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.

When R&D Taxonomies Travel: Target-Population Validity in the Selection of Young Technology-Oriented SMEs

hana kim

hanakim@kaist.ac.kr

Korea Advanced Institute of Science and Technology

Research Article

Keywords: technology transfer, R&D taxonomy, target-population validity, young SMEs, screening architecture, selection validity

Posted Date: July 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10434455/v1

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

Additional Declarations: No competing interests reported.

When R&D Taxonomies Travel: Target-Population Validity in the Selection of Young Technology-Oriented SMEs

Abstract

Technology-transfer and innovation-support systems often use sector taxonomies to reduce the number of firms requiring costly, detailed assessment. This study argues that a taxonomy can be descriptively valid yet selection-invalid when the population used to construct it differs from the population it is used to screen. It develops the concept of target-population validity and tests it with a proprietary Korean bank panel. Sector screens are fixed with 2012–2015 information and evaluated on later outcomes for 116,193 young SMEs in 54 industries. Three near-capacity-matched rules are compared: the OECD whole-industry R&D taxonomy, an R&D screen recalibrated on the SME population, and an SME multi-signal screen adding registered-patent and recognised-intangible indicators. The main downstream criterion combines 2016–2019 sales growth of at least 10% with a BBB-or-better rating. After controlling for age, size, prior growth, profitability, leverage, liquidity, debt, legal form, prior technological activity, macro-sector, and timing, the SME R&D and multi-signal screens are associated with 2.23 and 2.39 percentage-point higher joint-success probabilities; the OECD screen's adjusted difference is 0.85 points. Industry-clustered, wild-bootstrap, logit, attrition-weighted, placebo, and threshold analyses support the pattern. Recalibrating the same R&D construct on the target SME population explains nearly all practical improvement; adding patent and intangible signals contributes little at the margin. A dual-route architecture raises recall, but only when the value of avoiding false negatives exceeds added evaluation costs. The findings distinguish descriptive, target-population, and selection validity in upstream technology-transfer screening.

Keywords: technology transfer; R&D taxonomy; target-population validity; young SMEs; screening architecture; selection validity

JEL Classification: C52; L25; L26; O31; O32; O38

1. Introduction

Technology transfer and innovation support begin before detailed due diligence. Universities, public agencies, policy banks, development organisations, and investors must first decide which projects or firms deserve further attention. Because comprehensive assessment is costly, initial candidate generation commonly relies on coarse signals: technology domains, industry labels, programme categories, or R&D-based definitions of high technology. Such signals are useful because they compress a large search space, and consequential because candidates excluded at this stage may never reach the technical, market, financial, or sustainability assessments that follow.

Recent work in *The Journal of Technology Transfer* illustrates the breadth of this upstream selection problem. McCarthy et al. (2026) develop a rapid tool for evaluating early technology-transfer projects when information is sparse and detailed assessment is premature. Micol et al. (2026) identify deep-tech start-ups at scale to analyse the regional ecosystem factors associated with their presence across Europe. Gafrej et al. (2026) use interpretable machine learning to predict post-campaign states in entrepreneurial finance. Foss (2026) argues that evaluation architectures do not merely reveal quality; aggregation rules and institutional constraints shape the signals and allocations they produce. Across these studies, a shared theme is that technology-transfer outcomes depend partly on how candidates are initially represented and filtered.

Sector taxonomies are one such filter; the OECD classification based on business R&D intensity, for example, is transparent, internationally comparable, and substantively valuable (Galindo-Rueda & Verger, 2016). It is designed to describe average formal R&D effort across economic activities. Yet the population embedded in that ordering is not necessarily the population facing a particular selection decision. Whole-industry R&D aggregates can be influenced by large incumbents, whereas many support, finance, and commercialisation systems seek operating SMEs or young firms. Applying a whole-industry taxonomy to a young-SME target population therefore involves an untested transport: the measure is carried from the population in which the sector ordering was constructed to a different population in which selection occurs.

This study develops target-population validity to analyse that transport. The argument builds on use-oriented theories of measurement validity. A score is not validated in the abstract; validity concerns the interpretation and use attached to it (Messick, 1995; Kane, 2013). A taxonomy can accordingly possess descriptive validity for its original construct and reference population while failing to preserve a decision-relevant ordering in another population. This problem is distinct from ordinary construct undercoverage. Even if formal R&D is the correct construct, the same construct may generate a different sector ordering when measured among the firms that a programme actually seeks to evaluate. The central theoretical distinction is therefore among: (1) descriptive validity, whether a taxonomy measures its stated construct in its reference population; (2) target-population validity, whether the ordering remains relevant after

transport to the population being screened; and (3) selection validity, whether the resulting candidate set performs well against the downstream objective (Fig. 1; Table 1).

The empirical setting is a proprietary panel from the Industrial Bank of Korea (IBK). The analysis focuses on firms younger than seven years in the 2012–2015 early window and evaluates their 2016–2019 outcomes. The early population contains 116,193 young operating SMEs across 54 two-digit industries; 63,640 have the information required for the main later outcome. Three sector screens are constructed from this panel. The first is the OECD whole-industry R&D rule. The second retains the R&D construct but recalibrates it using early-window R&D intensity in the observed SME population. The third expands the SME measure with a reported registered-patent stock and recognised intangible assets. The SME screens are chosen to produce candidate pools close to the OECD rule's capacity, so performance differences cannot be attributed simply to examining more firms.

The main outcome is deliberately downstream and non-mechanical: a firm must achieve annualised sales growth of at least 10% in 2016–2019 and an average rating of BBB or better. The two components represent upside and a minimum viability condition. The study also examines high growth alone, the rating threshold alone, a stricter 20% growth criterion, later R&D entry among firms without early R&D, and employment growth. These outcomes do not enter construction of the sector screens.

First, at nearly equal candidate capacity, both SME-calibrated screens select more firms that later meet the joint growth-viability criterion. The OECD screen selects 29,358 early candidates and later contains 4,261 joint-success firms. The SME R&D screen selects 29,769 candidates and contains 4,535 successes; the multi-signal screen selects 28,928 and contains 4,457. Their raw precision rates are 24.98%, 26.37%, and 26.41%, respectively.

Second, the pattern survives extensive adjustment: the main linear probability models control for age and age squared, assets, sales, employment, prior sales growth, profitability, leverage, liquidity, borrowings, legal form, observation history, prior R&D and patent participation, firm R&D intensity, recognised intangibles, macro-sector, and early-window timing. Standard errors are clustered by the 54 industries because screen status is assigned at the industry level, and Webb wild-cluster inference is reported. In the technology-adjusted models, the SME R&D and multi-signal screens are associated with 2.23 and 2.39 percentage-point higher joint-success probabilities, while the OECD coefficient is 0.85 points and statistically imprecise. Pairwise horse-race models show that the SME screens outperform the OECD screen on the variation where they disagree; the difference between the two SME screens is not distinguishable from zero.

Third, the gain reflects population alignment rather than indicator proliferation. Moving from the OECD whole-industry ordering to the SME R&D ordering adds 274 later joint successes and increases precision by 1.39 percentage points. Moving from SME R&D to the multi-signal screen changes the candidate composition but produces almost no precision gain. The same pattern appears in direct technology-related evidence: both SME screens are associated with an approximately 6.2 percentage-point higher probability of later R&D entry, but the multi-signal extension does not materially outperform SME R&D. Thus, the major selection gain comes from estimating the familiar R&D construct in the population to which it will be applied.

This study makes three contributions. First, it extends technology-transfer evaluation research by introducing target-population validity as a separate requirement between construct measurement and downstream selection. Second, it provides an empirical decomposition of two frequently conflated remedies: population alignment and construct expansion. The evidence shows that a richer indicator set is not necessarily the main solution when the deeper problem is that a valid measure was calibrated on the wrong population. Third, it connects measurement validity to selection architecture. Drawing on Sah and Stiglitz (1986, 1988), the analysis shows that a dual-route rule increases recall but also expands evaluation cost; plural screening is efficient only under identifiable cost conditions.

The remainder of this study is organised as follows. Section 2 develops the theoretical framework and hypotheses. Section 3 describes the data, screens, outcomes, controls, and inferential design. Section 4 reports the results. Section 5 discusses the theoretical and practical implications, limitations, and research agenda. Section 6 concludes.

2. Theory and hypotheses

2.1. Upstream selection in technology-transfer systems

Technology transfer is commonly represented as a sequence of stages: opportunity recognition, disclosure or search, preliminary assessment, resource mobilisation, development, transfer, and commercialisation (Bozeman, 2000; Vohora et al., 2004; Markman et al., 2005; Phan & Siegel, 2006; Bozeman et al., 2015). The exact sequence varies across university, public-sector, financial, and corporate settings, but a common organisational problem appears early. Detailed evaluation of technical feasibility, market potential, intellectual property, team capability, financial viability, sustainability, or partner fit is costly. Institutions therefore use preliminary rules to reduce the number of candidates that receive deeper attention.

Early filters improve decision efficiency but can also create false negatives that later evaluators cannot correct, because excluded candidates never reach the next stage. McCarthy, Holland, and Shapira's (2026) early rapid sustainability assessment is designed to inform subsequent evaluation rather than replace it, illustrating the value of a low-cost screen that preserves issues for later scrutiny. Foss (2026) similarly emphasises that evaluation architectures shape outcomes through aggregation and constrained collective choice. An initial rule should therefore be judged on the downstream candidates it suppresses, not just on simplicity or face validity.

Sector taxonomies enter technology-transfer and innovation-support systems through eligibility, portfolio search, strategic targeting, benchmarking, and the identification of technology-based firms. Industry screens are attractive because sector codes are broadly available before detailed firm information is collected. They are particularly appealing for programmes whose mandates are framed around high technology or R&D. However, a sector label is an aggregate property assigned to all firms within a category. Using that label to select individual firms creates an ecological inference problem: a relationship or ordering at the industry level need not hold for the firm population relevant to the decision (Robinson, 1950; Peneder, 2010).

None of this implies that sector screening is inappropriate; without a preliminary rule, evaluation costs may be prohibitive. The relevant question is whether the screen's information base matches the population and purpose of the decision, which shifts attention from taxonomy construction alone to the validity of taxonomy use.

2.2. From descriptive validity to target-population validity

Measurement validity is use-dependent: Messick (1995) defines validity as an integrated judgment about the adequacy of interpretations and actions based on scores. Kane (2013) similarly treats validation as an argument linking observed scores to their proposed interpretation and use. These perspectives imply that evidence supporting one use does not automatically justify another. A measure can be reliable and substantively meaningful yet still be unsuitable for a transported decision.

Let a sector score depend on a construct or measurement family, a reference population, and an aggregation rule:

$$T_j = T_j(M, P_R, A),$$

where j indexes industries, M is the measured construct, P_R is the reference population used to calculate it, and A is the aggregation rule. The OECD R&D taxonomy is clear about M: business R&D intensity. Its reference evidence reflects whole-industry activity across participating economies, and its aggregation weights firms through sectoral expenditure and value added. This is a defensible descriptive system for comparing formal R&D effort.

A selection organisation, however, applies the score to a target population P_T - for example, young operating SMEs. The induced screening rule is:

$$S_i^{(R)}(b) = 1\{j(i) \in \mathcal{J}^{(R)}(b)\},$$

where $\mathcal{J}^{(R)}(b)$ is the set of industries admitted by the reference-population rule under evaluation capacity b. Target-population validity concerns whether the sector ordering remains relevant for P_T:

$$T_j(M, P_R, A) \overset{?}{\sim} T_j(M, P_T, A).$$

The two scores need not coincide. Whole-industry R&D can be dominated by large incumbents, while the distribution of formal R&D among SMEs may differ because of scale, specialisation, supply-chain position, reporting practices, or the presence of knowledge-intensive services. Population transport therefore introduces a form of dataset shift analogous to external-validity and transportability problems in causal and predictive analysis (Stuart et al., 2011; Pearl & Bareinboim, 2014). The key distinction is that the construct can remain unchanged while its sector ordering changes.

This study defines reference-population mismatch as the divergence that arises when a sector taxonomy calibrated on P_R is used to select individual firms from P_T. A simple industry-level representation is:

$$D_j = \text{rank}\{T_j(M, P_T, A)\} - \text{rank}\{T_j(M, P_R, A)\}.$$

Large values of $|D_j|$ indicate sectors whose position is sensitive to the population being represented. The selection consequence depends not only on this divergence but also on how many target firms are located in the affected sectors and how those firms perform after selection.

H1. At comparable candidate capacity, an R&D screen aligned with the young-SME target population has greater downstream selection validity than a whole-industry R&D screen.

2.3. Population alignment versus construct expansion

A common response to the limitations of R&D intensity is to add patents, intangible assets, innovation expenditure, skilled labour, digital intensity, or survey measures. This strategy addresses construct mismatch: formal R&D may

underrepresent innovation that occurs through engineering, software, design, equipment, interaction, or non-patented knowledge (Pavitt, 1984; Gallouj & Weinstein, 1997; Tether, 2005; Santamaria et al., 2009). Composite indicators may therefore improve coverage, although they also introduce weighting, transformation, and interpretation choices (OECD & Joint Research Centre, 2008; Grupp & Schubert, 2010).

Construct expansion and population alignment solve different problems. Consider three sector scores:

$$T_j^{OECD} = T_j(R\&D, P_R, A),$$

$$T_j^{SME-RD} = T_j(R\&D, P_T, A),$$

$$T_j^{SME-M} = T_j(R\&D,\ patents,\ intangibles, P_T, A).$$

The change from T_j^{OECD} to T_j^{SME-RD} isolates target-population alignment while holding the central construct fixed. The change from T_j^{SME-RD} to T_j^{SME-M} approximates construct expansion within the same target population. If the principal problem is indicator undercoverage, the multi-signal rule should materially outperform SME R&D. If the principal problem is reference-population mismatch, recalibrating R&D alone should capture most of the gain.

The distinction has both theoretical and practical stakes, because an apparent failure of R&D screening may be attributed too quickly to the narrowness of R&D as an innovation indicator when R&D is in fact the right signal measured in the wrong population. Conversely, a richer composite can add complexity without improving the candidate set if its extra dimensions do not change the relevant industry boundary.

H2. The improvement in selection validity produced by target-population alignment is larger than the incremental improvement produced by adding patent and intangible-asset signals within the target population.

2.4. Selection architecture: hierarchy, polyarchy, and evaluation cost

A single sector rule creates a hierarchy: candidates must pass one upstream gate before reaching firm-level assessment. Sah and Stiglitz (1986, 1988) show that organisational architectures differ systematically in their error profiles. Hierarchies tend to reject projects when any required stage rejects them, reducing some false positives but increasing the risk that promising projects are lost early. Polyarchies allow multiple routes to acceptance, reducing false negatives while admitting more candidates for evaluation.

In the present setting, a dual-route screen admits firms if either the established OECD rule or a target-population SME rule selects their industry:

$$S_i^{dual} = \max(S_i^{OECD}, S_i^{SME}).$$

This architecture weakly increases recall because it cannot exclude a success selected by either route. It also weakly increases candidate volume. Let $TP(S)$ be the number of later successes admitted by screen S, $N(S)$ the number of early candidates, and c the evaluation cost expressed as a fraction of the value v attached to one successful candidate. Net benefit is:

$$NB(S; c, v) = v\,TP(S) - c\,N(S).$$

A dual route is preferred to a single SME route when:

$$\frac{c}{v} < \frac{TP(S^{dual}) - TP(S^{SME})}{N(S^{dual}) - N(S^{SME})}.$$

The framework does not assume that polyarchy is always superior. Its value depends on the relative cost of deeper evaluation and the cost of a false negative. This decision-theoretic framing follows work on statistical treatment rules and welfare-sensitive classification, where a rule is evaluated by the decisions it induces rather than by fit alone (Manski, 2004; Kitagawa & Tetenov, 2018).

H3. A dual-route architecture increases recall relative to a single target-population screen, but improves net benefit only when the cost of evaluating additional candidates is below the value of the false negatives it avoids.

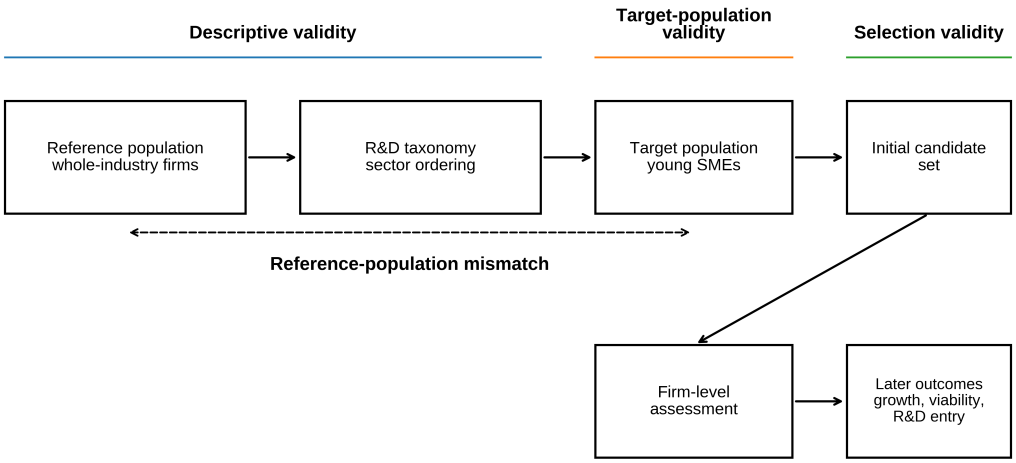


Fig. 1 Descriptive validity, target-population validity, and selection validity in an upstream screening architecture

3. Data and methods

3.1. Data source, population boundary, and timing

The empirical analysis uses a proprietary IBK financial and administrative panel covering fiscal years 2012-2019. IBK is a Korean policy-oriented bank with a large SME customer base. The source population consists of operating firms with an IBK relationship and usable financial or administrative information. It is not a census of Korean firms and does not represent pre-revenue ventures without financial statements.

The analysis separates screen construction and outcome evaluation. The early window is 2012–2015. Sector scores and all baseline firm covariates are fixed using information available in that period. The late window is 2016–2019. A firm is classified as young when its age is between zero and six years at its last retained early-window observation. Placeholder founding dates, including 1111-01-01, are treated as missing rather than as extremely old firms. Negative ages and records without a usable two-digit industry are excluded.

The resulting early population contains 116,193 young firms in 54 industries. The main joint outcome can be calculated for 63,640 firms. Later sales growth is observed for 67,693 firms, later ratings for 76,493, and later R&D entry for 77,744 firms that report no positive R&D in the early window. The industry classification is stable within firms in the underlying panel, reducing ambiguity about which sector rule applies.

Table 1 Validity concepts, failure modes, and empirical tests

Validity concept	Question	Failure mode	Empirical test in this study
Descriptive validity	Does the score represent its stated construct in its reference population?	The taxonomy does not measure what it claims to measure.	OECD R&D groups are retained as the established benchmark; no claim is made that they are incorrect descriptions of formal R&D.
Target-population validity	Does the same construct preserve a useful sector ordering in the population being screened?	Reference-population mismatch between whole-industry calibration and young-SME application.	OECD whole-industry R&D is compared with R&D intensity recalculated in the early SME population.
Construct coverage	Does a broader indicator family add decision-relevant information?	Formal R&D omits patent- or intangible-intensive activity.	SME R&D is compared with an SME multi-signal score based on R&D, registered patents, and recognised intangibles.
Selection validity	Does the rule admit candidates that later satisfy the downstream objective?	A descriptively credible screen creates firm-level false negatives.	Capacity-matched precision, recall, controlled joint outcomes, later R&D entry, and architecture net benefit are compared.

3.2. Three sector screens

3.2.1. OECD whole-industry R&D screen

The benchmark screen selects industries assigned to the OECD high or medium-high R&D-intensity groups. At the available two-digit level, 11 retained industries meet this condition. The rule includes chemicals, pharmaceuticals, electronics and communication equipment, electrical equipment, machinery, motor vehicles, other transport equipment, publishing and software, computer programming, information services, and scientific R&D. The rule is fixed; it is not re-estimated from the IBK data.

3.2.2. SME-population R&D screen

The target-population R&D score is computed from early-window IBK SMEs. For industry j, the ratio-of-sums measure is:

$$RD_j^{SME} = \frac{\sum_{i \in j, t \in 2012:2015} R\,D_{it}}{\sum_{i \in j, t \in 2012:2015} Sales_{it}}.$$

Industries are ranked by RD_j^{SME}. Starting from the highest rank, industries are added until the candidate count is closest to the OECD screen's early capacity. This produces a 12-industry screen with 29,769 young candidates, compared with 29,358 under the OECD rule. Nine industries overlap. The SME R&D rule replaces motor vehicles and other transport equipment with medical and precision instruments, architecture and engineering, and other professional scientific and technical services (Appendix A, Table A1).

3.2.3. SME multi-signal screen

The multi-signal score uses early-window R&D intensity, an audited reported registered-patent stock per employee, and recognised intangible assets relative to total assets. The patent field is nearly time-invariant within firm and is therefore treated as a stock or snapshot counted once per firm rather than as annual patent flow. Each non-negative indicator is transformed by $\ln(1 + x)$, standardised across industries, and entered into principal component analysis. The first component is oriented so that all loadings are positive and used as the common technology-activity level.

As with SME R&D, industries are added from the top of the ranking until candidate capacity is closest to the OECD rule. The resulting 12-industry screen contains 28,928 young candidates. It differs from SME R&D in only one marginal substitution: postal and telecommunications replaces other professional scientific and technical services. The two SME screens have a Jaccard overlap of 0.846 and become identical at a 13-industry threshold (Table 2).

Table 2 Screen design and analytical sample

Item	OECD whole-industry R&D	SME-population R&D	SME multi-signal
Selected industries	11	12	12
Early young-firm candidates	29,358	29,769	28,928
Firms with main later outcome	17,056	17,199	16,874
Common industries with OECD rule	-	9	9
Main population represented	Whole-industry international ordering	Bank-observed SME R&D	Bank-observed SME R&D, patent stock, and recognised intangibles
Main comparison purpose	Established reference rule	Population alignment	Population alignment plus construct expansion

3.3. Downstream outcomes and selection metrics

Late sales growth is calculated from the first and last retained sales observations in 2016–2019. For firm i observed from a_i to b_i:

$$CAGR_i = \left(\frac{Sales_{i,b_i}}{Sales_{i,a_i}}\right)^{1/(b_i - a_i)} - 1.$$

The principal downstream outcome combines a growth threshold and a minimum viability threshold:

$$Y_i^{10} = 1\{CAGR_i \geq 0.10\} \times 1\{Rating_i^{late} \geq BBB\}.$$

The rating uses the actual provider ladder. No BBB-minus category exists in the source data. The late score is the mean of available fiscal-year ratings during 2016–2019. The joint outcome is not a realised equity return, technology-transfer completion, or causal effect of support. It is a decision-relevant criterion that requires both observable upside and a minimum financial profile after the screen has been fixed.

Alternative outcomes are: (1) sales CAGR of at least 10%; (2) a BBB-or-better late rating; (3) CAGR of at least 20% combined with BBB or better; (4) later positive R&D among firms with no positive early R&D; and (5) continuous

sales and employment CAGR. The R&D-entry outcome is particularly relevant to the technology-transfer setting because it captures movement into formal knowledge-investment activity after initial screening.

For screen s, precision and recall are:

$$Precision_s = \frac{\sum_i S_{is}\, Y_i}{\sum_i S_{is}\, 1\{Y_i\ observed\}},$$

$$Recall_s = \frac{\sum_i S_{is}\, Y_i}{\sum_i Y_i}.$$

Precision measures the yield among observed selected candidates; recall measures the share of all observed later successes admitted by the screen. The analysis also reports the number of early candidates, later successes, mean late growth, rating viability, and later R&D entry.

3.4. Controls and inferential models

Raw screen performance is descriptive. Controlled models assess whether selection differences remain after adjusting for observable firm composition. The main specification is a linear probability model:

$$Y_i^{10} = \alpha + \beta S_{ij} + X_i'\gamma + \delta_{m(i)} + \tau_{e(i)} + \varepsilon_i,$$

where S_{ij} is one of the industry-level screen indicators, X_i contains firm-level controls, $\delta_{m(i)}$ are broad macro-sector fixed effects, and $\tau_{e(i)}$ are fixed effects for the firm's last early-window year.

The baseline controls are age and age squared; log assets, sales, and employment; early ROA; leverage; liquidity; borrowings relative to assets; early sales CAGR and a missing-growth indicator; corporate legal form; and the number of early observations. The technology-adjusted specification additionally controls for early R&D participation, early patent participation, firm R&D intensity, and the recognised-intangible ratio. Continuous variables are winsorised at the 1st and 99th percentiles and standardised. These controls are measured in the early period and therefore precede the outcome window.

Screen status is assigned at the industry level. Treating tens of thousands of firms as independent would understate uncertainty. Standard errors are therefore clustered by the 54 industries with finite-cluster correction (Cameron & Miller, 2015). Because the number of clusters is moderate and some screens are concentrated in relatively few sectors, Webb six-point wild-cluster bootstrap p-values are also reported (Cameron et al., 2008; MacKinnon & Webb, 2017).

The primary models estimate one screen at a time. Pairwise horse-race models include two screen indicators simultaneously:

$$Y_i^{10} = \alpha + \beta_a S_{ia} + \beta_b S_{ib} + X_i'\gamma + \delta_{m(i)} + \tau_{e(i)} + \varepsilon_i,$$

and test $H_0: \beta_b - \beta_a = 0$. Because the screens overlap substantially, this contrast is identified by the industries where their selection decisions differ. Cluster-robust and wild-cluster inference are both reported.

Robustness analyses include: cluster-robust logit average marginal effects; inverse-probability-of-censoring weights for later outcome availability (Wooldridge, 2007); leave-one-industry-out re-estimation; thresholds from the top 5 to top 20 SME industries; 30,000 random 12-industry portfolios with the nearest 1,000 candidate-count matches retained as a placebo distribution; and alternative outcomes. No model is interpreted causally, because sector-screen assignment is not random and no actual programme rule change is observed.

3.5. Selection-architecture analysis

Single-route performance is compared with dual-route rules that admit an industry if either the OECD or SME rule selects it. For a candidate-evaluation cost c and unit value of one later joint success, the empirical net-benefit calculation is:

$$\widehat{NB}_s(c) = TP_s - cN_s.$$

The break-even cost for adding the OECD route to the SME route is the incremental success count divided by the incremental candidate count. This calculation is not a social-welfare estimate; it makes the architecture trade-off transparent under a common unit-value normalisation.

4. Results

4.1. Near-equal capacity and raw selection performance

Table 3 and Fig. 2 compare the three screens at similar early candidate capacity. The OECD rule includes 29,358 young firms. SME R&D includes 29,769, 1.4% more, while the multi-signal rule includes 28,928, 1.5% fewer. The candidate

volumes are sufficiently close that selection yield can be compared without imposing a mechanical advantage from a substantially broader pool

The OECD screen contains 4,261 firms that later meet the joint growth-viability criterion. SME R&D contains 4,535, an increase of 274; the multi-signal rule contains 4,457, an increase of 196 relative to OECD. Raw precision is 24.98% under OECD, 26.37% under SME R&D, and 26.41% under the multi-signal screen. Recall is 27.91%, 29.70%, and 29.19%, respectively. Mean late CAGR is approximately 7.0% for both SME-selected groups and 6.72% under OECD.

Table 3 Near-capacity-matched raw selection performance

Screen	Early candidates	Later observed	Later joint successes	Precision (%)	Recall (%)	Mean late CAGR (%)	Later R&D-entry rate (%)
OECD whole-industry R&D	29,358	17,056	4,261	24.98	27.91	6.72	12.03
SME-population R&D	29,769	17,199	4,535	26.37	29.70	7.03	12.83
SME multi-signal	28,928	16,874	4,457	26.41	29.19	7.03	12.97

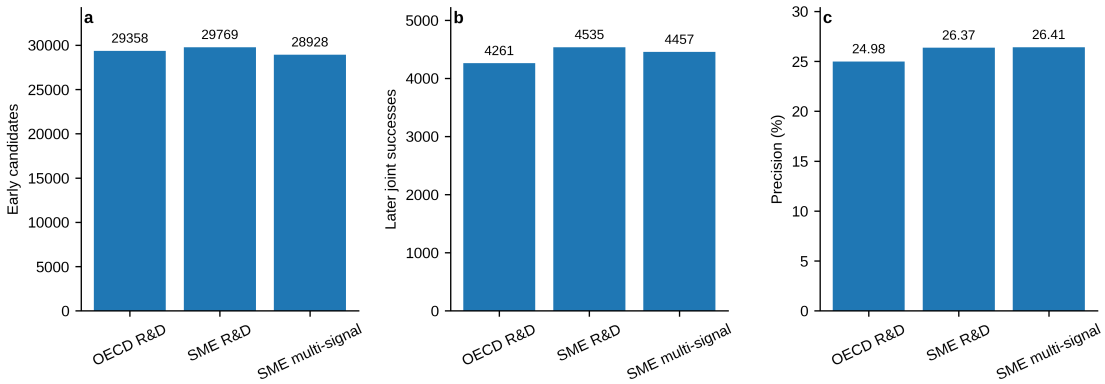


Fig. 2 Near-capacity-matched raw performance under the three sector screens. Panel a reports early candidate counts, panel b reports later joint-success counts, and panel c reports precision

The candidate-count-matched placebo reinforces the comparison. Among the 1,000 random 12-industry portfolios nearest to the multi-signal screen's candidate volume, mean precision is 23.96% and the 95th percentile is 25.57%. The actual multi-signal precision of 26.41% lies at the 99.7th percentile (Fig. 5, Appendix B). The result does not prove that the chosen screen is globally optimal, but it shows that its yield is unusual relative to similarly sized random sector portfolios.

These descriptive results are consistent with H1, but firm composition differs across selected industries. The next section introduces controls and industry-level inference.

4.2. Controlled selection validity

Table 4 reports the main adjusted estimates. Under baseline controls, the SME R&D and multi-signal coefficients are 3.57 and 3.73 percentage points, with wild-cluster p-values of 0.007 and 0.004. The OECD coefficient is 1.85 points and is less precisely estimated (wild p = 0.148)

The technology-adjusted models are the more demanding specifications because they control for prior firm-level R&D and patent participation, R&D intensity, and recognised intangibles in addition to financial and demographic covariates (Table C1 in Appendix C reports the full coefficient sets). The SME R&D coefficient is 2.23 percentage points (cluster p = 0.004; wild p = 0.019), and the multi-signal coefficient is 2.39 points (cluster p = 0.002; wild p = 0.015). The OECD coefficient falls to 0.85 points (cluster p = 0.331; wild p = 0.415). Cluster-robust logit average marginal effects are nearly identical: 2.22 points for SME R&D, 2.38 for the multi-signal screen, and 0.86 for OECD.

Table 4 Controlled association between screen inclusion and later joint success

Controls	Screen	Adjusted difference (pp)	Cluster SE	95% CI	Cluster p	Webb wild p
Baseline	OECD whole-industry R&D	1.85	1.04	[-0.20, 3.89]	0.076	0.148
Baseline	SME-population R&D	3.57	0.92	[1.77, 5.38]	<0.001	0.007
Baseline	SME multi-signal	3.73	0.93	[1.91, 5.55]	<0.001	0.004
Technology-adjusted	OECD whole-industry R&D	0.85	0.88	[-0.87, 2.57]	0.331	0.415
Technology-adjusted	SME-population R&D	2.23	0.78	[0.71, 3.75]	0.004	0.019

Controls	Screen	Adjusted difference (pp)	Cluster SE	95% CI	Cluster p	Webb wild p
Technology-adjusted	SME multi-signal	2.39	0.78	[0.85, 3.92]	0.002	0.015

Notes: The dependent variable equals one when late sales CAGR is at least 10% and the mean late rating is BBB or better. Baseline controls include age, age squared, log assets, log sales, log employment, ROA, leverage, liquidity, borrowings/assets, early sales CAGR and missingness, corporate form, number of early observations, macro-sector fixed effects, and last-early-year fixed effects. Technology-adjusted models additionally include early R&D and patent participation, firm R&D intensity, and recognised intangible assets. N = 63,640; 54 industry clusters.

Pairwise horse-race models provide a stricter test. In the technology-adjusted specification, the coefficient contrast is 6.35 percentage points for SME R&D relative to OECD (cluster p = 0.007; wild p = 0.082) and 7.43 points for the multi-signal screen relative to OECD (cluster p = 0.001; wild p = 0.053). The multi-signal minus SME R&D contrast is 4.88 points in the coefficient metric but is not statistically distinguishable under wild-cluster inference (p = 0.747). Because these models include both screen indicators, the contrast reflects the exclusive industries rather than the large common core.

The controlled results support H1. The same R&D construct becomes more selection-relevant when recalculated within the SME target population.

4.3. Mechanism decomposition: population alignment dominates construct expansion

The practical gain can be decomposed sequentially. Moving from OECD whole-industry R&D to SME-population R&D adds 411 early candidates and 274 later joint successes. Precision rises by 1.39 percentage points and recall by 1.79 points. Moving from SME R&D to the multi-signal screen reduces the candidate count by 841 and the observed success count by 78; precision changes by only 0.05 points and recall falls by 0.51 points.

The threshold analysis reaches the same conclusion. At a top-13 threshold, SME R&D and the multi-signal rule select exactly the same industry set and therefore have identical precision and recall. Across thresholds from 5 to 20 industries, their precision-recall curves are nearly superimposed. Patent and intangible information remains useful for describing technological profiles, but it does not materially improve the first-stage selection boundary in this application.

Fig. 3a displays the decomposition. Target-population alignment generates nearly all incremental performance, while construct expansion has essentially zero marginal effect on precision. This supports H2.

This runs against a common diagnosis: the weakness of a whole-industry R&D screen is less that R&D is too narrow than that its sector ordering represents a different firm population. Formal R&D remains useful, but its decision value changes when the denominator and firm composition are aligned with the population being screened.

4.4. Alternative outcomes, direct technology evidence, and robustness

The outcome decomposition in Fig. 3b clarifies what the screen differences represent. After technology-adjusted controls, none of the three screens has a statistically precise association with high sales growth alone (Table 5). The coefficients range from 0.34 to 0.97 percentage points. By contrast, the BBB-or-better rating difference is 2.23 points for OECD, 4.15 for SME R&D, and 4.33 for the multi-signal screen. The stricter joint outcome requiring 20% growth and BBB or better is also positive for the SME screens: 1.87 and 1.99 points, compared with 1.01 for OECD.

The direct technology-related outcome is later R&D entry among firms reporting no positive early R&D. The adjusted difference is 4.68 percentage points for the OECD screen, 6.18 for SME R&D, and 6.22 for the multi-signal screen. All three are statistically significant, and the SME estimates are larger. This evidence is important for JTT's domain: the target-population screens do not merely select firms with stronger ratings; they also admit more firms that subsequently enter formal R&D activity.

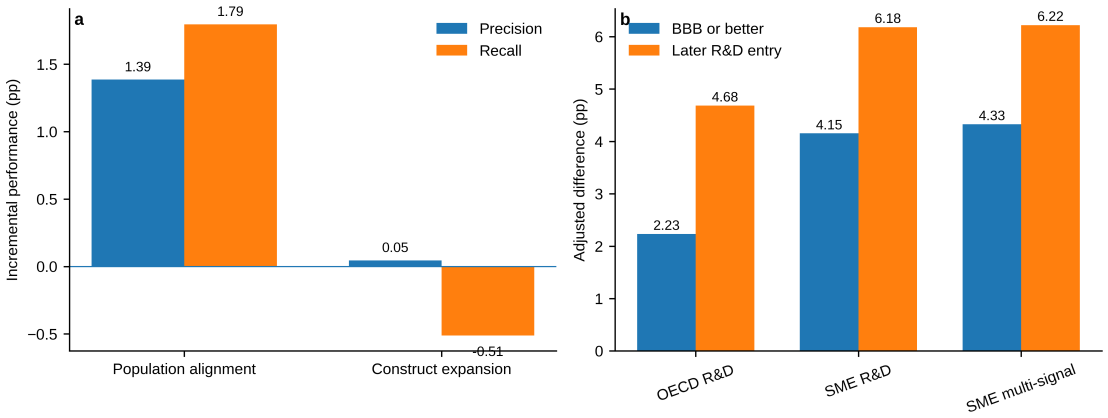


Fig. 3 Mechanism decomposition and alternative outcomes. Panel a decomposes incremental precision and recall into target-population alignment and construct expansion; panel b reports adjusted differences in BBB-or-better ratings and later R&D entry

Table 5 Alternative outcomes and robustness evidence

Test	OECD whole-industry R&D	SME-population R&D	SME multi-signal	Interpretation
High growth >=10%, adjusted difference (pp)	0.34 (p=.758)	0.97 (p=.340)	0.94 (p=.368)	No screen predicts growth alone after controls.
BBB or better, adjusted difference (pp)	2.23 (p=.047)	4.15 (p<.001)	4.33 (p<.001)	SME screens identify stronger later viability profiles.
CAGR >=20% and BBB, adjusted difference (pp)	1.01 (p=.113)	1.87 (p<.001)	1.99 (p<.001)	Results extend to a stricter joint criterion.
Later R&D entry, adjusted difference (pp)	4.68 (p=.024)	6.18 (p=.001)	6.22 (p=.002)	All screens have technology relevance; SME alignment is stronger.
Cluster-robust logit AME for main outcome (pp)	0.86 (p=.323)	2.22 (p=.003)	2.38 (p=.002)	Functional-form robustness.
IPCW technology-adjusted estimate (pp)	1.24 (p=.191)	2.59 (p=.003)	2.76 (p=.002)	Result survives correction for observed attrition.
Leave-one-industry-out coefficient range (pp)	[0.26, 1.13]	[1.87, 2.50]	[2.03, 2.67]	SME coefficients remain positive in all 54 omissions.
Candidate-matched random placebo percentile	-	-	99.7th	Multi-signal precision is unusually high among random sector sets.

Continuous outcome models are consistent with this decomposition. After controls, screen inclusion does not raise mean sales or employment CAGR. It does raise the mean late rating by 0.21 points for OECD, 0.34 for SME R&D, and 0.35 for the multi-signal rule. The study therefore does not claim that sector screening causes growth. Its defensible claim is narrower: target-population alignment improves the yield of a joint screening objective and is associated with later R&D entry and financial viability.

4.5. Single and dual-route selection architectures

A dual route admits firms selected by either the OECD or the SME rule. As predicted, recall rises. The SME multi-signal single route evaluates 28,928 young candidates and captures 4,457 later joint successes. The OECD-or-SME multi-signal route evaluates 32,610 candidates and captures 4,861 successes. Thus, 3,682 additional evaluations recover 404 additional successes, or 9.11 additional evaluations per extra success. Table 6 reports the resulting precision-recall trade-off.

Table 6 Selection architecture and cost-recall trade-off

Architecture	Industries	Early candidates	Later successes	Precision (%)	Recall (%)
OECD single route	11	29,358	4,261	24.98	27.91
SME R&D single route	12	29,769	4,535	26.37	29.70
SME multi-signal single route	12	28,928	4,457	26.41	29.19
OECD OR SME R&D	14	33,451	4,939	25.69	32.35
OECD OR SME multi-signal	14	32,610	4,861	25.71	31.84

Under the unit-value normalisation, the dual route has higher net benefit only when the cost of one additional firm evaluation is below 10.97% of the value assigned to one additional successful candidate. Fig. 4a presents the precision-recall frontier across SME thresholds; Fig. 4b shows the cost-sensitive architecture comparison.

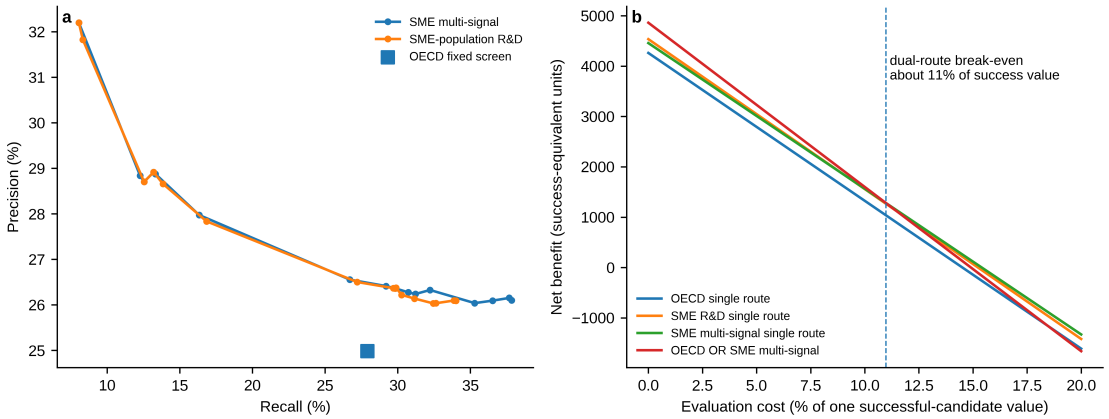


Fig. 4 Selection architecture. Panel a plots the precision-recall frontier across SME-screen thresholds; panel b compares the cost-sensitive net benefit of single- and dual-route architectures

The result supports H3. Polyarchy is not intrinsically superior. It purchases recall at the price of a larger assessment pool. The preferred architecture depends on the institutional cost of detailed evaluation and the opportunity cost of missing a candidate.

5. Discussion

5.1. Target-population validity as a technology-transfer problem

The central theoretical result is that a taxonomy can be correct about its construct and still be unsuitable as a transported selection rule. The OECD taxonomy remains a valid description of whole-industry formal R&D intensity. The failure appears when that ordering is used as a first-stage filter for a target population whose sectoral composition differs. This study therefore shifts the critique from "R&D is a bad measure" to a more precise proposition: the validity of an R&D taxonomy depends on the population represented by its aggregation.

This distinction adds a missing step to technology-transfer evaluation. Prior work has analysed transfer channels, organisational practices, intermediary roles, commercialisation processes, and project assessment (Siegel et al., 2003; Vohora et al., 2004; Markman et al., 2005; Bozeman et al., 2015; Clayton et al., 2018). Recent JTT studies extend this agenda to early sustainability screening (McCarthy et al., 2026), identification of deep-tech ecosystems (Micol et al., 2026), evaluation distortion (Foss, 2026), and predictive entrepreneurial-finance tools (Gafrej et al., 2026). The present contribution concerns the layer before those richer evaluations: the taxonomy that defines the candidate universe.

The framework also sharpens the relationship between classification and validity. Descriptive validation usually asks whether a score converges with a recognised benchmark, is stable across specifications, or has plausible content. Those tests remain necessary, but they are insufficient when the score is used for allocation. Selection validity asks an additional question: which candidates become reachable by subsequent evaluators? Target-population validity explains why a score can pass descriptive tests and still fail that question.

The empirical decomposition makes the distinction tangible: recalculating R&D intensity in the SME population produces almost all of the practical improvement. Adding patents and recognised intangibles changes some sector positions but contributes little to selection performance at the relevant capacity. This is not evidence that patents and intangibles lack conceptual value; it shows that, for this decision, who is represented in the sector aggregate matters more than how many technology indicators are combined.

5.2. Implications for identifying technology-oriented firms

Research on new technology-based and deep-tech firms emphasises that firm types, resource bases, regional capabilities, and development paths are heterogeneous (Storey & Tether, 1998; Colombo & Grilli, 2005; Mokhtar et al., 2026; Micol et al., 2026). Yet many empirical and policy definitions still begin with sector membership. The present evidence shows that a whole-industry high-tech label and an SME-specific technology screen need not identify the same young-firm population.

This matters because target firms are distributed within sectors, not represented by a sector average. In the present data, the SME R&D and OECD screens share nine core industries. The performance difference comes from substitutions at the boundary. The SME R&D rule includes medical and precision instruments, architecture and engineering, and other professional scientific services, while the OECD rule includes motor vehicles and other transport equipment. The result should not be interpreted as a universal ranking of those industries; it demonstrates that the SME segment of a

conventionally high-R&D industry can be less selection-relevant than the SME segment of an industry with a lower whole-industry R&D classification.

The direct R&D-entry result is especially informative. Among early non-R&D firms, later entry into positive R&D is substantially more likely inside the SME-selected sectors, even after controlling for size, financial position, prior growth, and timing. This links target-population screening to subsequent formal knowledge investment rather than only to a financial rating. It also complements Cocozza et al. (2026), who show that the payoffs to R&D differ over time and across firms, and Audretsch, Belitski, and Caiazza (2026), who distinguish general entrepreneurial activity from high-tech entrepreneurship. A population-aligned screen helps locate the firms for which later technological investment is more likely to emerge.

5.3. Implications for screening design and managerial practice

For technology-transfer offices, public programmes, technology-finance units, and innovation intermediaries, the principal implication is procedural. A sector taxonomy should be treated as a low-cost candidate-generation rule, not as a final assessment of firm quality. Before a taxonomy is embedded in eligibility, its reference population should be compared with the programme's target population, and three design questions follow.

First, what population generated the sector score? A whole-industry measure can be appropriate when the programme targets national technological capacity or large R&D performers. It may be less appropriate when the programme targets young SMEs, regional suppliers, or operating small firms.

Second, what is the downstream objective? A screen designed to locate formal R&D performers may differ from one designed to locate firms likely to enter R&D, grow while remaining viable, commercialise technology, or generate employment. The present data show that screen differences are not strong for growth alone but are material for viability, joint success, and R&D entry.

Third, what is the cost of an additional evaluation? When detailed technical and commercial assessment is inexpensive relative to the value of avoiding a false negative, a dual-route rule is attractive. When evaluation is costly, the target-population SME route has higher yield at similar capacity. The break-even calculation converts a general preference for "inclusive screening" into an explicit managerial condition.

These implications are consistent with early-assessment research. McCarthy et al. (2026) present their tool as a primer for subsequent analysis rather than a final decision. Likewise, the SME sector screen should not determine transfer or investment. Its role is to define a candidate set whose members can then be assessed on technology, intellectual property, team, market, sustainability, and finance.

5.4. Why the controls and robustness design matter

The screen indicators are industry-level, while outcomes are firm-level. Without adjustment, differences could reflect the age, scale, financial health, legal form, or initial technology profile of firms concentrated in selected sectors. The technology-adjusted models therefore impose a demanding comparison: firms are compared after conditioning on early R&D participation, patent participation, R&D intensity, intangible assets, prior growth, and a broad set of financial controls.

The fact that SME screen coefficients remain positive after these controls means the result is not mechanically generated by selecting firms that already disclose more technology activity. The later R&D-entry analysis further restricts the sample to firms with no early R&D. At the same time, the absence of a controlled growth-alone effect is substantively important: it prevents the paper from making an inflated "high-growth prediction" claim. The selection gain combines a more favourable later viability profile with a direct increase in later R&D entry, and it improves the joint growth-viability yield.

Industry clustering and wild-bootstrap inference address a second concern. The effective source of treatment variation is 54 industries, not 63,640 independent firms. The more conservative wild-cluster p-values are therefore the appropriate reference for the screen comparisons. Pairwise contrasts remain positive at conventional 10% inference for SME R&D versus OECD and near 5% for the multi-signal versus OECD comparison. Leave-one-industry-out estimates remain positive for all SME screens, showing that no single sector drives the result.

5.5. Boundary conditions and limitations

Several limitations define what this study establishes.

First, the analysis evaluates simulated sector screening rules, not an observed programme that formally used the OECD taxonomy to approve or reject firms. It therefore demonstrates selection consequences under explicit rules, not the causal effect of institutional adoption. Future work could link programme eligibility documents, application records, or policy changes to actual decisions.

Second, the IBK panel is a bank-observed population of operating SMEs. It excludes firms without an IBK relationship, many pre-revenue ventures, and firms without usable financial records. This is an informative target population for policy and finance, but not a representation of all Korean technology ventures. Replication in university spin-offs, grant applicants, venture portfolios, and regional innovation programmes is needed.

Third, the main outcome is a proxy. The joint criterion combines sales growth and a bank rating; it is not realised investor return, patent commercialisation, licensing revenue, productivity, or survival. Later R&D entry provides a more direct technology-related check, but does not prove successful transfer. Future research should examine technology licensing, public-support receipt, equity funding, exports, product innovation, and long-run productivity.

Fourth, the sector scores remain aggregates. Target-population alignment reduces one ecological mismatch but does not eliminate within-industry heterogeneity. Firm-level technical assessment is still required. The screening rule is valuable only as an upstream reduction of the search space.

Fifth, Korea's industrial structure and IBK's mandate may shape the results. The relative importance of population alignment and construct expansion could differ where SME R&D reporting is weaker, patenting is more informative, or services are classified differently.

Finally, the control-rich models remain associational. Observable adjustment, IPCW, logit robustness, wild clustering, placebo portfolios, and timing separation strengthen the evidence but do not create random assignment. The contribution is to selection-system evaluation rather than causal treatment-effect estimation.

6. Conclusion

Technology-transfer systems often use taxonomies before they use detailed evidence. This study shows that the validity of such taxonomies depends on the population they represent and the decision they induce, not only on what they measure.

A whole-industry R&D taxonomy can be descriptively valid and still be selection-invalid for young SMEs. In a Korean bank panel, recalculating the same R&D construct within the target SME population improves the yield of later joint growth-viability outcomes and later R&D entry at nearly equal candidate capacity. Adding patent and intangible signals changes the description of technological activity but contributes little additional selection performance at the margin. The main error is therefore reference-population mismatch rather than simple R&D undercoverage.

This result introduces target-population validity into technology-transfer screening. It also establishes a practical sequence for validation: clarify the construct, identify the reference population, align or test the target population, evaluate downstream selection consequences, and choose a single- or dual-route architecture according to evaluation cost. Before asking whether a taxonomy ranks industries correctly, technology-transfer organisations should ask whether it represents the firms they intend to find.

Statements and declarations

Competing interests

The author declares no competing interests.

Data availability

The protected IBK firm-level source data cannot be publicly redistributed. The accompanying replication package contains non-identifiable firm-level analytical data, industry-level screen inputs, analysis code, output tables, figures, a data dictionary, automated tests, and a secure-environment reconstruction script.

References

Arundel, A., & Kabla, I. (1998). What percentage of innovations are patented? Empirical estimates for European firms. *Research Policy, 27*(2), 127-141.

Audretsch, D. B., Belitski, M., & Caiazza, R. (2026). From democracy to high-tech entrepreneurship. *The Journal of Technology Transfer*. Advance online publication.

Audretsch, D. B., Lehmann, E. E., Paleari, S., & Vismara, S. (2016). Entrepreneurial finance and technology transfer. *The Journal of Technology Transfer, 41*(1), 1-9.

Berger, A. N., & Udell, G. F. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking & Finance, 22*(6-8), 613-673.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.

Bozeman, B. (2000). Technology transfer and public policy: A review of research and theory. *Research Policy, 29*(4-5), 627-655.

Bozeman, B., Rimes, H., & Youtie, J. (2015). The evolving state-of-the-art in technology transfer research: Revisiting the contingent effectiveness model. *Research Policy, 44*(1), 34-49.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics, 90*(3), 414-427.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317-372.

Castellacci, F. (2008). Technological paradigms, regimes and trajectories: Manufacturing and service industries in a new taxonomy of sectoral patterns of innovation. *Research Policy, 37*(6-7), 978-994.

Clayton, P., Feldman, M., & Lowe, N. (2018). Behind the scenes: Intermediary organizations that facilitate science commercialization through entrepreneurship. *Academy of Management Perspectives, 32*(1), 104-124.

Coad, A., & Rao, R. (2008). Innovation and firm growth in high-tech sectors: A quantile regression approach. *Research Policy, 37*(4), 633-648.

Cocozza, R., Gallo, S., Megaravalli, A. V., & Verdoliva, V. (2026). Short-run pain, long-term gain? R&D investment and firm growth in Indian pharma. *The Journal of Technology Transfer*. Advance online publication.

Colombo, M. G., & Grilli, L. (2005). Founders' human capital and the growth of new technology-based firms: A competence-based view. *Research Policy, 34*(6), 795-816.

Corrado, C., Hulten, C., & Sichel, D. (2009). Intangible capital and U.S. economic growth. *Review of Income and Wealth, 55*(3), 661-685.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology, 24*, 313-343.

Foss, N. J. (2026). The myth of clean evaluation: Collective choice, politics, and signal distortion in science and innovation awards. *The Journal of Technology Transfer*. Advance online publication.

Gafrej, I., Gafrej, O., & Bouzgarrou, H. (2026). Predicting post-campaign scenarios of equity crowdfunded firms: A machine learning approach with interaction effect analysis. *The Journal of Technology Transfer*. Advance online publication.

Gallouj, F., & Weinstein, O. (1997). Innovation in services. *Research Policy, 26*(4-5), 537-556.

Galindo-Rueda, F., & Verger, F. (2016). OECD taxonomy of economic activities based on R&D intensity. *OECD Science, Technology and Industry Working Papers, 2016/04*. OECD Publishing.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature, 28*(4), 1661-1707.

Grupp, H., & Schubert, T. (2010). Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy, 39*(1), 67-78.

Henrekson, M., & Johansson, D. (2010). Gazelles as job creators: A survey and interpretation of the evidence. *Small Business Economics, 35*(2), 227-244.

Hollenstein, H. (2003). Innovation modes in the Swiss service sector: A cluster analysis based on firm-level data. *Research Policy, 32*(5), 845-863.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Kitagawa, T., & Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica, 86*(2), 591-616.

Kleinknecht, A., van Montfort, K., & Brouwer, E. (2002). The non-trivial choice between innovation indicators. *Economics of Innovation and New Technology, 11*(2), 109-121.

MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics, 32*(2), 233-254.

Malerba, F. (2002). Sectoral systems of innovation and production. *Research Policy, 31*(2), 247-264.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica, 72*(4), 1221-1246.

Markman, G. D., Gianiodis, P. T., Phan, P. H., & Balkin, D. B. (2005). Innovation speed: Transferring university technology to market. *Research Policy, 34*(7), 1058-1075.

McCarthy, A., Holland, C., & Shapira, P. (2026). Developing an early rapid sustainability assessment tool for responsible innovation. *The Journal of Technology Transfer*. Advance online publication.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Micol, F., Leendertse, J., & van Rijnsoever, F. (2026). Building deep-tech entrepreneurial ecosystems: Factors that promote the presence of deep-tech startups in European regions. *The Journal of Technology Transfer*. Advance online publication.

Mokhtar, J., Knockaert, M., & Vanacker, T. (2026). Academic spin-offs versus new technology-based firms: Growth differences and the role of regional human capital. *The Journal of Technology Transfer*. Advance online publication.

Muller, E., & Doloreux, D. (2009). What we should know about knowledge-intensive business services. *Technology in Society, 31*(1), 64-72.

OECD, & Joint Research Centre. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.

Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy, 13*(6), 343-373.

Pearl, J., & Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science, 29*(4), 579-595.

Peneder, M. (2010). Technological regimes and the variety of innovation behaviour: Creating integrated taxonomies of firms and sectors. *Research Policy, 39*(3), 323-334.

Phan, P. H., & Siegel, D. S. (2006). The effectiveness of university technology transfer. *Foundations and Trends in Entrepreneurship, 2*(2), 77-144.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*(3), 351-357.

Sah, R. K., & Stiglitz, J. E. (1986). The architecture of economic systems: Hierarchies and polyarchies. *American Economic Review, 76*(4), 716-727.

Sah, R. K., & Stiglitz, J. E. (1988). Committees, hierarchies and polyarchies. *Economic Journal, 98*(391), 451-470.

Santamaria, L., Nieto, M. J., & Barge-Gil, A. (2009). Beyond formal R&D: Taking advantage of other sources of innovation in low- and medium-technology industries. *Research Policy, 38*(3), 507-517.

Siegel, D. S., Waldman, D., & Link, A. (2003). Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: An exploratory study. *Research Policy, 32*(1), 27-48.

Stam, E., & Wennberg, K. (2009). The roles of R&D in new firm growth. *Small Business Economics, 33*(1), 77-89.

Storey, D. J., & Tether, B. S. (1998). New technology-based firms in the European Union: An introduction. *Research Policy, 26*(9), 933-946.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., & Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A, 174*(2), 369-386.

Tether, B. S. (2005). Do services innovate (differently)? Insights from the European Innobarometer Survey. *Industry and Innovation, 12*(2), 153-184.

Vohora, A., Wright, M., & Lockett, A. (2004). Critical junctures in the development of university high-tech spinout companies. *Research Policy, 33*(1), 147-175.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics, 141*(2), 1281-1301.

Appendix A. Screen membership

Table A1 lists the industries admitted by each near-capacity-matched screen and reports the two target-population ranks used to explain disagreements at the selection boundary

Table A1 Industries selected by the three near-capacity-matched screens

Code and industry	OECD R&D group	SME R&D rank	SME multi-signal rank	Screen membership
C20 Chemicals	Medium-high	12	12	OECD; SME R&D; multi-signal
C21 Pharmaceuticals	High	4	4	OECD; SME R&D; multi-signal
C26 Electronics and communication equipment	High	7	6	OECD; SME R&D; multi-signal
C27 Medical, precision and optical instruments	Mixed	3	3	SME R&D; multi-signal
C28 Electrical equipment	Medium-high	10	10	OECD; SME R&D; multi-signal
C29 Other machinery and equipment	Medium-high	11	11	OECD; SME R&D; multi-signal
C30 Motor vehicles and trailers	Medium-high	24	22	OECD only
C31 Other transport equipment	High	21	24	OECD only
J58 Publishing and software	Medium-high	2	2	OECD; SME R&D; multi-signal
J61 Postal and telecommunications	Mixed	13	9	Multi-signal only
J62 Computer programming and systems integration	Medium-high	8	8	OECD; SME R&D; multi-signal
J63 Information services	Medium-high	6	7	OECD; SME R&D; multi-signal
M70 Scientific research and development	High	1	1	OECD; SME R&D; multi-signal

Code and industry	OECD R&D group	SME R&D rank	SME multi-signal rank	Screen membership
M72 Architecture, engineering and technical testing	Medium	5	5	SME R&D; multi-signal
M73 Other professional, scientific and technical services	Medium	9	13	SME R&D only

Notes: The two SME screens admit the 12 highest-ranked industries under their respective early-window scores. The OECD rule consists of the 11 unambiguously mapped high and medium-high R&D groups. C30 and C31 are not reclassified as technologically unimportant. They are excluded from both target-population screens because their early SME R&D ranks (24 and 21) and multi-signal ranks (22 and 24) lie well outside the 12-industry capacity boundary. Their source samples are large—7,526 firms for C30 and 3,316 for C31—so the result is not a small-sample artifact. The pattern is consistent with whole-industry R&D strength being concentrated in large original-equipment manufacturers or major suppliers outside the observed SME segment, but the present data do not directly observe large firms and therefore do not establish that mechanism.

Appendix B. Random-sector placebo

Fig. 5 compares the actual multi-signal precision with the nearest 1,000 random 12-industry portfolios by early candidate count. The actual result is at the 99.7th percentile.

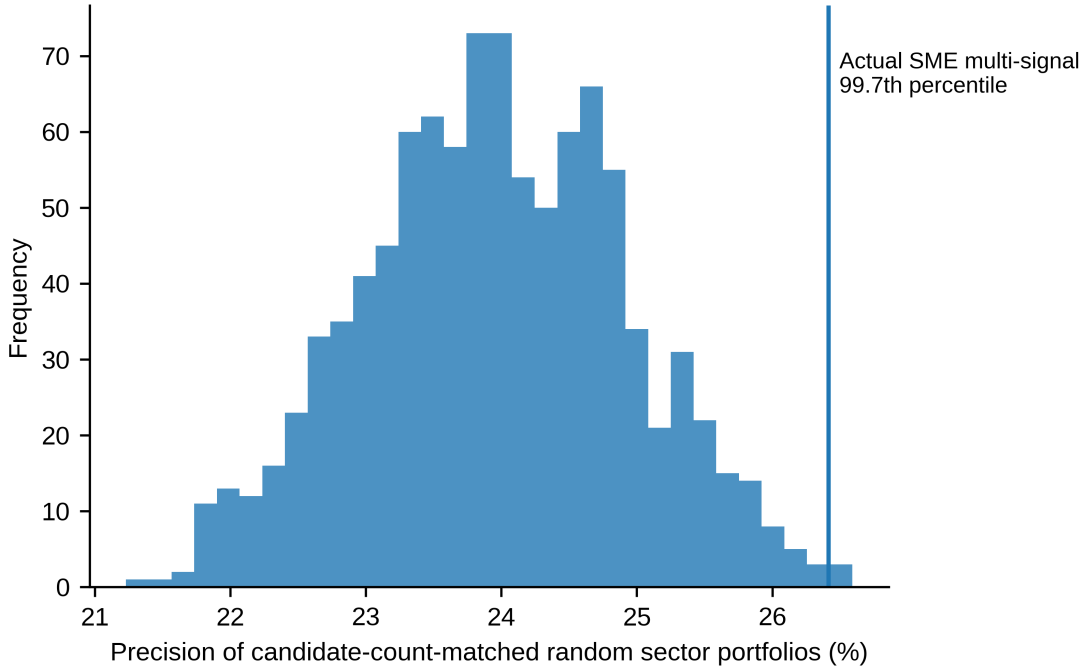


Fig. 5 Candidate-count-matched random-sector placebo distribution. The vertical line marks the observed SME multi-signal precision among the nearest 1,000 random 12-industry portfolios by candidate count

Appendix C. Full controlled models

Table C1 reports the complete coefficient set for the technology-adjusted linear probability models underlying Table 4. All continuous covariates are winsorised at the 1st and 99th percentiles and standardised before estimation. Coefficients are percentage-point changes in the probability of the joint later outcome. Macro-sector and last-early-year fixed effects are included but not displayed

Table C1 Full technology-adjusted linear probability models

Variable	OECD whole-industry R&D	SME-population R&D	SME multi-signal
Screen indicator	0.85 (0.88)	2.23*** (0.78)	2.39*** (0.78)
Age	-3.05*** (0.75)	-3.12*** (0.75)	-3.12*** (0.75)
Age squared	1.84*** (0.69)	1.88*** (0.69)	1.88*** (0.69)
Log assets	-4.14*** (0.47)	-4.08*** (0.47)	-4.08*** (0.47)
Log sales	-0.77* (0.47)	-0.68 (0.47)	-0.67 (0.47)
Log employment	1.69*** (0.23)	1.67*** (0.22)	1.67*** (0.22)
ROA	0.31 (0.28)	0.30 (0.28)	0.30 (0.28)
Leverage	-0.07 (0.82)	-0.08 (0.82)	-0.08 (0.82)
Liquidity	0.98*** (0.16)	0.98*** (0.16)	0.97*** (0.16)

Variable	OECD whole-industry R&D	SME-population R&D	SME multi-signal
Borrowings/assets	-2.90*** (0.70)	-2.86*** (0.69)	-2.86*** (0.69)
Early sales CAGR	0.47** (0.19)	0.48** (0.19)	0.48** (0.19)
Early CAGR missing	0.42 (0.45)	0.44 (0.45)	0.44 (0.45)
Corporate legal form	2.26** (0.93)	2.12** (0.93)	2.11** (0.93)
Number of early observations	-1.33*** (0.30)	-1.34*** (0.30)	-1.34*** (0.30)
Early R&D participation	1.69** (0.78)	1.50** (0.73)	1.50** (0.72)
Early patent participation	10.73*** (0.65)	10.51*** (0.58)	10.49*** (0.57)
Firm R&D intensity	0.48** (0.24)	0.41* (0.23)	0.41* (0.23)
Recognised-intangible ratio	-0.77*** (0.25)	-0.84*** (0.26)	-0.85*** (0.26)
Macro-sector fixed effects	Yes	Yes	Yes
Last-early-year fixed effects	Yes	Yes	Yes
Observations	63,640	63,640	63,640
Industry clusters	54	54	54
R-squared	0.0261	0.0264	0.0264

*Notes: Industry-clustered standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10. The dependent variable equals one when later sales CAGR is at least 10% and the mean later rating is BBB or better. The screen indicator is assigned by two-digit industry; all other displayed regressors are measured at firm level in the early window.*