Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Concerning Quality: How Maternal Perceptions Before, During, and After Daycare Entry Shape Care and Work Hours

Franz Neuberger

fneuberger@dji.de

German Youth Institute https://orcid.org/0000-0003-3427-0298

Mariana Grgic

German Youth Institute https://orcid.org/0000-0002-5448-7075

Johannes Wieschke

German Youth Institute https://orcid.org/0000-0001-7375-8603

Susanne Kuger

German Youth Institute https://orcid.org/0000-0002-7874-7556

Research Article

Keywords: Early childhood education and care, Perceived quality, Maternal time allocation, Mundlak-Hybrid within/between decomposition, Germany

Posted Date: June 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9985837/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Concerning Quality: How Maternal Perceptions Before, During, and After Daycare Entry Shape Care and Work Hours

Franz Neuberger* Mariana Grgic Johannes Wieschke Susanne Kuger

German Youth Institute (DJI), Nockherstr. 2, 81541 Munich, Germany

*Corresponding author: neuberger@dji.de

ORCID Franz Neuberger: 0000-0003-3427-0298
ORCID Mariana Grgic: 0000-0002-5448-7075
ORCID Johannes Wieschke: 0000-0001-7375-8603
ORCID Susanne Kuger: 0000-0002-7874-7556

Preprint | June 9, 2026

Abstract. Does perceived daycare quality shape maternal time allocation, and if so, on which margin? Using the KiBS Long Panel 2021–2024, a representative German parent survey (N = 8,967 mothers, 21,572 person-years), we follow perceived quality along the three-stage order families face: reasons for not using daycare before entry, perceived quality at entry, and within-person changes in perceived quality among continuing users (Mundlak-Hybrid within/between decomposition). Demand-side preferences and quality concerns are the strongest barriers to daycare and employment entry. Prior quality concerns also predict lower perceived structural quality and accessibility after entry. Among continuing kindergarten users, only the child's interaction experience moves behaviour within-person, and it moves daycare hours, not work hours; the cross-sectional quality–work association is mostly selection. Unplanned centre closures shift care (long closures compress daycare hours; short closures push mothers onto grandparent care) but not work hours. We read these patterns as a trust-based care-time mechanism, with policy implications that differ between the under-3 entry margin and the post-entry intensive margin.

Keywords: Early childhood education and care • Perceived quality • Maternal time allocation • Mundlak-Hybrid within/between decomposition • Germany.

JEL Classification: J13 • J22 • J16 • I21 • D13.

1 Introduction

Germany's quantitative expansion of early childhood education and care (ECEC; in German: *Kindertagesbetreuung*, covering both centre-based care for children under three and kindergarten-age provision) since the 2013 legal entitlement for under-threes has clearly raised maternal employment (Bauernschuster and Schlotter, 2015; Müller and Wrohlich, 2020; Neuberger et al., 2022). Much of the additional employment is part-time, concentrated in narrow weekly-hour bands, and the headline participation rates have not been matched on the quality side. Staff shortages translate into reduced opening hours, partial (single-class) closures, and unplanned full-centre closures that now affect more than 40% of all families (Grgic and Friederich, 2023; Neuberger et al., 2025a,b). If parents perceive the care their child receives as inadequate or unreliable, the employment-promoting gains of the expansion can be undercut. Conversely, if perceived quality is itself a margin on which the maternal employment decision moves, then women's labour-force participation can be supported not just through more slots but through better-trusted ones. Whether that second possibility holds empirically is the question this paper takes on.

The German evidence base on the perceived-quality side is thin. The quantity link — access to daycare and maternal employment — has been mapped extensively (Bauernschuster and Schlotter, 2015; Hermes et al., 2023; Zimmert, 2019; Zoch and Hondralis, 2017). The quality–behaviour link is younger: Stahl and Schober (2020) test the quality–hours association on SOEP cross-sections, and Philipp et al. (2025) show experimentally that perceived quality shapes employment norms, but longitudinal evidence that exploits within-person change in how the same mother perceives the quality of her child's care is largely absent. Without that within-variation, cross-sectional designs cannot distinguish a causal quality effect from a selection pattern in which mothers with stronger labour-market attachment also report higher quality.

This paper uses the KiBS Long Panel 2021–2024, a representative annual parent survey conducted by the German Youth Institute (DJI), to track the role of perceived quality along the three-stage order families actually face: a non-use reason *before* entry, a perceived quality *at* entry, and a continuously measured quality perception *after* entry, alongside maternal employment and informal-care arrangements. We use a Mundlak-Hybrid within/between decomposition to isolate the within-person quality–behaviour signal from selection, and a barrier-elimination counterfactual to translate the upstream estimates into policy-relevant magnitudes. The substantive finding, previewed in the abstract above, is that the quality margin works as a trust-based *care-time* mechanism rather than a labour-supply lever: the work-side response so often presupposed in policy debate is absent in our within-person evidence.

The analytic chain has three stages. Each operates on its own sub-sample, defined by where the mother currently sits in the empirical sequence. **Stage 1** (transition models) takes mothers who are neither using daycare nor employed and asks how their non-use reasons predict daycare and employment entry one year later; a barrier-elimination counterfactual translates these estimates into policy-relevant magnitudes. **Stage 2** (quality-selection) takes mothers who are about to enter and asks whether their pre-entry reasons predict the quality they then perceive after entry. **Stage 3** (Mundlak-Hybrid within/between decomposition) takes continuing kindergarten users and asks which dimensions of their changing quality perceptions move their weekly daycare hours, work hours, and informal-care hours within-person, net of selection. The three sub-samples are deliberately disjoint, because the questions are different.

Section 2 develops the trust framework and the perceived-quality construct; Section 3 states the four hypotheses and maps them onto the stages; Section 4 introduces the data and constructs; Section 5 specifies the three-stage estimator; Section 6 reports the results stage by stage; Section 7 reads them through the trust frame; Section 7.1 discusses scope conditions; Section 8 concludes.

2 Theoretical Framework

2.1 Institutional Context: From Expansion to the Quality Question

The expansion of ECEC since the Day Care Expansion Act (2005) and the legal entitlement for under-threes (2013) has raised the under-3 participation rate from below 14% (2006) to over 36% (2023), with substantial East–West disparities persisting (BMBFSFJ, 2025a; Kayed et al., 2025). Macroeconomic studies confirm that this expansion has significantly increased the female employment rate (Bauernschuster and Schlotter, 2015; Neuberger et al., 2022).

In parallel, the focus has shifted from slot availability to quality (Dreyer et al., 2025). Federal quality investment through the KiQuTG (Good Daycare Act) has been substantial in scale, but its translation into observable improvements at the centre level has been uneven: the most recent monitoring report concludes that progress on key structural quality indicators is heterogeneous across federal states and only partially convergent (BMBFSFJ, 2025b), while demand for trained professionals continues to outpace supply (Meiner-Teubner et al., 2024). Meta-analytic evidence qualifies what structural improvements alone can achieve: across 49 studies in 14 countries covering 18,049 ECEC classrooms, Hausladen et al. (2026) show that the child–staff ratio is significantly negatively associated with process quality overall (Fisher's $z = -0.12$), but the effect is concentrated in under-3 settings ($z = -0.14$, $p < .01$) and mixed-age groups ($z = -0.32$, $p < .001$), with essentially no measurable association in kindergarten-age classrooms ($z = -0.06$, $p = .11$). The staff shortage also translates into emotional exhaustion in the pedagogical workforce (Oppermann et al., 2025) and into service restrictions up to full-day closures (Grgic and Friederich, 2023; Weimann-Sandig and Kalicki, 2024). Over 40% of families are affected by unplanned centre closures (Neuberger et al., 2025a,b), with a documented dose–response gradient: longer closures reduce well-being, mediated through deteriorating quality perceptions (Neuberger et al., 2025a). The well-being footprint of these disruptions is gendered: Huebener et al. (2024) show stronger negative effects for mothers in Germany during COVID-19 closures, and Hiekel and Kühn (2024) document a gender mental-health gap that emerged among German parents during the same period; international evidence from Japan parallels the pattern (Ito et al., 2024). Centres serving high proportions of disadvantaged children are disproportionately affected (Schieler and Menzel, 2024); children with migration backgrounds are cared for under qualitatively less favourable conditions (Stahl, 2015). If perceived quality declines or centres become unreliable, the employment-promoting effect of expansion can be undercut.

2.2 Trust as the Link Between Quality and Employment Decisions

The central connection between perceived quality and employment decisions runs, we argue, through trust. Parents entrust two things to the centre at once. The first is their child, whose daily experience they can observe only fragmentarily. The second is their own everyday family

organisation: a mother who takes up or extends employment is betting that the centre will function reliably enough for her own schedule to hold. Both bets align her time allocation with an expectation about the centre, and both can fail. Mayer et al. (1995) define trust as the willingness to be vulnerable to the actions of another party, based on perceived Ability, Benevolence, and Integrity (the ABI model). Trust in this sense is not binary but a continuum that can diverge across dimensions (Camehl et al., 2018): parents may attribute high Ability to staff while remaining uncertain on Benevolence, or trust the relationship with their child's teacher while losing trust in the institutional reliability of the centre.

The trust question is particularly salient in the ECEC context because the pedagogical day remains opaque to parents. Nentwig-Gesemann and Hurmaci (2020) speak of the *black-box problem*: quality judgements depend less on objective indicators than on perceived fit between parenting values and institutional culture, a milieu-specific mechanism that can generate distrust where habitual congruence between staff and family is lacking. This transparency deficit gives key importance to the interaction moments parents do observe — drop-off, pick-up, and the routine exchange with the pedagogical staff. Hummel et al. (2022a), based on 735 families in 169 centres, show that informal communication is the strongest trust predictor, explaining 44% of trust variance; educational background shows no association, while non-German family language and children's emotional symptoms systematically lower trust. Hummel et al. (2022b) complement this with the finding that opportunities for parents to participate in the centre's everyday life and decision-making strengthen trust, as does parent education, while structural conditions (preparation time, qualification levels) shape interaction quality at the parent–teacher interface (Hummel et al., 2022c).

The partnership ideal of professional discourse, however, masks asymmetries: Betz (2022) shows that pedagogical staff initiate interactions while parents act as recipients. Toward disadvantaged and migrant families, a deficit perspective persists (Betz, 2015; Betz and Bollig, 2023), hindering trust formation precisely where it would be most needed (Betz, 2025).

2.3 Perceived Quality and Parental Decisions

Two literatures need to be kept separate here. One measures *observed* quality through professional, externally validated instruments rated by trained observers. The German tradition has relied mostly on the structure-oriented rating scales (KES-R, ECERS-R); internationally, the interaction-focused Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) is the more widely used instrument, and meta-analytic evidence indicates that such interaction-focused measures predict children's academic development somewhat more strongly than structure-oriented scales (Ulferts et al., 2019). Work in this tradition includes Slot (2018), Schreiber (2004), and Camehl et al. (2015). The other literature measures *perceived* quality from parental reports: Camehl et al. (2018), Hummel et al. (2022a,b). The two correlate weakly (Camehl et al., 2015; Schreiber, 2004), and their relevance for parental behaviour is asymmetric. Following the Thomas theorem, situations defined as real are real in their consequences: when a mother decides whether to enrol her child, to extend his or her hours of attendance, or to take up work herself, the quantity that enters that decision is the quality *she perceives* the centre to have, not the quality that a trained observer would code on a process-quality instrument she has never seen. Two implications follow that we want to flag at the outset, because they shape how the rest of the paper is built. First, "perceived quality" here is not a global rating of the kind "how

good is your child's daycare"; it is built from a battery of survey items that ask parents about specific, observable features of their child's care — the staffing situation, the relationship with the pedagogical staff, the child's everyday experience, the reliability of opening hours, and so on. The construct is therefore a parental *assessment of specific aspects of the centre*, not a holistic impression. Second, the integrative model of Anders and Oppermann (2024) provides the substantive scaffolding for this multi-aspect design by combining staff professional competencies with the quality of pedagogical processes; we use this model as the underlying theoretical map, not as a forward-reference to the empirical dimensions, which are introduced only in Section 4.

Parents act on quality *perceptions*: Camehl et al. (2018) document considerable information asymmetries between observed and perceived quality. Philipp et al. (2025) provide the closest experimental evidence to the present design: in a factorial survey experiment in the pairfam panel, high (vs. medium) hypothetical daycare quality leads respondents to normatively recommend greater daycare use and roughly one additional hour per week of maternal working time — a normative-response finding whose contrast with revealed-behaviour evidence in panel data we revisit in the Discussion.

Non-use reasons serve as behavioural markers of underlying barriers: Jessen et al. (2020) distinguish supply-side from demand-side reasons; Burghardt (2018) shows that parental attitudes explain more variance in under-3 take-up than sociodemographic variables; Geier and Riedel (2009) demonstrates independent explanatory power of parental attitudes beyond SES. The persistent weight of demand-side preferences in German parental decision-making is consistent with a familialist legacy in continental welfare-state arrangements (Deindl, 2024): cultural scripts of exclusive early maternal care have liberalised only gradually since the 1970s and remain influential alongside formal expansion (Berth et al., 2025), and personal attitudes diverge from perceived public opinion in patterned ways across East and West Germany (Beringer et al., 2025).

Theoretically, two mechanisms can be distinguished and tested separately on the data: (1) A **trust mechanism** operates through the ABI dimensions: trust is built primarily through everyday communication (Hummel et al., 2022a) and strengthened by opportunities for parents to participate in the centre (Hummel et al., 2022b); on this view, the empirical signal should be strongest on the quality dimensions parents themselves directly observe in their interaction with their child and the staff. (2) A **reliability mechanism**: structural quality and accessibility signal planability, and when reliability declines parents must make compensatory arrangements; on this view, signals should travel via centre-level structural conditions rather than via parent–staff interaction. Whether the two mechanisms reach different decision margins (entry, care hours, work hours, informal-care substitution) is the empirical question Section 6.4 settles.

2.4 Formal and Informal Care

Whether daycare and grandparent care operate as substitutes or complements is empirically ambiguous. Gambaro et al. (2024) find little support for a pure crowding-out hypothesis; informal and formal care are increasingly combined. Igel and Szydlik (2011) show that public daycare investment "crowds in" grandparental willingness but "crowds out" intensity; family and state complement each other. The substitution possibility is socially unequally distributed: families with migration backgrounds have less access to informal arrangements (Gambaro et al., 2024); Steinberg (2025) documents that resource-rich families are better placed to compensate when

institutional care fails, though they are not the only ones who do.

2.5 Extensive versus Intensive Margin

Labor economics distinguishes between the extensive margin (whether to work) and the intensive margin (how many hours), a separation that is central to the analysis of quality effects because both decisions follow different logics.

The extensive margin (entering employment) is a discrete decision with high thresholds: it requires a job, a care solution, and the overcoming of institutional barriers (Morrissey, 2017). Empirically, Jessen et al. (2020) show that the basic care-usage decision was, in earlier waves of the expansion period, driven primarily by structural factors, with quality aspects playing a subordinate role. Whether that ranking still holds in the more recent, perceived-quality-crisis period is, as we argue in the Research Questions below, an open empirical question we explicitly address.

The intensive margin is a continuous adjustment: quality improvements may enable more daycare hours and, in parallel, potentially more work hours. Gambaro et al. (2016) document an increase of 2.5 hours/week through afternoon care. The closest evidence on causal labour-supply effects of daycare expansion comes from interventions that operate on the *availability* side rather than on perceived quality: Hermes et al. (2023, 2024) use a randomised information intervention that reduces application barriers and document a sizeable employment effect, and Humphries et al. (2024) show that full-day-care availability raises maternal earnings by over 20%, persistent over six years. These designs identify the consequence of expanding access; whether *perceived quality* of the care that mothers actually use translates similarly into hours and earnings is a separate question that requires within-person variation in perceived quality. Cascio (2013) show that expanding high-quality daycare increases employment only among low-SES families; Schober and Stahl (2016) document that full-day care is positively associated with life satisfaction in East Germany but not in West Germany.

3 Research Questions and Hypotheses

From the theoretical considerations, the overarching research question emerges: Does perceived daycare quality influence employment and care decisions, and if so, through which mechanisms and on which decision margin?

We approach this question with four exploratory hypotheses. They are deliberately broad: we expect perceived quality and reliability to enter at multiple decision margins (entry, daycare hours, work hours, informal care), but make no commitment in advance to which margin or which dimension dominates. The decomposition of within- versus between-person variation is treated as an open test rather than as a directional prediction, in keeping with the post-hoc role of Mundlak-Hybrid in disentangling effect from selection.

H1 – Quality concerns and entry. Quality-related non-use reasons before entry lower the probability that a non-user enters daycare in the next period and, indirectly, the probability of subsequent employment entry. The institutional context has changed since Jessen et al. (2020): against the backdrop of staff shortages, partial and full closures, and the broader perceived-quality crisis (Neuberger et al., 2025a,b), Neuberger et al. (2026) document a sharp rise in

non-user quality concerns over the 2021–2024 window. Whether quality remains a subordinate driver of usage decisions in this newer context is therefore an empirical question, not a settled one, and our analysis tests it directly by disaggregating non-use reasons into ten categories of which quality concerns is one. We do not commit ex ante to whether the entry barrier is stronger for children under three or for kindergarten-age children.

H2 – Perceived quality among users. Among mothers who already use daycare, perceived quality affects the intensive margin on both the care side (weekly daycare hours, grandparent-care hours) and the work side (work hours and employment entry conditional on having a slot). The direction is expected to be positive on the care side; whether quality moves work hours, and through which dimensions, is left open.

H3 – Reliability shocks (closures). Unplanned centre closures act as reliability shocks. We expect them to register at *some* margin – weekly daycare hours, informal care substitution, or work hours – but make no a-priori prediction about which margin absorbs short versus long closures.

H4 – Effect versus selection (Mundlak-Hybrid). Cross-sectional quality–behaviour associations may reflect causal within-person responses, between-person selection, or both. The Mundlak-Hybrid decomposition is used to test, with an *open expected outcome*, whether within-effects survive controlling for between-person heterogeneity. We do not assume a priori that selection dominates the work-hours association; we test it.

Mapping onto the analytic chain. H1 is tested in Stage 1 (transition models on the non-user / non-employed risk set) and translated into policy-relevant magnitudes by the counterfactual derived from it. The Stage 2 quality-selection model is the empirical bridge between H1 and the post-entry hypotheses: it tests whether the same concerns that block entry also predict the quality mothers then perceive once they enter. H2 and H4 are jointly tested in Stage 3 Panel a (Mundlak-Hybrid on continuing kindergarten users, all five quality dimensions). H3 is tested in Stage 3 Panel b (the closure-window specification on the same population). The four hypotheses therefore correspond to the four cuts of the empirical chain (entry, post-entry quality, within-person quality response, reliability-shock response) without redundancy.

4 Data

The analyses are based on the KiBS Long Panel 2021–2024, an annual representative parent survey (CAWI) conducted by the German Youth Institute (DJI) with samples drawn from population registries. The core analytic panel consists of mothers with children in a daycare centre and at least two panel waves; Stage 1 additionally draws on the upstream pool of non-user person-years to build the transition risk set (see Methods). Starting from 26,585 observations of 10,802 mothers, sequential exclusions (school-age children at ≥ 7 years, age inconsistencies, winsorisation with hours > 60/week capped at 60, parental-leave disambiguation, births during the panel) reduce the core panel to 21,572 observations of 8,967 mothers. The KiBS sample is selective: 82% of mothers hold the *Abitur*, the German university-entrance qualification at the upper end of the ISCED 3 / ISCED 3A range; 48% report high household income (top tercile); and the employment rate is 93% (though a substantial share of that employment is part-time). This

selectivity constrains policy-level generalisation and is revisited in the limitations (Section 7.1). Figure 1 summarises the sample-construction cascade and the analytic chain.

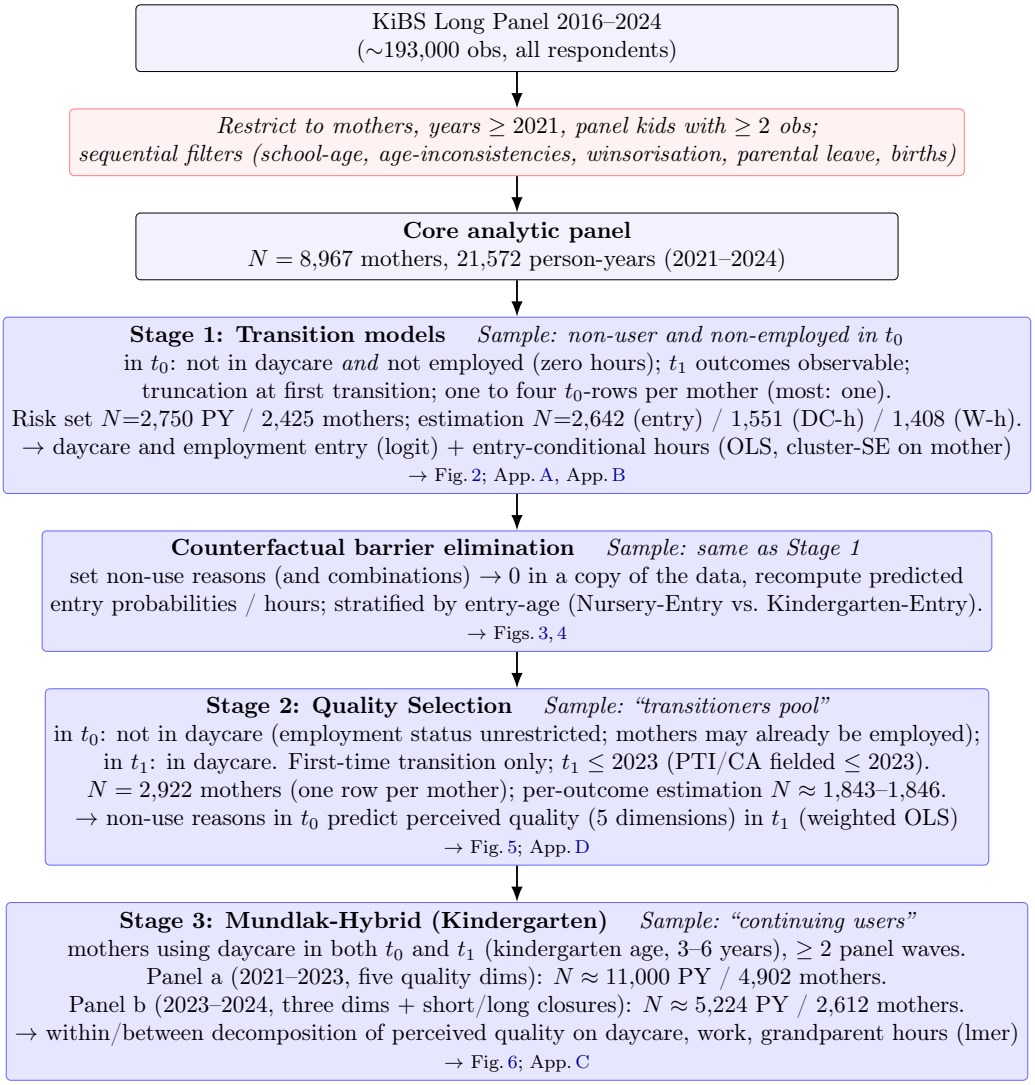


Figure 1: Sample construction and analytic chain. Blue boxes: analytic samples and stages; red box: sequential exclusion step. The core analytic panel (N = 8,967 mothers, 21,572 person-years) is built from 26,585 observations on 10,802 mothers via sequential restrictions (school-age exclusion at ≥ 7, age-inconsistency removal, hours-winsorisation at 60/week, parental-leave disambiguation, births during the panel). **Each stage operates on its own sub-sample of the core panel**, and the sub-samples are deliberately different because the questions are different: Stage 1 conditions on the non-user and non-employed state in t_0 (neither daycare nor employment, observed at risk of both transitions); the counterfactual is computed on the same sample by re-prediction from the Stage 1 models; Stage 2 conditions on a non-user-in-t_0 to user-in-t_1 transition (employment status in t_0 unrestricted, because the question is about the perceived quality after entry, not a labour-market transition); Stage 3 conditions on continuing kindergarten users, because within-person decomposition requires repeated quality measurements among already-using mothers. Panel a (2021–2023) covers all five quality dimensions; Panel b (2023–2024) adds short and long closure indicators and is restricted to the three quality dimensions that retain genuine 2023–2024 within-variation (PTI and CA were fielded only through 2023).

4.1 Outcome Variables

Six outcomes capture the extensive and intensive margins of maternal employment and of formal and informal care:

Employment Binary: currently employed (1) or not (0).

Work hours Weekly employment hours (metric, 0–60, winsorised at 60), conditional on employment.

Daycare use Binary: child currently in daycare (1) or not (0).

Daycare hours Weekly hours the child actually attended the daycare centre (parent-reported, metric, 0–60; conditional on use). This is *used* hours, not the contracted slot length: contracted slots are typically discrete bands (e.g. 25/35/45 h) and were not collected as a metric variable before 2024; the used-hours measure is the substantively closer quantity to the behavioural question of how much exposure the mother allocates to the centre.

Grandparent care Binary: child receives any grandparent care (1) or not (0).

Grandparent hours Weekly care hours from grandparents (metric, 0–80; conditional on any).

Throughout the paper, *daycare* refers to centre-based ECEC. Home-based childminding is a distinct and, in this age range, much smaller form of formal care; "daycare use" and "non-use" here concern centre-based attendance specifically, and we use the term consistently in this sense.

4.2 Stage 1 Predictors: Non-Use Reasons

Mothers who do *not* use daycare are asked to indicate, in a fixed list of ten items, why they currently do not use a centre-based care slot. We treat these as the empirically upstream piece of the quality story: the first decision is whether to enter at all, and quality-related considerations are one dimension of that decision alongside demand-side preferences and supply-side constraints. Five reasons are supply-side (**quality concerns**, **no slot available**, **no offer nearby**, **cost**, and **hours don't fit**) and five are demand-side (**child too young**, **want to raise myself**, **good experience at home**, **grandparents available**, and **not an option** for me). Each reason is coded 1 if cited, 0 otherwise. Reasons are measured in each mother's last year of non-use (t_0) and treated as time-constant predictors in the transition and quality-selection models.

4.3 Stage 2–3 Predictors: Perceived Quality Dimensions

Among the mothers who *do* use daycare, KiBS fields five indices of perceived quality. The items, scales, and reliability statistics follow the construction in Neuberger et al. (2026) on the parent KiBS cross-section; we use the indices as defined there, but estimate them within the Long Panel and on a narrower age range. All five are indices of parent-perceived centre attributes; none is an observer-rated pedagogical-quality measure (see Limitations). They map onto the integrative quality model of Anders and Oppermann (2024) at differing proximities to the child–staff relationship: CIE and PTI capture the child's and the parent's direct interaction experience, CA the parent's inference about staff competence, and structural quality and accessibility the institutional reliability framing in which those interactions take place.

Structural quality Satisfaction with group size, number of teachers, reliability, and stability of pedagogical staff (4 items, 6-point scale 0–5 rescaled so that higher = better; available 2021–2024; $\alpha = 0.77$). A reliability / contract-keeping signal in the ABI framework.

Accessibility Satisfaction with opening hours and temporal flexibility of the centre (2 items, 6-point scale; available 2021–2024; $\alpha = 0.53$). The item on cost satisfaction was excluded because fee exemptions and fee-free federal states distort it systematically. Note: the low α is a consequence of the 2-item structure; results involving this dimension should be interpreted with corresponding caution.

Child interaction experience (CIE) Parent-reported four-item index of *the child's experience at the centre*: "my child feels well in daycare", "my child has a trusting relationship with at least one teacher", "the teachers have enough time for my child", and "my child has playmates" (4-point scale; available 2021–2024 with genuine within-person variation in each year-to-year transition; $\alpha = 0.74$). A fifth item on participation rights was dropped for low item-total correlation. This is the dimension most directly anchored in what parents themselves can observe about their child's experience, and it captures both *Benevolence* ("has enough time for my child") and the felt quality of the teacher–child relationship (trusting relationship, well-being) in the ABI trust model. It is not a drop-off/pick-up rating but a general parental assessment of the child's integration and felt security in the centre.

Parent–teacher interaction (PTI) Quality of communication with the pedagogical staff: "good relationship with the teachers", "cooperation is difficult" (reverse-coded), "approachable when problems arise", "open to parental concerns" (4 items, 4-point scale; fielded 2021–2023 only; $\alpha = 0.87$). The communicative foundation through which parental trust is built (Hummel et al., 2022a).

Competence attribution (CA) Perceived competence of the pedagogical staff: "correctly judge my child's developmental stage", "skilled in handling my child", "know what my child needs" (3 items, 4-point scale; fielded 2021–2023 only; $\alpha = 0.85$). A second-order parental inference about staff *Ability* rather than a first-order observation, and therefore at one remove from the child's experience captured by CIE.

All Cronbach's α values are computed on the analysis subsample (mothers, daycare users, 2021–2023), so that the reliability estimate matches the data the models are estimated on rather than the full prep sample.

The five KiBS dimensions map onto the ABI trust model (Mayer et al., 1995) at differing proximities to the child–teacher relationship: CIE is the most proximate (a parent assessment of whether the child feels well, the candidate *Benevolence* signal); CA is the same dimensions at one remove (a parental inference about staff *Ability*); PTI represents the communicative *Integrity* signal at the parent–teacher interface; and structural quality and accessibility operate as the *Integrity*-as-reliability channel (the contract-like expectation that the centre delivers what it promises). All five dimensions contribute genuine within-person variation to Panel a of the Mundlak-Hybrid model (2021–2023, Section 5.4). For the 2023–2024 closure specification (Panel b), PTI and CA are excluded because the items were not fielded in 2024 and therefore carry no 2023–2024 within-variation; structural quality, accessibility, and CIE retain genuine 2023–2024 within-variation and enter Panel b. Table 2 displays the year-by-year coverage.

4.4 Controls

Time-varying controls include child age (in months), parental-leave dummies (employed while on parental leave; not employed while on parental leave), and year fixed effects. The regional

care gap at the federal-state level captures local supply constraints: it is the difference between parental demand and the current under-three or kindergarten-age participation rate, computed for the age band that matches the child (Kindertagesbetreuung Kompakt 2025; BMBFSFJ, 2025a).

Time-constant level-2 controls comprise education (Abitur dummy as defined above), household income (terciles), East/West Germany, region type (urban/suburban/rural, following the settlement-structural region typology of the German Federal Institute for Research on Building, Urban Affairs and Spatial Development; BBSR),[1] number of children, migration background (defined as the child and/or at least one parent born outside Germany), social risk (welfare receipt), and single-parent status. Single-parent cases are a small share of every analytic sample, accounting for about 1.9% in the Stage 1 risk set and roughly 4% in the broader long panel. The dummy nonetheless enters all four Stage 1 outcomes and the Stage 2–3 specifications without convergence problems or rank deficiency. Estimates are therefore conditional on partnership status. We do not interpret the single-parent coefficient substantively given the small subgroup, but including it blocks an obvious omitted-variable channel and removes a potential source of confounding for the remaining controls. No panel weights are used in the main specifications. Quality-selection models additionally report survey-weighted estimates as a sensitivity check.

Tables 1 and 2 provide descriptive statistics by subsample and survey year, respectively. Notable trends across years include rising employment rates (78.6% in 2021 to 91.2% in 2024 among users), declining structural quality ratings, and increasing income selectivity of the panel (42.7% to 61.6% high income on the user side), the latter likely reflecting differential panel attrition.

[1] BBSR settlement-structural region types (*Siedlungsstrukturelle Regionstypen*): https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/regionen/siedlungsstrukturelle-regionstypen/regionstypen.html.

Table 1: Descriptive statistics by subsample

	Under 3 (ages 0–2)		Age 3+ (ages 3–6)	
Variable	Users	Non-users	Users	Non-users
Outcomes				
Work hours/week	24.5 (13.3)	4.6 (10.1)	24.5 (13.3)	11.4 (14.4)
Employed (%)	84.5	36.5	85.6	49.7
Daycare hours/week	33.1 (8.1)	—	33.7 (7.0)	—
Grandparent hours/week	2.4 (4.6)	2.4 (5.7)	2.5 (4.8)	2.5 (5.6)
Uses grandparent care (%)	37.8	30.4	38.3	26.9
Non-use reasons (non-users only)				
Quality concerns (%)	—	12.2	—	20.3
Wants to raise child herself (%)	—	61.1	—	58.5
Good home-care experience (%)	—	63.2	—	62.7
Wanted place, got none (%)	—	16.4	—	24.2
No offer nearby (%)	—	13.8	—	22.4
Cost barrier (%)	—	15.9	—	15.2
Hours don't fit (%)	—	5.4	—	8.6
Child too young (%)	—	87.7	—	24.7
Grandparents available (%)	—	34.3	—	36.7
Not an option (%)	—	28.5	—	35.1
Quality				
Structural quality	0.2 (0.9)	—	-0.1 (1.0)	—
Accessibility	0.0 (0.9)	—	0.0 (1.0)	—
Child interaction exp. (CIE)	0.1 (0.9)	—	0.0 (0.9)	—
Parent-teacher interaction (PTI)	0.1 (0.9)	—	-0.0 (1.0)	—
Competence attribution (CA)	0.2 (0.9)	—	0.0 (1.0)	—
Closures				
Closure ≤1 week (%)	30.7	—	29.0	—
Closure >1 week (%)	12.9	—	10.2	—
Controls				
Child age (months)	26.3 (6.7)	13.8 (7.8)	56.4 (12.1)	54.2 (15.4)
Abitur (%)	85.9	73.9	82.1	57.2
Income (high tercile, %)	51.9	29.7	49.8	18.7
East Germany (%)	28.6	16.1	26.9	15.5
Number of children	1.7 (0.7)	1.7 (0.8)	2.0 (0.7)	2.3 (1.2)
Care gap (PP, federal state)	13.2 (5.3)	14.7 (4.6)	6.0 (1.7)	6.3 (1.6)
Migration background (%)	3.1	8.3	3.6	21.0
Social risk (%)	4.7	8.5	4.1	19.5
Single parent (%)	2.9	3.0	4.5	8.2
Employed + parental leave (%)	16.2	26.8	7.3	6.6
Not employed + par. leave (%)	9.2	57.3	7.8	23.7
N (person-years)	4,855	13,889	16,717	1,534
N (mothers)	3,270	12,424	7,936	1,438
Use rate (%)	25.9		91.6	

Note: Cells report mean (SD) for continuous variables and percent for binary indicators. The table crosses **age group** (under 3 / age 3+) with **daycare-use status** (Users / Non-users). Quality dimensions and closure items are collected only from daycare users and therefore appear only in the User columns; non-use reasons are collected only from non-users and appear only in the Non-user columns. The *Use rate* row at the bottom is the share of each age-specific sample currently using daycare, computed as $N_{\text{users}}/(N_{\text{users}} + N_{\text{non-users}})$. Closure items were only fielded in 2023/2024. KiBS Long Panel 2021–2024.

Table 2: Descriptive statistics by survey year: users vs. non-users

	2021		2022		2023		2024	
Variable	User	Non-user	User	Non-user	User	Non-user	User	Non-user
Outcomes								
Work hours/week	22.6 (14.6)	6.7 (11.9)	24.4 (13.5)	5.1 (10.8)	24.9 (12.8)	4.9 (10.4)	26.1 (11.7)	4.9 (10.4)
Employed (%)	78.6	42.9	84.9	38.0	87.3	36.0	91.2	36.1
Daycare hours/week	33.4 (7.8)	—	33.5 (7.4)	—	33.4 (7.3)	—	34.1 (6.4)	—
Grandparent hours/week	2.7 (5.0)	1.9 (4.8)	2.5 (5.0)	2.5 (5.7)	2.3 (4.8)	2.6 (6.0)	2.2 (4.2)	2.6 (6.0)
Uses grandparent care (%)	43.6	24.7	37.6	31.0	36.1	31.1	35.6	31.4
Non-use reasons (non-users only)								
Quality concerns (%)	—	9.7	—	10.5	—	12.8	—	17.5
Wants to raise child herself (%)	—	64.2	—	60.2	—	60.2	—	60.0
Good home-care experience (%)	—	64.5	—	61.2	—	64.1	—	63.4
Wanted place, got none (%)	—	16.7	—	16.3	—	19.1	—	16.4
No offer nearby (%)	—	14.8	—	14.4	—	15.9	—	13.5
Cost barrier (%)	—	12.7	—	17.4	—	15.1	—	16.9
Hours don't fit (%)	—	5.4	—	6.2	—	5.7	—	5.4
Child too young (%)	—	74.4	—	84.6	—	81.3	—	85.0
Grandparents available (%)	—	42.5	—	30.5	—	34.6	—	33.3
Not an option (%)	—	33.3	—	28.1	—	28.2	—	28.4
Quality								
Structural quality	0.1 (0.9)	—	-0.0 (1.0)	—	-0.0 (1.0)	—	-0.1 (1.0)	—
Accessibility	0.1 (0.9)	—	0.0 (1.0)	—	0.1 (1.0)	—	-0.0 (1.0)	—
Child interaction exp. (CIE)	0.0 (0.9)	—	0.1 (0.9)	—	0.1 (0.9)	—	0.0 (1.0)	—
Parent-teacher interaction (PTI)	-0.1 (1.0)	—	-0.0 (1.0)	—	0.1 (0.9)	—	—	—
Competence attribution (CA)	0.0 (1.0)	—	0.0 (1.0)	—	0.1 (1.0)	—	—	—
Closures								
Closure ≤1 week (%)	—	—	—	—	30.4	—	27.9	—
Closure >1 week (%)	—	—	—	—	10.8	—	10.6	—
Controls								
Child age (months)	45.4 (14.6)	23.3 (17.7)	49.8 (16.7)	16.7 (14.3)	48.9 (17.8)	17.5 (14.5)	55.5 (16.1)	15.8 (13.3)
Abitur (%)	83.1	69.4	82.2	71.3	82.8	72.6	84.3	74.5
Income (high tercile, %)	42.7	20.0	46.3	25.6	52.9	28.6	61.6	37.4
East Germany (%)	30.3	15.3	27.6	16.3	25.8	15.3	25.3	17.0
Number of children	2.0 (0.7)	1.8 (1.0)	1.9 (0.7)	1.7 (0.9)	1.9 (0.7)	1.7 (0.9)	1.9 (0.7)	1.7 (0.9)
Kindergarten (vs. nursery, %)	72.8	18.0	79.4	8.8	73.9	9.0	85.4	6.8
Care gap (Bundesland, PP)	7.9 (4.5)	13.2 (5.2)	7.5 (4.1)	14.0 (5.0)	7.9 (4.4)	14.1 (5.0)	7.0 (3.5)	14.0 (4.9)
Migration background (%)	3.5	7.9	3.6	9.7	3.5	10.8	3.3	9.3
Social risk (%)	5.2	12.0	4.3	9.7	3.6	8.9	3.8	8.4
Single parent (%)	3.7	3.6	4.4	2.9	4.1	3.9	4.5	3.7
Employed + parental leave (%)	11.4	24.9	9.6	25.7	9.5	24.1	6.1	24.4
Not employed + par. leave (%)	12.7	48.0	8.7	54.0	7.1	55.3	3.3	56.5
N (obs.)	4,780	2,799	6,639	4,334	6,084	3,882	4,069	4,408

Note: Each year column is split into **User** (current daycare users) and **Non-user** (current non-users). Cells report mean (SD) for continuous variables and percent for binary indicators. Cells marked — indicate the item is not applicable for that subsample: daycare hours, the five quality dimensions, and closure items are fielded only for users; non-use reasons are fielded only for non-users. PTI (parent–teacher interaction) and CA (competence attribution) were fielded only 2021–2023 and are therefore dashed in 2024. Grandparent hours for 2021 are imputed from each respondent's neighbouring-year observations because the grandparent-care-hours item was not fielded in 2021: when both 2020 and 2022 values are available, the arithmetic mean is used; otherwise the single available neighbour-year value is used. This recovers 100.0% of 2021 User cases and 100.0% on the Non-user side. N (obs.) is person-years per wave; unique mothers across all four waves: 8,967 (users), 13,716 (non-users). KiBS Long Panel 2021–2024.

5 Methods

The empirical question — does perceived daycare quality shape maternal time allocation, and if so, on which margin — requires the analyst to follow three sequential decisions a mother actually makes, and to place each decision in front of the model that can answer it. We therefore split the core analytic panel into three disjoint sub-samples and run three corresponding models. **Stage 1** (transition models) operates on mother-years observed in the non-user and non-employed state and tests whether the ten reported non-use reasons predict daycare and employment entry one year later. A **counterfactual barrier-elimination simulation** is derived directly from the Stage 1 models and translates the estimates into policy-relevant magnitudes (predicted entry probabilities and hours under hypothetical absence of each reason and combinations thereof). **Stage 2** (quality-selection) operates on the subsample of mothers who actually transitioned from non-use to use and tests whether the same pre-entry reasons predict the quality they perceive after entry. **Stage 3** (Mundlak-Hybrid within/between decomposition) operates on continuing kindergarten users and decomposes the quality–behaviour association into a within-person (effect-of-change) and a between-person (selection) component, on two complementary windows. Figure 1 visualises the chain and the sub-sample relations; this section specifies each stage in detail.

5.1 Stage 1: Transition Models

Time indices. Throughout the paper, t_0 denotes a *baseline* survey wave at which a mother is observed in a given state (e.g., non-using, non-employed), and $t_1 = t_0+1$ denotes the immediately following wave at which we measure her outcomes (entry, hours, perceived quality). Indices t_0 on the right-hand side of a model and t_1 on the left-hand side therefore mark the temporal direction of the prediction (one-year lead).

Risk set and 4-outcome design. We model how non-use reasons in t_0 predict, in t_1, (1) daycare entry, (2) weekly daycare hours conditional on entry, (3) employment entry, and (4) work hours conditional on employment entry. All four models are estimated on a single, internally consistent risk set: mother-years in which the mother is observed in t_0 as both a daycare *non-user* (no daycare attendance) and as *not employed* (zero contracted weekly hours), with t_1 outcomes observable in the data. Each mother's record is truncated at the year of her first observed transition; censored mothers (no observed transition) contribute up to four t_0-years. The joint non-user / non-employed restriction defines a clean "baseline at risk for both kinds of transition" so that the four outcome models share their identifying variation.

Non-use reasons (NU). The ten non-use reasons are reported by mothers in each year of non-use; the indicators $\mathbf{NU}_{it}$ are time-varying within mother (a mother's reasons can shift between consecutive t_0-years). They cover both demand-side reasons ("want to raise child myself", "good home experience", "grandparents available", "not an option", "child too young") and supply-side reasons ("no slot available", "quality concerns", "no offer nearby", "cost", "hours don't fit"). Each model is estimated *stepwise* (one reason at a time, with controls) and *jointly* (all ten simultaneously); the joint specification is the main report.

Extensive-margin models (logit, cluster-robust SE).

$$\operatorname{logit}\big(P(y_{i,t_1}=1)\big) = \boldsymbol{\gamma}'\mathbf{NU}_{i,t_0} + \boldsymbol{\delta}'\mathbf{Z}_{i,t_0} + \lambda_{t_0}, \quad y_{i,t_1} \in \{\,\text{daycare entry},\ \text{employment entry}\,\} \tag{1}$$

estimated as binomial logits with standard errors clustered on the mother.

Intensive-margin models (OLS, cluster-robust SE). On the entry-conditional subset (rows with $y_{i,t_1}^{\text{entry}} = 1$ and y_{i,t_1}^{hours} observed):

$$y_{i,t_1}^{\text{hours}} = \boldsymbol{\gamma}'\mathbf{NU}_{i,t_0} + \boldsymbol{\delta}'\mathbf{Z}_{i,t_0} + \lambda_{t_0} + \varepsilon_{i,t_1}, \quad y^{\text{hours}} \in \{\text{ daycare hours/week, work hours/week }\} \quad (2)$$

fitted with OLS and the same cluster-robust SE as the logits. The cluster-robust sandwich SE on the mother captures the within-mother correlation that arises for mothers with multiple consecutive non-using and non-employed t_0-years.

Controls. $\mathbf{Z}_{i,t_0}$ comprises child age in months, education (Abitur dummy), household income (terciles), East/West Germany, region type, number of children, migration and social-risk indicators, single-parent status, parental-leave status (employed-while-on-leave / not-employed-while-on-leave), and the regional care gap at federal-state level. λ_{t_0} is a t_0-year fixed effect.

Samples. The shared risk set comprises N = 2,750 person-years on 2,425 mothers; after complete-case filtering on controls the four estimation samples are:

- Daycare entry (logit): N = 2,642 person-years.
- Employment entry (logit): N = 2,641 person-years (one row drops on missing t_1 employment status).
- Daycare hours (OLS): N = 1,551 entry-year rows (the 58.7% of risk-set rows where the mother entered daycare in t_1).
- Work hours (OLS): N = 1,408 entry-year rows (the 53.4% of risk-set rows where the mother entered employment in t_1).

Stratified estimates use the entry-age cutoff: *Under-3 Entry* (= Nursery-Entry in figure labels) covers children aged 0–1 in t_0 and entering at age ≤ 2; *Age-3+ Entry* (= Kindergarten-Entry in figure labels) covers children aged ≥ 2 in t_0 and entering at age ≥ 3. The two subsample names refer to the *age at entry*, not to the institutional type, and are used throughout to keep the figure labels readable.

5.2 Counterfactual Simulations (from Stage 1 models)

Counterfactual simulations set individual non-use reasons (and combinations thereof) to zero in a copy of the data and recompute predicted entry probabilities and hours from the joint Stage 1 models. Subgroup-specific counterfactuals are obtained by re-fitting the joint models on the Under-3 Entry and Age-3+ Entry subsamples separately, so that predicted effects reflect age-specific heterogeneity. The counterfactuals are partial-equilibrium (they do not model how the removal of a barrier would itself be achieved) and therefore represent an upper bound on expected effects.

5.3 Stage 2: Quality Selection

Do non-use reasons in t_0 predict the perceived quality of the chosen daycare centre in t_1? Weighted OLS for each of five quality dimensions (z-standardised):

$$\text{Quality}_{j,t_1} = \boldsymbol{\gamma}'\mathbf{NU}_{i,t_0} + \boldsymbol{\delta}'\mathbf{Z}_i + \lambda_t + \varepsilon_i \quad (3)$$

Sample. First-time transitioners: mothers who are observed as non-user in t_0 and as user in t_1, taking only the first such transition per mother. *Crucially, Stage 2 does not condition on employment status in t_0.* A mother may already be employed at t_0 and only *her child* transitions into daycare; she may also be non-employed at t_0 and on a joint trajectory. Conditioning on the non-employed state (as in Stage 1) would amount to asking a different question — a labour-market transition, not a perceived-quality outcome — and would also throw away most of the transitioners. Quality outcomes are restricted to $t_1 \leq 2023$ because PTI and CA were fielded only through 2023, so a $t_1 = 2024$ observation would have no measurement for these two dimensions. The deduplicated transitioner pool comprises 2,922 mothers; after restricting to $t_1 \leq 2023$ and applying complete-case filtering on controls and on the respective quality dimension, the per-outcome estimation samples lie in the range $N = 1{,}843$–$1{,}846$ (one row per mother). Subsample-stratified estimates use the same entry-age cutoff as Stage 1: Under-3 Entry (≤ 2 years at t_1, per-outcome estimation $N \approx 1{,}302$–$1{,}304$) and Age-3+ Entry (≥ 3 years at t_1, per-outcome estimation $N \approx 541$–542). The unrestricted transitioner pool of 2,922 mothers splits into roughly 2,166 Under-3 Entry and 756 Age-3+ Entry transitioners before the $t_1 \leq 2023$ restriction is applied.

5.4 Stage 3: Mundlak-Hybrid Within/Between Decomposition

Where Stages 1 and 2 took the entry margin and the immediate post-entry quality respectively, Stage 3 takes the question one step further: once a mother is a continuing user, do *changes in her perceived quality from year to year* move her weekly daycare hours, her work hours, or her grandparent-care hours? This is a within-person question that cannot be asked of the Stage 1 / Stage 2 samples, because those are defined around the entry event. We therefore move to a third, structurally different sub-sample of the core panel.

Sample restriction. Stage 3 estimates only on *kindergarten-age users* (children aged 3–6) who are observed in daycare in both t_0 and t_1. The within-person interpretation needs repeated quality ratings for the same child in a comparable care context across consecutive waves. Two features of the under-three phase make it ill-suited for this. First, the observation window is short: most children enter during the under-three years and are then observed for only one or two waves before they reach kindergarten age, which leaves little within-person variation to identify from. Second, the move from the under-three to the kindergarten context shifts the frame of reference for several quality items even when the child stays in the same institution, and KiBS does not record the exact timing of that move — it may occur before, at, or some months after the third birthday. Restricting Stage 3 to children already in the kindergarten phase holds the care context broadly constant and keeps the within-comparison clean. The restriction is therefore substantive, not merely sample-pragmatic; a within-design for the under-three phase would require denser measurement than the panel provides and is left for future work.

Specification. On this sample, we ask: does perceived quality predict behaviour *within* the same mother, net of time-constant heterogeneity? The post-selection question is addressed as a Mundlak-Hybrid decomposition:

$$y_{it} = \sum_k \beta_{W,k} x_{kit} + \sum_k \beta_{B,k} \bar{x}_{ki} + \boldsymbol{\delta}' \mathbf{Z}_i + \lambda_t + u_i + \varepsilon_{it} \tag{4}$$

where x_{kit} = time-varying within-person predictors, $\bar{x}_{ki}$ = person-means (Mundlak terms). The coefficient $\beta_{W,k}$ is the within-estimator, i.e. the effect of intra-individual *change* net of time-constant heterogeneity (Schunck, 2013); $\beta_{B,k}$ captures between-person selection. Outcomes: daycare hours, work hours, and grandparent hours (lmer, random intercept at mother level).

We report two complementary specifications:

Panel a (main, 2021–2023, all five quality dimensions). The full quality vector enters as x_{kit}. The window ends in 2023 because PTI and CA were fielded only through 2023 and therefore contribute no within-variation to a 2023–2024 specification. *Closures are not controlled for in Panel a.* The closure items were fielded only from the 2023 wave onward, so within the 2021–2023 window closures are observed in just one of the three years; they carry no usable within-person variation across this window and cannot enter as a within-person predictor. This holds even for a simple did-it-happen indicator, and a fortiori for the short-versus-long duration distinction on which our shock analysis turns. The closure analysis is therefore carried out separately in Panel b, on the two waves (2023–2024) for which closures are observed.

Panel b (closure specification, 2023–2024, three quality dimensions with 2023–2024 coverage + closures). The window is restricted to the two closure-covered waves. x_{kit} includes the three quality dimensions with genuine 2023–2024 within-variation (structural quality, child interaction experience (CIE), and accessibility), together with two binary closure indicators: short (≤ 1 week) and long (> 1 week) unplanned closures within the past 12 months (reference category: no closure). This panel provides the cleanest within-person identification of closure shocks.

Sample. Both panels are estimated on mothers of kindergarten-age children (3–6 years) who are observed as users in at least two consecutive panel waves within the respective window.

- *Panel a (2021–2023, five quality dimensions, no closures).* Starting from 13,243 kindergarten-age user person-years on 7,110 mothers, the $\geq$ 2-wave panel filter retains 11,035 person-years on 4,902 mothers. Per-outcome estimation samples after complete-case filtering on controls: N = 11,010 for daycare hours, N = 10,996 for work hours, N = 11,010 for grandparent hours.
- *Panel b (2023–2024, three quality dimensions + closures).* Starting from 7,968 kindergarten-age user person-years on 5,356 mothers, the $\geq$ 2-wave panel filter retains 5,224 person-years on 2,612 mothers. Per-outcome estimation samples after complete-case filtering: N = 5,217 for daycare hours, N = 5,212 for work hours, N = 5,217 for grandparent hours.
- *Employment-entry sub-risk-set (E1b).* A glmer logit on mothers not yet employed at panel start, derived from Panel a, comprises 2,303 person-years on 1,007 mothers; estimation N = 2,258 on 1,003 mothers after complete-case filtering.

The Stage 3 sample is therefore disjoint from the Stage 1 risk set: Stage 1 conditions on *non-users in* t_0, Stage 3 conditions on *continuing users in both* t_0 *and* t_1. The two stages address complementary questions on complementary populations.

6 Results

6.1 Stage 1: Non-Use Reasons and Transition into Daycare and Employment

Figure 2 shows the transition models estimated separately for the overall sample and for the two entry-age subsamples (Under-3 Entry, i.e. Nursery-Entry; and Age-3+ Entry, i.e. Kindergarten-Entry).

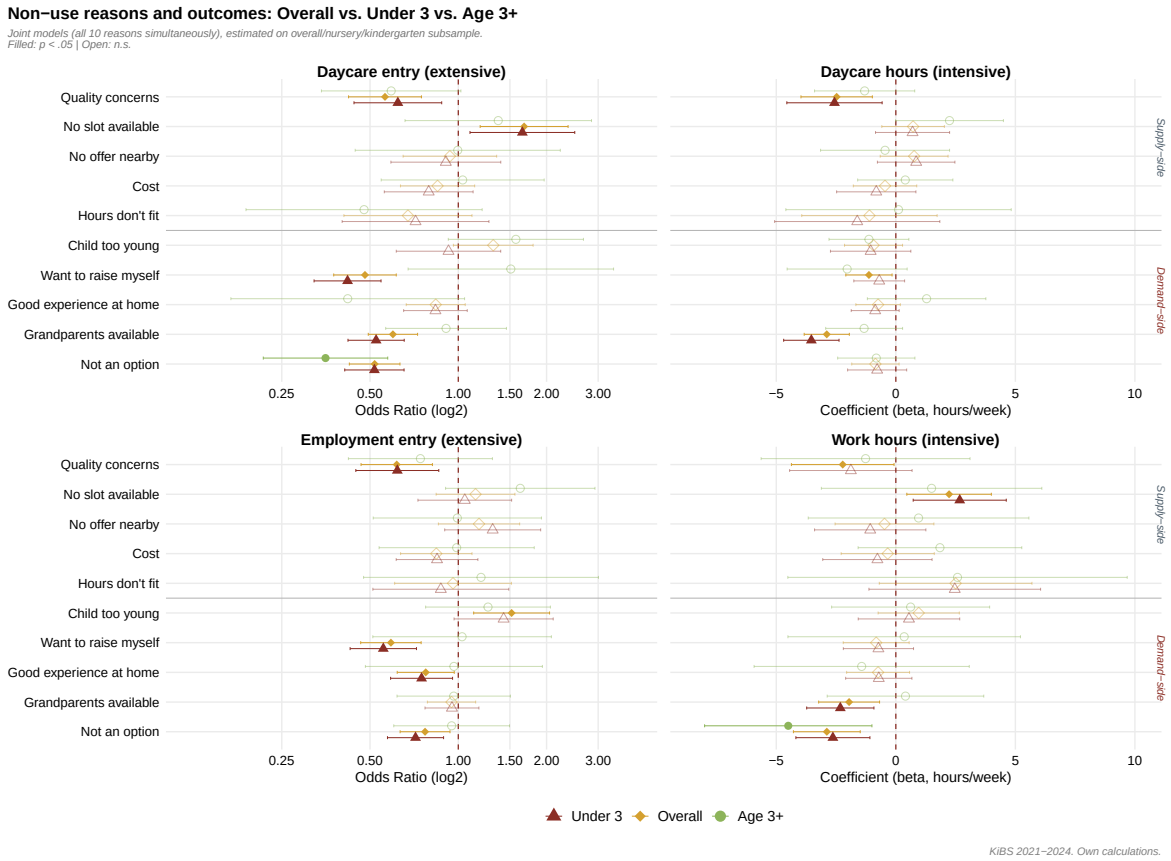


Figure 2: Non-use reasons and outcomes: overall vs. Nursery-Entry vs. Kindergarten-Entry. Joint models (all 10 reasons simultaneously) on the Stage 1 risk set. **2×2 layout.** Rows: *Daycare* (top) and *Employment* (bottom). Columns: *Extensive margin* (left, logit, OR on log2 scale) and *Intensive margin* (right, OLS, β in h/week). ▲ Nursery-Entry, ♦ Overall, • Kindergarten-Entry. Filled and full opacity: $p < .05$; open and faded (45% opacity): n.s. KiBS Long Panel 2021–2024.

Extensive margin: daycare entry (joint logit, $N = 2{,}642$ person-years). "Want to raise myself" is the strongest single demand-side barrier (OR = 0.48, $b = -0.732$, $p < .001$), followed by "not an option" (OR = 0.52, $p < .001$), quality concerns (OR = 0.56, $b = -0.577$, $p < .001$), and "grandparents available" (OR = 0.60, $p < .001$). The cost and hours-don't-fit dummies do not retain conditional significance once the demand-side reasons are partialled out. "No slot available" carries a *positive* coefficient (OR = 1.68, $p < .01$) and "child too young" a marginal positive (OR = 1.31, $p < .10$): mothers who cited these reasons in earlier years are more likely to transition once the binding constraint resolves, a catch-up pattern consistent with waitlisting and with the temporal nature of the age constraint. In stepwise specifications, effect magnitudes are larger because correlated reasons are absorbed (quality concerns alone OR = 0.40, raise myself alone OR = 0.28); the full table is in Appendix A.

Extensive margin: employment entry (joint logit, $N = 2{,}641$ person-years). The

same non-use reasons predict not just whether the mother enters daycare but whether she enters employment. "Want to raise myself" (OR = 0.59, $p < .001$) and quality concerns (OR = 0.62, $p < .001$) are the strongest demand-side barriers; "good experience at home" (OR = 0.77, $p < .05$) is a third demand-side reason that reaches conditional significance for employment (and barely for daycare). "Child too young" inverts to OR = 1.52 ($p < .01$), again a catch-up pattern. The pattern broadly parallels daycare entry, with the demand-side reasons doing the bulk of the work; quality concerns thus operate as a barrier to *both* care and labour-market re-entry, not merely to daycare itself.[2]

Intensive margin: daycare hours conditional on entry (OLS, $N = 1{,}551$ entry-year rows). Among mothers who do enter daycare, those who cited quality concerns in t_0 book -2.47 **h/week** fewer ($p < .001$) than those who did not. "Grandparents available" has the largest single demand-side effect on used hours (-2.89 h, $p < .001$), consistent with informal-care substitution. "Want to raise myself" is a smaller but significant pull toward shorter usage (-1.12 h, $p < .05$). On the supply-side, none of the access-related reasons (cost, hours don't fit, no offer nearby) survives the joint specification once demand-side reasons are controlled. The intensive-margin signal is therefore *not* dominated by supply-side hour caps but by demand-side preferences and informal-care availability, plus quality concerns.

Intensive margin: work hours conditional on employment entry (OLS, $N = 1{,}408$ entry-year rows). The work-hours pattern partly mirrors and partly inverts the daycare-hours pattern. Quality concerns remain a negative work-hours predictor (-2.22 h, $p < .05$), reinforcing the read that quality concerns continue to compress maternal time allocation even *after* the entry barrier is overcome. "Grandparents available" (-1.95 h, $p < .01$) and "not an option" (-2.89 h, $p < .001$) are similarly negative. "No slot available" *flips positive* on work hours ($+2.23$ h, $p < .05$): mothers who cite "no slot" but do enter employment in the next year work *more* once they are employed, possibly because slot-constrained re-entry tracks already strong labour-market attachment.

Reading the Under-3 Entry vs. Age-3+ Entry split. The transition models are reported in three columns — overall, Under-3 Entry, and Age-3+ Entry — for transparency. (Figure labels carry the older "Nursery-Entry" and "Kindergarten-Entry" names; throughout the body text we use the age-based labels.) The Under-3 Entry column carries the substantive load. Two reasons. First, sample size: Under-3 Entry contributes about $N = 2{,}200$ person-years on $\approx 2{,}000$ mothers, Age-3+ Entry only $N \approx 545$ on ≈ 495 mothers, because the German daycare attendance rate is already over 92% at age 3 and the Stage 1 pool of non-using and non-employed mothers is correspondingly thin in that age group. Second, composition: relative to Under-3 Entry, the Age-3+ Entry subsample concentrates migration background (9.8% vs. 4.2%, $p < .001$), welfare receipt (14.5% vs. 6.6%, $p < .001$), and low household income (bottom tercile 51.1% vs. 32.8%); has a markedly lower employment-entry rate (32.4% vs. 58.2%); and contains more long-term non-employed mothers (Group C $\approx$ 29% in Age-3+ Entry vs. $\approx$ 7% in Under-3 Entry). Age-3+ Entry is therefore not a parallel main result but a small, persistently disadvantaged late-entry subgroup. The pronounced "not an option" effect that surfaces in this subgroup is

[2] An individual-level cross-tabulation in the Stage 1 risk set confirms that the two transition outcomes are positively but not perfectly correlated ($\phi = 0.385$): of the $N = 2{,}750$ person-years, 40.4% transition into both daycare and employment in t_1, 17.9% into daycare only, 12.6% into employment only, and 29.0% into neither. Among the 311 quality-concerned mothers, the no-daycare / no-employment cell is markedly over-represented (42.4% vs. 29.0% overall), and the both-transitions cell is under-represented (21.9% vs. 40.4%). Quality concerns therefore concentrate in the same individuals on both barrier outcomes, rather than in two disjoint sub-populations.

consistent with a residual stratum of categorical non-users: mothers who, by the time the child is three, still report daycare as fundamentally not under consideration, beyond the ordinary preference dimensions captured by "want to raise myself". Subsample-specific coefficient differences are best read as sensitivity rather than as substantive subgroup heterogeneity (full subsample comparison: Appendix B). The within-person quality story for continuing kindergarten users is the natural complement and is reported separately in Section 6.4 (Stage 3), on a much larger sample.

6.2 Counterfactual (from Stage 1): What If Barriers Were Eliminated?

Translating the transition estimates into policy-relevant magnitudes: what would happen to daycare entry and to work hours if the non-use barriers identified above were eliminated? Figures 3 and 4 show aggregate and individual counterfactual scenarios, separately for Under-3 Entry and Age-3+ Entry.

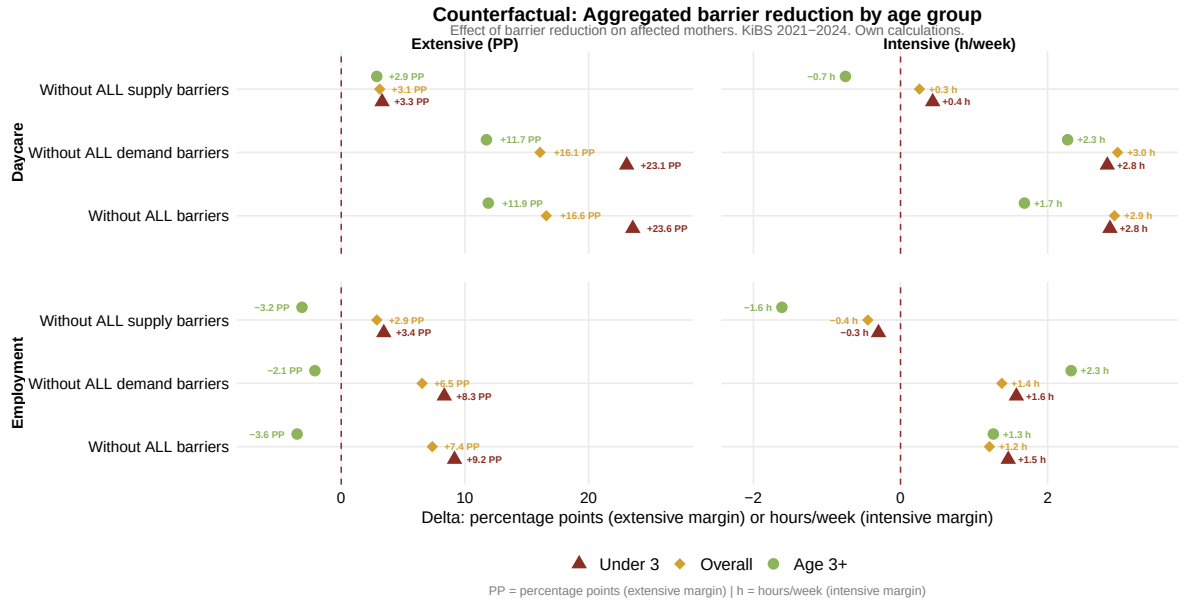


Figure 3: Counterfactual: aggregated barrier reduction by entry-age subsample. **2×2 layout.** Rows: *Daycare* (top) and *Employment* (bottom). Columns: *Extensive margin* (left, Δ in PP) and *Intensive margin* (right, Δ in h/week). ▲ Nursery-Entry (under-3), ♦ Overall, ● Kindergarten-Entry. In the substantive Nursery-Entry subsample ($N \approx 2{,}200$ PY), removing all barriers raises entry by +23.6 PP and adds +2.9 h daycare and +1.5 h work per affected mother per week. The Kindergarten-Entry subsample ($N \approx 545$ PY) is small and selective (Appendix B); effects there are smaller and noisier. KiBS Long Panel 2021–2024.

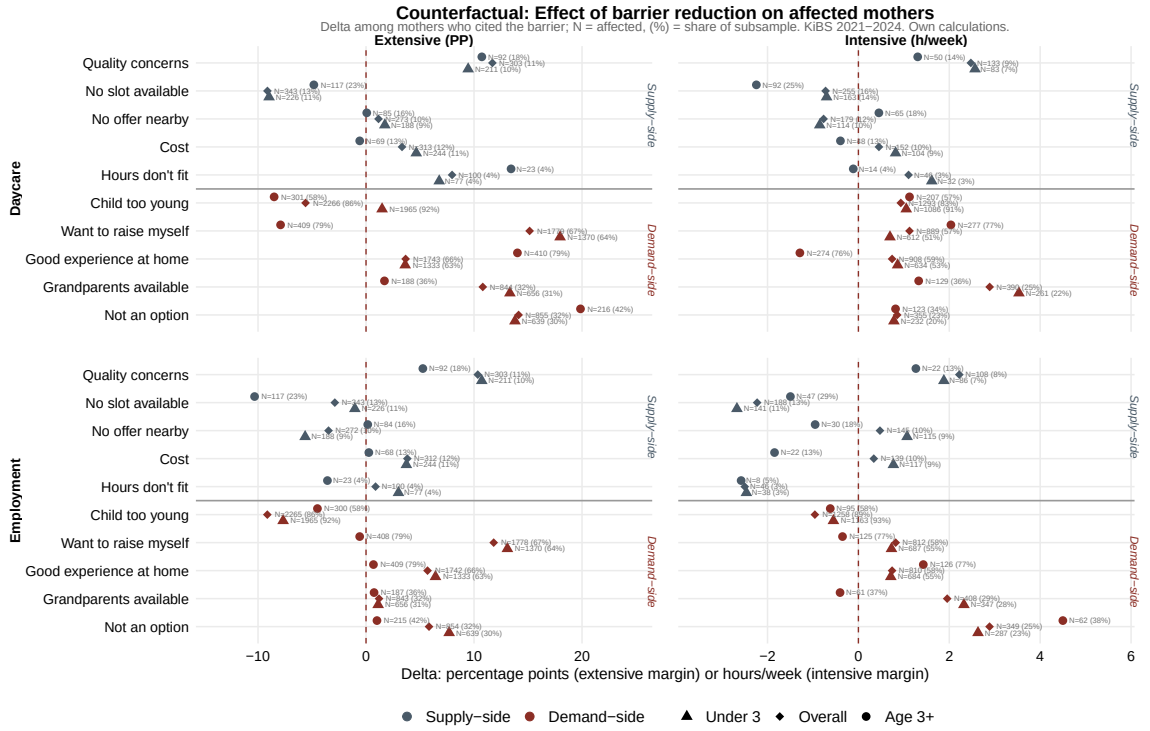


Figure 4: Counterfactual: individual non-use reasons, effect on affected mothers. **2×2 layout.** Rows: *Daycare* (top) and *Employment* (bottom). Columns: *Extensive margin* (left, Δ in PP) and *Intensive margin* (right, Δ in h/week). N = number of affected mothers. Negative values for "no slot" and "child too young" reflect that the catch-up effect disappears when the reason is removed. ▲ Nursery-Entry, ♦ Overall, ● Kindergarten-Entry. Blue = supply-side, red = demand-side. KiBS Long Panel 2021–2024.

Aggregated effects (overall Stage 1 sample, N = 2,642 person-years). Removing *all* supply-side barriers raises daycare entry by +1.0 PP; removing all demand-side barriers by +15.4 PP; removing both by +16.4 PP. In hours terms, removing all barriers adds +2.9 h daycare and +1.2 h work per mother on average, and raises employment entry by +7.3 PP. The bulk of the effect is on the demand side; the supply-side aggregate is small in the joint specification because correlated demand-side reasons absorb most of the variation. The strongest single demand-side lever is "want to raise myself" (about +10 PP overall, 67% affected, +15 PP among affected). The strongest single supply-side lever is quality concerns (about +12 PP among affected, 11% affected), quantifying precisely how binding this constraint is for the mothers who cite it.

A note on the catch-up reasons. "No slot available" (joint OR = 1.68) and "child too young" (OR = 1.32) carry positive coefficients in the joint logit. Setting them to zero in the counterfactual mechanically removes that positive contribution and therefore *reduces* predicted entry and employment for the mothers who cited them. This is the source of the negative employment-PP signs visible in some of the small-sample cells (most strikingly the −3.6 PP for Age-3+ Entry employment) and of the modest aggregate work-hours figures. We retain these reasons in the counterfactual rather than excluding them on the grounds that they are real reasons for non-use as reported by the mothers; the reader should read the aggregate as a partial-equilibrium thought experiment in which all reported barriers are simultaneously absent, including those whose individual conditional contribution is positive (waitlisting, age maturation). Figure 4 disaggregates the contribution of each reason so that the catch-up cases are visible separately.

Under-3 Entry vs. Age-3+ Entry. The two entry-age subsamples respond differently to barrier elimination, but the difference is now quantitative, not categorical. *Under-3 Entry ($N \approx 2{,}200$ person-years).* Removing all barriers raises daycare entry by +23.6 PP, weekly daycare hours by +2.9 h, employment entry by +9.2 PP, and work hours by +1.5 h. The under-3 barrier-elimination policy is therefore predicted to act on *all four margins simultaneously*: more daycare use, more weekly hours of attendance, more women in employment, and modestly more weekly work hours. *Age-3+ Entry ($N \approx 545$ person-years).* Removing all barriers raises daycare entry by +11.9 PP, daycare hours by +1.7 h, and work hours by +1.3 h, but *lowers* employment-entry rate by −3.6 PP. The Age-3+ Entry sample is small and concentrates disadvantaged mothers (Appendix B); we read these magnitudes as a sensitivity check rather than as a parallel main result. Earlier versions of this paper, on a broader Stage 1 risk set that included also already-employed non-using mothers, reported a starker Under-3-vs.-Age-3+ contrast (with under-3 barrier elimination not translating into work hours and age-3+ effects on work hours of $\approx +3.8$ h). On the Stage 1 risk set used here, that contrast is reduced: barrier elimination produces work-hour gains in both age windows, but the action is in the under-3 entry margin, which is where the bulk of the risk-set mothers are.

The Under-3 Entry counterfactual qualifies the common policy narrative that expansion at the under-3 margin is primarily a family-policy instrument: in the Stage 1 population, barrier elimination is predicted to raise both care use *and* employment, consistent with the reading that the constraint that keeps these mothers out of daycare is also the constraint that keeps them out of employment. The subsequent stages ask *why*: Stage 2 (Quality Selection, Section 6.3) tests whether the quality concerns that drive non-use also shape what mothers perceive once they enter, and Stage 3 (Mundlak-Hybrid, Section 6.4) tests whether perceived quality changes within the same mother move care or work hours.

6.3 Stage 2: Quality Selection

Stage 1 showed that quality concerns keep mothers out of daycare. Stage 2 asks what happens to those concerns *after* entry: do mothers who entered worried about quality end up in centres they then rate as better, worse, or no different? The naive "vote with your feet" expectation is that quality-conscious parents sort into better provision, so prior concerns should, if anything, predict *higher* perceived quality after entry. Testing this is informative precisely because the perception is subjective: if prior concerns instead predict *lower* post-entry ratings, that can mean either that the concern tracked a real local supply constraint the mother could not sort around, or that a critical disposition carries forward into how she rates whatever centre she ends up in. The two have different policy implications, so it is worth separating them as far as the data allow. Figure 5 shows the estimates.

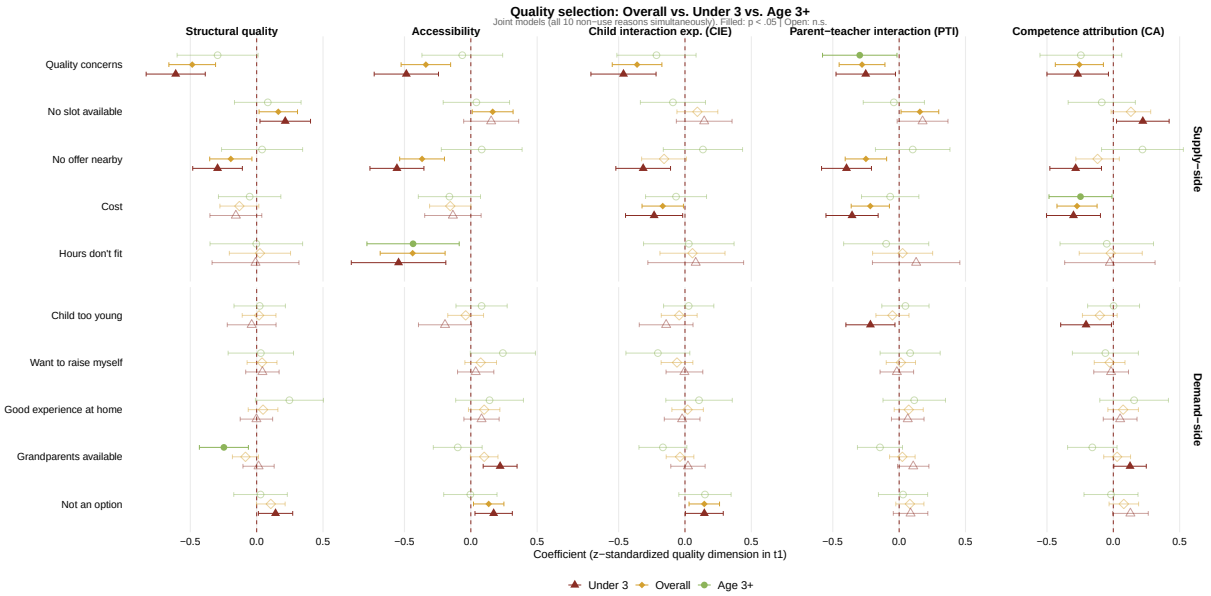


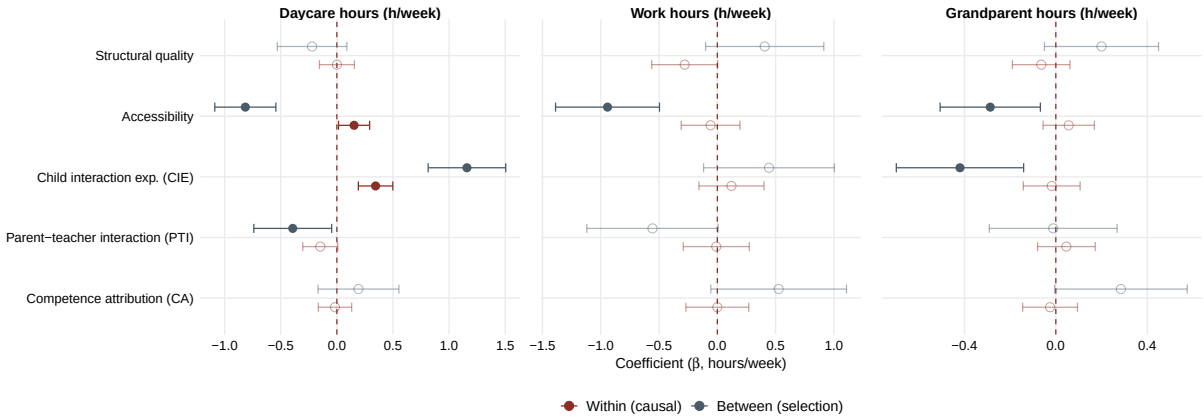
Figure 5: Quality selection: overall vs. nursery vs. kindergarten. Joint models. 5 quality dimensions (z-standardised). Weighted OLS. ▲ Nursery, ♦ Overall, • Kindergarten. Filled: $p < .05$; open: n.s. Stepwise comparison: Appendix Figure 7. KiBS 2021–2024.

Quality concerns in t_0 predict *lower* perceived structural quality and accessibility after entry (stepwise: $\beta = -0.45$ and -0.32 for structural quality and accessibility, respectively, $p < .001$). In the joint specification, conditional on the other nine non-use reasons, the quality-concerns effect on structural quality remains large and negative ($\beta = -0.49$, $p < .001$), and accessibility tightens slightly ($\beta = -0.34$, $p < .001$). All five quality dimensions are reported on the survey-weighted z-standardised scale, so coefficients are directly comparable in SD-units.

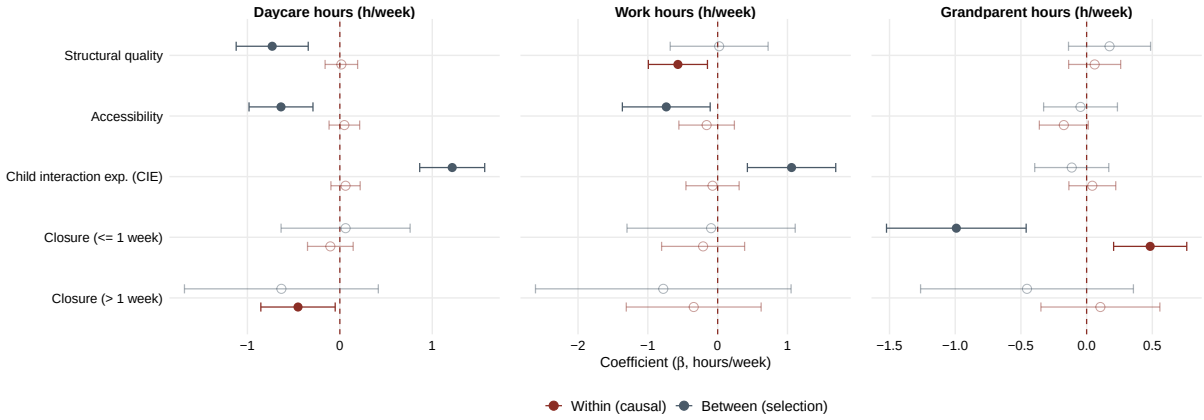
This runs against the naive "vote with your feet" expectation: prior concerns predict *lower*, not higher, post-entry quality ratings. Two readings fit the pattern, and the data as we have them cannot quantify how much each contributes. The first is *structural segregation*: the local supply environment that motivated the concern in the first place was still binding at entry; concerned mothers could not sort around a thin or overburdened supply, and they end up in the centres whose quality they were right to worry about. The Under-3-vs.-Age-3+ difference in the figure is consistent with this — the effect is larger in the under-3 subsample, where the regional supply gap is objectively tighter. The second reading is a *disposition* effect: mothers who came in with concerns hold more demanding or more concrete expectations and rate whatever centre they end up in more critically, a tendency familiar from satisfaction research (people with very specific prior expectations are hard to satisfy anywhere). On this reading the lower rating is not a surprise but the expected continuation of a stable disposition. We cannot put numbers on the relative weight of the two channels — that would need observer-rated quality for the same centres, which KiBS does not provide — and we flag this in the Limitations. What does not depend on the split is the policy-relevant conclusion: quality concerns before entry are not neutralised by choice; the post-entry perception of quality stays lower where the prior concern was higher. If the segregation channel carries real weight, closing the *structural* supply gap is what would let sorting actually work.

6.4 Stage 3: Within/Between Decomposition — Quality, Closures, and Maternal Time Use

Mundlak-Hybrid panel models decompose the quality–behaviour association into within-effects (intra-individual change over time) and between-effects (cross-sectional selection). We report two specifications. Figure 6 (panel a) covers the full 2021–2023 panel with all five quality dimensions but without closure indicators (closures are only observed from 2023 onward). Figure 6 (panel b) re-estimates the model on the 2023–2024 window, adding short ($\leq$ 1 week) and long ($>$ 1 week) closures as additional within/between predictors and restricting the quality vector to the three dimensions with 2023–2024 repeated measurement (see Table 2).



(a) Main specification. Kindergarten 2021–2023, all five quality dimensions, no closure indicators.



(b) Closure specification. Kindergarten 2023–2024, three quality dimensions with 2023–2024 coverage plus closure indicators (short/long).

Figure 6: Mundlak-Hybrid: Within vs. Between effects across two specifications. Red = within (intra-individual change); blue = between (cross-sectional selection). Filled: $p < .05$; open: n.s. Panel (a): $N \approx 11{,}000$ person-years, 4,902 mothers (2021–2023). Panel (b): $N \approx 5{,}200$ person-years, 2,612 mothers (2023–2024). KiBS Long Panel.

Panel a: Main specification (2021–2023). *Daycare hours.* Two within-effects are significant at $\alpha = .05$: child interaction experience ($\beta_W = +0.35$, $p < .001$) and accessibility ($\beta_W = +0.15$, $p = .034$). Both are consistent with the trust/reliability mechanisms posited in Section 2: when a mother's perception of her child's well-being improves, she uses more hours; when accessibility improves, she uses more hours. Structural quality and competence attribution

show no within-effect. Between-effects are large and in partly opposite directions: child interaction experience ($\beta_B = +1.16$, $p < .001$), accessibility ($\beta_B = -0.82$, $p < .001$), parent–teacher interaction ($\beta_B = -0.39$, $p = .025$). The accessibility *sign flip* (negative between, positive within) is the clearest illustration of why decomposition matters. Cross-sectionally, high-accessibility centres are used by mothers with shorter usage patterns — a selection effect: mothers who need only a few hours sort into the centres that make short, flexible attendance easy. Longitudinally, when accessibility improves for the same mother, she then uses more hours — a behavioural response: better temporal fit lets her rely on the centre for a larger share of the week. The two components point in opposite directions because they answer different questions (which mothers use which centres, versus what a given mother does when her own centre becomes easier to rely on).

Work hours. No quality dimension produces a within-effect significant at $\alpha = .05$. Structural quality has a marginally negative within-coefficient ($\beta_W = -0.28$, $p = .051$); we do not interpret this as causal. The positive cross-sectional CIE–work-hours association is not replicated once the between-component is separated ($\beta_B^{\text{CIE}} = +0.44$, $p = .12$; $\beta_W^{\text{CIE}} = +0.12$, $p = .40$). Accessibility shows a large negative between-effect ($\beta_B = -0.94$, $p < .001$), consistent with selection into short-hour centres by mothers with short-hour work schedules.

Grandparent hours. No within-effects of quality are significant. Between-effects: accessibility $\beta_B = -0.29$ ($p = .009$); child interaction experience $\beta_B = -0.42$ ($p = .003$); cross-sectional patterns consistent with selection, not within-person quality-driven substitution.

Panel b: Closure specification (2023–2024). Adding closures to the 2023–2024 window ($N = 5{,}224$ person-years, 2,612 mothers) produces the cleanest within-person identification of closure effects: the short- and long-closure indicators have genuine within-variation across the two waves, so the Mundlak decomposition separates shock responses from cross-sectional closure frequency.

Daycare hours. **Long closures (> 1 week) reduce weekly daycare hours** within-person: $\beta_W = -0.45$ ($p = .030$). Short closures have no within-effect. The quality within-effects are not replicated in this shorter window: structural quality and CIE are n.s. within. Between-effects repeat the pattern from panel a (structural quality and accessibility negative, CIE positive, all $p < .001$).

Work hours. **No closure effect on work hours** (both short and long closures: n.s. within and between). Structural quality shows a negative within-effect ($\beta_W = -0.57$, $p = .007$); given the absence of such an effect in panel a (and against a reliability-based prior), we treat this as exploratory rather than causal evidence. The CIE between-effect on work hours now reaches significance ($\beta_B = +1.06$, $p = .001$), mirroring the selection pattern for daycare hours.

Grandparent hours. **Short closures substitute into grandparent care**: within-person, a short closure adds $\beta_W = +0.48$ hours of grandparent care per week ($p < .001$). The long-closure within-effect is small and not significant ($\beta_W = +0.11$). The short-closure *between*-effect flips negative and large ($\beta_B = -0.99$, $p < .001$): mothers in centres with persistently frequent short closures average less grandparent care. A plausible reading of this between-pattern is socio-spatial: centres with persistently high closure frequency are disproportionately located in disadvantaged areas serving families with migration backgrounds, who in turn have less access to local intergenerational support (Gambaro et al., 2024; Steinberg, 2025). The negative between-

effect would then be a selection artefact rather than a within-person substitution null, reinforcing the within-person interpretation of the positive short-closure effect.

Combined reading. The two panels together support four claims that map onto our four exploratory hypotheses. First, *within* continuing ECEC attendance, the only robust quality within-effect is CIE on daycare hours (panel a), consistent with the care-side branch of **H2**: perceived quality moves care time within-person. Second, on **H4** (effect vs. selection): the cross-sectional quality–work-hours association is dominated by between-person selection and does not survive the within/between decomposition; the within-effects on work hours are indistinguishable from zero. The decomposition therefore answers the open test in **H4** clearly: the cross-sectional quality–work-hours link reflects between-person selection (which mothers use which centres), not a within-person behavioural response to changing quality. Third, **H3** (closures register at some margin): long closures displace daycare hours and short closures substitute into grandparent care; neither shifts work hours. The reliability-shock channel is therefore on care, not on work. Fourth, the work-side branch of **H2** is *not* supported: no quality dimension moves work hours within-person at $\alpha = .05$. The grandparent-substitution channel posited by H2 only shows up through the reliability-shock channel of H3, not as a within-person response to perceived quality. The absence of a work-hours response is not, however, the same as indifference: mothers may need to work despite dissatisfaction with daycare, simply because the household budget or the employment contract leaves no room to reduce hours. On that reading, the work-hours null is a footprint of economic compulsion rather than of preference indifference, and the behavioural cost of staying in a low-perceived-quality arrangement is then plausibly absorbed on other margins that the present design does not measure (informal patches, in-home compensations, parental stress); we revisit this constraint reading in the Limitations.

7 Discussion

Trust begins long before enrolment. Two decades of German ECEC-and-labour-supply research have focused on the quantity side of the expansion. The quality side has received less attention, but it is now the side on which families are making their decisions. Before a mother asks how many hours to book and whether to take up work, she asks whether to use an institutional slot at all, and that answer depends on more than price and availability: it depends on whether she trusts the centre enough to leave her child there. Where that trust does not form, the family stays out of daycare. In our Stage 1 data, this looks observationally identical to a supply-side shortage: the slot is unused, regardless of whether none was available or none was wanted. We see the role of quality concerns as an entry barrier as the first substantive finding of the paper. In the joint specification on the Stage 1 risk set (mothers neither using daycare nor employed in t_0, $N = 2{,}642$ person-years, mostly post-birth parental-leave window), the demand-side preference "want to raise the child myself" is the strongest single barrier (OR = 0.48), with quality concerns the second strongest (OR = 0.56). The Under-3-targeted counterfactual shows that removing these barriers raises daycare entry by +23.6 PP and lifts weekly daycare hours by +2.9 h and work hours by +1.5 h. The same barriers also predict employment entry on the same sample (quality concerns: OR = 0.62), so the trust deficit is not specific to daycare but extends to the work choice daycare is supposed to enable. Two parallel KiBS analyses inform the interpretation

here: Neuberger et al. (2026) report that vulnerable families consistently rate relational and interaction quality lower, and Neuberger et al. (2025a) show that unplanned closures have risen sharply over 2021–2024, with a dose–response gradient on maternal well-being that runs largely through degraded process-quality perceptions. The expansion story no longer ends at the slot, because the slot does not reliably deliver what it is supposed to.

Concerns do not dissolve on entry; they shape what is then perceived. The second stage asks whether mothers who entered with quality concerns then managed to sort into better centres. They did not. Prior concerns predict *lower* perceived structural quality and accessibility after entry ($\beta = -0.49$ and -0.34, $p < .001$ in the joint specification), and the effect is strongest in the under-3 subsample, where the regional supply gap is tightest. Two readings fit this pattern. The first is structural segregation: the thin supply that motivated the concern in the first place was still binding at entry, so concerned mothers could not sort around it. Two German studies support this reading. Bader and Scholz (2024) document, from TALIS Starting Strong, that slot-allocation criteria at the centre level are themselves inequality-relevant: in the multi-level governance setting, centre directors exercise discretion on admission criteria, and that discretion correlates with socio-economic outcomes. Carbuccia et al. (2025) take this from association to causation in a randomised controlled trial: administrative burden and information costs produce socially unequal access to early childcare. The second reading is a stable critical disposition: mothers who came in worried hold more demanding or more concrete expectations and rate whatever centre they end up in more critically — the lower rating is then not a surprise but the continuation of a disposition, as satisfaction research would lead one to expect. Our data cannot put numbers on how much each channel contributes (that would need observer-rated quality for the same centres, which KiBS lacks), but the policy reading does not hinge on the split: the post-entry experience of care is not the self-correcting outcome of preference-driven choice that a "vote with your feet" logic would predict (Burghardt, 2018; Jessen et al., 2020). Concerned mothers stay concerned, whether because they could not move or because they were right the first time.

Within users, the trust signal moves care time, not work hours. The third stage takes only the mothers who did enter and asks which dimensions of their quality perceptions move their behaviour within-person, over time. Among the five dimensions, the signal that most reliably moves within-person behaviour is *child interaction experience* (CIE): a parent-reported four-item index of whether the child feels well, has a trusting relationship with at least one teacher, has enough time from the staff, and has playmates. This is not a drop-off rating but a general assessment of how the child is doing at the centre, and it is the dimension most tightly anchored in what parents themselves can observe. In ABI terms (Mayer et al., 1995) it loads on Benevolence ("enough time for my child") and on the felt quality of the teacher–child bond; in Nentwig-Gesemann and Hurmaci (2020)'s vocabulary, it is the information parents pull out of the black box when the pedagogical day is otherwise opaque to them. The within-estimate is $\beta_W = +0.35$ ($p < .001$) on *daycare hours*. One caveat applies to the care-hours channel itself: parents do not always choose their booked hours freely either. Some centres offer only fixed booking modules or only full-day places (in parts of Germany public provision was for years full-day-only), so the within-person care-hours response is plausibly driven by the

subset of mothers who do have room to adjust, or it runs through switching between centres rather than adjusting hours at a given centre. There is also a possible confound in the other direction — longer-hours or full-day places may themselves come with higher perceived quality. We read the CIE care-hours effect as robust but, if anything, conservatively estimated, and not as evidence of a frictionless market in hours. On work hours, no quality dimension reaches a within-effect significant at $\alpha = .05$. So the cross-sectional quality–work association is largely selection: for CIE on daycare hours, $|\beta_B/\beta_W| = 3.4$, with the between-component more than three times the size of the within-component. A good portion of the cross-sectional quality–behaviour associations reported in earlier work (e.g. Camehl et al., 2018; Stahl and Schober, 2020) is therefore between-person selection rather than a behavioural response to changing quality. The asymmetry across dimensions (significant within-effect on CIE, null on the structural and reliability dimensions) fits the meta-analytic finding of Hausladen et al. (2026) that child–staff ratios, the most operationalisable structural lever, have essentially no measurable association with process quality in kindergarten-age classrooms ($z = -0.06$, n.s.). The Stage 3 sample is kindergarten-only by design, so we cannot compare under-three and kindergarten-age within-effects in our data; whether the under-three part of the Hausladen pattern translates into perceived quality and behaviour is something we can only speculate about. For the kindergarten age range we can say that the chain "structural change → observable process quality → perceived quality → used hours" has limited within-person traction: structural variation in kindergarten-age classrooms is unlikely to propagate to observable process quality (Hausladen et al., 2026), and what does move within-person is not the structural dimensions but CIE, which parents observe directly at the child rather than infer from structure. This is a narrower claim than "quality matters". First, quality moves the care margin, not the work margin: within the same mother, perceived quality goes up, daycare hours go up, work hours do not. Second, the mechanism is Benevolence, not Ability: what shifts behaviour is the parent's sense that the staff are attentive to the child, not her inference that they are professionally competent (the competence-attribution dimension, where we find no within-effect). This is itself informative about what parents read as quality. The dimension that moves their behaviour is the warm, responsive side of care — whether the child feels secure and is given time — rather than the narrower educational mission that ECEC also carries. For the parental understanding of quality, and for the role centres are expected to play, this suggests that warmth and sensitivity are what draw parents in, with the educational function recognised less, or valued as an addition rather than the core. Experimental vignette evidence (Philipp et al., 2025) suggests that high perceived quality raises *normatively recommended* work hours; our revealed-behaviour evidence shows that the norm does not translate into actual hours, at least in the middle-class KiBS population where employment is already near saturation and intensive-margin elasticities are correspondingly low. The Limitations make this explicit, but it is worth flagging here: the work-hours margin is structurally narrow in Germany. Many centres offer only a few fixed booking lengths, and most employment contracts prescribe weekly hours within a narrow band; a mother dissatisfied with quality can register that dissatisfaction without being able to act on it through work-hours adjustment. The absence of a within-person work-hours signal is therefore not evidence that quality does not matter to mothers, but evidence that the work-hours margin is the wrong place to look. Where the dissatisfaction goes instead is plausibly twofold: into the daily search for stop-gap arrangements (grandparents, private patches, extra trips), and into parental stress and reduced well-being. A mother who can neither

cut her work hours nor shorten her child's attendance stays exposed to the problem, and the cost surfaces as strain rather than as a visible change in time use. Neuberger et al. (2025a) document, on the same data, that closures lower maternal well-being largely through degraded process-quality perceptions; their well-being effect and our null on work hours are two readings of the same constraint, from different sides. One further qualification guards against over-reading the result as the outcome of free individual choice. How long a child is in care, and how readily a mother returns to or extends work, are shaped by regional norms as much as by individual preference: in East Germany the supply gap is smaller, opening hours are traditionally longer, and full-day attendance and full-time maternal employment carry more social acceptance than in the West. The behavioural margins we observe are therefore embedded in normative and institutional context, not the free play of a market in hours.

Reliability is latent in steady state and becomes visible when it breaks. The fourth result complements the CIE within-effect from the opposite side. The reliability-signalling dimensions (structural quality, accessibility) produce no robust within-person work-hours response, but the closure indicators do move care: long closures compress weekly daycare hours ($\beta_W = -0.45$) and short closures substitute into grandparent care ($\beta_W = +0.48$). Neither reaches work hours. The natural reading is that reliability is taken for granted in steady-state operation. Parents do not update their behaviour on reliability signals they do not register, so the behavioural footprint of reliability only surfaces where reliability actually breaks. Neuberger et al. (2025a) approach the same channel from the breakdown side and show that closure effects on maternal well-being run largely through degraded perceptions of process quality. Their mediation estimates and our within-estimates describe the same mechanism from opposite directions. The policy implication is that the steady-state null coefficients on the reliability dimensions are easy to misread as "reliability does not matter". They actually say that as long as the centre runs as promised, parents have nothing to react to, and the within-coefficient sits near zero. The footprint of reliability becomes visible when reliability breaks; that is what our closure findings pick up. Reading the steady-state coefficients alone would lead policymakers to underestimate how much stable operation matters.

Policy consequence is age-specific and structural, not graduated. In the Stage 1 risk set, removing non-use barriers in the Under-3 Entry subsample raises entry sharply (+23.6 PP) and lifts both care and work margins (+2.9 h daycare used per week, +1.5 h work hours). The Stage 1 work-hour gain may at first look as if it contradicts the Stage 3 null on work hours, but it does not: Stage 1 estimates the extensive-margin barrier (mothers neither using daycare nor employed at t_0, where removing the barrier opens both transitions at once), whereas Stage 3 estimates the intensive margin among continuing users, where employment is already near saturation and the room for further work-hours adjustment to perceived-quality change is correspondingly small. The two estimates are therefore consistent: quality concerns are a barrier where the barrier is still binding, and quality perceptions move care but not work where the barrier has been overcome. Effects in the small Age-3+ Entry subsample are quantitatively smaller and operationally less informative because the subsample is selective on disadvantage (Appendix B): migration background is twice as common, low household income is over 50%, and the residual labour-market re-attachment is structurally weaker after three years of non-

employment. The action is therefore in the under-3 entry margin, where the bulk of the risk-set mothers are concentrated; once they have entered, both their weekly daycare hours and their work hours move upward. The within-person quality story for continuing kindergarten users (Section 6.4, Stage 3) is the natural complement: it operates on a different sub-sample and asks a different question (does perceived quality move behaviour *within* the same mother), and is where the trust-based care-time mechanism becomes visible.

Hypotheses revisited. The four exploratory hypotheses from Section 3 come out with mixed but interpretable verdicts. **H1** (quality concerns lower entry) is *supported*: on the Stage 1 risk set, quality concerns reduce the odds of daycare entry (OR = 0.56, $p < .001$) and of employment entry (OR = 0.62, $p < .001$), with only "want to raise the child myself" larger in magnitude. **H2** (perceived quality among users moves both care and work margins) is *partially supported*: on the care side, CIE moves daycare hours within-person ($\beta_W = +0.35$, $p < .001$); on the work side, no dimension reaches $\alpha = .05$ within-person. H2 therefore holds on the care side, but not the work side. **H3** (closures register at some margin) is *supported with a specific pattern*: long closures compress weekly daycare hours, short closures substitute into grandparent care, neither reaches work hours. The reliability-shock channel is on care, not on work. **H4** (effect vs. selection, no prior commitment) is *settled*: the cross-sectional quality–work-hours association is largely between-person selection and does not survive the Mundlak-Hybrid decomposition. On the daycare-hours channel the within-component is non-trivial (notably for CIE), so selection is not the whole story there. Taken together, the hypotheses confirm a trust-based care-time mechanism, identify selection as the dominant explanation for the cross-sectional quality–work-hours pattern, and make clear that perceived quality does not *directly* move maternal work hours within-person in this population. "Directly" is the operative word: we estimate the year-to-year within-person response, not whether a sustained, society-wide erosion of trust in the system would shift labour supply over the medium term, which our design cannot speak to.

7.1 Limitations

1. *Selective sample.* The KiBS Long Panel is markedly selective on socio-economic status: 82% of mothers hold the Abitur, 48% are in the top household-income tercile, and the employment rate is 93%. Income selectivity grows across waves (42.7% → 61.6% high income from 2021 to 2024 among users). Policy-relevant populations (low-education, low-income, migration background) are under-represented; our estimates therefore best describe the middle- and upper-middle-class KiBS population. Effect magnitudes, in particular counterfactual PP, are upper bounds for that population and may be non-transferable to disadvantaged subgroups. Meta-analytic evidence indicates that universal ECEC tends to be an equaliser, with stronger gains for children from low-SES households (Schmutz, 2024), and that the inequality-reducing potential of center-based formal childcare can be undone if other childcare arrangements widen gaps in parallel (Ghirardi and Gioachin, 2025). To the extent that disadvantaged families benefit more from quality improvements than the KiBS-overrepresented middle class, our middle-class effect sizes are plausibly conservative for population-level policy projections.
2. *Single mothers enter the samples but cannot be separated out.* Single mothers are part of the estimation samples throughout. The single-parent dummy is included as a control in

all stages, and the models converge cleanly without rank deficiency. The single-mother subgroup is, however, small in the analytic samples: only N=51 in the Stage 1 risk set, about 1.9%. The reduction relative to the population share (around 4%) reflects the fact that single mothers are more often employed and therefore drop out of the Stage 1 pool. With this subgroup size we can adjust for partnership status but cannot reliably estimate effect heterogeneity *within* the single-mother subsample. The single-mother stratum is arguably the subgroup for whom a "daycare as labour-market policy" argument is most directly relevant. Recovering that subgroup will require dedicated over-samples or administrative linkage.

3. *Subjective quality measures only.* No observational quality indicators (ECERS, CLASS) are available; perceived quality correlates weakly with objective quality (Camehl et al., 2015; Schreiber, 2004).
4. *Accessibility scale reliability.* Cronbach's $\alpha = 0.53$ is at the lower bound of acceptability. Results involving this dimension should be interpreted with caution; the scale may be better treated as a formative index in future work.
5. *Mundlak-Hybrid temporal coverage.* The main within-estimates for the full five-dimensional quality vector rest on 2021–2023, because PTI and CA were fielded only in 2023 and therefore contribute no within-variation to the 2023–2024 panel. Closure indicators are only observed from 2023 onward and therefore enter only in the 2023–2024 specification (Panel b), which is restricted to the three quality dimensions with 2023–2024 repeated measurement. In both windows, at most two year-on-year within-differences per child are available, limiting power for within-effects.
6. *Associations, not strict causality.* The within-estimator controls for time-invariant confounding but not for time-varying confounders (e.g., simultaneous life events).
7. *Stage 1 model choice.* All four Stage 1 outcome models (two extensive-margin logits, two intensive-margin OLS) are estimated with cluster-robust standard errors on the mother rather than with a random mother intercept. Most mothers contribute only one t_0-row to the Stage 1 risk set, so a random-intercept specification would estimate the intercept variance at the boundary; the cluster-robust sandwich captures the within-mother correlation that does exist (for the minority of mothers with multiple consecutive non-employed and non-using years before transition) without forcing a mixed-model spec.
8. *No panel weights in the main specifications.* Quality-selection models use survey weights as a sensitivity check; transition and Mundlak-Hybrid models are unweighted, following the lack of longitudinal weights in KiBS.
9. *Time-constancy of non-use reasons.* Reasons are taken from each mother's last year of non-use; for mothers with multiple entry/exit episodes (a small subgroup), this reflects the latest reasons rather than contemporaneous ones. Robustness: models restricted to mothers with a single transition yield comparable coefficients (not shown).
10. *Stage 1 sample is dominated by parental-leave mothers.* The Stage 1 risk set is approximately 88% Group D (not employed, on parental leave) and 11% Group C (long-term not employed, not on leave); Group B (employed on leave) is residual at 0.9%. Stage 1 results therefore describe the parental-leave-exit transition rather than a generic non-user-to-user transition. We retain all groups and control via parental-leave dummies (reference: Group C); a robustness specification excluding Groups B and D would identify only the long-term non-employed population (Group C) and is left for future work because the re-

sulting sample would be small. The Stage 3 within-person quality analysis is unaffected because it operates on continuing kindergarten users.

11. *Age-3+ Entry subsample is small and selective.* The Stage 1 Age-3+ Entry subsample ($N \approx 545$ person-years) is roughly one-fourth of the Under-3 Entry subsample. It is over-represented on migration background and welfare receipt, has a markedly lower top-tercile income share, and under-represented on current parental-leave status (because statutory parental leave is normally exhausted by age 3). Age-3+ Entry results should be read as a sensitivity check on the under-3-dominant main story; see Appendix B for the full subsample comparison.
12. *Partial-equilibrium counterfactuals.* Setting a reason to zero assumes the mother's other reasons and the broader choice context remain constant; general-equilibrium effects (e.g., how expanded supply would change centre quality) are not modeled.
13. *Multiple comparisons.* The paper reports roughly 200 tested coefficients across transition, quality-selection, and Mundlak-Hybrid models. We do not apply a global FDR or Bonferroni correction; readers should interpret marginal findings (p between .01 and .05) as exploratory rather than confirmatory.
14. *Constrained behavioural margins and the small-effect interpretation.* Two features of the German daycare-and-employment landscape compress the margin on which perceived quality could plausibly act, and we read the small within-person work-hours effect in this light. (i) Daycare bookings are not free-flowing. Many centres offer only a small number of fixed slot lengths (typically 25/35/45 hours per week), so a mother who would prefer to extend or compress the hours her child actually attends often cannot do so without changing the centre itself, which is in turn supply-constrained. (ii) Employment contracts in Germany typically prescribe weekly hours within a narrow band; a mother dissatisfied with the quality of care her child receives cannot just trim her work hours by one or two per week to compensate, because contracts and the household budget pin down the level. Under these constraints, the room to translate quality concerns into a work-hours response is narrow even where the perception is strong. So the pattern we observe (quality concerns register before entry, predict lower perceived quality after entry, and move within-person daycare hours through CIE, but not work hours) is consistent with a population that perceives quality and is partly trapped by it: aware of the issue, dissatisfied with it, but structurally unable to act on it through the work-hours channel earlier literature has tended to privilege. The small within-person work-hours effect is not evidence that quality does not matter to mothers; it is evidence that the work-hours margin is the wrong place to look for the footprint of quality dissatisfaction in this institutional setting. Where does that footprint go instead? Plausibly to the well-being margin: a mother who cannot shorten her work hours or her child's attendance stays exposed to the problem, and what shows up is chronic strain rather than time-use change. Three lines of evidence support this absorption reading. Neuberger et al. (2025a), on the same data, document that degraded process-quality perceptions translate into lower maternal well-being. Hiekel and Kühn (2024) show that the pandemic produced a gender mental-health gap specifically among parents in Germany, consistent with childcare-related strain accumulating on the mental rather than the time-use side. Ito et al. (2024) document the parallel international pattern: daycare closures in Japan during COVID-19 raised parental stress (measured by the Kessler Psychological Distress Scale), with the effect

concentrated among mothers and non-regular workers. Bertogg et al. (2026) extend the timeline and show that the gendered well-being aftermath of the pandemic persists into the medium term. Together, these strands suggest that quality dissatisfaction surfaces on the well-being margin when the work-hours margin is not available.

8 Conclusion

For the population this panel can speak to, the quality margin of German ECEC works as a trust-based care-time mechanism, not as a labour-supply lever. Among continuing kindergarten users (Stage 3), perceived quality moves care hours within the same mother but not her work hours, and it does so primarily through what she sees of her child's everyday experience. At the upstream entry margin (Stage 1), demand-side barriers, especially the preference to raise the child oneself and quality concerns about available care, weigh heavily on entry above and beyond the reported shortage of slots. (Our measure is the relative weight of parent-reported reasons, not the objective availability of places: for families who never sought a slot, we cannot observe whether one would have been available.) Removing those barriers in the Under-3 Entry subsample (Stage 1 counterfactual) produces movement on all four margins: entry, daycare hours, employment entry, and work hours. The Age-3+ Entry subsample is too small and too selective on disadvantage to carry a parallel main result. Reliability enters the same picture from the opposite side (Stage 3 Panel b): it is latent in steady-state operation, but unplanned closures reveal the same Benevolence channel through which quality moves care, with short closures pushing mothers onto grandparent care and long closures compressing used hours. Neither reaches work hours. Extrapolation to populations under-represented in KiBS (low-education, low-income, migration-background, and single-mother households) will require complementary data and should not be read off these estimates. We see the contribution of the paper as identifying the channel: in the middle- and upper-middle-class population the expansion has successfully reached, the behavioural response to quality is located on care time, and its micro-foundation is trust in the Benevolence sense, visible both when quality improves and when reliability fails.

Statements and Declarations

Funding. The KiBS study (DJI-Kinderbetreuungsstudie / DJI Child Care Study), within which the data analysed here were collected, is funded by the German Federal Ministry of Education and Family Affairs, Senior Citizens, Women and Youth (BMBFSFJ) and conducted at the German Youth Institute (DJI). No additional external grants were received for the analyses reported in this manuscript.

Competing Interests. The authors have no competing financial or non-financial interests to disclose. All four authors are employed at the German Youth Institute (DJI), the institution that fields the KiBS survey. The survey instrument and the analyses reported here are independent research products; the views expressed are those of the authors and do not necessarily reflect the views of the DJI.

Author Contributions. **Franz Neuberger** (FN): conceptualisation, methodology, formal analysis, software (R), visualisation, writing – original draft, writing – review and editing. **Mariana Grgic** (MG): conceptualisation, writing – review and editing, supervision. **Johannes Wieschke** (JW): data preparation, conceptualisation, writing – review and editing. **Susanne Kuger** (SK): conceptualisation, writing – review and editing, supervision. All authors read, revised, and approved the final manuscript.

Data Availability. The KiBS Long Panel data underlying this study are currently restricted to in-house use because the survey is still ongoing and the data-protection framework for external access is under development. The data will be made available to the scientific community via the DJI Research Data Centre (FDZ) from 2026, subject to a data-use agreement and the standard FDZ access procedures. The R analysis code that reproduces all results, figures, and tables in the paper from the prepared analysis dataset can be made available to reviewers and editors on request. Regional supply-side indicators (*KitaKompakt 2025*) and the federal monitoring data cited are publicly available from the sources referenced in Section 4.

Ethics Approval and Consent. The analyses use anonymised secondary survey data collected within the KiBS study at the DJI under the standing ethical and data-protection framework of the institute. KiBS participants are informed about the study purpose, the voluntary nature of participation, and the anonymised use of their responses for scientific research at the point of recruitment; no individual-level identifiers were available to the authors during analysis. No new primary data were collected for the present manuscript.

Use of Generative AI and AI-Assisted Tools. During the preparation of this manuscript, the authors used the large language model *Claude* (Anthropic) as a writing and coding assistant. Specifically, Claude was used to (i) draft and refactor portions of the R analysis code, with all outputs verified by the authors against the underlying data; (ii) copy-edit the manuscript for English-language style and consistency; (iii) cross-check internal consistency of numbers, hypothesis-to-stage mappings, and cross-references between sections. The model was not used to generate research questions, theoretical claims, empirical results, or substantive interpretations.

The authors verified all AI-assisted output, take full responsibility for the content of this publication, and explicitly affirm that no AI tool is listed as an author. Use of AI was disclosed here in line with the journal's policy on AI-assisted writing.

References

Anders, Yvonne and Elisa Oppermann (2024). "Quality in early childhood education and care: A theoretical research synthesis and an integral framework of pedagogical quality". In: *Zeitschrift fur Erziehungswissenschaft* 27.2, pp. 551–577. ISSN: 18625215. DOI: 10.1007/s11618-024-01218-7.

Bader, Samuel and Antonia Scholz (2024). "De facto-Kategorisierung und Ungleichheit bei der Kita-Platzvergabe in Deutschland: Der Rechtsanspruch auf Frühe Bildung und seine Realität". In: *Zeitschrift für Sozialreform* 70.3, pp. 1–26. ISSN: 0514-2776. DOI: 10.1515/zsr-2024-0010.

Bauernschuster, Stefan and Martin Schlotter (Mar. 2015). "Public child care and mothers' labor supply—Evidence from two quasi-experiments". In: *Journal of Public Economics* 123, pp. 1–16. DOI: 10.1016/j.jpubeco.2014.12.013. URL: https://doi.org/10.1016/j.jpubeco.2014.12.013.

Beringer, Samira, Martin Bujard, and Sabine Diabate (2025). "Conflicting views? How personal attitudes and perceived public opinion towards formal childcare differ". In: *Families, Relationships and Societies*. DOI: 10.1332/20467435Y2025D000000062. URL: http://beringer_2025_conflicting_views_childcare_attitudes_.pdf.

Berth, Felix, Dominik Hank, and Leonie Kleinschrot (2025). "Das westdeutsche Kitasystem im Wandel: Einstellungen, Diskurse und Angebote zur Kleinkindbetreuung seit den 1970er-Jahren". In: *Zeitschrift für Erziehungswissenschaft*. DOI: 10.1007/s11618-025-01322-2.

Bertogg, Ariane, Tiziana Nazio, and Matteo Piolatto (2026). "A Gendered Aftermath? Medium-Term Implications of COVID-19 for Affective Well-Being Among Mothers and Fathers in Finland, Germany, Italy, and the Netherlands". In: *Social Indicators Research* 182.1, p. 24. DOI: 10.1007/s11205-026-03825-9. URL: http://bertogg_2026_a_gendered_aftermath_medium_term_covid_.pdf.

Betz, Tanja (2015). "Ungleiche Kindheiten - ungleiche Kitas: Zur Reproduktion sozialer Ungleichheit im vorschulischen Bildungssystem". In: *Zeitschrift für Soziologie der Erziehung und Sozialisation* 35.2, pp. 149–168.

— (2022). "Elternschaft im Kontext frühkindlicher Bildung: Zwischen Partnerschaft und Defizitzuschreibung". In: *Zeitschrift für Erziehungswissenschaft* 25.3, pp. 661–678. URL: https://link.springer.com/article/10.1007/s11618-022-01112-3.

— (2025). "Das Verhältnis zwischen Kindertageseinrichtung und Familie: Positionen und Perspektiven von Eltern in der Frühen Bildung, Betreuung und Erziehung". In: *Bildung und Erziehung* 78.3, pp. 367–384. DOI: 10.7788/buer-2025-780307. URL: https://www.vr-elibrary.de/doi/10.7788/buer-2025-780307.

Betz, Tanja and Sabine Bollig (2023). "Bildungs- und Erziehungspartnerschaften in der frühkindlichen Bildung. Doing Collaboration als Konzept zur Erforschung eines Programms". In: *Qualitätsentwicklung in der Frühen Bildung. Akteure - Organisationen - Systeme*. Ed. by Regine Schelle, Kristine Blatter, Stefan Michl, and Bernhard Kalicki. DOI: 10.25656/01:28082.

BMBFSFJ (2025a). *Kindertagesbetreuung Kompakt. Ausbaustand und Bedarf 2024*. Tech. rep. Berlin: Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend.

— (2025b). *Monitoringbericht zum KiQuTG 2025. Monitoringbericht 2025 des Bundesministeriums fuer Bildung, Familie, Senioren, Frauen und Jugend nach § 6 Absatz 2 KiQuTG*

fuer die Berichtsjahre 2023 und 2024. Tech. rep. Berlin. URL: http://bmbfsfj_2025_monitoringbericht_kiqutg_2025_.pdf.

Boll, Christina and Till Nikolka (2026). "The impact of a parental leave benefit reform on parents' leave-taking, labor supply and childcare arrangements". In: *Review of Economics of the Household*. DOI: 10.1007/s11150-026-09845-1. URL: http://boll_2026_impact_parental_leave_benefit_reform_.pdf.

Burghardt, Lars (2018). "Elterliche Einstellungen zum Krippenbesuch und deren Einfluss auf die Inanspruchnahme einer Krippe". PhD thesis. Bamberg. DOI: 10.20378/irbo-53970. URL: https://fis.uni-bamberg.de/handle/uniba/44587.

Camehl, Georg F., Pia S. Schober, and C. Katharina Spiess (Nov. 2018). "Information asymmetries between parents and educators in German childcare institutions". In: *Education Economics* 26.6, pp. 624–646. ISSN: 0964-5292. DOI: 10.1080/09645292.2018.1463358. URL: https://www.tandfonline.com/doi/full/10.1080/09645292.2018.1463358.

Camehl, Georg F, Juliane F Stahl, Pia S Schober, and C Katharina Spieß (2015). "Elternzufriedenheit mit Kitas: Höhere Qualität, höhere Zufriedenheit". In: *DIW Wochenbericht* 82.50, pp. 1185–1192.

Carbuccia, Laudine, Sciences Po Paris, and Arthur Heim (2025). "Investigating how administrative burden and information costs affect social inequalities in early childcare access: evidence from a randomised controlled trial". In: *OSF Preprint* April. DOI: 10.31219/osf.io/w2ey7. URL: https://osf.io/preprints/osf/w2ey7_v5.

Cascio, Elizabeth U (2013). "The Impacts of Expanding Access to High-Quality Preschool Education". In: *Brookings Papers on Economic Activity* 2013.2, pp. 127–192. DOI: 10.1353/eca.2013.0012.

Deindl, Raphael (2024). "Familialism reloaded? On the transformation of welfare state and family in Austria". In: *Osterreichische Zeitschrift fur Soziologie* 49.3, pp. 357–376. ISSN: 18622585. DOI: 10.1007/s11614-024-00576-0.

Dreyer, Rahel, Peter Cloos, Nina Hogrebe, and Ina Kaul (2025). *Memorandum zur Frühkindlichen Bildung, Betreuung und Erziehung in Deutschland (Kurz- und Langfassung)*. Tech. rep.

Gambaro, Ludovica, Clara Schäper, and C. Katharina Spiess (2024). "Crowded-out? Changes in informal childcare during the expansion of formal services in Germany". In: *Social Policy and Administration* February, pp. 1–16. ISSN: 14679515. DOI: 10.1111/spol.13067.

Gambaro, Ludovica, C Katharina Spieß, and Katharina Wrohlich (2016). "Nachmittagsbetreuung und Müttererwerbstätigkeit: Institutionalisierte Betreuung erhöht Beschäftigungswahrscheinlichkeit und Arbeitszeit". In: *DIW Wochenbericht* 83.47, pp. 1123–1131.

Geier, Boris and Birgit Riedel (2009). "Ungleichheiten der Inanspruchnahme öffentlicher frühpädagogischer Angebote. Einflussfaktoren und Restriktionen elterlicher Betreuungsentscheidungen". In: *Frühpädagogische Förderung in Institutionen*. Ed. by Hans-Günther Roßbach and Hans-Peter Blossfeld. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 11–28. ISBN: 978-3-531-16291-1. DOI: \url{10.1007/978-3-531-91452-7{\textunderscore}2}.

Ghirardi, Gaia and Filippo Gioachin (2025). "Center-based formal childcare closes gaps, others may widen them: How childcare arrangements shape early educational inequality". In: *Acta Sociologica*, pp. 1–13. DOI: 10.1177/00016993251389750.

Grgic, Mariana and Tina Friederich (2023). "Innovative ways of handling staff shortages in ECEC? Increasing flexibility in the regulatory frameworks for non-traditional and non-German

professional staff qualifications". In: *International Journal of Child Care and Education Policy* 17.1. ISSN: 22886729. DOI: 10.1186/s40723-023-00126-3. URL: https://doi.org/10.1186/s40723-023-00126-3.

Hausladen, Kristina, Katrin Wolf, Lars Burghardt, and Yvonne Anders (2026). "The relationship between the child-staff ratio and process quality in early childhood education institutions: A meta-analysis". In: *Early Childhood Research Quarterly* 76, pp. 556–567. DOI: 10.1016/j.ecresq.2026.05.002. URL: http://hausladen_2026_child_staff_ratio_process_quality_meta_.pdf.

Hermes, Henning, Marina Krauß, Philipp Lergetporer, Frauke Peter, and Simon Wiederhold (2023). "Early Child Care, Maternal Labor Supply, and Gender Equality: A Randomized Controlled Trial". In: *SSRN Electronic Journal* December. ISSN: 1556-5068. DOI: 10.2139/ssrn.4316005. URL: https://www.ssrn.com/abstract=4316005.

Hermes, Henning, Philipp Lergetporer, Fabian Mierisch, Guido Schwerdt, and Simon Wiederhold (2024). "Information about inequality in early child care reduces polarization in policy preferences". In: *Journal of Economic Behavior & Organization* 228, p. 106780. DOI: 10.1016/j.jebo.2024.106780.

Hiekel, Nicole and Mine Kühn (2024). "Lessons from the pandemic : Gender inequality in childcare and the emergence of a gender mental health gap among parents in Germany N icole Hiekel Mine Kühn". In: *DEMOGRAPHIC RESEARCH* 51.July. DOI: 10.4054/DemRes.2024.51.3.

Huebener, Mathias, Sevrin Waights, and C. Katharina Spiess (2024). "Well-Being Throughout the COVID-19 Pandemic in Germany: Gendered Effects of Daycare and School Closures". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4788221.

Hummel, Theresia G, Franziska Cohen, and Yvonne Anders (2022a). "Parents' Trust in Their Child's Preschool: Associations with Child and Family Characteristics and Aspects of Parent-Preschool Communication". In: *Early Education and Development* 34.6, pp. 1267–1284. DOI: 10.1080/10409289.2022.2101344.

— (2022b). "The role of partnership practices in strengthening parental trust". In: *Early Child Development and Care* 193.4, pp. 571–588. DOI: 10.1080/03004430.2022.2093868.

Hummel, Theresia Gabriele, Franziska Cohen, Juliane Gessulat, and Yvonne Anders (2022c). "Measuring interaction quality between parents and professionals and its relation to preschool characteristics". In: *International Journal of Educational Research Open* 3.April. ISSN: 26663740. DOI: 10.1016/j.ijedro.2022.100195.

Humphries, John Eric, Fabian Kosse, Lukas Pape, Sophia Schmitz, and Simon Wiederhold (2024). "Parents' Earnings and the Returns to Universal Pre-K". In: *NBER Working Paper* 33042. DOI: 10.3386/w33042.

Igel, Corinne and Marc Szydlik (2011). "Grandchild care and welfare state arrangements in Europe". In: *Journal of European Social Policy* 21.3, pp. 210–224. ISSN: 0958-9287. DOI: \url{10.1177/0958928711401766}.

Ito, Tsubasa, Michio Naoi, Kazuto Sumita, and Qing Ye (2024). "The Impact of Daycare Closures Owing to COVID-19 on Parental Stress: The Case of Japan". In: *Journal of Family and Economic Issues* 45.3, pp. 579–595. ISSN: 15733475. DOI: 10.1007/s10834-023-09929-5.

Jessen, Jonas, Katharina Spieß, Sevrin Waights, and Andrew Judy (2020). "Gründe für unterschiedliche Kita-Nutzung von Kindern unter drei Jahren sind vielfältig". In: *DIW Wochenbericht* 87.14, pp. 267–275. ISSN: 0012-1304.

Kayed, Theresia, Johannes Wieschke, and Susanne Kuger (2025). *Elterlicher Bedarf und Ungleichheiten im Zugang*. Tech. rep. München. DOI: 10.36189/DJI202530. URL: https://www.dji.de/KiBS.

Mayer, Roger C, James H Davis, and F David Schoorman (1995). "An Integrative Model of Organizational Trust". In: *Academy of Management Review* 20.3, pp. 709–734. DOI: 10.5465/amr.1995.9508080335.

Meiner-Teubner, Christiane, Melanie Böwing-Schmalenbrock, Ninja Olszenka, and Thomas Rauschenbach (2024). "The demand for professionals in early childhood education and care – a forecast up to the year 2035". In: *Zeitschrift fur Erziehungswissenschaft* October. ISSN: 18625215. DOI: 10.1007/s11618-024-01279-8.

Morrissey, Taryn W (Mar. 2017). "Child care and parent labor force participation: a review of the research literature". In: *Review of Economics of the Household* 15.1, pp. 1–24.

Müller, Kai Uwe and Katharina Wrohlich (2020). "Does subsidized care for toddlers increase maternal labor supply? Evidence from a large-scale expansion of early childcare". In: *Labour Economics* 62.July 2018. ISSN: 09275371. DOI: 10.1016/j.labeco.2019.101776.

Nentwig-Gesemann, Iris and Adeline Hurmaci (2020). *KiTa-Qualität aus der Perspektive von Eltern*. Tech. rep. Gütersloh: Bertelsmann Stiftung. DOI: 10.11586/2020061.

Neuberger, Franz, Mariana Grgic, Kerstin Lippert, and Susanne Kuger (2026). "Wenn es erst an Personal und dann an Vertrauen mangelt: Wahrgenommene Qualität, Schließungserfahrungen und der Wandel der Nicht-Nutzungsgründe in der Kindertagesbetreuung". In: *DJI-Preprint*. DOI: 10.36189/DJI2026042. URL: https://www.dji.de/fileadmin/user_upload/dasdji/publikationen/Broschueren_2026/DJI_Preprint_1-2026_Qualitaetskrise.pdf.

Neuberger, Franz, Mariana Grgic, Johannes Wieschke, and Susanne Kuger (2025a). "Seeing the Forest because of the Trees: New Perspectives on Daycare Center Closures, Maternal and Child Well-being and Early Childhood Education Quality through Machine Learning". URL: https://doi.org/10.31235/osf.io/cxtgm_v1.

— (2025b). "(Spät-)Folgen der Strukturprobleme in den Kitas – Kitaschließungen und ihre Auswirkung auf Kinder". In: *Diskurs Kindheits- und Jugendforschung/Discourse. Journal of Childhood and Adolescence Research* 2, pp. 1–56.

Neuberger, Franz, Tobias Rüttenauer, and Martin Bujard (2022). "Where does public childcare boost female labor force participation? Exploring geographical heterogeneity across Germany 2007–2017". In: *Demographic Research* 46.April, pp. 693–722. ISSN: 14359871. DOI: 10.4054/DEMRES.2022.46.24.

Oppermann, Elisa, Theresia Gabriele Hummel, and Yvonne Anders (2025). "Preventing emotional exhaustion among early childhood professionals through a supportive organizational climate". In: *Zeitschrift fur Erziehungswissenschaft*. ISSN: 18625215. DOI: 10.1007/s11618-025-01353-9.

Philipp, Marie-Fleur, Silke Büchau, Pia S Schober, Viktoria Werner, and C Katharina Spiess (Mar. 2025). "How perceived daycare quality shapes norms around daycare use and parental employment: experimental evidence from Germany". In: *European Sociological Review*. ISSN:

0266-7215. DOI: 10.1093/esr/jcaf009. URL: https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcaf009/8069435.

Pianta, R. C., K. M. La Paro, and B. K. Hamre (2008). *Classroom Assessment Scoring System: Manual K-3*. Baltimore: Paul H. Brookes Publishing Co.

Schieler, Andy and Daniela Menzel (2024). "Kitas 2. Klasse? Mehrfachbelastungen von Kitas mit Kindern aus sozioökonomisch benachteiligten Familien". In: *FES diskurs*.

Schmutz, Rita (2024). "Is universal early childhood education and care an equalizer? A systematic review and meta-analysis of evidence". In: *Research in Social Stratification and Mobility* 89.October, p. 100859. ISSN: 02765624. DOI: 10.1016/j.rssm.2023.100859. URL: https://doi.org/10.1016/j.rssm.2023.100859.

Schober, Pia S and Juliane F Stahl (2016). "Expansion of Full-Day Childcare and Subjective Well-Being of Mothers: Interdependencies with Culture and Resources". In: *European Sociological Review* 32.5, pp. 593–606. ISSN: 0266-7215. DOI: \url{10.1093/esr/jcw006}.

Schreiber, Norbert (2004). "Qualität von was? Qualität wozu? Zur Perspektivität von Eltern- und Erzieherinnenurteilen". In: *Zeitschrift für Pädagogik* 50.4, pp. 555–579. URL: https://www.pedocs.de/volltexte/2012/5605/.

Schunck, Reinhard (2013). "Within and between estimates in random-effects-models: Advantages and drawbacks of correlated random effects and hybrid models". In: *The Stata Journal* 1, pp. 65–76.

Slot, Pauline (2018). "Structural characteristics and process quality in early childhood education and care: A literature review". In: *OECD Education Working Paper* 176, p. 65. URL: http://www.oecd.org/education/school/earlychildhoodeducationandcare.htm..

Stahl, Juliane (2015). "Wer nutzt welche Qualität? Zusammenhänge zwischen sozioökonomischer Herkunft und Kita-Qualität". In: *DIW Roundup* 73.

Stahl, Juliane F. and Pia S. Schober (Feb. 2020). "Early education and care quality: Does it matter for maternal working hours?" In: *Social Science Research* 86, p. 102378. ISSN: 0049089X. DOI: 10.1016/j.ssresearch.2019.102378. URL: https://linkinghub.elsevier.com/retrieve/pii/S0049089X18309244.

Steinberg, Hannah Sinja (2025). *Soziale Ungleichheiten in der frühkindlichen Betreuung: Eine Analyse von Betreuungsentscheidungen und -verläufen*. Wiesbaden: Springer VS. DOI: 10.1007/978-3-658-48120-9.

Ulferts, Hannah, Katrin M. Wolf, and Yvonne Anders (2019). "Impact of Process Quality in Early Childhood Education and Care on Academic Outcomes: Longitudinal Meta-Analysis". In: *Child Development* 90.5, pp. 1474–1489. ISSN: 0009-3920. DOI: 10.1111/cdev.13296.

Weimann-Sandig, Nina and Bernhard Kalicki (2024). "Nur Teilzeit in der Kita? Arbeitszeitumfang und Beschäftigungspotentiale in der Kindertagesbetreuung". In: *Working Paper Forschungsförderung* 331, pp. 1–17.

Zimmert, Franziska (2019). "Early child care and maternal employment: empirical evidence from Germany". In: *IAB-DISCUSSION PAPER*.

Zoch, Gundula and Irina Hondralis (Sept. 2017). "The Expansion of Low-Cost, State-Subsidized Childcare Availability and Mothers' Return-to-Work Behaviour in East and West Germany". In: *European Sociological Review* 33.5, pp. 693–707. DOI: 10.1093/esr/jcx068. URL: https://doi.org/10.1093/esr/jcx068.

A Screening Table: Stepwise Coefficients

Table 3: Screening: non-use reasons and outcomes after daycare entry

	Extensive Margin (OR)		Intensive Margin (β)	
Non-use reason	Daycare entry	Employment	Daycare hours	Work hours
No slot available	2.29***	1.45**	+2.20***	+3.03***
Quality concerns	0.40***	0.48***	-3.23***	-3.08**
No offer nearby	1.46**	1.35*	+1.91**	+1.53
Cost	0.72*	0.74*	-1.21†	-0.43
Hours don't fit	0.69†	0.88	-1.16	+2.98†
Child too young	0.80	1.07	-2.42***	-1.02
Want to raise myself	0.28***	0.43***	-2.96***	-3.02***
Good experience at home	0.39***	0.51***	-2.68***	-2.72***
Grandparents available	0.46***	0.77**	-3.56***	-2.81***
Not an option	0.33***	0.56***	-2.51***	-4.24***

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Risk set (logits): $N = 2,750$ PY, 2,425 mothers; entry-conditional samples: D-Hours $N = 1,606$, W-Hours $N = 1,458$.

B Subsample Comparison: Under-3 Entry vs. Age-3+ Entry

Background: who is in the Stage 1 risk set?

The Stage 1 risk set is built from mother-years in which the mother is observed both as a daycare non-user and as not employed in t_0. Two structural features of the German ECEC system shape who ends up in this pool. First, the daycare attendance rate is strongly age-graded: from approximately 30% at age 1 to 60% at age 2 and over 92% at age 3 (BMBFSFJ, 2025a, KitaKompakt 2025). By age 3, the residual non-user pool is therefore small, and a non-using mother of a kindergarten-aged child is by construction a selective late-entry case. Second, statutory parental leave (*Elternzeit*, henceforth EZ) is the typical institutional bridge between birth and daycare entry; mothers on EZ are the population at risk of the daycare-entry-and-employment-return transition that Stage 1 models. Both features narrow the substantive interpretation of the Stage 1 sample.

Sample composition by parental-leave status

We classify mothers into four parental-leave groups: A = employed not on leave, B = employed on leave (Group B), C = not employed not on leave, D = not employed on leave (Group D). The joint non-user / non-employed restriction removes Group A by construction (employed in t_0). Empirically, the surviving sample is dominated by Group D: of the N = 2,750 Stage 1 person-years, approximately 88% are in Group D (not employed and currently on parental leave), about 11% are in Group C (long-term not employed, not currently on leave), and Group B contributes a residual 0.9% (employed-with-contract on leave; this group's hours-based employment indicator can register as zero during parental leave and therefore enters the Stage 1 pool, but only marginally). Substantively, Stage 1 should therefore be read primarily as a model of *the EZ-exit transition*: out of parental leave, into daycare use and (re-)employment. This is the same population margin that recent reform-evaluation work targets: Boll and Nikolka (2026) use a

difference-in-differences and regression-discontinuity design to show that the ElterngeldPlus reform shifted leave-taking, labor supply re-entry, and childcare arrangements precisely along this EZ-exit transition. The barrier-elimination counterfactual we estimate is therefore positioned on the same behavioural margin that policy is empirically known to be able to move.

Under-3 Entry vs. Age-3+ Entry comparison

Within the Stage 1 risk set, the two age-of-entry subsamples (Under-3 Entry: child aged 0–1 in t_0; Age-3+ Entry: child aged ≥ 2 in t_0) differ systematically. Table 4 reports descriptive cells for the overall risk set and for both subsamples in parallel, with p-values testing the Under-3 Entry vs. Age-3+ Entry contrast. "Migration background" is defined as the child or both parents being born abroad.

Table 4: Sample composition: Stage 1 risk set (non-user and non-employed in t_0), by entry-age subsample

Variable	Overall N=2,750 PY 2,425 mothers	Under-3 Entry N=2,205 PY 2,003 mothers	Age-3+ Entry N=545 PY 495 mothers	p
Outcomes				
Daycare entry rate (%)	58.4	55.6	69.9	0.000***
Employment entry rate (%)	53.1	58.2	32.4	0.000***
Sociodemographics				
Child age (months)	16.06 (11.36)	11.39 (5.11)	34.96 (9.99)	0.000***
Mother has Abitur (%)	76.4	78.1	69.4	0.000***
Income high tercile (%)	24.1	26.3	15.4	0.000***
Income low tercile (%)	36.4	32.8	51.1	0.000***
East Germany (%)	14.7	16.0	9.5	0.000***
Number of children	1.83 (0.95)	1.69 (0.81)	2.38 (1.22)	0.000***
Risk groups				
Migration background (%)	5.3	4.2	9.8	0.000***
Welfare receipt (%)	8.2	6.6	14.5	0.000***
Parental leave				
On parental leave, employed (Group B, %)	0.9	1.0	0.6	0.287
On parental leave, not employed (Group D, %)	88.0	92.3	70.5	0.000***
Not on parental leave, not employed (Group C, %)	11.1	6.6	29.0	0.000***
Supply context				
Regional care gap (PP)	14.10 (4.77)	14.56 (4.55)	12.21 (5.16)	0.000***

Note: Stage 1 risk set: mothers neither using daycare nor employed in t_0, split into Under-3 Entry (child aged 0–1 in t_0, entry at age ≤ 2) and Age-3+ Entry (child aged ≥ 2 in t_0, entry at age ≥ 3). Cells: percent for binary indicators, mean (SD) for continuous variables. p-values test Under-3 Entry vs. Age-3+ Entry (two-sample χ^2 for proportions, two-sample t-test for continuous). Migration background = child and/or both parents born abroad. Parental-leave groups (sum to 100%): Group B = employed and on parental leave; Group C = not employed and not on parental leave; Group D = not employed and on parental leave. Group A (employed not on leave) is 0 by construction (the risk set conditions on being not employed at t_0). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$. KiBS Long Panel 2021–2024.

Reading the table

The Age-3+ Entry subsample is systematically the more disadvantaged of the two on every dimension that prior work has linked to delayed daycare entry. (1) **Sample sizes are highly unbalanced:** Under-3 Entry contributes N = 2,205 person-years on 2,003 mothers, Age-3+ Entry only N = 545 person-years on 495 mothers. The ratio of about four-to-one mirrors the institutional reality that most daycare entries occur during the under-3 window. (2) **Migration background is over twice as common in Age-3+ Entry** (9.8% vs. 4.2%, $p < .001$), and welfare receipt is similarly more than doubled (14.5% vs. 6.6%, $p < .001$). (3) **Income polarises**

along the same axis: the high-income share (top tercile) drops from 26.3% in Under-3 Entry to 15.4% in Age-3+ Entry, while the low-income share rises from 32.8% to 51.1%. (4) **Educational and regional composition differ:** the Under-3 Entry pool is more highly educated (Abitur 78.1% vs. 69.4%) and more East-German (16.0% vs. 9.5%); the Age-3+ Entry pool has on average more children per mother (2.38 vs. 1.69). (5) **EZ-status differs categorically.** The Under-3 Entry pool is saturated by Group D (92.3% on parental leave without contract), as expected for mothers of children under 2; in Age-3+ Entry, Group D coverage drops to 70.5% (statutory parental leave is normally exhausted by age 3), and the residual is mostly Group C (long-term non-employed without active leave). (6) **Outcome rates flip by age:** the daycare-entry rate is higher in Age-3+ Entry (69.9% vs. 55.6%, the eventual catch-up at age 3) but the employment-entry rate is markedly *lower* (32.4% vs. 58.2%), consistent with the reading that mothers who have not entered employment for three years past birth are structurally further from a labour-market return than mothers earlier in the EZ window. Age-3+ Entry is therefore not just smaller but *a categorically different population*: persistently disadvantaged late-entry mothers whose labour-market re-attachment is weaker. We retain the subsample for completeness, but the substantive Stage 1 narrative rests on Under-3 Entry.

Implications for inference

We retain the full Stage 1 sample including all parental-leave groups, controlling for status via two dummies (employed-while-on-leave and not-employed-while-on-leave; reference: Group C). Group B is too small ($N = 26$) to influence point estimates, so its pooling with the rest is innocuous. The substantive caveat is that Stage 1 estimates correspond to the EZ-exit margin rather than to a generic non-user-to-user transition; we revisit this in the Limitations (Section 7.1). The Stage 3 within-person quality analysis (Section 6.4) operates on a different sample (continuing kindergarten users) and is therefore not affected by these EZ-dynamics.

C Robustness: Mundlak-Hybrid Technical Notes

As a robustness check for the post-selection results (Section 6), we estimate Mundlak-Hybrid panel models on the kindergarten sample, restricted to 2021–2023 because PTI and CA were fielded only in 2023 and therefore contribute no within-variation to a 2023–2024 specification. Daycare closures are excluded as controls because closure data is only available from 2023 onward. The sample comprises observations from mothers with ≥ 2 panel waves. For each time-varying quality predictor x_{kit}, the person-mean $\bar{x}_{ki}$ is included simultaneously; the coefficient of x_{kit} then corresponds to the within-estimator $\hat{\beta}_W$, while $\bar{x}_{ki}$ absorbs between-variation (Schunck, 2013). Four outcomes: daycare hours (I2b, lmer), work hours (I1b, lmer), grandparent hours (I3b, lmer), and employment entry (E1b, glmer logit, restricted to mothers not employed at panel start).

Key within-effects (summary):

- **I2b Daycare hours:** Child interaction experience $\beta_W = +0.35^{***}$ and accessibility $\beta_W = +0.15^{*}$ increase daycare hours. No other dimension significant within.
- **I1b Work hours:** No dimension reaches $\alpha = .05$ within; structural quality has a negative point-estimate that does not cross conventional thresholds ($\beta_W = -0.28$, $p = .051$).
- **I3b Grandparent hours:** No significant within-effects.

- **E1b Employment entry:** No significant within-effects (low power; $N = 2{,}258$ person-years, 1,003 mothers not employed at panel start).

Key between-effects (summary):

- **I2b Daycare hours:** Child interaction experience $\beta_B = +1.15^{***}$; accessibility $\beta_B = -0.82^{***}$; parent–teacher interaction $\beta_B = -0.40^{*}$.
- **I1b Work hours:** Accessibility $\beta_B = -0.95^{***}$; other dimensions n.s. (PTI and CA have point-estimates near $p = .05$ but do not cross conventional thresholds).
- **I3b Grandparent hours:** Accessibility $\beta_B = -0.29^{**}$; child interaction experience $\beta_B = -0.42^{**}$.

Between/Within ratios for dimensions with significant within-effects (daycare hours only): $|\beta_B/\beta_W| = 3.4$ (CIE, $|1.16/0.35|$) and 5.3 (accessibility, $|-0.82/0.15|$). For work hours, the ratio is not meaningfully computable because within-effects are not distinguishable from zero at $\alpha = .05$.

Accessibility flip: Accessibility shows a *negative* between-effect ($\beta_B = -0.82^{***}$ for daycare hours) but a *positive* within-effect ($\beta_W = +0.15^{*}$). This mirrors the post-selection coefficient (negative) and reveals it as between-contaminated: mothers who rate accessibility highly tend to book fewer hours (selection into convenience), but when accessibility *improves* for the same mother, she books more hours (reliability mechanism).

D Quality Selection: Stepwise vs. Joint Model

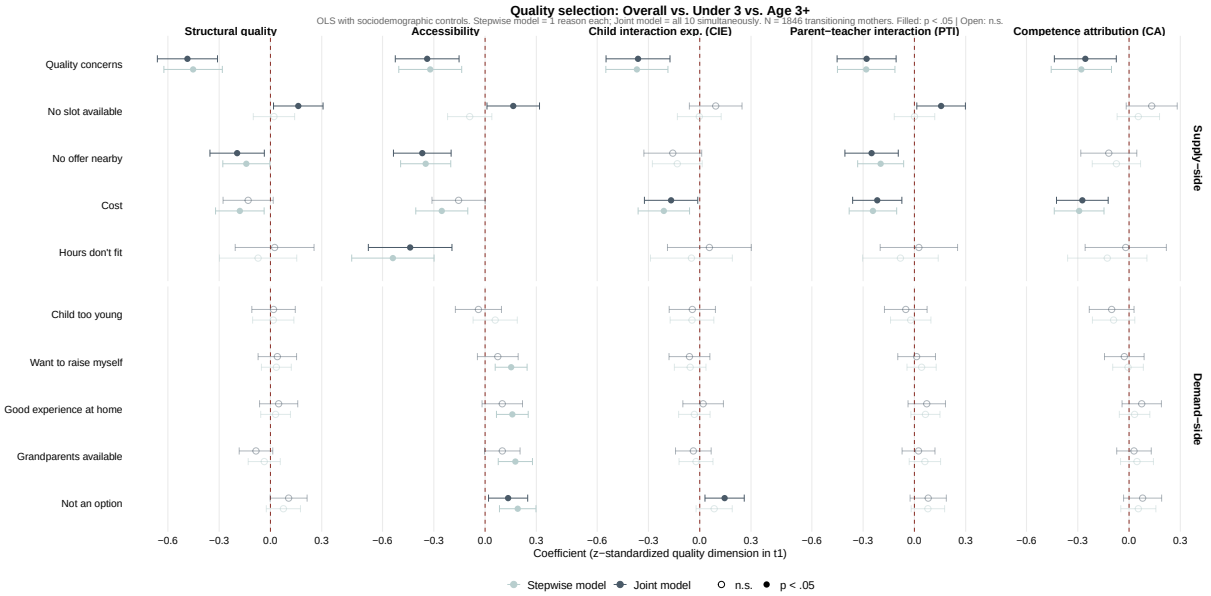


Figure 7: Quality selection (overall): stepwise (light blue) vs. joint model (dark blue). KiBS 2021–2024.