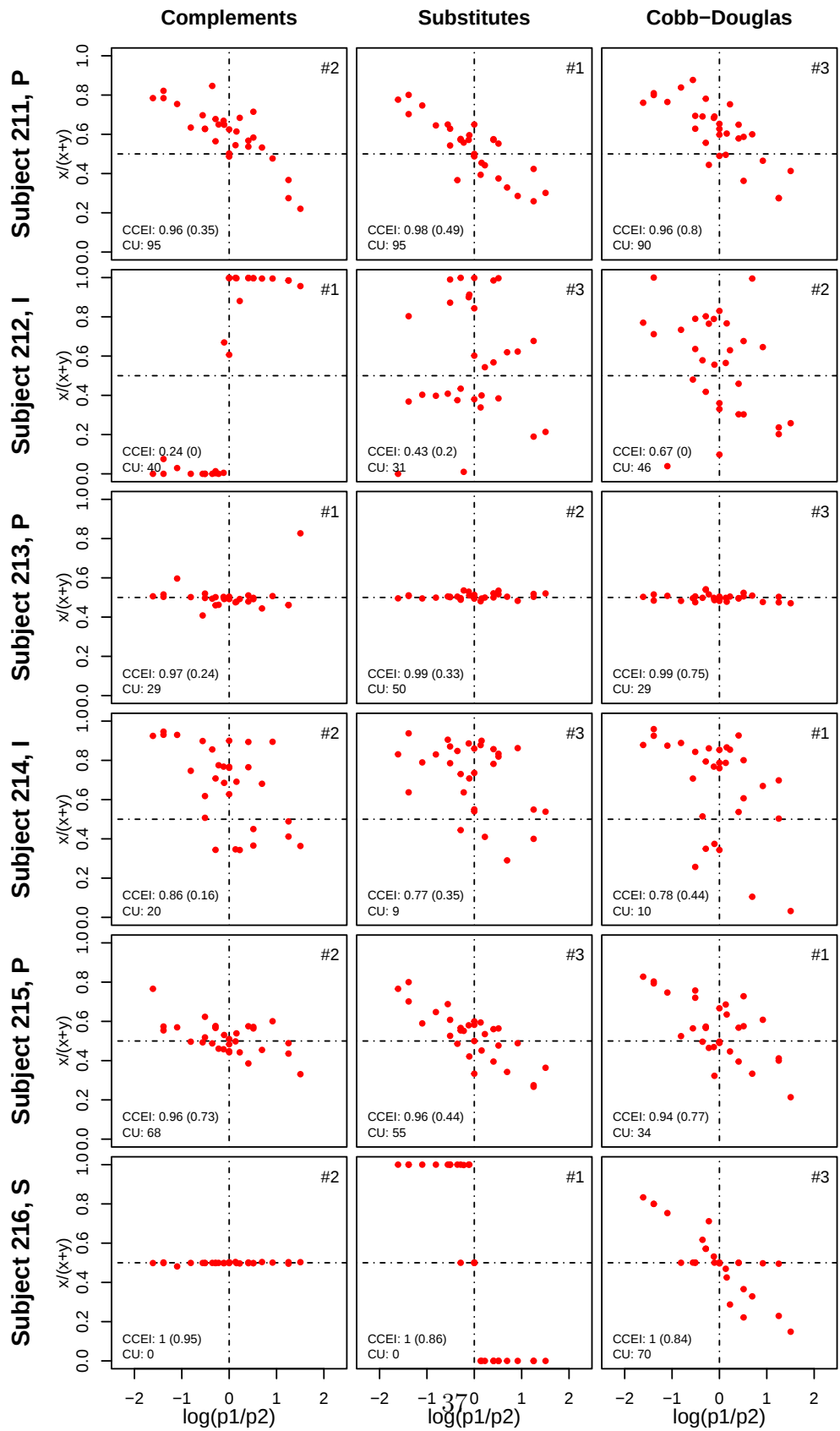


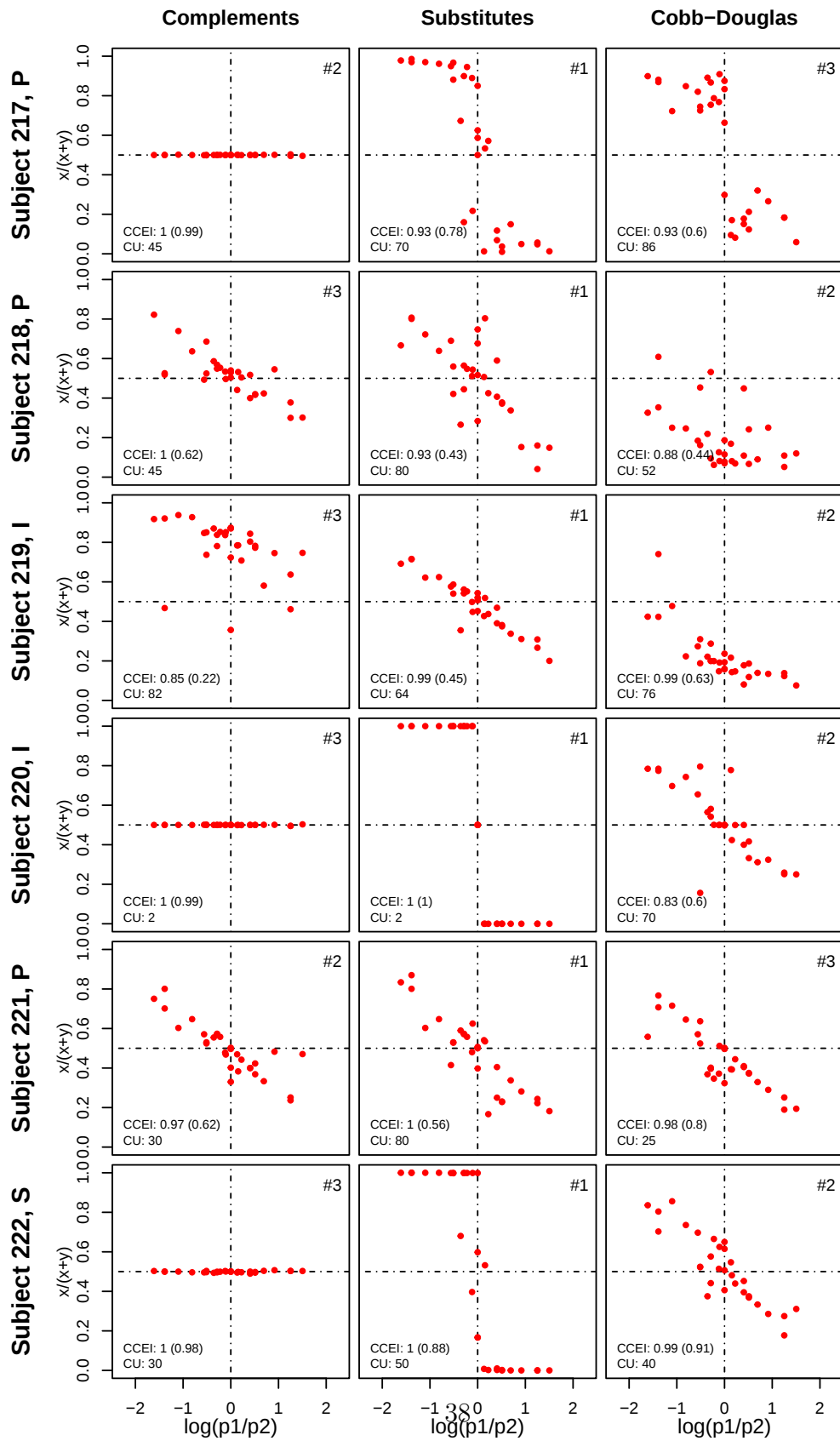


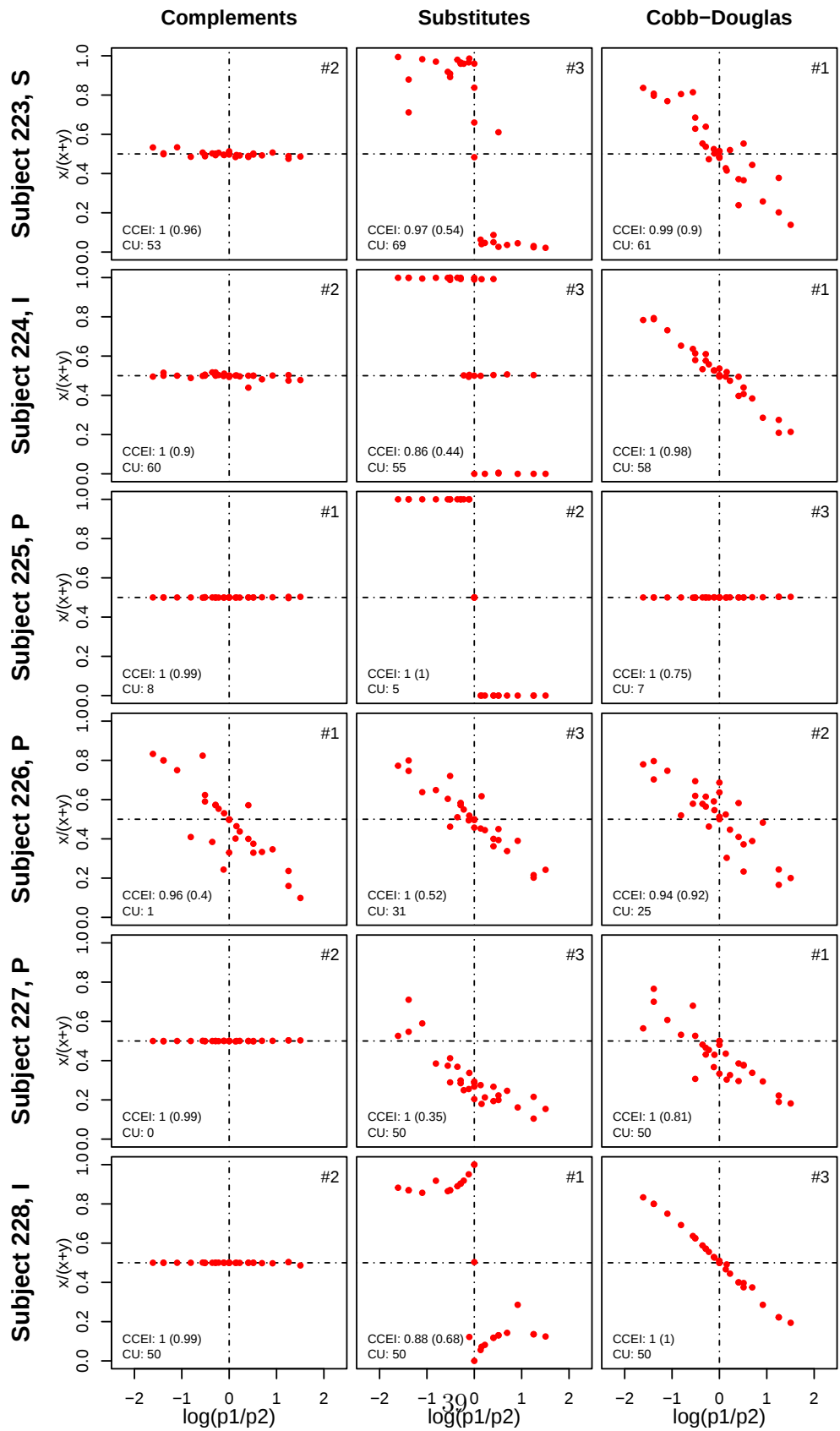


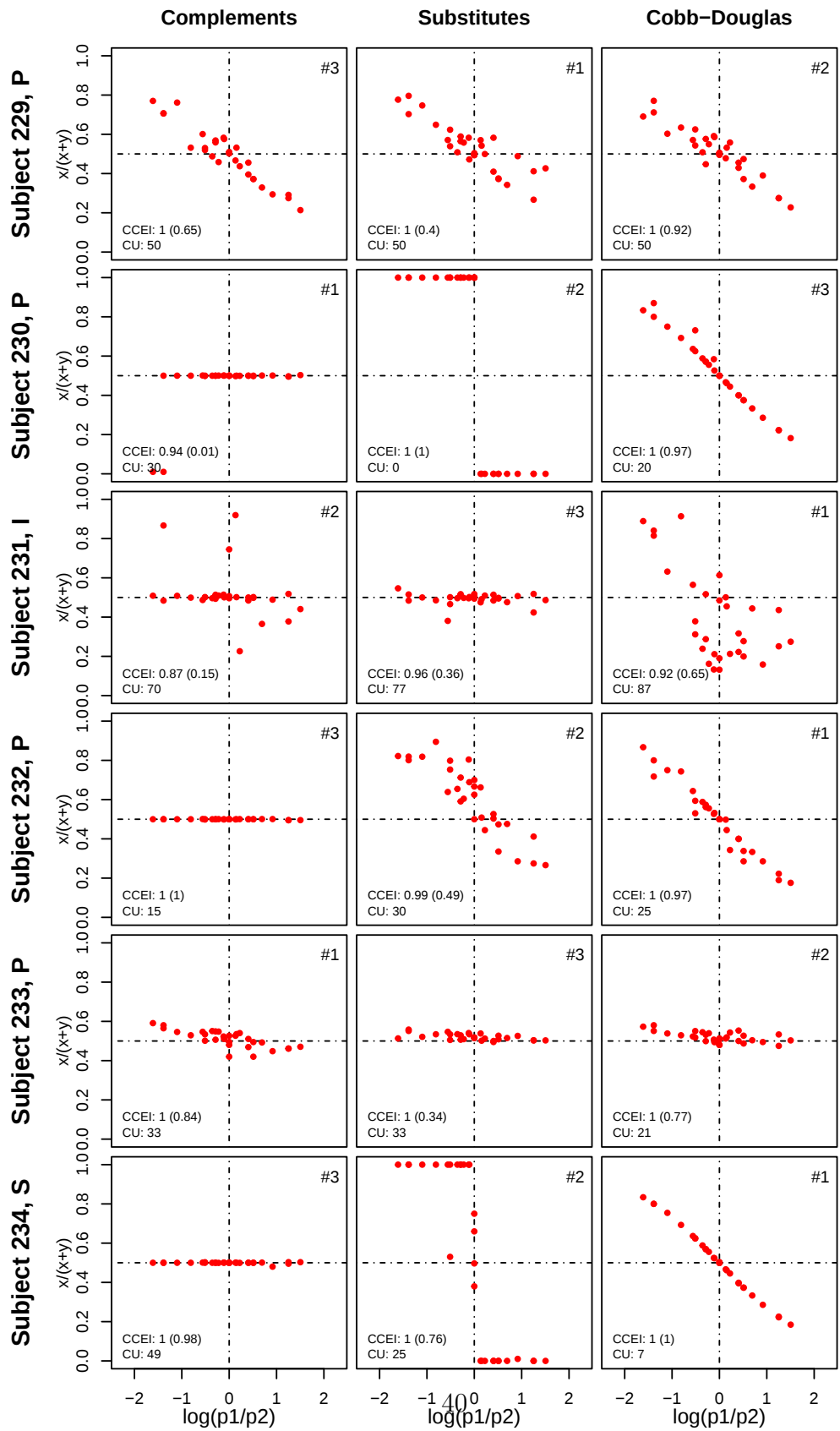


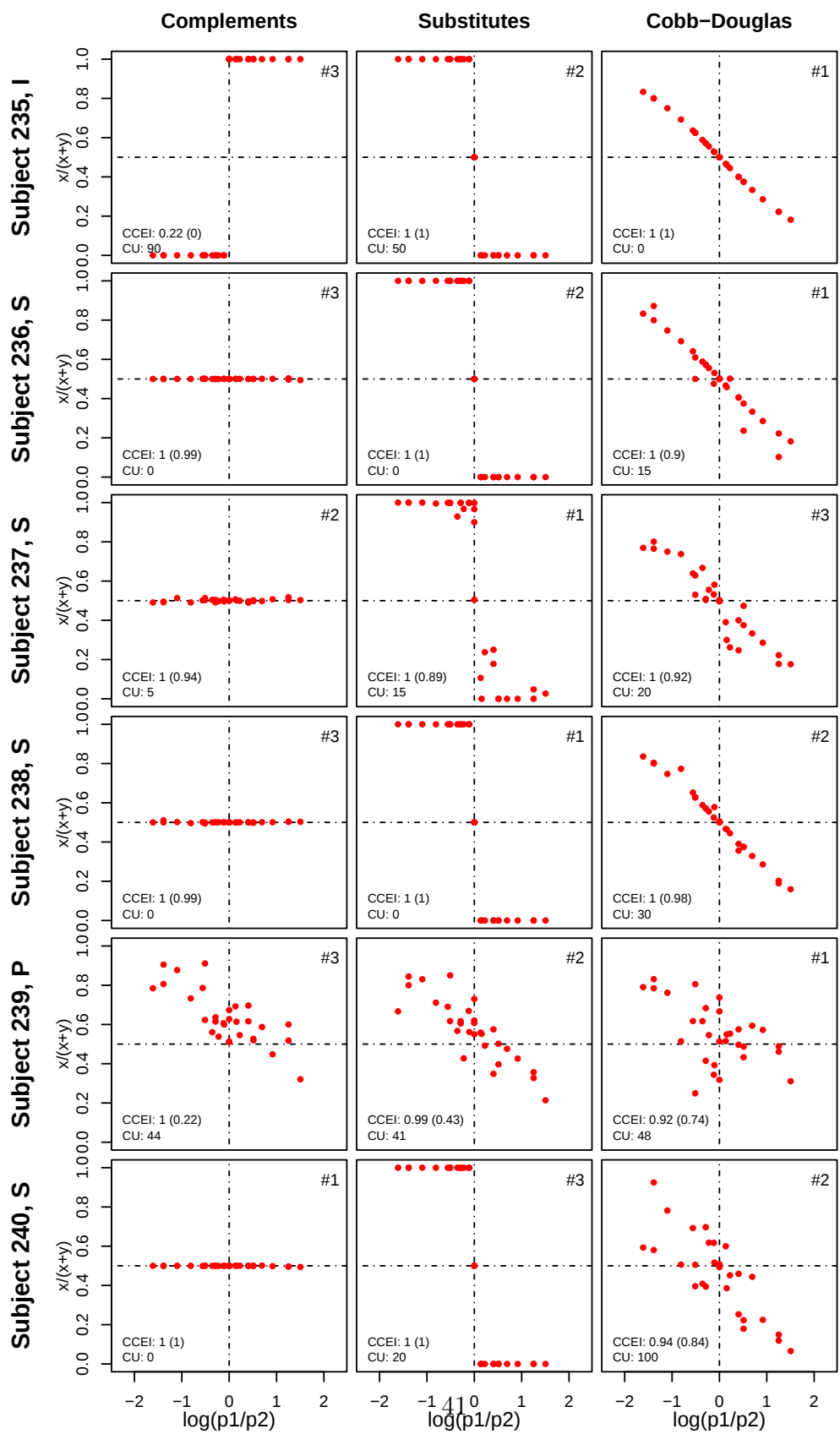


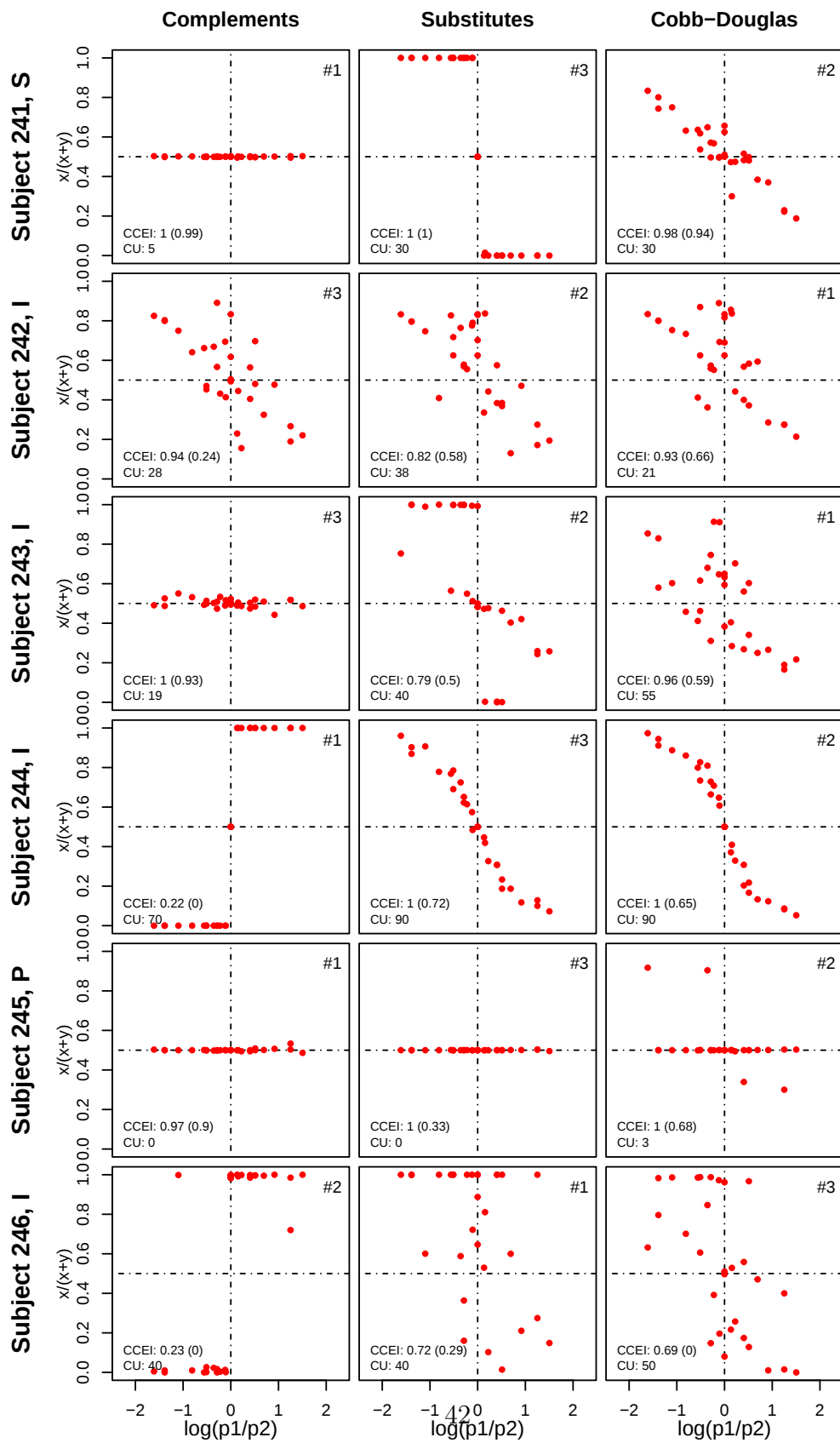


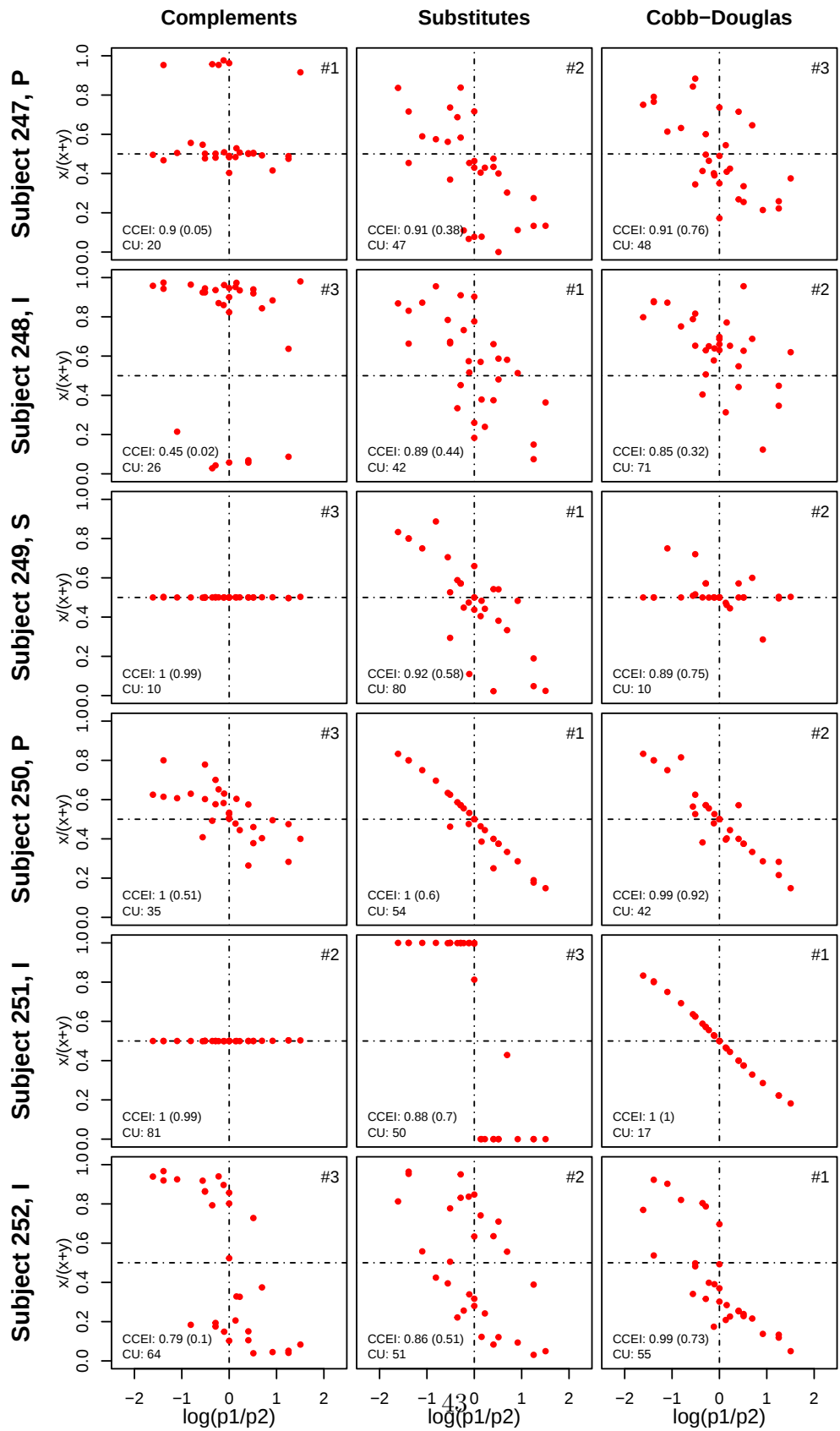


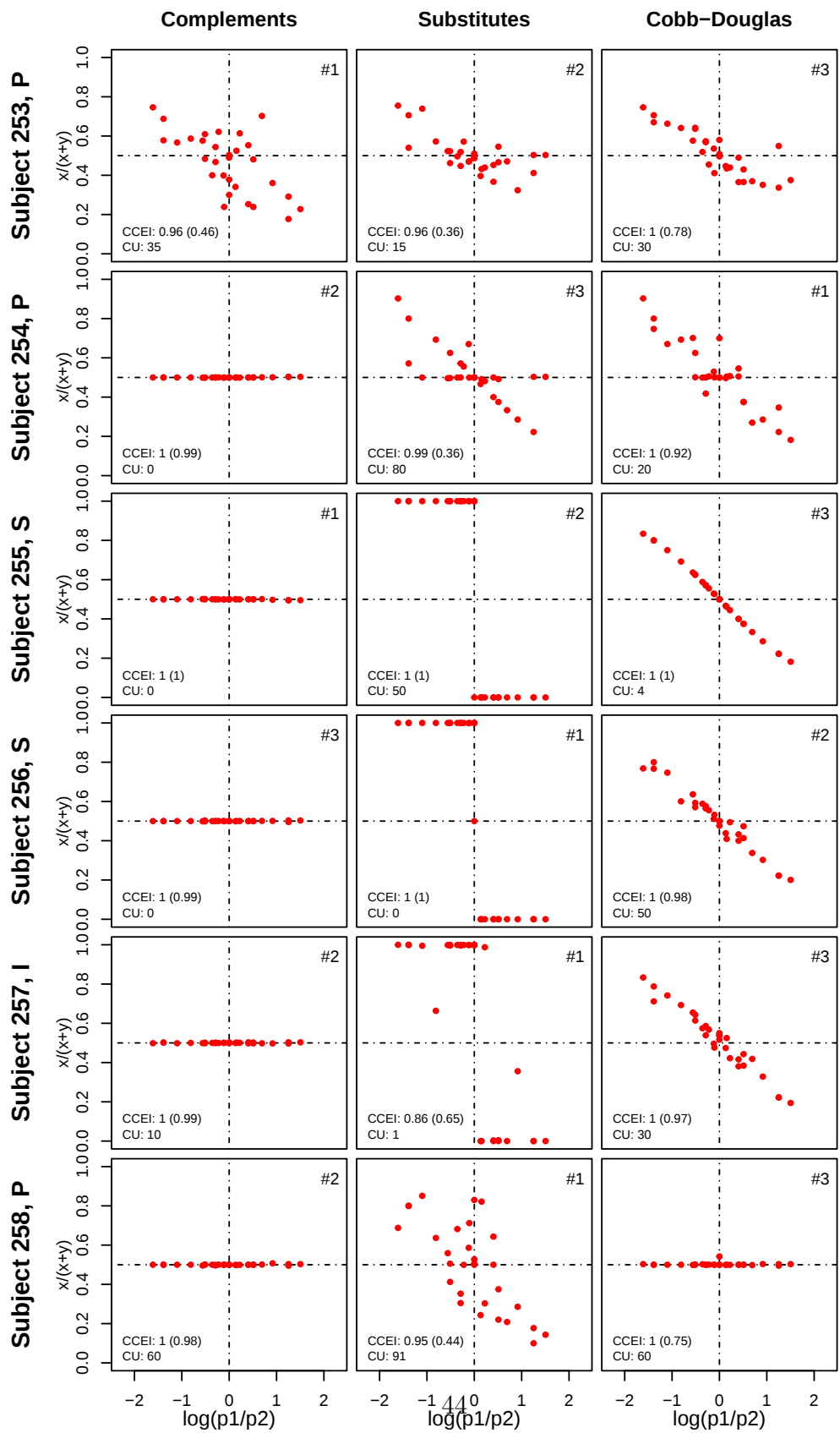


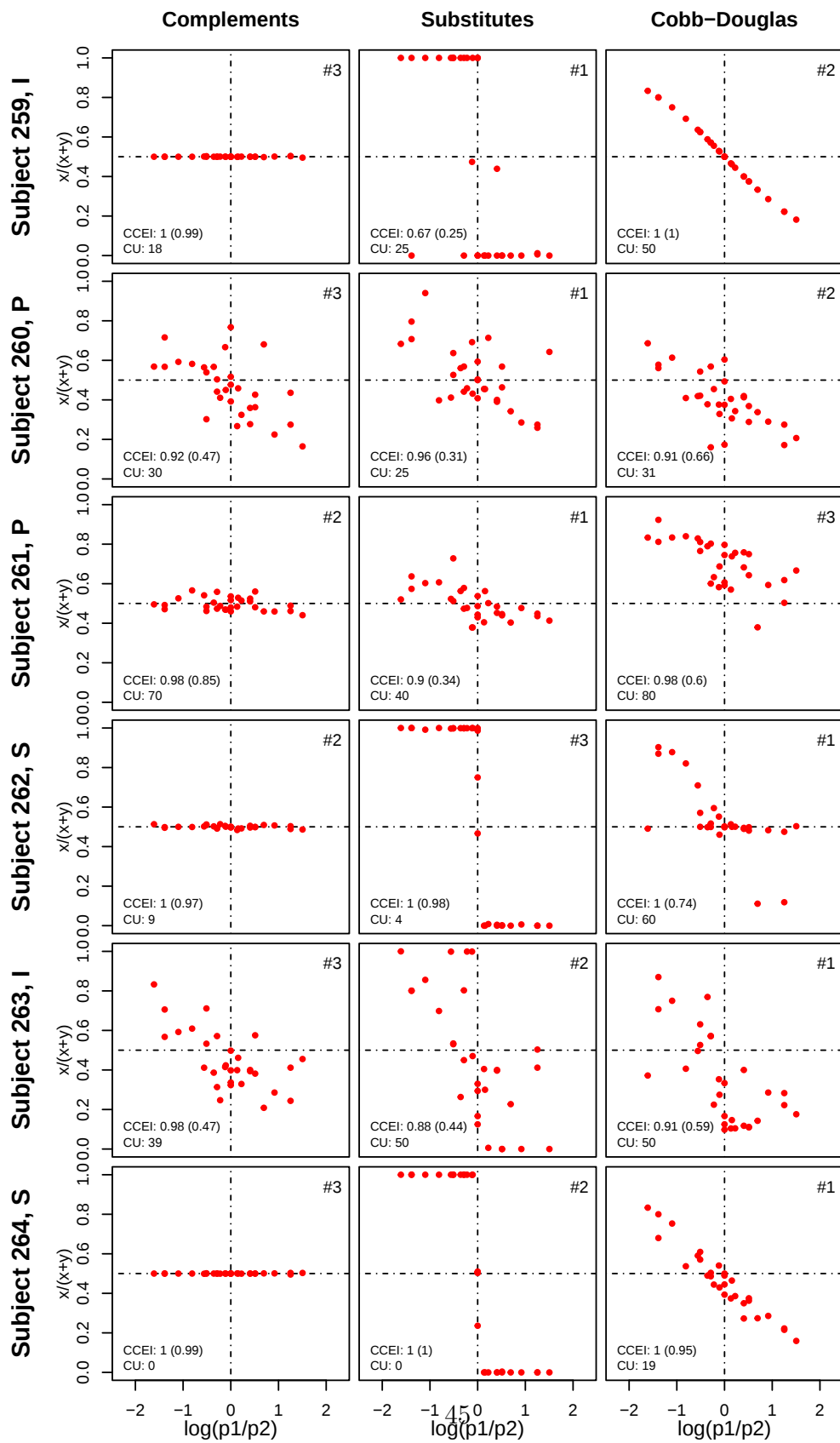


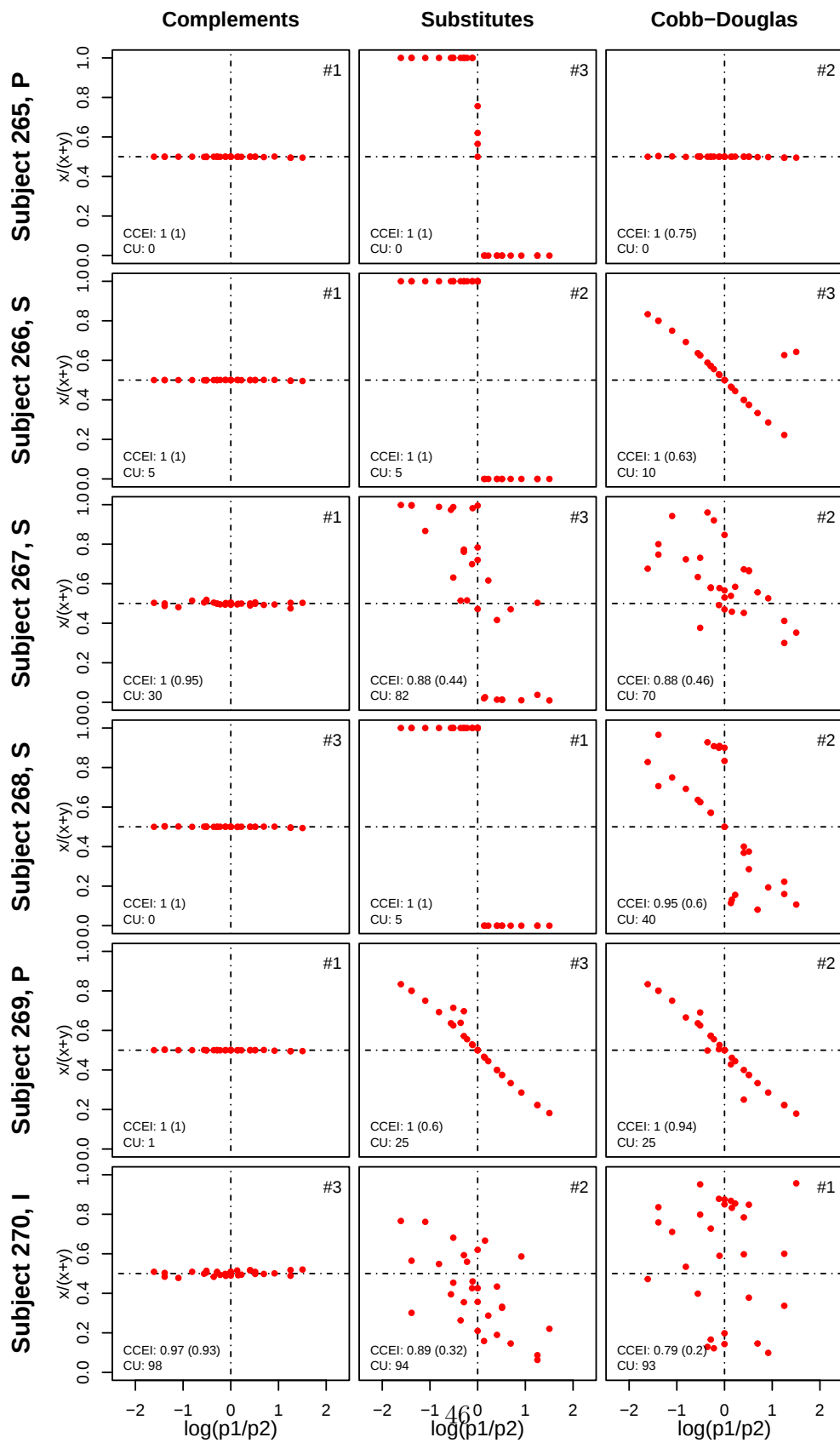


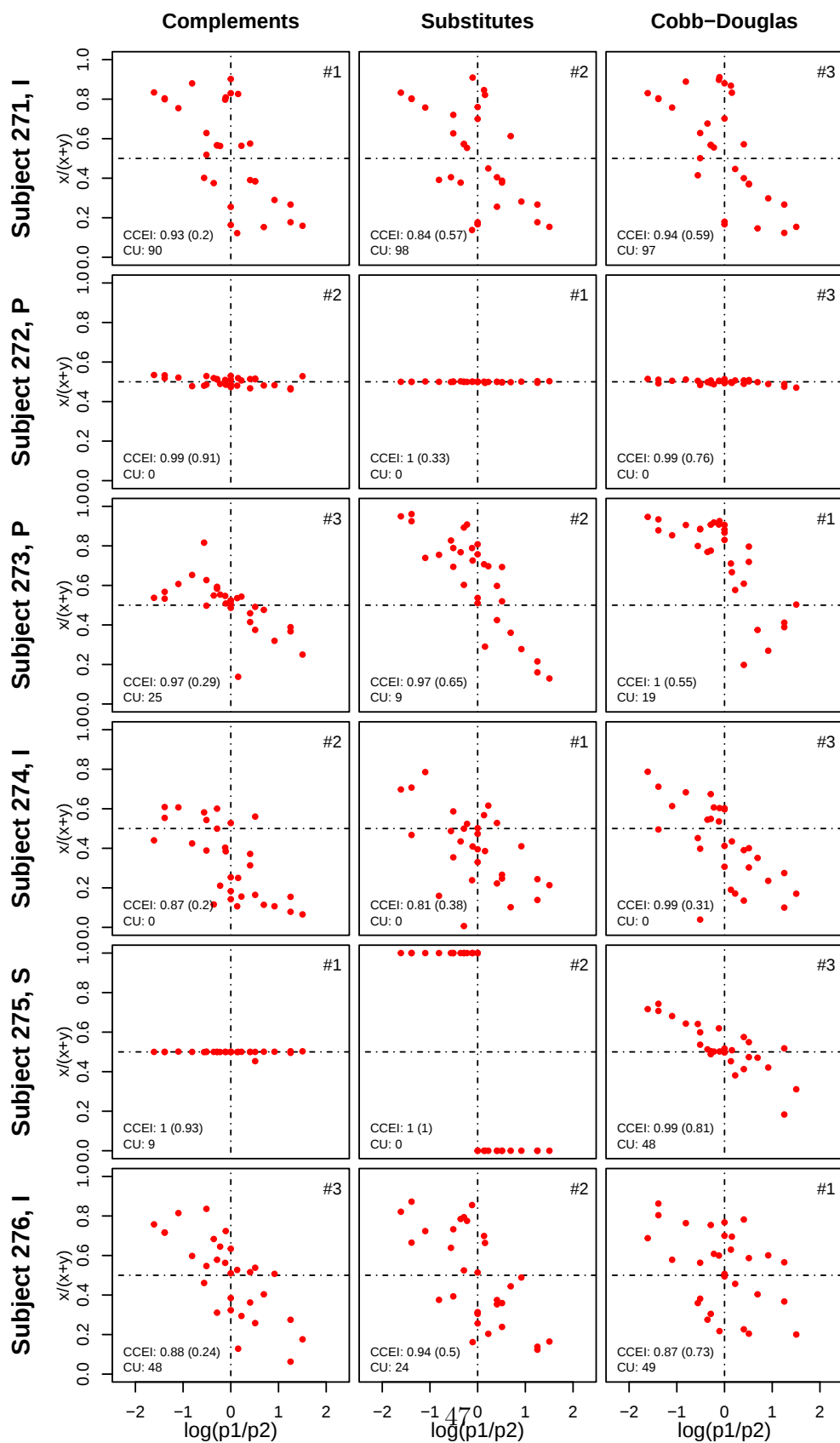


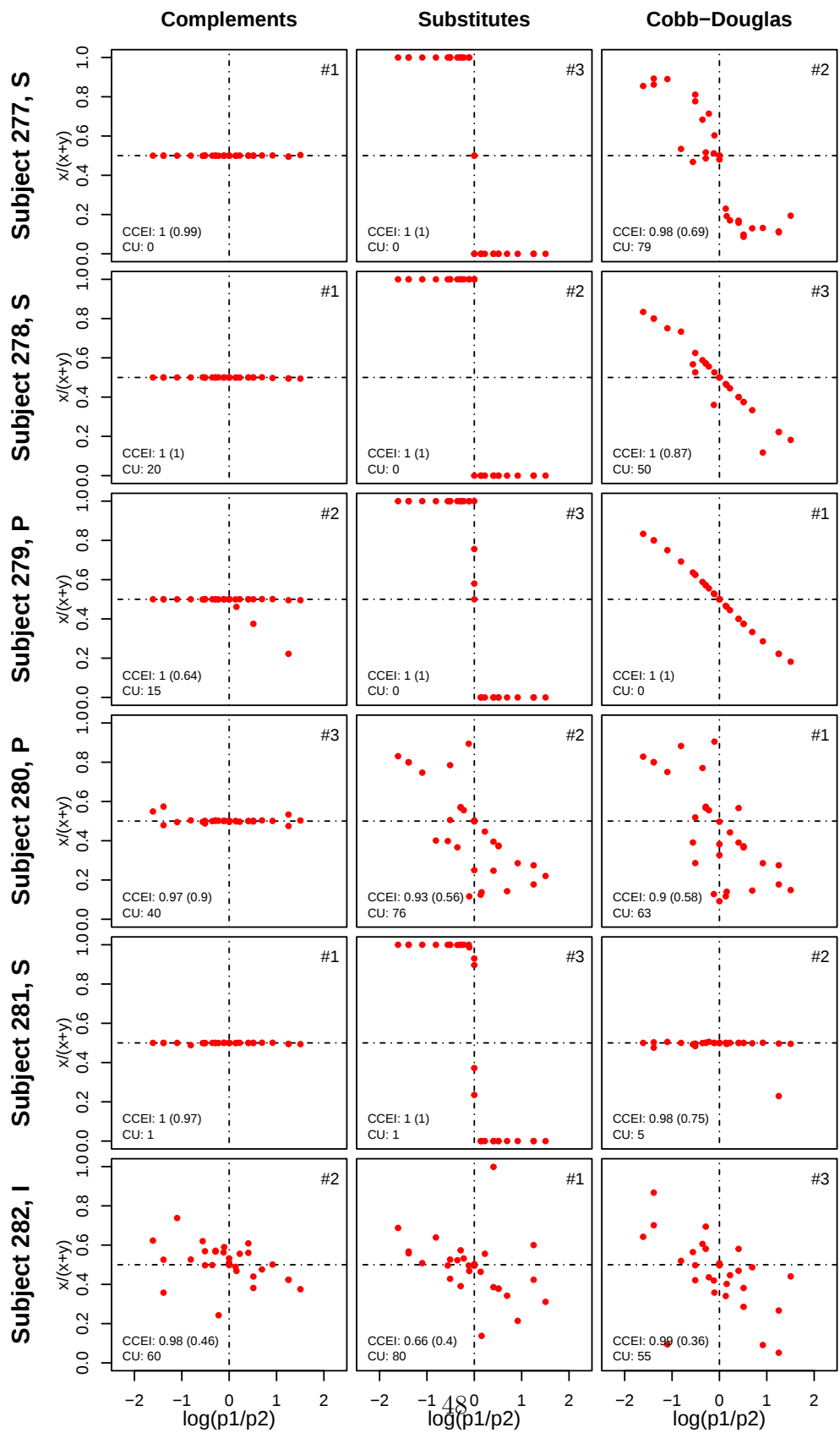


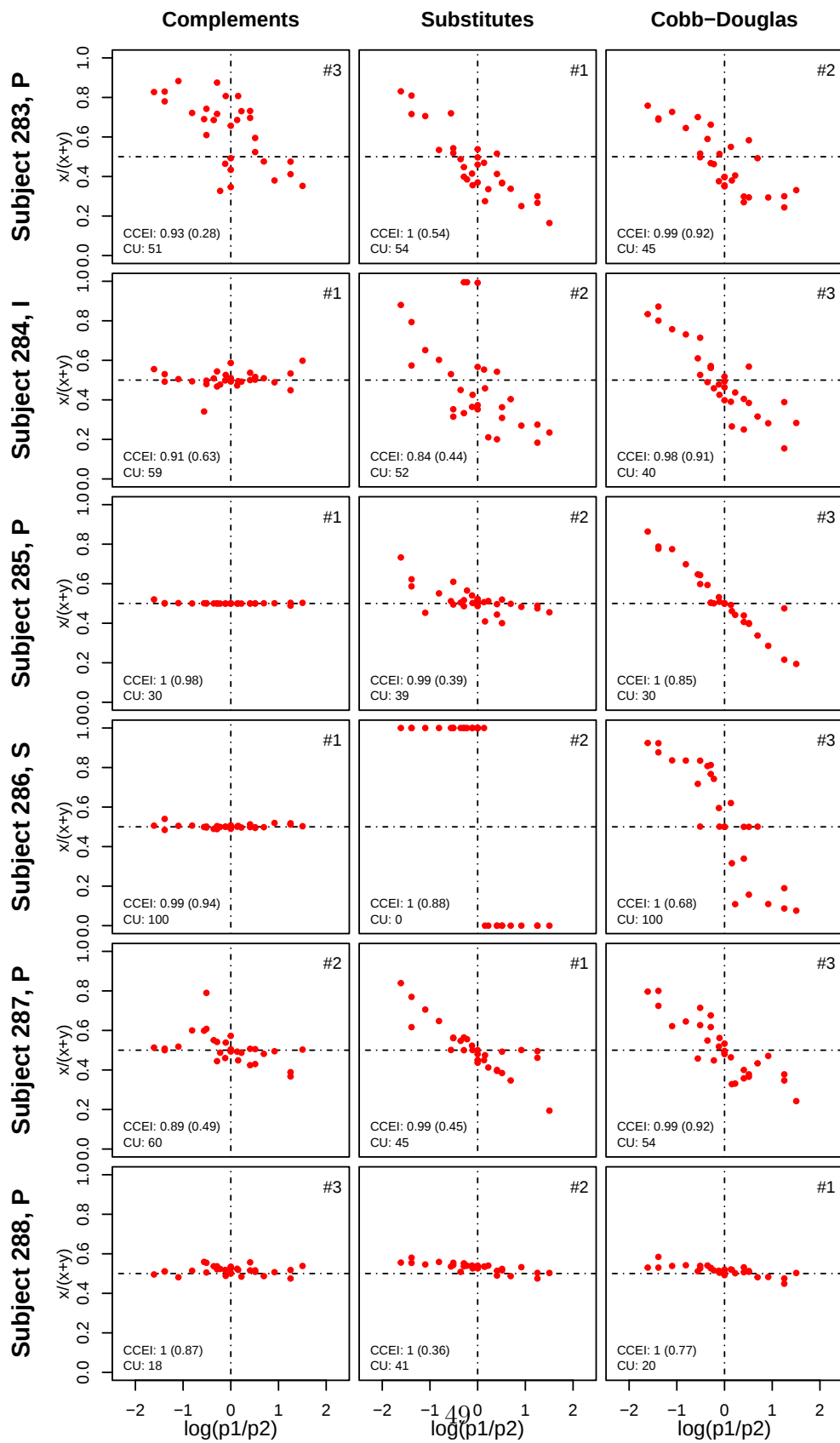


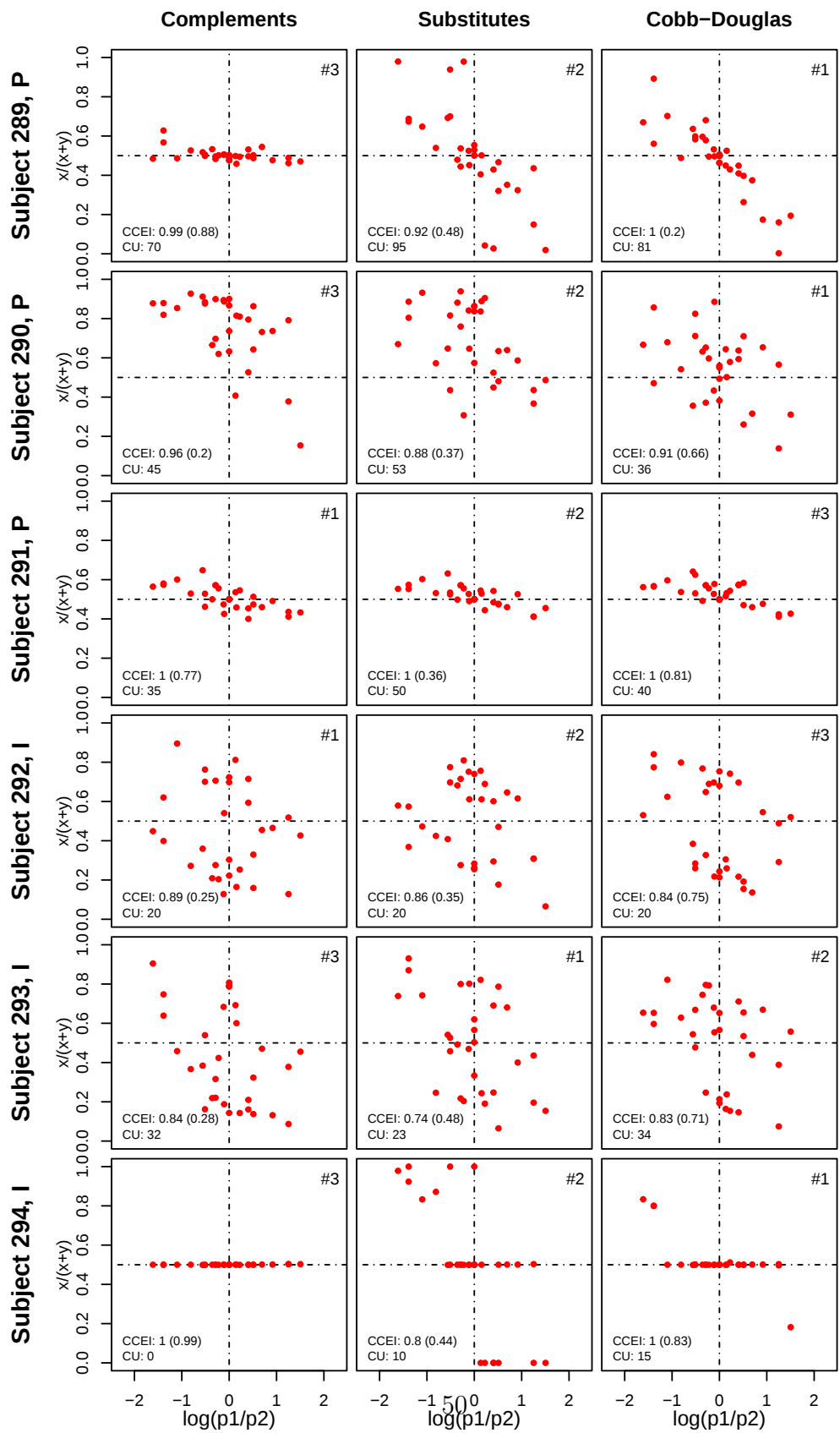


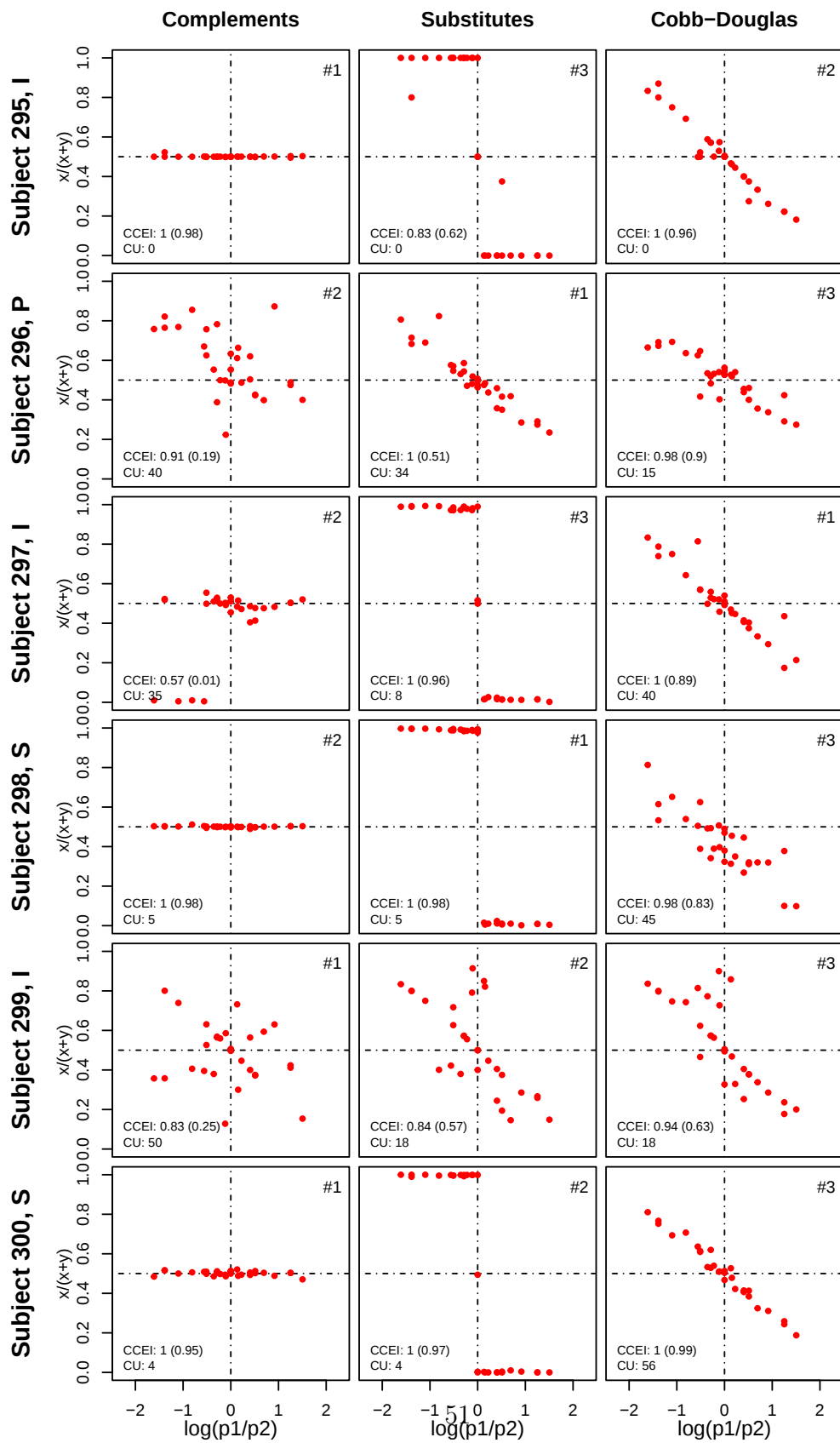


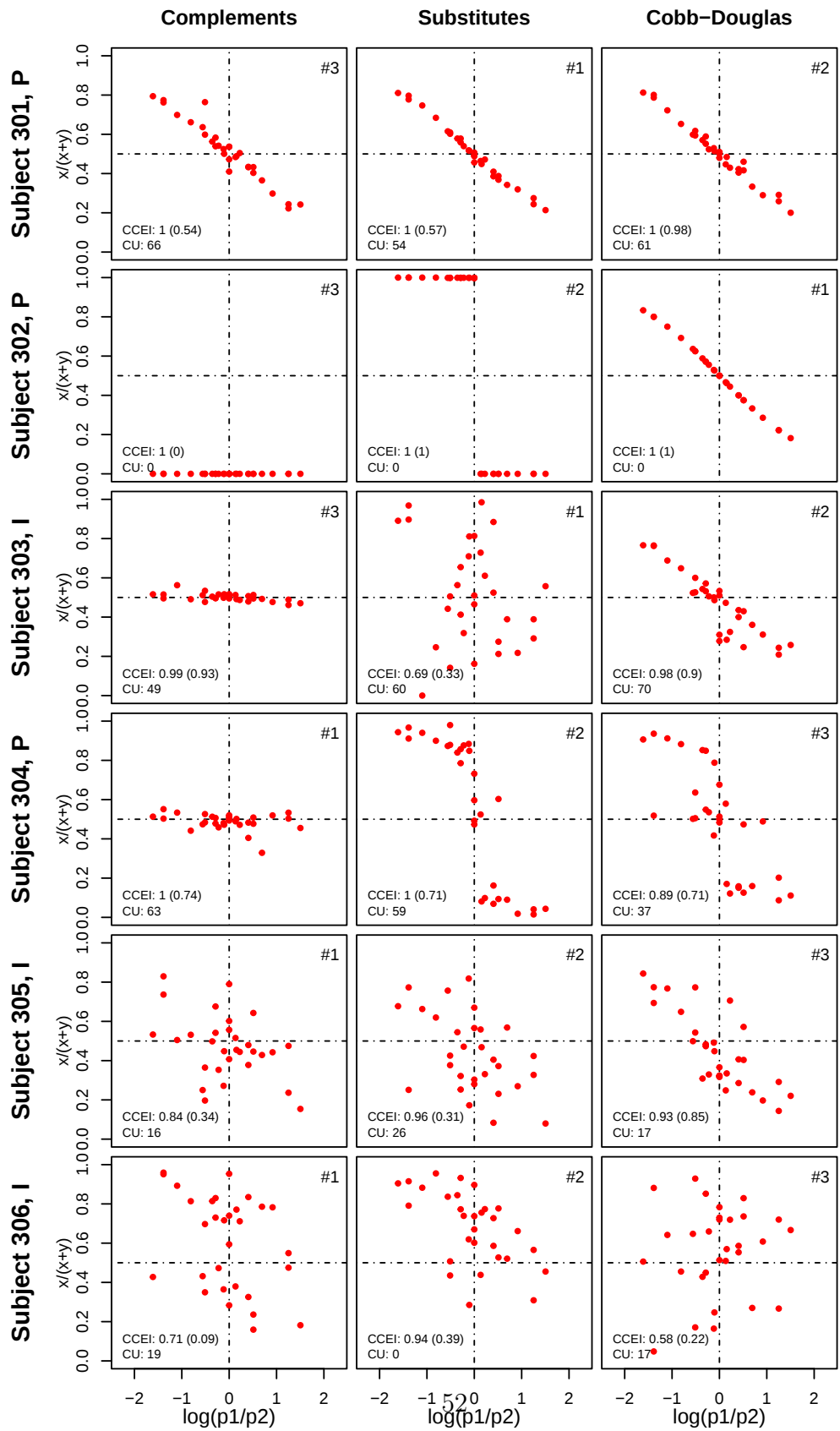


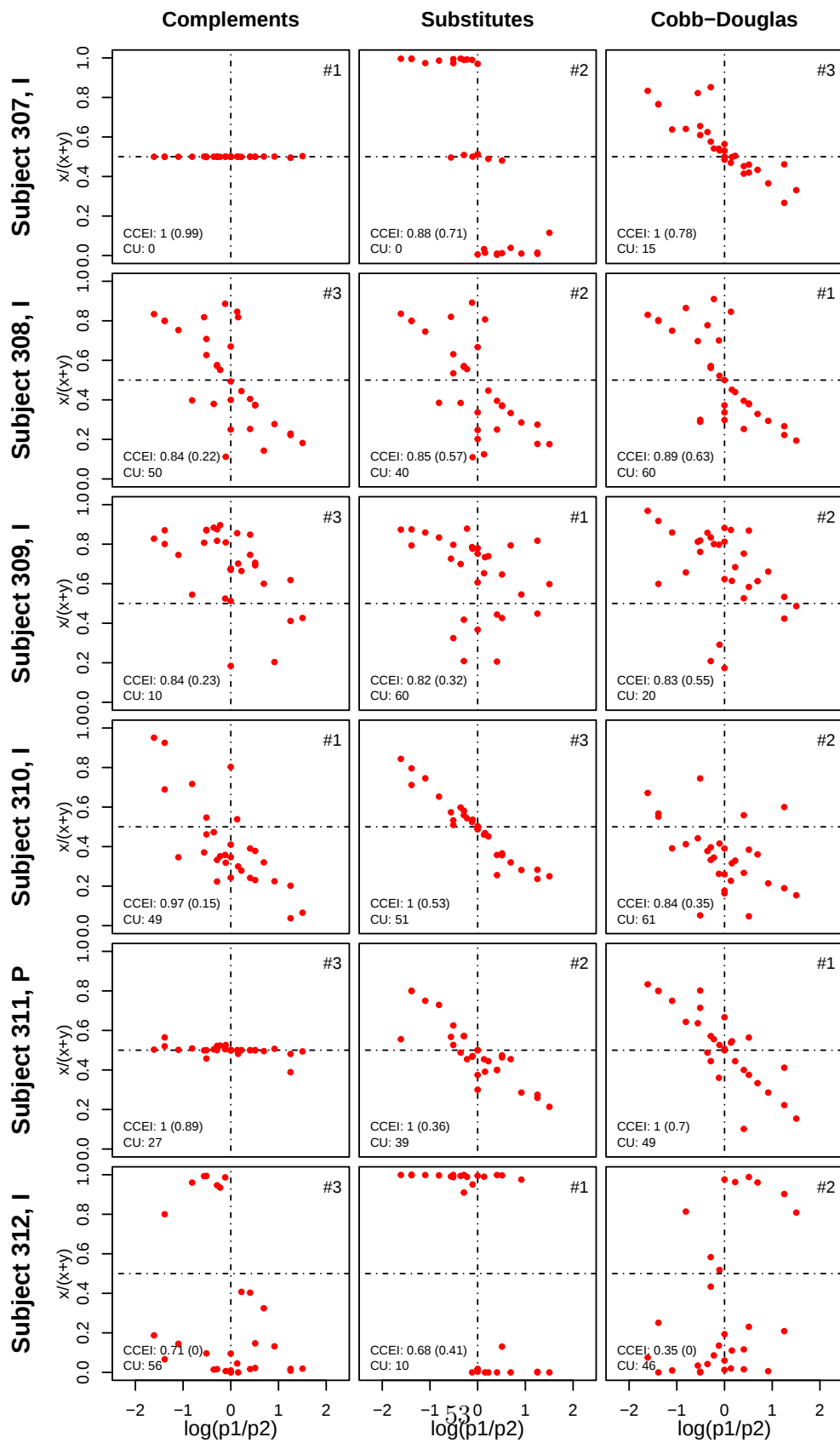


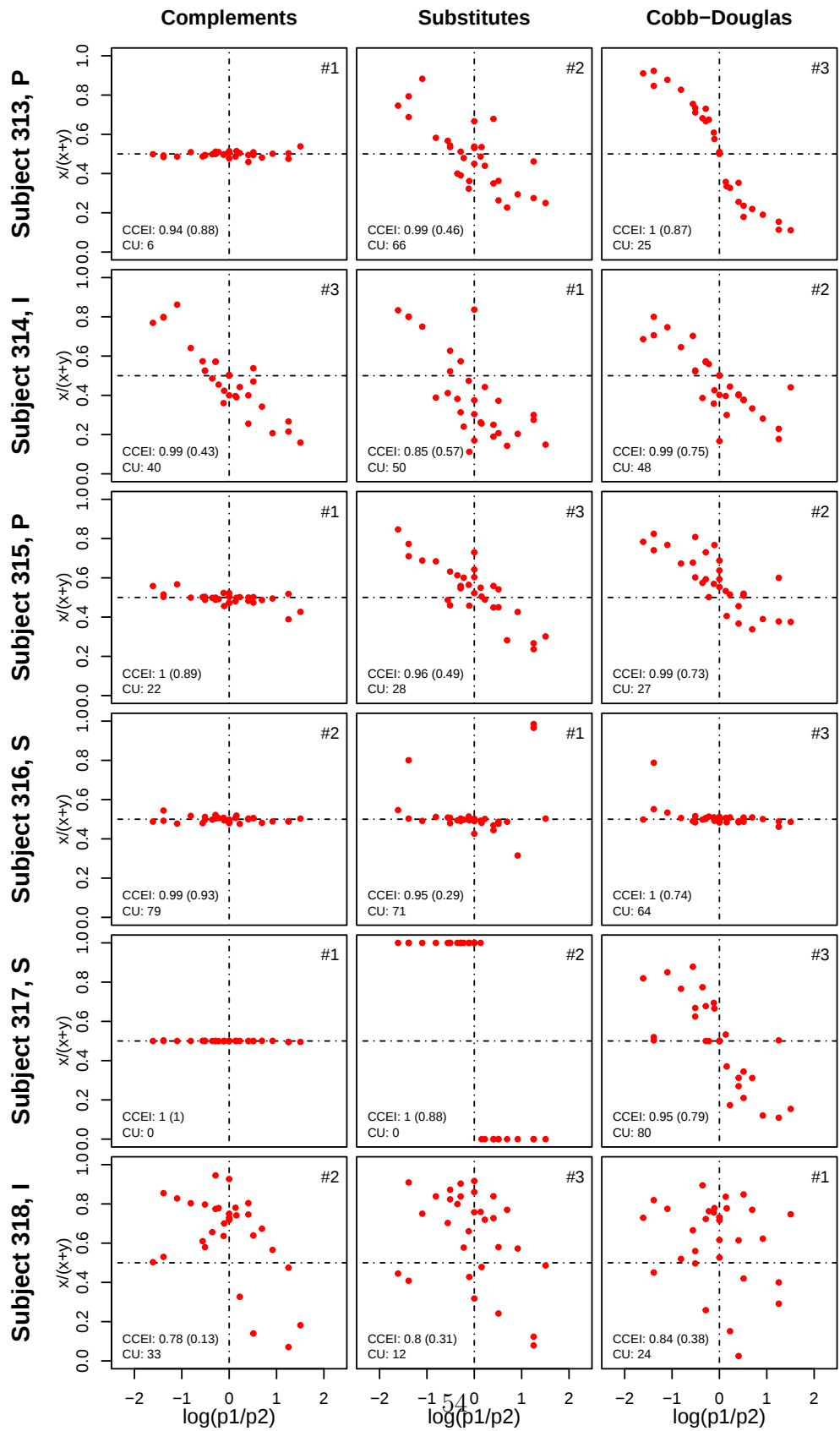


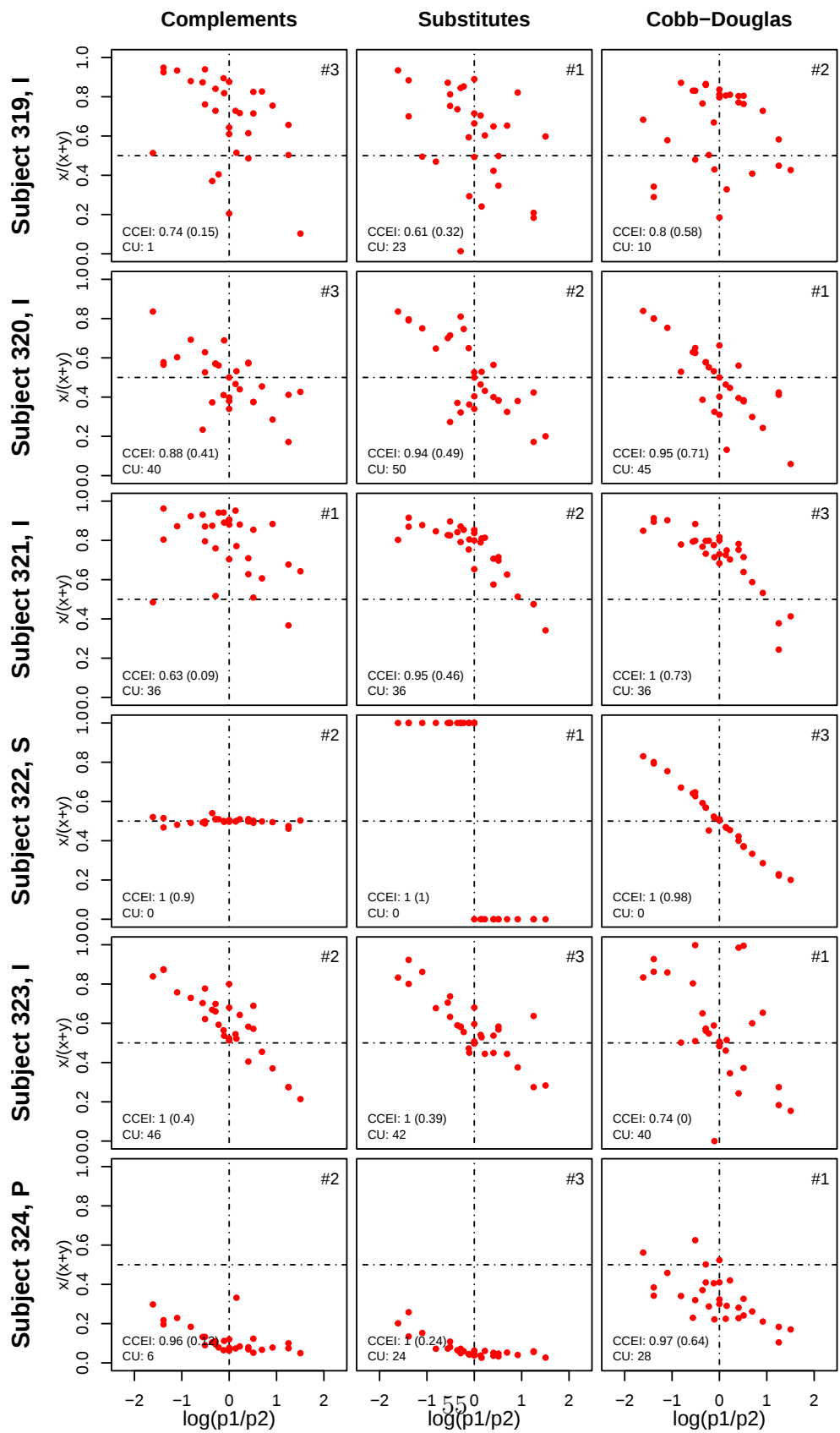


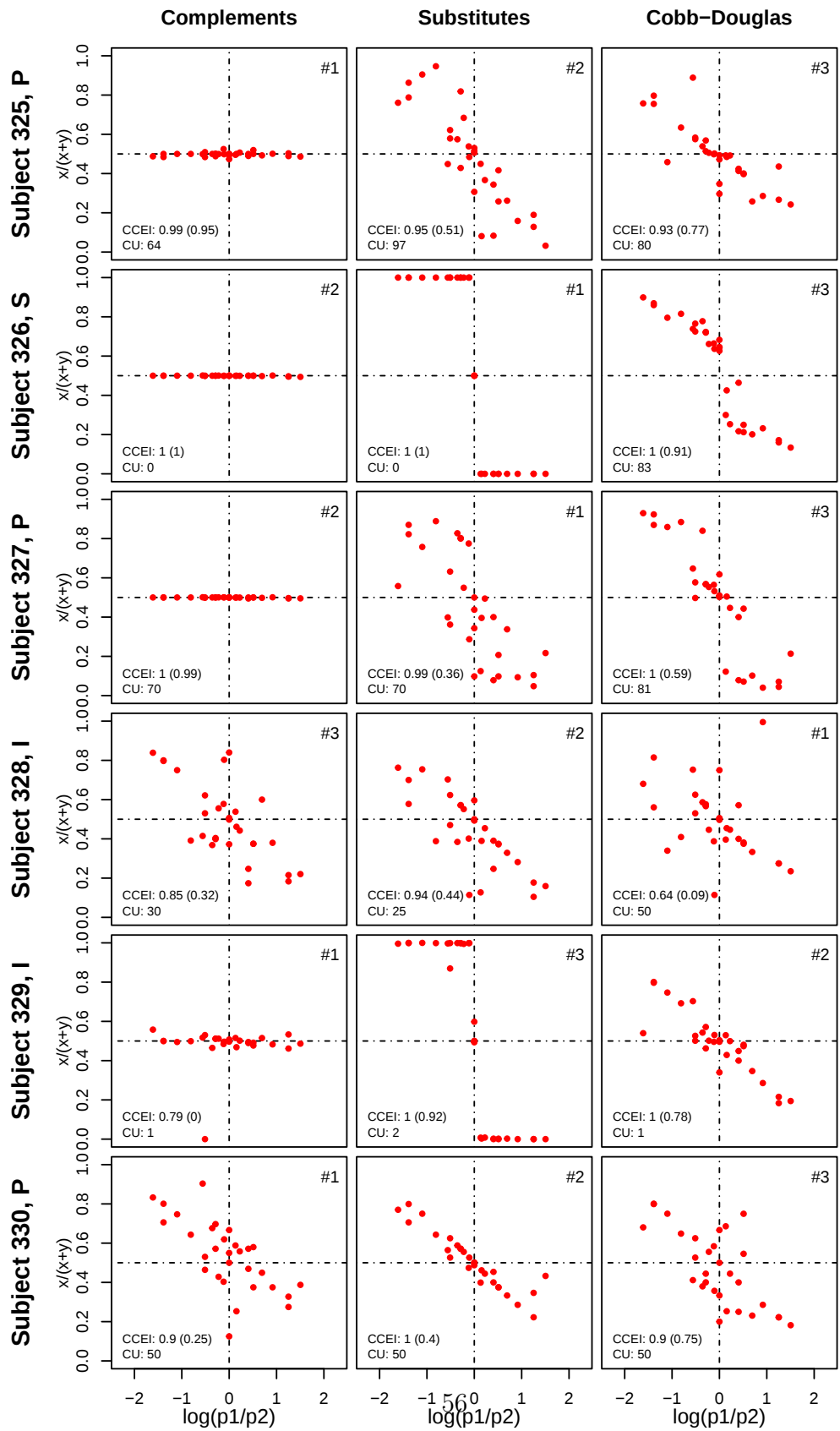


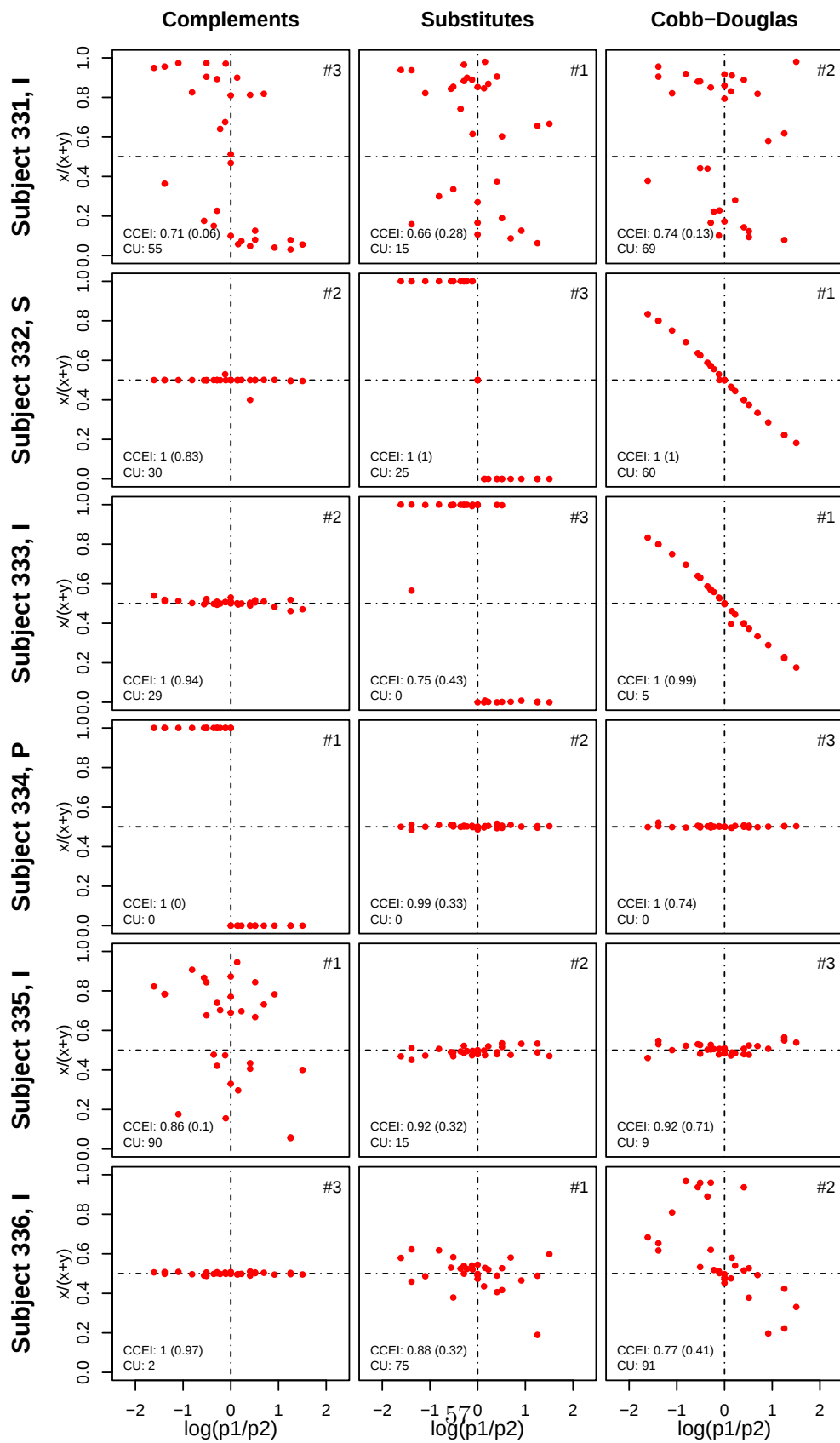


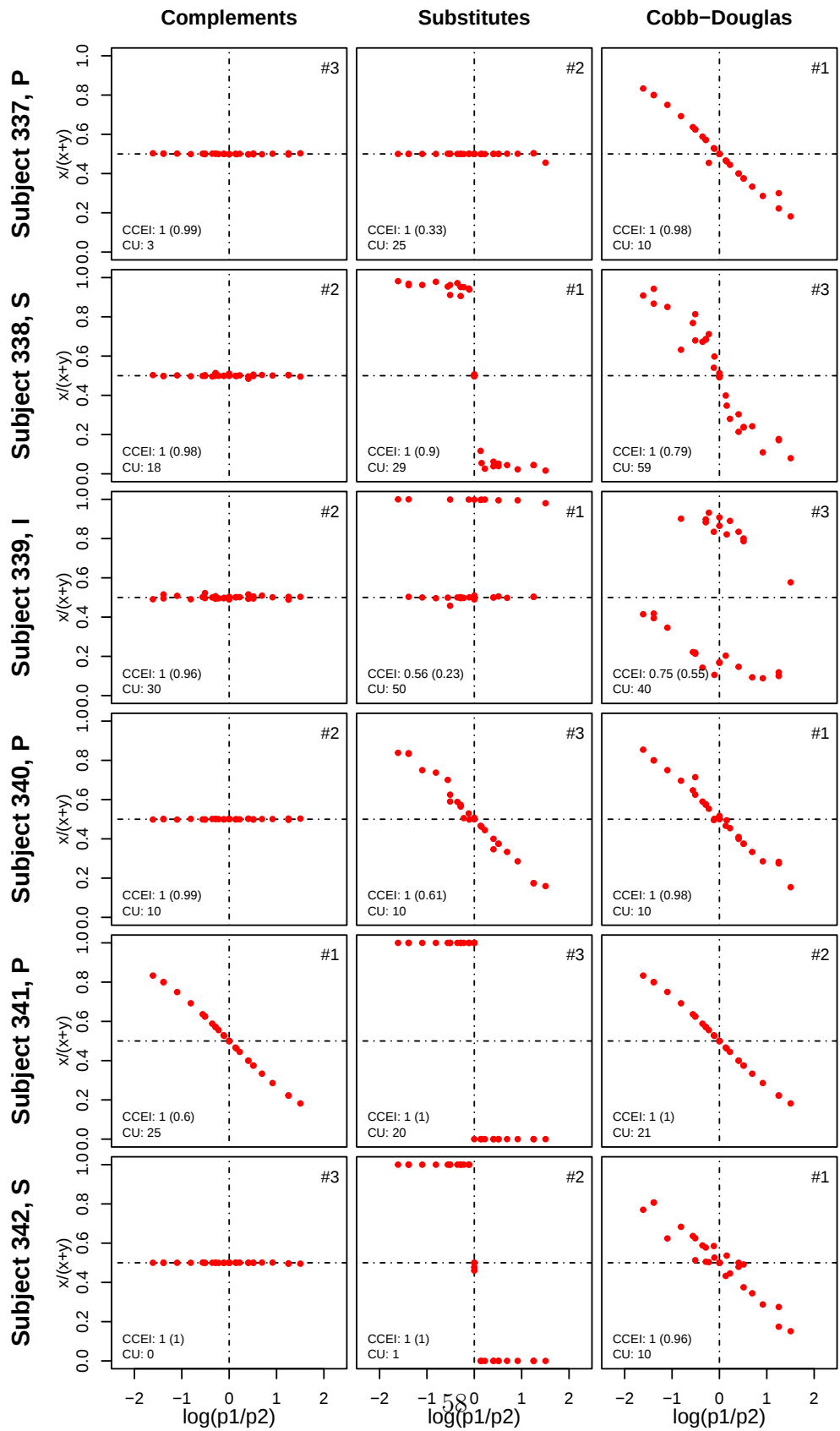


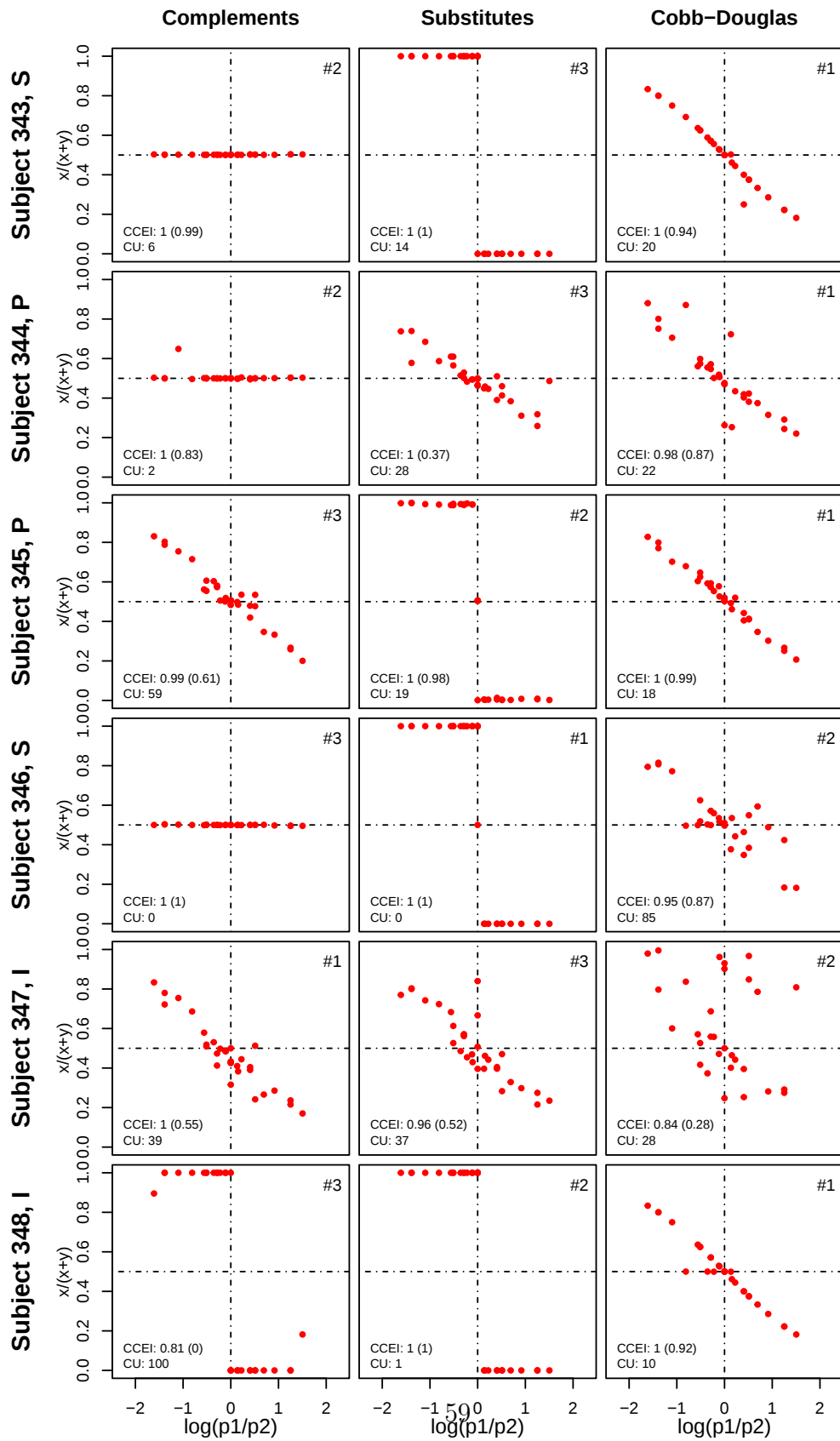


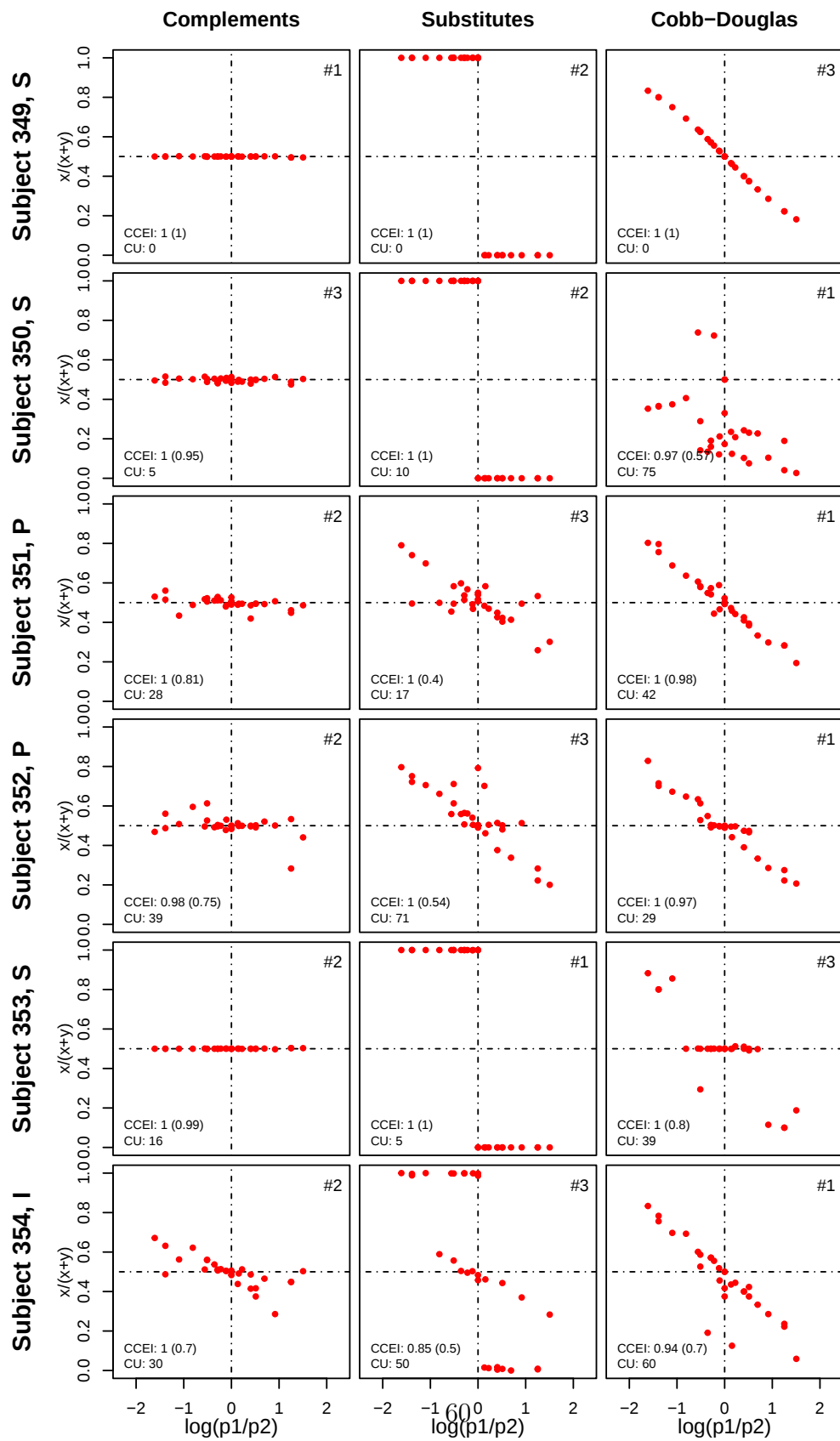


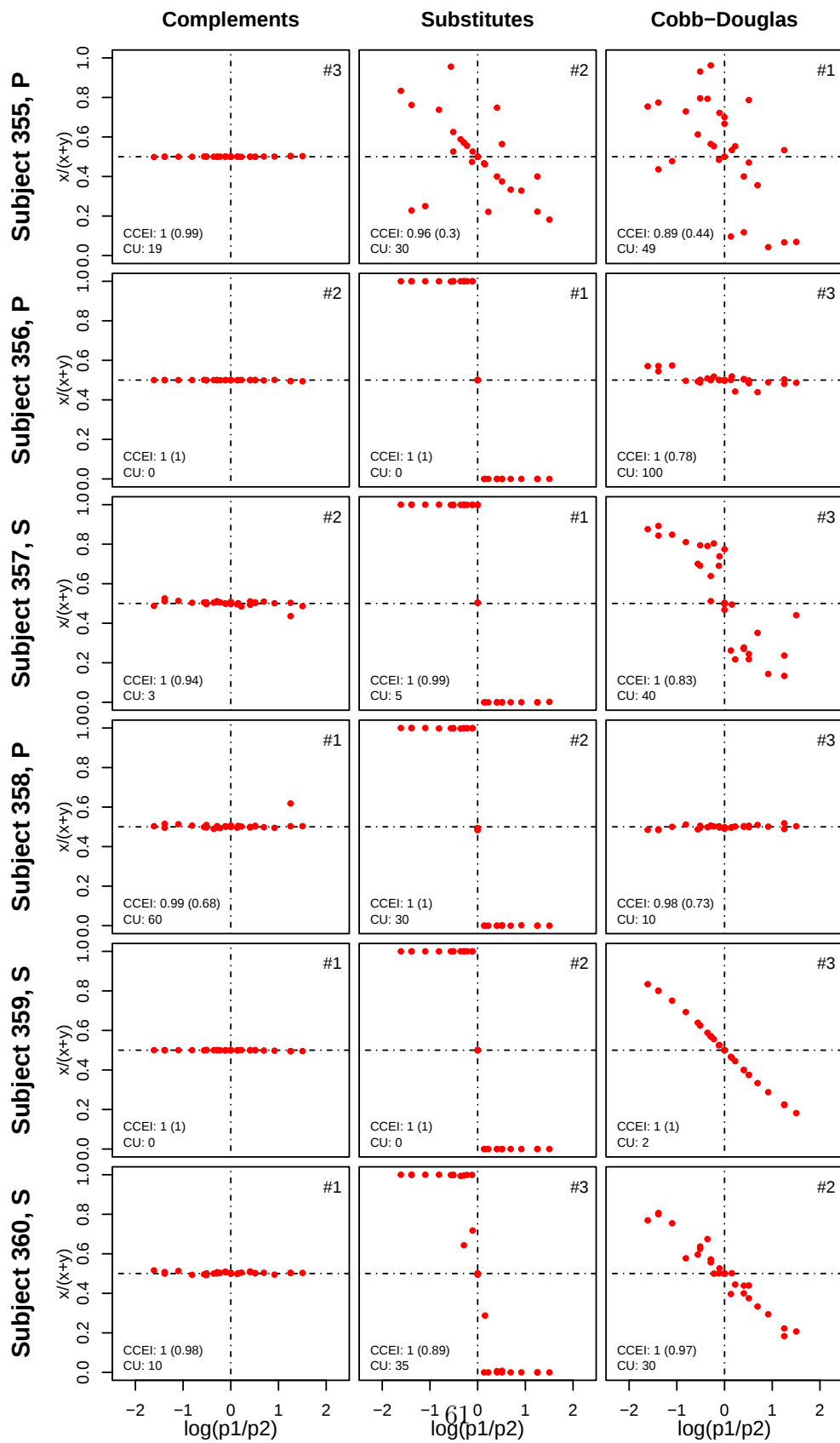


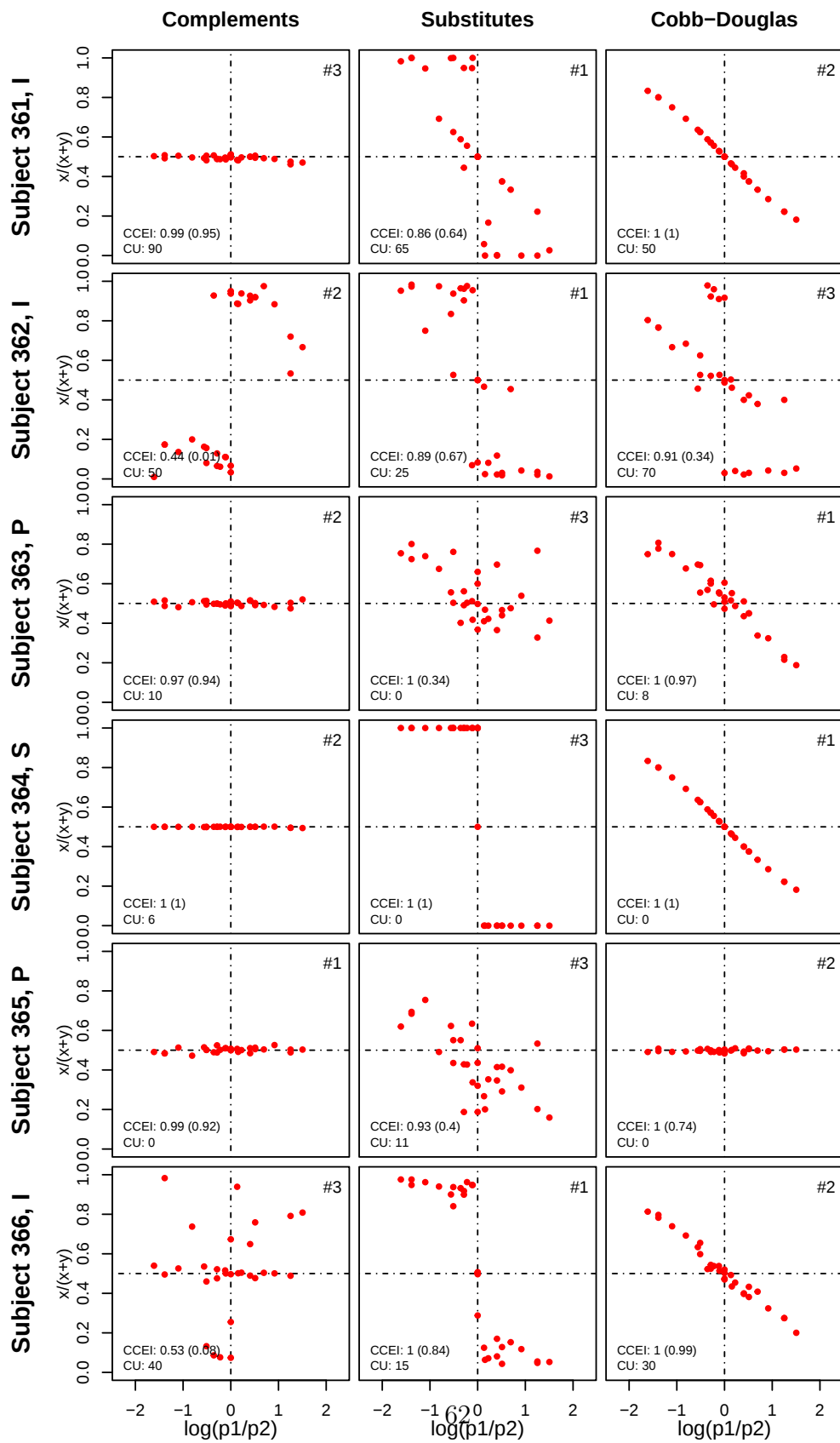


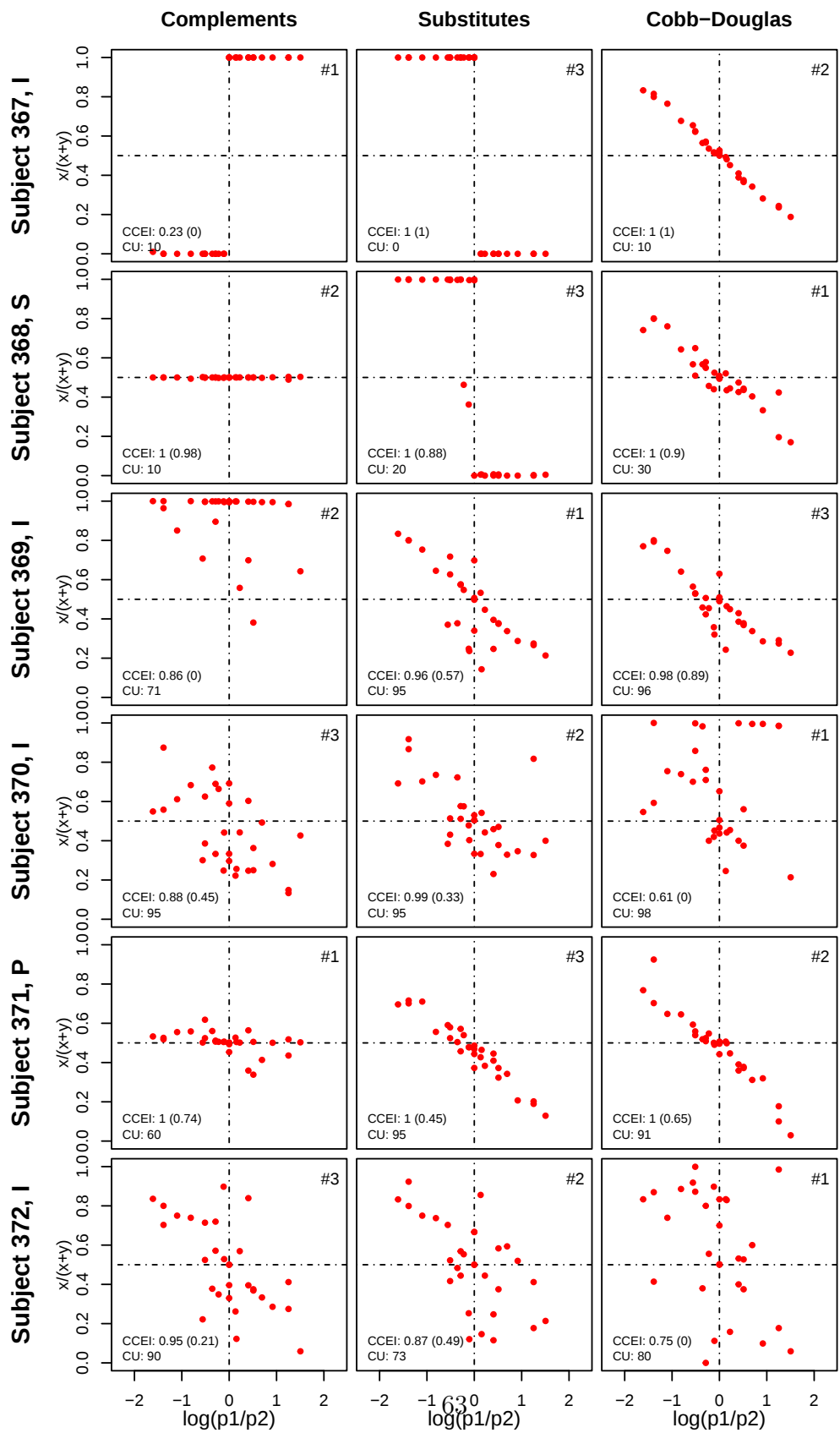


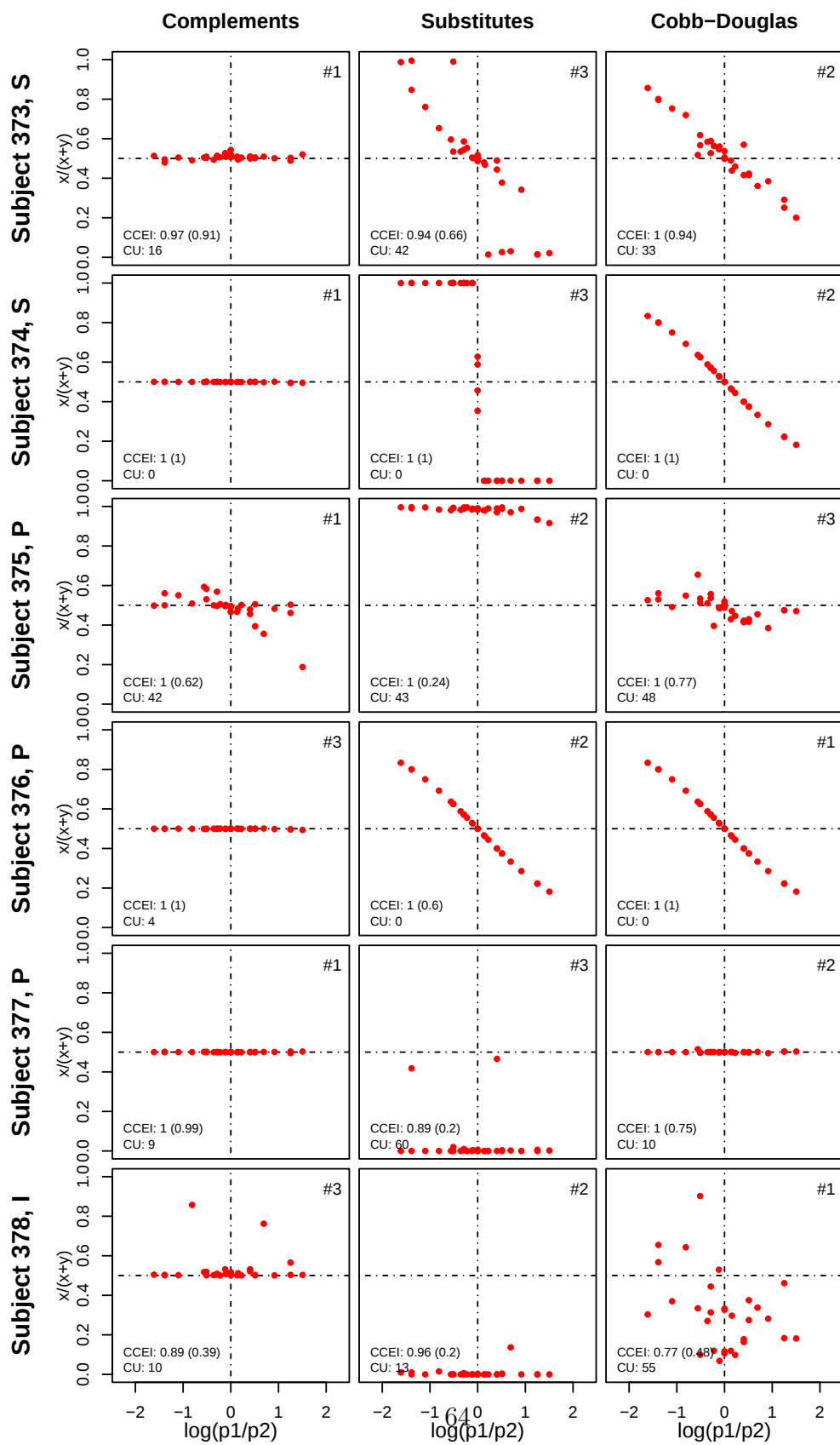


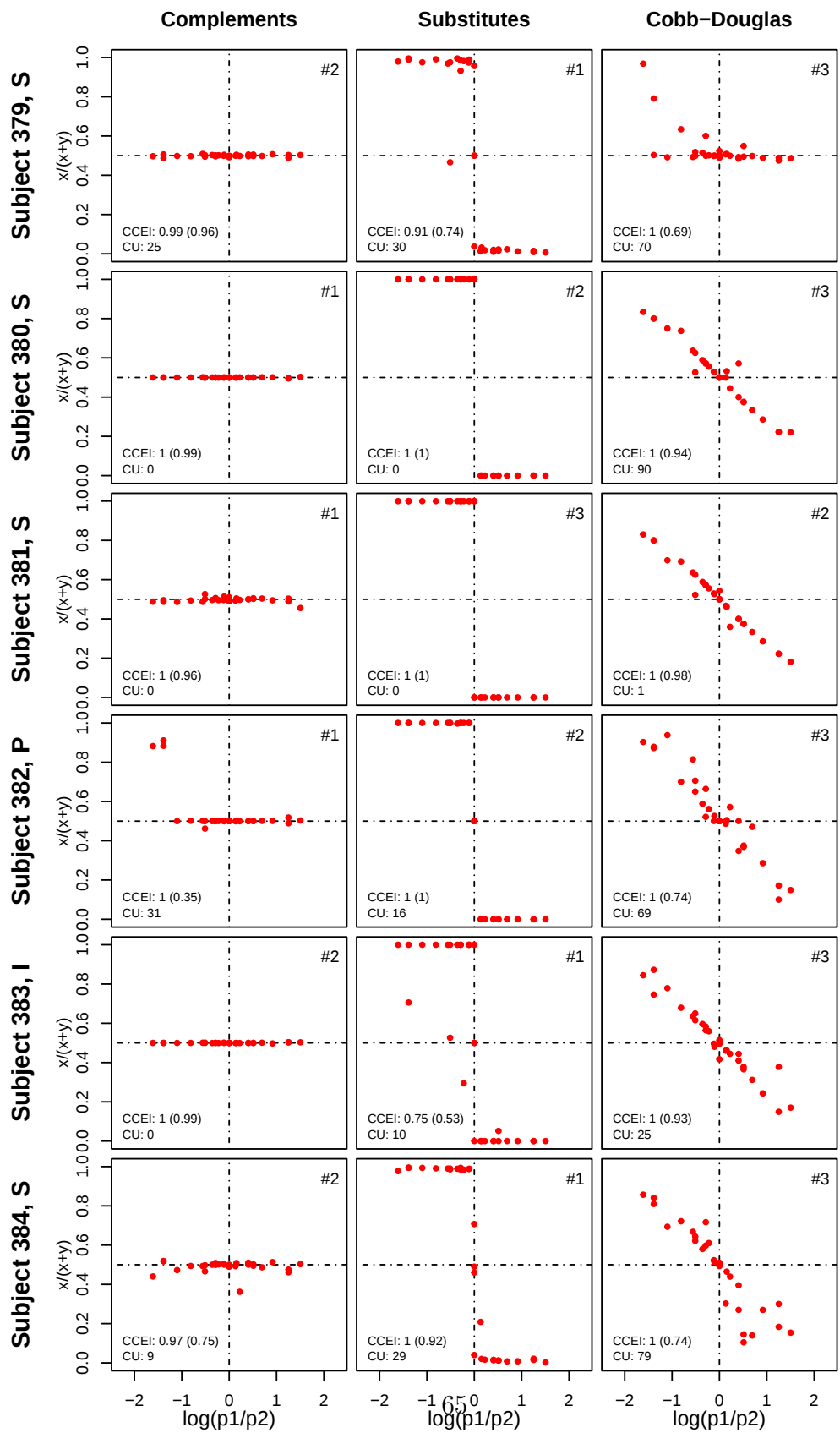


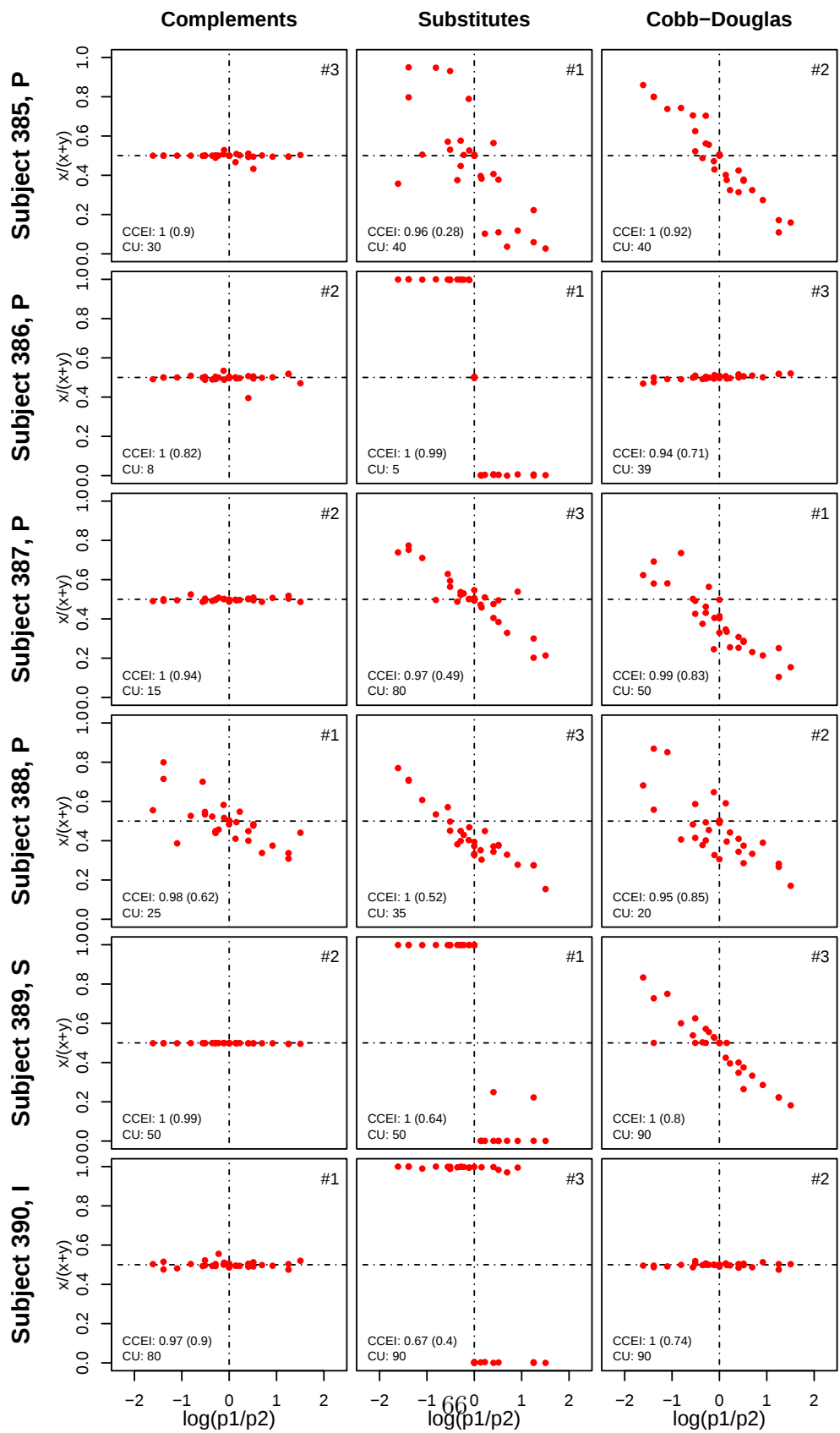


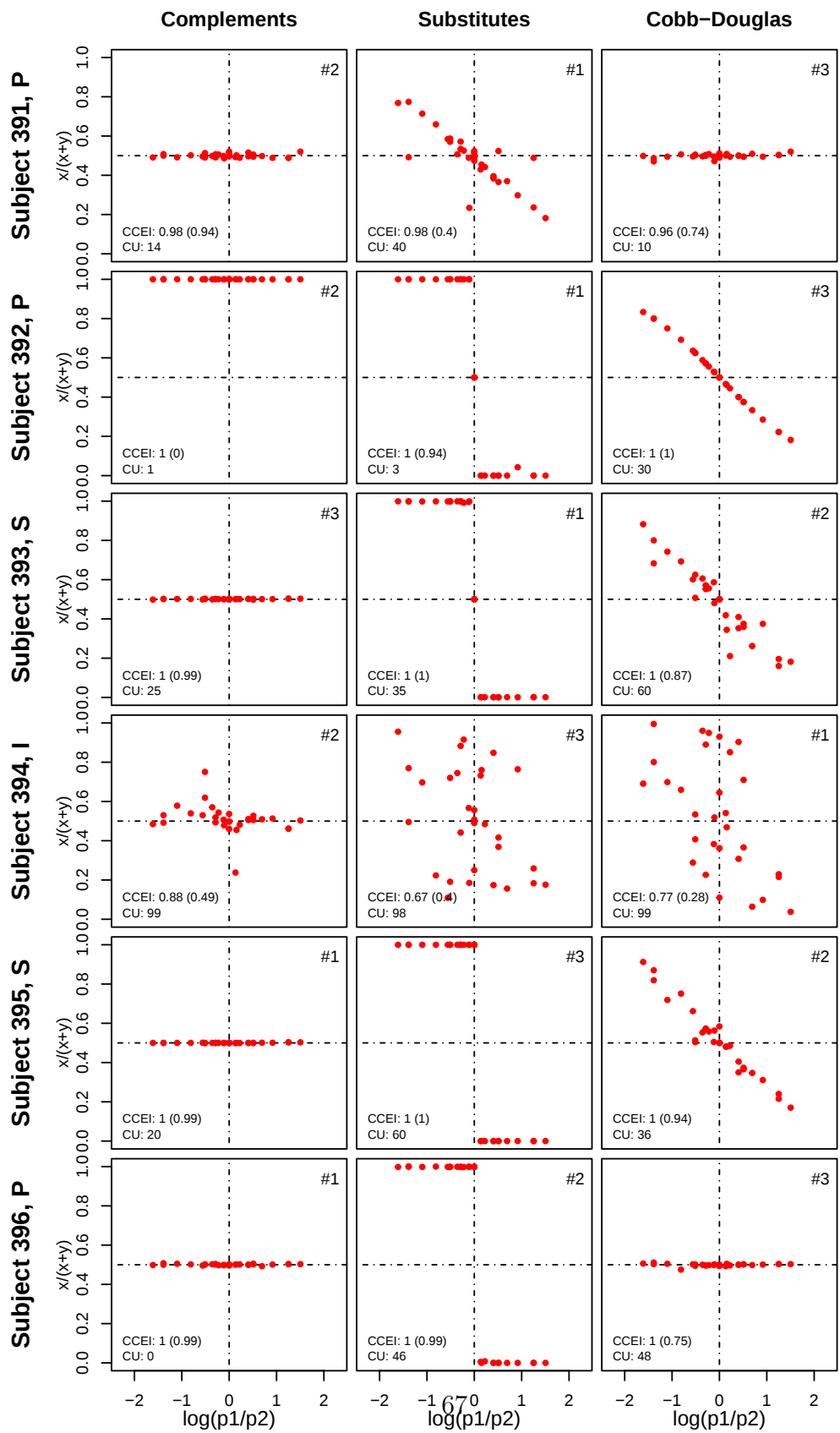


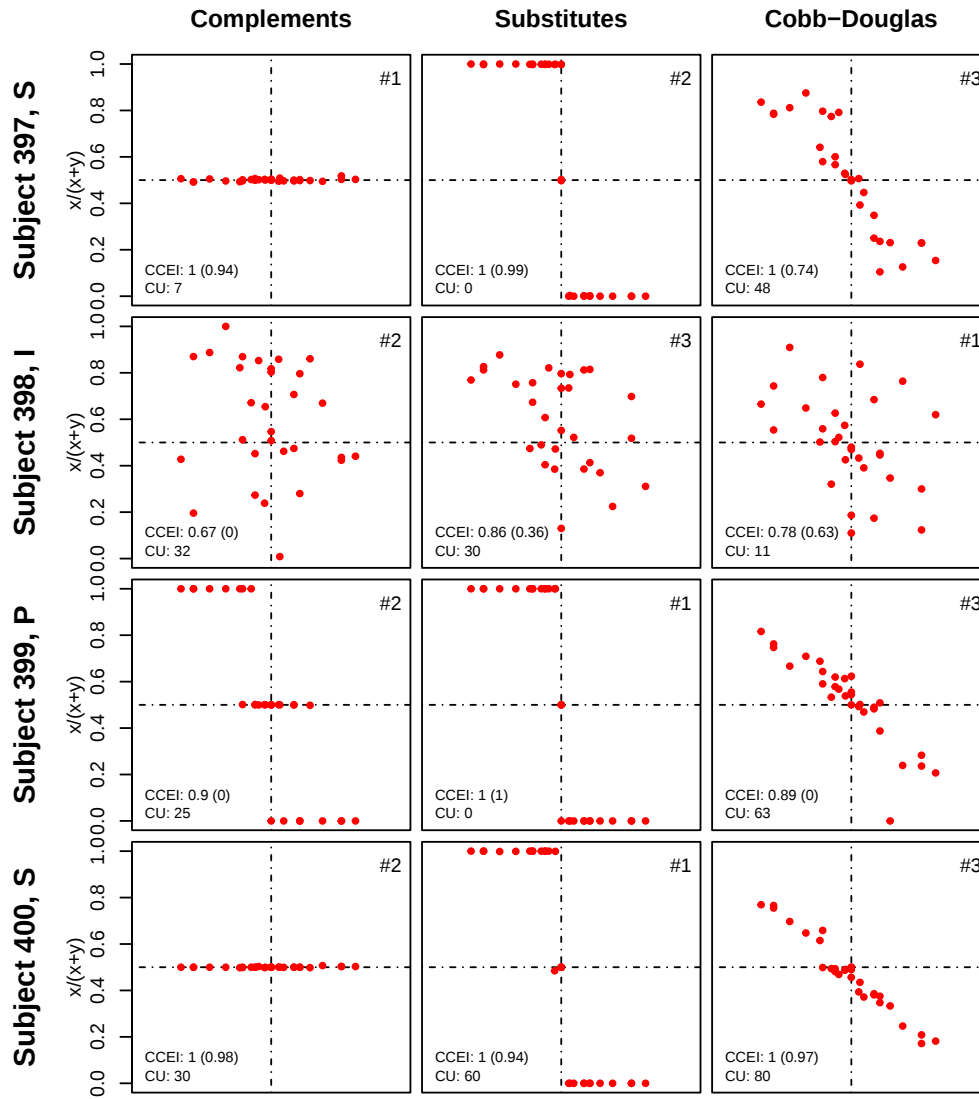












OA.2 Screenshots of the Experimental Interface

We include screenshots of instructional screens, quizzes, representative decision screens for each induced preference, and the CRT questions used. We show the budget interface before clicking, and include a post-clicking example to illustrate that display. We similarly show a post-clicking example for the confidence question.

Introduction

- We will start by providing you with **instructions** for the study.
- Please read and follow the instructions closely and carefully.
- If you complete the main parts of the study, you will receive a **Sure Payment** of \$10.
- We will ask you **comprehension questions** about the instructions. You will receive extra payment if you answer these correctly the first time.
- 20% of participants will be randomly selected to be paid **Additional Earnings** (on top of the sure payment), which depend on your decisions and chance.

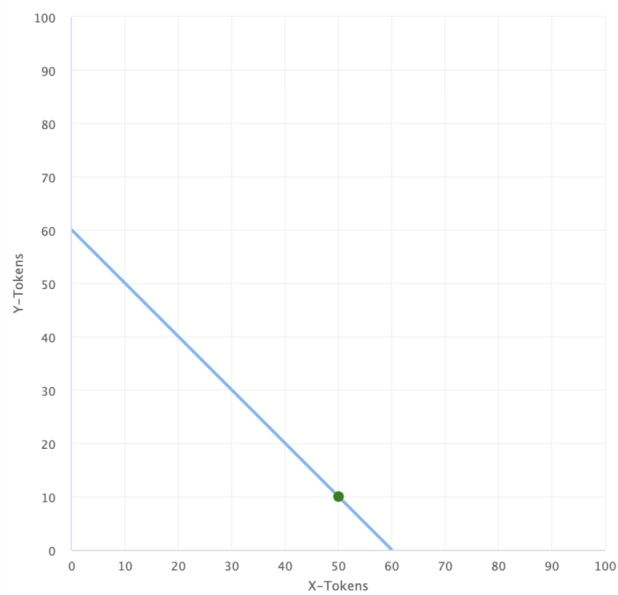


Tokens and Earnings Points

- This experiment consists of **3 Parts**, each with **30 Tasks**.
- In each Task, you can acquire two different types of Tokens: **X-Tokens** and **Y-Tokens**
- Each Part starts by telling you the **Earnings Formula** that will apply in the ensuing Tasks.
- The Earnings Formula explains how the number of X-Tokens and Y-Tokens you acquire in a task translates to **Earnings Points**.
- At the end of the survey, 20% of participants are randomly selected to be paid **Additional Earnings based on the decision they made in one randomly drawn Task**.
- **The more Earnings Points you get in a task, the larger your potential Additional Earnings will be.**

Making Your Choice

In each task you will decide how many X-Tokens and Y-Tokens to acquire using a screen like the one below.



X-Tokens: 50.00 Y-Tokens: 10.00



X-Tokens: 50.00 Y-Tokens: 10.00

- On the x-axis (the horizontal axis) is a range of possible X-Token amounts; on the y-axis (the vertical axis) is a range of possible Y-Token amounts.
- The **blue line** on the graph shows all of the **possible combinations** of X-Tokens and Y-Tokens you can **choose from** for this Task.
- To select a combination, simply click on the **blue line**. The higher (further to the left) you click, the more Y-Tokens you're selecting; the lower (further to the right) you click, the more X-Tokens you're selecting. After clicking, you can also use right/left **arrow keys** to fine tune, or click somewhere else.
- The current selection will be shown as a **green dot**. The **exact numbers** of X- and Y- Tokens associated with the selection are listed just below the graph. For instance, the green dot in the graph above corresponds to 50 X-Tokens and 10 Y-Tokens.



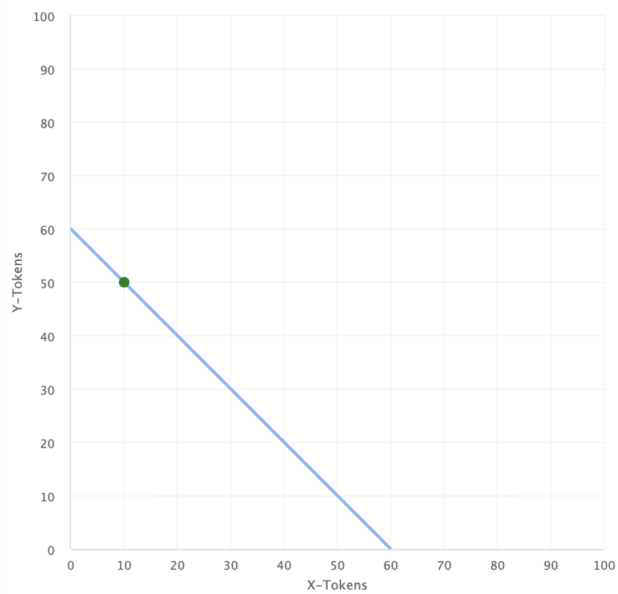
Finalizing Your Decisions in a Task

- You can click on the line and press the left/right arrow keys as many times as you want to **fine tune** your choices.
- When you've finalized your decision, click the "Submit Your Choice" button to **SUBMIT** your choice and proceed to the next Task.
- The **blue line** shown on your screen will change across Tasks.



Comprehension Questions

If you answer the following questions correctly **on the first try** we will give you an **extra \$1** in sure payment. The questions on this page involve the following Figure:



X-Tokens: 10.00 Y-Tokens: 50.00

Question #1: The current choice is shown on the graph by

- ☐ The current choice isn't shown on the graph
- ☐ The blue line
- ☐ The gray lines
- ☐ The green dot

Question #2: How do you select how many X- and Y-Tokens to acquire?

- ☐ By pressing the space bar.
- ☐ By clicking on the blue line or using arrow keys to fine tune
- ☐ By clicking on the x- or y-axis.
- ☐ By pressing the Enter key.

Question #3: In the example above, how many X-Tokens are selected?

☐ 50

☐ 100

☐ 0

☐ 10

Question #4: In the example above, how many Y-Tokens are selected?

☐ 50

☐ 100

☐ 10

☐ 0



Comprehension Quiz on Earnings Formulas

- To confirm your understanding of Earnings Formulas, we ask you to calculate some simple examples.
 - You get **an extra \$1** in sure payment if you do all calculations right the first time.
(Enter only numeric values in each response box.)
-

Multiply Earnings Formula: your X-Tokens **times** your Y-Tokens.

This means the computer will multiply your X-Tokens times your Y-Tokens to determine your Earnings Points.

In other words, your Earnings Points will be: X-Tokens times Y-Tokens.

Question #1: If you have 4 X-Tokens and 2 Y-Tokens, how many Earnings Points will the computer give you under the **Multiply Earnings Formula**? Enter the number in the box below.

Average Earnings Formula: the **average** of your X-Tokens and your Y-Tokens.

This means the computer will add up your chosen X-Tokens and Y-Tokens and divide this sum in half to determine your Earnings Points.

In other words, your Earnings Points will be: (X-Tokens + Y-Tokens)/2

Average Earnings Formula: the **average** of your X-Tokens and your Y-Tokens.

This means the computer will add up your chosen X-Tokens and Y-Tokens and divide this sum in half to determine your Earnings Points.

In other words, your Earnings Points will be: $(X\text{-Tokens} + Y\text{-Tokens})/2$.

Question #2: If you have 4 X-Tokens and 2 Y-Tokens, how many Earnings Points will the computer give you under the **Average Earnings Formula**? Enter the number in the box below.

Minimum Earnings Formula: your X-Tokens **or** your Y-Tokens, whichever is **smaller**.

*This means the computer will compare the number of X-Tokens to the number of Y-Tokens you have selected and pick whichever is **smaller** to be your Earnings Points.*

In other words, your Earnings Points will be the minimum of your X-Tokens and your Y-Tokens.

Question #3: If you have 4 X-Tokens and 2 Y-Tokens, how many Earnings Points will the computer give you under the **Minimum Earnings Formula**? Enter the number in the box below.



Part 1: The **Minimum** Earnings Formula

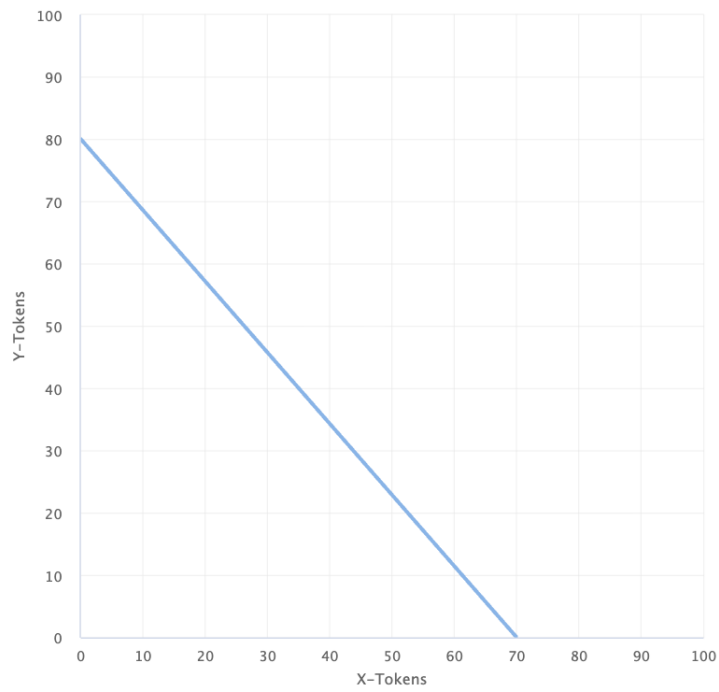
- In each of the next 30 Tasks, the Earnings Formula will be:

Earnings Points = Your X-Tokens or your Y-Tokens, whichever is smaller.

- This means the computer will compare the number of X-Tokens to the number of Y-Tokens you have selected and pick whichever is **smaller** to be your Earnings Points.
- In other words, your Earnings Points will be the minimum of your X-Tokens and your Y-Tokens.
- If one of these Tasks is chosen for payment, you will get Additional Earnings equal to 25 cents for each Earnings Point you get in that Task.



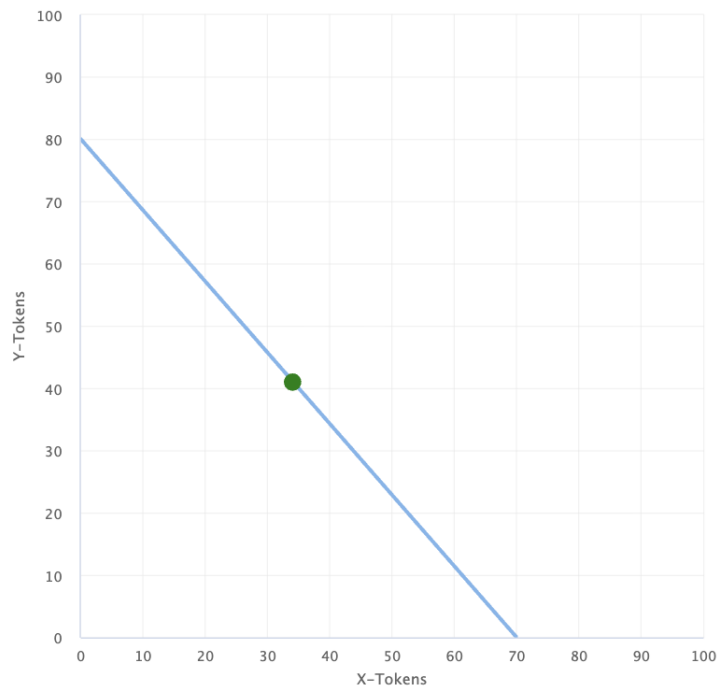
Part 1 : Task 1 of 30



X-Tokens: ? Y-Tokens: ?

Earnings Points: Your X-Tokens **or** your Y-Tokens, whichever is **smaller**

Part 1 : Task 1 of 30

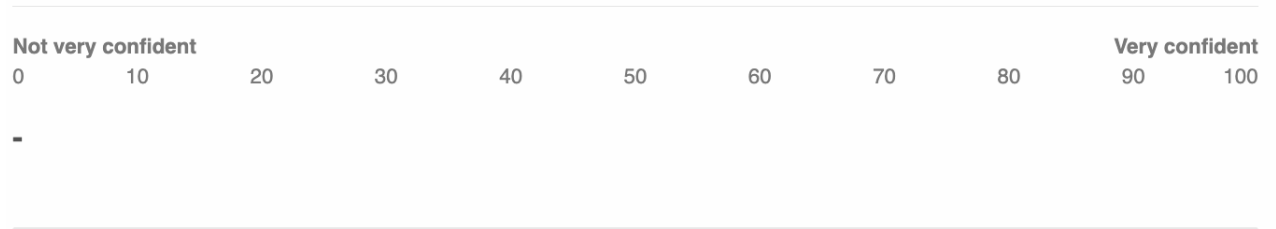


X-Tokens: 34.00 Y-Tokens: 41.10

Earnings Points: Your X-Tokens or your Y-Tokens, whichever is smaller

Submit Your Choice

On average, for Part 1 (the 30 tasks you've just completed under the current Earnings Formula), how confident are you (in percentage terms between 0% and 100%) that your token choice was no more than 5 X-Tokens, and no more than 5 Y-Tokens, away from the best possible choice given your options? Click and drag on the slider to make your choice.



On average, for Part 1 (the 30 tasks you've just completed under the current Earnings Formula), how confident are you (in percentage terms between 0% and 100%) that your token choice was no more than 5 X-Tokens, and no more than 5 Y-Tokens, away from the best possible choice given your options? Click and drag on the slider to make your choice.

Not very confident

0102030405060708090100

Very confident

-

50



Part 2: The **Multiply** Earnings Formula

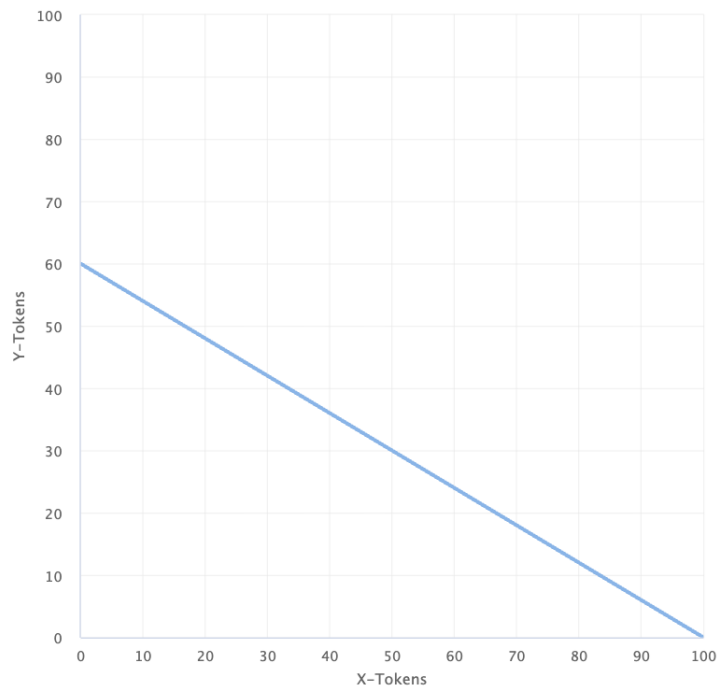
- In each of the next 30 Tasks, the Earnings Formula will be:

Earnings Points = Your X-tokens times your Y-Tokens.

- This means the computer will multiply your X-Tokens times your Y-Tokens to determine your Earnings Points.
- In other words, your Earnings Points will be: X-Tokens times Y-Tokens.
- If one of these Tasks is chosen for payment, we will first take the square root of your Earnings Points in that Task. You will get Additional Earnings equal to 25 cents for each of the resulting points. This means, the more Earnings Points you get, the higher your Additional Earnings will be.



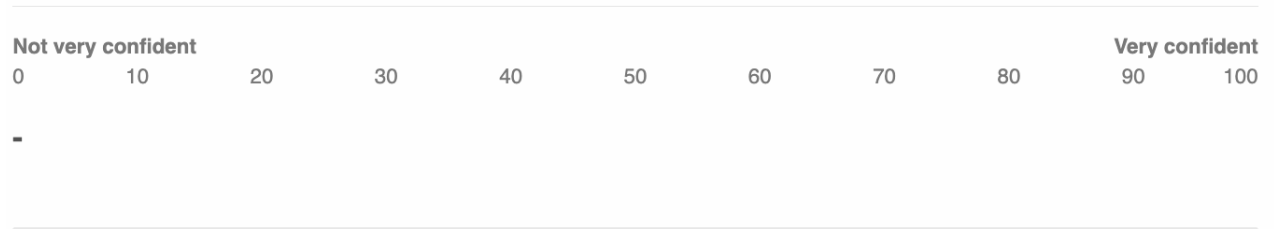
Part 2 : Task 1 of 30



X-Tokens: ? Y-Tokens: ?

Earnings Points: Your X-Tokens **times** your Y-Tokens.

On average, for Part 2 (the 30 tasks you've just completed under the current Earnings Formula), how confident are you (in percentage terms between 0% and 100%) that your token choice was no more than 5 X-Tokens, and no more than 5 Y-Tokens, away from the best possible choice given your options? Click and drag on the slider to make your choice.



Part 3: The **Average** Earnings Formula

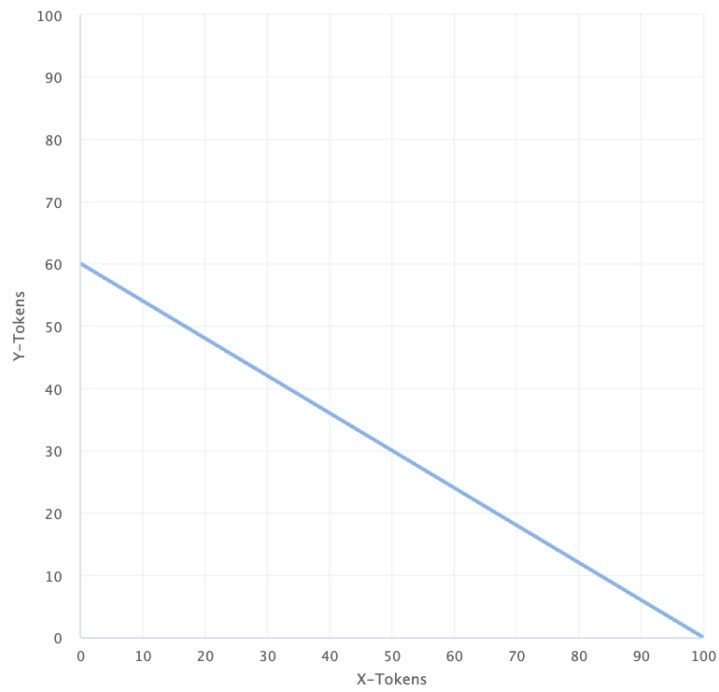
- In each of the next 30 Tasks, the Earnings Formula will be:

Earnings Points = The **average of your X-Tokens and your Y-Tokens.**

- This means the computer will add up your chosen X-Tokens and Y-Tokens and divide this sum in half to determine your Earnings Points.
- In other words, your Earnings Points will be: $(X\text{-Tokens} + Y\text{-Tokens})/2$
- If one of these Tasks is chosen for payment, you will get Additional Earnings equal to 25 cents for each Earnings Point you get in that Task.



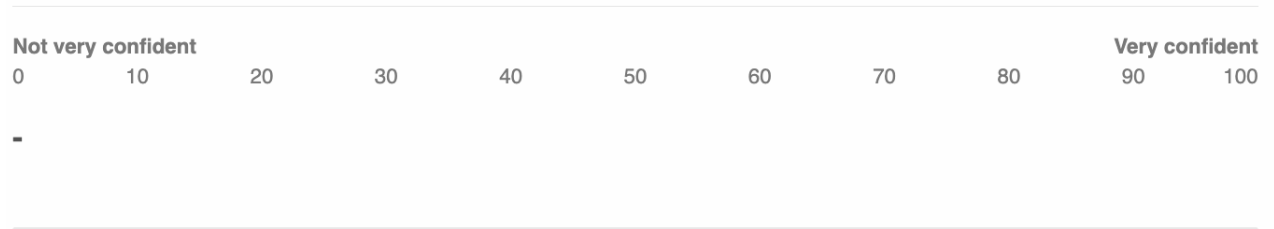
Part 3 : Task 1 of 30



X-Tokens: ? Y-Tokens: ?

Earnings Points: The **average** of your X-Tokens and your Y-Tokens.

On average, for Part 3 (the 30 tasks you've just completed under the current Earnings Formula), how confident are you (in percentage terms between 0% and 100%) that your token choice was no more than 5 X-Tokens, and no more than 5 Y-Tokens, away from the best possible choice given your options? Click and drag on the slider to make your choice.



In the next part of the experiment, we will ask you a short series of questions. Please answer them to the best of your ability.



If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how many days would it take them to drink one barrel of water together?

Enter the number of days it would take below.



A man buys a pig for \$60, sells it for \$70, buys it back for \$80, and sells it finally for \$90. How much has he made?

Enter the number of dollars he made below (no dollar sign).



Simon decided to invest \$8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, how many dollars does Simon have from this investment?

Enter the number of dollars Simon has below (no dollar sign).

