

Chancen & Grenzen kontrafaktischer Wirkungsanalysen für die österreichische FTI-Politik – eine Situationsanalyse

fteval Kooperationsstudie 2025–2026

Jakob Kofler, KMU Forschung Austria

Jürgen Streicher, Joanneum Research Forschungsgesellschaft mbH

Verena Régent, WPZ Research GmbH

Isabella Wagner, Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung

Beitragende & Qualitätssicherung:

Sascha Sardadvar, WPZ Research GmbH

Peter Kaufmann, KMU Forschung Austria

Inhalt

Executive Summary	4
Executive Summary (EN).....	6
1 Einleitung	8
2 Literaturanalyse	9
2.1 Rolle kausalanalytischer Ansätze in FTI-Evaluierungen	9
2.2 Das methodische Instrumentarium zur Schätzung kausaler Effekte.....	12
2.2.1 Experimentelle und quasi-experimentelle Verfahren	12
2.2.2 Methodische Qualitätsstufen und Evidenzhierarchien	13
2.2.3 Methodische Weiterentwicklungen.....	14
2.3 Grenzen und Herausforderungen der Kausalanalyse.....	16
3 Beispiele internationaler Evaluierungen von F&E-Förderprogrammen und Möglichkeiten zur Umsetzung in Österreich	18
3.1 SkatteFUNN, Norwegen	18
3.2 Zentrales Innovationsprogramm Mittelstand (ZIM), Deutschland	19
3.3 Small Business Innovation Research (SBIR), USA.....	20
4 Analyse des Status Quo	21
4.1 Daten und Methodik	21
4.1.1 Datengrundlage und Identifikationsstrategie	21
4.1.2 Suchalgorithmus und Operationalisierung.....	21
4.1.3 Selektionsprozess.....	22
4.1.4 Analyseraster	22
4.2 Ergebnisse der Analyse des Status Quo	23
4.2.1 Zeitlicher Verlauf und Studiencharakteristika	25
4.2.2 Maßnahmencharakteristika und Untersuchungseinheiten.....	25
4.2.3 Analyse der methodischen Güte	26
5 Kontextbedingungen und Methodenwahl kausalanalytischer Wirkungsanalysen	29
5.1 Struktureller und institutioneller Kontext	29
5.1.1 Formaler Rahmen und Datenzugang	29
5.1.2 Evaluierungskultur und -praxis.....	30
5.1.3 Evaluierungsmarkt und Vergabep Praxis.....	31
5.2 Befunde aus den Interviews.....	31
5.2.1 Verständnis und Relevanz	32
5.2.2 Erwartungen und Nutzen.....	33
5.2.3 Grenzen, Herausforderungen und Voraussetzungen.....	34
5.2.4 Perspektive auf Methodenwahl und Einsatz.....	35

5.2.5 Datenlage und methodische Anforderungen.....	36
5.3 Kontextbedingungen – Fazit.....	37
6 Synthese und Handlungsempfehlungen	38
7 Literaturverzeichnis.....	42
Annex I – Codebuch	48
Abkürzungsverzeichnis	50

Executive Summary

Kontrafaktische Wirkungsanalysen gelten international als Goldstandard, um zu zeigen, ob eine politische Maßnahme tatsächlich wirkt. Wie verbreitet sie in der österreichischen FTI-Evaluierung sind, welche Rahmenbedingungen ihren Einsatz erleichtern oder erschweren und was daraus für die Praxis folgt, untersucht diese Studie auf Basis einer Literaturanalyse, einer systematischen Auswertung österreichischer Evaluierungen und Interviews mit zentralen Akteur:innen des Systems.

17 österreichische Evaluierungen haben kontrafaktische Wirkungsanalysen eingesetzt. Der zeitliche Verlauf ist dabei ungleichmäßig: Auf drei Pionierarbeiten zwischen 2000 und 2004 folgte bis 2016 nur eine einzige weitere Studie, danach 13 im letzten Jahrzehnt. Für ein Forschungssystem dieser Größe ist das eine beachtliche Zahl: Büchele et al. (2024) identifizieren für Deutschland 14 Evaluierungen mit zumindest teilweise kausalen Designs (2009–2023), Collin et al. (2022) für Schweden elf. Gemessen an der Gesamtzahl der in Österreich durchgeführten FTI-Evaluierungen bleibt ihr Anteil dennoch sehr gering, und nur sechs der 17 österreichischen Studien erreichen den methodischen Mindeststandard für belastbare kausale Evidenz. Im internationalen Vergleich ist die Zahl damit hoch, in der heimischen Evaluierungspraxis bleiben kontrafaktische Wirkungsanalysen aber die Ausnahme und kein Standardinstrument.

Drei strukturelle Engpässe erklären diese begrenzte Verbreitung: **1) Kleine Fallzahlen:** Das österreichische FTI-System ist dicht und spezialisiert. Einzelne Programme fördern oft wenige Organisationen. Hinzu kommt, dass Programme regelmäßig angepasst werden. Damit schrumpft die Zahl statistisch vergleichbarer Förderfälle kontinuierlich. **2) Förderkumulation:** Organisationen nehmen gleichzeitig mehrere Förderungen aus verschiedenen Programmen in Anspruch. Die Wirkung eines einzelnen Programms lässt sich dadurch schwer isolieren – besonders, wenn nicht alle Förderdaten zugänglich sind. **3) Zeitverzögerungen:** Kontrafaktische Analysen brauchen Sekundärdaten. Diese erscheinen mit bis zu zwei Jahren Verzögerung (z. B. die Innovationserhebung). Innovationen entstehen erst ein bis zwei Jahre nach Projektabschluss. Evaluierungen laufen meist nur ein Jahr. Das Ergebnis: Vom ersten Förderbeginn bis zur messbaren Wirkung vergehen rund 6,5 Jahre – oft länger als die Evaluierung selbst.

Drei Entwicklungen verändern diese Ausgangslage: **1) Regulatorischer und fachlicher Druck:** Das EU-Beihilferecht verlangt für große Programme kontrafaktische Wirkungsanalysen. Darüber hinaus setzen sich in der akademischen Literatur zur angewandten Ökonometrie zunehmend kontrafaktische Verfahren durch. **2) Erweitertes Methodenrepertoire:** Neuere methodische Entwicklungen, darunter Synthetic Control und Dynamic Matching, ermöglichen kausale Designs auch dort, wo kleine Fallzahlen oder laufende Programme sie bisher verhindert haben. **3) Besserer Datenzugang:** Das Austrian Micro Data Center (AMDC) erleichtert die Nutzung administrativer Mikrodaten und deren Verknüpfung. Damit ist eine Datengrundlage verfügbar, an der kontrafaktische Designs in Österreich bisher regelmäßig gescheitert sind. Der Zugang wird jedoch aktuell als aufwendig, kostenintensiv sowie rechtlich und institutionell anspruchsvoll wahrgenommen.

Literaturauswertung und Interviews zeigen: Die geringe Verbreitung kontrafaktischer Designs beruht nicht auf einer grundsätzlichen Ablehnung. Interesse und methodische Kompetenz wachsen. Neben den genannten methodischen und datenbezogenen Restriktionen wirken auch institutionelle Faktoren, etablierte Evaluierungsroutinen und Vergabepaxen, die knappe Zeitpläne und Budgets vorgeben. Kausalanalysen eignen sich demnach insbesondere dann, wenn Fragestellung, Wirkungslogik, Datenzugang und Ressourcen zusammenpassen, häufig in Kombination mit qualitativen oder theoriegeleiteten Ansätzen.

Was jetzt zu tun ist: Konkrete Empfehlungen

Evaluator:innen

- Methodenkompetenz durch Training breiter streuen und zugleich an der methodischen Weiterentwicklung anschlussfähig bleiben
- Neue methodische Ansätze erproben, um das Anwendungsfeld kontrafaktischer Verfahren unter österreichischen Kontextbedingungen zu erweitern
- Level 3 der Maryland-Skala als Referenz-Mindeststandard anstreben, wo die Kontextbedingungen es zulassen; Identifikationsannahmen transparent dokumentieren und mit Robustheitschecks absichern
- Schätzwerte stets mit ihrer Unsicherheit ausweisen und im Zusammenhang mit den übrigen Evaluierungsbefunden interpretieren, nicht als abschließenden Beleg einer Programmwirkung

Auftraggeber:innen

- Eigene Methodenkompetenz aufbauen, um einschätzen zu können, in welchen Kontexten kontrafaktische Verfahren anwendbar sind und welche Aussagekraft ihre Ergebnisse haben
- Evaluierung bereits beim Programmdesign mitdenken: Wirkungsindikatoren früh identifizieren, Daten von Beginn an erheben
- Kostenrealismus: kontrafaktische Analysen sind methodisch und datenseitig aufwendiger und entsprechend zu budgetieren, wo ein methodischer Mehrwert zu erwarten ist
- Vollständigen Zugang zu allen Förderdaten gewährleisten, um Programmeffekte trotz Förderkumulation isolieren zu können

Umfeldakteur:innen und Datenbereitsteller:innen

- AMDC-Zugang vereinfachen, beschleunigen und kostengünstiger gestalten; FTI-relevante Datensätze dauerhaft verknüpfen statt für jedes Projekt neu
- Für eine differenzierte Nutzung der Ergebnisse sensibilisieren: Messbarkeit ist kein Indikator für Bedeutung, qualitative Befunde sind gemeinsam mit kausalen Schätzwerten zu verorten. Das gilt besonders für Interessenvertretungen, beratende Gremien und die Politik, die Evaluierungsergebnisse in Entscheidungen übersetzen

Executive Summary (EN)

Counterfactual impact evaluations are regarded internationally as the gold standard for demonstrating whether policy measures actually work. This study investigates how widespread such approaches are in Austrian R&I evaluation, which contextual conditions facilitate or impede their application, and what implications follow for practice. The analysis rests on a literature review, systematic assessment of Austrian evaluations, and interviews with key stakeholders.

Seventeen Austrian evaluations have deployed counterfactual methods. The temporal trajectory is uneven: three pioneering studies between 2000 and 2004 were followed by only one additional study through 2016, then thirteen in the past decade. For a research system of Austria's size, this is a noteworthy figure. Büchele et al. (2024) identify fourteen evaluations with at least partial causal designs for Germany (2009–2023); Collin et al. (2022) document eleven for Sweden.

Measured against the total number of R&I evaluations conducted in Austria, however, their share remains very small, and only six of the seventeen Austrian studies meet the methodological minimum standard for robust causal evidence. The absolute count is high by international comparison, but in domestic evaluation practice counterfactual impact analyses remain the exception rather than a standard instrument.

Three structural bottlenecks explain this limited uptake. **First, small sample sizes:** Austria's R&I system is densely structured and specialized. Individual programmes support relatively few organizations, and programmes are regularly adjusted. This continuously shrinks the pool of statistically comparable funded cases. **Second, funding accumulation:** organizations receive multiple grants from different programmes simultaneously. The effect of any single programme becomes difficult to isolate, particularly when not all funding data are accessible. **Third, temporal lags:** counterfactual analyses depend on secondary administrative data, which are published with delays of up to two years (for example, the Community Innovation Survey). Innovation outcomes typically emerge one to two years after project completion. Evaluations typically run for one year. The result: from initial funding disbursement to measurable impact in secondary data, approximately 6.5 years elapse – often longer than the evaluation period itself.

Three developments are changing this landscape. **First, regulatory and disciplinary pressure:** EU state aid regulations now mandate counterfactual impact analyses for large programmes. Beyond this, the academic literature on applied econometrics increasingly adopts counterfactual approaches. **Second, expanded methodological toolkit:** recent methodological innovations, including synthetic control and dynamic matching, enable causal designs even where small sample sizes or ongoing programmes previously precluded them. **Third, improved data access:** the Austrian Micro Data Center (AMDC) streamlines the use of administrative microdata and their linkage. This provides data infrastructure that counterfactual designs in Austria have regularly lacked. Current access, however, is perceived as administratively demanding, costly, and legally complex.

Literature analysis and interviews reveal that limited uptake of counterfactual designs does not stem from fundamental rejection. Interest and methodological competence are growing. Beyond the methodological and data-related constraints noted above, institutional factors, established evaluation routines, and procurement practices that impose tight timelines and budgets also matter. Causal analyses prove most suitable when research questions, theories of change, data access, and resources align – frequently in combination with qualitative or theory-driven approaches.

What Needs to Happen Now: Concrete Recommendations

Evaluators

- Broaden methodological training whilst remaining aligned with methodological innovation;
- Pilot new approaches to expand the applicability of counterfactual methods under Austrian context conditions;
- Pursue Maryland Scale Level 3 as a reference minimum standard where context permits, documenting identifying assumptions transparently and securing them with robustness checks;
- Always present estimates alongside uncertainty ranges and interpret them within the broader evaluation context, not as definitive proof of programme effect.

Commissioners

- Build internal methodological capacity to assess when counterfactual approaches are feasible and what their findings warrant;
- Integrate evaluation planning from programme design onwards, identifying impact indicators early and beginning data collection from inception;
- Budget realistically, recognising that counterfactual analyses are methodologically and informationally more demanding and should be funded accordingly where analytical value is expected;
- Guarantee complete access to all funding data so that programme effects can be isolated despite funding accumulation.

Ecosystems and data providers

- Simplify, accelerate, and reduce costs for AMDC access;
- Establish permanent linkage of R&I-relevant datasets rather than creating new linkages for each project;
- Build awareness for differentiated use of findings, recognising that measurability is not an indicator of significance and that qualitative findings must be co-located with causal estimates;
- Engage especially with interest groups, advisory bodies, and policymakers who translate evaluation results into decisions.

1 | Einleitung

Evaluierungen sind ein wesentliches Instrument zur Überprüfung und Verbesserung von Forschungs-, Technologie- und Innovationspolitik (FTI-Politik). Sie ermöglichen es Policymaker:innen und Fördergebern, die Wirksamkeit ihrer Maßnahmen zu verstehen und evidenzbasierte Entscheidungen für die Weiterentwicklung von Programmen zu treffen. Eine zentrale Herausforderung für Evaluierungen besteht darin, kausale Effekte von Politikinterventionen zu identifizieren: Lassen sich beobachtete Veränderungen tatsächlich auf die untersuchte Maßnahme zurückführen, oder wären diese auch ohne die Intervention eingetreten?

Kontrafaktische Wirkungsanalysen bieten einen methodischen Zugang zu dieser Frage. Sie ermöglichen es, durch strukturierte Vergleiche zwischen begünstigten und kontrafaktischen Vergleichsgruppen Rückschlüsse auf kausale Effekte zu ziehen. In anderen Politikbereichen (etwa der Arbeitsmarkt- und Bildungspolitik) sind solche Methoden etabliert; in der FTI-Politik hingegen bleiben sie bislang unterrepräsentiert. Die Gründe dafür sind vielfältig: Sie reichen von Herausforderungen bei der Datenverfügbarkeit über institutionelle Rahmenbedingungen bis hin zu Fragen der wirtschaftlichen und zeitlichen Machbarkeit.

Im Jahr 2024 veröffentlichte die deutsche Expertenkommission Forschung und Innovation (EFI) eine Studie, die systematisch dokumentierte, in welchem Ausmaß kontrafaktische Methoden in der deutschen FTI-Evaluierungspraxis Anwendung finden. Das österreichische Bundesministerium für Frauen, Wissenschaft und Forschung (BMFWF) erkannte das Potenzial einer vergleichbaren Analyse für den österreichischen Kontext und unterstützte die Idee einer eigenständigen Situationsanalyse. Die Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung (fteval) schrieb daraufhin eine Kooperationsstudie aus, um diese Forschungsfragen gemeinsam mit interessierten Mitgliedsinstitutionen zu bearbeiten. Das Konsortium – bestehend aus Jakob Kofler (KMU Forschung Austria), Verena Grubmüller-Régent (WPZ Research GmbH), Jürgen Streicher (Joanneum Research Forschungsgesellschaft mbH) und Isabella Wagner (Plattform fteval) – führte die Studie zwischen Juli 2025 und Juli 2026 durch.

Ziel der vorliegenden Studie ist es, die Anwendung dieser kontrafaktischen Methoden in der Evaluierung von FTI-Politikmaßnahmen in Österreich systematisch zu untersuchen. Das Vorgehen orientiert sich methodisch an einer aktuellen deutschen Untersuchung (Büchele et al., 2024) und wird für den österreichischen Kontext angepasst und erweitert:

1. Zunächst wird die einschlägige akademische Literatur zu kausalanalytischen Ansätzen in der Evaluierungsforschung aufgearbeitet (Abschnitt 2 | Literaturanalyse).
2. Darauf aufbauend analysiert die Studie bestehende österreichische Evaluierungen aus dem FTI-Bereich, die kontrafaktische Methoden bereits eingesetzt haben (Abschnitt 4 | Analyse des Status Quo).
3. Ergänzend wurden internationale Best-Practice-Beispiele untersucht (SkatteFUNN in Norwegen, ZIM in Deutschland, SBIR in den USA), um Strukturen und Voraussetzungen zu identifizieren, die den erfolgreichen Einsatz kausalanalytischer Methoden ermöglichen (Abschnitt 3 | Beispiele internationaler Evaluierungen von F&E-Förderprogrammen und Möglichkeiten zur Umsetzung in Österreich).
4. Schließlich werden die Bedingungen und Einflussfaktoren diskutiert, die den Einsatz solcher Methoden in der österreichischen Praxis begünstigen oder behindern (Abschnitt 5 | Kontextbedingungen und Methodenwahl kausalanalytischer Wirkungsanalysen).

Abschließend werden die Ergebnisse zusammengeführt, um Handlungsempfehlungen abzuleiten, wie die Nutzung robuster kausalanalytischer Ansätze in der Evaluierung von FTI-Politikmaßnahmen weiter gestärkt werden kann.

2 | Literaturanalyse

Evidenzbasierte Politikgestaltung erfordert, dass Maßnahmen systematisch untersucht und ihre Wirkungen mit wissenschaftlich fundierten Methoden überprüft werden (Edler et al., 2013; Kugler et al., 2014). Hierfür steht ein breites Spektrum quantitativer und qualitativer Evaluierungsansätze zur Verfügung, die je nach Kontext, Datenlage und Zielsetzung unterschiedlich kombiniert werden. Die valide kausale Zuschreibung (Attribution) von Wirkungen stellt dabei eine zentrale methodische Herausforderung dar (Edler et al., 2016).

Als methodischer „Goldstandard“ zur Lösung dieses Attributionsproblems gilt das experimentelle Design, insbesondere in Form von Randomized Controlled Trials (RCTs). Die zentrale Stärke liegt hierbei in der Randomisierung: Durch die zufällige Zuteilung in Interventions- und Kontrollgruppen werden systematische Unterschiede zwischen den Gruppen eliminiert. Dies ermöglicht die Konstruktion einer validen kontrafaktischen Situation – also die Beantwortung der Frage, was ohne die Maßnahme geschehen wäre –, da Ergebnisunterschiede unter diesen Idealbedingungen isoliert auf die Intervention zurückgeführt werden können (Shadish et al., 2002).

In der Praxis der Forschungs-, Technologie- und Innovationspolitik (FTI-Politik) lässt sich diese ideale Versuchsanordnung jedoch kaum realisieren. Oft läuft die Intervention bereits, eine zufällige Zuteilung ist aus rechtlichen oder ethischen Gründen nicht möglich und die Daten sind häufig unvollständig (Frontier Economics, 2024; Reiner & Smoliner, 2012). Zudem treten volkswirtschaftliche Wirkungen erst mit großer zeitlicher Verzögerung ein. Um unter diesen komplexen Feldbedingungen dennoch belastbare Aussagen zu Wirkungen treffen zu können, stellt eine qualitativ hochwertig umgesetzte kontrafaktische Wirkungsanalyse anhand von Quasi-Experimenten eine valide methodische Alternative dar (Athey & Imbens, 2017). Während sich dieser Ansatz in anderen Politikbereichen mit ähnlichen Herausforderungen bereits etabliert hat, ist er in der FTI-Politik noch vergleichsweise wenig verbreitet. Dies steht im deutlichen Gegensatz zu den expliziten Forderungen der Europäischen Kommission sowie nationaler Expert:innen, die seit Jahren eine stärkere Evidenzbasierung und methodische Fundierung in der Evaluierung von FTI-Politik anmahnen (Hünernmund & Czarnitzki, 2019).

Die kontrafaktische Wirkungsanalyse ermöglicht es, die entscheidende Frage zu beantworten: *Was wäre ohne die untersuchte politische Intervention geschehen?* Nur durch die Simulation dieses Szenarios lässt sich der tatsächliche Nettoeffekt einer Maßnahme realitätsnah schätzen (Lach, 2002; Ramberg & Knell, 2012). Unabhängig von der politischen Programmatik ist dieser Aspekt zentral, da sich nur so bestimmen lässt, welche *zusätzlichen* Wirkungen (Additionalität) durch den Einsatz öffentlicher Mittel tatsächlich ausgelöst wurden (Buisseret et al., 1995; David et al., 2000).

2.1 | Rolle kausalanalytischer Ansätze in FTI-Evaluierungen

Die Nutzung kausalanalytischer Methoden in der Evaluierungsforschung ist keine neue Entwicklung, hat aber seit den 1990er Jahren deutlich an Bedeutung gewonnen (Zúñiga, 2024; Athey & Imbens, 2017). Insbesondere in den letzten zwei Jahrzehnten ist ein starker Anstieg empirischer Studien zu verzeichnen, die fortgeschrittene ökonometrische und quasi-experimentelle Verfahren einsetzen, um die Wirkungen politischer Maßnahmen zu identifizieren (Fougère & Jacquemet, 2019). Diese Entwicklung wird von Angrist und Pischke (2010) treffend als „Credibility Revolution“ in der empirischen Ökonomie beschrieben. Fortschritte in der Methodik – etwa bei Matching-Verfahren, Instrumentvariablen (IV) oder Regressionsdiskontinuitätsdesigns (RDD) – sowie eine verbesserte Verfügbarkeit von Mikrodaten haben die Anwendungsmöglichkeiten stark erweitert und zu einer signifikanten Professionalisierung der Wirkungsanalyse beigetragen (Huber, 2021; Crato & Paruolo, 2019).

Kausalanalytische Ansätze zielen darauf ab, den Zusammenhang zwischen Interventionen und beobachteten Ergebnissen so zu modellieren, dass Korrelationen tatsächlich als Ursache-Wirkungs-Beziehungen interpretiert werden können. Im Kern versuchen sie, den kontrafaktischen Zustand – also das Szenario ohne Intervention – statistisch zu rekonstruieren. Dies ermöglicht es, den Effekt einer Maßnahme (z. B. auf Beschäftigung, Produktivität oder Innovationsleistung) als Differenz zwischen tatsächlichem und kontrafaktischem Ergebnis zu schätzen (Fougère & Jacquemet, 2019). Diese Verfahren korrigieren dabei systematische Verzerrungen durch Selbstselektion oder unbeobachtete Heterogenität und sind damit unverzichtbar, um den spezifischen Effekt eines Programms von jenen externen Einflüssen und Trends zu entflechten, die das Ergebnis auch bei Abwesenheit der Intervention verändert hätten (Ardito et al., 2025).

In der Verwendung dieser Ansätze zur Wirkungsprüfung zeigen sich erhebliche Unterschiede zwischen Politikfeldern. Während in der Arbeitsmarkt- und Bildungspolitik oder in der Entwicklungszusammenarbeit experimentelle und quasi-experimentelle Designs längst zum Standard gehören, hinkt ihre Verbreitung im Bereich der Forschungs-, Technologie- und Innovationspolitik (FTI-Politik) deutlich hinterher (Hünormund & Czarnitzki, 2019).

In der FTI-Politik werden Förderentscheidungen meist auf Grundlage von Qualität und Relevanz der Projektanträge getroffen. Damit unterscheiden sich geförderte Unternehmen systematisch von nicht geförderten, was die Identifikation kausaler Effekte erschwert. Würde man diese strukturellen Unterschiede ignorieren, bestünde die Gefahr, die bessere Performance der Unternehmen fälschlicherweise der Förderung zuzuschreiben, obwohl sie oft auf deren ohnehin höhere Leistungsfähigkeit zurückzuführen ist. Eine valide Analyse erfordert daher zwingend die Konstruktion einer Vergleichsgruppe, die den Begünstigten in allen relevanten Aspekten gleicht, um diesen Selektionseffekt vom tatsächlichen Programm-Effekt zu isolieren (Ardito et al., 2025).

Wie Hünormund und Czarnitzki (2019) betonen, stellt dieses Selektionsproblem die zentrale methodische Hürde dar. Deshalb müssen Evaluator:innen auf Beobachtungsdaten zurückgreifen und geeignete ökonometrische Strategien entwickeln, um Verzerrungen durch unbeobachtete Heterogenität zu minimieren. Die Autoren illustrieren dies am sogenannten „Confounding Problem“: Unterschiede zwischen geförderten und nicht geförderten Unternehmen umfassen oft nicht nur beobachtbare Merkmale (z. B. Größe, Branche), sondern auch schwer messbare Faktoren wie Managementqualität oder Innovationsorientierung. Ohne statistische Korrektur würden diese Unterschiede dazu führen, dass positive Korrelationen fälschlich als kausale Effekte der Förderung interpretiert werden.

In den letzten zehn Jahren finden vermehrt rigorose quasi-experimentelle Designs Anwendung in der FTI-Evaluierung, etwa Regressionsdiskontinuitäts- und Instrumentvariablenansätze, die helfen, das Kontrafaktum präziser zu identifizieren (Einiö, 2014; Bronzini & Iachini, 2014; Howell, 2017). Dennoch bleibt die Evidenzlage heterogen und die Generalisierbarkeit der Ergebnisse begrenzt. Hünormund und Czarnitzki (2019) kommen in ihrer europäischen Studie etwa zu dem Schluss, dass F&E-Förderungen im Durchschnitt keinen einheitlichen Effekt auf Beschäftigung, Umsatz oder Patente zeigen, wobei erhebliche Unterschiede je nach Projekttyp und Programmdesign bestehen.

Parallel zu den methodischen Fortschritten wächst der politische Druck, die Evidenzbasis staatlicher Interventionen zu verbessern. Dabei ist vorauszuschicken, dass kausalanalytische Designs nicht den alleinigen Anspruch auf Robustheit erheben (Stern et al., 2012). Insbesondere für das Verständnis komplexer Wirkungsmechanismen und Kontextbedingungen sind qualitative und theoriebasierte Ansätze (Theory-Based Evaluation) unverzichtbar (Rogers, 2008; Rogers et al., 2000). Sie zielen primär darauf ab, den Beitrag (Contribution) einer Intervention in einem komplexen Ursachengefüge plausibel nachzuzeichnen, ohne zwingend den Anspruch zu erheben, diesen von anderen Einflussfaktoren statistisch vollständig zu isolieren (Mayne, 2012).

Kritisch wird in der Fachdebatte jedoch das spezifische Fehlen jener Verfahren beurteilt, die über diese Plausibilisierung hinaus eine valide Quantifizierung des Nettoeffekts (Attribution) erlauben. Während die

Contribution Analysis fragt, ob die Maßnahme einen entscheidenden Einfluss hatte, liefert nur die Attribution die Antwort auf die Frage, wie groß dieser Anteil im Vergleich zum kontrafaktischen Szenario exakt ist (Stern et al., 2012).

Da diese kausalanalytischen Standards in der FTI-Politik jedoch noch keineswegs flächendeckend verankert sind, haben sich zuletzt vereinzelte Untersuchungen dem Ziel gewidmet, den Status quo der Anwendungspraxis in verschiedenen Ländern empirisch zu vermessen.

Ein einschlägiges Beispiel hierfür ist die Arbeit von Collin, Sandström und Wennberg (2022) zur schwedischen Evaluierungslandschaft. Die Autoren untersuchten die Methodennutzung in 110 innovationspolitischen Evaluierungen und stellten dabei fest, dass kontrafaktische Ansätze nur äußerst selten zur Anwendung kommen: Lediglich in elf Fällen (10 %) konnte ein solches Design nachgewiesen werden. Dieser Befund verdeutlicht exemplarisch, dass der Einsatz kausalanalytischer Methoden zur Isolierung von Nettoeffekten selbst in innovationsstarken Ländern bislang die Ausnahme darstellt.

Für das deutsche Innovationssystem liefert eine aktuelle Studie von Büchele et al. (2024), die im Kontext der Expertenkommission Forschung und Innovation (EFI) entstand, ähnliche Befunde. Sie belegt, dass kontrafaktische Wirkungsanalysen in der deutschen FTI-Politik bislang eine untergeordnete Rolle spielen: Lediglich 14 der 81 (17 %) untersuchten Evaluierungen nutzten Methoden, die zumindest bedingt kausale Rückschlüsse zulassen. Als zentrale strukturelle Hürden identifizieren die Autoren neben der eingeschränkten Datenverfügbarkeit und Transparenz insbesondere den Umstand, dass entsprechende methodische Ansätze in den Ausschreibungen oft nicht explizit eingefordert werden.

Eine vergleichbare Untersuchung von Santos und Coad (2023) beleuchtet die Anwendung kontrafaktischer Methoden in der Evaluierung der EU-Kohäsionspolitik. Die Ergebnisse zeigen, dass seit 2015 lediglich rund 8 % der in den Mitgliedsstaaten durchgeführten Evaluierungen auf solchen kausalanalytischen Ansätzen basieren. In Evaluierungen, die spezifisch das Politikfeld „Forschung und Innovation“ betreffen, liegt dieser Anteil sogar nur bei etwa 3 %.

Vor dem Hintergrund dieser geringen Verbreitung stellt das europäische Beihilferecht den derzeit stärksten institutionellen Treiber für kausalanalytische Designs dar, wenngleich nur für ein sehr spezifisches Segment. Im Zuge der State Aid Modernisation (SAM) und der überarbeiteten Allgemeinen Gruppenfreistellungsverordnung (AGVO) hat die Europäische Kommission verbindliche Vorgaben für die Legitimierung staatlicher Markteingriffe verankert (Europäische Kommission, 2014b, 2014c).

Für die FTI-Politik bedeutet dies: Sobald Fördermaßnahmen ein Durchschnittsvolumen von 150 Mio. Euro pro Jahr überschreiten, ist eine Wirkungsanalyse mittels kausalanalytischer Methoden zwingend vorgeschrieben (Europäische Kommission, 2014a). Da diese hohe Budgetschwelle jedoch nur von wenigen großvolumigen Programmen erreicht wird, hat diese regulatorische Verpflichtung die absolute Anzahl an kontrafaktischen Evaluierungen bislang nur marginal erhöht.

Die verpflichtende Anwendung kausalanalytischer Methoden stellte für viele europäische Länder ein Novum dar. Entsprechende Kompetenzen mussten in den Mitgliedsstaaten und der Evaluierungs-Community erst aufgebaut werden, zumal diese Verfahren in der FTI-Politik noch wenig etabliert waren und sich auch die notwendige Dateninfrastruktur erst langsam entwickelte. Zur Unterstützung dieses Kapazitätsaufbaus beauftragte die Kommission das *Competence Centre on Microeconomic Evaluation* (CC-ME) am Joint Research Centre (JRC), welches seither methodische „Toolboxes“, Workshops, ein Repositorium von Evaluierungen mit kausalanalytischen Methoden und regelmäßige Qualitätsprüfungen bereitstellt.

Der aktuelle Bericht zur Qualitätsüberprüfung (Ardito et al., 2025) konstatiert einen positiven Trend, der sich in signifikanten Qualitätsverbesserungen der Berichte sowie einer zunehmenden Verbreitung valider methodischer Ansätze wie *Difference-in-Differences* oder *Matching* manifestiert. Nichtsdestotrotz bestehen weiterhin Herausforderungen hinsichtlich der Datenverfügbarkeit und -qualität sowie der eindeutigen Identifikation kausaler Wirkungszusammenhänge.

2.2 | Das methodische Instrumentarium zur Schätzung kausaler Effekte

Die Schätzung kausaler Effekte in der Evaluierung von FTI-Maßnahmen beruht auf einem breiten Spektrum an Methoden, deren Auswahl wesentlich durch Datenverfügbarkeit, Evaluierungsdesign und theoretisches Vorwissen bestimmt wird. Im Zentrum steht dabei die Frage, wie aus beobachteten Zusammenhängen belastbare Schlussfolgerungen über Ursache-Wirkungs-Beziehungen gezogen werden können (Czarnitzki, 2023).

Konzeptionell wurde die kontrafaktische Grundfrage bereits eingeführt; methodisch lässt sie sich auf das von Holland (1986) formulierte „fundamentale Problem der Kausalinferenz“ zurückführen (Athey & Imbens, 2017): Für jede Untersuchungseinheit ist stets nur das Ergebnis des tatsächlich eingetretenen Interventionszustands beobachtbar, nicht aber jenes des alternativen Szenarios. Kausale Effekte können daher nie direkt gemessen, sondern nur indirekt über den Vergleich unterschiedlich behandelter Einheiten geschätzt werden. Damit wird die Unterscheidung von Korrelation und Kausalität zum zentralen methodischen Problem, an dem die im Folgenden dargestellten Verfahren ansetzen.

2.2.1 | Experimentelle und quasi-experimentelle Verfahren

Als „Goldstandard“ der Kausalanalyse gilt das randomisierte kontrollierte Experiment (Randomised Controlled Trial, RCT). Dabei werden die betrachteten Einheiten per Zufall auf eine Gruppe mit bspw. Förderung (Interventionsgruppe) und eine Gruppe ohne Förderung (Kontrollgruppe) aufgeteilt. Dadurch gleichen sich – im Durchschnitt – andere Unterschiede zwischen den Gruppen aus, sodass beobachtete Unterschiede in den Ergebnissen plausibel auf die Intervention zurückgeführt werden können (Krabel & Wagner, 2024; Athey & Imbens, 2017). In der Praxis von FTI-Maßnahmen ist eine solche zufällige Zuteilung jedoch selten möglich, weil Förderentscheidungen meist auf Gutachten, strategischen Prioritäten sowie ethischen und politischen Vorgaben beruhen.

Aus diesem Grund dominieren in der Evaluierungspraxis der FTI-Politik quasi-experimentelle Verfahren, die versuchen, die kontrafaktische Situation statistisch zu rekonstruieren. Die wichtigsten Ansätze sind in

Tabelle 1 kurz beschrieben; für eine vertiefende Diskussion der methodischen Details und Implementierung sei auf Czarnitzki (2023) verwiesen. Die tatsächliche Nutzung dieser Verfahren ist jedoch an strikte Voraussetzungen gebunden, die in vielen Evaluierungskontexten nicht erfüllt werden können.

Empirisch zeigt sich eine starke Konzentration auf zwei Hauptverfahren. Wie Daten des Joint Research Centre (JRC, 2025) belegen, basieren 17 der 24 gelisteten FTI-Evaluierungen auf Difference-in-Differences (DiD) und elf auf Propensity Score Matching (PSM). Die häufige gleichzeitige Nennung deutet auf die Verbreitung hybrider Strategien hin (z. B. Conditional DiD), welche die Stärken beider Ansätze vereinen. Designs wie RDD oder IV hingegen, die exogene Variationen in der Zuteilung voraussetzen (z. B. Schwellenwerte), bleiben mit zwei bzw. null Anwendungen die Ausnahme.

Tabelle 1 | Wichtigste quasi-experimentelle Verfahren in der Evaluierungspraxis

Methode	Kernidee	Zentrale Anforderung
Propensity Score Matching (PSM)	„Statistische Zwillinge“: Für jedes geförderte Unternehmen wird ein möglichst ähnliches nicht-gefördertes Unternehmen gesucht (Rosenbaum & Rubin, 1983).	Umfangreiche Daten zu beobachtbaren Merkmalen (Größe, Branche, Vorjahresumsatz etc.) sind notwendig.
Difference-in-Differences (DiD)	„Doppelte Differenz“: Vergleicht die <i>Entwicklung</i> (Trend) der Geförderten mit der Kontrollgruppe, um zeitlich stabile Unterschiede herauszurechnen (Angrist & Pischke, 2009).	Parallel Trend Assumption: Ohne Förderung hätten sich beide Gruppen gleich entwickelt.
Regression Discontinuity (RDD)	„Der Schwellenwert“: Nutzung harter Fördergrenzen (z. B. Bewertungsscore > 80 Punkte). Knapp abgelehnte Unternehmen dienen als Vergleich für knapp angenommene (Imbens & Lemieux, 2008).	Ein Cut-off, der die Zuteilung bestimmt und nicht manipulierbar ist.
Instrumental-Variablen (IV)	„Der externe Hebel“: Nutzung eines externen Zufallsfaktors (Instrument), der die Förderung beeinflusst, aber nicht das Ergebnis (Angrist & Krueger, 2001).	Eine unabhängige Variable, die zwar die Teilnahme an der Maßnahme beeinflusst, aber nicht vom Erfolg oder den Eigenschaften des Unternehmens abhängt (meist schwer zu finden).

2.2.2 | Methodische Qualitätsstufen und Evidenzhierarchien

Sobald die prinzipielle Anwendbarkeit der verschiedenen Methoden geklärt ist, gilt es zu bestimmen, welche der möglichen Verfahren die validesten Ergebnisse liefern. Bei der quantitativen Bestimmung kausaler Wirkungszusammenhänge gibt es hier eine etablierte Hierarchie, die Forschende oder Auftraggeber:innen bei der Auswahl unterstützt.

Um die Belastbarkeit empirischer Studien im wirtschaftspolitischen Kontext vergleichbar zu machen, hat sich die Adaption der Maryland Scientific Methods Scale (SMS) durch das What Works Centre for Local Economic Growth (2016) als relevanter Standard etabliert. Diese Skala klassifiziert Evaluationsdesigns in fünf Qualitätsstufen, wobei Level 3 als methodischer Mindeststandard gilt, um von kausaler Evidenz sprechen zu können:

1. Level 1 (Einfache Korrelation): Entweder ein reiner Vorher-Nachher-Vergleich der Interventionsgruppe ohne Kontrollgruppe, oder ein einfacher Querschnittsvergleich zwischen geförderten und nicht geförderten Einheiten, jeweils ohne jegliche Kontrollvariablen zur statistischen Bereinigung von Gruppenunterschieden.
2. Level 2 (Statistische Kontrolle): Verwendet beobachtbare Kontrollvariablen, um Störfaktoren zu isolieren. Dies kann in zwei Szenarien erfolgen: (1) Entweder werden in einem Querschnittsvergleich beobachtbare Unterschiede zwischen den Gruppen durch Regressionen ausgeglichen, oder es werden in einem Vorher-Nachher-Vergleich zeitliche Veränderungen (z. B. makroökonomische

- Faktoren) durch Kontrollvariablen statistisch isoliert. (2) Propensity Score Matching fällt dann unter Level 2, wenn es ausschließlich im Querschnitt ohne Zeitverlaufskomponente eingesetzt wird: in Kombination mit einem Vorher-Nachher-Vergleich kann es Level 3 erreichen.
3. Level 3 (Vergleichsgruppen-Designs im Zeitverlauf): Kombiniert den Vorher-Nachher-Vergleich der Interventionsgruppe mit einer expliziten Vergleichsgruppe als Kontrafaktum (z. B. Difference-in-Differences). Die zentrale Anforderung ist die Plausibilität der Annahme paralleler Trends: Ohne die Intervention müssten sich Interventions- und Vergleichsgruppe zeitlich gleichläufig entwickelt haben. Diese Annahme kann durch Vorperiodenvergleiche gestützt, aber nicht abschließend bewiesen werden. Regressionen oder Propensity Score Matching dienen der zusätzlichen Justierung beobachtbarer Unterschiede: unbeobachtete Unterschiede bleiben als Restrisiko bestehen.
 4. Level 4 (Quasi-Zufälligkeit): Designs, die eine quasi-zufällige Variation in der Zuteilung zur Intervention nutzen, sodass glaubhaft gemacht werden kann, dass sich Interventions- und Kontrollgruppe nur durch die zufällige Exposition unterscheiden. Dies beinhaltet meist die Nutzung von Instrumentvariablen (IV) oder Schwellenwerten (Regression Discontinuity Design), bei denen externe Ereignisse oder institutionelle Regelungen eine exogene Variation erzeugen, die einer Zufallszuteilung nahekommmt.
 5. Level 5 (Randomisierung): Forschungsdesigns mit expliziter zufälliger Zuteilung (Randomisierung) in Interventions- und Kontrollgruppen (Randomised Controlled Trials, RCTs). Es muss nachgewiesen werden, dass keine signifikanten Unterschiede zwischen den Gruppen bestehen und dass Probleme wie selektiver Ausfall (Attrition) vernachlässigbar bzw. durch Sensitivitätsanalysen adressiert sind.

Anhand dieser Skala lässt sich zeigen, dass die Evaluierungspläne und Berichte aus dem Repositorium des *Joint Research Centre* (2025) trotz gesteigerter Qualität vielfach nicht die Mindeststandards für Kausalevidenz erfüllen. Lediglich jene Arbeiten, die einen *Difference-in-Differences*-Ansatz in Kombination mit Matching anwenden, erreichen Stufe 3. Die Anforderungen der Stufe 4 werden im gesamten Sample nur von den zwei Studien erfüllt, die ein *Regression Discontinuity Design* (RDD) nutzen.

2.2.3 | Methodische Weiterentwicklungen

Fortschritte in statistischen Verfahren

Jenseits der klassischen Designs haben sich spezifische statistische Weiterentwicklungen etabliert, um komplexere Interventionsverläufe abzubilden. So lässt sich die Sequenzanalyse mit Matching-Verfahren kombinieren, wenn detaillierte Längsschnittdaten vorliegen. Im FTI-Kontext bedeutet dies, spezifische Muster in der Unternehmenshistorie – wie etwa den vorherigen oder nachfolgenden Erhalt anderer Förderungen – abzubilden, um die Vergleichbarkeit von Interventions- und Kontrollgruppe zu präzisieren (Deiana et al., 2019).

Eine grundsätzliche Schwierigkeit für Kontrollgruppenvergleiche stellen dauerhaft laufende Maßnahmen dar, bei denen alle anspruchsberechtigten Unternehmen jederzeit eine Förderung beantragen können. Ein bekanntes österreichisches Beispiel dafür ist die steuerliche Forschungsprämie. Hier eignet sich der Kreis der Nicht-Geförderten aus zwei Gründen nur eingeschränkt als Kontrollgruppe. Erstens sind viele dieser Unternehmen schon deshalb nicht vergleichbar, weil sie überhaupt nicht forschen. Zweitens ist bei den forschenden Unternehmen, die keine Förderung beantragt haben, anzunehmen, dass dies kein Zufall ist: Sie dürften sich systematisch von den Geförderten unterscheiden – etwa durch schlechteren Informationszugang, geringere Managementqualität oder schwächere administrative und strategische Kapazitäten. In beiden Fällen würde ein einfacher Vergleich mit den Geförderten den tatsächlichen Effekt der Förderung verzerren. Einen Ausweg bietet das sogenannte *Dynamic Matching*, das die unterschiedlichen Zeitpunkte der Erstantragstellung nutzt: Unternehmen, die erst später eine Förderung erhalten, dienen, solange sie selbst noch nicht gefördert sind, als Vergleichsgruppe für früher geförderte

Unternehmen mit ähnlicher Vorgeschichte. Geschätzt wird damit der Effekt, zu einem bestimmten Zeitpunkt gefördert zu werden – verglichen mit einer (noch) nicht erfolgten Förderung. Voraussetzung sind zwei Annahmen: erstens, dass sich Geförderte und Vergleichsunternehmen bei gleichen beobachtbaren Merkmalen und gleichem bisherigen Förderverlauf nicht systematisch unterscheiden, und zweitens, dass die Unternehmen den Antragszeitpunkt nicht in Erwartung künftiger Ergebnisse wählen (Fredriksson & Johansson, 2008; Deiana et al., 2019). Gerade die zweite Annahme ist im Förderkontext kritisch: Verschieben Unternehmen ihren Antrag gezielt auf ein Jahr, in dem sie ohnehin viel forschen, ist sie verletzt und der geschätzte Effekt verzerrt.

An ihre Grenzen stoßen klassische Methoden auch in stark konzentrierten Branchen mit wenigen Akteuren, etwa im Flugzeug- oder Automobilbau, in denen zudem die meisten Unternehmen gefördert werden und deshalb keine natürliche Vergleichsgruppe existiert. Für solche Fälle wurde die Methode der synthetischen Kontrollgruppe (Synthetic Control Method) entwickelt – ursprünglich von Abadie & Gardeazabal (2003), methodisch ausgearbeitet von Abadie, Diamond & Hainmueller (2010, 2015). Statt ein einzelnes reales Vergleichsunternehmen zu suchen, wird aus mehreren nicht geförderten Einheiten eine gewichtete, künstliche Vergleichseinheit zusammengesetzt, die den Verlauf des geförderten Unternehmens vor der Maßnahme möglichst genau nachzeichnet; der Effekt ergibt sich aus der Differenz beider Verläufe danach (Deiana et al., 2019). Vorteilhaft sind der nachvollziehbare, datengestützte Vergleichsmaßstab, die anschauliche grafische Darstellung und die Absicherung über Placebo-Tests schon bei einer einzigen behandelten Einheit. Voraussetzung bleibt jedoch, dass genügend nicht geförderte Einheiten als Bausteine vorhanden sind. Sind nahezu alle Akteure gefördert, stößt auch dieses Verfahren an seine Grenze. Hinzu kommen der hohe Bedarf an Daten zu einer möglichst langen, gut passenden Vorlaufperiode und das Fehlen üblicher Signifikanztests. Auf diese Schwächen hat die Forschung mit zahlreichen Weiterentwicklungen reagiert, die Abadie (2021) zusammenfassend einordnet.

Maschinelles Lernen und künstliche Intelligenz in der Wirkungsforschung

Zur präziseren Quantifizierung kausaler Effekte erweitern Ansätze des maschinellen Lernens (ML) das methodische Instrumentarium erheblich. Nichtparametrische Verfahren wie Random Forests oder künstliche neuronale Netze ersetzen dabei nicht die klassische parametrische Regression, sondern werden in ein übergeordnetes, semiparametrisches Rahmenmodell – insbesondere das Double Machine Learning (DML; Chernozhukov et al., 2018) – eingebunden: Dort übernehmen sie die flexible Schätzung von Störgrößen wie dem Einfluss von Störfaktoren (Confoundern) oder der Teilnahmewahrscheinlichkeit, während der kausale Zielparameter weiterhin identifiziert und inferenzstatistisch interpretierbar bleibt (Chernozhukov et al., 2024; Matthay, 2025). Ein zentraler Vorteil dieser datengetriebenen Ansätze liegt in der Verarbeitung hochdimensionaler Daten: Insbesondere regularisierende Verfahren wie das Lasso ermöglichen eine robustere Kontrolle von Störfaktoren, indem sie aus einer Vielzahl potenzieller Kovariaten automatisch die relevantesten Einflussgrößen auswählen und so Verzerrungen durch ausgelassene Variablen verringern (Huber, 2021; Matthay, 2025). Darüber hinaus erlauben auf Entscheidungsbäumen beruhende und neuronale Verfahren die flexible Abbildung komplexer Interaktionen und nicht-linearer Beziehungen sowie die Aufdeckung von Wirkungsheterogenität (Heterogeneous Treatment Effects). Mittels Causal Forests (Wager & Athey, 2018) lässt sich datengetrieben schätzen, wie unterschiedlich eine Intervention in verschiedenen Subgruppen wirkt (konditionale durchschnittliche Behandlungseffekte, CATE), und so identifizieren, welche Subgruppen am stärksten von ihr profitieren, ohne diese a priori definieren zu müssen. Dies schafft die Grundlage für weiterführende Mediationsanalysen und eröffnet zugleich die Möglichkeit, künftige Fördermaßnahmen zielgenauer auf jene Gruppen auszurichten, bei denen die höchste Wirksamkeit zu erwarten ist (Huber, 2021).

Ein weiterer entscheidender Fortschritt liegt in der Nutzung unstrukturierter Daten. Chernozhukov et al. (2024) demonstrieren, wie KI-Modelle genutzt werden können, um Informationen aus komplexen Textquellen – etwa Projektbeschreibungen oder Patenttexten – zu extrahieren, die in rein tabellarischen Daten fehlen. Durch die Umwandlung dieser Texte in numerische Repräsentationen (Embeddings) und deren Einbindung in Double-Machine-Learning-Verfahren lassen sich Verzerrungen durch ausgelassene Variablen (Omitted Variable Bias) reduzieren – allerdings nur, soweit die relevanten Störfaktoren im Text tatsächlich enthalten sind. Dieser Vorbehalt ist konzeptionell wichtig: Das Verfahren verschiebt das Identifikationsproblem, löst es aber nicht vollständig. Die Kausalinterpretation bleibt an die Annahme gebunden, dass außerhalb des Textmaterials keine wesentlichen unbeobachteten Störgrößen verbleiben.

Neben der Datenaufbereitung bergen Large Language Models (LLMs) auch erhebliches Potenzial zur konzeptionellen Unterstützung kausaler Analysen. Han (2024) weist insbesondere auf die Eignung von LLMs bei der Identifikation von Instrumentvariablen (IV) hin. Die Herausforderung dabei ist, Variablen zu finden, die zwar die Teilnahme an einer Maßnahme beeinflussen (Relevanz), das Ergebnis aber nur über diesen Weg und nicht direkt (Ausschlussrestriktion). Da geeignete IVs selten offensichtlich sind, erfordert dieser Prozess in der Regel die Exploration eines breiten Suchraums sowie eine argumentative Plausibilisierung der Kausalannahmen. Der spezifische Vorteil von LLMs liegt darin, diesen Suchraum durch die Auswertung disziplinübergreifender Literatur und institutioneller Zusammenhänge systematisch zu erweitern und den Prozess deutlich zu beschleunigen. Die beiden Bedingungen müssen jedoch weiterhin belegt werden: Die Relevanz ist empirisch prüfbar, die Ausschlussrestriktion bleibt eine argumentativ zu begründende, nicht direkt testbare Annahme. LLMs können diesen Schritt unterstützen, aber nicht ersetzen.

2.3 | Grenzen und Herausforderungen der Kausalanalyse

Die Aussagekraft kausaler Schlussfolgerungen hängt wesentlich davon ab, welche theoretischen und empirischen Voraussetzungen im jeweiligen Evaluierungskontext erfüllt sind. Hünermund und Bareinboim (2025) weisen darauf hin, dass drei Faktoren die Belastbarkeit kausaler Schätzungen erheblich einschränken können: unbeobachtete Störfaktoren (also Einflüsse, die nicht in den Daten erfasst sind), eine selektive Datenverfügbarkeit sowie eine strukturelle Heterogenität der untersuchten Population. Betroffen sind dabei sowohl die *interne Validität* – die Frage, ob ein Effekt im untersuchten Fall korrekt geschätzt ist – als auch die *externe Validität*, also die Übertragbarkeit auf andere Kontexte.

Deaton und Cartwright (2018) präzisieren, dass selbst randomisierte Experimente keine generalisierbaren Ergebnisse garantieren. Unter idealen Bedingungen liefern sie zwar unverzerrte Schätzungen des durchschnittlichen Behandlungseffekts (Average Treatment Effect, ATE) innerhalb der Studienpopulation. Ob sich dieser Wert aber auf andere Populationen, Kontexte oder Zeitpunkte übertragen lässt, ist eine davon getrennte Frage. Sie lässt sich durch ein noch so sauberes experimentelles Design allein nicht beantworten.

Über diese statistischen Hürden hinaus besteht eine grundsätzlichere, erkenntnistheoretische Grenze: Kontrafaktische Kausalanalysen – Verfahren, die einen beobachteten Zustand mit einem hypothetischen „Was wäre ohne die Maßnahme geschehen?“ vergleichen – sind primär darauf ausgelegt, die Existenz und Stärke eines Effekts zu isolieren (das Dass). Den dahinterliegenden Wirkungsmechanismus (das Wie und Warum) können sie meist nicht empirisch rekonstruieren. Es gibt zwar quantitative Ansätze zur Analyse der Wirkmechanismen, etwa die Mediationsanalyse (Imai et al., 2011). Diese setzen jedoch zusätzliche, oft starke Annahmen über den Ablauf der kausalen Kette voraus, die im Feld der FTI-Politik selten vollständig erfüllbar sind. In der angewandten Evaluierungsforschung bleibt der

Wirkungsmechanismus daher häufig im Inneren der „Black Box“ (Pawson & Tilley, 1997), was die Erklärungskraft der Ergebnisse und ihren Nutzen für die Politikgestaltung begrenzt.

Hinzu kommt, dass die passende Herangehensweise an Kausalität auch von der Natur des untersuchten Gegenstands abhängt. Das kontrafaktische Modell, das den meisten quantitativen Verfahren zugrunde liegt, funktioniert dort gut, wo eine Maßnahme klar abgegrenzt und stabil ist. Bei komplexen Politikfeldern stößt es jedoch an Grenzen: Hier wirken oft mehrere Instrumente und Rahmenbedingungen zusammen, und ein Ergebnis kann über unterschiedliche Wege zustande kommen. Ein Beispiel für ein alternatives Kausalitätsverständnis ist der konfigurationale Ansatz: Statt den durchschnittlichen Effekt einer einzelnen Maßnahme zu schätzen, fragt er, welche Kombinationen von Faktoren gemeinsam zu einem Ergebnis führen. Damonte und Negri (2023) veranschaulichen dies an der Tabakpolitik – um das Rauchen einzudämmen, können Regierungen auf Aufklärungskampagnen, Warnhinweise, Werbeverbote, Tabaksteuern oder Rauchverbote setzen, einzeln oder in Kombination. Welche Wirkung eintritt, hängt davon ab, welches Bündel an Maßnahmen unter welchen Bedingungen greift. Solche Ansätze erlauben es, Wirkungskonstellationen abzubilden, die sich im kontrafaktischen Rahmen nur schwer fassen lassen (Damonte & Negri, 2023; Maxwell, 2012).

Abgesehen von diesen allgemeinen methodischen Einschränkungen sind die Rahmenbedingungen für kausale Evaluierung in der innovationspolitischen Praxis besonders anspruchsvoll. Dies lässt sich beispielhaft an fünf zentralen strukturellen Beschränkungen kontrafaktischer Verfahren in der FTI-Evaluierung zeigen (Edler et al., 2016; Frontier Economics, 2024; Reiner & Smoliner, 2012):

- **Kleine Stichproben in spezialisierten Programmen:** Bei themen- oder sektorspezifischen Maßnahmen (z. B. Quantentechnologie) stoßen quantitative Methoden an statistische Grenzen. In solchen Nischen ist sowohl die Zahl der geförderten als auch der vergleichbaren nicht geförderten Einheiten häufig zu gering für belastbare kontrafaktische Schätzungen.
- **Komplexes Förderökosystem:** Mehrere sich überlagernde Förderinstrumente auf nationaler und europäischer Ebene wirken gleichzeitig. Das erschwert die klare Zuordnung von Effekten ebenso wie die Bildung sauberer, unbeeinflusster Kontrollgruppen.
- **Immaterielle und schwer fassbare Outputs:** Zentrale Wirkungsdimensionen wie neu generiertes Wissen, Netzwerkbeziehungen oder strategische Neuausrichtungen lassen sich kaum direkt messen. Kausalanalysen weichen deshalb häufig auf messbare Hilfsgrößen aus – etwa Patente oder Publikationen –, die nur einen Ausschnitt der tatsächlich relevanten Effekte abbilden.
- **Schiefe Effektverteilung:** FTI-Maßnahmen erzeugen typischerweise wenige sehr wirkungsstarke Erfolge und eine breite Masse geringer oder nicht nachweisbarer Effekte (eine stark rechtsschiefe Verteilung). Verfahren, die auf Mittelwerten beruhen, insbesondere ATE-Schätzungen, können diese Struktur systematisch verdecken. Für politisch zentrale Fragen nach Exzellenz und Selektivität sind sie daher nur begrenzt aussagekräftig.
- **Langzeitwirkungen und Zuordnungsproblem:** Die strukturellen Effekte von FTI-Politiken, etwa auf technologische Entwicklungspfade oder Innovationsökosysteme, entfalten sich über Zeiträume, die übliche Beobachtungsfenster überschreiten. Mit wachsendem zeitlichem Abstand wird es zugleich methodisch immer schwieriger, den Effekt der Maßnahme von externen Schocks und parallel laufenden Politikmaßnahmen zu trennen.

In der Praxis der Evaluierung von FTI-Politik kommt daher der Triangulation und Mixed-Methods-Designs zentrale Bedeutung zu. Die Kombination kontrafaktischer Kausalschätzung mit qualitativen Verfahren erlaubt eine belastbarere Interpretation sowohl der Effektstärken als auch der zugrundeliegenden Mechanismen und hilft, quantitative Evidenz in ihren Kontext einzuordnen (Damonte & Negri, 2023).

3 | Beispiele internationaler Evaluierungen von F&E-Förderprogrammen und Möglichkeiten zur Umsetzung in Österreich

Der akademische Diskurs im vorangegangenen Abschnitt dokumentiert, unter welchen Bedingungen kausalanalytische Methoden tragfähig eingesetzt werden können. Die folgenden Beispiele aus Norwegen, Deutschland und den USA zeigen, wie drei Länder diese methodischen Anforderungen institutionell verankert haben und welche strukturellen Voraussetzungen sie geschaffen haben, um rigorose Designs in der Praxis zu ermöglichen.

3.1 | SkatteFUNN, Norwegen¹

Das Programm SkatteFUNN wurde im Jahr 2002 eingeführt, um F&E im norwegischen Privatsektor durch steuerliche Anreize zu erhöhen² – in dieser Hinsicht hat es einen ähnlichen Charakter wie die österreichische Forschungsprämie³, wenngleich es im Unterschied dazu Obergrenzen der Förderung gibt. Etwa die Hälfte der Begünstigten sind Unternehmen mit weniger als 10 Beschäftigten, 80 % beschäftigen weniger als 50 Personen. Rund 50 % der Begünstigten hatten zuvor keine F&E-Erfahrung; für über 40 % war SkatteFUNN die einzige in Anspruch genommene öffentliche Förderung.

Die gegenständliche Evaluierung des Programms stützt sich auf eine Regressions-Diskontinuitäts-Analyse (RDD) sowie einen Differenz-in-Differenz-Ansatz (DiD) mit Matching – mit dieser Kombination an quasi-experimentellen Methoden soll Endogenitäts- bzw. Selektionsproblematiken entgegengewirkt werden.

Die Regressions-Diskontinuitäts-Analyse nutzt die Fördergrenze für steuerlich absetzbare F&E-Ausgaben und vergleicht Unternehmen knapp ober- und unterhalb dieses Schwellenwerts, wobei angenommen wird, dass diese Unternehmen vergleichbar sind und sich primär durch den marginalen Förderanreiz unterscheiden. Die Ergebnisse bestätigen einen positiven, wenn auch moderateren Effekt: Gefördert werden vor allem zusätzliche marginale F&E-Investitionen, und Unternehmen werden angeregt, ihre F&E-Aktivitäten fortzuführen.

Mittels des ergänzenden Differenz-in-Differenz-Ansatzes wurde das Investitionsverhalten von begünstigten und nicht-begünstigten Unternehmen vor und nach Einführung der Maßnahme sowie über Förderkohorten und Änderungen im Programmdesign verglichen. Die Kontrollgruppe wurde über ein Matching-Verfahren gebildet und umfasst in der vorliegenden Anwendung überwiegend Unternehmen ohne eigene F&E-Tätigkeit.⁴ Die Ergebnisse zeigen eine signifikante positive Input-Additionalität der Förderung, deren Ausmaß jedoch über verschiedene Kohorten und Zeiträume variiert und tendenziell mit der Zeit abnimmt. Insgesamt verdeutlicht die kombinierte Anwendung von RDD und DiD, wie unterschiedliche quasi-experimentelle Ansätze genutzt werden können, um die Robustheit kausaler Effektschätzungen trotz nicht-zufälliger Programmteilnahme zu erhöhen.

Insofern bietet die Evaluierung wertvolle methodische Anknüpfungspunkte und kann insbesondere für ähnliche Fragestellungen bzw. bei geeigneter Stichprobenbildung als Modell für F&E-Evaluierungen in Österreich dienen. Eine Replikation mit österreichischen Daten ist für klassische Förderinstrumente

¹ Benedictow, Andreas, Emil Cappelen Bjørn, Fernanda Winger Eggen, Marthe Norberg-Schulz, Marina Rybalka und Rolf Røtnes (2018): Evaluation of SkatteFUNN, Samfunnsøkonomisk analyse AS Report no. 18-2018. <https://kudos.dfo.no/documents/12299/files/12437.pdf>

² Kleine und mittlere Unternehmen können bis zu 20 % der förderfähigen F&E-Kosten als Steueranreiz erhalten, größere Unternehmen bis zu 18 %

³ Deren Evaluierung von 2017 wird im Bericht (Benedictow et al., 2018) ab S. 34f. ausführlich zitiert.

⁴ Im Bericht (S. 60) gibt es eine kurze Abhandlung über F&E-betreibende Unternehmen, die keine Förderung beantragen, allerdings gibt es keine Angaben darüber, wie viele es sind. Da die Förderung an keine weiteren Bedingungen geknüpft ist, ist anzunehmen, dass sich überwiegend Unternehmen ohne eigene F&E-Tätigkeit in der Kontrollgruppe befinden. Diese Auswahl der Kontrollgruppe kann zwar kritisch hinterfragt werden, wird jedoch durch die ergänzende Anwendung des RDD-Ansatzes methodisch weitgehend abgedeckt. Allerdings muss betont werden, dass die F&E-Ausgaben nicht als Kontrollvariable fungieren.

durchaus realistisch. Die zentralen Wirkungsindikatoren zu F&E-Ausgaben werden in Norwegen in einer eigenen Erhebung erfasst. Diese deckt allerdings lediglich etwa 30 % der F&E-treibenden Unternehmen ab und erfasst insbesondere kleine F&E-treibende Unternehmen nur unzureichend. Aus diesem Grund musste die Evaluierung, um die Datengrundlage zu verbessern, zusätzliche Datenquellen heranziehen, beispielsweise andere erhaltene Förderungen, und die fehlenden Daten für F&E-Ausgaben schätzen.

Im Gegensatz dazu erfasst die F&E-Erhebung in Österreich alle F&E-treibenden Unternehmen, sodass eine Anwendung der Methodik ohne Ungenauigkeiten bei wichtigen Wirkungsindikatoren und somit mit soliden Ergebnissen möglich wäre. Einschränkungen für eine Anwendung im österreichischen Kontext ergeben sich hingegen für spezifische methodische Ansätze wie die in der Evaluierung eingesetzte Regressions-Diskontinuitäts-Analyse (vgl. Benedictow et al., 2018, S. 50). Ein Beispiel hierfür ist die Forschungsprämie, bei der eine nutzbare Obergrenze fehlt, sodass andere Quellen exogener Variation gefunden werden müssen. Am ehesten bieten sich die über die Zeit geänderten Rahmenbedingungen an, etwa die mehrfach angehobenen Fördersätze (zuletzt 14 %) oder der geänderte pauschale Unternehmer:innenlohn, die sich als zeitliche Politikänderungen für quasi-experimentelle Designs (z. B. Difference-in-Differences um den jeweiligen Änderungszeitpunkt) nutzen lassen.

3.2 | Zentrales Innovationsprogramm Mittelstand (ZIM), Deutschland⁵

In dieser Studie wird ein Programm evaluiert, das das erklärte Ziel hat, private F&E-Investitionen, speziell von KMU, zu stärken. Das Programm ist themen- und technologieoffen und umfasst primär nicht rückzahlbare Zuschüsse für marktnahe F&E-Projekte. Es gliedert sich in mehrere Projektformen wie F&E-Einzelprojekte, F&E-Kooperationsprojekte u.a. und ist vom Charakter her in Österreich am ehesten mit dem Basisprogramm der FFG vergleichbar. Im Unterschied zu SkatteFUNN ist nicht jedes F&E betreibende Unternehmen automatisch qualifiziert. Die Evaluierung geht u.a. der Frage nach, welche Wirkungen sich bei den Zuwendungsempfänger:innen beobachten lassen und inwiefern diese kausal dem Förderstimulus zuzurechnen sind. Die kausale Programmwirkung wird mithilfe von konditionalen Differenz-in-Differenz-Analysen bestimmt.

Interessant ist die Bildung der Kontrollgruppe, bei der auf – ansonsten übliche – Matching-Verfahren verzichtet wird (S. 95). Stattdessen kommt das Entropie-Balancing zum Einsatz, wobei die Gesamtheit potenzieller Kontrollgruppen-Unternehmen herangezogen und jedes Kontrollgruppen-Unternehmen so gewichtet wird, dass die Verteilung der Kontrollgruppe insgesamt statistisch identisch ist mit der Verteilung der Gruppe der geförderten Unternehmen in Bezug auf ein festgelegtes Set von Variablen. In der Kontrollgruppe berücksichtigt werden alle Unternehmen mit unter 1.000 Beschäftigten, auch solche, die keine eigene (interne) F&E betreiben, allerdings ist F&E-Tätigkeit als Variable in der Auswahl enthalten.⁶

Im Rahmen der Schätzungen werden die folgenden fünf abhängigen Variablen herangezogen: (i) F&E-Ausgaben, (ii) F&E-Beschäftigung, (iii) Beschäftigung insgesamt, (iv) Umsatz und (v) Umsatz je Beschäftigten. Dabei werden Informationen aus fünf Datenquellen zusammengeführt: (i) Angaben aus ZIM-Anträgen, (ii) Angaben aus der eigenen Primärdatenerhebung, (iii) Angaben aus dem Mannheimer Innovationspanel, (iv) Angaben aus dem Mannheimer Unternehmenspanel, (v) Angaben zu erhaltenen F&E-Förderungen.

⁵ Stehnen, Thomas, Isabel Schöfl, Thomas Danneil, Michael Astor, Christian Rammer, Bettina Peters, Alexander Ehrlich, Kornelius Kraft (2024): Evaluation des „Zentralen Innovationsprogramms Mittelstand“ (ZIM), Studie der ZEW und der Prognos AG im Auftrag des Bundesministeriums für Wirtschaft und Klimaschutz. https://www.zew.de/fileadmin/FTP/gutachten/ZIM_Evaluation_Endbericht_extern.pdf

⁶ Die Strukturgleichheit wurde über die Merkmale Größe, Alter, Wirtschaftszweig, Standort, Bonität, F&E-Tätigkeit, erhaltene F&E-Förderungen sowie Umsatz- und Beschäftigungsdynamik im Zeitraum vor der ZIM-Förderung hergestellt.

Methodisch überzeugt die Evaluierung durch den konsequenten Einsatz kausaler Verfahren und eines klar strukturierten Forschungsdesigns. Insbesondere wird die kausalanalytische Wirkungsanalyse in ein breiteres Mixed-Method-Design eingebettet. Konkret umfasst diese umfangreiche quantitative Analysen auf Basis kombinierter Datenquellen (Förderdaten, Primärbefragung, Sekundärdaten zu Unternehmenskennzahlen und Innovationsaktivität) sowie qualitative Erhebungen, um auch zugrundeliegende Kontextfaktoren zu erfassen.

Für den österreichischen Kontext ist die Studie als besonders relevant einzuschätzen. Das zugrundeliegende Evaluierungsdesign ist grundsätzlich sehr gut übertragbar. Voraussetzung ist die Verknüpfung administrativer Förderdaten mit Unternehmens- und Innovationsdaten, wie sie in Österreich etwa bei FFG- oder aws-Programmen vorliegen. Entsprechende Dateninfrastrukturen wie das AMDC und Verknüpfungsroutinen werden in Österreich bereits vereinzelt genutzt, wobei hier der Vorteil besteht, dass die F&E-Erhebung und die Innovationserhebung von der Statistik Austria durchgeführt werden, die auch das AMDC betreibt, wodurch eine direkte Verknüpfung der Mikrodaten auf Unternehmensebene ermöglicht wird. Andere Möglichkeiten der Datenverbindung werden schon länger genutzt, etwa indem Förderdaten mit Unternehmensdatenbanken verbunden werden.⁷

3.3 | Small Business Innovation Research (SBIR), USA⁸

Das gegenständliche Programm ist ein seit 1982 bestehendes Förderinstrument der US-Bundesregierung zur Unterstützung kleiner Unternehmen in Forschung und Entwicklung. Es zielt darauf ab, private F&E-Aktivitäten zu stimulieren, technologische Innovationen voranzutreiben und die Kommerzialisierung von Forschungsergebnissen zu fördern. Die Förderung erfolgt über wettbewerbliche, nicht rückzahlbare Zuschüsse, die von mehreren Bundesbehörden in zweistufigen Verfahren vergeben werden, und gilt als eines der zentralen innovationspolitischen Instrumente in den USA.

In der Evaluierung des SBIR-Programms wird ein Regressions-Diskontinuitäts-Ansatz gewählt. Dieser eignet sich besonders, da die Zuweisung der Förderung nicht zufällig erfolgt, sondern an einen klar definierten numerischen Schwellenwert gekoppelt ist: Die Autorin nutzt ihren Zugang zur Reihung der Förderanträge nach einer Bewertungspunktzahl, um anhand des Cut-offs – oberhalb dessen Anträge gefördert werden, unterhalb nicht – eine Kontrollgruppe zu bilden. Der zentrale methodische Gedanke besteht darin, dass Unternehmen unmittelbar um diese Diskontinuität herum als quasi-zufällig gefördert bzw. nicht gefördert gelten können, sodass potenzielle Selektionsverzerrungen weitestgehend vernachlässigt werden können. Auf diese Weise lässt sich testen, ob die Förderung verschiedene Innovationsvariablen beeinflusst, darunter Patentierungen, Zitierungen, Finanzierungen und Umsatz. Die methodische Stringenz dieses Ansatzes spiegelt sich auch in der Publikation der Evaluierung in einer renommierten Fachzeitschrift wider.

Das untersuchte zweistufige Programm SBIR unterscheidet sich erheblich von den Programmen in Österreich. Nichtsdestoweniger kann die Evaluierung für Österreich als Vorbild dienen, sofern es gelingt, die Reihung bei der Vergabe bestimmter Förderungen nachvollziehbar abzubilden. Zu berücksichtigen ist jedoch, dass entsprechende Diskontinuitäten nicht in allen Förderprogrammen vorliegen – etwa dann nicht, wenn keine standardisierte, quantitative Bewertung mit klar definiertem Schwellenwert besteht, ab dem eine Förderung gewährt wird.

⁷ Zum Beispiel bei den Evaluierungen der FFG-Programme „Kleinprojekt“ (<https://www.wpz-research.com/evaluierung-der-ffg-programmlinie-kleinprojekt/>) und „FFG-Innovationsscheck mit Selbstbehalt“ (<https://www.wpz-research.com/evaluierung-des-ffg-innovationsschecks-mit-selbstbehalt-2/>)

⁸ Howell, Sabrina T. (2017): Financing Innovation: Evidence from R&D Grants, American Economic Review 107(4), 1136–1164. <https://doi.org/10.1257/aer.20150808>

4 | Analyse des Status Quo

Dieser Abschnitt folgt einer systematischen Analyse, die sich methodisch an der Studie von Büchele et al. (2024) orientiert. Diese Studie wurde im Kontext der deutschen Expertenkommission Forschung und Innovation (EFI) durchgeführt und untersucht das Ausmaß, in dem kontrafaktische Wirkungsanalysen in Evaluierungen der deutschen Forschungs- und Innovationspolitik eingesetzt werden. Die vorliegende Analyse überträgt diesen Ansatz auf den österreichischen Kontext und folgt dabei den von der EFI-Studie entwickelten Kategorien und Bewertungskriterien, um konzeptionelle Vergleichbarkeit herzustellen.

Untersucht werden österreichische Evaluierungen der letzten 30 Jahre bis zum Stichtag 15.10.2025, in denen kontrafaktische Methoden bereits zur Anwendung kamen. Ziel ist es, den aktuellen Status quo der kausalanalytischen Evaluierungspraxis in der österreichischen FTI-Politik zu dokumentieren und dabei zu erfassen, unter welchen Bedingungen solche Methoden eingesetzt werden und welche methodischen Ansätze tatsächlich Verwendung finden.

4.1 | Daten und Methodik

Um diese Frage zu beantworten, kombinieren wir zwei Zugänge: Erstens eine dokumentarische Analyse aller relevanten österreichischen Evaluierungen der letzten 30 Jahre, zweitens ein tieferes Verständnis der Kontextfaktoren, die Entscheidungen über Methodenwahl beeinflussen. Um den Status quo der Nutzung kausalanalytischer Ansätze in der österreichischen FTI-Politikevaluierung empirisch zu erfassen, wurde ein mehrstufiges Selektions- und Analyseverfahren angewandt. Dieses Kapitel beschreibt die Datenquellen, die Suchstrategie sowie das Analyseraster.

4.1.1 | Datengrundlage und Identifikationsstrategie

Die primäre Datenbasis dieser Untersuchung bildet das Repositorium der österreichischen Plattform für Forschungs- und Technologiepolitikevaluierung (fteval). Diese Datenbank fungiert als zentrales Archiv für Evaluierungsberichte im österreichischen FTI-Sektor. Sie umfasst in erster Linie veröffentlichte Evaluierungsberichte, aber auch Studien der eigenen wissenschaftlichen Zeitschrift sowie graue Literatur. Der durchsuchte Bestand deckt einen Zeitraum von 30 Jahren ab und umfasst zum Zeitpunkt der Untersuchung insgesamt 796 Einträge.

Um das Risiko einer systematischen Untererfassung relevanter Studien im Repositorium zu minimieren, wurde ergänzend zur Datenbankrecherche eine Community-Umfrage unter den Mitgliedern der fteval durchgeführt (Auftraggebende und Evaluierende). Diese Kombination der Suchstrategien erwies sich als effektiv: Aus der Umfrage gingen fünf spezifische Evaluierungen hervor, die im Repositorium nicht auffindbar waren und direkt in den Screening-Prozess aufgenommen wurden.

4.1.2 | Suchalgorithmus und Operationalisierung

Die Identifikation relevanter Studien im Repositorium erfolgte mittels einer systematischen Schlagwortsuche. Der entwickelte Suchstring operationalisiert die Anwendung von Kausalanalyseverfahren durch eine umfassende Kombination aus methodenspezifischen Fachtermini sowie konzeptionellen Begriffen der Kausalinferenz.

Hinsichtlich der Sprachwahl wurde berücksichtigt, dass ein Großteil der im Repositorium verzeichneten Quellen in deutscher Sprache vorliegt. Um jedoch Vollständigkeit zu gewährleisten und auch englischsprachige Einträge nicht auszuschließen, wurden ergänzend die entsprechenden englischen Termini berücksichtigt.

Der angewandte Suchstring lautet im vollen Wortlaut:

"Causal analysis" OR "Kausalanalyse" OR "Causal inference" OR "Kausale Inferenz" OR "Counterfactual" OR "Kontrafaktisch" OR "Treatment effect" OR "Maßnahmeneffekt" OR "Control group" OR "Kontrollgruppe" OR "Comparison group" OR "Vergleichsgruppe" OR "Quasi-experimental" OR "Quasi-experimentell" OR "Randomized control trial" OR "RCT" OR "Randomisierte kontrollierte Studie" OR "Difference-in-Differences" OR "Differenz-von-Differenzen" OR "Regression Discontinuity Design" OR "Regressions-Diskontinuitäts-Design" OR "Instrumental Variable" OR "Instrumentenvariable" OR "Matching" OR "Matching-Verfahren" OR "Synthetic Control Method" OR "Synthetische Kontrollmethode" OR "Fixed Effects model" OR "Fixed-Effects-Modell" OR "Selection bias" OR "Selektionsverzerrung"

4.1.3 | Selektionsprozess

Der Selektionsprozess orientierte sich am PRISMA 2020-Statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Page et al., 2021). Wie im PRISMA-Flussdiagramm dargestellt (siehe Abbildung 1), führte die Anwendung des Suchstrings sowie die Inklusion der Umfrage-Ergebnisse zu einer initialen Treffermenge von N = 269 Quellen (264 aus der Datenbanksuche, 5 aus der Umfrage).

Diese 269 Quellen wurden zunächst mithilfe eines Large-Language-Model-gestützten Verfahrens (Google NotebookLM) vorab klassifiziert. Dazu wurden die Volltexte als Dokumentenkörper hochgeladen und über einen einheitlichen, strukturierten Prompt analysiert. Das Modell fasste für jedes Dokument Hauptthema und zentrale Ergebnisse zusammen, stufte die Relevanz für die Fragestellung anhand definierter Kriterien als „relevant“ oder „nicht relevant“ ein und begründete diese Einstufung. Als relevant galt ein Dokument, wenn es spezifische kausalanalytische Identifikationsstrategien – etwa kontrafaktische Analysen, RCTs, Difference-in-Differences, Regression-Discontinuity-Designs, Matching, Instrumentvariablen, synthetische Kontrollgruppen oder Fixed-Effects-Modelle – tatsächlich anwendet oder methodisch erörtert; rein deskriptive, qualitative oder formative Wirkungsanalysen ohne statistische Inferenz wurden als nicht relevant eingestuft. Da die Ausgaben von Large Language Models nicht deterministisch sind und Fehlklassifikationen nicht auszuschließen sind, wurde die KI-Auswahl anschließend manuell validiert. Die Übereinstimmung war hoch; in wenigen Grenzfällen wich die manuelle Beurteilung ab, sodass N = 45 Studien verblieben.

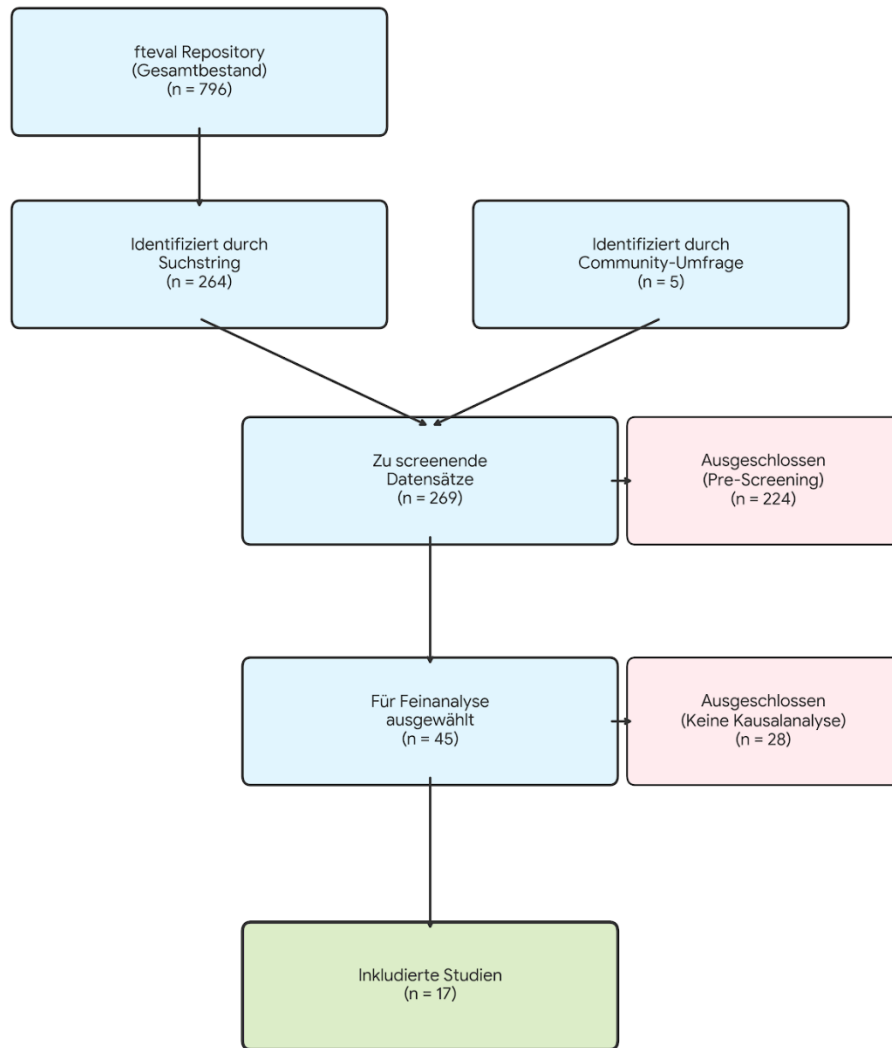
Im anschließenden manuellen Screening dieser 45 Treffer wurde geprüft, ob tatsächlich fundierte kausalanalytische Methoden angewendet wurden oder lediglich deskriptive Vergleiche (z. B. mit einer Kontrollgruppe ohne statistische Inferenz) stattfanden. Aus dem Selektionsprozess resultierten 15 Berichte; da ein fteval-Artikel drei eigenständige RCT-Evaluierungen der FFG bündelt, umfasst der finale Korpus 17 Evaluierungen.

4.1.4 | Analyseraster

Für die systematische Auswertung der identifizierten Studien wurde ein erweitertes Kodierschema entwickelt (siehe Annex I – Codebuch). Dieses basiert grundlegend auf der Klassifikation von Büchele et al. (2024), um eine strukturelle Einordnung der Evaluierungsgegenstände und -designs vorzunehmen.

Um die Qualität der kausalen Inferenz noch differenzierter bewerten zu können, wurde das Raster um die Maryland Scientific Methods Scale (SMS) ergänzt (What Works Centre for Local Economic Growth, 2016). Während das Basisschema bereits eine Einschätzung der internen Validität anhand einer dreistufigen Skala vorsieht, ermöglicht die Integration der Maryland SMS eine standardisierte Beurteilung mit klar definierten Abgrenzungskriterien. Dies erlaubt eine präzisere Bestimmung, inwieweit die angewandten Identifikationsstrategien (z. B. Randomisierung, Matching, Difference-in-Differences) in der Lage waren, Selektionsverzerrungen zu minimieren und valide Nettoeffekte zu isolieren.

Abbildung 1 | PRISMA-Flussdiagramm des Selektionsprozesses



Quelle: Eigene Darstellung in Anlehnung an Page et al. (2021); Datengrundlage: fteval-Repository und eigene Erhebung.

4.2 | Ergebnisse der Analyse des Status Quo

Der Selektionsprozess resultierte in einem finalen Korpus von N = 17 Studien, die eine kontrafaktische Wirkungsanalyse anwenden. Die Ergebnisse zeigen eine zunehmende Verbreitung und methodische Professionalisierung über die Zeit.

Tabelle 2 | Österreichische Evaluierungen mit kausalanalytischen Verfahren, 2000–2024

Jahr	Evaluierungsgegenstand	Auftraggebende	Autor:innen	Kausalanalytisches Verfahren	SMS
2000	ITF-Schwerpunkt FlexCIM	BMWV	Geyer et al.	Ökonometrischer Kontrollgruppenansatz (Probit)	2
2002	Sonderaktion Holzforschung	FFF	Mahlich et al.	Dynamisches Paneldatenmodell (GMM)	2
2004	FFF-Unternehmensförderungen	BMVIT	Schibany et al.	Fixed-Effects-Panelregression	2
2013	Kompetenzzentren Kplus, K_ind/K_net	BMWFI, FFG	Schibany et al.	DiD auf Paneldaten, ohne Matching	2
2016	Förderprogramme der FFG	AK Wien, BMVIT	Bock-Schappelwein et al.	Propensity Score Matching	2
2016	Sparkling Science	BMWFW	Manahl et al.	Matching-basierte Kontrollgruppe	2
2017	FWF International Programmes	FWF	Degelsegger-Márquez et al.	Matching, bibliometrische Analysen	2
2019	IWB/EFRE-Regionalprogramm 2014–2020	ÖROK	Kaufmann et al.	Konditionale DiD mit PSM	3
2019	Frontrunner Programm	BMVIT	Warta et al.	Fixed-Effects-Panelregression	2
2020	FWF Spezialforschungsbereiche (SFB)	FWF	Dinges et al.	Propensity Score Matching, bibliometrisch	2
2021	Öffentliche F&E-Ausgaben / FFG-Förderungen	BMDW	Keuschnigg et al.	Propensity Score Matching	2
2022	Impact Innovation (Trial InnoCAP)	FFG	Landon & Hochreiter	RCT, zwei Treatmentarme, ohne Kontrollgruppe	5
2022	FFG Basisprogramm (Trial FeedS First)	FFG	Landon & Hochreiter	RCT mit Kontrollgruppe (Benchmark-Feedback)	5
2022	Social Crowdfunding (Trial SIM Crowd)	FFG	Landon & Hochreiter	RCT, Randomisierung auf dreiarmig, auf Individuen	5
2023	FFG Basisprogramm	ohne Auftrag	Widmann	IV-Schätzung (2SLS)	4
2023	FFG Programmlinie Kleinprojekt	FFG	Sardadvar & Ecker	Propensity Score Matching	2
2024	Öko-Scheck	BMK, FFG	Firpo	Randomisierte Vergabe (Losverfahren)	5

Anmerkung: SMS = Maryland Scientific Methods Scale. Level 2 = statistische Kontrolle, Level 3 = fortgeschrittenes quasi-experimentelles Design, Level 4 = Schwellenwert- bzw. Instrumentvariablen-Design, Level 5 = randomisiertes Experiment (RCT).

Quelle: eigene Erhebung KMU Forschung Austria auf Basis des fteval-Repositoriums und der Community-Umfrage, n = 17 Studien in 15 Publikationen.

4.2.1 | Zeitlicher Verlauf und Studiencharakteristika

Die Analyse der zeitlichen Verteilung verdeutlicht eine signifikante Zunahme kausalanalytischer Designs in der jüngeren Vergangenheit. Während auf die letzten zehn Jahre insgesamt 13 Publikationen entfallen, stammen lediglich drei Arbeiten aus den frühen 2000er Jahren. Diese drei frühen Studien (publiziert 2000, 2002 und 2004) können als methodische Pionierarbeiten im österreichischen Kontext gewertet werden, die bereits früh versuchten, kausalanalytische Ansätze in der FTI-Evaluierung anzuwenden. Auffällig ist jedoch die daran anschließende Diskontinuität: Im Zeitraum von 2004 bis 2016 konnte nur eine einschlägige Veröffentlichung identifiziert werden.

Hinsichtlich des Studiencharakters lässt sich der Korpus in drei Gruppen differenzieren:

1. **Klassische Ex-post-Evaluierungen (n=11):** Diese bilden die Mehrheit und untersuchen die Wirkung abgeschlossener Programmlaufzeiten.
2. **Themenspezifische Wirkungsanalysen (n=3):** Hierbei handelt es sich um breiter angelegte Wirkungsanalysen, die losgelöst von einem spezifischen Evaluierungszyklus oder einem spezifischen Programm ökonomische Effekte (z. B. Verhaltensadditionality) untersuchen.
3. **Policy-Experimente (n=3):** Randomisierte Pilotstudien, bei denen die Zufallszuteilung eigens als Experiment in den administrativen Prozess integriert wurde (RCTs); bislang beschränkt auf niederschwellige Mikro-Interventionen.

4.2.2 | Maßnahmencharakteristika und Untersuchungseinheiten

Die identifizierten Studien lassen sich anhand ihrer Untersuchungseinheiten und der Konstruktion des kontrafaktischen Szenarios in drei Gruppen gliedern:

Maßnahmen der angewandten Forschungsförderung

Der größte Teil der identifizierten Studien bezieht sich auf Maßnahmen der anwendungsorientierten Forschung, die Förderungen an Unternehmen und Forschungseinrichtungen vergeben. Zu den evaluierten Maßnahmen zählen hierbei sowohl themenoffene Förderprogramme (z. B. *FFG Basisprogramme*, *Frontrunner Programm*) als auch technologiespezifische Schwerpunkte (z. B. *ITF-Schwerpunkt FlexCIM*, *Sonderaktion Holzforschung*) sowie Kompetenzzentrenprogramme (*Kplus*, *K_ind/K_net*). Die Operationalisierung der Kausalanalyse erfolgt für diese Maßnahmen fast ausschließlich auf Unternehmensebene. Zur Bildung des kontrafaktischen Szenarios werden geförderte Unternehmen mit einer Kontrollgruppe verglichen. Diese besteht aus abgelehnten Antragsteller:innen, statistischen Zwillingunternehmen, die keinen Antrag gestellt haben, sowie in einem Fall aus geförderten Unternehmen eines vergleichbaren deutschen Programms. Diesem Cluster sind auch themenspezifische Wirkungsanalysen zuzuordnen (z. B. zur Verhaltensadditionality von Forschungsförderung). Eine analytische Ausnahme bildet hierbei jedoch die Untersuchung von Beschäftigungseffekten: Hier wurde neben der Vergleichsgruppensituation zwischen geförderten und nicht geförderten Unternehmen eine zweite Situation auf Personenebene umgesetzt, um die Wirkungen der Innovationsförderung auf einzelne Beschäftigte zu untersuchen.

Maßnahmen der Wissenschaftsförderung

Eine zweite Gruppe umfasst Förderungen im akademischen Bereich, repräsentiert durch die Evaluierungen der *International Programmes* und der *Spezialforschungsbereiche* des Wissenschaftsfonds (FWF) sowie des Programms *Sparkling Science*. Im Gegensatz zur Unternehmensförderung verschiebt sich die Operationalisierung der Kausalanalyse hier von der Institution auf die Ebene des einzelnen Forschungsprojekts oder der Publikation. Bei den FWF-Programmen fungieren reguläre Stand-alone-Projekte ohne die spezifische Zusatzmaßnahme als Vergleichsgruppe, um den Mehrwert des speziellen Förderinstruments zu isolieren. Beim Programm *Sparkling Science* wurde ein Matching-Verfahren auf Publikationsebene angewandt: Hier dienten wissenschaftliche Publikationen aus demselben

Fachbereich, derselben Institution und demselben Jahr als Kontrollgruppe, um die wissenschaftliche Performance zu kontrastieren.

Administrative Interventionen und Mikro-Maßnahmen

Eine dritte Kategorie bilden spezifische, niederschwellige Interventionen, die weniger die Forschungstätigkeit selbst, sondern den Antragsprozess oder Beratungsleistungen adressieren (z. B. *Innovationsschecks*, *Feedback-Schleifen*, *Crowdfunding-Incentivierung*). Bei diesen Maßnahmen, die im Rahmen von Policy-Experimenten untersucht wurden, entspricht die Untersuchungseinheit dem antragstellenden Akteur zum Zeitpunkt der Transaktion. Die Zuweisung zur Vergleichsgruppensituation (Treatment- und Kontrollgruppe) erfolgte hierbei randomisiert direkt im administrativen Prozess der Antragstellung.

4.2.3 | Analyse der methodischen Güte

Die Bewertung der Studien erfolgte anhand der *Maryland Scientific Methods Scale* (SMS), die als Maßstab für die interne Validität dient. Sie klassifiziert, wie robust das gewählte Forschungsdesign gegen Störfaktoren und alternative Erklärungen ist. Die Analyse zeigt, dass von den 17 identifizierten Arbeiten lediglich sechs den methodischen Mindeststandard für belastbare kausale Evidenz (Level 3 oder höher) erfüllen. Der Großteil der Evaluierungen verbleibt auf Level 2, was bedeutet, dass zwar statistische Kontrollverfahren angewandt wurden, die Belastbarkeit kausaler Schlussfolgerungen jedoch sehr eingeschränkt ist.

Level 2 Statistische Kontrolle (Matching und Regression)

Die größte Gruppe der Studien (n=11) befindet sich auf Level 2. Methodisch lassen sich hierbei zwei unterschiedliche Herangehensweisen differenzieren:

- **Matching-Verfahren (n=7):** Die Mehrheit dieser Gruppe setzt auf Matching-Ansätze (insb. Propensity Score Matching), um die Vergleichbarkeit zwischen geförderten und nicht-geförderten Einheiten ex post herzustellen. Hierbei wird für jedes geförderte Unternehmen (oder Projekt) ein „statistischer Zwilling“ aus der Kontrollgruppe gesucht, der in zentralen Eigenschaften (z. B. Branche, Unternehmensgröße, Gründungsjahr) identisch ist. Ziel ist es, die Unterschiede in den beobachtbaren Charakteristika zwischen den Gruppen zu minimieren, um den reinen Fördereffekt zu isolieren.
- **Panel- und Regressionsmodelle (n=4):** Vier Studien verzichten auf ein explizites Matching und nutzen stattdessen ökonometrische Regressionsmodelle. Hierbei werden Verfahren wie Fixed-Effects-Panelregressionen oder GMM-Schätzer angewandt. Der Fördereffekt wird geschätzt, indem der Einfluss anderer Kovariaten rechnerisch kontrolliert wird. Panel-Modelle haben hierbei den Vorteil, dass sie zumindest zeitkonstante Heterogenität der Unternehmen berücksichtigen können.

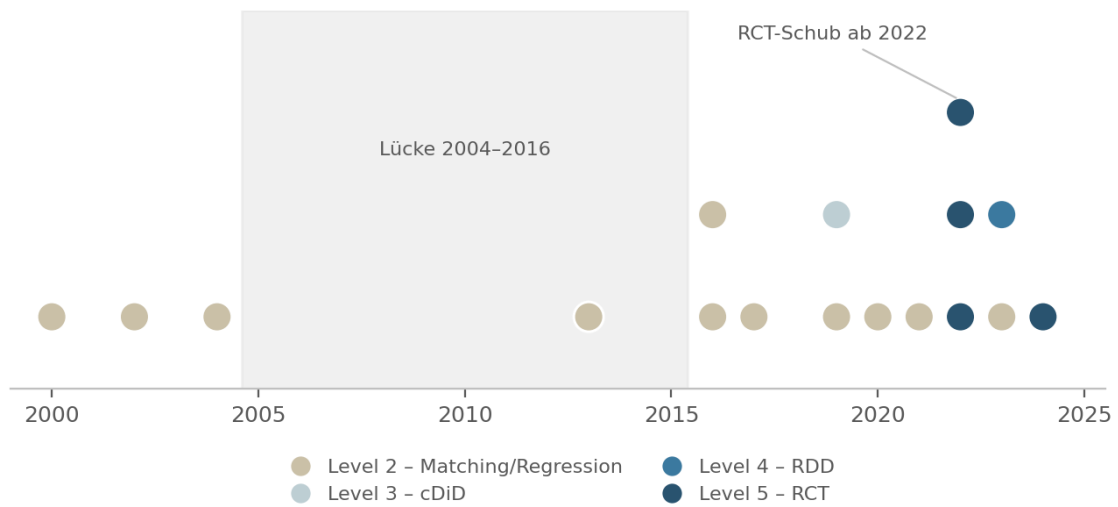
Das zentrale Defizit beider Level-2-Ansätze liegt in der fehlenden Kontrolle unbeobachteter, zeitvariabler Einflussfaktoren. Dazu zählen beispielsweise intrinsische Innovationsneigung, die Managementqualität oder die strategische Marktorientierung. Diese sind in den administrativen Daten nicht enthalten. Da diese Faktoren oft positiv mit der Wahrscheinlichkeit einer Förderzusage oder auch Förderbewerbung korrelieren, besteht das Risiko einer Verzerrung der Ergebnisse durch unbeobachtete Variablen (Omitted Variable Bias): Der gemessene Effekt ist folglich nicht rein kausal auf die Förderung zurückzuführen, sondern durch unbeobachtete Faktoren verzerrt, was zu einer systematischen Über- oder Unterschätzung der tatsächlichen Programmwirkung führen kann.

Level 3–5 Robuste Designs und experimentelle Ansätze

Jene sechs Studien, die den methodischen Mindeststandard der Maryland-Skala erfüllen, wenden Identifikationsstrategien an, die spezifisch darauf ausgelegt sind, auch diese unbeobachteten Störfaktoren zu neutralisieren.

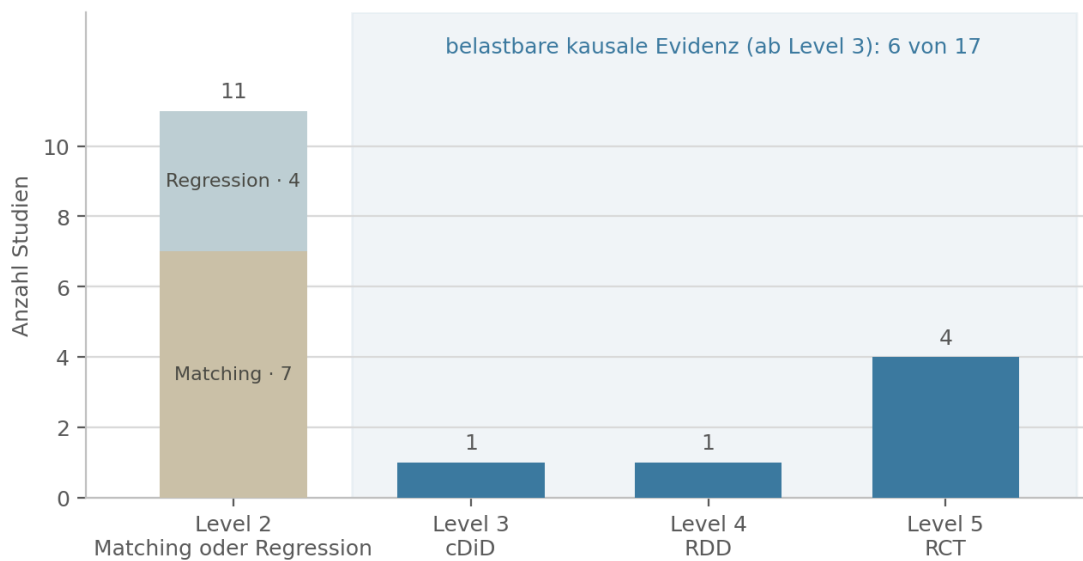
- **Quasi-Experimentelle Designs (Level 3 & 4):** Zwei Studien nutzen fortgeschrittene ökonometrische Ansätze zur Isolierung von Selektionseffekten:
 - **Conditional Difference-in-Differences (cDiD, Level 3) (n=1):** Eine Evaluierung im Bereich der Regionalförderung wendet diesen hybriden Ansatz an, der die Vorteile von Matching-Verfahren und Panel-Analysen kombiniert. Während klassische Fixed-Effects-Modelle (Level 2) zwar bereits zeitkonstante, unbeobachtete Störfaktoren herausrechnen, geht das cDiD einen entscheidenden Schritt weiter: Durch ein vorgeschaltetes Matching wird sichergestellt, dass die verglichenen Gruppen auch hinsichtlich ihrer beobachtbaren Merkmale strukturell ähnlich sind. Diese Kombination erhöht die Robustheit der Schätzung signifikant, da die für das DiD notwendige „Parallel-Trend-Annahme“ in einer gematchten Stichprobe weitaus plausibler ist als beim bloßen Vergleich unebener Gruppen.
 - **Instrumentvariablen-Schätzung (RDD, Level 4) (n=1):** Die methodisch anspruchsvollste quasi-experimentelle Arbeit setzt an der internen technischen Bewertung von Anträgen im FFG-Basisprogramm an. Ein förmlicher Schwellenwert existiert dort nicht, die Bewilligungswahrscheinlichkeit steigt jedoch in einem schmalen Punktbereich sprunghaft an, weil Anträge ab einer bestimmten Bewertung regelmäßig zur Förderung empfohlen werden. Dieser nichtlineare Anstieg dient als Instrument, während der lineare Zusammenhang zwischen Bewertung und Ergebnis separat kontrolliert wird – eine Identifikationsstrategie, die der Logik eines Fuzzy-RDD folgt, ohne dessen Voraussetzung eines definierten Cut-offs. Der Ansatz kommt ohne die Annahme paralleler Trends aus, auf der das cDiD beruht, und ist insofern gegenüber zeitvariablen Störfaktoren unempfindlich. Er verlagert die Identifikationslast allerdings auf zwei ebenfalls nicht prüfbare Annahmen: dass die Projektqualität im fraglichen Punktbereich nicht ebenso sprunghaft ansteigt, und dass die funktionale Form des direkten Zusammenhangs korrekt spezifiziert ist. Der geschätzte Effekt gilt zudem nur lokal für Anträge im Bereich des Anstiegs.
- **Experimentelle Designs (Level 5) (n=4):** Die höchste methodische Güte weisen vier Studien mit randomisiertem Design auf. Drei davon sind eigens angelegte Policy-Experimente; die vierte ist eine Ex-post-Evaluierung, die eine aus administrativen Gründen eingeführte Randomisierung nachträglich für die Wirkungsschätzung nutzt (Öko-Scheck). Durch die randomisierte Zuteilung der Intervention (Randomized Controlled Trial, RCT) werden Selektionsverzerrungen designbedingt ausgeschlossen. Da Treatment- und Kontrollgruppe rein zufällig gebildet werden, sind sie im Erwartungswert hinsichtlich beobachtbarer und unbeobachteter Merkmale strukturell identisch. Messbare Unterschiede in den Zielgrößen lassen sich somit eindeutig kausal auf die Maßnahme zurückführen. Diese Designs stellen im österreichischen Kontext jedoch eine noch junge Entwicklung dar (alle publiziert ab 2022) und beschränken sich bislang auf administrative Mikro-Interventionen.

Abbildung 2 | Kausalanalytische Studien im Zeitverlauf und Maryland Scientific Methods Scale



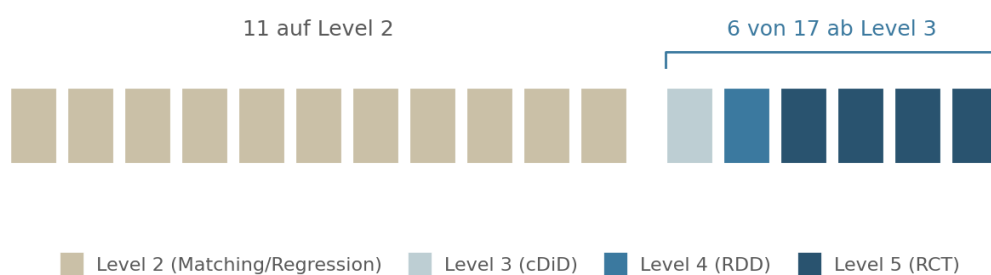
Quelle: eigene Darstellung auf Basis der kodierten Studien

Abbildung 3 | Kausalanalytische Studien nach Maryland Scientific Methods Scale



Quelle: eigene Darstellung auf Basis der kodierten Studien

Abbildung 4 | Kausalanalytische Studien nach Maryland Scientific Methods Scale (Version 2)



Quelle: eigene Darstellung auf Basis der kodierten Studien

Diese Erhebung zeigt: 17 Evaluierungen mit kausalanalytischen Methoden sind eine signifikante, aber noch begrenzte Zahl. Die nächste Frage lautet: Warum? Welche Bedingungen begünstigen oder hemmen solche Designs? Die folgenden Kapitel analysieren diese strukturellen, institutionellen und marktbezogenen Rahmenbedingungen basierend auf Interviews und Dokumentenanalyse.

5 | Kontextbedingungen und Methodenwahl kausalanalytischer Wirkungsanalysen

Das nachfolgende Kapitel untersucht, welche Rolle kausalanalytische Wirkungsanalysen im österreichischen Forschungs-, Technologie- und Innovationssystem (FTI) spielen, welche Bedingungen hierfür herrschen und welche Faktoren die Wahl entsprechender Methoden in der Evaluierungspraxis prägen. Zunächst werden auf Basis relevanter Literatur zentrale Kontextbedingungen diskutiert, gefolgt von einer Zusammenschau von Befunden aus Interviews mit Vertreter:innen der österreichischen FTI-Szene. Den Abschluss bilden eine Zusammenfassung und ein Fazit der Ergebnisse.

5.1 | Struktureller und institutioneller Kontext

Wirkungsanalysen und damit verbundene Ansätze können alleinstehend durchgeführt werden, sind aber oft Teil von Evaluierungen. Diese entstehen im österreichischen FTI-Bereich in einem Zusammenspiel aus formalen Vorgaben, verfügbaren Daten, institutionellen Routinen und marktbezogenen Rahmenbedingungen. Diese Faktoren bestimmen nicht nur, ob Programme, Maßnahmen oder Initiativen evaluiert werden, sondern auch, welche methodischen Zugänge im konkreten Fall realistisch erscheinen. Auf diese Aspekte wird nachfolgend kurz eingegangen.

5.1.1 | Formaler Rahmen und Datenzugang

Der formale Rahmen für Evaluierungen und vergleichbare Untersuchungen im FTI-Bereich wird vor allem durch das Bundeshaushaltsrecht⁹, ressortspezifische Vorgaben und förderbezogene Anforderungen geprägt. Mit der Einführung der wirkungsorientierten Haushalts- und Verwaltungsführung haben Evaluierungen auf Bundesebene an Bedeutung gewonnen (Pichler & Steyer, 2017). Die entsprechenden Vorgaben tragen dazu bei, dass relevante Programme und Maßnahmen regelmäßig evaluiert werden. Sie lassen aber offen, wie streng oder wie detailliert Effekte und Wirkungen gemessen werden sollen. Dadurch bleibt ein gewisser Spielraum: Kontrafaktische oder andere kausalanalytische Verfahren werden in den Vorgaben nicht ausdrücklich gefordert.

Die über die Jahre gewachsene Struktur des österreichischen FTI-Systems, das durch eine Vielzahl an Programmen, Maßnahmen und Förderlogiken geprägt ist (Biegelbauer et al., 2020, OECD, 2018, Streicher, 2017), spielt zusätzlich eine wichtige Rolle. In diesem Kontext dienen Evaluierungen weniger häufig einer abschließenden Wirkungsanalyse, sondern eher der Begleitung, Fortschreibung oder Anpassung laufender Aktivitäten. Es sind dann oft Zwischenevaluierungen bzw. ähnlich gelagerte Untersuchungen, die stärker auf Umsetzungsfragen, aktuelle Zielerreichung und Steuerungsbedarf ausgerichtet sind als auf eine (kausale) Wirkungsbestimmung. Das reduziert nicht grundsätzlich die Relevanz methodisch anspruchsvollerer Verfahren, begrenzt aber vielfach deren praktische

⁹ Vgl. insbesondere die Bestimmungen des BHG 2013 zur wirkungsorientierten Folgenabschätzung und internen Evaluierung sowie die Bestimmungen der ARR 2014 zur Gestaltung und Evaluierung von Förderungsprogrammen; ergänzend sind, abhängig vom jeweiligen Evaluierungsgegenstand, das Forschungs- und Technologieförderungs-gesetz (FTFG), das Forschungsfinanzierungsgesetz (FoFinaG) und das Forschungsorganisationsgesetz (FOG) relevant.

Anschlussfähigkeit, wenn Programme ressourcenmäßig klein und heterogen sind oder laufend angepasst werden.

Eine zentrale Voraussetzung für kontrafaktische Wirkungsanalysen ist die Verfügbarkeit geeigneter Daten. In Österreich war dieser Bereich lange durch fragmentierte Datenbestände, begrenzte Verknüpfbarkeit und institutionelle Zugangsbarrieren geprägt (Schwarz et al., 2020; Stampfer & Strassnig, 2020). Evaluierungen im FTI-Bereich greifen häufig auf Förderdaten, Monitoringinformationen und eigene empirische Erhebungen, insbesondere Befragungen oder Interviews, zurück. Solche Daten können für viele Evaluierungsfragen passgenau sein, stoßen aber dort an Grenzen, wo Vergleichsgruppen, längere Beobachtungszeiträume oder Verknüpfungen mit Register- und Administrativdaten notwendig wären.

Österreichische Förderagenturen, wie die FFG oder aws, verfügen über interne Datenbestände, die für Monitoring, programmspezifische Auswertungen, aber auch für Wirkungsanalysen relevant sein können. Dazu zählen etwa Informationen zu Förderprojekten, Volumina, thematischen/inhaltlichen Schwerpunkten und Hintergrundinformationen zu beteiligten Organisationen bzw. Unternehmen. Diese Daten sind jedoch meist für operative Zwecke angelegt und daher nicht immer unmittelbar für komplexere Wirkungsanalysen geeignet. Für kontrafaktische Analysen ist daher zunächst zu klären, ob die Förderdaten mit weiteren Datenquellen verknüpft werden können. Darüber hinaus müssen Programmgestaltung, Fallzahlen und Beobachtungszeiträume die Bildung belastbarer Vergleichsgruppen ermöglichen.

Mit dem Austrian Micro Data Center (AMDC) der Statistik Austria haben sich die Voraussetzungen für datenintensive Analysen zuletzt verbessert. Das AMDC bietet einen geregelten Onlinezugang zu Mikrodaten für wissenschaftliche Zwecke und stellt damit eine zentrale Infrastruktur für die Nutzung von Statistik-, Register- und Administrativdaten dar (Statistik Austria, AMDC). Damit verbessert das AMDC die technischen und rechtlichen Möglichkeiten datenbasierter Analysen, macht kontrafaktische Designs aber nicht automatisch zum Standard.

5.1.2 | Evaluierungskultur und -praxis

Neben formalen Vorgaben prägen institutionelle Routinen und Erwartungen der Auftraggebenden, etwa der Programmeigner:innen (z. B. Ministerium), die praktische Ausgestaltung von Evaluierungen in Österreich. Auftraggeber:innen entwickeln im Laufe der Zeit Vorstellungen darüber, was eine Evaluierung leisten kann oder soll, welche Fragestellungen als interessant und/oder relevant und welche Methoden tragfähig erscheinen. Diese Erwartungen sind häufig durch frühere Evaluierungen, etablierte Ausschreibungsroutinen und eine pragmatische Verwaltungskultur geprägt (Streicher, 2017).

Unter Berücksichtigung der existierenden formalen Rahmenbedingungen mag es nicht verwundern, dass Evaluierungen im FTI-Bereich als fester Bestandteil administrativer Praxis wahrgenommen werden. Programmeigner:innen und umsetzende Agenturen verstehen sie als reguläres Instrument der Steuerung, Rechenschaftslegung und Weiterentwicklung. Diese Routinisierung trägt zur Stabilität der Evaluierungspraxis bei, kann aber zugleich dazu führen, dass Evaluierungen eher als etablierter Prozess denn als Ausgangspunkt grundlegender methodischer oder strategischer Neuorientierung behandelt werden (OECD, 2018; Streicher, 2017). Ergebnis sind dann häufig standardisierte Leistungsbeschreibungen, die bewährte Fragestellungen und Designs wiederholen und nur begrenzten Raum für methodisch aufwändigere Verfahren lassen.

Die österreichische FTI-Evaluierungspraxis ist insgesamt stark durch methodische Offenheit und ein pluralistisches Evaluierungsverständnis geprägt. Eine diesbezüglich besondere Rolle spielt die Plattform fteval, die sich seit vielen Jahren für eine professionelle Evaluierungspraxis im FTI-Bereich in Österreich einsetzt. Sie gibt Standards heraus, organisiert Austausch und fördert ein gemeinsames Verständnis von Qualität. Dabei betont die Plattform methodische Offenheit und legt Wert darauf, unterschiedliche

Zugänge sichtbar zu machen. Anspruchsvollere kausalanalytische Verfahren können Teil eines Evaluierungsdesigns (für FTI in Österreich vgl. Kohlweg, 2019) sein, müssen aber zur Fragestellung, zur Datenlage und zu den institutionellen Rahmenbedingungen passen.

5.1.3 | Evaluierungsmarkt und Vergabepaxis

Die konkrete Methodenwahl wird schließlich auch durch die Struktur des österreichischen FTI-Evaluierungsmarktes beeinflusst. Dieser Markt ist klein und spezialisiert, zugleich aber wettbewerbsintensiv. Eine einschlägige Untersuchung zur Marktsituation im Bereich der FTI-Evaluierung in Österreich aus dem Jahr 2020 zeigte, dass Evaluierungen in einem spezifischen Zusammenspiel aus fachlicher Spezialisierung, begrenzter Anbieterstruktur und öffentlicher Beauftragung stattfinden (Streicher et al., 2020). Die Vergabe von Evaluierungen folgt der Logik öffentlicher Auftragsvergabe. Viele Evaluierungen werden mit begrenzten Budgets und kurzen Laufzeiten ausgeschrieben oder vergeben. In der Praxis beeinflussen daher nicht nur fachliche Erwägungen die Wahl des Designs, sondern auch Budgetrahmen, Fristen, Verfahrensart und die erwartete Umsetzbarkeit. Anspruchsvollere Designs, insbesondere kontrafaktische Wirkungsanalysen, erfordern jedoch frühzeitige Planung, geeignete Datenzugänge und längere Beobachtungszeiträume. Unter knappen finanziellen und zeitlichen Bedingungen werden solche Designs daher seltener nachgefragt oder angeboten.

Für Auftragnehmer:innen entsteht daraus ein Spannungsverhältnis zwischen methodischem Anspruch und wirtschaftlicher Machbarkeit. Einerseits können kontrafaktische Verfahren ein fachliches Qualitätsmerkmal darstellen. Andererseits müssen Angebote so kalkuliert werden, dass sie im Wettbewerb bestehen und innerhalb der verfügbaren Ressourcen umsetzbar bleiben. Für Auftraggeber:innen stellt sich umgekehrt die Frage, ob der erwartete zusätzliche Erkenntnisgewinn den höheren Aufwand rechtfertigt und ob die notwendigen Daten- und Abstimmungsprozesse innerhalb des vorgesehenen Evaluierungsrahmens realistisch sind.

Die beschriebenen Markt- und Vergabebedingungen wirken damit indirekt auf die methodische Praxis zurück. Sie setzen Anreize für Designs, die kalkulierbar, anschlussfähig und innerhalb kurzer Zeiträume realisierbar sind. Kontrafaktische Wirkungsanalysen erscheinen unter solchen Bedingungen vor allem dann machbar, wenn sie bereits bei der Programmentwicklung mitgedacht werden, die Datenvoraussetzungen geklärt sind und Auftraggeber:innen bereit sind, den zusätzlichen Aufwand in Ausschreibung, Budget und Zeitplan abzubilden.

5.2 | Befunde aus den Interviews

Während die vorherige Analyse die formalen und administrativen Rahmenbedingungen kurz skizziert, ermöglichen qualitative Interviews mit Auftraggeber:innen, Evaluierenden und Förderinstitutionen einen Blick auf die Motivationen, Wahrnehmungen und praktischen Entscheidungsprozesse, die über Methodenwahl entscheiden. Die nachfolgenden Befunde beruhen auf insgesamt 15 Interviews mit österreichischen Akteur:innen, die mit Evaluierungen im nationalen FTI-Bereich befasst sind und/oder deren Rahmenbedingungen mitprägen. Das Sampling folgte zwei Zugängen: Einerseits wurden gezielt Personen mit Erfahrung bzw. Expertise in kausalanalytischen Wirkungsanalysen angesprochen; andererseits wurden Organisationen berücksichtigt, die im österreichischen FTI-System regelmäßig mit Evaluierungen und Förderprogrammen zu tun haben. Die Stichprobe umfasst daher viele einschlägig informierte Perspektiven, jedoch nicht durchgehend Personen mit spezifischer Erfahrung in kausalanalytischen Ansätzen. Die Interviews wurden thematisch ausgewertet. Die Kategorien orientierten sich am Erkenntnisinteresse der Studie und wurden anhand wiederkehrender Muster im Material verdichtet.

In der nachfolgenden Auswertung werden drei Gruppen unterschieden: Auftraggeber:innen von Evaluierungen und damit verbundene Intermediäre wie Agenturen, Auftragnehmer:innen von Evaluierungen sowie weitere Beratungs- und Umfeldakteur:innen. Letztere umfassen Vertreter:innen von Einrichtungen, die keine Evaluierungen beauftragen oder durchführen, jedoch im erweiterten FTI-Umfeld tätig sind. Diese Unterscheidung ermöglicht es, Gemeinsamkeiten und Unterschiede zwischen administrativer Beauftragungspraxis, evaluierender Umsetzungsperspektive und übergreifender fachlicher Einordnung sichtbar zu machen.

Tabelle 3 | Übersicht der Interviewpartner:innen

Gesamt	Auftraggeber und Intermediäre	Auftragnehmer bzw. Evaluator:innen	Beratungs- & Umfeldakteure
15	6	4	5

Die nachfolgende Ergebnisdarstellung folgt fünf thematischen Schwerpunkten. Zunächst geht es um das Verständnis und die Relevanz kausalanalytischer Wirkungsanalysen. Daran anschließend werden Erwartungen und Nutzen, Grenzen und Machbarkeit, akteursspezifische Perspektiven auf Methodenwahl und Einsatz sowie Fragen der Datenlage und methodischen Anforderungen behandelt.

5.2.1 | Verständnis und Relevanz

Bei **Auftraggeber:innen und Intermediären** ist der Begriff weitgehend präsent, wird aber unterschiedlich abgegrenzt. Zumeist wird er mit dem Einsatz von Vergleichsgruppen sowie mit dem Versuch assoziiert, kausale Zusammenhänge möglichst belastbar nachzuweisen. Der größere Teil in dieser Gruppe fasst kausalanalytische Wirkungsanalysen eng und verbindet sie primär mit quantitativen bzw. ökonometrischen Verfahren. Ein kleinerer Teil vertritt ein eher weites Verständnis, das auch qualitative Zugänge einschließt, sofern diese geeignet sind, kausale Schlüsse zu stützen; zugleich wird in dieser Gruppe betont, dass nicht die Methode an sich, sondern die Belastbarkeit und Nutzbarkeit der Schlussfolgerungen entscheidend sei.

Kausalanalysen werden von **Auftragnehmer:innen** von Evaluierungen als zentrales Aufgabenfeld, vereinzelt auch als zentraler „methodischer Referenzpunkt“ oder auch „Goldstandard“, beschrieben. Dabei wird allerdings davor gewarnt, den Begriff auf kontrafaktische Designs oder rein quantifizierbare Effekte zu verengen. Es wird betont, dass je nach Fragestellung unterschiedliche methodische Zugänge geeignet sein können. Ökonometrische und datengetriebene Modelle spielen eine wichtige Rolle, können aber durch theoriegeleitete oder qualitative Ansätze ergänzt werden, wenn relevante Wirkungsdimensionen anders nicht angemessen erfasst werden.

Beratungs- und Umfeldakteur:innen verbinden kausalanalytische Ansätze zunächst ebenfalls häufig mit quantitativen Verfahren und statistisch nachweisbaren Zusammenhängen. Gleichzeitig wird auch in dieser Gruppe ein breiteres Verständnis sichtbar. Kausalanalysen werden nicht als alleiniger Erkenntniszugang verstanden, sondern als ein Element in einem umfassenderen Evidenzsystem. Entscheidend ist, ob die Analyse dazu beiträgt, Wirkungszusammenhänge besser zu verstehen und evidenzbasierte Politikgestaltung zu unterstützen. Weniger explizit, aber dennoch deutlich, erfolgt in dieser Gruppe der Verweis auf den mit solchen Ansätzen verbundenen Aufwand.

Die Relevanz kausalanalytischer Wirkungsanalysen wird über alle Gruppen hinweg grundsätzlich anerkannt. Sie nimmt im Kontext wachsender Wirkungsorientierung, steigenden Budget- und Effizienzdrucks in Politik und Verwaltung sowie internationaler Entwicklungen zu. Für Österreich überwiegt jedoch die Wahrnehmung, dass kausalanalytische Wirkungsanalysen in einem strengen Sinn (also „quantitativ“) bislang nur begrenzt im FTI-Bereich eingesetzt werden. Die Anerkennung der

Relevanz ist jedoch nicht gleichbedeutend mit einer uneingeschränkten Anwendungserwartung. Kausalanalytische Verfahren erscheinen dann als sinnvoll, wenn Fragestellung, Wirkungslogik, Datenlage, Ressourcen und Evaluierungsdesign dies zulassen. Zu den diesbezüglichen Grenzen und Voraussetzungen siehe 5.2.3 | [Grenzen, Herausforderungen und Voraussetzungen](#).

5.2.2 | Erwartungen und Nutzen

Auftraggeber:innen und Intermediäre erwarten von kausalanalytischen Wirkungsanalysen in erster Linie belastbare und kommunizierbare Aussagen zur Wirkung von FTI-Maßnahmen. Besonders wichtig ist ihnen die Möglichkeit, Wirkungen in Zahlen zu erfassen und dadurch nach außen prägnanter und verständlicher darzustellen. Kausalanalytische Ergebnisse können so zur Legitimation von Maßnahmen beitragen, insbesondere gegenüber der Politik, aber auch der breiten Öffentlichkeit oder anderen Stakeholdern. Darüber hinaus ermöglichen sie eine bessere Einschätzung der Effizienz einer Maßnahme, indem sie geförderte und nicht geförderte Vorhaben bzw. Entwicklungen gegenüberstellen.

Ein zentrales Nutzenversprechen besteht in der Abschätzung von Additionalität und Mitnahmeeffekten. Die Interviewten verknüpfen Kausalanalysen mit der Frage, ob eine beobachtete Wirkung tatsächlich auf die Förderung zurückzuführen ist oder auch ohne Förderung eingetreten wäre. Gerade dieses kontrafaktische Erkenntnisinteresse unterscheidet kausalanalytische Ansätze aus Sicht vieler Befragter von rein deskriptiven Evaluierungszugängen. Gleichzeitig wird der Nutzen nicht unkritisch gesehen. Einige Vertreter:innen dieser Gruppe weisen darauf hin, dass die Qualität der Analyse entscheidend ist und Ergebnisse im FTI-Bereich wegen komplexer Rahmenbedingungen vorsichtig interpretiert werden müssen. Lernprozesse, steuerungsrelevante Informationen oder die konkrete Weiterentwicklung von Maßnahmen werden aus dieser Gruppe hingegen kaum ausdrücklich als Erwartung genannt. Der Nutzen wird stärker über Legitimation, Kommunikation und Nachweisbarkeit beschrieben.

Auftragnehmer:innen von Evaluierungen teilen die Erwartung belastbarer quantitativer Erkenntnisse im Großen und Ganzen, setzen aber stärker einen methodischen Vorbehalt. Auch aus ihrer Sicht liegt eine besondere Stärke kausaler Wirkungsanalysen in der quantitativen Bestimmung von Additionalität und Mitnahmeeffekten; zugleich betonen sie, dass das gesamte Wirkungsspektrum einer Maßnahme selten vollständig quantitativ erfassbar ist. Qualitative und theoriegeleitete Methoden bleiben notwendig, um Wirkungsdimensionen wie Prozesse, Vernetzung, institutionelle Veränderungen oder nicht intendierte Effekte erfassen zu können. Aus Sicht der Evaluator:innen entsteht ein Risiko, wenn Legitimationsdruck dazu führt, dass kausalanalytische Ergebnisse politisch oder kommunikativ stärker beansprucht werden, als es die Daten- und Kontextbedingungen zulassen. Der erwartete Nutzen kausaler Evidenz sollte daher immer an die methodische Belastbarkeit gekoppelt bleiben.

Beratungs- und Umfeldakteur:innen betonen besonders die Objektivierungsfunktion. Je belastbarer die Evidenz zu einer Maßnahme ist, desto schwieriger lassen sich ungünstige Befunde im politischen Diskurs relativieren. Kausalanalysen stärken evidenzbasierte Politikgestaltung und erhöhen die legitimierende Funktion von Evaluierungen; zugleich ordnen diese Akteur:innen kausalanalytische Analysen stärker in einen breiteren Erkenntnisrahmen ein. Sie sehen sie als ein wichtiges Element, aber nicht als alleinigen Zugang zur Bewertung von Forschungs-, Technologie- und Innovationspolitik. Im Unterschied zu den Auftraggeber:innen nennen Beratungs- und Umfeldakteur:innen explizit auch Lernprozesse und die Weiterentwicklung von Maßnahmen als wichtigen Nutzen. Besonders relevant erscheint dies dort, wo bisher wenig darüber bekannt ist, wie Forschungsförderung langfristig wirkt. Vorhandene Daten und Analysen erfassen häufig kurzfristige Wirkungen; jede belastbare Erkenntnis zu längerfristigen Effekten wird daher als wertvoll angesehen. Abschnitt 5.2.4 | behandelt, wie diese Erwartungen in Ausschreibungen, Methodenwahl und Entscheidungsprozesse übersetzt werden.

5.2.3 | Grenzen, Herausforderungen und Voraussetzungen

Über alle Interviewgruppen hinweg besteht weitgehend Einigkeit, dass kausalanalytische Ansätze insbesondere für viele klassische FTI-Förderprogramme grundsätzlich geeignet sein können. Wie **Auftraggeber:innen bzw. Intermediäre** betonen, gilt dies vor allem dort, wo Wirkungen auf Ebene von Organisationen, Unternehmen oder Projekten mit gut messbaren Indikatoren beschrieben werden können und wo relativ einfache Hypothesen zu Wirkungen einzelner Maßnahmen formulierbar sind. Für **Beratungs- und Umfeldakteur:innen** erscheint ein kausaler Nachweis von Additionalität und Wirkung besonders bei Instrumenten mit hohem Mitteleinsatz relevant. In diesem Zusammenhang wurde in den Interviews einige Male auf die steuerliche Forschungsförderung in Österreich verwiesen.

Grenzen für kausalanalytische Ansätze werden von allen Interviewgruppen dort ausgemacht, wo Fördermaßnahmen komplexe, langfristige oder systemische Wirkungen entfalten sollen. Im FTI-Bereich wirken häufig viele Faktoren gleichzeitig, was eine eindeutige Isolation einzelner Förderwirkungen erschwert. Besonders anspruchsvoll sind Förderprogramme mit wenigen Einzelprojekten, Cluster- und Koordinationsmaßnahmen sowie Programme, die stark auf Prozesse oder Vernetzung abzielen. In solchen Fällen kann eine Verengung auf kausale quantitative Effekte die Wirkungen nur unzureichend abbilden.

Eine zentrale methodische Herausforderung besteht in der Bildung geeigneter Vergleichs- oder Kontrollgruppen. Mit Blick auf z. B. abgelehnte Antragsteller:innen bei Förderprogrammen stellen sich **Auftraggeber:innen bzw. Intermediäre** die Frage, ob diese tatsächlich eine valide Vergleichsgruppe darstellen. Werden jedoch auch Organisationen ohne Antragstellung einbezogen, entstehen andere methodische und rechtliche Fragen; zugleich beschreiben **Auftragnehmer:innen** von Evaluierungen die Suche nach geeigneten Kontrollgruppen als kreativen Teil der kausalanalytischen Vorgehensweise, wobei dann je nach Kontext auch eine pragmatische Auswahl sinnvoll sein kann.

Neben methodischen Grenzen spielen Ressourcen eine zentrale Rolle. Kausalanalytische Wirkungsanalysen erfordern häufig umfangreiche Vorarbeiten: Genannt werden die Prüfung verfügbarer Daten, Datenaufbereitung, Modellkonzeption, Plausibilisierung und Sensitivitätsanalysen. **Auftragnehmer:innen** berichten, dass die Auswahl geeigneter Analyseverfahren oft erst nach einer genaueren Sichtung der Daten möglich ist. Die dafür notwendigen zeitlichen und finanziellen Spielräume sind in Evaluierungsaufträgen jedoch nicht immer vorgesehen. **Auftraggeber:innen und Intermediäre** verweisen in diesem Zusammenhang auf begrenzte Evaluierungsbudgets und Zeitdruck. Der zeitliche Aufwand wird differenziert bewertet: Sowohl qualitative als auch quantitative Verfahren können zeitaufwendig sein, sodass sich für kausalanalytische Analysen kein pauschaler zeitlicher Mehraufwand ableiten lässt. Entscheidend bleibt, ob der zusätzliche Erkenntnisgewinn den organisatorischen, technischen und finanziellen Aufwand rechtfertigt.

Beratungs- und Umfeldakteur:innen heben neben methodischen und praktischen Hürden auch kulturelle und institutionelle Barrieren hervor. Zwar habe sich die Offenheit der relevanten Akteur:innen gegenüber kausalanalytischen Verfahren in Österreich laut den Interviews erhöht, doch werden solche Ansätze im Design von Maßnahmen nicht systematisch mitgeplant. Häufig fehlen dafür eigene Budgets, klare Outcomes, explizite Wirkungslogiken oder frühzeitig geplante Vergleichsdesigns, etwa in Form von Vergleichs- oder Kontrollgruppen.

Eine weitere Barriere betrifft den Umgang mit den Ergebnissen selbst. Die Verdichtung von Wirkungen auf wenige Kennzahlen birgt laut den Interviews eine Gefahr: Werden komplexe Effekte einer Maßnahme auf einzelne quantitative Kennzahlen heruntergebrochen und allein diese in der Debatte aufgegriffen, geraten differenzierte oder längerfristige Wirkungen leicht aus dem Blick. Fallen die ausgewiesenen Werte zudem geringer als erwartet oder negativ aus, werden (vermeintliche) Defizite unmittelbar sichtbar und den verantwortlichen Akteuren direkt zugerechnet. Der damit verbundene politische Rechtfertigungsdruck kann die Bereitschaft verringern, kausalanalytische Verfahren frühzeitig

einzuplanen, deren Ergebnisse offen zu kommunizieren oder grundlegende Anpassungen einer Maßnahme vorzunehmen.

5.2.4 | Perspektive auf Methodenwahl und Einsatz

In Ausschreibungen und Leistungsbeschreibungen werden laut **Auftraggeber:innen und Intermediären** konkrete Methoden in der Regel nicht vorgegeben. Die Wahl der Methode wird primär den Expert:innen überlassen. Die im Zuge von Ausschreibungsverfahren eingeholten Ansätze werden in Bewertungs- und Auswahlprozessen geprüft. Eine Ausnahme stellt die laufende Evaluierung der Basisprogramme der FFG dar, die, von mehreren Interviewpartner:innen angeführt, explizit einen kausalanalytischen Ansatz vorsieht und von der auch neue Erkenntnisse für den Umgang mit AMDC-Daten erwartet werden.

Ein Teil der **Auftraggeber:innen und Intermediäre** berichtet zudem, dass mögliche Ansätze und praktikable Methoden bereits im Vorfeld von Ausschreibungen intern diskutiert werden. Leitfäden, Checklisten und Guidance-Dokumente kommen vereinzelt zum Einsatz, beziehen sich aber überwiegend auf die technische Abwicklung, vergaberechtliche Rahmenbedingungen und die Formulierung von Evaluierungsfragen, weniger bis gar nicht auf konkrete methodische Vorgaben. Spielräume für präzisere Anforderungen bestehen vor allem dort, wo die ausschreibende Stelle Zugriff auf relevante Daten hat bzw. im Vorfeld weiß, was möglich ist, und Budget- sowie Datenanforderungen eigenständig gestalten kann.

Auftragnehmer:innen bzw. Evaluator:innen beobachten, dass kausalanalytische Ansätze im europäischen Kontext zunehmend zum Standard werden und in Ausschreibungen fallweise konkret eingefordert werden. Auf nationaler Ebene wird eine solche Entwicklung nicht wahrgenommen. Gleichzeitig besteht der Eindruck, dass in Österreich Angebote mit kausalanalytischen Elementen häufig positiv aufgenommen werden – auch wenn nicht ausdrücklich verlangt – wenn sie plausibel begründet sind. Die für die Durchführung von Kausalanalysen notwendigen Kompetenzen werden in den Einrichtungen der befragten Auftragnehmer:innen grundsätzlich als vorhanden beschrieben.

Bei der Auswahl von Evaluierungsteams achten **Auftraggeber:innen und Intermediäre** auf fachliche Qualifikation, methodische Erfahrung, Glaubwürdigkeit und teilweise den Ruf der Anbieter:innen; zugleich wird die eigene interne Kompetenz zur vertieften Beurteilung kausaler Methoden von einigen Interviewpartner:innen als begrenzt eingeschätzt. Daraus ergibt sich ein Bedarf an methodischer Orientierung und Kompetenzaufbau, um Vorschläge angemessen beurteilen und kausalanalytische Optionen früh im Auswahlprozess einschätzen zu können.

Während **Auftragnehmer:innen bzw. Evaluator:innen** überwiegend den Eindruck haben, dass Auftraggeber:innen ihre Bedarfe meist relativ klar formulieren können, orten **Beratungs- und Umfeldakteure** teilweise Unsicherheiten im Umgang mit Studienergebnissen und Evaluierungserwartungen. Sie verweisen auf begrenzte Datenkompetenz in politischen Entscheidungsprozessen und auf die Gefahr, dass bestimmte Ansätze vorschnell als nicht möglich angesehen werden. Gleichzeitig wird betont, dass politische Entscheidungsprozesse vor allem knappe, handlungsorientierte Schlussfolgerungen benötigen. Der Weg zu diesen Schlussfolgerungen muss methodisch solide sein; zugleich müssen die Ergebnisse für die Nutzung in politischen Entscheidungsprozessen verständlich aufbereitet werden.

Die Entscheidung für oder gegen kausalanalytische Ansätze hängt aus Sicht aller drei Gruppen zusammenfassend wesentlich von Zielsetzung und Erkenntnisinteresse der Evaluierung ab. Kausalanalytische Ansätze sollten vor allem dann in Betracht gezogen werden, wenn belastbare Evidenz zur Wirkung einer Maßnahme benötigt wird und deskriptive oder weniger aufwendige Zugänge nicht ausreichen. Gegen ihren Einsatz sprechen wiederum fehlende Machbarkeit, hoher Aufwand, begrenzte Budgets, kurze Beobachtungszeiträume, kleine Fallzahlen, fehlende Vergleichsgruppen, eingeschränkte Datenverfügbarkeit und politische Sensibilität möglicher Ergebnisse.

5.2.5 | Datenlage und methodische Anforderungen

Auftraggeber:innen und Intermediäre betonen, dass grundsätzlich zahlreiche relevante Datenquellen existieren. Besonders das AMDC wird vereinzelt als echter „Gamechanger“ beschrieben, weil es den Zugang zu Mikrodaten erleichtert bzw. erleichtern soll und neue Möglichkeiten der Verknüpfung unterschiedlicher Datenquellen eröffnet; zugleich wird darauf hingewiesen, dass sich die tatsächlichen Effekte dieser Infrastruktur erst über einen längeren Zeitraum zeigen werden und dass die Nutzung derzeit mit vergleichsweise hohen Kosten verbunden ist.

Neben dem AMDC werden weiters die F&E-Erhebung, die Transparenzdatenbank sowie Datenbestände zentraler Fördereinrichtungen wie FWF und FFG von **Auftraggeber:innen** als relevante Quellen genannt. Die F&E-Erhebung bietet als Vollerhebung einen umfassenden und regelmäßigen Einblick in die Forschungsaktivität von Unternehmen. Die Transparenzdatenbank ermöglicht es, nachzuvollziehen, welche unterschiedlichen Förderungen einzelne Organisationen in Anspruch nehmen. Daten von FWF und FFG sind für Evaluierungen wertvoll, wurden aber eher für administrative Zwecke konzipiert und sind daher nicht immer ohne Weiteres für vergleichende oder kausalanalytische Analysen nutzbar. Trotzdem bestehen laut Einschätzung der Interviewpartner:innen erhebliche Datenlücken. So fehlt eine öffentlich zugängliche, umfassende Projektdatenbank, und insbesondere bei kleineren Förderformaten sind Daten häufig unvollständig. Auch die systematische Erfassung von Projektergebnissen und Wirkungen wird insgesamt als verbesserungsbedürftig beschrieben.

Über alle drei Gruppen hinweg heben die Interviewpartner:innen die Verknüpfung von Förder-, Antrags-, Register- und Wirkungsdaten als eine zentrale Voraussetzung kausalanalytischer Designs hervor, welche aber zugleich eine der größten praktischen Hürden darstellt. Fehlende einheitliche Identifikationssysteme, institutionelle Fragmentierung und unterschiedliche Datenabgrenzungen erschweren die Zusammenführung. Auch institutionell getrennte Zuständigkeiten und Datenbestände zwischen verschiedenen Förderinstrumenten, etwa zwischen der steuerlichen Forschungsförderung und der direkten Förderung, erschweren eine analytisch sinnvolle Verknüpfung der Daten.

Auftragnehmer:innen bzw. Evaluator:innen beschreiben den Datenzugang trotz verbesserter Infrastruktur mittels AMDC als insgesamt schwierig und teuer. Evaluierende können Daten oft nicht eigenständig verknüpfen und sind auf institutionelle Prozesse angewiesen. Insbesondere für Längsschnittanalysen wären über die Zeit vergleichbare und auf Ebene einzelner Untersuchungseinheiten verknüpfbare Daten erforderlich. Eine stärkere Nutzung vorhandener administrativer Daten wäre aus Sicht einzelner Auftragnehmer technisch naheliegend und könnte zusätzliche Erhebungen bei Unternehmen, Forschungseinrichtungen und Fördernehmenden vermeiden. Sie wird jedoch durch rechtliche und institutionelle Hürden begrenzt. Gleichzeitig verweisen einige Interviewpartner:innen darauf, dass einer erweiterten Datennutzung mitunter auch kulturelle oder governancebezogene Vorbehalte gegenüberstehen könnten, insbesondere weil eine höhere Datentransparenz potenziell auch Leistungsvergleiche zwischen Institutionen erleichtern würde.

Beratungs- und Umfeldakteur:innen bewerten das AMDC ebenfalls als sehr wichtiges Werkzeug, verweisen aber auf Fragen der ökonomischen Effizienz, der personellen Ausstattung und der rechtlichen Umsetzbarkeit. Im internationalen Vergleich, etwa zu den Niederlanden oder skandinavischen und baltischen Ländern, wird Österreich teilweise wenig forschungsfreundlichkeit im Zugang zu Mikrodaten attestiert.

Im Hinblick auf einzelne Evaluierungen lautet ein wiederkehrender Befund von **Auftragnehmenden** wie auch **Beratungs- und Umfeldakteur:innen**, dass die Datenverfügbarkeit häufig erst im Kick-off oder im Verlauf einer Evaluierung geklärt wird. Dies kann zu erheblichen Problemen führen, wenn erwartete Datenzugänge (doch) nicht realisierbar sind. Daher: Datenbezogene Anforderungen müssen laut den Vorschlägen dieser Gruppen idealerweise bereits bei der Konzeption von Programmen, bei der Formulierung von Evaluierungsfragen und vor der Ausschreibung zu einer Evaluierung oder

Wirkungsanalyse mitgedacht werden. Langfristig erscheint es einigen Interviewpartner:innen zentral, Datenstrukturen so zu gestalten, dass zukünftige Evaluierungen systematisch ermöglicht werden. Dazu gehören Forschungsdatenmanagement, rechtliche Klärungen, Budgetierung, Datenharmonisierung, Pseudonymisierung und ein realistischer Abgleich zwischen Erkenntnisinteresse und Datenlage.

Als positive Entwicklungen werden über alle Interviewpartner:innen hinweg Initiativen wie eine geplante API des FWF, offene Dateninfrastrukturen wie OpenAlex sowie internationale Standardisierungs- und Pseudonymisierungsansätze, etwa im Rahmen von RoRI bzw. der Thunder Data-Plattform, genannt. Zudem wird auf das wachsende Potenzial von KI-Methoden verwiesen, beispielsweise zur Klassifizierung von Projekten und Ergebnissen oder zur Unterscheidung unterschiedlicher Innovationsarten. Diese Entwicklungen können die Datenlandschaft verbessern, ersetzen aber nicht die Notwendigkeit konsistenter Datenstrukturen und klarer Nutzungsregeln.

Rechtliche Rahmenbedingungen und Datenschutz werden in allen Interviewgruppen wiederum als zentrale Einschränkungen genannt. Die Interviewten beschreiben ein Spannungsfeld zwischen dem Bedarf an umfassender Datenverfügbarkeit für Evaluierungen und den Anforderungen an Datenschutz, Datensparsamkeit und institutionelle Vertraulichkeit. Besonders sensibel sind Daten zu abgelehnten Projekten, die für kontrafaktische Analysen hochrelevant wären („Was ist hier ohne die Förderung passiert?“). Pseudonymisierungsansätze im AMDC können hier nach Hinweisen einiger Befragter teilweise helfen, lösen aber nicht alle rechtlichen und institutionellen Fragen. Überhaupt ist zu klären, wie solche Daten von Beginn an in Systeme aufgenommen und weiter gepflegt werden können.

Schließlich wurden in allen drei Gruppen auch ethische Fragen angesprochen, insbesondere bei experimentellen oder quasi-experimentellen Designs. Laut einzelnen Interviewpartner:innen bleibt etwa offen, welche Informationsrechte Kontrollgruppen haben, wo Grenzen zulässiger Datennutzung liegen und wie Datenschutz, Transparenz und wissenschaftliches Erkenntnisinteresse angemessen ausbalanciert werden können. Diese Fragen sollten nicht erst im Einzelfall entschieden werden, sondern Teil einer breiteren Diskussion über Standards kausalanalytischer Evaluierungen im FTI-Bereich sein.

5.3 | Kontextbedingungen – Fazit

Das Interesse an kausalanalytischen und kontrafaktischen Wirkungsanalysen steigt im österreichischen FTI-System. Ihr Mehrwert wird insbesondere dort anerkannt, wo es um belastbare Aussagen zu Wirkung, Additionalität, Mitnahmeeffekten und Effizienz geht. Neben gut kommunizierbaren Ergebnissen wird jedoch auch die Relevanz einer transparenten Herleitung und Einordnung ihrer Aussagekraft betont.

In der Diskussion der Vorteile, Grenzen und Einsatzmöglichkeiten setzen die befragten Akteursgruppen unterschiedliche Schwerpunkte. Auftraggeber:innen und Intermediäre heben vor allem die Legitimations- und Kommunikationsfunktion kausaler Evidenz sowie Fragen der Machbarkeit und des Kosten-Nutzen-Verhältnisses hervor. Auftragnehmer:innen betonen die methodische Belastbarkeit, die Anforderungen an die Datenverfügbarkeit und die Grenzen der Aussagekraft. Beratungs- und Umfeldakteur:innen verweisen neben der Objektivierung politischer Diskussionen insbesondere auf Lernprozesse, die Weiterentwicklung von Maßnahmen und die Unterstützung evidenzbasierter Politikgestaltung. Das AMDC wird gruppenübergreifend als wichtige Verbesserung der Dateninfrastruktur eingeschätzt; der Datenzugang bleibt jedoch aufwendig und rechtlich wie institutionell anspruchsvoll.

Die derzeit geringe Verbreitung kontrafaktischer Designs ist daher nicht als grundsätzliche Ablehnung zu verstehen. Sie erklärt sich vielmehr aus institutionellen Routinen, administrativen Erwartungen und den konkreten Anforderungen an ihre Umsetzung. Auf Basis der Einschätzungen aus den Gesprächen erscheinen kausalanalytische Wirkungsanalysen besonders geeignet bei klar abgegrenzten Maßnahmen, einer ausreichenden Zahl an geförderten und vergleichbaren Untersuchungseinheiten, nachvollziehbaren Wirkungslogiken sowie belastbaren Datenzugängen. Voraussetzung dafür ist,

Outcomes, Interventionslogik, Datenverfügbarkeit, Identifikatoren, rechtliche Nutzungsbedingungen sowie Budget und Zeitplan frühzeitig zu klären. Wie in den Gesprächen betont wurde, sollten kausalanalytische Verfahren nicht losgelöst vom jeweiligen Kontext eingesetzt werden, sondern dort, wo sie zur Fragestellung passen und gegenüber alternativen Designs einen zusätzlichen, klaren Erkenntnisgewinn versprechen. In vielen Fällen liegt ihr besonderer Nutzen in der Kombination mit qualitativen, deskriptiven oder theoriegeleiteten Zugängen, um sowohl belastbare Aussagen zu Wirkungen als auch ein angemessenes Verständnis der jeweiligen Förderlogik zu ermöglichen.

Die Analyse der Kontextbedingungen offenbart ein komplexes Bild: Methodische Kompetenz wächst, politisches Interesse ist vorhanden, aber strukturelle, marktbezogene und finanzielle Hürden begrenzen den flächendeckenden Einsatz.

6 | Synthese und Handlungsempfehlungen

Aus der Verbindung der Erkenntnisse der Literaturanalyse, der Erhebung und Analyse bisheriger kontrafaktischer Evaluierungen sowie der Untersuchung der Stakeholderperspektiven und Rahmenbedingungen ergibt sich ein klares Bild der bisherigen Nutzung der Methode in der österreichischen FTI-Politik.

Mit 17 identifizierten Studien, die eine kontrafaktische Wirkungsanalyse auf österreichische FTI-Maßnahmen anwenden, bleibt die Methode ein Nischenphänomen – allerdings eines mit wachsender Bedeutung: 13 der 17 Arbeiten stammen aus den letzten zehn Jahren. Im internationalen Vergleich ist die Nutzung dabei durchaus beachtlich. Für Schweden identifizieren Collin, Sandström und Wennberg (2022) 11 kontrafaktische Designs und damit rund 10 % der von ihnen untersuchten innovationspolitischen Evaluierungen, für Deutschland Büchele et al. (2024) 14 Evaluierungen mit zumindest teilweise kausalen Designs (2009–2023) bzw. rund 17 % der untersuchten Studien; in der EU-Kohäsionspolitik liegt der Anteil im Politikfeld Forschung und Innovation bei lediglich rund 3 % (Santos & Coad, 2023). Gemessen an der Größe des österreichischen Evaluierungsmarkts fällt die Nutzung damit vergleichsweise hoch aus – wobei der direkte Vergleich der Anteile aufgrund unterschiedlicher Erhebungs- und Selektionsverfahren nur indikativ möglich ist. Zugleich zeigt die Bewertung anhand der Maryland Scientific Methods Scale, dass nur sechs der 17 Studien den Mindeststandard für belastbare kausale Evidenz (Level 3 oder höher) erreichen (eigene Erhebung, AP2, siehe Abschnitt 4).

Die begrenzte Anzahl an Evaluierungen mit kontrafaktischen Methoden lässt sich durch mehrere, teils zusammenhängende Faktoren erklären:

Ein erster Faktor sind die **kleinen Fallzahlen auf Programmebene**. Evaluierungen werden im österreichischen FTI-System mit wenigen Ausnahmen auf Programmebene durchgeführt. Programme weisen aufgrund begrenzter Budgets oder thematischer Eingrenzung vielfach nur eine aus statistischen Gesichtspunkten geringe Zahl geförderter Organisationen auf und ebenso eine begrenzte Anzahl vergleichbarer nicht geförderter Unternehmen. Hinzu kommt die dynamische Entwicklung der Programme: Nicht zuletzt aufgrund von Evaluierungen werden regelmäßig Anpassungen vorgenommen, die die Zahl jeweils vergleichbarer Förderfälle weiter verkleinern, da geförderte Organisationen unterschiedliche Förderbedingungen vorfinden und die Wirkung einer spezifischen Maßnahmenausgestaltung schwer zu isolieren ist. Aus demselben Grund sind internationale kausale Befunde nur eingeschränkt übertragbar. Österreichspezifische Studien sind weiterhin notwendig und müssen bei wesentlichen Änderungen des Programmdesigns wiederholt werden.

Eng damit verbunden ist die **Förderkumulation im dichten Unterstützungssystem**. Die österreichische Förderlandschaft ist gut ausgebaut, und es besteht ein breites Angebot an Förderungen im FTI-Bereich. Organisationen nehmen vielfach im selben Zeitraum mehrere Förderungen aus unterschiedlichen Programmen in Anspruch, was die Isolierung der Wirkung eines einzelnen Programms zusätzlich

erschwert. Dies gilt insbesondere, wenn nicht alle Daten zu Förderfällen zugänglich sind, wie bisher bspw. bei der Forschungsprämie, und daher unklar bleibt, wo und in welchem Umfang eine Organisation bereits Unterstützung erhalten hat.

Ein dritter Faktor sind die **langen Zeiträume bis zur Sichtbarkeit von Wirkungen in Sekundärdaten**. Kontrafaktische Ansätze sind auf Sekundärdaten angewiesen, da Förderprogramme keine Daten zu nicht geförderten Unternehmen erheben – geschweige denn über einen längeren Zeitraum vor und nach der Förderung. Diese Daten werden meist mit einer Verzögerung von bis zu zwei Jahren veröffentlicht; die Innovationserhebung der Statistik Austria (CIS) für 2024 erschien beispielsweise erst im Juli 2026. In Verbindung mit der Ausschreibungspraxis – Evaluierungen laufen selten länger als ein Jahr und werden bei Ex-post-Evaluierungen meist während oder unmittelbar nach der letzten Ausschreibung eines Programms beauftragt – ergibt sich ein sehr enges Fenster für beobachtbare Wirkungen. In der FTI-Politik werden zudem meist Langzeitwirkungen wie die Umsetzung von Innovationen angestrebt, mit denen frühestens ein bis zwei Jahre nach Projektabschluss zu rechnen ist. Dies hat auch einen förderrechtlichen Hintergrund: Projekte, die zu nahe an der Marktumsetzung stehen, dürfen nach EU-Beihilferecht nicht gefördert werden. Ein Rechenbeispiel verdeutlicht die Konsequenz: Bewilligt ein Programm erste Förderfälle 2020, beträgt die Projektlaufzeit zwei Jahre und vergehen danach weitere zwei Jahre bis zur Marktumsetzung, sollten sich erste Wirkungen im CIS 2024 abzeichnen – der Mitte 2026 erscheint. Vom Förderbeginn bis zur Sichtbarkeit der Wirkungen in der Sekundärstatistik vergehen somit rund 6,5 Jahre. Läuft ein Programm von 2020 bis 2024 und sind die letzten Projekte 2026 abgeschlossen, sind die Wirkungen aller Projekte erst Mitte 2030 in der Sekundärstatistik sichtbar.

Hinzu kommt der **eingeschränkte Zugang zu Mikrodaten**. Sekundärstatistiken müssen für kontrafaktische Analysen auf Organisationsebene vorliegen. Die beiden für den FTI-Bereich relevantesten Erhebungen werden von der Statistik Austria durchgeführt; auf Basis des Bundesstatistikgesetzes waren diese Daten auf Mikroebene bislang jedoch nur sehr schwer zugänglich und für Evaluierungen kaum nutzbar. Dies erklärt zum Teil auch die geringe kausale Validität vieler bisheriger kontrafaktischer Analysen.

Diesen Restriktionen stehen jedoch **positive Entwicklungen** gegenüber. In den letzten Jahren lässt sich ein klar erkennbarer Aufwärtstrend bei der Anwendung der Methodik beobachten, der von mehreren Entwicklungen getragen wird. Erstens fordern verschiedene Akteure die Nutzung vermehrt ein: die Europäische Kommission über das Beihilferecht, teilweise auch Rechnungshöfe sowie der akademische Diskurs in der angewandten Ökonomie im Zuge der „Credibility Revolution“ (Angrist & Pischke, 2010). Zweitens ist mit Randomized Controlled Trials ein im österreichischen Kontext neuer Ansatz für die Gestaltung und Entwicklung von FTI-Politikmaßnahmen aufgekommen: Seit 2022 wurden vier Policy-Experimente publiziert, die die randomisierte Zuteilung direkt in das Design administrativer Mikro-Interventionen integrieren. Die Gründung des Austrian Micro Data Center (AMDC) dürfte die positive Entwicklung in den kommenden Jahren weiter befeuern, insbesondere, wenn die Zugänglichkeit durch geringere Kosten sowie optimierte und schnellere Prozesse künftig weiter vereinfacht wird. Hinzu kommt die methodische Weiterentwicklung selbst: Synthetic-Control-Ansätze, Dynamic Matching und zunehmend Machine-Learning-gestützte Verfahren verbessern den Umgang mit kleinen Vergleichs- und Kontrollgruppen und dürften die Zahl belastbarer kontrafaktischer Analysen künftig weiter steigern.

Die **Stakeholderperspektiven** bestätigen dieses Bild. Die befragten Stakeholdergruppen bewerten den vermehrten Einsatz positiv. Auftraggeber:innen geben in Ausschreibungen in der Regel keine Methoden vor; die Methodenwahl liegt damit faktisch bei den Evaluator:innen. Auftraggeber:innen können jedoch die Voraussetzungen schaffen, unter denen kontrafaktische Designs überhaupt realistisch sind: über entsprechenden Datenzugang, Budgets und Zeithorizonte für Evaluierungen. Zugleich besteht über alle Akteursgruppen hinweg Konsens, dass differenzierte Evaluierungen ein breites Wirkungsspektrum abdecken und auch Aufschluss über Wirkungszusammenhänge geben müssen, was ein breites Methodenset einschließlich qualitativer und theoriebasierter Verfahren erfordert. Kontrafaktische Analysen können die Existenz und Stärke einer Wirkung isolieren, nicht aber den dahinterliegenden

Wirkungsmechanismus empirisch überprüfen. Sie stellen somit stets einen Teilaspekt einer Evaluierung dar, der unmittelbar von den anderen angewendeten Methoden profitiert, da sich Wirkungen besser interpretieren und validieren lassen.

Um Nutzung und Mehrwert kontrafaktischer Analysen in der österreichischen FTI-Evaluierung weiter zu steigern, werden folgende **Empfehlungen** gegeben. Sie richten sich spezifisch an die jeweilige Akteursgruppe.

Evaluator:innen

- Kompetenzaufbau durch Training, um die Methodenkompetenz einerseits breiter zu streuen und andererseits an der Spitze der methodischen Weiterentwicklung zu bleiben.
- Methodische Weiterentwicklungen gezielt für österreichische Kontextbedingungen einsetzen: Synthetic-Control- und Dynamic-Matching-Ansätze adressieren kleine Fallzahlen bzw. dauerhaft laufende Maßnahmen und damit genau die oben beschriebenen Engpässe.
- Level 3 der Maryland-Skala als Referenz-Mindeststandard anstreben, wo die Kontextbedingungen es zulassen; Identifikationsannahmen (z. B. Parallel-Trend-Annahme) transparent dokumentieren und mit Robustheitschecks absichern.

Auftraggeber:innen

- Kompetenzaufbau, um Vor- und Nachteile besser zu verstehen und die Anwendbarkeit in spezifischen Kontexten einschätzen zu können; Sensibilisierung für die richtige Interpretation der Ergebnisse, die durch eine verkürzte Darstellung mehr Schaden als Nutzen stiften können. Die Verfügbarkeit einer präzisen Zahl für eine bestimmte Wirkung darf nicht dazu führen, dass nicht oder nur schwer quantifizierbare Wirkungen als weniger relevant gelten oder aus dem Blick geraten. Messbarkeit ist kein Indikator für Bedeutung. Kontrafaktische Schätzwerte sollten daher stets mit der ihnen zugrunde liegenden Unsicherheit ausgewiesen und im Zusammenhang mit den übrigen Evaluierungsbefunden interpretiert werden, nicht als alleiniger oder abschließender Beleg einer Programmwirkung.
- Evaluierung bereits beim Programmdesign mitdenken und den potenziellen Einsatz von Kausalanalysen frühzeitig berücksichtigen: relevante Wirkungsindikatoren identifizieren und Daten frühzeitig sammeln.
- Kostenrealismus ist geboten, da kontrafaktische Analysen methodisch und datenseitig aufwendig und damit häufig kostenintensiver als andere Verfahren sind. Wo bei spezifischen Programmen ein methodischer Mehrwert zu erwarten ist, sollte dieser Aufwand bei der Budgetierung der Evaluierung eingeplant werden.
- Vollständigen Zugang zu allen Förderdaten gewährleisten, um die Effekte einzelner Programme isolieren zu können.

Umfeldakteur:innen und Datenbereitsteller:innen

- Datenzugang beim AMDC sollte verbessert und kostengünstiger werden, da die in Österreich vergleichsweise geringen Evaluierungsbudgets durch hohe Datenkosten belastet werden – was insbesondere dann nicht sinnvoll erscheint, wenn es sich faktisch um eine interne Weitergabe innerhalb der öffentlichen Hand handelt. Darüber hinaus sollten Zugangsprozesse optimiert und Bearbeitungszeiten verkürzt werden. Da die einzelnen Register derzeit für jedes Projekt einzeln verknüpft werden müssen, sollte eine kontinuierlich aktualisierte Verknüpfung FTI-relevanter Datensätze angestrebt werden. In Deutschland wird dies derzeit im Projekt EVALDAT erprobt, das eine dauerhafte Dateninfrastruktur für die Evaluierung von FTI-Maßnahmen aufbaut; durch den Mikrodatenzugang zu öffentlichen Registern hätte ein solcher Ansatz in Österreich noch deutlich mehr Potenzial.

- Eine differenzierte Nutzung der Ergebnisse durch Sensibilisierungsmaßnahmen ist geboten, weil kontrafaktische Analysen wesentliche Wirkungen auf einzelne Kennzahlen herunterbrechen und dadurch anfällig für eine undifferenzierte Nutzung im politischen Diskurs sind. Qualitative Erkenntnisse sollten dadurch nicht in den Hintergrund gedrängt, sondern die Ergebnisse in einem breiten, durch verschiedene Methoden analysierten Wirkungsspektrum verortet werden. Dies gilt für Interessenvertretungen und beratende Gremien in besonderem Maße, da sie Evaluierungsergebnisse in den wirtschaftspolitischen Diskurs übersetzen. Es gilt ebenfalls für die Politik selbst, die sie als Grundlage für Entscheidungen heranzieht.

7 | Literaturverzeichnis

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4), 69–85. <https://doi.org/10.1257/jep.15.4.69>.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. <https://doi.org/10.1515/9781400829828>.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3–30. <https://doi.org/10.1257/jep.24.2.3>.

Ardito, C., Barbaglia, L., Bernhofer, J., Berton, F., Martinez Santos, F., & Sasso, A. (2025). *Quality assessment of evaluation plans and reports in the area of State aid (EVALSA III): The 2024 activity report* (JRC141213). Publications Office of the European Union. <https://doi.org/10.2760/1600170>.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), 3–32. <https://doi.org/10.1257/jep.31.2.3>.

Benedictow, Andreas, Emil Cappelen Bjørn, Fernanda Winger Eggen, Marthe Norberg-Schulz, Marina Rybalka und Rolf Røtnes (2018): Evaluation of SkatteFUNN, Samfunnsøkonomisk analyse AS Report no. 18-2018. <https://kudos.dfo.no/documents/12299/files/12437.pdf>.

Biegelbauer, P., Hartmann, C., Polt, W., Wang, A. & Weber, M. (2020). *Mission-Oriented Innovation Policies in Austria – a case study for the OECD*. Graz/Wien. [10.22163/fteval.2020.493](https://doi.org/10.22163/fteval.2020.493).

Bock-Schappelwein, J., Eppel, R., Famira-Mühlberger, U., Kügler, A., Mahringer, H., Unterlass, F., & Zulehner, C. (2016). *Die Wirkung von Innovationsaktivitäten geförderter österreichischer Unternehmen auf die Belegschaft*. Studie im Auftrag der Kammer für Arbeiter und Angestellte für Wien und des Bundesministeriums für Verkehr, Innovation und Technologie. Österreichisches Institut für Wirtschaftsforschung.

Bronzini, R., & Iachini, E. (2014). Are incentives for R&D effective? Evidence from a regression discontinuity approach. *American Economic Journal: Economic Policy*, 6(4), 100–134. <https://doi.org/10.1257/pol.6.4.100>.

Büchele, S., Buenstorf, G., Cantner, U., Dreier, L., Meurer, P., & Neumann, L. P. (2024). *Commissioned project evaluations of research and innovation policy in Germany: A review* (Studien zum deutschen Innovationssystem, No. 11-2024). Expertenkommission Forschung und Innovation (EFI). <https://hdl.handle.net/10419/285357>.

Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., & Syrgkanis, V. (2024). *Applied causal inference powered by ML and AI*. arXiv. <https://doi.org/10.48550/arXiv.2403.02467>.

Collin, E., Sandström, C., & Wennberg, K. (2022). Evaluating evaluations of innovation policy: Exploring reliability, methods, and conflicts of interest. In K. Wennberg & C. Sandström (Eds.), *Questioning the entrepreneurial state: Status-quo, pitfalls, and the need for credible innovation policy* (Vol. 53, pp. 157–173). Springer. https://doi.org/10.1007/978-3-030-94273-1_9.

Crato, N., & Paruolo, P. (2019). The power of microdata: An introduction. In N. Crato & P. Paruolo (Eds.), *Data-driven policy impact evaluation* (pp. 1–14). Springer. https://doi.org/10.1007/978-3-319-78461-8_1.

Czarnitzki, D. (2023). *Guidebook for conducting counterfactual impact evaluations of State aid schemes for research, development and innovation*. Publications Office of the European Union. <https://doi.org/10.2760/814181>.

Damonte, A., & Negri, F. (Eds.). (2023). *Causality in policy studies: A pluralist toolbox*. Springer. <https://doi.org/10.1007/978-3-031-12982-7>.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>.

Degelsegger-Márquez, A., Wagner, I., Kroop, S., Rigby, J., & Cox, D. (2017). *Portfolio evaluation FWF international programmes: Final report*. Zentrum für Soziale Innovation. <https://doi.org/10.5281/zenodo.1194558>.

Deiana, C., Geraci, A., & Meroni, E. (2019). *Advanced counterfactual evaluation methods: Guidance document*. Publications Office of the European Union. <https://doi.org/10.2767/464242>.

Dinges, M., Heller-Schuh, B., Kalcik, R., Scherngell, T., Wang, A., Glänzel, W., & Thijs, B. (2020). *Evaluation FWF special research programmes (SFB): Final report*. AIT Austrian Institute of Technology. <https://repository.fteval.at/531/>.

Edler, J., Cunningham, P., Gök, A., & Shapira, P. (2013). *Impacts of innovation policy: Synthesis and conclusions* (Nesta Working Paper No. 13/21). Nesta. <https://www.nesta.org.uk/report/impacts-of-innovation-policy-synthesis-and-conclusions/>.

Edler, J., Gök, A., Cunningham, P., & Shapira, P. (2016). Introduction: Making sense of innovation policy. In J. Edler, P. Cunningham, A. Gök, & P. Shapira (Eds.), *Handbook of innovation policy impact* (pp. 1–17). Edward Elgar Publishing. <https://doi.org/10.4337/9781784711856.00007>.

Einiö, E. (2014). R&D subsidies and company performance: Evidence from geographic variation in government funding based on the ERDF population-density rule. *The Review of Economics and Statistics*, 96(4), 710–728. https://doi.org/10.1162/REST_a_00410.

European Commission. (2014a). *Common methodology for State aid evaluation* (Commission Staff Working Document SWD(2014) 179 final). Publications Office of the European Union. https://competition-policy.ec.europa.eu/system/files/2021-04/modernisation_evaluation_methodology_en.pdf.

European Commission. (2014b). Commission Regulation (EU) No 651/2014 of 17 June 2014 declaring certain categories of aid compatible with the internal market in application of Articles 107 and 108 of the Treaty. *Official Journal of the European Union*, L 187, 1–78. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0651>.

European Commission. (2014c). You can't improve what you can't measure: State aid evaluation. *Competition Policy Brief*, 2014(7). <https://doi.org/10.2763/59207>.

Firpo, T. (2024). *Öko-Scheck evaluation report: Effects of the introduction of randomization to select projects*. Humboldt Innovation.

Fougère, D., & Jacquemet, N. (2019). Causal inference and impact evaluation. *Economie et Statistique / Economics and Statistics*, (510–512), 181–200. <https://doi.org/10.24187/>.

Frontier Economics. (2024). *Counterfactuals for research and innovation evaluation: Rapid evidence review of counterfactual methodologies to evaluate research and innovation programmes*. UK Research and Innovation. <https://www.ukri.org/wp-content/uploads/2024/04/UKRI-300424-Counterfactuals-for-Research-and-Innovation-Evaluation-Report.pdf>.

Geyer, A., Rammer, C., Pointner, W., Polt, W., Hollenstein, H., Donzé, L., Arvanitis, S., & Jörg, L. (2000). *Evaluierung des ITF-Programms FlexCIM*. Endbericht (OEFZS–S-0102). Österreichisches Forschungszentrum Seibersdorf. <https://repository.fteval.at/299/>.

Han, S. (2024). *Mining causality: AI-assisted search for instrumental variables*. arXiv. <https://doi.org/10.48550/arXiv.2409.14202>.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960. <https://doi.org/10.1080/01621459.1986.10478354>.

Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review*, 107(4), 1136–1164. <https://doi.org/10.1257/aer.20150808>.

Huber, M. (2021). An introduction to flexible methods for policy evaluation. In N. Hashimzade & M. A. Thornton (Eds.), *Handbook of research methods and applications in empirical microeconomics* (pp. 82–111). Edward Elgar Publishing. <https://doi.org/10.4337/9781788976480.00010>.

Hünormund, P., & Bareinboim, E. (2025). Causal inference and data fusion in econometrics. *The Econometrics Journal*, 28(1), 41–82. <https://doi.org/10.1093/ectj/utad008>.

Hünormund, P., & Czarnitzki, D. (2019). Innovation policy and causality. *ifo DICE Report*, 17(4), 3–6. <https://www.ifo.de/DocDL/dice-report-2019-4-huenermund-czarnitzki-january.pdf>.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615–635. <https://doi.org/10.1016/j.jeconom.2007.05.001>.

Joint Research Centre (JRC). (2025). *Evaluation report repository*. European Commission. Abgerufen am 17. Dezember 2025, von <https://microeconomicvaluation.jrc.ec.europa.eu/evaluation-report>.

Kaufmann, P., Dorr, A., Enichlmair, C., Hosner, D., Mollay, U., Robubi, A., & Zlatev, T. (2019). *IWB/EFRE-OP AT 2014–20. Begleitende Evaluierung Leistungspaket 2: Kleine und mittlere Unternehmen (KMU) – Endbericht*. ÖIR; convolop – cooperative knowledge design; KMU Forschung Austria; ÖAR; ÖGUT; Spatial Foresight. <https://repository.fteval.at/id/eprint/638/>.

Keuschnigg, C., Gogola, G., Johs, J., Kritzing, M., & Sardadvar, S. (2021). *Wirkung von Forschungsausgaben*. Studie im Auftrag des Bundesministeriums für Digitalisierung und Wirtschaftsstandort. WPZ Research. <https://repository.fteval.at/566/>.

- Kohlweg, K. (2019). „IMPACT“ EIN DISKUSSIONSBEITRAG. fteval - Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung, Wien. [10.22163/fteval.2019.309](https://doi.org/10.22163/fteval.2019.309).
- Krabel, S., & Wagner, V. (2024). *Unsere Maßnahme wirkt. Oder? Methoden zur Analyse kausaler Effekte mit Kontroll- und Vergleichsgruppen* (iit-Studie). Institut für Innovation und Technik (iit). https://doi.org/10.23776/2024_11.
- Kugler, F., Schwerdt, G., & Wößmann, L. (2014). Ökonometrische Methoden zur Evaluierung kausaler Effekte der Wirtschaftspolitik. *Perspektiven der Wirtschaftspolitik*, 15(2), 105–132. <https://doi.org/10.1515/pwp-2014-0013>.
- Lach, S. (2002). Do R&D subsidies stimulate or displace private R&D? Evidence from Israel. *The Journal of Industrial Economics*, 50(4), 369–390. <https://doi.org/10.1111/1467-6451.00182>.
- Landon, T., & Hochreiter, H. (2022). Randomised controlled trials and other experimental approaches in the Austrian Research Promotion Agency (FFG) – Experience, learnings and outlook. *fteval Journal for Research and Technology Policy Evaluation*, (53), 160–168. <https://doi.org/10.22163/fteval.2022.554>.
- Mahlich, J., Ohler, F., Gamsjäger, C., & Puwein, W. (2002). *Evaluierung der Sonderaktion Holzforschung*. Technopolis Forschungs- und Beratungsgesellschaft. <https://doi.org/10.22163/fteval.2002.239>.
- Manahl, C., Wagner, I., & Schuch, K. (2016). *Evaluierung des wissenschaftlichen Impacts von „Sparkling Science“*. Zentrum für Soziale Innovation. [10.22163/fteval.2016.175](https://doi.org/10.22163/fteval.2016.175).
- Matthay, E. C., Neill, D. B., Titus, A. R., Desai, S., Troxel, A. B., Cerdá, M., Díaz, I., Santacatterina, M., & Thorpe, L. E. (2025). Integrating artificial intelligence into causal research in epidemiology. *Current Epidemiology Reports*, 12(1), Article 6. <https://doi.org/10.1007/s40471-025-00359-5>.
- Maxwell, J. A. (2012). *A realist approach for qualitative research*. SAGE Publications.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270–280. <https://doi.org/10.1177/1356389012451663>.
- OECD (2018). *OECD Reviews of Innovation Policy: Austria 2018*. OECD Reviews of Innovation Policy, OECD Publishing, Paris. [10.1787/9789264309470-en](https://doi.org/10.1787/9789264309470-en).
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372(71). <https://doi.org/10.1136/bmj.n71>.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. SAGE Publications.
- Pichler, R. & Steyer, M. (2017). Evaluierung und Wirkungsorientierung in Österreich. Zur Rolle von Evaluierung im neuen Haushaltsrecht und in der Forschungsförderung. *Zeitschrift für Evaluation*, 16(2), 121–139.
- Ramberg, I., & Knell, M. (2012). *Challenges measuring effects of research and innovation policy interventions in ex-post impact evaluations: A synthesis report*. Technopolis Group. https://www.researchgate.net/publication/260552896_Challenges_measuring_effects_of_research_and_innovation_policy_interventions_in_ex-post_impact_evaluations_A_synthesis_report.

- Reiner, C., & Smoliner, S. (2012). *Outputorientierte Evaluierung öffentlich geförderter FTI-Programme – Möglichkeiten und Grenzen* (Studie im Auftrag des Bundesministeriums für Verkehr, Innovation und Technologie). Joanneum Research; Zentrum für Soziale Innovation (ZSI). <https://doi.org/10.22163/fteval.2012.132>.
- Rhomberg, W., Steindl, C., & Weber, M. (2006). *Neue Entwicklungen im Bereich der Wirkungsanalyse und -abschätzung FTI-politischer Maßnahmen: Endbericht* (Report No. ARC-sys-0108). Austrian Research Centers (ARC). <https://doi.org/10.22163/fteval.2006.167>.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>.
- Santos, A. M., & Coad, A. (2023). Monitoring and evaluation of transformative innovation policy: Suggestions for improvement. *Socio-Economic Planning Sciences*, 90, Article 101714. <https://doi.org/10.1016/j.seps.2023.101714>.
- Sardadvar, S., & Ecker, B. (2023). *Evaluierung der FFG-Programmlinie Kleinprojekt*. WPZ Research. <https://doi.org/10.22163/fteval.2023.671>.
- Schibany, A., Dinges, M., Reiner, C., Reidl, S., Hofer, R., Marbler, F., Leitner, K.-H., Dachs, B., Zahradnik, G., Weber, M., Schartinger, D., & Edler, J. (2013). *Ex-post Evaluierung der Kompetenzzentrenprogramme Kplus und K_ind/K_net*. Endbericht im Auftrag des Bundesministeriums für Verkehr, Innovation und Technologie (BMVIT) und des Bundesministeriums für Wirtschaft, Familie und Jugend (BMWFI). Joanneum Research POLICIES; AIT Austrian Institute of Technology; University of Manchester. <https://repository.fteval.at/96/>.
- Schibany, A., Streicher, G., Gretzmacher, N., Falk, M., Falk, R., Knoll, N., Schwarz, G., & Wörter, M. (2004). *Evaluation FFF – Impact analysis: Background report 3.2* (InTeReg Research Report No. 22-2004). Joanneum Research Forschungsgesellschaft mbH.
- Schwarz, G., König, T., Oberhofer, H. & Strassnig, M. (2020). *Zugang zu Register- und Individualdaten für die wissenschaftliche Forschung in Österreich*. fteval Journal for Research and Technology Policy Evaluation, 50, 11–15. [10.22163/fteval.2020.464](https://doi.org/10.22163/fteval.2020.464).
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Stampfer, M. & Strassnig, M. (2020). *Evaluierung neu denken. Studie im Auftrag der Plattform für Forschungs- und Technologiepolitikevaluierung*. Endbericht, Wien. [10.22163/fteval.2020.529](https://doi.org/10.22163/fteval.2020.529).
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations* (DFID Working Paper 38). Department for International Development. <https://repository.fteval.at/126/>.
- Storey, D. J. (2017). Six steps to heaven: Evaluating the impact of public policies to support small businesses in developed economies. In D. L. Sexton & H. Landström (Eds.), *The Blackwell handbook of entrepreneurship* (pp. 176–197). Blackwell Publishers. <https://doi.org/10.1002/9781405164214.ch9>.

Streicher, J. (2017). *Evaluations, Actors and Institutions. The Case of Research, Technology and Innovation Policy in Austria*. Dissertation, Wirtschaftsuniversität Wien, Wien. <https://doi.org/10.57938/3776f66f-bfeb-47cd-b12a-18c7f9257674>.

Streicher, J., Polt, W. & Unger, M. (2020). Eine Untersuchung der Marktsituation im Bereich der FTI-Evaluierung in Österreich. *fteval Journal for Research and Technology Policy Evaluation*, 50, 72–81. DOI: [10.22163/fteval.2020.472](https://doi.org/10.22163/fteval.2020.472).

Warta, K., Dudenbostel, T., Gassler, H., Rammer, C., & Köhler, M. (2019). *Evaluierung der Frontrunner-Initiative*. Endbericht im Auftrag des Bundesministeriums für Verkehr, Innovation und Technologie (BMVIT). Technopolis Group Austria. <https://repository.fteval.at/389/>.

What Works Centre for Local Economic Growth. (2016). *Guide to scoring evidence: Using the Maryland Scientific Methods Scale*. https://whatworksgrowth.org/wp-content/uploads/16-06-28_Scoring_Guide.pdf.

Widmann, R. (2023). *The behavioral additionality of government research grants* (Discussion Paper No. 417). Collaborative Research Center Transregio 190 – Rationality and Competition. <https://hdl.handle.net/10419/282109>.

Zúñiga, P. (2024). *The impact and effectiveness of innovation policy: Evidence from middle-income countries* (Policy Research Working Paper No. 10692). World Bank. <https://thedocs.worldbank.org/en/doc/d3d62fd8bf435a032da07aca3f3aab67-0050022024/original/Impact-and-Effectiveness-of-Innovation-Policy-Pluvia-Zuniga.pdf>.

Annex I – Codebuch

Stufe 1

Variablenname	Erläuterung	Ausprägung	Kodierung
Id	Entspricht dem Dateinamen	BMBF_X oder BMWK_X	
auftraggeber	Gibt das verantwortliche Ministerium bzw. den Auftraggeber an.	BMBF	
		BMWi / BMWK	
evaluator	Gibt den Studiennehmer an.	<i>Freitext: Bspw. Prognos</i>	
titel_r	Titel des untersuchten Reports.	<i>Freitext: Bspw. Evaluation des HSP</i>	
fassung_r_1	Fassung des untersuchten Reports hinsichtlich Vollständigkeit.	<i>Freitext: Bspw. „Langfassung“ oder „Kurzfassung“</i>	
fassung_r_2	Fassung des untersuchten Reports hinsichtlich des Zeitpunkts der Erstellung.	<i>Freitext: Bspw. „Endbericht“ oder „Zwischenbericht“</i>	
jahr_r	Jahr der Fertigstellung des Reports.	Jahr	4-stellige Jahreszahl
titel_p	Gibt den Titel des Evaluierungsgegenstandes an.	<i>Freitext: Bspw. Titel der Fördermaßnahme</i>	
typ_p	Gibt den Typ des Evaluierungsgegenstandes an.	Fördermaßnahme	1
		Gesetz	2
		Institution	3
start_p	Gibt das Jahr des Starts des Evaluierungsgegenstandes an.	Jahr	4-stellige Jahreszahl
end_p	Gibt das Jahr des Endes des Evaluierungsgegenstandes an.	Jahr	4-stellige Jahreszahl
anspruch_eva	Typ der durchgeführten Evaluierung.	Wirkungsevaluation	1
		Andere Evaluierung (bspw. Befragungen)	2
		Kosten-Nutzen-Analyse	3
		Monitoringbericht	4
	Mehrfachnennung möglich.		
wirk_eva	Behauptung von Wirkungen und tatsächlicher Wirkungsevaluation.	Es werden keine Wirkungsbeziehungen ¹⁰ dargestellt und keine Evaluierungsmethoden ¹¹ angewendet.	0
	Stufe 2 erfolgt nur bei Kodierungsschritt 2 oder 3! D. h. es muss mindestens mit einer Kontrollgruppe gearbeitet werden.	Es werden Wirkungsbeziehungen dargestellt und keine Evaluierungsmethoden angewendet.	1
		Es werden Wirkungsbeziehungen dargestellt und Evaluierungsmethoden angewendet.	2

¹⁰ Beispielsweise wird im Rahmen der Studie von „Effekten oder Wirkungen des Evaluationsgegenstandes“ gesprochen. Dabei kann vor allem die Gesamtdarstellung (bspw. in Einleitung oder Schluss) herangezogen werden.

¹¹ Methoden der Evaluationsforschung sind in diesem Kontext alle experimentellen oder quasi-experimentellen Vorgehensweisen (Bspw. Gruppenvergleiche, Matching-Ansätze, Diff in Diff)

		Es werden keine Wirkungsbeziehungen dargestellt, obwohl Evaluierungsmethoden angewendet wurden.	3
--	--	---	---

Stufe 2

Variablenname	Erläuterung	Ausprägung	Kodierung
kg_vor	Eine Kontrollgruppe ist vorhanden.	Ja	1
		Nein	0
kg_art	Erfasst die Art der Kontrollgruppe.	<i>Freitext: Beispielsweise „nicht geförderte Maßnahmenbewerber“</i>	
eva_meth_1	Erfasst das experimentelle Design der Evaluierung hinsichtlich der Randomisierung.	Randomisiertes Experiment	1
		Quasi-experimentelle Methoden	2
eva_meth_2	Erfasst das experimentelle Design der Evaluierung hinsichtlich des Verlaufs.	Querschnitt	1
		Panel oder repeated cross-section	2
eva_quasi	Genaue Evaluierungsmethode bei quasi-experimentellem Vorgehen	<i>Freitext: Bspw. „Matching-Ansätze“ oder „Diff in Diff“</i>	

Stufe 1 + 2

limit	Es wird auf Limitationen des Forschungsdesigns und/oder der Ergebnisse hinsichtlich der kausalen Interpretation hingewiesen.	Ja	1
		Nein	0
limit_x	Genannter Grund für die Limitation.	<i>Freitext: Bspw. Datenverfügbarkeit</i>	
zus_qual	Gibt an, ob (zusätzlich) eine qualitative Vorgehensweise im Report vorhanden ist Mehrfachnennung möglich!	Keine	0
		Interviews	1
		Dokumentenanalyse	2
		Andere	3
int_val	Einschätzung des Raters zur internen Validität der Evaluierungsstudie.	Gering	1
		Moderat	2
		Hoch	3
effektiv	Die Maßnahme war effektiv.	Nein	0
		Teils/Teils	1
		Ja	2

Abkürzungsverzeichnis

AGVO	Allgemeine Gruppenfreistellungsverordnung
AK Wien	Arbeiterkammer Wien
AMDC	Austrian Micro Data Center
API	Application Programming Interface
ARC	Austrian Research Centers (historisch)
aws	Austria Wirtschaftsservice
BMBWF	Bundesministerium für Bildung, Wissenschaft und Forschung (historisch)
BMDW	Bundesministerium für Digitalisierung und Wirtschaftsstandort (historisch)
BMFWF	Bundesministerium für Frauen, Wissenschaft und Forschung
BMK	Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (historisch)
BMVIT	Bundesministerium für Verkehr, Innovation und Technologie (historisch)
BMWFJ	Bundesministerium für Wirtschaft, Familie und Jugend (historisch)
BMWV	Bundesministerium für Wissenschaft und Verkehr (historisch)
cDiD	Conditional Difference-in-Differences
CIS	Community Innovation Survey
DiD	Difference-in-Differences
DML	Double Machine Learning
EFI	Expertenkommission Forschung und Innovation
EU	Europäische Union
FFF	Forschungsförderungsfonds für die gewerbliche Wirtschaft (historisch)
FFG	Österreichische Forschungsförderungsgesellschaft
fteval	Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung
FTI	Forschung, Technologie und Innovation
FWF	Österreichischer Wissenschaftsfonds
GmbH	Gesellschaft mit beschränkter Haftung
GMM	Generalized Method of Moments
IV	Instrumentvariablen
JRC	Joint Research Centre
KI	Künstliche Intelligenz
KMU	Kleine und mittlere Unternehmen
ML	Machine Learning
OECD	Organisation for Economic Co-operation and Development
ÖROK	Österreichische Raumordnungskonferenz
PSM	Propensity Score Matching
RCT	Randomized Controlled Trial
RDD	Regression Discontinuity Design
SBIR	Small Business Innovation Research
SAM	State Aid Modernisation
SMS	Scientific Methods Scale
ZIM	Zentrales Innovationsprogramm Mittelstand
ZSI	Zentrum für Soziale Innovation