

Review

Outcome Measurement, Costing, and Practical Guidance for Evaluating Supported Employment for People with Severe Mental Illness: A Methodological Review

Kuo-Yi Jade Chang ^{1,*} , Jennifer Smith-Merry ¹ , Ancheng Koh ¹, Yao Yao ² and Ying Li ²

¹ Centre for Disability Research and Policy, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW 2006, Australia; jennifer.smith-merry@sydney.edu.au (J.S.-M.); koh.ancheng@sydney.edu.au (A.K.)

² Faculty of Medicine and Health, School of Public Health, The University of Sydney, Sydney, NSW 2006, Australia; yyao0384@uni.sydney.edu.au (Y.Y.); yili0735@uni.sydney.edu.au (Y.L.)

* Correspondence: jade.chang@sydney.edu.au

Abstract

Supported employment programs are widely implemented to improve vocational outcomes for people with severe mental illness, yet evaluations of these programs employ highly heterogeneous approaches to costing and outcome measurement. This methodological review examines how costs and outcomes have been identified, measured, and reported in randomized controlled trials of supported employment interventions conducted in high-income countries. Eligible trials were drawn from a recent rapid scoping review and analysed using narrative synthesis, with data extracted on analytical perspectives, outcome domains, costing methods, and data collection approaches. Thirty-two trials were included, of which seven incorporated an economic analysis. Vocational outcomes were consistently prioritised, while clinical, functioning, recovery-oriented, service use, and implementation outcomes were variably assessed. Economic analyses differed substantially in perspective, time horizons, cost identification and valuation, and transparency, limiting comparability and transferability. While methodological diversity reflected differing research questions and policy contexts rather than inconsistent intervention effectiveness, incomplete reporting constrained interpretation and evidence use. Based on these findings, the review outlines practical guidance for evaluators, emphasising alignment of study design with decision-making needs, transparent and perspective-consistent costing, theory-informed outcome selection, and explicit documentation of contextual and methodological choices to enhance the policy relevance and interpretability of supported employment evaluations.

Keywords: supported employment; severe mental illness; individual placement and support; economic evaluation; costing; outcome measurement; program evaluation



Academic Editor: Joaquim J.F. Soares

Received: 27 January 2026

Revised: 30 May 2026

Accepted: 31 May 2026

Published: 2 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

For people with severe mental illness (SMI), employment represents more than a source of income. Individuals with SMI consistently report that work helps structure daily life, strengthens personal identity, and enables meaningful participation in mainstream social roles [1–4]. Beyond these individual benefits, employment may reduce reliance on disability benefits and institutional care, while contributing to broader patterns of community inclusion [4].

Despite strong preferences for employment, competitive employment rates among people with SMI typically range between 10% and 20%, far below those of the general population [5]. This gap reflects multiple, intersecting barriers, including environmental and personal factors. At the personal level, cognitive impairments affect learning, skill acquisition, and workplace problem-solving [6]. Psychological factors, including expectations of failure that undermine motivation and engagement, further limit vocational participation [7]. These factors can also lead to a lack of work readiness, including limited work experience, vocational training, and educational attainment, which then further compound pre-existing vulnerabilities [8]. Structurally, people with SMI face discrimination in hiring and within workplaces [9] and the structure of welfare payments can mean that individuals are concerned about the impact of work on disability support payments [10]. These barriers operate cumulatively, making sustained employment difficult to achieve even among those who express a strong desire to work.

In response to these challenges, supported employment programs—most prominently those based on the Individual Placement and Support (IPS) model—have been developed and increasingly adopted across OECD countries [4]. In contrast to traditional “train-and-place” vocational rehabilitation approaches, IPS includes a rapid job search for competitive employment, individualised support aligned with participant preferences, and ongoing assistance following job placement [11]. In addition to standard IPS, some programs incorporate adjunct components—including cognitive remediation (CR), work-focused social skills training, and cognitive behavioural interventions—to address cognitive impairments and defeatist beliefs that may limit vocational outcomes [12–14]. Other approaches include transitional work programs offering time-limited, non-competitive placements [15] and the Clubhouse model, which combines support for employment with broader social support [3].

A substantial body of research has evaluated supported employment programs for people with SMI, with IPS being the most extensively studied model. However, the resulting evidence base is highly heterogeneous. Studies vary widely in the outcomes selected, the costing approaches applied, the analytical perspectives adopted (e.g., health system, welfare system, societal), and the data sources and instruments used to measure both costs and outcomes. Economic analyses further differ in their choice of efficiency metrics—ranging from cost per job gained, cost per hour worked, and cost per quality-adjusted life year (QALY) gained to net benefit or return on investment—as well as in time horizons and approaches to uncertainty [16].

This heterogeneity should not be interpreted as a flaw in the evidence base or as an indication that supported employment is inconsistently effective. Rather, it reflects the fact that evaluations have been conducted to address different research questions, policy objectives, and knowledge gaps, often within distinct institutional, service, labour market, and fiscal contexts. Methodological diversity is therefore both expected and, in many cases, appropriate. Attempts to impose uniformity across evaluations risk obscuring context-specific priorities and limiting the relevance of findings for local decision-making.

Challenges arise, however, when methodological differences—and their implications—are insufficiently transparent. Choices regarding study perspective, inclusion or exclusion of cost components, outcome measurement, and analytical design can materially influence both findings and their interpretation. Two studies may both conclude that IPS is “cost-effective” yet be answering fundamentally different questions about value for different decision-makers, using different definitions of benefit [17]. The extent to which such findings are transferable is further shaped by contextual factors, including labour market conditions and local welfare systems. Without explicit consideration of how methodological choices shape results, evidence synthesis becomes difficult and policy translation uncertain.

This places a particular responsibility on researchers. To generate evidence that is meaningful for policy and service planning, researchers must be equipped not only to apply appropriate costing and outcome measurement techniques, but also to understand how different methodological approaches influence findings, introduce potential bias, and constrain interpretation. This includes the ability to assess study quality, evaluate sources of uncertainty, and design evaluations using methods that are aligned with the specific decision contexts in which evidence will be applied. Methodological rigour, in this sense, is not about adopting a single “best” approach, but about making informed, transparent, and context-appropriate methodological choices.

Viewed through this lens, this methodological review aims to (1) systematically examine how costs and outcomes have been identified, measured, and reported in evaluations of supported employment for people with SMI; (2) assess the implications of different methodological choices for interpretation, comparability, and policy relevance; and (3) provide practical guidance to support program evaluators in designing, implementing, and reporting evaluations that generate robust, decision-relevant evidence. In doing so, this review also seeks to enable policymakers and practitioners to better judge the relevance of existing evidence to their own contexts and to support the design of future evaluations that inform real-world decision-making, rather than producing results that are technically sound but difficult to apply.

2. Materials and Methods

This paper presents a methodological review rather than a systematic review or meta-analysis of intervention effectiveness, both of which have been conducted previously [4,18,19]. We do not evaluate effect sizes or conduct assessments of overall study quality or risk of bias across included trials. Rather, our analysis examines the methodological choices researchers make in identifying and measuring costs and outcomes, the data collection approaches employed, and the transparency with which these choices are reported.

2.1. Study Selection

Eligible randomised controlled trials (RCTs) were identified from a recently completed rapid scoping review of community-based psychosocial interventions for adults with SMI [20]. The parent scoping review followed the Cochrane Rapid Review Methods Guidance and searched MEDLINE and CINAHL for studies published between January 2009 and May 2024. Full details of the search strategy, eligibility criteria, and study selection process are reported elsewhere [20].

For the present methodological review, we extracted studies from the scoping review dataset that met the following criteria:

- Study design: RCTs, including secondary analyses and trial-based economic evaluations. Secondary analyses were retained because they frequently make independent methodological choices that are directly relevant to this review’s aims.
- Intervention type: Supported employment or vocational rehabilitation delivered as a primary intervention.
- Population: Adults with SMI living in high-income countries.

Studies were included regardless of whether supported employment was augmented with additional therapeutic components (e.g., cognitive or behavioural interventions), provided that supported employment remained a central element of the intervention.

2.2. Data Extraction

Data extraction was conducted independently by two reviewer pairs. A.K. and Y.L. extracted data related to outcome measurement, while A.K. and Y.Y. extracted data related to costing and economic analysis. This domain-specific division was adopted to leverage the complementary methodological expertise of the review team, given that outcome measurement and economic analysis require distinct technical knowledge. Within each domain, reviewers extracted data independently before comparing and reconciling their records. Discrepancies within each pair were resolved through discussion; where consensus could not be reached, K.-Y.J.C. adjudicated. K.-Y.J.C. also reviewed and validated all extracted data to ensure consistency and accuracy.

For each included study, information was extracted on key study characteristics, including country of publication, study aims, sample size, study population, and descriptions of the intervention and comparator. Outcome-related data included the outcomes assessed, the measurement instruments used, and the data sources and collection methods employed.

For studies that incorporated a costing component or economic evaluation, additional data were extracted on the type of economic analysis conducted, analytical perspective, time horizon, use of discounting, and approaches to uncertainty or sensitivity analysis. Detailed information was also collected on cost categories and items, methods used to measure costs in physical units, cost valuation sources, and data collection procedures.

2.3. Methodological Quality Assessment

The methodological quality of cost and outcome measurement was assessed using a modified subset of items from the Consensus Health Economic Criteria (CHEC) list [21]. Five domains were selected from the 19-item CHEC list, corresponding to the domains most directly relevant to this review's focus on cost and outcome measurement practices: (i) whether all important and relevant outcomes were fully identified in relation to the stated perspective and research question (CHEC item 10); (ii) whether outcome measurement was directly based on the identified outcomes, with valid and clearly stated instruments (CHEC item 11)—this criterion concerns the alignment between identified and measured outcomes, not data collection procedures, which are documented separately in the Supplementary Materials; (iii) whether all important and relevant costs were fully identified in relation to the stated perspective and research question (CHEC item 7); (iv) whether costs were measured appropriately using physical units of resource use (CHEC item 8); and (v) whether sources of cost valuation and reference years were clearly reported (CHEC item 9). The remaining CHEC items—covering general study design, incremental analysis, discounting, sensitivity analysis, reporting quality, and ethical considerations—were outside the scope of this methodological review and are addressed narratively in the results where relevant rather than as formal quality criteria. Each item was rated using a four-point response scale: Yes, Partially, No, or Unclear, reflecting whether the criterion was fully met, partially met, not met, or unable to be assessed due to insufficient reporting. Quality assessment was conducted independently by the reviewers responsible for data extraction, with any discrepancies resolved through discussion and adjudication by K.-Y.J.C. To standardise the interpretation of each item and guide decisions in ambiguous cases, assessors referred to the instructions provided in the CHEC assessment guidelines [21].

2.4. Data Synthesis and Analysis

Findings were synthesized using a narrative approach, led by K.-Y.J.C., to address the study aims.

3. Results

3.1. Overview of Included Studies

This methodological review included 32 RCTs examining supported employment interventions for people with SMI (Table 1). Of these, seven studies incorporated a costing component and conducted an economic analysis.

Table 1. Characteristics of the included randomized control trials, 2009–2023.

First Author, Publication Year	Country	Study Aims	Sample Size	Population	Intervention	Comparator
Areberg et al., 2013 [22]	Sweden	To examine the effectiveness of IPS, as compared to TVR, in terms of occupational engagement, work-motivation, empowerment, and quality of life among people with SMI.	120	People with SMI, such as psychosis	IPS	TVR
Au et al., 2015 [12]	Hong Kong	To investigate the synergistic effects of CR training on Integrated Supported Employment for people with schizophrenia or schizoaffective disorder.	90	People with schizophrenia or schizoaffective disorder	Integrated Supported Employment (ISE) plus CR training: ISE blends individual placement support service with work-related social skills training, and CR training involves individualized, visual-based computer-assisted cognitive exercises	Integrated Supported Employment alone
Bejerholm et al., 2015 [23]	Sweden	To determine the effectiveness of IPS on vocational outcomes among people with SMI in a Swedish context. A secondary aim was to evaluate a community integration effect.	120	People with SMI	IPS	TVR
Bejerholm et al., 2017 [24]	Sweden	To study the effectiveness of a newly developed Individual Enabling and Support model adapted for people with affective disorders compared to TVR.	61	People with affective disorders (depression and bipolar disorder)	Individual Enabling and Support—a program that integrates motivational, cognitive, and time-use strategies with supported employment	TVR
Bell et al., 2014 [13]	USA	To determine whether augmenting supported employment with cognitive remediation can improve vocational outcomes, particularly for participants with lower community functioning.	175	People with schizophrenia or schizoaffective disorder	Supported employment + CR	Supported employment only
Bell et al., 2018 [25]	USA	To compare the effects of focused cognitive training performed in a cognitive laboratory with sham control using portable brain games on vocational, neurocognitive, and functional outcomes for participants with psychotic disorders in vocational rehabilitation.	77	People with schizophrenia or other psychosis	CR + vocational rehabilitation	Sham control (portable brain games) + vocational rehabilitation
Bond et al., 2013 [26]	USA	To compare job matching rates for clients with severe mental illness enrolled in two types of employment programs and examine the occupational matching hypothesis.	187	People with SMI, such as schizophrenia, bipolar disorder, or other psychotic disorder	IPS	Diversified Placement Approach
Bond et al., 2015 [27]	USA	To examine the effectiveness of the IPS model of supported employment for people with severe mental illness who have a history of arrest or incarceration.	87	People with SMI, such as schizophrenia, bipolar disorder, psychotic disorder, and justice involvement	IPS	Work Choice (a job club approach)
Bond et al., 2016 [28]	USA	To investigate the effects of IPS on young adults with severe mental illness.	109	Young adults with SMI	IPS	Alternative vocational services

Table 1. Cont.

First Author, Publication Year	Country	Study Aims	Sample Size	Population	Intervention	Comparator
Burns et al., 2009 [29]	UK, Germany, Italy, Switzerland, Bulgaria, The Netherlands	To explore associations between IPS, returning to work, and clinical and social outcomes.	312	People with SMI	IPS	Traditional vocational services
Christensen et al., 2019 * [14]	Denmark	To investigate the effects of IPS vs. IPS with enhancements vs. service as usual on a population of individuals with severe mental illness in Denmark.	720	People with SMI, such as Schizophrenia spectrum disorder, bipolar disorder, recurrent depression	IPS, IPS with enhancements (IPS + CR and work-focused social skills training)	Service as usual—vocational rehabilitation at job centres
Davis et al., 2012 [30]	USA	To examine the outcome of evidence-based supported employment for veterans with PTSD.	85	Veterans with PTSD	IPS	Veterans' Health Administration Vocational Rehabilitation Program
Davis et al., 2015 [31]	USA	To evaluate the feasibility, acceptability, and preliminary effectiveness of a mindfulness intervention for improving work outcomes in individuals with schizophrenia	34	People with schizophrenia or schizoaffective disorder	Mindfulness-Based Stress Reduction adapted for schizophrenia	Intensive Support
Davis et al., 2018 [15]	USA	To determine whether individual placement and support (IPS)—supported employment is more effective than stepwise vocational rehabilitation involving transitional work assignments at helping veterans with PTSD attain steady, competitive employment.	119	Veterans with PTSD	IPS	Treatment-as-usual: vocational rehabilitation services, including transitional work
DeTore et al., 2023 [32]	USA	To explore the effects of CR on employment outcomes for Black and White participants with SMI.	284	Black and White individuals with SMI, such as schizophrenia, schizoaffective disorder	CR + vocational rehabilitation	Vocational rehabilitation only
Drake et al., 2013 [33]	USA	To investigate whether a comprehensive package of insurance coverage, mental health treatments, vocational services, and elimination of disability reviews might enable Social Security Disability Insurance beneficiaries with psychiatric impairments to return to work and, secondarily, to improve their mental health status and quality of life.	2059	Social Security Disability Insurance beneficiaries with schizophrenia, bipolar disorder, or major depression	Multifaceted intervention: team-based supported employment, systematic medication management, other behavioral health services, complete health insurance coverage, suspension of disability reviews.	Usual services
Gold et al., 2016 [3]	USA	To evaluate whether participation in a certified Clubhouse supported employment program, relative to a vocationally integrated Assertive Community Treatment program, is associated with greater improvements in quality of life among individuals who obtain competitive employment, and whether longer duration of competitive employment within Clubhouse programs is associated with greater quality-of-life improvement.	167	People with SMI who are engaging in a competitive job	Certified Clubhouse	Vocationally Integrated Program of Assertive Community Treatment
Heslin et al., 2011 * [34]	England, UK	To investigate the effectiveness and cost-effectiveness of IPS in England in patients followed up for 2 years.	219	People with SMI	IPS	Local vocational services

Table 1. Cont.

First Author, Publication Year	Country	Study Aims	Sample Size	Population	Intervention	Comparator
Hoffmann et al., 2014 * [5]	Switzerland	To evaluate the long-term effects and social return on investment (SROI) of SE over a 5-year period.	100	People with SMI	SE	TVR
Holmås et al., 2021 * [35]	Norway	To provide empirical evidence on the long-term effect and cost effectiveness of IPS compared to relevant train-and-place programs.	327	People with moderate to severe mental illness, including psychotic or bipolar disorder, affective disorders, depression, and anxiety disorders	IPS	Treatment as usual
Howard et al., 2010 * [36]	UK	To investigate the effectiveness and cost-effectiveness of IPS in the UK.	219	People with SMI, such as psychotic or chronic affective disorder	IPS	TVR
Humm et al., 2014 [37]	USA	To assess the effectiveness of a virtual-reality role-play system in improving job interview skills and confidence for adults with serious mental illnesses.	96	People with SMI, such as autism spectrum disorder, schizophrenia, PTSD	Job Interview Training with Molly Porter, a virtual-reality role-play system	Wait-list Control group
Juurink et al., 2020 [38]	The Netherlands	To explore whether personality disorders moderate the effectiveness of IPS on employment outcomes.	146	People with SMI, including personality disorders	IPS	TVR
Kern et al., 2018 [39]	USA	To examine the efficacy of adding errorless learning, a behavioral training intervention, to evidence-based supported employment to improve SMI work outcomes.	162	People with SMI	Errorless learning combined with IPS—a method that minimizes errors during learning to improve job performance and tenure for individuals with serious mental illness	IPS alone
Knapp et al., 2013 * [40]	The Netherlands, UK, Italy, Bulgaria, Germany, and Switzerland	To evaluate the cost-effectiveness, budget impact, and overall economic impact of IPS.	312	People with SMI, such as schizophrenia and schizophrenia-like disorders, bipolar disorder, or depression with psychotic features	IPS	Standard vocational rehabilitation
Mervis et al., 2017 [41]	USA	To examine the impact of IVIP on defeatist beliefs, motivation for work, and work outcomes in serious mental illness.	64	People with schizophrenia-spectrum disorders	The CBT-based Indianapolis Vocational Intervention Program (IVIP) designed to reduce defeatist beliefs related to work functioning	Support therapy group with participant-led support and discussion about work-related issues and concerns
Michon et al., 2014 [42]	The Netherlands	To evaluate the effectiveness of IPS in the Dutch Socioeconomic context.	151	People with SMI	IPS	TVR
Mueller et al., 2019 [43]	USA	To evaluate the effectiveness of IPS on functioning in work or school, social, and interpersonal areas as specifically related to the symptoms of PTSD.	541	Veterans with PTSD	IPS	Supported employment involving transitional work
Mueser et al., 2011 [44]	USA	To compare the effectiveness of the IPS model of supported employment to control vocational rehabilitation programs for improving competitive work outcomes of people with severe mental illness and co-occurring substance use disorder.	106	People with SMI and co-occurring substance use disorder (dual disorder)	IPS	Conventional vocational rehabilitation programs
Mueser et al., 2014 [45]	USA	To compare supported employment with 2 other vocational rehabilitation programs in 3 ethnic/racial groups of participants with severe mental illness: Latinos, non-Latino African Americans, and non-Latino Whites.	204	Latinos with SMI	Supported employment	Psychosocial clubhouse program, standard services

Table 1. Cont.

First Author, Publication Year	Country	Study Aims	Sample Size	Population	Intervention	Comparator
Penk et al., 2010 [46]	USA	To compare the effectiveness of transitional work experience services and job placement services for veterans with co-morbid substance use and non-psychotic psychiatric disorders.	89	Veterans with co-morbid substance use and non-psychotic psychiatric disorders	Transitional Work Experience: Structured work settings typically in Veterans Administration facilities or private companies, with compensation for work	Job placement
Yamajguchi et al., 2017* [47]	Japan	To evaluate the cost-effectiveness of a cognitive remediation and supported employment program versus traditional vocational services in improving vocational outcomes and cognitive function in people with mental illness and low cognitive function.	111	People with mental illness and low cognitive function	CR + supported employment	TVS

* Studies include economic analyses. CR = Cognitive remediation. IPS = Individual placement and support. PTSD = Post-traumatic stress disorder. TVR = Transitional vocational rehabilitation. SMI = Severe mental illness.

Most included studies were conducted in the United States ($n = 18$, 56%). European countries also contributed substantially to the evidence base, with studies conducted in Sweden ($n = 3$), the Netherlands ($n = 2$), the United Kingdom ($n = 2$), Denmark ($n = 1$), Norway ($n = 1$), Switzerland ($n = 1$). Two studies were conducted in Asia: in Japan and Hong Kong. In addition, two studies were large-scale, multi-country European collaborative trials [29,40].

Across the included studies, the primary aim was to evaluate the effectiveness of supported employment interventions—most commonly IPS—compared with traditional vocational rehabilitation services or treatment as usual. Approximately one quarter of studies ($n = 8$) examined augmented or enhanced supported employment approaches. These studies investigated whether the addition of cognitive remediation [12–14,25,32,41], cognitive-behavioural therapy [41], work-focused social skills training [14], mindfulness-based interventions [31], or errorless learning techniques [39] could improve outcomes beyond supported employment alone.

Three-quarters of studies ($n = 24$) recruited participants with broadly defined SMI, typically including schizophrenia spectrum disorders, bipolar disorder, and major depression with psychotic features or a chronic course. Several studies focused on specific populations, including veterans with post-traumatic stress disorder (PTSD) [15,30,43], individuals with co-occurring mental illness and substance use disorders [44,46], people with personality disorders [38], young adults [28], Latino populations [45], individuals with justice involvement [27], and Social Security Disability Insurance (SSDI) beneficiaries [33].

Sample sizes varied considerably, ranging from 34 participants in a pilot feasibility study [31] to 2059 participants in a large-scale multi-site demonstration project [33]. The median sample size was approximately 120 participants, with most studies ($n = 20$, 63%) recruiting between 60 and 200 participants. Studies incorporating economic analyses tended to have larger sample size (median $n = 312$), likely reflecting the statistical requirements for robust cost-effectiveness analyses.

3.2. Outcome Selection: Domains, Priorities, and Conceptual Alignment

Across the 32 included trials, outcome selection reflected a strong and consistent prioritization of vocational outcomes, alongside substantial variability in the selection of secondary outcomes. Seven broad outcome domains were identified: (1) vocational outcomes; (2) recovery-oriented outcomes; (3) clinical or symptom outcomes; (4) cognitive

and neurocognitive outcomes; (5) functional outcomes; (6) support needs, service use, and welfare outcomes; and (7) process or implementation outcomes (Table 2). A detailed report of outcomes measured, measurement tools, and data collection methods is available in Supplementary Material S1: Complete data extraction on study characteristics, outcome measurement, economic analyses, and quality assessment.

3.2.1. Vocational Outcomes

Vocational outcomes were the dominant outcome domain and were assessed in all 32 included studies. Employment acquisition—typically defined as attainment of competitive paid employment—was the most frequently reported outcome, measured in 23 (72%) studies. Measures of work intensity and volume, such as hours worked per week, number of weeks worked, or total hours worked, were assessed in 21 studies, providing quantitative estimates of employment engagement beyond binary job attainment. Employment earnings were reported in 17 studies, usually as wages or salary. The majority of studies reported participants' actual earnings, whereas a smaller number—such as Christensen et al. (2020) [14] and Knapp et al. (2013) [40]—applied average market wages. Actual earnings capture the individual-level financial impact of employment, while market wages are intended to reflect the social value of production. Although the use of actual earnings data has strong face validity, it may underestimate social value when wages fall below marginal productivity as a result of labor market imperfections or discrimination against people with SMI. Notably, no study examined whether employment earnings were sufficient to lift participants above poverty thresholds or financial hardship benchmarks, despite the relevance of such indicators to social and economic policy objectives. Job tenure or employment duration was assessed in 13 studies and provided insight into employment stability and job retention—a key dimension of successful vocational rehabilitation. Time to first employment was examined in 10 studies, reflecting the IPS principle of rapid job placement and enabling differentiation from traditional “train-and-place” vocational models, which typically involved extended periods of pre-vocational preparation.

Less frequently assessed vocational outcomes included job satisfaction, job match, type of employment obtained, work performance or behavior, motivation to work, and job interview skills, each appearing in fewer than 10% of studies. These omissions are consequential: they privilege externally observable indicators of success over participant-centred, subjective experiences of work. A supported employment programme might achieve high rates of job acquisition while placing people in poorly matched, unsatisfying roles that fail to support recovery.

3.2.2. Quality-of-Life and Recovery-Orientation Outcomes

Recovery-oriented outcomes—constructs central to contemporary recovery frameworks in mental health—were assessed with substantial variability across studies. Quality of life was the most commonly measured construct within this domain, assessed in 11 studies (34%), but measurement approaches varied considerably, with six different instruments used. Self-esteem was measured in seven studies, all using the Rosenberg Self-Esteem Scale, representing one of the few areas of consistency in outcome measurement. Other recovery-oriented constructs—including empowerment, self-efficacy, intrinsic motivation, subjective well-being, recovery attitudes, and defeatist beliefs—were each measured in only one study.

Table 2. Outcome Measures and Instruments Used in Supported Employment Studies.

Outcome Domain	Outcome Measure	Psychometric Instrument	Studies Using This Measure
Vocational Outcome	Employment acquisition (competitive, paid) (<i>n</i> = 23)	NA	[3,5,13,15,22–25,28–30,32–36,38,40,42–44,46,47]
	Time to first employment (<i>n</i> = 10)	NA	[15,23,24,28,35,38,42–44,46]
	Job tenure (<i>n</i> = 13)	NA	[5,12,23,27–29,35,36,39,43–45,47]
	Work steadiness (<i>n</i> = 1)		[15]
	Income (salary, wage, hourly rate) (<i>n</i> = 17)	NA	[3,5,12,15,23–25,28,30,32,35,36,39,44–46]
	Job termination (<i>n</i> = 1)	NA	[12]
	Quantity and intensity of work (hours worked per week, number of weeks worked, total working hours) (<i>n</i> = 21)	NA	[3,5,13–15,23–26,30,32,35,36,39–46]
	Motivation to work (<i>n</i> = 2)	Worker Role Interview—Swedish version (WRI-S)	[22]
		Treatment Self-Regulation Questionnaire (TSRQ), Autonomous Motivation subscale	[41]
	Education, internships, prevocational activities (<i>n</i> = 2)	NA	[23,24]
	Job match (<i>n</i> = 2)	NA	[15,26]
	Type of employment obtained (<i>n</i> = 3)	Dartmouth Vocational Update Form	[3,15,27]
	Number of employers (<i>n</i> = 1)	NA	[35]
	Job satisfaction (<i>n</i> = 3)	Indiana Job Satisfaction Scale (IJSS)	[26,32,36]
	Attitudes toward employment (<i>n</i> = 1)	Employment Attitude Survey (EAS)	[41]
	Job interview skills (<i>n</i> = 1)	NA	[37]
	Work performance/behavior (<i>n</i> = 3)	Work Behavior Inventory (WBI)	[31,41,46]
Recovery-Oriented Outcome	Empowerment (<i>n</i> = 1)	The Empowerment Scale (ES)	[22]
	Intrinsic motivation (<i>n</i> = 1)	Derived from Quality-of-Life Scale (QLS)	[13]
	Self-efficacy (<i>n</i> = 1)	General Self-Efficacy Scale	[14]
	Self-esteem (<i>n</i> = 7)	Rosenberg Self-esteem Scale (RSES)	[3,12,14,34,36,41,42]
	Subjective well-being (<i>n</i> = 1)	Chinese version of Personal Well-Being Index—Adult (PWI)	[12]
	Quality of Life (<i>n</i> = 11)	Manchester Short Assessment of QOL (MANSA)	[22,24,34,42]
		Lancashire Quality of Life Profile-European Version (LQoLP-EU)	[29,36]
		Quality of Life Interview	[33]
		Lehman’s Brief Version of Quality-of-Life Interview (QOLI)	[3]
		Wisconsin Quality of Life Index (W-QLI)	[5]
		Quality of Life Scale (QLS)	[13,25]
	Self-reported recovery (<i>n</i> = 1)	Recovery Assessment Scale	[27]

Table 2. Cont.

Outcome Domain	Outcome Measure	Psychometric Instrument	Studies Using This Measure
Clinical/Symptom Outcome	Recovery attitudes ($n = 1$)	Recovery Process Inventory	[5]
	Self-perceived confidence ($n = 1$)	NA	[5]
	Defeatist beliefs ($n = 1$)	Dysfunctional Attitude Scale (DAS)	[41]
	Anxiety and/or depression ($n = 3$)	Hospital Anxiety and Depression Scale (HADS)	[29]
		Montgomery–Asberg Depression Self Rating Scale	[24]
		Hamilton Rating Scale for Depression (HAM-D)	[47]
	Psychiatric symptoms ($n = 8$)	18 item Brief Psychiatric Rating Scale (BPRS)	[12,34,36]
		Positive and Negative Syndrome Scale (PANSS)	[13,29,41,47]
		PTSD Checklist for DSM-5	[15]
	Mental or physical health ($n = 2$)	Mental health inventory-5 (MHI-5)	[33]
		12-item Short Form Health Survey	[42]
Cognitive/ Neurocognitive Outcome	Global cognition ($n = 5$)	MATRICES Consensus Cognitive Battery (MCCB)	[12,25,32]
		Brief Assessment of Cognition in Schizophrenia (BACS)	[14,47]
	Executive and attentional process ($n = 3$)	Wisconsin Card Sorting Test (WCST)	[12,13]
	Learning and memory ($n = 4$)	Wechsler Adult Intelligence Scale (WAIS), Trail Making A test	[13]
		Hongkong List Learning Test (HKLLT)	[12]
		Hopkins Verbal Learning Test (HVLTL)	[13]
		Brief Visuospatial Memory Test (BVMT)	[13]
		Wechsler Memory Scale (WMS)	[13]
	Overall functioning ($n = 5$)	Global Assessment of Functioning (GAF)	[12,29,34,36,47]
Functional Outcome	Social function/disability ($n = 3$)	Groningen Social Disability Scale (GSDS)	[29]
		Personal and Social Performance Scale (PSP)	[14]
		Social and Occupational Functioning Assessment Scale (SOFAS)	[41]
	PTSD-related functional impairment ($n = 1$)	Posttraumatic stress related inventory (PFRI)	[43]
	Occupational engagement ($n = 1$)	Profile of Occupational Engagement in people with SMI (POES)	[22]
Service Needs/ Utilisation Outcome	Met or unmet needs across life domains ($n = 2$)	Camberwell Assessment of Need (CAN)	[34,36]
	Mental health care utilisation (e.g., inpatient, outpatient) ($n = 9$)	NA	[5,14,27,29,34–36,40,47]
	Benefits received (e.g., social assistance, unemployment benefits, disability pension, and work assessment allowance) ($n = 1$)	NA	[35]

Table 2. *Cont.*

Outcome Domain	Outcome Measure	Psychometric Instrument	Studies Using This Measure
Process/ Implementation Outcome	Fidelity (<i>n</i> = 4)	Mindfulness Fidelity Scale (MFS)	[31]
		IPS Fidelity monitor	[15]
		15-item Individual Placement and Support Fidelity Scale	[33]
		Quality of Supported Employment Implementation Scale (QSEIS)	[42]
	Service satisfaction (<i>n</i> = 2)	Service Evaluation Questionnaire (SEQ)	[3]
		Client Satisfaction Questionnaire (CSQ-8)	[31]

NA = Not applicable.

European studies showed a greater tendency to include quality-of-life and recovery-oriented outcomes than US-based studies—quality of life was assessed in seven of twelve (58%) European studies compared with three of eighteen (17%) US studies, possibly reflecting differences in evaluation traditions and policy frameworks. The inclusion of recovery-oriented outcomes is theoretically and clinically important in the context of supported employment, as these interventions aim not only to increase employment but also to enhance autonomy, self-worth, social inclusion, and subjective personal recovery. Omitting these outcomes risks underestimating the full impact of supported employment and reinforces a narrow, productivity-focused conception of effectiveness.

3.2.3. Clinical, Cognitive, and Functional Outcomes

Clinical outcomes—primarily psychiatric symptom severity—were assessed in 11 studies, most commonly using the Positive and Negative Syndrome Scale (PANSS) or the Brief Psychiatric Rating Scale (BPRS). Cognitive and neurocognitive outcomes were measured in six studies, predominantly in trials evaluating cognitive remediation or related augmentation strategies. Functional outcomes—assessing activities of daily living, social functioning, or independent living skills—were assessed in eight studies. The Global Assessment of Functioning (GAF) dominated this domain, used in five studies.

The limited and uneven inclusion of clinical, cognitive, and functional outcomes suggests that these domains were often treated as explanatory or exploratory rather than core components of effectiveness. This restricts examination of bidirectional relationships—for example, whether employment influences symptoms or functioning over time, or whether changes in symptoms or functioning affect job retention and stability.

3.2.4. Support Needs, Service Utilisation and Welfare Receipt

Support needs were assessed in two studies, both using the Camberwell Assessment of Need (CAN). Only one study measured changes in welfare receipt as a result of supported employment. Service utilisation—such as inpatient stays, outpatient visits, and medication use—was measured in nine studies; however, most framed these as “costs” (or cost savings, indicating reduced service utilisation) within economic analyses rather than as indicators of supported employment effectiveness. This framing risks misinterpreting increased service contact as a negative consequence, rather than as a potential indicator of appropriate care-seeking, treatment engagement, or improved access to services.

3.2.5. Implementation Outcomes

Implementation outcomes received minimal attention, with only four studies assessing programme fidelity and two assessing participant satisfaction. It is possible that process

evaluations were conducted at different stages of programme development, such as in earlier feasibility studies, or reported in separate publications not captured in the primary trial reports. It is worth noting that the limited integration of process measures within efficacy trials has important implications. For complex, multi-component interventions such as IPS, fidelity assessment helps determine whether the intervention was implemented as intended and may help explain inconsistent outcomes across sites and contexts. In addition, implementation outcomes can provide evidence on acceptability, adoption, appropriateness, penetration, and sustainability of supported employment programmes in real-world settings beyond controlled trial conditions.

3.3. Economics Analysis: Analytical Approaches, Costing Methods, and Implications

Table 3 presents a summary of the economic analysis methods, costing approaches, and data collection processes. The following observations are descriptive and illustrative, mapping the costing approaches employed across the seven studies that incorporated an economic component; no pooled synthesis is presented given the methodological heterogeneity across studies.

3.3.1. Types of Economic Analysis

Only seven of the included studies incorporated an economic analysis alongside the randomized controlled trial of supported employment interventions [5,14,34–36,40,47]. Cost-effectiveness analysis (CEA) was the most frequently employed approach, used in four studies [14,34,40,47]. CEA measured effectiveness using natural units, such as competitive employment attainment or number of days worked. While these analyses are intuitively aligned with supported employment objectives, they do not allow comparison with interventions targeting different health or social outcomes and therefore have limited application for cross-sectoral resource allocation decisions [48].

One study extended CEA to include a cost–utility analysis (CUA), expressing outcomes in quality-adjusted life years (QALYs) [14]. By paring intervention-related outcomes with a generic health utility measure, this combined approach enables comparison across intervention types while retaining sensitivity to outcomes central to supported employment. Such dual reporting is often recommended in health economic evaluation guidelines [48,49].

Although labelled as a CEA, the economic analysis conducted by Howard et al. (2010) [36] is more accurately classified as a cost–consequence analysis (CCA), as it reports costs alongside multiple disaggregated outcomes—including employment outcomes, job satisfaction, unmet needs, self-esteem, quality of life, psychiatric state, and global functioning—without synthesizing them into a single summary measure such as an incremental cost-effectiveness ratio (ICER) or net benefit statistic. While CCA preserves the full complexity of multi-dimensional outcomes and allows decision-makers to apply their own value judgements about the relative importance of different outcomes, it requires stakeholders to evaluate which combinations of costs and effects constitute good value for money within their specific policy context—a task that may be challenging without explicit cost-effectiveness thresholds or benchmarks.

Table 3. Characteristics of economic analyses in supported employment studies.

Author, Publication Year	Type of Economic Analysis	Study Perspective	Time Horizon	Discount Rate	Uncertainty Analysis	Cost Categories and Cost Items	Cost Measurement (in Physical Units)	Cost Valuation	Data Collection Methods
Christensen et al., 2020 [14]	CEA & CUA	Societal perspective	18 months	NA	Parameter uncertainty: Probabilistic sensitivity analysis	C1. Intervention costs, incl. IPS and IPSE C2. Hospital costs C3. Primary health care costs C4. Pharmaceuticals C5. Labor market interventions (e.g., counselling, mentor support, vocational rehabilitation) C6. Municipal social interventions (e.g., counselling, psychosocial initiatives, personal assistance) C7. Productivity costs	C1: Per minute, plus an additional €600 per patient for IPSE C2: Inpatient care per day; other services per visit C3: Per visit C4: Per prescription C5: Per hour C6: Per group C7: Per hour	C1: Mean costs for the site C2: DRG tariffs C3: National service tariffs C4: Full price of prescription drugs (regardless of subsidies), as purchased in Danish pharmacies C5: Hourly costs obtained from the Danish Agency for Labor Market and Recruitment C6: Mean costs per group obtained from municipal data C7: Hours in competitive employment multiplied by the average wage	C1: Patient registration recorded by programme staff Data collected from various agencies: C2: The National Patient Register C3 & C4: The National Health Service Register C5: The Danish Agency for Labor Market and Recruitment C6: Municipal administrative data C7: Electronic income register from the Danish Agency for Labor Market and Recruitment
Heslin et al., 2011 [34]	CEA	Not explicitly stated by the authors; inferred from cost categories as a health and social care system perspective.	24 months	Not reported	Sampling uncertainty: bootstrapping to produce CIs Decision uncertainty: CEAC	C1. Intervention costs C2. Health & social care costs (e.g., psychiatric inpatient, general inpatient, psychiatrist, occupational therapist, district nurse, social care, day care/education) C3. Medications (for psychosis, mania, depression, etc. and for treating side effects)	C1: Per contact and time spent C2: Not specifically reported C3: Self-reported dose or lowest recommended dose	C1 & C2: Curtis (2007), Unit Costs of Health and Social Care for health and social care services [50] C3: Prices obtained from the March 2006 British National Formulary	C1–C3: Collected via the Client Service Receipt Inventory (CSRI)
Hoffmann et al., 2014 [5]	CBA	Societal perspective/social return on investment (SROI)	5 years	Not reported	Not reported	C1. Total vocational program costs C2. Total mental health treatment costs, incl. in psychiatric hospitals and day treatment, outpatient mental health services, and medication	C1 & C2. Per person	Not reported	Collected monthly from participants

Table 3. Cont.

Author, Publication Year	Type of Economic Analysis	Study Perspective	Time Horizon	Discount Rate	Uncertainty Analysis	Cost Categories and Cost Items	Cost Measurement (in Physical Units)	Cost Valuation	Data Collection Methods
Holmås et al., 2021 [35]	CBA	Not explicitly stated by the authors; inferred from cost categories as a government fiscal perspective	43 months	Not applied	Not reported	C1. Intervention costs Monetary benefits: B1. Increased labor income B2. Reduced welfare expenditures, incl. social assistance, unemployment benefits, work assessment allowance, and disability pension	C1, B1 & B2. Per person	C1. Unit costs of IPS were obtained from the Directorate of Labor and Welfare and the Directorate of Health; unit costs of traditional active labor market programmes were derived from administrative budget and programme expenditure data. B1. Labor income was obtained from NAV's State Register of Employers and Employees (SREE) and income registers. B2. Welfare expenditures were obtained from the Directorate of Labor and Welfare.	C1, B1, B2. Data were obtained through linkage to government administrative data, recorded on a monthly basis over the observation period.
Howard et al., 2010 [36]	CCA	Not explicitly stated by the authors; inferred from cost categories as a health and social care system perspective.	12 months	NA	Sampling uncertainty: non-parametric bootstrapping (10,000 replications) of cost data	C1. Intervention costs, including direct and indirect client time C2. Health care services, including psychiatric and general inpatient care; psychiatrist, general practitioner, doctor, psychologist/therapy, OT, district nurse, and community mental health nurse C3. Social care services, including day care/education and social care	C1: Per hour C2 & C3: Per contact	C1: Based on the employment worker's annual salary and a ratio of direct to indirect client time of 0.67 C2 & C3: Not explicitly reported in the paper, but referenced to Curtis (2007), Unit Costs of Health and Social Care, Personal Social Services Research Unit [50]	C1: Administrative data collected alongside the RCT C2 & C3: Collected through interviews using the Client Service Receipt Inventory (CSRI) over 12 months

Table 3. Cont.

Author, Publication Year	Type of Economic Analysis	Study Perspective	Time Horizon	Discount Rate	Uncertainty Analysis	Cost Categories and Cost Items	Cost Measurement (in Physical Units)	Cost Valuation	Data Collection Methods
Knapp et al., 2013 [40]	CEA & partial CBA	Health and social care system perspective	18 months	NA	Sampling uncertainty: non-parametric bootstrapping Decision uncertainty: CEACs	CI. Intervention costs, incl. labor, overhead costs C2. Mental health costs, incl. inpatient services, outpatient services, community-based services/profession, medication C3. Social care costs, incl. accommodation	C1–C3. Per contact	C1. UK unit costs to human resources. Where necessary, costs from different countries were converted into British pounds using purchasing power parities and then adjusted for inflation so that all costs are expressed in 2003 UK price. C2 & C3. unit costs for 2003 were taken from the annual Personal Social Services Research Unit (PSSRU) volume for England.	C1. Collected locally from trial sites throughout the trial. C2 & C3. Collected using Client Socio-demographic and Service Receipt Inventory—European Version (CSSRI-EU) at baseline, and 6, 12, and 18 months later.
Yamaguchi et al., 2017 [47]	CEA	Healthcare and social care system perspective	12 months	NA	Sampling uncertainty analyses: bootstrapping, Decision uncertainty: CEAC	C1. Intervention—CR C2. Intervention—SE, TVS C3. Medical services (e.g., out-patient care, in-patient care) C4. Medication C5. Other social services (e.g., sheltered workshops, other community services, public sector services, accommodation)	Not reported	C2 & C5. Unit cost information from the Service and Supports for Persons with Disabilities Act (SSPDA), incl. salaries, overheads, and training. C1 & C3. Prices on the receipts; unit costs from National Health Insurance, or SSPDA C4. Prices on the receipts	C1–C5 Service utilization was collected by the care manager using CSRI-J during monthly interviews

CBA = Cost–benefit analysis. CCA = Cost–consequence analysis. CEA = Cost-effectiveness analysis. CEAC = Cost-effectiveness acceptability curve. CI = Confidence interval. CUA = Cost–utility analysis. NA = Not applicable. RCT = Randomized controlled trial.

Three studies conducted cost–benefit analyses (CBA) [5,35,39], monetizing both costs and benefits to estimate net benefits or benefit–cost ratios. However, there was substantial heterogeneity in how the benefits were defined and valued. Holmås et al. (2021) calculated net benefits as the sum of increased labor income and reduced benefits payments minus intervention costs, reflecting a narrow fiscal perspective centred on government budgetary impacts [35]. Knapp et al. (2013) focused primarily on gross wage gains as benefits, explicitly excluding health and quality-of-life outcomes, and therefore characterized their analysis as a “partial” CBA [40]. Hoffmann et al. (2014) framed their evaluation as a social return on investment (SROI) analysis, calculating ratios of earnings relative to vocational programme costs and mental health service expenditures [5]. These differences substantially affect conclusions about value for money and limit comparability across studies.

3.3.2. Analytical Perspective and Cost Identification

Choices regarding analytical perspective varied widely and were not always explicitly reported. Cost identification should be aligned with both the analytical perspective and the research question. The analytical perspective establishes the boundary of the evaluation—specifically, whose costs are considered relevant—while the research questions further inform key costing decisions. These include the prioritization of cost components (e.g., primary vs. secondary), the level of detail (micro-costing versus macro-costing), and the scope of analysis with respect to population, intervention, comparator, geographical coverage, and time horizon. Within this framework, any exclusion of key cost items should be explicitly justified. Accordingly, we assessed the quality of cost identification by asking whether authors fully identify all important and relevant costs in relation to the stated perspective and research question.

Christensen et al. (2020) [14] adopted the most comprehensive societal perspective analysis, aiming to investigate the cost–utility and cost-effectiveness of IPS with and without CR and social skills training for people with SMI. Although the authors identified major cost drivers, including healthcare sector costs, labor market services, municipal social services, and productivity impacts, they did not include informal care or participant out-of-pocket costs, which would be expected under a societal perspective.

Both Knapp et al. (2013) [40] and Yamaguchi et al. (2017) [47] adopted a health and social care perspective. Each study captured costs related to hospital inpatient care, outpatient care, medication, community-based services, accommodation, and vocational employment services. Yamaguchi et al. (2017) [47] also included public sector services, such as city offices, unemployment offices, and ‘sheltered workshops’, and presented more granular cost categories overall.

Hoffmann et al. (2014) [5] explicitly framed their analysis as an assessment of SROI, implying a societal perspective that would incorporate all relevant costs and benefits accruing to society. In practice, however, the implemented analysis was substantially narrower. Benefits were limited to participants’ earnings from competitive and non-competitive employment over a five-year period, while costs were confined to intervention costs and mental health treatment and medication costs. Notably absent from this purportedly societal analysis were other healthcare costs beyond mental health services, social care costs, impacts on family caregivers, productivity losses during unemployment, and other broader societal impacts. This mismatch between the stated SROI framework and the restricted set of included cost and benefit components represents a significant methodological inconsistency that undermines comparability with other SROI analyses and risks misleading policymakers who may draw on the reported figure without recognizing the limitations of its scope.

Heslin et al. (2011) [34], Holmås et al. (2021) [35] and Howard et al. (2010) [36] did not explicitly report their analytical perspective. Therefore, it is not possible to assess whether all important and relevant costs were appropriately identified in relation to the perspective and research question.

3.3.3. Cost Measurement and Reporting Transparency

Best practice in economic evaluations requires reporting physical quantities of resources consumed and their respective unit costs separately, rather than aggregated totals alone. This disaggregated approach enhances transferability across settings and time periods for several reasons. First, physical quantities are relatively stable and transferable, whereas unit costs vary substantially across countries, regions, and time due to differences in wages, overhead structures, pricing policies, and currency values. Second, separate reporting enables verification of cost calculations and assessment of their plausibility. Third, disaggregation distinguishes intervention effectiveness (reflected in resource quantities) from economic efficiency (reflected in unit costs), clarifying whether high costs arise from high resource utilization, high unit costs, or both. Finally, disaggregated reporting facilitates sensitivity analyses by allowing alternative unit costs to be applied under different assumptions, settings, or time horizons—analyses that are impossible with aggregated totals alone. Accordingly, we assessed the quality of cost measurement by examining whether authors appropriately measured costs using physical units of resource use.

Across the seven included economic analysis studies, only one study (Howard et al., 2010 [36]) demonstrated full transparency, reporting physical quantities of resource use (e.g., mean number of contacts) alongside total costs for each service category, with clear references to external sources for unit cost estimates. Three studies demonstrated substantial but incomplete transparency. Knapp et al. (2013) [40] and Christensen et al. (2020) [14] reported costs by service category with referenced unit cost sources, but did not present underlying physical quantities (e.g., number of visits). Back-calculation of physical quantities is not feasible because their reported categories aggregate heterogeneous services with different unit costs. Yamaguchi et al. (2017) [47] reported “mean times of service utilisation per participant” in Supplementary Materials but did not provide unit costs; moreover, the cited Japanese-language sources limit international transparency and reproducibility. The remaining three studies [5,34,35] demonstrated inadequate transparency, reporting only aggregated cost totals without physical quantities or unit cost information.

3.4. Data Sources and Data Collection Approaches

Across the included studies, data sources and data collection approaches varied systematically by outcome and cost domain, reflecting differing trade-offs between objectivity, feasibility, respondent burden, and analytic scope (see Supplementary Material S1 for data collection methods used by each included study). Four broad categories of data collection approaches were identified: (1) administrative data sources, (2) participant self-report methods, (3) researcher-administered assessments, and (4) direct observation and tracking. The majority of studies ($n = 27$, 90%) employed multiple approaches in combination, selecting methods reflected pragmatic considerations of data availability, resource constraints, and the specific outcomes prioritized in each evaluation.

3.4.1. Administrative Data Sources

Administrative data, accessed via government-maintained registers, welfare databases, healthcare administrative records, and organizational record systems, served as primary source for vocational outcomes and for service utilisation and cost measurement. For example, Holmås et al. (2021) [35] constructed employment outcomes algorithmically from Norway’s State Register of Employers and Employees on a monthly basis over

43 months. They also obtained healthcare utilisation and welfare benefit receipt data from administrative health care records and national welfare registers, respectively.

The primary strength of administrative data lies in objectivity and reduced susceptibility to recall bias. Because these data are collected continuously as part of routine operations, they enable comprehensive, long-term follow-up without ongoing participant burden. However, data quality is dependent on the accuracy of initial data entry and coding practices, which may vary across systems and over time. Access to administrative data was also constrained by institutional, legal, and technical barriers, with many studies lacking the data-sharing agreements or linkage infrastructure necessary to utilize these sources. Consequently, reliance on administrative data was feasible primarily in jurisdictions with well-developed registry systems and established protocols for research access.

3.4.2. Participant Self-Report Methods

Participant self-report methods involve data collection directly from study participants, either through self-administered questionnaires, telephone or in-person interviews, or participant-maintained records such as job logs and diaries. Self-report methods served as the primary source for subjective and experiential outcomes, including job satisfaction and workplace experiences that extend beyond objective labor market indicators. Recovery-oriented outcomes, such as quality of life, self-esteem, empowerment and perceived social functioning, were also self-assessed.

For service utilisation and cost measurement, self-report methods were employed both as primary source when administrative data were unavailable and as a supplementary source to capture services and costs outside formal systems, including informal supports and out-of-pocket costs. For example, in Heslin et al. (2011) [34] and Howard et al. (2010) [36], a validated instrument, the Client Service Receipt Inventory (CSRI), was used for retrospective service-use reporting.

While self-report methods are necessary to capture participants' own perspectives, they are also subject to recall bias, particularly when recall periods are extended or when participants have cognitive impairments that affect memory. Some studies attempted to address this issue by validating self-reported records against administrative data. Self-report measures may also be prone to social desirability bias, limited insight, or difficulty in quantifying complex experiences.

3.4.3. Researcher-Administered Assessments

Researcher-administered assessments involve trained personnel conducting standardized evaluations using validated psychometric instruments, delivered through structured interviews, performance-based testing, or clinical examinations. These assessments served as the primary source for clinical, cognitive, and functional outcomes, with independent assessors often blinded to intervention allocation to reduce the risk of measurement bias.

Vocational readiness, work-related social skills, and job interview competencies were also evaluated by trained personnel capable of administering complex protocols and scoring participant performance using standardized criteria. For example, Humm et al. (2014) [37] conducted in-person role-plays at baseline and post-intervention to assess work-related social skills, while Kern et al. (2018) [39] administered semi-structured interviews at baseline and after completion of the intervention to evaluate vocational readiness.

The key strength of research-administered assessments lies in their standardization and the use of validated instruments with established psychometric properties, which reduce random measurement error. Blinding of assessors to treatment allocation, when feasible, reduces the risk of systematic bias in outcome measurement. These assessments

also enable the measurement of complex constructs that require clinical judgement or specialized expertise, such as neurocognitive functioning.

However, these methods are resource-intensive, requiring trained personnel, ongoing quality assurance, and substantial time commitment from both assessors and participants. Data quality depends heavily on assessor training and adherence to standardized protocols. Several studies provided incomplete reporting of blinding procedure or inter-rater reliability, limiting the ability to assess potential measurement bias.

3.4.4. Direct Observation and Tracking

Direct observation and tracking involve systematic documentation by program staff, performance observation, employer verification, and fidelity assessment. Program staff tracking of vocational outcomes was among the most common data collection approaches for employment-related outcomes. In addition, direct observation was used to assess work performance, work behavior, and skill demonstration in structured or naturalistic settings. Davis et al. (2015) [31] collected supervisor assessments using the Work Behavior Inventory and tracked work performance longitudinally at multiple time points, whereas Mervis et al. (2017) [41] rated work performance based on direct observation at work therapy sites.

Intervention fidelity was assessed through systematic observation, document review, and staff interviews to ensure treatment integrity. Davis et al. (2015) [31] employed session videotaping combined with independent fidelity ratings to assess intervention delivery and participant engagement.

The primary advantages of direct observation and program tracking include real-time or near-real-time data capture, which reduces recall bias, and integration into routine service delivery, which minimizes additional participant burden. Program staff who are in regular contact with participants are well positioned to document employment changes. However, the quality of data collection through program tracking and observation is heavily dependent on staff training, adherence to documentation protocols, and consistency across sites and over time. Staff turnover can introduce discontinuities in data collection and variation in documentation practices. Intensive observation may also disrupt routine workflows, increase resource requirements, and may not be feasible in all settings.

3.4.5. Temporal Considerations

Follow-up durations varied considerably across studies, ranging from relatively short follow-ups (6 months) to extended observation periods (5 years), with most studies employing 12- to 24-month observation periods. This variation has important implications for both cost and outcome measurement.

Vocational outcomes may show different patterns over time, with initial job acquisition potentially followed by job loss, re-employment, or sustained work particularly where inappropriate job matching has occurred. Studies with longer follow-ups were better positioned to assess employment sustainability and trajectory, while shorter studies primarily captured immediate post-intervention effects. Clinical, functional, and psychosocial outcomes may also evolve over time, with potential for delayed benefits as individuals adjust to employment or conversely, initial improvements that attenuate with sustained work stress.

The cost implications of temporal variation are substantial. Studies with longer follow-up periods capture more complete cost trajectories, including potential cost offsets from reduced health or social service use or, conversely, delayed costs associated with service re-engagement. In contrast, studies with shorter follow-up periods may either underestimate long-term costs if initial cost reductions prove unsustainable, or overestimate costs if upfront intervention expenditures are not offset by subsequent savings.

Table 4 summarizes the key methodological limitations across the three domains examined in this review—outcome measurement, economic analysis, and data collection—and outlines their implications for the interpretation, comparability, and policy relevance of supported employment evaluations.

Table 4. Summary of methodological limitations identified across supported employment studies and their implications for evidence quality and policy translation.

Methodological Domain	Descriptive Findings	Methodological Implications
Outcome measurement		
Outcome selection	- Vocational outcomes prioritised consistently across all studies - Recovery-oriented, cognitive, functional, and implementation outcomes measured inconsistently	- Overemphasis on employment attainment privileges externally observable indicators over participant-centred experiences of work - A programme may achieve high job acquisition rates while overlooking participant-centred experiences of work and recovery
Recovery-oriented and quality-of-life outcomes	- Quality of life assessed in approximately one-third of studies, using six different instruments - Constructs such as empowerment, self-efficacy, and defeatist beliefs rarely included - European studies more likely to include these outcomes than US studies	- Varied instruments limit comparability across studies - Omitting recovery-oriented outcomes risks underestimating the full impact of supported employment and reinforces a narrow, productivity-focused conception of effectiveness
Clinical, cognitive, and functional outcomes	- Treated as explanatory or exploratory rather than core components of effectiveness - Included mainly when the intervention specifically targeted these domains	- Limits examination of bidirectional relationships between employment, symptoms, and functioning - Cannot determine whether employment influences symptoms or functioning over time, or whether changes in symptoms or functioning affect job retention and stability
Service utilisation and welfare outcomes	- Service use typically framed as a cost variable rather than an outcome indicator - Welfare benefit dependency and transitions off benefits were largely unexamined	Framing service contact as a cost risks misinterpreting appropriate care-seeking, treatment engagement, or improved service access as a negative outcome
Implementation and fidelity outcomes	- Programme fidelity and participant satisfaction assessed in very few studies - Process evaluations often absent from trial publications	- Without fidelity data, it is difficult to determine whether the intervention was delivered as intended, or to explain inconsistent outcomes across sites - Evidence on real-world acceptability, adoption, and sustainability is limited
Economic analysis		
Coverage and type of economic analysis	- Minority of studies incorporated any economic analysis - Multiple analytical types used (CEA, CUA, CBA, CCA, SROI), with inconsistent definitions of costs and benefits	- CEA using natural units does not permit comparison with interventions targeting different outcomes, limiting application for cross-sectoral resource allocation - Heterogeneity of methods prevents direct comparison of value for money across studies
Analytical perspective and cost identification	- Analytical perspective frequently unstated or inconsistently applied - Stated perspective sometimes broader than the cost categories actually included	- Impossible to verify whether all relevant costs were captured in relation to the stated perspective and research question - Mismatch between stated and implemented perspective may mislead policymakers about the scope of evidence
Cost measurement and reporting transparency	Few studies disaggregated costs into physical resource quantities and unit costs	- Aggregated reporting prevents verification of cost calculations and assessment of their plausibility - Without disaggregation, sensitivity analyses using alternative unit costs are not possible, limiting transferability across settings and time periods
Data collection		
Data collection approaches	- Administrative data feasible mainly in jurisdictions with established registry systems - Self-report used widely for psychosocial outcomes; reporting of blinding and inter-rater reliability often incomplete	- Variation in registry infrastructure and data availability across jurisdictions limits comparability - Incomplete reporting of measurement procedures makes assessment of potential bias difficult

Table 4. *Cont.*

Methodological Domain	Descriptive Findings	Methodological Implications
Follow-up duration	• Most studies used short to medium follow-up periods • Few studies observed participants long enough to assess employment sustainability or downstream cost offsets	• Short follow-up may underestimate long-term costs if initial reductions prove unsustainable, or overestimate costs if upfront expenditures are not offset by subsequent savings • Longer-term employment sustainability and transitions off welfare benefits may not be captured within typical study windows
Contextual documentation	• Limited reporting of local labour market, welfare system, and healthcare delivery context • Studies concentrated in the US, with limited representation of other welfare and labour market contexts • Without explicit contextual documentation, the extent to which findings are transferable to other settings cannot be assessed	

4. Discussion

4.1. Summary of Principal Findings

This methodological review examined how costs and outcomes have been identified, measured, and reported in RCTs of supported employment programs for people with SMI in high-income countries. The principal finding is that methodological heterogeneity across studies is extensive, encompassing substantial variation in analytical perspectives, outcome selection, costing methods, data collection approaches and temporal design. This heterogeneity is neither unexpected nor inherently problematic. Rather, it reflects the fact that evaluations have been designed to address different research questions, inform different policy decisions, and operate within distinct institutional, labor market, and fiscal contexts. For instance, studies designed to test a specific augmentation mechanism—such as cognitive remediation—appropriately included neurocognitive outcome measures, while the use of ‘work without wage subsidies’ as an outcome, as employed by Holmås et al. (2021) [35], reflects the Norwegian welfare system’s policy objective of reducing long-term benefit expenditure rather than simply achieving job placement. Imposing methodological uniformity across all such evaluations would risk producing evidence that is technically standardized but contextually irrelevant, thereby limiting its utility for local policy and service planning.

However, this review also identified critical gaps in the transparency and completeness of reporting that substantially limit the interpretability and usability of the evidence base. Many studies provided insufficient documentation of key methodological choices and their rationale, incomplete specification of cost components and valuation methods, limited description of data collection procedures, and inadequate discussion of how methodological decisions affect transferability to other contexts. Few studies followed the established reporting guidelines such as CONSORT (Consolidated Standards of Reporting Trials) for RCTs [51] or CHEERS (Consolidated Health Economic Evaluation Reporting Standards) for economic evaluations [49]. These reporting deficiencies create challenges for critical appraisal, evidence synthesis, and appropriate application of findings [52].

4.2. Practical Guidance for Program Evaluators

Building on the methodological strengths and weaknesses identified across the 32 included trials (Sections 3.2–3.4), and drawing on established evaluation methodology standards, we propose practical guidance for practitioners and evaluators designing evaluations of supported employment programs for people with SMI. This guidance is organized around three key domains: (i) aligning study design with decision-making needs, (ii) transparent costing, and (iii) selecting and measuring outcomes. A summary is provided in Table 5. This guidance is not intended to offer advice on formal economic evaluation or

statistical analysis; for such work, we encourage collaboration with health economists and statisticians as early as the design phase.

Table 5. Practical guidance for evaluating supported employment programmes for people with SMI.

Domain	Key Design Question	Practical Guidance	Common Pitfalls Addressed
1. Decision alignment	Who will use the evidence, and for what decision?	Identify intended decision-makers (e.g., funders, policymakers, service providers, participants) and the specific decision the evaluation is intended to inform (e.g., continuation, scale-up, redesign). Engage decision-makers early to clarify evidence needs and constraints.	Misalignment between study design and policy or funding decisions
	What analytical perspective is appropriate?	Explicitly state and justify the study perspective (service user, employment services, health system, welfare, societal). Ensure that all costs and outcomes included are consistent with this perspective.	Implicit or inconsistent perspectives leading to incomplete costing
	What time horizon is required?	Justify the evaluation time horizon based on intervention mechanisms and decision needs. Minimum follow-up should capture stabilization of primary outcomes (typically ≥ 12 –18 months). Where feasible, extend follow-up to 24–36 months to assess sustainability, cost offsets, and longer-term trajectories.	Follow-up too short to capture durable effects or downstream costs
	How will contextual transferability be supported?	Document intervention delivery context, provider characteristics, organizational setting, fidelity monitoring, participant clinical and employment histories, and community-level contextual factors (e.g., labor markets, welfare systems, healthcare context). Discuss implications for transferability.	Poor external validity and limited interpretability across settings
2. Transparent costing	Which cost categories are relevant?	Identify all relevant cost components based on perspective, including intervention delivery costs and downstream health, welfare, and societal costs where applicable. Acknowledge omitted cost categories explicitly.	Selective or opaque cost inclusion
	How are costs measured and valued?	Separate resource quantities from unit costs. Use micro-costing for novel interventions and justified gross-costing for established programmes. Adjust costs for inflation and report reference year. Use PPP for international comparisons.	Non-reproducible or non-transferable cost estimates
	How is uncertainty addressed?	Conduct sensitivity analyses on key cost assumptions and unit costs. Clearly report assumptions and allocation methods for shared resources.	False precision and untested assumptions
3. Outcomes measurement	Which outcomes matter, and why?	Select outcomes based on theory of change and stakeholder priorities. Pre-specify primary and secondary outcomes and clearly label exploratory analyses. Avoid unjustified outcome proliferation.	Overburdened measurement and interpretive ambiguity
	How should outcomes be measured?	Match measurement method to outcome type and time frame. Prefer administrative data for employment and service use where available; supplement with validated self-report or qualitative measures for person-centred outcomes.	Mismatch between outcome constructs and measurement tools
	How is measurement quality ensured?	Specify assessor training, blinding, instrument versions, timing, handling of missing data, and reliability checks. Align measurement timing with expected intervention effects.	Measurement bias, missing data, and inconsistent timing

(i) Aligning study design with decision-making needs

Evaluators should begin by identifying who will use the evidence generated—whether policymakers, funders, service providers, or participants themselves—and what specific decision the evaluation needs to inform, such as funding allocation, program adoption, or service redesign. Early engagement with decision-makers is essential to understand their evidence needs, practical constraints, and decision contexts in which findings will be applied. As highlighted in Section 3.3.2, the study perspectives must be made explicit and justified based on the decision context, as perspective fundamentally determines which costs and outcomes are relevant [53,54].

As documented in Section 3.4.5, the majority of included studies applied follow-up periods of 12 months or less. The evaluation time horizon should be justified by both the

temporal dynamics of the intervention and the informational needs of decision-makers. Evidence from supported employment research indicates that employment outcomes do not stabilise rapidly. McGurk et al. (2016) [55] suggest that employment trajectories in IPS programmes often require 24 months or more to stabilise, and that follow-up periods of at least two years are necessary to assess programme durability and secondary outcomes such as service use and relapse. Consistent with this, Christensen et al. (2023) [56] found that employment effects observed at 18 months were sustained over a 30-month follow-up period. Moreover, longer registry-based follow-ups demonstrate that transitions off benefits and downstream cost implications may emerge only over extended time horizons and are therefore not captured in shorter studies [35]. Taken together, these findings suggest that while a minimum follow-up of 12–18 months may be sufficient to detect initial effects, robust evaluation of sustainability, cost offsets, and longer-term outcome trajectories generally requires observation periods of 24–36 months or longer. Pragmatic considerations, such as funding constraints, may limit follow-up duration in practice. In such cases, evaluators could consider modelling longer-term outcomes using sensitivity analyses or decision-analytic modelling approaches, while transparently documenting the assumptions underlying such projections [48,57].

During study design, comprehensive documentation of contextual factors is essential to support interpretation and assessment of transferability [48,49]. This includes detailed description of intervention delivery context, such as provider characteristics, organizational settings, and fidelity monitoring procedures. Participant characteristics should be reported comprehensively, including not only demographics but also community-level contextual characteristics, clinical features and employment histories that may moderate intervention effects. Local labor market conditions, welfare system structures, and healthcare context should also be described, with explicit discussion of how these contextual factors may affect the transferability of findings to other settings.

(ii) Transparent costing

As detailed in Section 3.3.2, transparent costing begins with clear identification of relevant cost items based on the analytical perspective adopted. All resources consumed in intervention delivery should be identified, including personnel time, training, supervision, materials, space, and overheads. The scope of costs captured beyond the intervention itself depends on perspective: health system perspectives include health service costs; welfare perspectives include benefit payments and employment service costs; societal perspectives encompass all costs regardless of who bears them, including participant out-of-pocket costs and caregivers' time. Omission of important cost categories should be acknowledged as a limitation rather than left unreported [48,49,58].

As highlighted in Section 3.3.3, the measurement and valuation of costs require separation of physical quantities from unit costs, rather than reporting only aggregated totals, enabling readers to verify calculations and adapt findings to different unit-cost contexts. For novel interventions, micro-costing approaches that measure resource use in detail are preferred. For established programs, gross-costing methods using standardized unit costs are acceptable, provided that unit costs are well-documented and appropriate to the context. When using unit costs from other jurisdictions, explicit justification is required. All costs should be adjusted for inflation using appropriate indices—consumer price indices for general costs or healthcare-specific indices for medical costs—and the reference year for all costs must be clearly reported. For international comparisons, purchasing power parity (PPP) conversion is preferred over simple exchange rates to account for differences in price levels across countries. Given uncertainty in cost estimates, sensitivity analyses on key unit-cost assumptions should be conducted to assess the robustness of findings [48,49,58].

Data sources for all cost components must be specified, along with any assumptions and allocation methods used for shared resources, such as capital costs and overheads. When available, high-quality administrative data should be prioritized because of their objectivity and completeness. When relying on self-report for cost data, validation against external sources should be undertaken when feasible [59].

Comprehensive reporting requires itemized presentation of cost components with quantities and unit costs in tables or Supplementary Materials under appropriate category headings (e.g., intervention costs, healthcare costs, welfare costs), rather than relegating this detail to cited references. This disaggregation enables readers to understand cost drivers and to recalculate findings from alternative perspectives or using different unit-cost assumptions, thereby enhancing the utility of cost estimates for decision-making in different contexts [48,49].

(iii) Selecting and measuring outcomes

As discussed in Section 3.2, outcome selection should be guided by the intervention's theory of change, identifying which outcomes should plausibly be affected by the intervention and over what time frame. Selection should also reflect the values and priorities of key stakeholders, such as participants, funders, and service providers, who may weigh different outcome domains differently. Proliferation of outcomes without theoretical justification should be avoided, as this increases respondent burden, multiplicity problems, and interpretative challenges. Primary and secondary outcomes should be pre-specified in study protocols, with any exploratory outcomes clearly identified as such in reporting [51].

As detailed in Sections 3.2.1 and 3.2.2, the measurement strategy for each outcome domain should match the nature of the outcome—such as whether it is objective or subjective, short- or long-term—and the available data sources. For vocational outcomes, administrative records are preferred where available, for their objectivity and completeness. However, depending on study objectives, these may need to be supplemented with self-report or qualitative measures to capture person-centred subjective dimensions—such as job satisfaction, perceived job match, and workplace experiences—that administrative data cannot capture [18]. Clinical and functional outcomes should be assessed using validated instruments with established psychometric properties, with complete specification of instrument version and scoring method. Health and social service utilisation is most reliably measured through administrative data linkage when accessible; when self-report is necessary, validated instruments such as the Client Service Receipt Inventory can be used [59], with recall periods minimized to reduce bias. Quality-of-life measurement may use generic instruments such as the EQ-5D [60] or SF-36 [61] to enable comparison across interventions.

As observed in Section 3.4.3, ensuring measurement quality requires attention to assessor training, blinding, reliability, missing data, and timing. Whether assessors are masked to treatment allocation should be specified; when blinding is not feasible, this should be acknowledged as a potential source of bias. Inter-rater reliability should be established. The extent and patterns of missing data should be reported transparently, with clear description of any imputation methods used or sensitivity analyses conducted to assess the impact of missing data on findings. Measurement timing should be selected to align with the expected onset of intervention effects and should be consistent across participants to enable meaningful comparison [51].

4.3. Balancing Methodological Rigor with Pragmatic Constraints

While the guidance outlined above represents methodologically sound approaches to evaluating supported employment programs, we acknowledge that real-world evaluations are inevitably shaped by resource constraints, data access limitations, and time pressures

that can make ideal designs infeasible. The pursuit of methodological perfection should not become a barrier to conducting evaluations that, while imperfect, can nonetheless generate useful evidence to inform decision-making. When faced with constraints, evaluators must make strategic choices that balance scientific rigor with feasibility while maximizing the decision relevance of findings. For example, evaluators may prioritize the outcome domains most relevant to the decision-makers, recognizing that it is preferable to collect a smaller set of high-quality data than to attempt comprehensive measurement that results in poor data quality across domains. Furthermore, the fixed duration of many research grants and the misalignment between funding timelines and appropriate follow-up periods may limit the feasibility of longer-term outcome assessment. In such cases, researchers should clearly articulate to funders the methodological importance of adequate follow-up and justify proposed follow-up periods explicitly within funding applications.

Regardless of the specific constraints encountered, transparent documentation is essential. Evaluators should explain clearly what was not possible and why, whether due to funding limitations, data access barriers, institutional constraints, or practical considerations. The implications of these constraints for the interpretation and generalizability of findings should be discussed explicitly rather than left for readers to infer. Where feasible, evaluators should suggest how future studies might overcome encountered limitations, contributing to the cumulative improvement of the evidence base. Critically, findings should be framed in light of the constraints under which they were generated rather than presented as universal truths. For example, a study concluding that an intervention is cost-effective based on 12-month follow-up should clearly state that this conclusion applies to the short-term horizon observed and that longer-term cost-effectiveness remains uncertain. Such transparent acknowledgment of constraints enhances credibility by demonstrating scientific integrity and enables appropriate interpretation by decision-makers operating under different conditions.

4.4. Evidence Gaps

This review identified several evidence gaps. First, no studies examined whether employment earnings lifted participants above poverty thresholds or reduced dependence on means-tested benefits, despite these being explicit goals of many welfare-to-work policies. Addressing this gap requires collecting household composition data alongside employment earnings, enabling comparison of total household income to contextually appropriate poverty thresholds (such as official poverty lines, living wage standards, or self-sufficiency benchmarks). Benefit dependency can be assessed through administrative data linkage to welfare and benefits systems where accessible, or through validated self-report questionnaires that capture receipt of disability benefits, means-tested income support, and in-kind assistance.

Second, wider psychosocial and well-being outcomes—such as community integration beyond health and employment—received minimal attention. This leaves important questions unexamined, including whether employment supports personal recovery or comes at the expense of other valued life domains. Outcome selection in the included studies reflects an implicit hierarchy in which economic productivity is prioritized over social participation and personal recovery, potentially misaligning with the holistic recovery goals articulated in contemporary mental health policy.

Third, few studies incorporated outcomes reflecting longer-term career trajectories or developmental goals. Career progression, skill development, educational advancement, and occupational mobility were rarely assessed, suggesting an implicit assumption that initial job placement—rather than career development—constitutes the primary indicator of success. This short-term focus may be pragmatically driven by funding constraints

and follow-up duration, yet it obscures important questions about whether supported employment facilitates sustainable career trajectories or predominantly produces entry-level, high-turnover placements.

Finally, the reviewed studies rarely assessed outcomes from explicitly defined stakeholder perspectives. While employment rates matter most to policymakers concerned with labor force participation and benefit expenditure, participants may prioritize job satisfaction, skill development, and alignment with personal goals. Family members may value reduced caregiving burden and enhanced family functioning, while employers may focus on retention, productivity, and workplace integration. The absence of stakeholder-specific outcome frameworks means that evaluation findings may incompletely address the information needs of key decision-makers and may misrepresent intervention success from different vantage points.

4.5. Review Limitations

Several limitations should be acknowledged. First, our inclusion criteria restricted the review to randomized controlled trials published between 2009 and 2024, identified through a parent rapid scoping review that searched only MEDLINE and CINAHL databases. This approach may have excluded relevant studies indexed elsewhere, as well as quasi experimental and observational studies examining supported employment in naturalistic settings, potentially limiting insights into methodological practices in real-world evaluations. Additionally, the inclusion of secondary analyses and trial-based economic evaluations alongside primary RCTs introduces some variability in study objectives and analytical scope; however, as this review evaluates methodological practices rather than intervention effects, this was considered appropriate and is unlikely to have materially affected the conclusions.

Second, this review did not conduct a formal quality assessment using comprehensive critical appraisal tools such as the Drummond checklist for economic evaluations or the Cochrane Risk of Bias tool for trials. While specific aspects of cost and outcome measurement quality were assessed using modified CHEC, broader study quality domains—such as allocation concealment, adjustment for baseline imbalances, selective outcome reporting, and appropriateness of statistical analysis—were not evaluated. This review therefore complements rather than replaces formal quality assessments.

Third, data extraction focused on information reported in published manuscripts and Supplementary Materials. When methodological details were referenced to earlier publications or protocols, these were not retrieved and reviewed. This approach may have missed details reported elsewhere, though it reflects the information typically accessible to readers of published evaluations.

Despite these limitations, this methodological review provides a comprehensive foundation for understanding current practices in costing and outcome measurement in supported employment evaluations, identifies key gaps in reporting transparency, and offers practical guidance to strengthen the decision relevance of future evaluations.

5. Conclusions

Supported employment programs for people with severe mental illness operate at the intersection of health, welfare, and labor market systems. Evaluations of these programs necessarily reflect diverse perspectives, priorities, and contexts, resulting in methodological heterogeneity that is often appropriate rather than problematic. However, this diversity creates challenges when methodological choices remain opaque, limiting decision-makers' ability to interpret findings, assess transferability, and synthesize evidence.

By making methodological choices explicit, documenting them transparently, and discussing their implications openly, researchers can produce evaluations that genuinely

inform policy and service planning decisions, even under real-world constraints. This review provides a foundation for such efforts, offering practical guidance to support the design, implementation, and reporting of evaluations that generate robust, decision-relevant evidence on the value of supported employment for people with SMI.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/soc16060180/s1>, Supplementary Material S1: Complete data extraction on study characteristics, outcome measurement, economic analyses, and quality assessment.

Author Contributions: Conceptualization, K.-Y.J.C. and J.S.-M.; methodology, K.-Y.J.C.; data extraction, A.K., Y.Y. and Y.L.; quality assessment, A.K., Y.Y. and Y.L.; validation, K.-Y.J.C.; formal analysis, K.-Y.J.C.; writing—original draft preparation, K.-Y.J.C.; writing—review and editing, J.S.-M.; funding acquisition, J.S.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Australian Research Council through an Industry Laureate Fellowship, grant number IL230100154.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. This methodological review did not generate new data. All information synthesized in this study was derived from previously published sources, which are cited in the manuscript.

Acknowledgments: Kuo-Yi Jade Chang acknowledges the support of the NHMRC-funded Centre of Research Excellence in Achieving Health Equity for All People with Disabilities (Grant #2035278).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CBA	Cost-benefit analysis
CBT	Cognitive behavioral therapy
CCA	Cost-consequence analysis
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CHEC	Consensus Health Economic Criteria
CI	Confidence interval
CR	Cognitive remediation
CSRI	Client Service Receipt Inventory
CUA	Cost-utility analysis
ICER	Incremental cost-effectiveness ratio
IPS	Individual placement and support
PPP	Purchasing power parity
PTSD	Post-traumatic stress disorder
QALY	Quality-adjusted life year
RCT	Randomized controlled trial
SMI	Severe mental illness
SROI	Social return on investment
TVR	Transitional vocational rehabilitation
VA	Veterans Administration

References

- Baxter, S.; Cullingworth, J.; Whitworth, A.; Runswick-Cole, K.; Clowes, M. Understanding interventions and outcomes in supported employment and individual placement support: A qualitative evidence synthesis. *Disabil. Health J.* **2024**, *17*, 101579. [\[CrossRef\]](#)
- Drake, R.E.; Wallach, M.A. Employment is a critical mental health intervention. *Epidemiol. Psychiatr. Sci.* **2020**, *29*, e178. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gold, P.B.; Macias, C.; Rodican, C.F. Does Competitive Work Improve Quality of Life for Adults with Severe Mental Illness? Evidence from a Randomized Trial of Supported Employment. *J. Behav. Health Serv. Res.* **2016**, *43*, 155–171. [\[CrossRef\]](#)
- Modini, M.; Tan, L.; Brinchmann, B.; Wang, M.J.; Killackey, E.; Glozier, N.; Mykletun, A.; Harvey, S.B. Supported employment for people with severe mental illness: Systematic review and meta-analysis of the international evidence. *Br. J. Psychiatry* **2016**, *209*, 14–22. [\[CrossRef\]](#)
- Hoffmann, H.; Jäckel, D.; Glauser, S.; Mueser, K.T.; Kupper, Z. Long-Term Effectiveness of Supported Employment: 5-Year Follow-Up of a Randomized Controlled Trial. *Am. J. Psychiatry* **2014**, *171*, 1183–1190. [\[CrossRef\]](#)
- McGurk, S.R.; Xie, H.; Bond, G.R.; Mueser, K.T. Impact of cognitive remediation on the prediction of employment outcomes in severe mental illness. *Schizophr. Res.* **2022**, *241*, 149–155. [\[CrossRef\]](#)
- Mervis, J.E.; Lysaker, P.H.; Fiszdon, J.M.; Bell, M.D.; Chue, A.E.; Pauls, C.; Bisoglio, J.; Choi, J. Addressing defeatist beliefs in work rehabilitation. *J. Ment. Health* **2016**, *25*, 366–371. [\[CrossRef\]](#)
- Weld-Blundell, I.; Shields, M.; Devine, A.; Dickinson, H.; Kavanagh, A.; Marck, C. Vocational Interventions to Improve Employment Participation of People with Psychosocial Disability, Autism and/or Intellectual Disability: A Systematic Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 12083. [\[CrossRef\]](#) [\[PubMed\]](#)
- Crestois, N.; Kempton, M.J.; Twumasi, R. A systematic review and meta-analysis of employer discrimination towards people living with psychosis. *Schizophr. Res.* **2025**, *278*, 35–46. [\[CrossRef\]](#) [\[PubMed\]](#)
- Koning, P.; Muller, P.; Prudon, R. Do disability benefits hinder work resumption after recovery? *J. Health Econ.* **2022**, *82*, 102593. [\[CrossRef\]](#)
- Wallstroem, I.G.; Pedersen, P.; Christensen, T.N.; Hellström, L.; Bojesen, A.B.; Stenager, E.; White, S.; Mueser, K.T.; Bejerholm, U.; van Busschbach, J.T.; et al. A Systematic Review of Individual Placement and Support, Employment, and Personal and Clinical Recovery. *Psychiatr. Serv.* **2021**, *72*, 1040–1047. [\[CrossRef\]](#) [\[PubMed\]](#)
- Au, D.W.H.; Tsang, H.W.H.; So, W.W.Y.; Bell, M.D.; Cheung, V.; Yiu, M.G.C.; Tam, K.; Lee, G.T.-H. Effects of integrated supported employment plus cognitive remediation training for people with schizophrenia and schizoaffective disorders. *Schizophr. Res.* **2015**, *166*, 297–303. [\[CrossRef\]](#)
- Bell, M.D.; Choi, K.H.; Dyer, C.; Wexler, B.E. Benefits of cognitive remediation and supported employment for schizophrenia patients with poor community functioning. *Psychiatr. Serv.* **2014**, *65*, 469–475. [\[CrossRef\]](#) [\[PubMed\]](#)
- Christensen, T.N.; Wallström, I.G.; Stenager, E.; Bojesen, A.B.; Gluud, C.; Nordentoft, M.; Eplov, L.F. Effects of Individual Placement and Support Supplemented with Cognitive Remediation and Work-Focused Social Skills Training for People with Severe Mental Illness: A Randomized Clinical Trial. *JAMA Psychiatry* **2019**, *76*, 1232–1240. [\[CrossRef\]](#) [\[PubMed\]](#)
- Davis, L.L.; Kyriakides, T.C.; Suris, A.M.; Ottomanelli, L.A.; Mueller, L.; Parker, P.E.; Resnick, S.G.; Toscano, R.; Scrymgeour, A.A.; Drake, R.E. Effect of Evidence-Based Supported Employment vs Transitional Work on Achieving Steady Work Among Veterans with Posttraumatic Stress Disorder: A Randomized Clinical Trial. *JAMA Psychiatry* **2018**, *75*, 316–324. [\[CrossRef\]](#)
- Park, A.L.; Rinaldi, M.; Brinchmann, B.; Killackey, E.; Aars, N.A.P.; Mykletun, A.; McDaid, D. Economic analyses of supported employment programmes for people with mental health conditions: A systematic review. *Eur. Psychiatry* **2022**, *65*, e51. [\[CrossRef\]](#)
- Hellström, L.; Kruse, M.; Christensen, T.N.; Trap Wolf, R.; Eplov, L.F. Cost-effectiveness analysis of a supported employment intervention for people with mood and anxiety disorders in Denmark—The IPS-MA intervention. *Nord. J. Psychiatry* **2021**, *75*, 389–396. [\[CrossRef\]](#)
- Mateen, B.A.; Doogan, C.; Hayward, K.; Hourihan, S.; Hurford, J.; Playford, E.D. Systematic Review of Health-Related Work Outcome Measures and Quality Criteria-Based Evaluations of Their Psychometric Properties. *Arch. Phys. Med. Rehabil.* **2017**, *98*, 534–560. [\[CrossRef\]](#)
- Patmisari, E.; Huang, Y.; Orr, M.; Govindasamy, S.; Hielscher, E.; McLaren, H. Supported employment interventions with people who have severe mental illness: Systematic mixed-methods umbrella review. *PLoS ONE* **2024**, *19*, e0304527. [\[CrossRef\]](#)
- Chang, K.-Y.J.; Yen, I.; Bobo, F.; Hollier, J.; Smith-Merry, J. Implementing Community-Based Psychosocial Interventions for Adults with Severe Mental Illness in High-Income Countries: A Rapid Scoping Review. *Community Ment. Health J.* **2026**. [\[CrossRef\]](#)
- Evers, S.; Goossens, M.; de Vet, H.; van Tulder, M.; Ament, A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int. J. Technol. Assess. Health Care* **2005**, *21*, 240–245. [\[CrossRef\]](#)
- Areberg, C.; Bejerholm, U. The effect of IPS on participants' engagement, quality of life, empowerment, and motivation: A randomized controlled trial. *Scand. J. Occup. Ther.* **2013**, *20*, 420–428. [\[CrossRef\]](#)

23. Bejerholm, U.; Areberg, C.; Hofgren, C.; Sandlund, M.; Rinaldi, M. Individual Placement and Support in Sweden—A randomized controlled trial. *Nord. J. Psychiatry* **2015**, *69*, 57–66. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Bejerholm, U.; Larsson, M.E.; Johanson, S. Supported employment adapted for people with affective disorders—A randomized controlled trial. *J. Affect. Disord.* **2017**, *207*, 212–220. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Bell, M.D.; Laws, H.; Pittman, B.; Johannesen, J.K. Comparison of focused cognitive training and portable “brain-games” on functional outcomes for vocational rehabilitation participants. *Sci. Rep.* **2018**, *8*, 1779. [\[CrossRef\]](#)
26. Bond, G.R.; Campbell, K.; Becker, D.R. A Test of the Occupational Matching Hypothesis for Rehabilitation Clients with Severe Mental Illness. *J. Occup. Rehabil.* **2013**, *23*, 261–269. [\[CrossRef\]](#)
27. Bond, G.R.; Kim, S.J.; Becker, D.R.; Swanson, S.J.; Drake, R.E.; Krzos, I.M.; Fraser, V.V.; O’Neill, S.; Frounfelker, R.L. A Controlled Trial of Supported Employment for People with Severe Mental Illness and Justice Involvement. *Psychiatr. Serv.* **2015**, *66*, 1027–1034. [\[CrossRef\]](#)
28. Bond, G.R.; Drake, R.E.; Campbell, K. Effectiveness of individual placement and support supported employment for young adults. *Early Interv. Psychiatry* **2016**, *10*, 300–307. [\[CrossRef\]](#)
29. Burns, T.; Catty, J.; White, S.; Becker, T.; Koletsis, M.; Fioritti, A.; Rössler, W.; Tomov, T.; van Busschbach, J.; Wiersma, D.; et al. The Impact of Supported Employment and Working on Clinical and Social Functioning: Results of an International Study of Individual Placement and Support. *Schizophr. Bull.* **2009**, *35*, 949–958. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Davis, L.L.; Leon, A.C.; Toscano, R.; Drebing, C.E.; Ward, L.C.; Parker, P.E.; Kashner, T.M.; Drake, R.E. A Randomized Controlled Trial of Supported Employment Among Veterans with Posttraumatic Stress Disorder. *Psychiatr. Serv.* **2012**, *63*, 464–470. [\[CrossRef\]](#)
31. Davis, L.W.; Lysaker, P.H.; Kristeller, J.L.; Salyers, M.P.; Kovach, A.C.; Woller, S. Effect of Mindfulness on Vocational Rehabilitation Outcomes in Stable Phase Schizophrenia. *Psychol. Serv.* **2015**, *12*, 303–312. [\[CrossRef\]](#)
32. DeTore, N.R.; Balogun-Mwangi, O.; Mueser, K.T.; McGurk, S.R. Comparison of Black and White participants with severe mental illness in response to cognitive remediation as an augmentation of vocational rehabilitation. *Schizophr. Res.* **2023**, *253*, 60–67. [\[CrossRef\]](#)
33. Drake, R.E.; Frey, W.; Bond, G.R.; Goldman, H.H.; Salkever, D.; Miller, A.; Moore, T.A.; Riley, J.; Karakus, M.; Milfort, R. Assisting Social Security Disability Insurance Beneficiaries with Schizophrenia, Bipolar Disorder, or Major Depression in Returning to Work. *Am. J. Psychiatry* **2013**, *170*, 1433–1441. [\[CrossRef\]](#)
34. Heslin, M.; Howard, L.; Leese, M.; McCrone, P.; Rice, C.; Jarrett, M.; Spokes, T.; Huxley, P.; Thornicroft, G. Randomized controlled trial of supported employment in England: 2 year follow-up of the Supported Work and Needs (SWAN) study. *World Psychiatry* **2011**, *10*, 132–137. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Holmås, T.H.; Monstad, K.; Reme, S.E. Regular employment for people with mental illness -- An evaluation of the individual placement and support programme. *Soc. Sci. Med.* **2021**, *270*, 113691. [\[CrossRef\]](#)
36. Howard, L.M.; Heslin, M.; Leese, M.; McCrone, P.; Rice, C.; Jarrett, M.; Spokes, T.; Huxley, P.; Thornicroft, G. Supported employment: Randomised controlled trial. *Br. J. Psychiatry* **2010**, *196*, 404–411. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Humm, L.B.; Olsen, D.; Be, M.; Fleming, M.; Smith, M. Simulated job interview improves skills for adults with serious mental illnesses. *Stud. Health Technol. Inform.* **2014**, *199*, 50–54. [\[PubMed\]](#)
38. Juurlink, T.T.; Lamers, F.; van Marle, H.J.F.; Michon, H.; van Busschbach, J.T.; Beekman, A.T.F.; Anema, J.R. Employment in Personality Disorders and the Effectiveness of Individual Placement and Support: Outcomes from a Secondary Data Analysis. *J. Occup. Rehabil.* **2020**, *30*, 255–262. [\[CrossRef\]](#)
39. Kern, R.S.; Zarate, R.; Glynn, S.M.; Turner, L.R.; Smith, K.M.; Mitchell, S.S.; Sugar, C.A.; Bell, M.D.; Liberman, R.P.; Kopelowicz, A.; et al. Improving Work Outcome in Supported Employment for Serious Mental Illness: Results from 2 Independent Studies of Errorless Learning. *Schizophr. Bull.* **2018**, *44*, 38–45. [\[CrossRef\]](#)
40. Knapp, M.; Patel, A.; Curran, C.; Latimer, E.; Catty, J.; Becker, T.; Drake, R.E.; Fioritti, A.; Kilian, R.; Lauber, C.; et al. Supported employment: Cost-effectiveness across six European sites. *World Psychiatry* **2013**, *12*, 60–68. [\[CrossRef\]](#)
41. Mervis, J.E.; Fiszdon, J.M.; Lysaker, P.H.; Nienow, T.M.; Mathews, L.; Wardwell, P.; Petrik, T.; Thime, W.; Choi, J. Effects of the Indianapolis Vocational Intervention Program (IVIP) on defeatist beliefs, work motivation, and work outcomes in serious mental illness. *Schizophr. Res.* **2017**, *182*, 129–134. [\[CrossRef\]](#)
42. Michon, H.; van Busschbach, J.T.; Stant, A.D.; van Vugt, M.D.; van Weeghel, J.; Kroon, H. Effectiveness of Individual Placement and Support for People with Severe Mental Illness in the Netherlands: A 30-Month Randomized Controlled Trial. *Psychiatr. Rehabil. J.* **2014**, *37*, 129–136. [\[CrossRef\]](#)
43. Mueller, L.; Wolfe, W.R.; Neylan, T.C.; McCaslin, S.E.; Yehuda, R.; Flory, J.D.; Kyriakides, T.C.; Toscano, R.; Davis, L.L. Positive Impact of IPS Supported Employment on PTSD-Related Occupational-Psychosocial Functional Outcomes: Results from a VA Randomized-Controlled Trial. *Psychiatr. Rehabil. J.* **2019**, *42*, 246–256. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Mueser, K.T.; Campbell, K.; Drake, R.E. The Effectiveness of Supported Employment in People with Dual Disorders. *J. Dual Diagn.* **2011**, *7*, 90–102. [\[CrossRef\]](#)

45. Mueser, K.T.; Bond, G.R.; Essock, S.M.; Clark, R.E.; Carpenter-Song, E.; Drake, R.E.; Wolfe, R. The Effects of Supported Employment in Latino Consumers with Severe Mental Illness. *Psychiatr. Rehabil. J.* **2014**, *37*, 113–122. [[CrossRef](#)] [[PubMed](#)]
46. Penk, W.; Drebing, C.E.; Rosenheck, R.A.; Krebs, C.; Van Ormer, A.; Mueller, L. Veterans Health Administration Transitional Work Experience vs. Job Placement in Veterans with Co-Morbid Substance Use and Non-Psychotic Psychiatric Disorders. *Psychiatr. Rehabil. J.* **2010**, *33*, 297–307. [[CrossRef](#)] [[PubMed](#)]
47. Yamaguchi, S.; Sato, S.; Horio, N.; Yoshida, K.; Shimodaira, M.; Taneda, A.; Ikebuchi, E.; Nishio, M.; Ito, J. Cost-effectiveness of cognitive remediation and supported employment for people with mental illness: A randomized controlled trial. *Psychol. Med.* **2017**, *47*, 53–65. [[CrossRef](#)]
48. Drummond, M.F.; Sculpher, M.J.; Claxton, K.; Stoddart, G.L.; Torrance, G.W. *Methods for the Economic Evaluation of Health Care Programmes*, 4th ed.; Oxford University Press: Oxford, UK, 2015.
49. Husereau, D.; Drummond, M.; Augustovski, F.; de Bekker-Grob, E.; Briggs, A.H.; Carswell, C.; Caulley, L.; Chaiyakunapruk, N.; Greenberg, D.; Loder, E.; et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: Updated reporting guidance for health economic evaluations. *BMJ* **2022**, *376*, e067975. [[CrossRef](#)]
50. Curtis, L. *Unit Costs of Health and Social Care*; Personal Social Services Research Unit: Canterbury, UK, 2007.
51. Schulz, K.F.; Altman, D.G.; Moher, D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ* **2010**, *340*, 698–702. [[CrossRef](#)]
52. Debruyne, P.R.; Knott, V.E.; Pattison, N.A. A call for rigorous research and transparent study reporting. *Eur. J. Cancer Care* **2013**, *22*, 421–422. [[CrossRef](#)]
53. Griffin, S.; Walker, S.; Sculpher, M.; Asaria, M. An Analytical Framework for Economic Evaluation of Interventions with Effects on Multiple Outcomes, Costs Falling on Different Budgets, and Involving More than One Decision Maker. *Value Health* **2018**, *21*, S10. [[CrossRef](#)]
54. Sittimart, M.; Rattavipapong, W.; Mirelman, A.J.; Hung, T.M.; Dabak, S.; Downey, L.E.; Jit, M.; Teerawattananon, Y.; Turner, H.C. An overview of the perspectives used in health economic evaluations. *Cost Eff. Resour. Alloc.* **2024**, *22*, 41. [[CrossRef](#)]
55. McGurk, S.R.; Mueser, K.T. Sustaining the Long-Term Effects of Supported Employment for Persons with Psychiatric Disabilities. *Am. J. Psychiatry* **2016**, *173*, 953–955. [[CrossRef](#)]
56. Christensen, T.N.; Wallström, I.G.; Stenager, E.; Hellström, L.; Bojesen, A.B.; Nordentoft, M.; Eplov, L.F. 30-Month Follow-Up of Individual Placement and Support (IPS) and Cognitive Remediation for People with Severe Mental Illness: Results from a Randomized Clinical Trial. *Psychiatry J.* **2023**, *2023*, 2789891. [[CrossRef](#)]
57. Chen, W.; Howell, M.; Cass, A.; Gorham, G.; Howard, K. Understanding modelled economic evaluations: A reader's guide for clinicians. *Med. J. Aust.* **2024**, *221*, 302–307. [[CrossRef](#)] [[PubMed](#)]
58. Sanders, G.D.; Neumann, P.J.; Basu, A.; Brock, D.W.; Feeny, D.; Krahm, M.; Kuntz, K.M.; Meltzer, D.O.; Owens, D.K.; Prosser, L.A.; et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA* **2016**, *316*, 1093–1103. [[CrossRef](#)]
59. Beecham, J.; Knapp, M. Costing psychiatric interventions. In *Measuring Mental Health Needs*, 2nd ed.; Thornicroft, G., Ed.; Gaskell: London, UK, 2001; pp. 200–224.
60. EuroQol Group. EuroQol—A new facility for the measurement of health-related quality of life. *Health Policy* **1990**, *16*, 199–208. [[CrossRef](#)] [[PubMed](#)]
61. Ware, J.E.; Sherbourne, C.D. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Med. Care* **1992**, *30*, 473–483. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.