



**Lernen aus dem Vergleich von  
Gruppen – Mögliche Ansätze und  
eine Anwendung für die Evaluation  
der Initiative Horizont Handwerk**

Stefan Krabel, Leo Wangler



Stefan Krabel, Leo Wangler

## Lernen aus dem Vergleich von Gruppen – Mögliche Ansätze und eine Anwendung für die Evaluation der Initiative Horizont Handwerk

Zur Analyse der kausalen Wirkung von Fördermaßnahmen empfohlen Expertinnen und Experten häufig, randomisierte kontrollierte Studien durchzuführen. Wenn die Voraussetzungen für eine solche Methode nicht gegeben sind – was in der Praxis oft der Fall ist, insbesondere in der Forschungs- und Wirtschaftsförderung – können sogenannte quasi-experimentelle Designs zum Einsatz kommen. Dabei wird versucht, möglichst vergleichbare Gruppen zu identifizieren, um die Wirkung von Maßnahme abzuleiten. In diesem Beitrag wird dieser Ansatz exemplarisch an der Initiative „Horizont Handwerk“ in Baden-Württemberg vorgestellt. Das Ergebnis: Die Wirkung der Förderung lässt sich kausal nachweisen.

### Einleitung

Dieser Beitrag befasst sich mit der Frage nach der Messung direkter Programmeffekte am Beispiel eines niedrighwelligen Förderprogramms, das auf die Unterstützung von Handwerksbetrieben ausgerichtet ist, der sogenannte „Horizont Handwerk“. Mit der Initiative wurde in Baden-Württemberg ein Angebot geschaffen, um das Handwerk beim Strukturwandel zu unterstützen. Die Initiative konzentriert sich auf vier zentrale Handlungsfelder: (1) Fachkräftegewinnung und -sicherung, (2) strategische Ausrichtung und Transformation, (3) Digitalisierung im Handwerk und (4) Nachhaltigkeit, Klimaschutz und Energiewende. Hierzu werden Unterstützungs-, Beratungs- und Wissensangebote für Handwerksbetriebe angeboten.

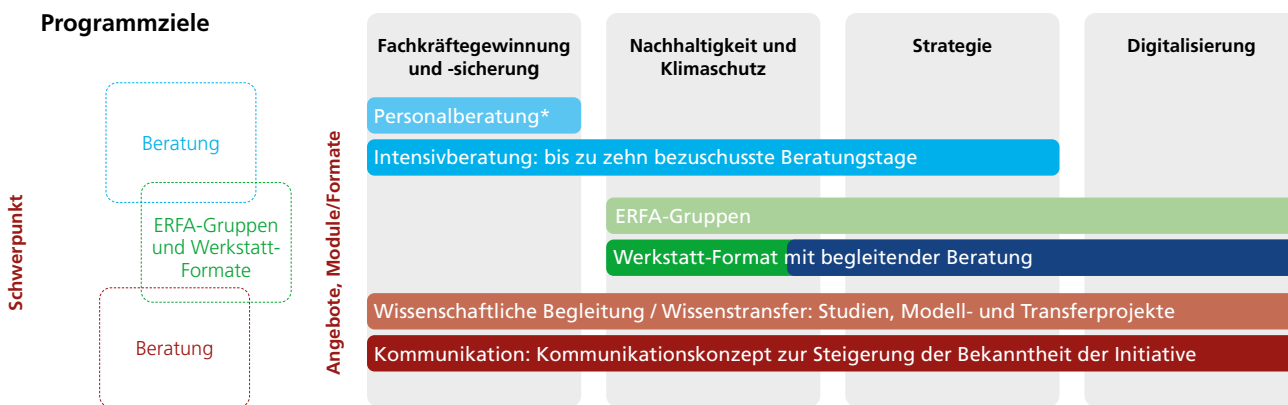
Die Unterstützung unterscheidet zwischen verschiedenen Maßnahmen in den Bereichen Beratung, Erfahrungsaustauschgruppen (ERFA-Gruppen) und Werkstatt-Formate sowie Transfer (vgl. Wangler et al. 2025). In allen Bereichen geht es darum, die Handwerksbetriebe so zu unterstützen, dass diese in der Lage

sind, besser mit bestehenden Herausforderungen umzugehen, und die nachhaltige Aufrechterhaltung der Geschäftstätigkeit der Handwerksbetriebe zu unterstützen.

Evaluationen sind darauf ausgelegt, die Wirkungen von FTI-Programmen zu erfassen<sup>1</sup>. In der Praxis werden FTI-Programme zielgerichtet so design und organisiert, dass das Angebot auf die Programmziele ausgerichtet ist. Die Evaluation ist zentraler Bestandteil der Wirkungsmessung, wenngleich das Programmdesign in der Regel nicht auf die Wirkungsmessung ausgerichtet ist, weshalb in der Praxis sogenannte reine Kontrollgruppenanalysen in den meisten Fällen nicht möglich sind.

1 „FTI-Prgramme“ steht für Förderprogramme in den Bereichen Forschung, Technologie und Innovation (FTI)

## Horizont Handwerk – Programmziele und Angebote



### Programmziele

**Stabstelle Gesamtkoordination und Nachhaltigkeit:** zentrale Anlaufstelle für Projektträger und Projektinteressierte aus Handwerkskammern, Fachverbänden und weiteren Handwerksorganisationen

**Stabstelle Personal:** einheitliche Struktur der Personalberatung für das gesamte Land sicherzustellen und den Projekterfolg nach einheitlichen Messgrößen bewertbar zu machen; bspw. Durch einen regelmäßigen Erfahrungsaustausch der Berater, die Erstellung von Informationsmaterialien, Leitfäden und Präsentationen

**Stabstelle Zukünftige Strategien und Geschäftsmodelle:** Ziel ist es, möglichen Strategiedefiziten der Betriebe entgegenzuwirken, bspw. Durch gezielte Sensibilisierung der Betriebe für das Thema Strategie und Entwicklung sinnvoller Instrumente für Beraterinnen und Berater sowie Betriebe (Werkzeugkoffer Strategie, Branchenlösungen)...

Abbildung 1: Horizont Handwerk – Programmziele und Angebote (Quelle: Wangler et al. 2025)

Trotz dieser Limitation gibt es „Second-best“-Lösungen. Die folgende Analyse befasst sich mit verschiedenen Ansätzen zur Quantifizierung von Programmeffekten. Hierzu werden verschiedene Methoden von Gruppenvergleichen vorgestellt, die häufig durchführbar sind. Darauf aufbauend wird am Beispiel der hier vorgestellten Initiative aufgezeigt, welche Analysemöglichkeiten im konkreten Fall von Horizont Handwerk möglich sind und welche Effekte nachgewiesen werden können. Der Beitrag schließt mit einer Zusammenfassung zentraler Ergebnisse und Erkenntnisse.

### 1. Kurzbeschreibung verschiedener Ansätze zu Gruppenvergleichen

Um die Wirkung von Fördermaßnahmen zu messen, empfehlen u. a. die Europäische Kommission und die Expertenkommission Forschung und Entwicklung (EFI) die Anwendung von randomisierten kontrollierten Studien – auf Englisch: Randomized Controlled Trials (RCT) – die eine kausale Messung der Maßnahme erlauben (siehe z. B. Europäische Kommission 2022, EFI

2024). Diese Methode setzt allerdings voraus, dass vor der Durchführung einer Fördermaßnahme eine zufällige Auswahl einer Gruppe von geförderten und einer Gruppe von nicht-geförderten Individuen oder Einrichtungen vorgenommen wird (siehe z. B. Krabel und Wagner 2024). Dies ist in der Praxis oft nicht der Fall, weil in der Regel eine Auswahl der besten Anträge über Gutachterbewertungen vorgenommen wird, um den Output der Förderung zu maximieren.

Tabelle 1: Ansätze zu Gruppenvergleichen mit Vor- und Nachteilen

	Beschreibung	Vorteil	Nachteil
1. Vergleich von Geförderten mit Nicht-Geförderten (aber Förderungs-bewerbern)	<b>Variante A:</b> Vergleich der gesamten Gruppe		
	Man vergleicht geförderte Einheiten mit solchen, die sich ebenfalls beworben haben, aber nicht gefördert wurden.	Beide Gruppen hatten ein Interesse an der Maßnahme. Daher gibt es <b>keine Selektionseffekte hinsichtlich der Teilnahme an dem Förderangebot.</b>	Die <b>Auswahl für die Förderung war evtl. nicht zufällig</b> – (z. B. wurden besonders vielversprechende Bewerbungen gefördert; dies führt zu Selektionseffekten in der Auswahl der Geförderten.
	<b>Variante B:</b> Regression Discontinuity Design		
	Bei dieser Methode wird ausgenutzt, dass <b>Förderentscheidungen häufig anhand eines Schwellenwerts getroffen werden</b> – (z. B. erhalten nur Anträge mit einer Bewertung über 80 Punkten eine Förderung. Einrichtungen, die knapp über und knapp unter dieser Grenze liegen, sind sich in vielen Merkmalen sehr ähnlich.	Ein Vergleich von Geförderten und Nicht-Geförderten mit ähnlicher Bewertung bzw. Punktzahl der Qualität des Antrags erlaubt eine <b>besonders verlässliche Abschätzung des Effekts der Förderung</b> , da die Qualität der Anträge in der Gruppe der betrachteten nicht-geförderten Projekte nahezu identisch ist mit der Qualität in der Gruppe der betrachteten geförderten Projekte. Diese Zuteilung sichert ähnlich wie in einem zufälligen Experiment eine gute Vergleichbarkeit der Gruppen und erfordert keine aufwändigen statistischen Kontrollverfahren.	Die Aussagekraft <b>bezieht sich nur auf Einrichtungen nahe am Schwellenwert („lokaler Effekt“)</b> – sie gilt also nicht unbedingt für alle Antragstellenden. Zudem braucht es ausreichend viele Fälle rund um den Schwellenwert und eine klare, strikt eingehaltene Fördergrenze.
2. Geförderte werden mit Einheiten verglichen, die nicht einmal die Möglichkeit kannten, sich zu bewerben	Geförderte werden mit Einheiten verglichen, die nicht gefördert wurden und <b>nicht einmal die Möglichkeit kannten, sich zu bewerben.</b>	Dieses Vorgehen erlaubt einen <b>Vergleich der Geförderten mit einer „neutralen“ Gruppe</b> , d. h. die Nicht-Geförderten haben weder eine bewusste strategische Entscheidung getroffen, die Förderung nicht in Anspruch zu nehmen, noch wurden sie anhand von Gütekriterien als weniger förderfähig bewertet (wie in Beispiel 1).	Es gibt die Gefahr der <b>starken systematischen Unterschiede</b> (Selektion auf unbeobachtete Variablen) zwischen den teilnehmenden und nicht teilnehmenden Einrichtungen. Es könnte beispielsweise sein, dass die Einrichtungen, denen die Förderung unbekannt war, systematisch schlechter vernetzt oder schlicht kleiner sind als die Geförderten und keine Ressourcen für die Sichtung passender Förderangebote haben – und sich daher von den Geförderten unterscheiden.
3. „Vorher-Nachher-Vergleich“ bei Geförderten	Man vergleicht die geförderten Einheiten <b>vor und nach der Maßnahme.</b>	Ein solcher Vergleich ist <b>oft einfach durchführbar und schließt Unterschiede zwischen betrachteten Einrichtungen aus</b> , da die Einrichtungen mit sich selbst verglichen werden.	Es gibt keine Kontrollgruppe und <b>keine Betrachtung von Zeittrends</b> oder anderen Einflüssen. So werden etwa mögliche zeitliche Einflüsse inklusive wichtiger Ereignisse oder rechtlicher Veränderungen auf die gesamte Branche nicht berücksichtigt.

Was kann also getan werden, wenn eine randomisierte Studie nicht mehr durchführbar ist? Auch wenn eine kontrollierte randomisierte Studie nicht mehr möglich ist, können Gruppenvergleiche zur Messung von Wirkungen sinnvoll sein. Im Folgenden sind einige Gruppenvergleiche und die jeweils damit verbundenen Vor- und Nachteile angeführt.

**2. Praxisbeispiel**

Datengrundlage sind die Ergebnisse der Online-Befragung zu „Horizont Handwerk“. Im Rahmen der Evaluation wurde eine Befragung unter Handwerksunternehmen, Verbänden und Innungen in Baden-Württemberg durchgeführt. Dabei wurden sowohl Betriebe erfasst, die die o. a. Angebote wahrgenommen haben, als auch Betriebe, die keine Unterstützung oder Beratung in Anspruch genommen haben. Mit Hilfe der Befragung konnte ermittelt werden, ob Betriebe, die keine Beratung in Anspruch genommen haben, die Initiative nicht kannten oder ob sie sich bewusst gegen eine Unterstützung entschieden haben. Der Datensatz umfasst 348 Teilnehmende.

Diese Unterscheidung erlaubt u. a. den Vergleich zwischen Geförderten und solchen, die die Maßnahme nicht kannten und daher keine Unterstützung in Anspruch genommen haben – Gruppenvergleich Nr. 2 aus vorherigem Abschnitt. Im Folgenden werden diese beiden Gruppen hinsichtlich der Aussichten auf Gewinn und Arbeitsplätze im folgenden Jahr verglichen, die über folgende Fragen erhoben wurden:

**Wie schätzen Sie die Aussichten Ihres Betriebs im Hinblick auf folgende Kenngrößen in den kommenden drei Jahren ein?“**

„Gewinn“			
1	2	3	4
steigend	gleichbleibend	fallend	Weiß nicht

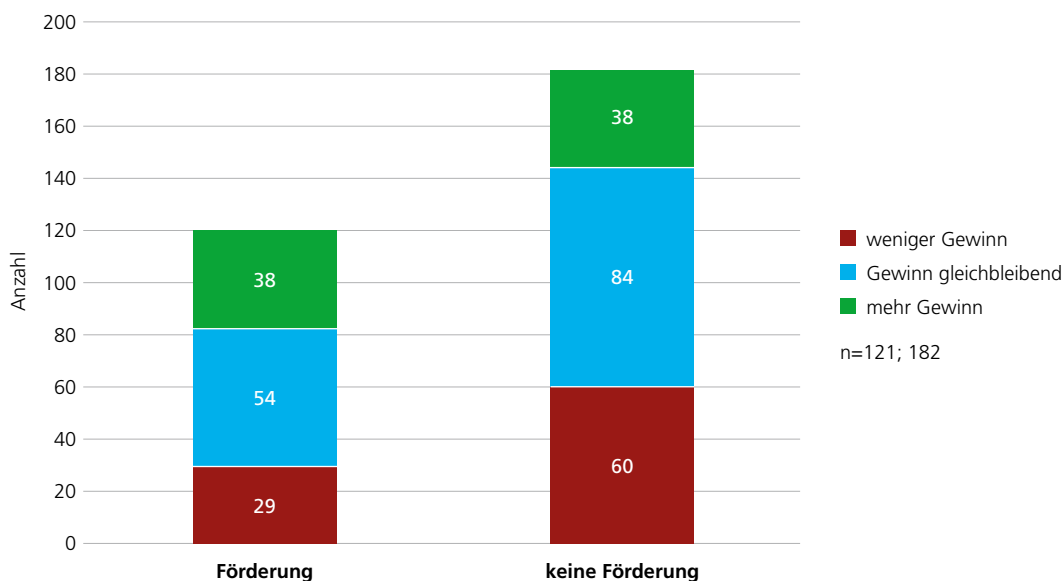
  

Arbeitsplätze			
1	2	3	4
steigend	gleichbleibend	fallend	Weiß nicht

**Darstellung 1:** Betriebsaussichten im Hinblick auf Gewinn und Arbeitsplätze in den kommenden drei Jahren (Quelle: Institut für Innovation und Technik, eigene Darstellung)

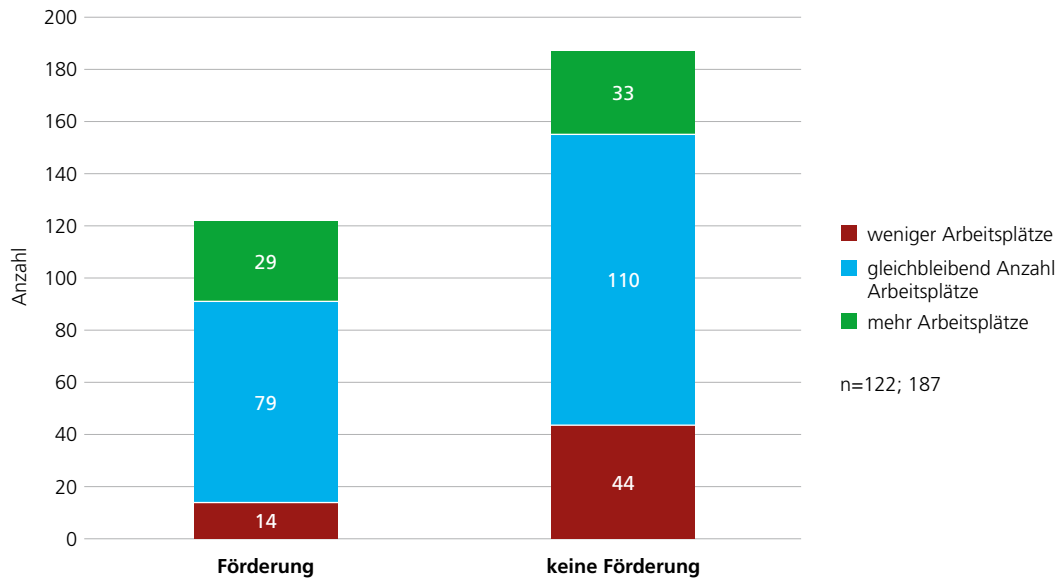
In der unten angeführten Abbildung 2 ist angeführt, wie die Verteilung der Antworten hinsichtlich der Aussichten auf Gewinn innerhalb der Gruppe der Betriebe verteilt ist, die Unterstützungsangebote wahrgenommen haben (n = 121; mehr Gewinn = 38, gleichbleibender Gewinn = 54, weniger Gewinn = 29) und Betriebe, denen die Förderung unbekannt war (n = 182; mehr Gewinn = 38, gleichbleibender Gewinn = 84, weniger Gewinn = 60).

**Aussicht auf Gewinn im Gruppenvergleich**



**Abbildung 2:** Aussichten auf Gewinn im Gruppenvergleich (Quelle: Institut für Innovation und Technik, eigene Darstellung)

**Aussichten Arbeitsplätze im Gruppenvergleich**



**Abbildung 3:** Aussichten Arbeitsplätze im Gruppenvergleich (Quelle: Institut für Innovation und Technik, eigene Darstellung)

In Abbildung 3 ist derselbe Vergleich für die Aussichten auf Arbeitsplätze dargestellt, mit Gruppengrößen von 122 bei denen, die die Förderangebote wahrgenommen haben (mehr Arbeitsplätze = 29, gleichbleibende Zahl Arbeitsplätze = 79, weniger Arbeitsplätze = 14) respektive 187 bei denen die die

Förderung nicht kannten (mehr Arbeitsplätze = 33, gleichbleibende Zahl Arbeitsplätze = 110, weniger Arbeitsplätze = 44). Betriebe, die mit „Weiß nicht“ geantwortet haben, sind in diesen Auswertungen nicht mitbetrachtet worden.

**Tabelle 2: Zweiseitige t-tests zur Analyse der Gruppenunterschiede hinsichtlich der Aussichten auf Gewinn und Arbeitsplätze**

Gruppe	M	SD	n	t	df	p
Aussichten auf Gewinn						
Behandlungsgruppe (Förderung)	1,925	0,743	121			
Kontrollgruppe (keine Förderung)	2,121	0,726	182	2,27	301	0,0238
Aussichten auf Arbeitsplätze						
Behandlungsgruppe (Förderung)	1,877	0,583	122			
Kontrollgruppe (keine Förderung)	2,059	0,640	187	2,52	307	0,0121

(Quelle: Institut für Innovation und Technik, eigene Darstellung)

Um einen validen statistischen Vergleich zu bekommen, wurden, t-Tests durchgeführt, die ermitteln, ob die dargestellten Unterschiede der Verteilungen statistisch signifikant sind. Das Signifikanzniveau ist ein Schwellenwert, der im Vorfeld der Analyse festgelegt wird, um zu entscheiden, welche Anforderung an den Niveauunterschied (Förderung vs. keine Förderung) gestellt werden.

### 3. Interpretation der Ergebnisse und Möglichkeit zur Überprüfung kausaler Effekte

Das dargestellte Beispiel zeigt statistisch signifikante Unterschiede zwischen der Gruppe der Geförderten und der Gruppe der Betriebe, denen die Förderung unbekannt war. Damit ist es sehr wahrscheinlich, dass die Förderung eine kausale Wirkung in Hinblick auf die untersuchten volkswirtschaftlichen Kennzahlen hatte – wenngleich ein kausaler Effekt damit aber auch nicht vollständig nachgewiesen ist. Denkbar ist beispielsweise, dass es bestimmte Faktoren (z. B. Anzahl Mitarbeitende, Alter der Firma) gibt, die sowohl die Bekanntheit der Förderung als auch die Aussichten auf steigenden Gewinn (respektive Wachstum der Anzahl der Arbeitsplätze) beeinflusst haben könnten. Sollte letzteres zutreffen, hätten derartige Faktoren sowohl für die Gruppeneinteilung als auch für die statistischen Unterschiede maßgeblichen Einfluss.

Um die Kausalität weiterführend zu validieren, bietet sich ein sogenanntes Matching-Verfahren an. Dieses kann dabei helfen, zwei Gruppen vergleichbar zu machen. Die Vorgehensweise ist dabei wie folgt: Für jedes Unternehmen (bzw. jede Person), das an der Maßnahme teilgenommen hat, wird ein möglichst ähnliches Unternehmen (bzw. eine ähnliche Person) gesucht, das nicht teilgenommen hat (sog. „Zwilling“). Die Paare „passen“ dann in relevanten Merkmalen zueinander. In unserem Beispiel konnten Paare nach Alter des Unternehmens, Größe des Unternehmens und Region gebildet werden, die sich anhand dieser Kriterien ähneln, sich jedoch in dem entscheidenden Merkmal des Erhalts der Förderung unterscheiden.

Ziel ist es, durch diesen Vorgang mit einer höheren Wahrscheinlichkeit kenntlich zu machen, ob Unterschiede zwischen den Gruppen wirklich an der Maßnahme liegen – und nicht einfach daran, dass die Personen vorher schon unterschiedlich waren. Damit könnte auch die Kausalwirkung einer Maßnahme untersucht werden. Für solche Matching-Verfahren gelten grob die folgenden Faustregeln.

#### Statistische Faustregeln bei Matching-Verfahren

Für einfache Gruppenvergleiche (z. B. Mittelwertunterschiede) gilt oft: Mindestens 30 Personen oder Unternehmen oder andere Einheiten sollten pro Gruppe enthalten sein – dies erlaubt grobe Schätzwerte.

Für Matching-Verfahren oder differenziertere Analysen sind oft über 100 Personen / Unternehmen / Einheiten pro Gruppe sinnvoll.

Je mehr Matching-Variablen benutzt werden, um Gruppeneinteilungen vorzunehmen, desto größer sollte die Gruppe sein.

Die Vergleichsgruppe sollte mindestens gleich groß oder größer sein als die Interventionsgruppe.

#### Infobox 1: Statistische Faustregeln bei Matching-Verfahren

Im oben angeführten Beispiel zeigt sich, dass Unterschiede zwischen den Gewinnaussichten und der Entwicklung der Anzahl der Mitarbeitenden zwischen den beiden Gruppen (Unternehmen, die sich an Angeboten der Förderung beteiligen vs. Unternehmen, denen die Förderung unbekannt ist) auch nach einem Matching-Verfahren signifikant bleiben. Konkret wurde ein Matching nach den Variablen „Alter des Unternehmens“, „Anzahl Mitarbeitende des Unternehmens“ und „Region des Unternehmens“ durchgeführt, wobei die Region über die Postleitzahl erfasst wurde. Nach dem Matching zeigen sich statistische Unterschiede im t-Test, die nur marginale Unterschiede in den Mittelwerten und ein ähnliches Signifikanzniveau (ebenfalls unter 5 %) aufweisen wie der Vergleich ohne Durchführung eines Matching-Verfahrens. Die Gruppengrößen reduzieren sich dabei allerdings deutlich, da nur die Beobachtungen in Betracht gezogen werden können, die Angaben zu den Matching-Variablen gemacht haben und zu denen ein guter „Zwilling“ in der anderen Gruppe gefunden werden konnte.

## Fazit und Ausblick

In dem angeführten Beispiel zeigt sich insgesamt folgendes Bild: In der Gruppe der Unternehmen, die am Unterstützungsangebot teilgenommen haben, sind die Aussichten auf zukünftigen Gewinn und steigende Anzahl an Mitarbeitenden in den kommenden Jahren statistisch signifikant höher als in der Gruppe der Unternehmen, die nicht am Angebot teilgenommen haben. Der Unterschied in den Gruppen bleibt auch dann robust, wenn ein Matching nach den Variablen ‚Alter des Unternehmens‘ sowie ‚Anzahl Mitarbeitende des Unternehmens‘ und ‚Region des Unternehmens‘ durchgeführt wird. Inwieweit diese Aussichten mit der tatsächlichen Entwicklung übereinstimmen, wird sich erst in den kommenden Jahren zeigen. In Bezug auf die abgefragten Aussichten ist der o.a. Effekt aber stabil.

Generell lässt sich festhalten: Auch wenn keine zufällige Zuteilung zu Gruppen möglich ist, lassen sich Wirkungen von Maßnahmen untersuchen – etwa durch Gruppenvergleiche mit Matching-Verfahren. Dabei werden Personen oder Unternehmen aus einer Vergleichsgruppe so ausgewählt, dass sie der behandelten Gruppe möglichst ähnlich sind (z.B. anhand von Alter, Geschlecht, Bildung). Solche Verfahren bieten eine gute Alternative, um Wirkungen auch in der Praxis belastbar einschätzen zu können. Matching-Verfahren können dabei helfen, Gruppen vergleichbarer zu machen – ersetzen aber keine echte Zufallszuweisung. Es ist ein praktischer Kompromiss, besonders bei Auswahlverfahren, in denen zufällige Zuteilungen nicht möglich oder nicht gewollt sind.

## Literatur

Expertenkommission für Forschung und Innovation (EFI) (2024): Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands – Gutachten 2024. Online unter: <https://www.e-fi.de/fileadmin/Assets/Gutachten/2024/>, zuletzt aufgerufen am 21.11.2025.

Europäische Kommission (2022): Entwicklung und Beauftragung von kontrafaktischen Wirkungsanalysen – Ein praktischer Leitfaden für ESF-Verwaltungsbehörden. Online unter: <https://op.europa.eu/de/publication-detail/-/publication/dd4a4fc7-42a3-11ec-89db-01aa75ed71a1>, zuletzt aufgerufen am 21.11.2025.

Krabel, Stefan; Wagner, Valentin (2024): Unsere Maßnahme wirkt. Oder? Methoden zur Analyse kausaler Effekte mit Kontroll- und Vergleichsgruppen, iit-Studie, Institut für Innovation und Technik (iit). Online unter: <https://www.iit-berlin.de/publikation/unsere-massnahme-wirkt-oder-methoden-zur-analyse-kausaler-effekte-mit-kontroll-und-vergleichsgruppen/>, zuletzt aufgerufen am 21.11.2025.

Wangler, Leo; Gerdes, Theresa; Justen, Roland; Kreibich, Miriam; Mangelsdorf, Axel; Palka, Silvia; Tatu, Diana-Cristina; Wesels, Jan; Robeck, Martin-Simon (2025): Evaluation der Initiative „Horizont Handwerk“, Evaluationsbericht im Auftrag des Ministeriums für Wirtschaft, Arbeit und Tourismus Baden-Württemberg. Online unter: <https://www.iit-berlin.de/publikation/evaluation-der-initiative-horizont-handwerk-2/>, zuletzt aufgerufen am 21.11.2025.

Wangler, Leo (2025): Evaluation von Forschungs-, Entwicklungs- und Innovationsbeihilfen: Zu einer praktischen Umsetzung von Vergleichsgruppenansätzen. In: Zeitschrift für Evaluation, Ausgabe 1/2025, 14. Jahrgang, S. 106-115. Online unter: [https://www.waxmann.com/zeitschrift/?tx\\_p2waxmann\\_zeitschriftenanzeige%5bartikel%5d=ART101621&tx\\_p2waxmann\\_zeitschriftenanzeige%5baction%5d=artikel](https://www.waxmann.com/zeitschrift/?tx_p2waxmann_zeitschriftenanzeige%5bartikel%5d=ART101621&tx_p2waxmann_zeitschriftenanzeige%5baction%5d=artikel), zuletzt aufgerufen am 21.11.2025.

### **Herausgeber**

Institut für Innovation und Technik (iit)  
in der VDI/VDE Innovation + Technik GmbH  
Steinplatz 1, 10623 Berlin

### **Zitation**

Krabel, Stefan; Wangler, Leo (2025): Lernen aus dem Vergleich von Gruppen – Mögliche Ansätze und eine Anwendung für die Evaluation der Initiative Horizont Handwerk. iit-perspektive Nr. 84. Institut für Innovation und Technik (iit), Berlin.

### **Autor:innen**

Dr. Stefan Krabel  
Tel.: +49 (0) 30 310078 506  
E-Mail: krabel@iit-berlin.de

Dr. Leo Wangler  
Tel.: +49 (0) 30 310078 434  
E-Mail: wangler@iit-berlin.de

### **iit perspektive Nr. 84**

Dezember 2025  
Layout: Poli Quintana  
DOI: 10.23776/2025\_15  
Bildnachweise: Abdul – stock.adobe.com

