systems

MDPI

Review

Assessing Impact in Europe: A Systematic Review of Evaluation Methodologies in Homeless Interventions

Juan Manuel Rodilla [1], Gloria Puchol [1] and Mercedes Botija [2,*]

[1] Sant Joan de Déu València, Calle Luz Casanova 8, 46009 Valencia, Spain; manuel.rodilla@sjd.es (J.M.R.); gloria.puchol@sjd.es (G.P.)
[2] Department of Social Work and Social Services, University of Valencia, Avenida Tarongers, 46022 Valencia, Spain
* Correspondence: mercedes.botija@uv.es

Abstract: Homelessness presents a complex societal challenge, necessitating evidence-based interventions. This paper conducts a PRISMA systematic review of impact evaluation methodologies in homeless interventions, examining the existence of standardized methodologies, and the role of theoretical frameworks, consensus on evaluation designs, and reliable outcome variables. Drawing from diverse studies, the research comprehensively analyzes impact evaluations with the goal of yielding valuable insights for practitioners and policymakers responding to the challenges and dynamics of the European context. The findings reveal a lack of standardized methodologies validated by regulatory agencies, particularly within Europe. Theoretical foundations guiding the evaluations vary widely, highlighting the need for a context-sensitive framework that considers the complexities of homelessness and socio-political factors across welfare states. While randomized controlled trials offer rigor, they are underutilized in Europe. The review advocates a mixed-methods approach for comprehensive insights to capture the multifaceted nature of homelessness interventions. Furthermore, the identification of suitable outcome variables emerges as a challenge, with inconsistent definitions hindering cross-study comparisons. The analysis underscores the significance of adopting standardized outcome variables, such as the ETHOS definition, to facilitate robust impact assessments. This review emphasizes the need for methodological refinement and collaboration, enabling comparability between programs and the generation of reliable evidence. Advocating standardized methodologies, robust frameworks, and comprehensive designs, it guides future research, evidence-based policymaking, and effective homeless interventions.

Keywords: homelessness; impact evaluation; systematic review; Europe

Citation: Rodilla, J.M.; Puchol, G.; Botija, M. Assessing Impact in Europe: A Systematic Review of Evaluation Methodologies in Homeless Interventions. *Systems* **2023**, *11*, 541. https://doi.org/10.3390/systems11110541

Academic Editors: Wayne Wakeland and Robert Y. Cavana

Received: 29 August 2023
Revised: 1 November 2023
Accepted: 3 November 2023
Published: 6 November 2023

1. Introduction

Addressing homelessness is a multifaceted societal challenge that demands evidence-based interventions. While significant research and evaluation has been conducted on responses to homelessness over the past few decades, the majority of this body of work has originated in the United States and Canada. Recognizing the need to assess the current disparate landscape of evaluations concerning homelessness interventions, this paper undertakes a systematic review of impact evaluation methodologies. To contextualize this evaluation, we begin by introducing the European Typology on Homelessness and Housing Exclusion (ETHOS) as a benchmarking tool. Our systematic review, drawing from a diverse range of studies, comprehensively analyzes impact evaluations with the goal of yielding valuable insights for practitioners and policymakers responding to the challenges and dynamics of the European context. This examination focuses on key parameters essential for designing a robust evaluation: methodology, theoretical framework, evaluation design, and outcome variables. Our objective is to identify essential findings that could promote consistency in the development of evaluations, enable more effective comparisons between

different interventions, and ultimately contribute to the ongoing effort to address and alleviate homelessness.

1.1. Homelessness and ETHOS

Homelessness is a social issue that has traditionally been characterized by difficulty in its definition and conceptualization [1]. A robust definition of homelessness is the necessary foundation for consistent data-acquisition mechanisms enabling informed policymaking [2]. For this purpose, the European Federation of National Organisations Working with the Homeless (FEANTSA) established the European Typology on Homelessness and Housing Exclusion (ETHOS) in 2005. ETHOS was created to improve homelessness understanding and measurement, and to facilitate the collation of statistics on homelessness in a more consistent manner across Europe [3].

ETHOS (Figure 1) identifies three domains as constituting a home: the physical domain, the social domain, and the legal domain. Living situations deficient in one or more of the domains are considered to constitute homelessness and housing exclusion [2]. Depending on the lack of the mentioned domains, four conceptual categories of living situation are identified: rooflessness, houselessness, insecure housing, and inadequate housing. The four conceptual categories are later developed into thirteen operational categories for policy purposes, such as mapping the problem using point-in-time counts or for monitoring and evaluating policies [4].

Conceptual Category		OPERATIONAL CATEGORY		LIVING SITUATION		GENERIC DEFINITION
	ROOFLESS	1	People Living Rough	1.1	Public space or external space	Living in the streets or public spaces, without a shelter that can be defined as living quarters
		2	People in emergency accommodation	2.1	Night shelter	People with no usual place of residence who make use of overnight shelter, low threshold shelter
	HOUSELESS	3	People in accommodation for the homeless	3.1	Homeless hostel	
				3.2	Temporary accommodation	Where the period of stay is intended to be short term
				3.3	Transitional supported accommodation	
		4	People in Women's Shelter	4.1	Women's shelter accommodation	Women accommodated to experience of domestic violence and where the period of stay is intended to be short term
		5	People in accommodation for immigrants	5.1	Temporary accommodation/reception centres	Immigrants in reception or short term accommodation due to their immigrant status
				5.2	Migrant workers accommodation	
		6	People due to be released from institutions	6.1	Penal institutions	No housing available prior to release
				6.2	Medical institutions	Stay longer than needed due to lack of housing
				6.3	Children's institutions/homes	No housing identified (e.g. by 18th birthday)
	INSECURE	7	People receiving longer-term support (due to homelessness)	7.1	Residential care for older homeless people	Long stay accommodation with care for formerly homeless people (normally more than one year)
				7.2	Supported accommodation for formerly homeless people	
		8	People living in insecure accommodation	8.1	Temporarily with family/friends	Living in conventional housing but not the usual place of residence due to lack of housing
				8.2	No legal (sub)tenancy	Occupation of dwelling with no legal tenancy illegal occupation of a dwelling
				8.3	Illegal occupation of land	Occupation of land with no legal rights
		9	People living under threat of eviction	9.1	Legal orders enforced (rented)	Where orders for eviction are operative
				9.2	Re-possession orders (owned)	Where mortagee has legal order to re-possess
		10	People living under threat of violence	10.1	Police recorded incidents	Where police action is taken to ensure place of safety for victims of domestic violence
	INADEQUATE	11	People living in temporary/ non-conventional structures	11.1	Mobile homes	Not intended as place of usual residence
				11.2	Non-conventional building	Makeshift shelter, shack or shanty
				11.3	Temporary structure	Semi-permanent structure hut or cabin
		12	People living in unfit housing	12.1	Occupied dwellings unfit for habitation	Defined as unfit for habitation by national legislation or building regulations
		13	People living in extreme over-crowding	13.1	Highest national norm of overcrowding	Defined as exceeding national density standard for floor-space or uesable rooms

Figure 1. FEANTSA 2005 European Typology on Homelessness and Housing Exclusion [4].

1.2. Lack of Impact Evaluations in Services for People Experiencing Homelessness

Diverse authors consider the lack of robustness of the scientific evidence in the European context to be the main topic to address in the strategic planning of homelessness research in Europe in the coming years [5–8]. This lack of evidence can be grouped under three main needs: large-scale data collection, the generation of research in the specific context of different European welfare states, and the orientation towards longitudinal and impact studies.

In specific reference to effectiveness and impact studies, the European Commission report "Fighting Homelessness and Housing Exclusion in Europe" [8] highlights that, in the vast majority of European countries, there is a clear lack of rigor, monitoring, and evaluation of the effectiveness of services for people experiencing homelessness. The report emphasizes that only two countries, Denmark and Finland, present a clear monitoring and evaluation framework linked to the implementation of programs. This lack of evaluation mechanisms can also be found in the publication of studies on the effectiveness of interventions in the European context [9]. In the Spanish context, several authors [10–12] emphasize the need to establish impact-measurement structures, generate evidence, and analyze the effectiveness of homelessness response programs.

1.3. Slow-Pace Improvements in the European Context

The progressive introduction of Housing First (HF) programs has positively promoted the establishment of evaluation procedures in some European countries. But, although Pleace and Bretherton [13] emphasize that there is sufficient scientific evidence to demonstrate the effectiveness of the Housing First programs in Europe, in specific countries, such as Spain, just two randomized controlled trials can be found corresponding to the same Habitat program: one conducted in 2014 with a sample size of 84 individuals [14,15] and one conducted in 2020 with n = 242 individuals [16]. This scarcity of impact evaluations becomes a problem, as it does not enable the comparison of the effectiveness between different intervention programs. The absence of comparisons with other programs [9], the methodological doubts stated by some authors [17,18], and the publication of some evaluations in gray literature rather than indexed research journals raise doubts about the optimality of the HF evaluation as a standard model and to encourage the improvement of methodological rigor in evaluations associated with homelessness.

The initial draft of the Spanish 2nd National Comprehensive Strategy for Homeless People 2023–2030 [19] states, "Despite the limited information available on the effectiveness of interventions carried out with homeless people, it can be considered that the current response structure is not very effective and rarely achieves the goal of social or residential inclusion". Marbán and Rodriguez [11] emphasize that, among the main limitations and fundamental priorities for the development of comprehensive policies in Spain, there is no information on the effectiveness of support services for people experiencing homelessness. The authors specifically highlight that very few organizations publish data on the effectiveness of their interventions for people experiencing homelessness.

1.4. Benefits and Lessons Learned from Impact Evaluation

Monitoring and evaluation, specifically impact evaluation, have been recommended based on their capacity to identify the effectiveness of the integrated strategic responses to homelessness [8]. Additionally, monitoring and evaluation would facilitate the identification of subgroups within the population experiencing homeless [20]. This approach facilitates the measurement of the impact of the diversity of interventions focused on addressing specific needs and characteristics of the various subgroups identified [21]. Various authors identify clear subgroups, with specific needs, among people experiencing homelessness. Youth homelessness [22], people with high levels of stressful life events [21], or those with severe trauma [23] are some of these subgroups. This opportunity aligns with the critiques of the new orthodoxy for treating people experiencing homelessness as a homogeneous group [7] instead of identifying internally homogeneous subgroups

that would help to find explanations for the problems of each subgroup in relation to homelessness [20].

Some recommendations on the best practices in quantitative methods for homelessness studies have already been published [24]. Specifically, randomized controlled trials (RCTs) have gained prominence in their application in the social field, partly due to the 2019 Nobel Prize in Economics awarded to Banerjee, Duflo, and Kremer for their excellent work in the fight against poverty, converting complex theoretical questions into bounded and evaluable analyses through mathematical methodologies. Their work can be found in the book "Poor Economics" [25], and, for the subject at hand, in their toolbox on randomized evaluation [26]. Currently, accessible methodological guides for conducting RCTs are available from international organizations [27], and European research has begun to generate some lessons learned from the application of RCTs to homelessness research [28].

The UK Centre for Homelessness Impact [28] highlights two facets essential for a solid evidence base to identify practices and interventions offering better outcomes for people experiencing homelessness. First, it emphasizes the need to enhance evaluation capacity in the academic sector dedicated to homelessness research, facilitating the execution of rigorous evaluations tailored to the European context. Second, it suggests moving away from relying on knowledge generated in other regions, mainly the United States and Canada, or from other disciplines such as medicine, advocating the generation of key learnings in the local context using appropriate data-collection tools and conducting studies that can inform policymakers on the best way to address homelessness.

The following sections will delve into the methodology and results of the systematic review, exploring the state-of-the-art impact evaluation methodologies in homeless response programs. The discussion and conclusions will try to identify the main findings useful for policymakers and practitioners, with a specific focus on the implications for the European context.

2. Materials and Methods

The main goal of the systematic review was to determine if there is a standardized evaluation method to measure the impact of housing programs on homelessness, assessing the degree to which the findings are transferable or useful to the European context. Systematic integrative reviews have direct applicability to practice and policy [29,30]. To capture the existing empirical literature that focuses on the evaluation of homelessness housing programs, we used the Preferred Reporting Items for Systematic reviews and Meta-Analyses for Protocols 2015 (PRISMA-P statement) [31]. PRISMA-P Guidelines, is a standardized methodology consisting of a 17-item checklist intended to provide methodological rigor and facilitate the preparation and reporting of a robust protocol for systematic review. The checklist contains 17 numbered items (26 including sub-items) categorized into three main sections: administrative information, introduction, and methods [31]. The protocol ensures that the systematic review is carefully planned, ensuring coherent procedures by the review team and fostering accountability and transparency. Our strategy for conducting each of these procedures is described below.

2.1. Criteria for Study Selection

The studies were included in the review if they met the following general criterion: they describe, either directly or indirectly, evaluation methods for housing programs responding to homelessness. By directly, we mean specific studies that focus on the analysis of certain evaluation methods. By indirectly, we mean research that carries out outcomes evaluations. For the records added through database searches, the inclusion criteria that were considered for the eligibility of the studies were as follows: (a) original articles published in journals with a peer-review process, regardless of study design (including narrative and systematic reviews); (b) studies published in the last 23 years (2000–2023); and (c) studies published in journals indexed in the Journal citation Reports (JCR) or Scopus, using Scimago Journal and Country Rank quartiles and impact factors as references. For

the studies identified through other sources, the inclusion criteria considered were (a) articles published in specialized homelessness journals with a peer-review process and gray literature (only reports). For both search methods, either the title or the abstract of the study had to point out that the study is based on the population experiencing homelessness.

The exclusion criteria were (a) the population studied were not predominantly experiencing homeless; (b) the study cases in which the participants were younger than 18 years of age; (c) the housing situation was not considered as an outcome indicator; (d) studies that were not written in either English or Spanish; (e) conference abstracts; (f) studies focusing on a specific population experiencing homelessness, such as veterans; (g) studies that did not evaluate public programs that include housing; and (h) empirical studies that did not include quantitative techniques (only mixed techniques and quantitative techniques were considered).

2.2. Search Strategies and Information Sources

A search strategy was developed by two researchers from the NGO Sant Joan de Deu (Valencia) in collaboration with the University of Valencia, specifically with the social research group of intervention and innovation (GESinn). Any conflicts emerging at the title and abstract screening and full-text review were resolved through discussion and consensus between the principal investigator and the other members of the research team.

The systematic review was conducted with the guide of several articles of systematic reviews and using PRISMA-P Guidelines [31]. We searched for articles in English and Spanish in three databases: Scopus, WoS, and Dialnet. Three groups of key words were identified: the first focused on the studied population, combined with the second concept which was "intervention", and the third group related to the main topic of the review (evaluation: impact and longitudinal). These combinations of words were inserted into the search fields for title, abstract, and keywords using Boolean operators as shown in Table 1.

Table 1. Terms used in the Boolean search.

Combination of Words Using Boolean Operators
"homelessness" or "homeless" and "impact" and "intervention"
"sinhogarismo" or "sin hogar" and "impacto" and "intervención"
"homelessness" or "homeless" and "longitudinal" and "intervention"
"sinhogarismo" or "sin hogar" and "longitudinal" and "intervención"
"homelessness" or "homeless" and "evaluation" and "intervention"
"sinhogarismo" or "sin hogar" and "evaluación" and "intervención"

In addition to this search, a search was conducted in specific journals on homelessness research to identify any additional studies not captured using our primary search strategy. The reason for this additional research is that most of the main homelessness literature in Europe (both journal articles and gray literature) is published in homelessness journals that are not indexed in JCR or Scopus. Specifically, we searched in the European Observatory on Homelessness, the European Journal of Homelessness, Housing First Europe Hub, and other European platforms for homelessness with gray literature (only reports).

2.3. Data Extraction

The search was made in June 2022 and updated in January 2023. The selection of studies, the data extraction and coding were carried out by two of the authors working separately. Later, the information collected was compared by both researchers to reach an agreement in the event of them having different opinions. We developed a data-extraction table to capture the searched for information in studies included in our review. The information collected in the table consisted of the following: country, year, study type, method type, evaluation type, study design, intervention type, population, sample size, follow up time, housing outcome variable, description of the variable, and observations.

2.4. Quality Evaluation of the Studies

The advancement of academic progress in a particular field of knowledge is influenced by reputable journals [32], serving as essential conduits for disseminating rigorous research and fostering scholarly dialogue. To ensure the robustness of the encompassed research, most of the studies that were included in the analysis were published in journals cataloged in JCR or Scopus, both of which are recognized as the most dependable indicators of quality and are highly regarded by research evaluation organizations. The number of studies included in the analysis but that are not published in journals cataloged in JCR or Scopus is detailed in Section 3.

To align our methodology with the recommended procedures for conducting and publishing systematic reviews, we adhered to the PRISMA-P statement [31] along with the PRISMA-P 17-point checklist. This checklist facilitated robust guidelines for the formulation and reporting of a comprehensive review protocol.

3. Results

A total of 813 potentially eligible documents were identified for the systematic literature review. These documents were reduced to 728 after removing duplicates. From the original 728 documents, 616 were rejected for not meeting the inclusion criteria. This left us with 112 articles, which were evaluated based on relevance from their title and abstract. From this evaluation, a total of 39 articles were selected and then analyzed through a full-text reading. A total of 19 articles were finally included in the qualitative synthesis. Of the 19 articles finally selected, 10 articles came from journals cataloged in JCR or Scopus, and 9 were articles or reports from journals not appearing in JCR or Scopus. The PRISMA flow diagram (Figure 2) detailing our selection process is shown in the following image:

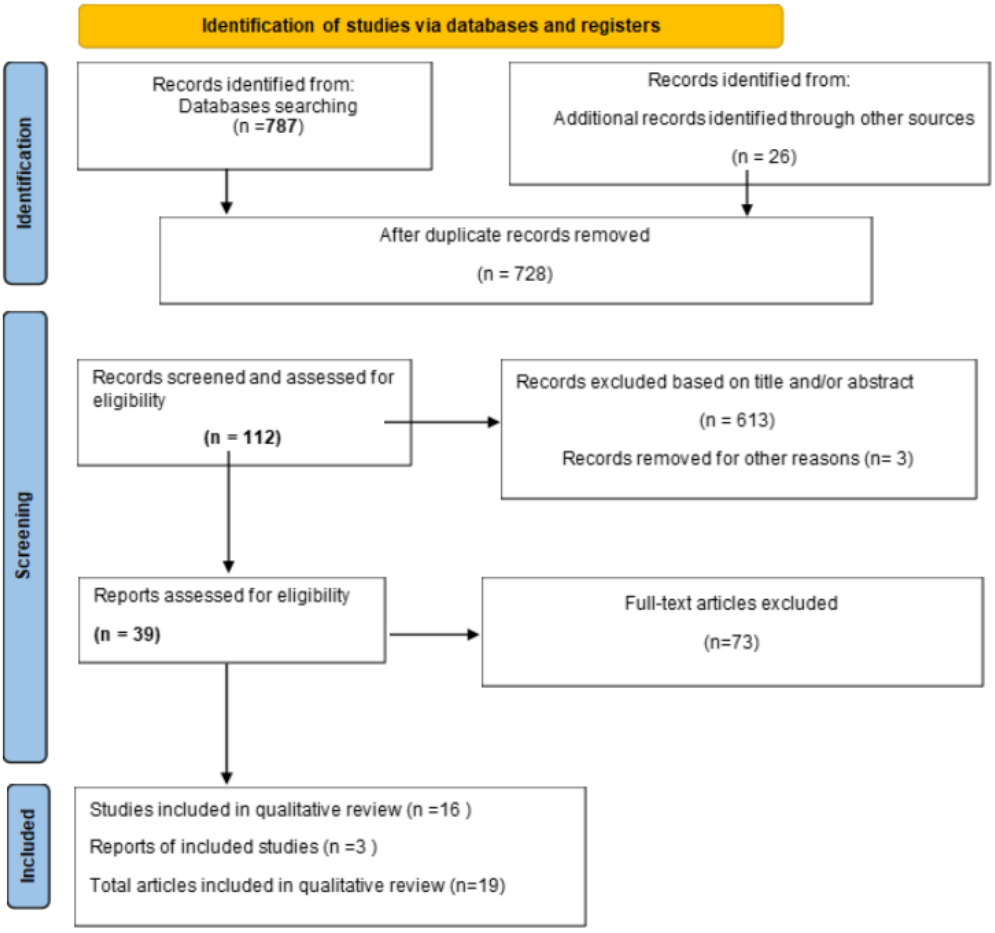

Figure 2. PRISMA 2009 flow diagram [33].

Table 2 summarizes the analysis of the 19 papers finally selected. The last three columns include a specific analysis focused on the housing-outcome variable used by each study; the definition of the variable provided by each document; and observations regarding the existence, or not, of a definition; the sources mentioned; and any possible incongruences within the definition or in comparison with the operational categories of the ETHOS definition.

The selected papers comprised four review studies, one report, and fourteen empirical studies. Of these, 37% (n = 7) of the studies were conducted in the USA, 15% in Canada (n = 3), 43% (n = 8) in Europe (Czech Republic, Germany, Ireland, Norway, Spain, and the United Kingdom), and one study in Australia (5%). Although the search was conducted in both Spanish and English, no articles written in Spanish were selected. More than half of the studies were published after 2012, with the first European article from this year onwards. Prior to 2012, all the included articles were originated from North American universities.

Of the 19 included studies, 10 (53%) were randomized controlled trials, 2 (11%) were quasi-experimental, and 3 (16%) were case studies. Among the included review studies, 3 of them encompassed randomized controlled trials, quasi-experimental, and case studies within the selected articles. Of the studies that used randomized controlled trials, 6 out of 10 were conducted in the USA, 2 out of 10 in Norway, 1 in Australia, and 1 in Spain. Studies with sample sizes exceeding 1000 were from the USA, Norway, and the United Kingdom. Regarding the follow up period, the maximum follow up duration was 72 months (6 years), and the minimum was 3 months.

Additionally, concerning the empirical studies, 3 out of 14 of them covered the entire population experiencing homelessness (within a specific region/country). The rest of the empirical studies focused on a subgroup, with individuals with severe mental illnesses experiencing homelessness being the most common subgroup. In some cases, studies focusing on severe mental illnesses also included the condition of chronic homelessness. Other subgroups studied were families and young women or individuals with addiction problems experiencing homelessness.

Regarding the housing intervention type, the most common intervention in the included studies was Housing First. Other housing interventions included transitional housing, supportive housing, short-term shelters, and rental subsidies. If we assess quality based on sample size, evaluation design (RCT), and follow up time for the evaluations, we find an overall higher quality in the United States, Canada, and Norway.

Table 2. Summary of information searched for in studies included in our review.

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
1	Lennon et al. (2005) [34]	US	Empirical	Quantitative	Outcome evaluation	RCT	Critical time intervention after shelter accommodation	Single homeless with mental health problems	96	18 months after program initiation	"Homeless"	"Our outcome measure consists of trajectories of homelessness over the observation period, divided into 18 months of 30 days each. A person was considered homeless within any 30-day period if he resided in a shelter, on the street, or in any other public place for just 1 night during that period".	Description is found. The definition corresponds to some of the ETHOS categories in a limited and oriented way.
2	Dostaler et al. (2003) [35]	Canada	Empirical	Mixed method	Outcome evaluation	Non-RCT (Case study)	Short-term shelter	Young women	40	3 months	"Housing"	No description provided. There is just a paragraph: "Housing. Most participants reported that housing had improved for them since they were now more stable where they lived. This stability in housing enabled them to start focusing on other areas of their lives".	No description is found.
3	Keenan et al. (2021) [36]	UK	Review	Quantitative	Outcome evaluations	RCT and Non-RCT	Accommodation-based interventions	Individuals experiencing or at risk of experiencing homelessness, irrespective of age and gender.	13.128	Different follow ups	"Homelessness" and "housing stability"	The article mainly uses Housing Stability.	No description is found.
4	Gubits et al. (2019) [37]	US	Empirical	Quantitative	Outcome evaluation	RCT	Long-term rent subsidies, short-term rent subsidies, and transitional housing	Families experiencing homelessness	2282	20 and 37 months later	"Homeless" and "housing stability"	"Measures of housing stability are (1) at least one night homeless or doubled up in the past six months, (2) any stay in emergency shelter in months 7 to 18 and months 21 to 32 after random assignment, (3) number of places lived in the past six months, and (4) at least one night homeless or doubled up in the past six months or any stay in emergency shelter in the past 12 months. Homeless was defined in the survey item to include living in a homeless shelter, in a place not typically used for sleeping, such as on the street, in a car, in an abandoned building, or in a bus or train station, or temporarily in an institution because the respondent had nowhere else to go".	Description is found for both homeless and housing stability.

Table 2. *Cont.*

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
5	Stefancic et al. (2007) [38]	US	Empirical	Quantitative	Outcome evaluation	RCT	Housing First: permanent and independent housing	Individuals with severe mental illness and co-occurring addictions and chronically homeless	260	20 months for control group and 47 months for experimental group	"Housing retention" and "housing status"	The information provided is "The first outcome, housing status, was a single point-in-time count of the number of persons housed within the two Housing First groups and the control group at 20 months. The second outcome, housing retention, consisted of housing retention rates for the two Housing First groups for a period of 47 months Rates of housing retention were calculated each month by dividing the number of consumers still maintaining housing by the number of consumers ever housed by the agency".	No description is found.
6	Sandu et al. (2021) [39]	Norway	Empirical	Mixed method	Outcome evaluation	RCT	Housing First: permanent and independent housing	Adults facing severe disadvantage: absolutely or precariously housed, and had a mental disorder	2141	24 months	"Housing stability"	RTLFB definition. They cite the source paper with the definition.	Description is found: RTLFB.
7	Wallace et al. (2018) [40]	Canada	Empirical	Quantitative	Outcome evaluation	Non-RCT (Case study)	Transitional housing	People experiencing homelessness	148	5 years (60 months)	"Housing outcome"	"Individuals entering the transitional program were emergency sheltered or provisionally sheltered as per the Canadian Definition of Homelessness [41]".	Description is found: Canadian Definition of Homelessness.
8	Bernad et al. (2016) [14]	Spain	Empirical	Quantitative	Outcome evaluation	RCT	Housing First: permanent and independent housing	Adults; homeless trajectory; facing mental health, substance abuse and/or a disability	69	24 months	"Housing retention"	They mention: "Housing retention as defined in the HF Europe". But no external source is mentioned. Researchers assume that "HF Europe" refers to the Housing First Guide [56]. The HF Guide provides three options to measure "Housing Sustainment": Length of time a Housing First service user has lived in the same home; Time spent in an apartment compared to time spent sleeping and living in other situations; or Individuals' feelings about their homes.	Description is not found.

Table 2. *Cont.*

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
9	Kuehnle et al. (2022) [42]	Australia	Empirical	Quantitative	Outcome evaluation	RCT	Housing First: permanent and independent housing	Chronically homeless	40	6 years (72 months)	"Housing"	Graphs show individuals who were "securely housed at the time of each survey". They also examine whether individuals "retained their housing after the program ended".	Description is not found. No clear definition is provided and multiple concepts are used throughout the article.
10	Lipton et al. (2000) [43]	US	Empirical	Quantitative	Outcome evaluation	RCT	Housing setting categories, categorized into high, moderate, or low intensity based on the amount of structure imposed and the degree of independence offered to tenants.	People with serious mental illness experiencing homelessness	2937	5 years (60 months)	"Tenure in housing"	"The outcome variable used in the study was tenure in housing. Individuals who became homeless, moved into unstable and marginal housing situations, or were imprisoned were considered to no longer be residing in stable housing and were classified as discontinuous placements. Those who remained in their initial housing or moved to settings regarded as stable housing were classified as being continuously housed. Individuals who were admitted to hospitals for physical conditions for extended times, who died, or who moved to appropriate housing but who could not be followed up by the Human Resources Administration were considered to be 'censored' at the time of the move and hence were not categorized as a discontinuous placement or as continuously housed".	Description is found. Nevertheless, the definition of "moving to setting regarded as stable housing" is not provided. It can be understood as several subcategories of the ETHOS considered homelessness (friend or family house).

Table 2. *Cont.*

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
11	Hoey et al. (2018) [44]	Ireland	Empirical	Quantitative	Outcome evaluation	Non-RCT (Case study)	Transitional housing	Families experiencing homelessness	288	6 months	"Secure housing"	"The majority of customers were living in either private rented accommodation (43%) or local authority housing (35%). Below is a breakdown of their current living situations, six months after disengaging from Focus Ireland services: 35% living in local authority housing. 33% in private rented sector accommodation with the housing assistance payment. 12% residing in Approved Housing Body housing. 5% in private rented sector accommodation with rent supplement. 4% residing in Focus Housing. 4% renting privately, independently without rental subsidies. 1% residing in privately-owned property. 1% living in transitional accommodation".	Description is not found. An explanation is made of where currently all the customers are, but no clear definition of "secure housing" is provided. Despite the lack of definition, it could be extracted from the long description of customers' current living situations. Some of the living situations seem to be considered "homeless" by the ETHOS, such as assisted accommodation, but cannot be identified without a proper definition.
12	Pleace (2013) [45]	UK	Review	Mixed methods	Mixed	RCT; quasi experimental and case studies	General	People experiencing homelessness	n.a.	n.a.	"Housing sustainment for potentially and formerly homeless people" (ending or preventing rooflessness and houselessness)	"Housing sustainability is defined as having the following characteristics: (1) Be affordable for potentially and formerly homeless people. (2) Be available for a long period or on an ongoing basis. Housing that is only available for a year or less, for example in the private rented sector in some EU countries, cannot provide a settled home and, by definition, cannot be sustained. (3) Be located in a neighborhood where there are acceptable levels of risks in terms of crime and nuisance behavior. (4) Be housing that is an acceptable state of repair and which offers acceptable space standards and basic amenities".	Description is found and an explanation of the basic principles contained in the concept of housing sustainability is present. Nevertheless, the definition is not used later to analyze the evaluations reviewed. Additionally, the definition hardly matches any of the existing definitions of homelessness.

Table 2. *Cont.*

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
13	Pauly et al. (2012) [46]	Canada	Review	Mixed methods	Mixed	RCT; quasi experimental and case studies	Permanent housing, transitional, monetary assistance and supportive housing	People experiencing homelessness	n.a.	Different follow ups	"Housing Status"	They define Housing status as: "Client's housing status (housed or homeless) and/or housing type before, during, and/or after the program; days spent homeless". They also use "Permanent independent housing refers to permanent, scattered site housing (not a single, dedicated building or housing project) as an intervention to end homelessness".	Description is found but "housed vs. homeless" is not defined. In fact, the paper mentions, "Secondly, in this project, we did not adopt a definition of homelessness. There are varied definitions of homelessness, with consensus definitions evolving in Europe, Australia, United States and Canada (European Federation of National Association Working with the Homeless (FEANTSA), 2007; Tipple & Speak, 2005)".
14	Glumbikova et al. (2020) [47]	Czech Republic	Empirical	Quantitative	Outcome evaluation	Quasi-experimental	Housing First: permanent and independent housing	People experiencing homelessness	147	12 months	They mention "Experience of need of housing", but they measure homeless-ness.	ETHOS. They state: "We have intentionally used the ETHOS typology, considering that it can provide a comparison of results in an international context. At the same time, however, it is necessary to be aware of the difference between the target group of people living in hostels and shelters; when homeless people from the category of roofless people can also be future clients of shelters rather than people living in hostels. This similarity between the two subgroups is further reflected in the data results themselves".	Description is found, using ETHOS. It is not clear how it is operationalized and the temporal evolution of the ETHOS depending on the assistance provided.

Table 2. *Cont.*

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
15	Busch-Geertsema (2013) [48]	Germany	Report	Mixed methods	Outcome evaluation	Non-RCT	Housing First: permanent and independent housing	People experiencing homelessness (different subgroups)	462	Different follow ups	"Housing Retention"	"In general, we have measured housing retention by the proportion of people who have been assigned housing by the HF project and have managed to sustain a tenancy (or to move to another tenancy) with the support of the project. If people have left the local programme in order to live in another apartment, this was generally seen as a positive case of housing retention. If people have died during their stay in the HF project, we have excluded such cases from the calculation of housing retention. It was more difficult to decide about those cases when people have moved from the HF project into a more institutionalized form of accommodation, like a long-term nursing home. In some cases, this was seen by service providers as the adequate form of accommodation given the support needs of the individual, but it cannot be seen as a success in relation to sustaining a tenancy, and in most cases we do not know as to what extent it was a desired solution by the person him- or herself. We have therefore opted for excluding those persons from the calculation of housing retention rates".	The paper mentions, "housing retention was measured in different ways at local level". On page 54, it is highlighted that housing retention rates are not comparable to the results in the US due to differences in the housing retention concept in Europe compared with those shown in Tsemberis et al. (2004 and 2012).

Table 2. *Cont.*

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
16	Munthe-Kaas (2016) [9]	Norway	Review	Quantitative	Outcome evaluation	RCT	Mixed	People who are homeless or at risk of becoming homeless	10.570	Different follow ups	Primary outcomes: "homeless-ness" and "residential stability"	"Number of days in stable housing, 12–24 months follow-up. The minimum follow up is 12 months after intake. Continuous data should describe the housing situation during specific periods, for instance, the past 30, 60, or 90 nights. This could be the mean number of nights, or the mean proportion of nights in a particular housing situation. Dichotomous data should involve the number of persons or the proportion of persons in different housing situations. Housing situations should be at least one of the following: homeless, unstable housing, or stable housing".	Description is found.
17	Hwang (2011) [49]	US	Empirical	Quantitative	Outcome evaluation	Quasi-experimental	Supportive housing program	People experiencing homelessness	112	18 months	"Residential stability"	"Residential stability was measured using the Residential Timeline Follow-Back Calendar, a validated method that allows for the collection of detailed and accurate information on housing history. Thirty-one participants' pattern of residences, hospital or prison stays, and periods of homelessness were recorded for each 6-month period. Residential stability was defined as living in one's own home or living with family or friends. Residential instability was defined as residing in jail, psychiatric hospital, drug treatment facility, or homeless shelters, or living in public places or on the street".	Description is found.

Table 2. *Cont.*

#	Author (Year)	Country	Study Type	Method Type	Evaluation Type	Study Design	Intervention Type	Population	N	Follow Up Time	Housing Outcome Variable	Definition of the Variable	Observations
18	Kertesz (2007) [50]	US	Empirical	Quantitative	Outcome evaluation	RCT	From permanent housing to emergency shelters	Treated cocaine-dependent people experiencing homelessness	99	12 months	"Stable housing"	"Days spent in the following settings [similar to Orwin's 'stably housed' category 30] counted toward stable housing: own apartment/house, parent/guardian's apartment/house, own single-resident occupancy (SRO), boarding house or board and care facility, group home and long-term alcohol/drug free facility. Settings such as shelter, treatment or recovery program (including those within shelters), corrections/halfway house, hospital, jail/prison, did not qualify".	Description is found. The paper explains the definition: "To provide policy-relevant information, categorical measures for stable housing and employment at one year were developed based on the treatment outcomes data".
19	Lim (2016) [51]	US	Empirical	Quantitative	Outcome evaluation	RCT	Supportive housing program	Youth experiencing homelessness (foster youth)	895	24 months after baseline	"Housing stability"	No housing stability definition is provided: "housing stability during the 2 years after baseline, which was defined as the first date a person became eligible for NYNY III. Sequence analysis was used to identify and define housing stability patterns (see a description of sequence analysis methods in the statistical analysis section)".	No description is found. The paper only identify patterns as a conclusion: "Three housing stability patterns (unstable housing, stable housing, and rare institutional dwelling patterns) were identified".

4. Discussion

In order to structure the main findings of the systematic review in a format useful for practitioners and policymakers responding to the challenges and dynamics of the European context, the discussion has been organized responding to the following specific questions: (a) Is there a standardized method for impact evaluation at a regional or national level for homelessness intervention programs? (b) Is there a standardized theoretical framework for impact evaluations? (c) Is there consensus on the best evaluation design? (d) Is there standardization in measuring the outcome variable? The four questions correspond to crucial components to consider when designing a robust evaluation: methodology, theoretical framework, evaluation design, and outcome variables (Figure 3).

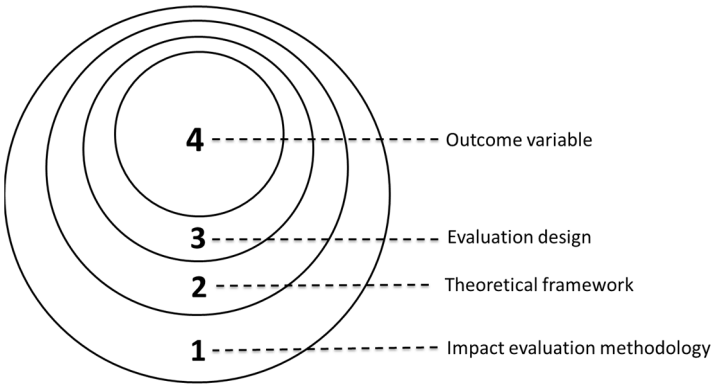

Figure 3. Structure of the discussion section.

4.1. Are There Established and Standardized Impact Evaluation Methodologies Validated by a Regulatory Agency and That Are Easily Replicable in Different Intervention Centers in Different Countries or That Are Established as a Regional or National Evaluation Methodology?

In this systematic review, no established and standardized impact evaluation methodologies, validated by a regulatory agency, which can be easily replicated in different centers or programs in different countries were found. Similar procedures and guidelines can be easily found in other fields, such as international development, by agencies, international organizations, or governments, such as, for example, OECD [52], World Bank [53,54], FAO [55], or USAID [56].

The evaluation methodology found in Housing First programs is the only impact evaluation methodology that aspires to be standardized and applied across various countries [48]. The methodology appears to be specifically designed to assess the impact of Housing First projects and has not been found to be applied in other programs or services, presenting certain methodological questions regarding the robustness of outcome variables for different welfare states, which will be further developed later in this section. However, Housing First evaluations must be recognized because they raise awareness of the need for impact evaluation and popularize randomized controlled trials among homelessness response structures. This observation seems to be consistent with the research by Baptista and Marlier [8] on the effectiveness of services for people experiencing homelessness in the European Union. "Housing First services, and to a lesser extent prevention services, are by far the area of provision where there is more evidence of available outcomes, allowing for a more solid and robust evaluation of the program's effectiveness".

In Pleace's review [45], five methodologies for evaluating homelessness programs are analyzed. The evaluated methodologies are the Danish national strategy for homelessness, the evaluation of the Finnish program "Name on the Door", the evaluation of homeless services in Dublin, the UK's "Homelessness Star", and the Dutch "Self-Sufficiency Matrix". The Finnish and Dublin cases will be set aside due to their focus on evaluating specific objectives of a service or particular strategy and not being developed with a standardized approach.

The Homelessness Star is oriented towards groups of individuals with high support needs experiencing homelessness. The tool monitors individual progress in 10 areas, through longitudinal scoring from 1 to 10, at the beginning of service utilization, during service use, and at the end of service use. The monitored areas are Motivation and Assumption of Responsibilities, Personal Care and Life Skills, Money and Personal Administration, Networks and Social Relationships, Drug and Alcohol Abuse, Physical Health, Emotional and Mental Health, Meaningful Use of Time, Housing and Accommodation Management, and Offending. The "Housing Maintenance" area is described as "Housing and Accommodation Management" and is scored from 1, which describes a homeless or potentially homeless person demanding to be left alone, to 10, which is a person who can live independently without support. Critics of the Outcome Star [57] highlight two main aspects. First, the inconsistency of the measurements, due to the reliance on professional teams' judgment instead of using validated scales. Second, that the tool assumes the need for behavioral change in the individual. The presence of outcome indicators such as "housing and accommodation management" or "money management", instead of the availability of money and housing, assume that the problem lies in the individual capacity of management.

The Dutch Self-Sufficiency Matrix (SSM) is not used specifically in homelessness response. Instead, the methodology serves to evaluate the outcomes of care services and to function as a tool to set goals and monitor progress in intervention for service providers. The SSM is intended to monitor 11 domains related to levels of self-sufficiency, scoring them (1 to 5) at three points in time: at the beginning of service utilization, during service use, and at the end of service use. A detailed guide is provided for scoring each domain. The domains include Income, Daytime Activities, Housing, Domestic Relationships, Mental Health, Physical Health, Addiction, Daily Life Skills, Social Network, Community Participation, and Judicial matters. The "Housing" domain is defined as "housing quality and likely duration of stay in current housing" and is scored from 1, acute problem, if someone is experiencing homelessness or in emergency accommodation, to 5, completely self-sufficient, if someone is fully responsible for all aspects of running their own home. One of the strengths of the tool is the detailed definition for scoring each of the domains. On the other hand, the tool appears to have a lack of solid theoretical framework that would enable the identification of both structural and individual factors that need to be addressed. Instead, it shows a focus on individual characteristics [45].

The rest of the reviewed documents do not provide any standardized methodology for impact evaluation. Nevertheless, certain trends in the applied methodologies can be identified. The first trend is the use of longitudinal approaches in most of the reviewed or empirical studies (95% of the documents). Longitudinal analysis enables the comparison of the evolution of controlled variables for the same subjects repeatedly over a period of time, thereby identifying the long-term effect of programs and enabling a deeper understanding of the complex processes leading people to experience homelessness and exit it [24]. The second trend is the use of randomized controlled trials (68% of the documents) as a methodology for evaluating homelessness response programs. The use of randomized controlled trials has been recommended by several authors, perhaps most prominently in terms of methodological dissemination, such as in the work of Duflo, Glennerster, and Kremer [26]. These trends appear consistent with the observations of Toro [5], Philippot [6], and Pleace [7] regarding the need to conduct large-scale longitudinal studies with control groups in the European context, allowing for the generation of knowledge in Europe and not relying primarily on findings produced by American and Canadian research.

The review also allows us to extract certain aspects that would be useful to consider in future evaluations. First, most evaluations do not use a standardized methodology or normalized guidelines, thus developing different evaluation methodologies in terms of techniques and variables. This characteristic complicates the possibility of subsequently comparing the results with other evaluations.

Second, given that the nature of homelessness and response structures is highly influenced by welfare states [58,59], and formal support structures, the methodologies aimed at evaluating interventions should also adapt to different welfare states. Consequently, and despite appearing to contradict the first point, the measurement areas and the variables that could be useful in one country or region may be irrelevant in another region. An example of a standard methodology solving this ambiguity can already be found in the European context [60].

Third, the weaknesses in identification of areas to measure and the clear focus on individual treatments or behaviors seem to be related to the lack of a solid theoretical framework. The importance of a robust theoretical framework in research has been highlighted by many authors in different fields of knowledge, "not only to derive specific and testable implications but also to provide a general direction of what the interesting questions are" [61].

Fourth, once the theoretical framework allows the definition of the areas to address and the specific variables to measure them, a detailed and coherent definition of each variable is needed to enable scoring of each domain. Not adequately defining the variables, even if they seem self-explanatory, would allow different evaluators to interpret the variables and scoring differently.

Fifth, the main outcome variable should be standardized to facilitate comparison between programs. For example, in evaluations of homelessness response programs, the fact that "residential situation" is not defined and is measured by different variables in each evaluation reduces the reliability of the results and also makes them difficult to compare. This recommendation makes even more sense in the European context when the ETHOS typology is available, which is conceptually robust, known in most European countries, and seems to adapt adequately to the different welfare states.

Sixth, and in conclusion, the factors mentioned above suggest a general weakness among the reviewed evaluations of homelessness response programs, especially in the European context.

4.2. Do Impact Evaluation Studies on Homelessness Response Present a Solid Theoretical Framework?

Evaluation approaches based on theoretical foundations should help us understand how programs are evaluated and serve as the basis for defining variables and indicators for impact evaluation. The necessary theoretical framework for the development of impact evaluations involves defining homelessness and also the identification of needs and resources within a specific socio-political and economic context, thus guiding the selection of areas to measure and the variables used.

The analysis of the different theoretical approaches used by the selected studies reveals two main characteristics: there is no single theoretical framework for the evaluation of homelessness response programs and there is a weakness in the definition of the theoretical framework used in the evaluations. This can lead to inconsistencies in the selection of variables used to measure both homelessness and the factors involved in its generation or resolution.

Below are examples of the diversity of theoretical frameworks used among the analyzed articles. In evaluations where homelessness is understood as a structural problem, the evaluation will be focused on the structural issues that lead to its occurrence. An example is the Canadian study by Pauly et al. [46], which argues that access to affordable housing or social rentals in a particular country or region is a critical factor in homelessness response programs. In the US context, where there is no universal healthcare system for citizens, the evaluation needs of homelessness services seem to be oriented towards assessing the prevention of recurrent hospitalizations or any other service that could increase public costs. The same argument applies in Lenon et al. [34]: Critical Time Intervention is a time-limited intervention that provides support and assistance to individuals during transitions between services or living situations, with the goal of preventing rehospitalization and reincarceration. In Stefancic et al. [38], a similar argument is used to advocate for Housing

First interventions for groups experiencing chronic homelessness: Housing First removes barriers to housing access and avoids the costs of care services consumed by individuals experiencing chronic homelessness. In contrast, for different vulnerable subgroups of individuals experiencing homelessness, such as families, the focus shifts, as explained in the US study by Gubits et al. [37]. In this case, a family experiencing homelessness is understood as an economic problem, so the evaluation shifts to comparing solutions related to different types of rental subsidies for families.

Furthermore, the European studies collected in this systematic review do not present a unified ideology, and it is difficult to identify a single conceptual framework beyond evaluations associated with Housing First. For example, the Norwegian article by Sandu et al. [39] argues that there is considerable variation in the impact on different populations receiving Housing First interventions, thus focusing on evaluating the assistance and additional support services alongside the provision of residential services. In the Czech Republic study by Glumbikova et al. [47], the main factor to evaluate is whether social housing results in savings in the provision of social work support. In the Irish evaluation study from 2008 described in Pleace's review, the understanding associated with homelessness is based on the idea that homelessness is better understood in terms of the pathways to housing [45]. The Spanish article [14] uses the Housing First evaluation approach, but it does not appear to consider any specific characteristics of the Spanish welfare system.

Despite the diversity of homelessness response programs that was found, 6 out of the 19 selected studies were evaluations of Housing First. In these cases, the evaluation framework was based on the new orthodoxy and principles of Housing First. Most of the selected studies based on this approach start from the premise that an individual housing intervention (housing as a right), which is unconditional (separation of housing and treatment) and sustained (people entering the program are considered out of homelessness, and support is available for as long as needed) is the best option for individuals experiencing homelessness. The starting point of these evaluations is to confirm that Housing First is an effective housing and treatment intervention that ends and prevents homelessness for individuals experiencing homelessness, which is especially effective for those with severe mental illness, co-occurring addictions, other health problems, and who have been homeless for years. The housing outcome variable typically used in Housing First evaluations is housing retention, since the approach's starting point is that significant barriers to housing entry are removed, and the main established goal is for the person to remain housed, even if it involves continued support from homelessness response programs in the long term.

This framework seems to be robust for cases of long-term homelessness with multiple complex pathologies within the current welfare structures and the levels of knowledge in social and healthcare fields. But this conceptual framework may not be the most coherent approach to apply in all homelessness response cases for several reasons.

First, because it is a conceptual framework created in the United States, and, as we have previously analyzed, based on its welfare system, it would require rethinking to adapt to the welfare systems and specific contexts of European countries. This adaptation might be necessary to address the availability and affordability of housing necessary for the models to be effective [62] or to pay more attention to the broader socio-economic and political context in which interventions are implemented [40].

Second, if we establish housing retention as the variable associated with the intervention's success, while simultaneously providing unconditional residential and/or economic support, there is a risk of institutionalizing people and denying their agency and the development of their capabilities, through which they could access a better life [63]. Findings like those from Rodilla et al. [64] demonstrate that income sources for homeless individuals from work earnings are more effective in reducing homelessness than unconditional government assistance.

Third, the theoretical framework of the analyzed Housing First evaluations appears to prioritize housing retention as the primary goal, rather than emphasizing the improvement of individuals' lives. The fundamental desired result of assisting individuals experiencing

homelessness should be to enhance their quality of life in various dimensions or central life capabilities [65]. By making housing retention the main target objective, there is a risk of assuming or allowing that merely keeping them housed is sufficient in the intervention process.

There is an obvious weakness in the theoretical frames that evaluations of homeless response programs use in the European context. In Europe, with countries of medium and high income and different housing contexts and welfare systems, homeless response services and strategies may work very differently depending on the context [36]. There is a need to apply evaluation approaches that take into account the broader socio-political and economic contexts in which the programs are being implemented.

4.3. Is There a Consensus on the Most Suitable Evaluation Design to Assess Homelessness Response Interventions?

At the academic level, there seems to be a consensus that the most rigorous evaluation design for assessing a specific program in terms of impact measurement is the randomized controlled trial (RCT), combined with long-term sample follow up and a large sample size that is representative of the studied population [26].

Among the analyzed articles, 53% followed an RCT design, with high-quality evaluations found in the US studies. Among the reviewed articles, there were no studies that followed an RCT, had a representative sample, and focused exclusively on the European population. This factor seems to be associated with one of the general conclusions observed in this review: the need, in European countries, for more comprehensive and rigorous evaluations to determine the most effective strategies and improve intervention outcomes for individuals experiencing homelessness.

Despite this overall consensus on the most effective designs for conducting impact evaluations, it is essential to remember that RCTs are expensive [66] and raise ethical issues [67]. RCT is a robust experimental design with certain limitations. Other study designs may be more suitable and easier to apply in specific applications. [68].

In parallel, methodologies alone do not determine the quality of an evaluation. Proper evaluation methodologies can contribute to the robustness of the evaluations conducted, but other factors should be taken into account, such as the conceptual framework and the areas and variables to be measured. The same evaluation methodology, with different areas and variables to be measured, can produce completely different results.

Finally, we must consider that we are analyzing quantitative methodologies, which facilitate their replicability and execution. However, to achieve higher levels of understanding of the results obtained, we should opt for mixed research methodologies. In this regard, Pleace [45] states that "no single method will provide a truly comprehensive evaluation of a service. A statistical analysis, which should be longitudinal, if possible, to test to what extent any benefit generated by a service for the homeless endures, is often very useful. However, statistics can only go so far in explaining how and why a service for the homeless is working the way it does. Real insights come from the homeless people using the service and the people providing it, making the case for a qualitative element in the service evaluation strong".

4.4. Is There a Consensus on the Most Suitable Housing Outcome Variables for Measuring the Impact in Homelessness Response Evaluations?

There does not seem to be a consensus on the most suitable outcome variables for measuring the impact in homelessness response evaluations. Among the main outcome variables used in different studies are homelessness, housing, housing stability, housing retention, housing status, housing situation at exit, tenure of housing, housing security, and housing maintenance. Some of these variables may seem conceptually similar; however, only 63% of the reviewed works provide a definition of the outcome variable, making comparison or equivalence creation impossible. Of these 63% of works that provide a definition, just 75% provide a concise definition that could be comparable to the ETHOS definition. This means that almost half of the works (47%) do not provide a solid definition

of the main outcome variable used in the evaluation, making it difficult to compare the results with later research or evaluations.

Regarding works that use a citation or external source for outcome variable definition, we can find three robust definitions. A single work [47] uses ETHOS (FEANTSA, 2020), providing a definition of homelessness. A single work [39] uses the Residential Time Line Follow Back Inventory [69], providing a definition of housing stability. A single work [40] uses the Definition of Homelessness in Canada [41].

In two cases, weaker external definition sources are used. The evaluation of the Housing First Habitat program in Spain [14] uses housing retention "defined as in Housing First Europe". However, no additional reference is provided, and the Housing First Guide [70] does not establish a specific and clear definition of housing retention. In its section "5.3. What to measure", it provides three options for measuring "Housing Maintenance": time a Housing First service user has lived in the same home, time spent in an apartment compared to time spent sleeping and living in other situations, and an individuals' feelings about their home.

The review of European Housing First projects by Busch-Geertsema [48] also uses housing retention but points out that "housing retention was measured in different ways at the local level" in the reviewed projects. It also notes that the housing retention rates of the projects reviewed in Europe are not comparable to the results in the United States due to differences in the conception of housing retention in Europe compared to that shown in Tsemberis et al. [71,72].

The origin of the research may play a significant role in the characteristics observed so far. The low use of ETHOS, as the European definition of homelessness, could be contributed to the fact that most of the reviewed works come from outside Europe or use population samples from outside Europe, as is the case with the Norwegian study. The lower level of definition in the outcome variable in European Housing First projects could be associated with the lower evaluation methodology quality standards in the European context.

However, the aforementioned problems in defining "housing retention" may not be the only ones associated with the use of this variable. We can identify three factors related to conceptual robustness, adaptation to the European context, and the purpose of measurement that recommends the use of ETHOS as the outcome variable in evaluating the impact of homelessness response.

First, both the European conception by Busch-Geertsema [48] and the American conception by Tsemberis et al. [69] consider certain individuals as having "housing retention" who, according to ETHOS, are still considered homeless. For instance, Item 26 of the Residential Time Line Follow Back Inventory, "Transitional Housing Program (long-term)", is considered "Stable", but according to ETHOS, it is considered homeless under the operational category 7, "People receiving longer-term support (due to homelessness)". This inconsistency was highlighted by Busch-Geertsema [73], who associated it with the different characteristics of welfare states, such as rental rights in programs for people experiencing homelessness and access to social housing and economic support which differ from country to country. However, taking into account the current European debate on deinstitutionalization, it seems incoherent to consider a person who is still dependent on a homeless support program to remain housed as being "out of homelessness".

Second, in the current context of the need for evidence-based research in Europe, it seems logical to base evaluations on the outcome variable using ETHOS, which was developed in Europe, rather than housing retention created in the United States. This is even more relevant considering that the ETHOS variable was designed to measure the housing status across different states and welfare systems within the European Union. The development of European tools and methods and the independence from American methodologies make even more sense following the recent wave of criticism towards Housing First regarding possible methodological weaknesses [18,74], potential biases [17], and doubts about the reliability of evidence-based policies [75].

Third, in terms of the conceptual framework used for designing the evaluation, as mentioned before, adopting housing retention as the outcome variable entails the risk of adopting a reductionist approach. Assuming that the primary objective of the intervention is to keep people housed, in a framework of understanding, such as the new orthodoxy, where homelessness is the result of a combination of individual traits and decisions within a set of structural conditions, changes in housing status should be associated with changes in different personal and structural dimensions. Under this conceptual framework, the selection of an outcome variable capable of monitoring change in multiple dimensions related to housing access (legal, physical, and social), such as ETHOS, becomes more appropriate.

4.5. Limitations of Study

This study acknowledges important limitations. While the search terms selected in the systematic review were intended to be comprehensive, they may not have ensured the full spectrum required for a meticulous investigation. Additionally, the potential for publication bias or alternative exclusion criteria is recognized. The search procedure, partly limited to scientific papers and articles from indexed journals with impact factors, may have unintentionally excluded relevant findings, frameworks and approaches from alternative studies. In addition, the incorporation of non-indexed articles and reports from specialized homelessness journals may have inadvertently excluded pertinent or more fitting findings.

The systematic review aimed to identify relevant research providing insights and learning that could apply beyond a specific local context. The goal was to identify impact evaluation methods for housing programs addressing homelessness on a global scale. Despite this global scope, the intent was to orientate the discussion and conclusions towards extracting meaningful insights specifically for practitioners and policymakers in Europe. This approach may have limited the number of relevant studies identified in the European context. Additionally, it is important to note that the systematic review was conducted exclusively in English and Spanish, potentially resulting in the omission of pertinent studies available in other local languages such as Italian, French, German, Portuguese, and others.

The systematic review specifically focused on evaluating standardized methods for assessing the impact of housing programs on homelessness. This focus raises the possibility that certain studies may not have been included if they did not explicitly pertain to housing programs. Additionally, the exclusion criteria resulted in the exclusion of research in which the housing situation was not considered as an outcome indicator. This factor raises concerns about potential articles that were excluded but may be relevant to the present discussion.

Moreover, the structure of Section 4 was organized around four components of evaluations: methodology, theoretical framework, evaluation design, and outcome variables. While this framework provides a systematic approach, it is important to acknowledge that alternative structures of analysis may have yielded different conclusions. A more diverse exploration may have provided additional perspectives and nuanced insights into the subject of study.

5. Conclusions

The systematic review of impact evaluation methodologies in homeless interventions reveals a lack of established and standardized approaches, particularly within the European context. The absence of a widely accepted regulatory-agency-approved methodology poses challenges to the comparability and generalizability of findings from different programs and countries. However, several important insights emerge from the review, shedding light on key considerations for future research and improvement in the field of impact evaluation of homeless interventions.

The analysis underscores the importance of a robust theoretical framework as the foundation for effective impact evaluations. The diversity of theoretical perspectives employed across the studies highlights the need for a more cohesive and context-sensitive approach that considers the unique socio-political and economic landscapes within which

interventions are implemented. Such an approach would enable a deeper understanding of the multifaceted factors contributing to entering and exiting homelessness and guide the selection of relevant areas and variables for measurement.

While randomized controlled trials (RCTs) are widely recognized as a rigorous evaluation design, their application remains limited, especially within the European setting. The lack of large-scale, robust, representative RCTs focused exclusively on the European population emphasizes the need for more comprehensive and rigorous evaluations that reflect the diverse socioeconomic realities of the region. To achieve a more comprehensive understanding of homelessness interventions, a mixed-methods approach is recommended, integrating quantitative methodologies with qualitative insights to capture a fuller picture of program effectiveness and impact.

One of the significant challenges highlighted by the review pertains to the lack of consensus on suitable outcome variables for measuring the impact of homelessness response interventions. This lack of uniformity in defining key variables, such as housing retention, hampers comparability and limits the potential for evidence-based policymaking. The review underscores the importance of adopting a standardized and comprehensive outcome variable, such as ETHOS, which is developed within the European context and encompasses multiple dimensions of housing access and stability.

In conclusion, while the systematic review reveals existing gaps in impact evaluation methodologies of homeless interventions, it also provides valuable insights for shaping future research and evaluation practices. Improvements in advocating a coherent theoretical framework, more consistent evaluation methodologies, and the adoption of standardized outcome variables should be addressed. By doing so, future evaluations can contribute to a more nuanced understanding of homelessness and to response programs that improve their effectiveness in addressing the multiplicity of cases behind the label of "homeless". Ultimately, a concerted effort towards methodological refinement and collaboration within the European context will enable evidence-driven decision making, leading to more impactful and tailored interventions for homeless individuals across diverse welfare states.

Author Contributions: Conceptualization, J.M.R. and G.P.; methodology, G.P. and M.B.; validation, M.B.; formal analysis, J.M.R. and G.P.; writing—original draft preparation, J.M.R. and G.P.; writing—review and editing, M.B.; supervision, M.B.; project administration, J.M.R. All authors have read and agreed to the published version of the manuscript.

Funding: Sant Joan de Déu Valencia, Spain (https://www.sanjuandediosvalencia.org/ accessed on 1 October 2023).

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article.

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

References

1. Pleace, N.; Hermans, K. Counting All Homelessness in Europe: The Case for Ending Separate Enumeration of 'Hidden Homelessness'. *Eur. J. Homelessness* **2020**, *14*, 35–62.
2. Amore, K.; Baker, M.; Howden-Chapman, P. The ETHOS Definition and Classification of Homelessness: An Analysis. *Eur. J. Homelessness* **2011**, *5*, 19–37.
3. Edgar, B. The ETHOS Definition and Classification of Homelessness and Housing Exclusion. *Eur. J. Homelessness* **2012**, *6*, 219–225.
4. FEANTSA. *European Typology of Homelessness and Housing Exclusion*; FEANTSA: Brussels, Belgium, 2005. Available online: https://www.feantsa.org/download/ethos2484215748748239888.pdf (accessed on 11 October 2023).
5. Toro, P.A. Toward an international understanding of homelessness. *J. Soc. Issues* **2007**, *63*, 461–481. [CrossRef]
6. Philippot, P.; Lecocq, C.; Sempoux, F.; Nachtergael, H.; Galand, B. Psychological research on homelessness in Western Europe: A review from 1970 to 2001. *J. Soc. Issues* **2007**, *63*, 483–503. [CrossRef]
7. Pleace, N. Researching homelessness in Europe: Theoretical perspectives. *Eur. J. Homelessness* **2016**, *10*, 19–44.
8. Baptista, I.; Marlier, E. *Fighting Homelessness and Housing Exclusion in Europe: A Study of National Policies. European Social Policy Network*; European Commission: Brussels, Belgium, 2019.
9. Munthe-Kaas, H.; Berg, R.C.; Blaasvær, N. Effectiveness of interventions to reduce homelessness: A systematic review and meta-analysis. *Campbell Syst. Rev.* **2016**, *14*, 1–281. [CrossRef]

10. Duque, J.M. Políticas de servicios sociales para la atención a personas en situación de exclusión residencial grave. Elementos para un diagnóstico y propuestas estratégicas. *Zerb. Rev. Serv. Soc.* **2014**, *55*, 35–46. [CrossRef]
11. Marbán, V.; Rodríguez, G. Las políticas sociales de lucha contra el sinhogarismo en la Unión Europea y España: Alcance, efectividad y principales limitaciones y prioridades. *Zerb. Rev. Serv. Soc.* **2020**, *72*, 5–18.
12. Departamento de Empleo y Políticas Sociales. *Estrategia Vasca para Personas sin Hogar 2018–2021*; Gobierno Vasco: Vitoria-Gasteiz, Spain, 2018.
13. Pleace, N.; Bretherton, J. The Case for Housing First in the European Union: A Critical Evaluation of Concerns about Effectiveness. *Eur. J. Homelessness* **2013**, *7*, 21–41.
14. Bernad, R.; Yuncal, R.; Panadero, S. Introducing the Housing First Model in Spain: First Results of the Habitat Programme. *Eur. J. Homelessness* **2016**, *10*, 53–82.
15. Bernad, R.; Cenjor, V.; Yuncal, R. Housing First Model in Spain: Habitat Programme 12 months-results. *Rev. Barc. Soc.* **2016**, *20*, 1–12.
16. Martínez-Cantos, J.L.; Martín-Fernández, J. Cost-effectiveness of a 'Housing First' programme implemented in Spain: An evaluation based on a randomised controlled trial. *Int. J. Soc. Welf.* **2023**, 1–17. [CrossRef]
17. Baxter, A.J.; Tweed, E.J.; Katikireddi, S.V.; Thomson, H. Effects of Housing First approaches on health and well-being of adults who are homeless or at risk of homelessness: Systematic review and meta-analysis of randomized controlled trials. *J. Epidemiol. Community Health* **2019**, *73*, 379–387. [CrossRef] [PubMed]
18. Groton, D. Are Housing First Programs Effective? A Research Note. *J. Sociol. Soc. Welf.* **2013**, *40*, 51–63. [CrossRef]
19. Ministerio de Derechos Sociales y Agenda 2030, Gobierno de España. *Diagnóstico inicial II Estrategia Nacional Integral para Personas sin Hogar 2023–2030*; Ministerio de Derechos Sociales y Agenda 2030, Gobierno de España: Madrid, Spain, 2023.
20. Somerville, P. Understanding Homelessness. *Hous. Theory Soc.* **2013**, *30*, 384–415. [CrossRef]
21. Muñoz, M.; Panadero, S.; Santos, E.P.; Quiroga, M.Á. Role of Stressful Life Events in Homelessness: An Intragroup Analysis. *Am. J. Community Psychol.* **2005**, *35*, 35–47. [CrossRef]
22. Hirsch, G.B.; Mosher, H.I. Using a System Dynamics Simulation Model to Identify Leverage Points for Reducing Youth Homelessness in Connecticut. *Systems* **2023**, *11*, 163. [CrossRef]
23. Racette, E.H.; Fowler, C.A.; Faith, L.A.; Geis, B.D.; Rempfer, M.V. Characteristics of Trauma among Women Experiencing Homelessness: An Exploratory Cluster Analysis. *N. Am. J. Psychol.* **2021**, *23*, 569.
24. Benjaminsen, L.; Muñoz, M.; Vázquez, C.; Panadero, S. Quantitative methods in Homelessness Studies: A critical guide and recommendations. In Proceedings of the Conference on Research on Homelessness in Comparative Perspective, Brussels, Belgium, 3–4 November 2005; Volume 3.
25. Banerjee, A.; Duflo, E. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*; PublicAffairs: New York, NY, USA, 2012.
26. Duflo, E.; Glennerster, R.; Kremer, M. Using randomization in development economics research: A toolkit. In *Handbook of Development Economics*; National Bureau of Economic Research: Cambridge, MA, USA, 2007; Volume 4. [CrossRef]
27. White, H.; Sabarwal, S.; De Hoop, T. *Ensayos Controlados Aleatorios*; UNICEF: New York, NY, USA, 2014.
28. Rodriguez-Guzman, G.; Hume, S.; Mackie, P.; Nwosu, C.; Piggott, H.; White, J. Accelerating Learning—Lessons and Reflections from the First Randomised Controlled Trials in Homelessness in the UK. *Eur. J. Homelessness* **2021**, *15*, 97–111.
29. Whittemore, R. Combining evidence in nursing research: Methods and implications. *Nurs. Res.* **2005**, *54*, 56–62. [CrossRef]
30. Whittemore, R.; Knafl, K. The integrative review: Updated methodology. *J. Adv. Nurs.* **2005**, *52*, 546–553. [CrossRef] [PubMed]
31. Moher, D.; Shamseer, L.; Clarke, M.; Ghersi, D.; Liberati, A.; Petticrew, M.; Stewart, L.A. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **2015**, *4*, 1–9. [CrossRef] [PubMed]
32. Judge, A.T.; Cable, D.M.; Colbert, A.E.; Rynes, S.L. What causes a management article to be cited—Article, author, or journal? *Acad. Manag. J.* **2007**, *50*, 491–506. [CrossRef]
33. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]
34. Lennon, M.C.; McAllister, W.; Kuang, L.; Herman, D.B. Capturing intervention effects over time: Reanalysis of a critical time intervention for homeless mentally ill men. *Am. J. Public Health* **2005**, *95*, 1760–1766. [CrossRef]
35. Dostaler, T.; Nelson, G. A Process and Outcome Evaluation of a Shelter for Homeless Young Women. *Psychol. Fac. Publ.* **2003**, *19*, 99–112. [CrossRef]
36. Keenan, C.; Miller, S.; Hanratty, J.; Pigott, T.; Hamilton, J.; Coughlan, C.; Mackie, P.; Fitzpatrick, S.; Cowman, J. Accommodation-based interventions for individuals experiencing, or at risk of experiencing, homelessness. *Campbell Syst. Rev.* **2021**, *17*, e1165. [CrossRef]
37. Gubits, D.; Shinn, M.; Wood, M.; Brown, S.R.; Dastrup, S.R.; Bell, S.H. What Interventions Work Best for Families Who Experience Homelessness? Impact Estimates from the Family Options Study. *J. Policy Anal. Manag.* **2018**, *37*, 735–766. [CrossRef]
38. Stefancic, A.; Tsemberis, S. Housing First for long-term shelter dwellers with psychiatric disabilities in a suburban county: A four-year study of housing access and retention. *J. Prim. Prev.* **2007**, *28*, 265–279. [CrossRef]
39. Sandu, R.D.; Anyan, F.; Stergiopoulos, V. Housing first, connection second: The impact of professional helping relationships on the trajectories of housing stability for people facing severe and multiple disadvantage. *BMC Public Health* **2021**, *21*, 249. [CrossRef]

40. Wallace, B.; Pauly, B.; Perkin, K.; Cross, G. Where's the Housing? Housing and Income Outcomes of a Transitional Program to End Homelessness. *J. Poverty* **2018**, *23*, 161–178. [CrossRef]
41. Gaetz, S.; Barr, C.; Friesen, A.; Harris, B.; Hill, C.; Kovacs-Burns, K.; Pauly, B.; Pearce, B.; Turner, A.; Marsolais, A. *Canadian Definition of Homelessness*; Canadian Observatory on Homelessness Press: Toronto, ON, Canada, 2012.
42. Kuehnle, D.; Johnson, G.; Tseng, Y. Making it home? Evidence on the long-run impact of an intensive support program for the chronically homeless on housing, employment and health. *J. Urban Econ.* **2023**, *133*, 103511. [CrossRef]
43. Lipton, F.R.; Siegel, C.; Hannigan, A.; Samuels, J.; Baker, S. Tenure in supportive housing for homeless persons with severe mental illness. *Psychiatr. Serv.* **2000**, *51*, 479–486. [CrossRef] [PubMed]
44. Hoey, D.; Sheridan, S. *Are You Still Ok? Housing Tenancy Sustainment among Focus Ireland Customers*; Focus Ireland: Dublin, Ireland, 2018.
45. Pleace, N. *Evaluating Homelessness Services and Strategies. A Review*; Centre for Housing Policy: York, UK; European Observatory on Homelessness: Brussels, Belgium, 2013.
46. Pauly, B.; Carlson, E.; Perkin, K. *Strategies to End Homelessness: Current Approaches to Evaluation*; Canadian Homelessness Research Network Press: Toronto, ON, Canada, 2012.
47. Glumbikova, K.; Rusnok, P.; Mikulec, M. Impact Evaluation of the Provision of Social Housing on the Use of Social Services by Homeless People in the Czech Republic. *Sustainability* **2020**, *12*, 10184. [CrossRef]
48. Busch-Geertsema, V. *Housing First Europe. Final Report*; European Union Programme for Employment and Social Security: Brussels, Belgium, 2013.
49. Hwang, S.; Gogosis, E.; Chambers, C.; Dunn, J.; Hoch, J.; Aubry, T. Health Status, Quality of Life, Residential Stability, Substance Use, and Health Care Utilization among Adults Applying to a Supportive Housing Program. *J. Urban Health Bull. N. Y. Acad. Med.* **2011**, *88*, 1076–1090. [CrossRef] [PubMed]
50. Kertesz, S.G.; Mullins, A.N.; Schumacher, J.E.; Wallace, D.; Kirk, K.; Milby, J.B. Long-term housing and work outcomes among treated cocaine-dependent homeless persons. *J. Behav. Health Serv. Res.* **2007**, *34*, 17–33. [CrossRef]
51. Lim, S.; Singh, T.P.; Gwynn, R.C. Impact of a Supportive Housing Program on Housing Stability and Sexually Transmitted Infections Among Young Adults in New York City Who Were Aging out of Foster Care. *Am. J. Epidemiol.* **2017**, *186*, 297–304. [CrossRef]
52. OECD. *Impact Evaluations and Development*; OECD: Washington, DC, USA, 2009. Available online: https://www.oecd.org/development/evaluation/dcdndep/47466906.pdf (accessed on 11 October 2023).
53. World Bank. *Impact Evaluation in Practice*; World Bank: Washington, DC, USA, 2016. Available online: https://openknowledge.worldbank.org/server/api/core/bitstreams/4659ef23-61ff-5df7-9b4e-89fda12b074d/content (accessed on 11 October 2023).
54. World Bank. *Handbook on Impact Evaluation*; World Bank: Washington, DC, USA, 2010. Available online: https://documents1.worldbank.org/curated/en/650951468335456749/pdf/520990PUB0EPI1101Official0Use0Only1.pdf (accessed on 11 October 2023).
55. FAO. *Tool for Designing, Monitoring and Evaluating Land Administration Programmes in Latin America*; FAO: Rome, Italy, 2023. Available online: https://www.fao.org/in-action/herramienta-administracion-tierras/introduction/presentation-tool/ar/ (accessed on 11 October 2023).
56. USAID. *Guide for Planning Long-Term Impact Evaluations*; USAID: Washington, DC, USA, 2018. Available online: https://pdf.usaid.gov/pdf_docs/PA00T9HJ.pdf (accessed on 11 October 2023).
57. Johnson, G.; Pleace, N. How do we measure success in homelessness services?: Critically assessing the rise of the homelessness outcomes star. *Eur. J. Homelessness* **2016**, *10*, 31–51.
58. Stephens, M.; Fitzpatrick, S. Welfare Regimes, Housing Systems and Homelessness: How are they linked? *Eur. J. Homelessness* **2007**, *1*, 201–212.
59. Benjaminsen, L. Homelessness in a Scandinavian welfare state: The risk of shelter use in the Danish adult population. *Urban Stud.* **2015**, *53*, 2041–2063. [CrossRef]
60. Rodilla, J.M.; Puchol, G.; Botija, M. LongHome: Desarrollo de una herramienta para la medición de impacto de los servicios de atención a personas sin hogar. *Pedagog. Treb. Social. Rev. Ciènc. Soc. Apl.* **2022**, *11*, 55–86. [CrossRef]
61. Duflo, E. Field experiments in development economics. *Econom. Soc. Monogr.* **2006**, *42*, 322.
62. Zerger, S.; Francombe Pridham, K.; Jeyaratnam, J.; Connelly, J.; Hwang, S.; O'Campo, P.; Stergiopoulos, V. The role and meaning of interim housing in housing first programs for people experiencing homelessness and mental illness. *Am. J. Orthopsychiatry* **2014**, *84*, 431. [CrossRef] [PubMed]
63. Sen, A. *Development as Freedom*; Alfred Knopf: New York, NY, USA, 1999.
64. Rodilla, J.M.; Puchol, G.; Botija, M. Sinhogarismo y fuente de ingresos: Estudio longitudinal sobre la influencia del apoyo formal y la renta del trabajo en la reducción del sinhogarismo. *OBETS Rev. Cienc. Soc.* **2023**, *18*, 173–188. [CrossRef]
65. Nussbaum, M. Capabilities as fundamental entitlements: Sen and social justice. *Fem. Econ.* **2003**, *9*, 33–59. [CrossRef]
66. Bhargava, A. Randomized controlled experiments in health and social sciences: Some conceptual issues. *Econ. Hum. Biol.* **2008**, *6*, 293–298. [CrossRef]
67. Goldstein, C.E.; Weijer, C.; Brehaut, J.C.; Fergusson, D.A.; Grimshaw, J.M.; Horn, A.R.; Taljaard, M. Ethical issues in pragmatic randomized controlled trials: A review of the recent literature identifies gaps in ethical argumentation. *BMC Med. Ethics* **2018**, *19*, 14. [CrossRef] [PubMed]

68. Grossman, J.; Mackenzie, F.J. The randomized controlled trial: Gold standard, or merely standard. *Perspect. Biol. Med.* **2005**, *48*, 516–534. [CrossRef]
69. Tsemberis, S.; McHugo, G.; Williams, V.F.; Hanrahan, P.; Stefancic, A. Measuring homelessness and residential stability: The residential time-line follow-back inventory. *J. Community Psychol.* **2007**, *35*, 29–42. [CrossRef]
70. Pleace, N. *Housing First Guide*; FEANTSA: Brussels, Belgium, 2016.
71. Tsemberis, S.; Gulcur, L.; Nakae, M. Housing First, consumer choice, and harm reduction for homeless individuals with a dual diagnosis. *Am. J. Public Health* **2004**, *94*, 651–656. [CrossRef]
72. Tsemberis, S.; Kent, D.; Respress, C. Housing Stability and Recovery Among Chronically Homeless Persons With Co-Occuring Disorders in Washington, DC. *Am. J. Public Health* **2012**, *102*, 13–16. [CrossRef] [PubMed]
73. Busch-Geertsema, V. Defining and Measuring Homelessness. In *Homelessness Research in Europe*; FEANTSA: Brussels, Belgium, 2010; pp. 19–39.
74. Kertesz, S.G.; Crouch, K.; Milby, J.B.; Cusimano, R.E.; Schumacher, J.E. Housing first for homeless persons with active addiction: Are we overreaching? *Milbank Q.* **2009**, *87*, 495–534. [CrossRef] [PubMed]
75. Stanhope, V.; Dunn, K. The curious case of Housing First: The limits of evidence–based policy. *Int. J. Law Psychiatry* **2011**, *34*, 275–282. [CrossRef] [PubMed]