A critical examination of the youth foyer model for alleviating homelessness: Strengthening a promising evidence base

Iris Levin, Brotherhood of St Laurence
Joseph Borlagdan, Brotherhood of St Laurence and University of Melbourne
Shelley Mallett, Brotherhood of St Laurence and University of Melbourne
Jehonathan Ben, Brotherhood of St Laurence and Deakin University
Corresponding author: jborlagdan@bsl.org.au

Abstract

This article assesses the quality of 15 primary studies that examined the effectiveness of youth foyer or foyer-like programs on the lives of young homeless people. The youth foyer model provides an integrated approach to tackling youth homelessness, connecting affordable accommodation to training and employment. In Australia, there is growing support from government for the development and funding of foyer programs. However to date, there has been very limited development and investment in the evidence base on the effectiveness of this model in Australia or internationally. Following an extensive literature search, we argue that there is a need to lift the standard of the evidence base of youth foyer effectiveness. We discuss two main issues: the difficulty studies had validating claims of foyer effectiveness, and limitations of research design and methodology. The implications of the lack of rigour in the research reviewed are three-fold. Firstly, youth foyer evaluation study quality could be improved by: clearer methodological and model documentation; post intervention follow-up design; comparison of data to non-randomised comparison groups; and a pre-publication peer-review process. This would be supported with clearer expectations from the research community regarding the production and assessment of grey literature. Secondly, while the standard of reporting needs to be raised, the 'gold standard' (i.e. randomised controlled trials) of research design in the scientific community is not a relevant benchmark in the field of homelessness research. This is due to the complexity of homelessness interventions and the inadequate funding of the homelessness research field. Greater investment in robust research and evaluation should accompany the substantial investment in youth foyer programs in order to accurately appraise the effectiveness of the youth foyer model. Thirdly, the lack of rigour in the studies reviewed suggests gaps in the service development of the youth foyers that were evaluated in the articles considered in this evidence review. The research reviewed was mostly unable to report key program mechanisms, pointing to a lack of program documentation. Ideally, strong service development practices would enable evaluative studies to explore the link between foyer model mechanisms and outcomes. Policy implications include putting in place a system for ensuring adequate program documentation with robust research design and methods, using a theoretical framework for the interpretation of findings, and adopting a peer-review process. To achieve this, public sector commissioners of youth foyers need to tighten the evaluation and research components when funding new foyers, recognising the critical relationship between service development and research.

The youth foyer model is a notable exception to standard youth homelessness services. foyers provide an integrated approach to tackling youth homelessness, combining affordable accommodation with training and employment opportunities (Anderson and Quilgars 1995).[1] In Australia and internationally, homeless young people stay in foyer

---

[1] The 2011 ABS Census estimated that 105,237 people were homeless on Census night. Of these, 26,238 were young people aged 12–24, although this figure was recognised by the ABS as an under-estimation of the rate of youth homelessness. Most of the homeless youth aged 12–18 years in 2011 in Australia were in 'severely' crowded dwellings (56%) or in supported accommodation for the

programs for up to 24 months with the aim of supporting them while they undertake further education, training or enter employment, and subsequently move onto independent living or move back with family members (Steen and Mackenzie 2013).[2] This innovative model addresses an education and employment need that current services of youth homelessness are not addressing. A range of evidence points to the relatively low rates of engagement in education and employment for young people experiencing homelessness or living in severely overcrowded housing. For example, of the young people using specialist homelessness services in the last quarter of 2011 who nominated their educational status (63 percent of those presenting), only 29 percent were enrolled in formal study or training (AIHW 2012; Mallett et al. 2014).

First developed in France during the mid-nineteenth century, and subsequently adapted for use in the UK in the early 1990s, there are now foyers in the US, Ireland, Romania, the Netherlands, Germany and Australia (Foyer Foundation 2015; Mallett et al. 2014, 3). Australian foyers were first established in the early 2000s to address youth homelessness in Sydney, Wollongong and Melbourne. In 2015 there are over 15 foyers or foyer-type services across all States and Territories except Tasmania and the Northern Territory.

While the causes and consequences of homelessness among young people are complex, service responses to youth homelessness in Australia have largely remained the same for the last 30 years. These services often struggle to provide systematic, integrated support in education, employment and community-based responses (Mallett et al. 2014). Despite relatively recent reforms to homelessness policy and programs instituted through the Australian government's white paper on homelessness (*The Road Home,* Commonwealth of Australia 2008), the core approach to youth homelessness service delivery remains crisis driven. These services typically support a young person in crisis, stabilise their circumstances in short term accommodation and then move them on as quickly as possible. Where longer support is required, homeless young people are moved into medium term transitional accommodation and from there to private rental, social housing, or accommodation with family and friends.

Over the past two decades, a diverse but relatively small body of national and international research on youth homelessness has emerged. However, much of this comprises descriptive studies detailing the pathways of the problem (conflict at home, personal mental health issues, drug or alcohol use (their own or another family member) or the desire for independence) (Mallett et al. 2010), or small scale evaluations of service effectiveness. Youth foyer evaluation research is no exception. The type, range and quality of the evaluation evidence on the effectiveness of the youth foyer model has been questioned. Some note that outcome measurement of the UK foyers is underdeveloped and in need of improvements (Quilgars et al. 2008, 103). Others observe that there is limited quantitative evidence available on the effectiveness of Australian foyers (Gronda 2009, 30).

While researchers have questioned the evidence base on the effectiveness of foyers, to date there has been no comprehensive, critical examination of the available research. Although evidence-based claims about the model indicate that it is effective in a range of ways, these findings remain untested.

We argue that there is a need to raise the quality of the existing evidence about the youth foyer model so that the effectiveness of the model can be assessed. During the

---

homeless (28%). While 8% of homeless people aged 12–18 years were staying temporarily with other households, this proportion increased to 14% for youth aged 19–24 years (ABS 2012a, b).

[2] The average stay in Australia is, however, 11 months as some young people leave the Foyer after a short stay (Hanover 2012).

process of selecting and reviewing studies evaluating its effectiveness, we have found that the lack of quality in the evidence base for youth foyers represents a significant gap in our knowledge. This makes it difficult to substantiate claims, either positive or negative, about the effectiveness of this model for young people experiencing homelessness. If we are to achieve evidence-based policy and effective service reform in this field it is critical that the evidence base on youth foyers keeps pace with the economic investment in this model. The question we ask in this paper is: How can strong evidence base be built so that we can assess the effectiveness of the youth foyer intervention?

We acknowledge that 'evidence' is a subjective term that cuts across political ideologies, beliefs and values. In many fields of social policy, there are often conflicting evidence and research findings about specific issues. While conflicting findings can become the topic of future research, there is still the question of which evidence is chosen for the purpose of policy and what 'counts as evidence' (Fopp 2014; Parsell et al. 2014, 69), as different forms of evidence can be important to policy formation.

However, a strong quality of evidence is imperative not only for foyer evaluation but for all other human services evaluations. Such relevant research that discusses implications for the mainstream service system can inform policy formation and assist in improving traditional services and developing new innovative services.

This paper aims to:

- appraise the scope and rigour of available research reporting on youth foyer's effectiveness;
- assess the quality and relevance of this research;
- identify potential gaps in the quality of this research; and,
- discuss potential implications and recommendations for future evaluations.

This review sought to examine a wide array of sources, including academic papers, but it was found that the majority of the evaluative research was in the form of grey literature (in this case, defined as reports that are not commercial and scholarly peer-reviewed publications) of varying levels of rigour. We identify the gaps in the quality of the assessable evidence and, most importantly, we suggest ways to improve the quality of studies developed for this medium, including the establishment of a peer review process, and identification of research design issues that could be improved.

We start with an explanation of how we have undertaken this assessment of the quality of evidence. We then present details about the available evidence on youth foyers, discussing with examples the two aspects of our findings: the difficulty of validating claims made in various studies about the programs' effectiveness; and the limitations of studies' research design and methodology. Finally, we discuss the implications and recommendations of this review for future research and policy development in the field of youth homelessness, as well as in other service systems.

Assessing the quality of the youth foyer evidence base

The initial aim of this study was to assess the effectiveness of youth foyers. However, early on in our search for literature on youth foyers, we realised that not only is the pool of studies limited, but also their quality is questionable. This is not an unusual finding in the field of homelessness, as well as broader welfare service research in Australia. Research and evaluation funding in Australia is very limited, and until relatively recently little attention has been paid by governments and policy makers to building the

quality of the evidence base for service interventions. Consequently, we shifted our focus from examining the extent to which programs are effective as reported in evaluation studies, to examining the quality of these evaluation studies and whether we can validate their claims about effectiveness.

We used various methods to source the international and national peer-reviewed evidence and grey literature on foyers. Due to the lack of research on foyers (Quilgars 2001; Ralphs 2004; Steen and Mackenzie 2013), we expected that a large portion of the literature would be grey literature, or 'material produced and published by organisations without recourse to the commercial or scholarly publishing industry' (Lawrence et al. 2014, 2). These reports largely comprised small scale evaluations commissioned and undertaken by community sector agencies delivering foyers.

First, we scanned and screened an existing youth homelessness and foyer endnote database developed by the authors as part of an ongoing larger evaluation project of Victorian 'Education First Youth Foyers'. Then we conducted an online database search of EBSCOhost, including the following databases: Academic One, SocINDEX, JSTOR Journals, ScienceDirect, ERIC and PsycINFO. We also searched Google Scholar and key international and national websites on homelessness research. The keyword foyer* had to be mentioned either in the title or in the abstract. Additionally, papers had to include one or more of the following keywords in the title or abstract: homeless*, housing, acommodat*, tenan*, residen*, youth*, young people*, adolescen*. We also scanned the reference list of relevant studies for additional studies. Finally, additional papers were retrieved by directly contacting experts, including the Foyer Federation and specific studies' authors.

Inclusion criteria

Due to the scarcity of studies evaluating the foyer model, we set broad inclusion criteria beyond the narrow research design set by the gold standard of randomised controlled trials (RCT). While the gold standard sits atop the hierarchy of research methods in scientific quantitative research (Leigh 2009), it is less relevant to the field of homelessness research. Although there have been examples of RCT studies in this field, such as the Journeys to Social Inclusion in Melbourne (J2SI, Johnson et al. 2015), and others in the US (see Hwang et al. 2005), this approach is rarely used for several reasons. First, RCTs require substantial economic resources (Johnson et al. 2015, 9) that are more commonly found in the health sciences. In the homelessness research field, policymakers and service providers rarely invest in research of the scale required by an RCT design. Moreover, there are ethical considerations that prevent the assigning of homeless people to experimental and control groups in order to observe intervention differences, thus effectively denying one or more groups access to a fundamental service (Commonwealth of Australia 2008, 60; Pleace 2013, 20). Furthermore, poor practices in collecting RCT evidence can also provide poor quality evidence of effectiveness. Conversely, non-RCT data can provide good quality evidence of effectiveness, for example by comparing program participants to a comparison group rather than to a randomly assigned control group (Rogers 2010).

'Effectiveness' was defined in a broad sense, beyond the limited definition of cost-effectiveness analysis, which measures the ratio between cost (measured in monetary values) and effectiveness (measured in physical units) (Kaplan 2014). Instead, effectiveness for the purpose of this review can mean a change in the quality of a characteristic and not only a change in its degree (e.g. a change in relationships' quality beyond their quantity), including various impacts and outcomes measured during and after the program. Since studies that address effectiveness of youth foyers are scarce,

this broad definition expanded the available pool of studies. Evidence on programs that specified 'foyer' or 'foyer-like' models was included and was not limited to any geographical locations or countries. The types of foyer or foyer-like programs were not limited. For example, unique programs tailored for particular groups (e.g. focused on mental health, parenting skills or employment skills) were included.

The present review includes published and unpublished empirical studies conducted between January 1990 and October 2014 that report primary (qualitative and/or quantitative) data on the effectiveness of youth foyers. Document types included in the review are journal articles, books, theses and dissertations, reports, and conference papers. Document types excluded were newspapers and magazines, financial reports, operational data, annual reviews, program descriptions not reporting original research, and literature reviews and other writings on foyers that do not contain original research. Only two of the included studies were peer-reviewed.[3] In emergent debates on the assessment of grey literature, the lack of peer-review processes has been identified as indicative of relatively poor rigour. However, critics point out that publication status is not a proxy for methodological quality, as biases in the peer-review journal process also occur (Bellefontaine and Lee 2014; Grayson 2002; Hopewell et al. 2005; Wilson 2009).

Following the search described above, we have employed a three stage process: the screening of studies and the selection of eligible studies for inclusion in this review; data extraction and quality appraisal of the selected studies; and coding and analysis of these studies.

First stage: Study selection

The online database search yielded over 300 papers that were screened for titles and abstracts by one reviewer to assess their eligibility for inclusion. Sixty five full texts of potentially eligible studies were obtained, screened by the same reviewer, and all uncertainties were resolved in discussion with a second reviewer. Of these, 25 studies met the inclusion criteria (as explained in the previous section). It was difficult to track down some studies that were no longer in print or for sale. We approached authors directly and used our networks to obtain some of them. We could only find 18 full texts for extraction.

Second stage: Data extraction and quality appraisal

Data were extracted by two reviewers according to variables pertaining to paper characteristics, program characteristics, sample characteristics at baseline, study background, methodology, results, conclusions, qualitative and quantitative quality appraisal, and mixed methods quality appraisal. For the full list of variables extracted see Table 1. All disagreements during the double extraction stage were resolved by consensus.

Three studies that were initially thought to meet the inclusion criteria were excluded after careful assessment, leaving a final list of 15 studies for review. One such example was a study on resilience which was carried out in five foyers across England (Carlin 2010). While this study set out to evaluate the effectiveness of foyers in enhancing resilience for their clients, it did not present conclusions about the level of program effectiveness but instead concluded with policy recommendations for practitioners.

---

[3]   Taylor et al. (2006) was published in a peer-reviewed journal, and Ralphs (2004) is a PhD thesis that in order to be conferred would have gone through a peer-review (examination) process.

Third stage: Coding and analysis

Following the extraction stage, a smaller number of codes were developed from the extraction categories in order to assess the remaining 15 studies. During the coding stage we decided against analysing the data according to themes around the quality of features within youth foyers (e.g. the quality of accommodation, employment, education, life skills, case management etc), which seems to be the prevailing approach in evidence reviews (e.g. Livingstone et al. 2014; Morris and Jamieson 2012).

Because most studies had not been through a peer-review process (or at least did not report they had), we had to assess their internal validity and their quality before the assessment of program effectiveness (which turned out to be beyond the scope of this paper). Therefore, the team analysed data according to two groups of codes, both relating to the quality of evidence of effectiveness (Rogers 2010): the difficulty to validate claims about the program's effectiveness (which relates to internal constraints); and the limitations of studies' research design and methodology (which relates to external constraints).

The first groups of codes included analysis of studies' claims regarding a program's 'effectiveness', 'effective' outcomes or 'effective' impact. With these codes we determined whether studies provided enough information for reviewers to assess whether claims about the program's effectiveness could be substantiated. We did not aim to assess the programs' effectiveness and impact on young people's lives. The second group of codes were used to analyse the quality of the studies by an assessment of the limitations of study design and methodology. The findings are presented below according to these two groups of codes.

We assessed studies using two quality appraisal tools that were adapted specifically for the purpose of this study. The framework suggested by Popay et al. (1998) was adapted and used for assessing qualitative studies. The Effective Public Health Practice Project's (EPHPP 2009) Quality Assessment Tool for Quantitative Studies was used for assessing quantitative studies. Mixed studies were assessed using these two instruments and an additional item on the integration of qualitative and quantitative data. For a description of these measures see Table 2. Only 5 of the 15 studies met most of the quality appraisal criteria.

The fifteen studies in the coding stage comprised one journal article, thirteen evaluation/research reports, and one PhD thesis. Ten were conducted in the UK (including one in Northern Ireland), three in Australia, one both in Australia and the UK and one in the US (we did not find studies from New Zealand that met the inclusion criteria). Of the fifteen studies, nine were published between 2005 and 2015, four were published between 1995 and 2004, and two had no publication date. Table 3 summarises some characteristics of the studies.

The studies investigated various types of foyer programs. The majority (12 of 15) evaluated the foyer model in one site or a number of sites, assessing different components (e.g. accommodation, employment, education, and/or cost analysis). Two studies focused on mental health programs within foyers, and one study explored discourses of risk and youth in one foyer. Table 4 presents a summary of the coding and analysis stage. The codes are further discussed with some examples below.

Validation of claims about the effectiveness of programs

The first stage of assessment of the foyer evidence base examined whether claims about the effectiveness of programs' outcomes could be substantiated by the evidence presented in the studies. This issue relates to internal constraints (e.g. rigorous research practices). We found that four studies claimed the program they evaluated had been effective, eight studies claimed that the program they evaluated had been somewhat effective and three studies claimed the program they evaluated had not been effective.

There were three critical issues that emerged during this stage, the first relating to terminology and the other two to the quality of the reporting. First, we found that almost a third of the studies (4 of 15) did not differentiate between measuring outcomes and outputs in their evaluation. While these studies claimed foyers were effective or somewhat effective, we could not substantiate any claims based on measuring outputs only, as outputs do not signify long-term or sustained impact. Second, we found that in a third of the studies (5 of 15), programs were not well-documented (to varying degrees) and it was impossible to identify the program logic, its key mechanisms, and how the program aimed to achieve its anticipated outcomes. Therefore, we could not validate claims about the program's effectiveness. Third, we found that in the majority of studies (9 of 15), research methods were not well-documented (again to varying degrees and irrespective of methods). This made it difficult to ascertain how the program was evaluated. Therefore, we could not validate claims of effective outcomes. Importantly, we do not argue that these programs had *not* been effective, only that we, as reviewers, could not validate the programs' effectiveness with the information available in these studies.

Studies did not differentiate between outcomes and outputs

The distinction between outputs and outcomes is important. In the context of youth foyers, Steen and Mackenzie (2013, 35) explain that outputs include quantitative data such as number of participants housed, training sessions provided, turnover of clients and percentage of ex-residents in sustainable housing over the long-term, as well as the number of unplanned exits. They are direct products of the program's activities, and provide evidence that the program was actually implemented. Outcomes relate to changes in behaviours, skills, knowledge and other attributes or level of functioning as a result of the program, and require the use of metrics and measurement techniques. In the context of youth foyers these can include the number of young people who have re-engaged with education or secured employment. In the field of program theory, there is a tendency to use these two terms interchangeably, although usually outputs are defined as completed activities or tangible products or services, whereas outcomes are defined as results in relation to benefits to the target group (and this can vary between organisations and countries) (Funnell and Rogers 2011). In order to substantiate claims of effectiveness, studies need to demonstrate that there has been a change over time and that outcomes – and not only outputs – have been achieved.

While four studies seem to have conflated outputs and outcomes, using them interchangeably, one study in particular provided a clear example of using 'outcomes' to refer also to 'outputs'. EJD Consulting and Associates (2013, 33–4) evaluated the South West Sydney Youth Hub Project (formerly the Miller Live 'N' Learn Campus). The seven objectives of the project addressed issues around the engagement and participation of young people, the delivery of services, the provision of transition pathways, and the identification of barriers (2013, 19). Findings for the first part of the study were presented with housing services and client outcomes, detailing outcomes that

imply a long-term change and address the objectives of the project (e.g. the number of people who obtained long-term accommodation), and also outputs that do not imply such a change (e.g. number of people who received assistance to look for long-term accommodation). This study was relatively robust with a well-documented program, but it failed to distinguish between long-term results of the program that signify change and short-term results that measure the number of people who received a particular service. This is problematic for evaluations of programs' effectiveness, because outputs do not indicate a long-term change. Therefore, claims of effectiveness for the long-term could not be substantiated by this study and others which similarly conflate outputs and outcomes (e.g. Anderson and Quilgars 1995; Grace et al. 2011). Most studies, however, did not conflate outcomes and outputs in their findings descriptions.

Studies did not document programs

In order to assess a program's effectiveness, it is important to understand the relationships between the services delivered and mechanisms that may have contributed to program outputs and outcomes (Funnell and Rogers 2011). Important details about the program include its aims, duration, how many people participated and who was eligible to participate, the service model, and the services provided. These details allow the reviewer to identify how the program aims to achieve effective outcomes or impact. Without clear program documentation it is difficult to validate any claims about its effectiveness. As Pawson and Tilley (1997, 216) explain:

> Evaluators need to focus on how the causal mechanisms which generate social and behavioural problems are removed or countered through the alternative causal mechanisms introduced in a social program.

Anderson and Quilgars (1995), for instance, identified in their evaluation of the Foyer Pilot Project a lack of detailed operational objectives of the program. Without a benchmark or goals against which to measure the project's performance, it was not possible for these researchers to assess whether the project was more or less successful than anticipated (Berry et al. 2003, 86). In this case, therefore, the program lacked a proper documentation, which affected the ability of Anderson and Quilgars to evaluate it. By closely examining programs' documentations, the reviewer can detect any issues with measures of performance. It is also important to identify mechanisms given the diverse operational contexts of youth foyers, as these can vary between sites (Lovatt and Whitehead 2006). A good example of program documentation is the study of Randolph and Wood (2005), which clarified the program objectives, identified key outcome areas and subsequently developed a number of key performance indicators. This documentation of the program logic and mechanism allowed us to validate claims this research had made regarding the effective outcomes of this program.

Studies did not document methods

In order to assess a research claim about a program's effectiveness it is critical for a reviewer to understand the research design and data collection methods employed for the evaluation of the program. When methods are not clearly explained it is difficult to assess any claims of effectiveness. An example of such a study is the evaluation report of the Engage program (no date), which summarised the development, implementation and outcomes of the Engage program in Northern Ireland. The report discussed staff training and a client profile, including quotes from participants and four case studies of young participants. It lacked a discussion of methods and rationale for research design

that would communicate what the researchers aimed to achieve, how they collected the data, and how they evaluated the program. It is not clear why the researchers chose to present four individual case studies and how they obtained their data. Yet, the final conclusions section (Engage Programme no date, 17) stated:

> In providing targeted bespoke training, education and support at the point of first contact, the program provides real opportunities for an early intervention that can improve [participants'] life chances, reduce dependency and create independent and confident citizens.

The conclusions made by the report are wide-ranging, referring to major aspects in the life of residents without discussing the specific changes that they have experienced. This statement could not be validated based on the report itself, which crucially lacked a documented rationale of the study and the methods of investigation. Another example is the study of the Foyer Health Programme 2007–10 (no date), which examined a program designed to increase the skills, resources and opportunities of disadvantaged young people regarding healthy eating and healthy living. The program was delivered in 105 foyers in the UK, but it is not clear from the study how many foyers were included in it. The study design was a cross-sectional collection at post-intervention (without pre-intervention or a comparison group) and without an appropriate statistical comparison. Among the 15 studies we found a tendency (e.g. Engage Programme no date; Foyer Health Programme no date; Roncile 2013) to draw conclusions which we could not verify because methods were not well documented. Some studies, though, provided a good documentation of methods and therefore we could validate claims they made about programs' effective outcomes.

Limitations of research design and methodology

The second stage of assessing the foyer evidence base quality was to examine whether studies have used a robust methodology and research design to substantiate their claims about effective outcomes. This issue relates to external constraints (e.g. funding), which often influence research design and methodology. We found that none of the 15 studies included a comparison group. The majority of studies (10 of 15) did not have a post-intervention follow-up and only presented findings from one time point of data collection (sometimes retrospectively). We also found that half the studies (7 of 15) identified methodological limitations in their studies that had varying implications for the validity of their own findings.

Studies did not have a comparison group

While we acknowledge that drawing a causal inference from a comparison group is much more difficult than from a control group, the use of a comparison group in an evaluation study should be considered if the study is unable to include a control group (Grace et al. 2011, 4). A comparison group of young people who are homeless or at a risk of becoming homeless but do not live at a foyer would provide researchers with a greater understanding of the mechanisms of the program components that are associated with various outcomes. We are then better able to distinguish between the possible impacts of external factors (e.g. the passing of time, policy changes, or increase in funding) compared to factors pertaining to the youth foyer intervention itself. The benefit of using a comparison group over a control group is that ethical issues are

diminished, as the comparison group is not being denied participation in the program. Moreover, the cost of using a comparison group in a study is significantly lower than that of a control group because it is possible to use a pre-determined group that has not been randomly assigned.

The pitfall of using a comparison group is the difficulty of selecting a relevant group. In the case of the evaluation of foyers this is due to the complexity of factors such as the variety of models and forms of delivery within youth foyers. For most research, however, it is difficult to undertake a study with a comparison group due to the significant investment needed to recruit a sizeable comparison group that allows for conclusions based on statistically meaningful differences in outcomes.

Studies did not have a post-intervention follow-up

Including in the research design a set of data collection time points with a post-intervention follow-up would determine the degree of impact of the foyer, and whether outcomes are sustained or only temporary. There is a need to evaluate outcomes in the long-term, after residents leave the foyers, to see whether positive outcomes have been sustained (Barker et al. 2012; Gronda et al. 2011; Steen and Mackenzie 2013). We acknowledge that it may be extremely difficult to undertake a follow-up study with ex-foyer residents. As a number of studies explain (e.g. Common Ground Community and Good Shepherd Services 2009; Commonwealth of Australia 2008; Grace et al. 2011; Worley and Smith 2001), successfully contacting ex-residents after they have left the foyer is very challenging, mainly due to the high mobility of some people experiencing homelessness. This acts as a barrier to recruiting and maintaining contact with research participants (Commonwealth of Australia 2008, 60). Foyers are not always capable of keeping up-to-date records of their former residents, and some residents do not wish to be contacted after leaving the foyer, for different reasons. Moreover, it is difficult to attain funding for longitudinal research that lasts beyond the lifespan of the program. Thus, many studies did not engage in such a complex research design.

One example is study of the West London YMCA Foyer (Worley and Smith 2001), in which the researchers were 'interested in establishing a qualitative method for the understanding of successful outcomes' (p. 12) and wanted to use methods that could be used with interviewees who spoke English as a first language and those who did not. They therefore conducted in-depth narrative interviews with 40 former Foyer residents and also used a pictorial tool, the 'blob tree'. The researchers met residents only once, after they had left the Foyer, with time out of Foyer varying between two months to four years. They used the blob tree as a tool to ask interviewees to retrospectively represent their changing place in society when they entered the Foyer, during their time at the Foyer, when they moved out of the Foyer, and at the time of the interview (Worley and Smith 2001).

The research presented many interesting and nuanced findings about residents' experiences and achievements during their time at the Foyer and since leaving it. However, it did not have a research design that included more than one data collection time point. Since interviewees self-reported their outcomes at only one time point (after leaving the Foyer), and the study did not examine a comparison group that did not live at the Foyer, it was difficult to attribute outcomes solely to their experiences within the Foyer. For some residents, data was collected only two months after they had left the Foyer; therefore it is not known whether for them some outcomes were sustained or only short-term.

There have been some examples of longitudinal studies with people who are or were homeless. These include the Australian Journeys Home study, which has tracked for

over two and a half years a national sample of 2719 individuals exposed to high levels of housing insecurity (Scutella et al. 2015, 9-11), the Journeys to Social Inclusion study in Melbourne (J2SI, Johnson et al. 2015), as well as other US longitudinal studies. The response rate for the Journeys Home study in Wave 1 was 62 percent (n=1682), and by the sixth wave, two and a half years later, the team interviewed almost 84 percent of the initial responding sample (n=1406). Similarly, the Journey to Social Inclusion followed 88 people over a period of 48 months. After 48 months, with the end of the J2SI program, the retention rate had dropped to 66%, although it had remained at or above 80% over the 36 preceding months. Low retention rates can systematically distort the findings because people who remain engaged with longitudinal studies are often quite different from those who drop out of the project (Johnson et al. 2015, 8-9). These examples show that when researchers have the resources, it is possible to follow up the majority of people after one, two or three years.

Studies had methodological limitations (self-reported by authors)

A number of authors reported limitations within their studies' methodology and research design. For example, in the study of the Melbourne City Mission Step Ahead program (Grace et al. 2011), which focused on the experiences of young people and their views about different aspects of the model and what made a difference to their lives, the intention of the research was to collect richly subjective data with an emphasis on a narrative approach, encouraging participants to tell stories about their time with Step Ahead and their lives since. The research used a mixed methods approach, including in-depth interviews with participants, a personally administrated survey, and a review of participants' case notes. The researchers aimed to contact all 63 previous clients of the program but could only contact 47, of whom only 29 agreed to participate. One of the limitations the authors noted is that they could not assert that this sample is representative, and that there may be differences between those who were contactable and those who were not, and between those who agreed to participate and those who did not (Grace et al. 2011, 18–9).

One attribute of a good qualitative study is its potential for assessing typicality, i.e. the claims that are being made for the generalisability of the findings to either other bodies of knowledge or other populations or groups (Popay et al. 1998, see Table 2). Using a reflexive form of reporting, Grace and colleagues (2011) discussed the potential for typicality and allowed the reviewer to evaluate its limitations. Despite these limitations the research is rigorous, as it met all the qualitative quality appraisal criteria and some of the quantitative criteria we applied during the extraction stage. Although the study stated that 'the evidence suggests that Step Ahead was effective in assisting young people to improve relationships with family' (2011, 5), it did not claim effectiveness in all areas of the program (e.g. housing, education and employment) and recognised the limitations of the program alongside its achievements.

This study, together with a few other studies (e.g. Common Ground Community and Good Shepherd Services 2009; Ralphs 2004; Randolph and Wood 2005; Steen and Mackenzie 2013; Taylor et al. 2006), is rigorous, its methods are robust, and there is enough detail on the research design to assess the level of creditability and to validate the findings and conclusions. Yet the researchers themselves recognised some limitations with their study design that prevented them from drawing unconditional conclusions regarding the effectiveness of the Step Ahead program. Some other studies, but not all, have recognised limitations to their evaluations in varying degrees of significance.

The arguments presented above examine the quality of evidence about effectiveness

(Rogers 2010) using two aspects: validation of claims studies make regarding program effectiveness, and limitations in studies' research design and methodology. In the concluding section we discuss why it was important to assess these two aspects, potential policy implications, and recommendations for a research agenda.

Strengthening the quality of the youth foyer evidence base

This review originally aimed to examine the evidence on the effectiveness of the youth foyer model for homeless young people. However, following our initial scan of the literature it was clear that we could not complete this task due to the lack of basic detail in the research reports identified for review. Instead we were compelled to examine the quality of the evidence base itself.

Many of the studies we examined drew encouraging conclusions regarding the effectiveness of youth foyers based on findings that indicated a successful implementation of the program and some tangible results. Of the 15 studies, 12 claimed effective or somewhat effective results. They did not, however, confirm long-term substantiated outcomes that signify change. For example, the study of EDJ Consulting and Associates (2013, 31) presented promising findings regarding positive housing outcomes, but these were short-term (the average tenancy stay was 28 weeks) and the study did not report on sustained longer-term outcomes. We were also often unable to validate claims regarding effective outcomes of youth foyers due to lack of documentation of programs and methods of evaluation. When there is a lack in program documentation it is difficult for the reviewer to assess the mechanism of the program, the program logic and how exactly the program has met its aims and objectives. In some cases, the lack of methods documentation prevented the reviewer from assessing how data were gathered and analysed. Of the studies we examined, some did not differentiate between outputs and outcomes, some did not document methods and programs, and others did not report on their research design and methodological limitations. This is certainly not unusual in research on homelessness services, particularly given the lack of investment in adequate, rigorous evaluation.

Our review also found that most studies conducted only one round of data collection, with few conducting a post-intervention follow-up. Generally, studies did not use benchmark comparisons to provide relational context to findings. The examination of studies thus revealed a disparity between the level of rigour and the conclusions that were drawn. Moreover, the vast majority of the 15 studies had not been peer-reviewed by external reviewers, and studies scored lowly on the quality appraisal criteria we applied in our examination. Both these issues indicate a lack of rigorous research practices.

The majority of the studies we reviewed were evaluation reports, often written by agencies delivering youth foyer programs or commissioned by these agencies. This is not surprising or unusual. Of the homelessness service evaluations that are publicly available, many are undertaken or commissioned by service agencies, often for advocacy purposes. These studies, including the foyer evaluations, can be classified as grey literature (reports that are not commercial and scholarly peer-reviewed publications).

Grey literature is one of the most important sources of information used by public policy makers and practitioners, as it is often readily accessible and relevant (Lawrence et al. 2014). However, in the field of youth foyer research, the over-reliance on grey literature as an evidence base signals a lack of investment in rigorous, peer reviewed

research reporting. This is not to suggest that the gold standard of research design is the only avenue to achieving valid and reliable reporting (Rogers 2010). We examined the quality of evidence about effectiveness through two interrelated aspects: validation of claims studies made regarding program effectiveness, and limitations of the studies' research design and methodology.

From this review, there are three key implications for policy development and future research in Australia to enhance the quality of the youth foyer evidence base. First, it is imperative to improve the rigour of studies and to lift the standard of evaluations. We acknowledge that the lack of high quality evaluations indicates inadequate funding and resources that otherwise may have enabled more rigorous research. Although we welcome recent large investments in youth foyers in Australia, this must be accompanied by a parallel investment in research and evaluation studies of high standard.

Second, to lift the standard of studies, ground rules need to be implemented and embraced by agencies and research bodies producing evaluation studies of youth foyers. These should include a system for ensuring an adequate research design, appropriate documentation of methods and programs, use of a theoretical framework for the interpretation of findings, and putting in place a peer-review process. Public sector commissioners of youth foyers should play an important role in tightening the quality of evaluations and research components when funding new foyers, recognising the critical relationship between service development and research. Although implementation of the gold standard design is not a realistic expectation in the homelessness research field, other methods of measurement can be employed. As the emerging debate around the quality and standard of grey literature shows, there have been some efforts to create review codes and checklists to assist in the evaluation of grey literature (see Lawrence (2014) and the University of Melbourne Library website (2015)). For example, the Grey Literature Strategies ARC Linkage project suggests that there are alternative methods to peer-review that can be used to scrutinise grey literature, such as review by an expert panel or board, internal review, and post publication review (Lawrence 2014). Another example is the Australian Housing and Urban Research Institute's (AHURI) system for peer-reviewing, in which every Final Report and Positioning Paper published after November 2007 is a peer-reviewed publication (AHURI 2015).[4] Although such a system admittedly would have high infrastructure costs, it can be similarly used in the case of youth foyer research and other service systems. This recommendation does not refer only to the small area of inquiry of homelessness research. Rather, ground rules for grey literature should be implemented, funded and monitored across all fields of social policy. One of the side effects of lack of investment is the lack of time for adequate research processes. Drawing on peer review expertise would no doubt have a positive impact on strengthening the quality of evidence.

However, although the above mentioned codes and checklists for grey literature have been developed, they seem most relevant for biomedical and public health research that upholds the gold standard (Barker et al. 2012; Schorr and Farrow 2011). This benchmark then continues to limit valid criteria for the appraisal of grey literature. In the social sciences, there is a need to develop appropriate codes for the assessment of grey literature which takes into account the specific constraints of the field. Moreover, there is a tendency for researchers to include significant grey literature in their reviews, but then exclude it due to its low quality research. Thus, there needs to be a framework to guide assessment of grey literature (Tyndall 2008).

---

[4]   A peer-reviewed AHURI report can be identified by an icon on its front cover.

Third, the lack of rigour in the youth foyer studies reviewed here also points to some service development gaps that are then mirrored in the corresponding research. The foyer studies examined often presented incomplete reporting program documentation. This may be because the service delivery tools such as program logics and theory of change frameworks do not exist for the programs examined, or the researchers did not use or have access to these documents. Consequently, the links between program outcomes and mechanisms cannot be established in the existing evidence. This has important implications for not only the level of research rigour in homelessness research, but also for hampering potential service development improvements based upon evidence informed research.

To conclude, this compilation of evaluation studies regarding the youth foyer model's effectiveness provides some recommendations for how this evidence base might be improved, especially with regard to studies' internal and external validity. Secondly, it highlights the need for more investment in rigorous research that accompanies investment in innovation in the field of homelessness. With appropriate funding and clear quality appraisal criteria for 'non gold-standard' research design, we can then begin to strengthen the evidence base.

References

ABS 2012a. 4922.0. *Information paper - a statistical definition of homelessness*, Australian Bureau of Statistics, Canberra, viewed 2 December 2015, <www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/B4B1A5BC17CEDBC9CA257A6E00186823/$File/49220_2012.pdf>.

ABS 2012b. *2049.0 - Census of Population and Housing: Estimating Homelessness, 2011*, Australian Bureau of Statistics, Canberra, viewed 7 August 2015, <www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/2049.0Main%20Features22011?opendocument&tabname=Summary&prodno=2049.0&issue=2011&num=&view=>.

AHURI 2015. Editorial Board Members, Australian Housing Research Institute, Melbourne, viewed 12 August 2015, <www.ahuri.edu.au/about/people/editorialboard>.

AIHW 2012. *Australia's Health 2012*, Australia's Health Series no.13. Cat. no. AUS 156, Australian Institute of Health and Welfare.

Anderson, I and Quilgars, D 1995. *Foyers for Young People: Evaluation of a Pilot Initiative*, Centre for Housing Policy, University of York.

Barker, J, Humphries, P, McArthur, M and Thomson, L 2012. *Literature Review: Effective Interventions for Working with Young People who are Homeless or are at Risk of Homelessness*, Australian Government Department of Families, Housing, Community Services and Indigenous Affairs, Canberra.

Bellefontaine, SP and Lee, CM 2014. Between black and white: Examining grey literature in meta-analyses of psychological research, *Journal of Child and Family Studies*, 23: 1378–88.

Berry, M, Chamberlain, C, Dalton, T, Horn, M and Berman, G 2003. *Counting the Cost of Homelessness: A Systematic Review of Cost Effectiveness and Cost Benefit Studies of Homelessness*, Prepared for the Commonwealth National Homelessness Strategy, AHURI, Melbourne.

Carlin, E 2010. *Feeling Good: Supporting Resilience in Young People in Foyers in England*, The Foyer Federation.

Common Ground Community and Good Shepherd Services 2009. *The Chelsea Foyer at the Christopher at Five Years: Lessons in Developing Stable Housing and Self-Sufficiency for Homeless Youth and Youth Exiting Foster Care*, Funded by the Annie E. Casey Foundation.

Commonwealth of Australia 2008. *The Road Home: A National Approach to Reducing Homelessness*, Homelessness Taskforce, Canberra.

EJD Consulting and Associates 2013. *South West Sydney Youth Hub Project: Incorporating the Foyer Model*, Department of Family and Community Services and Housing NSW, Sydney.

Engage Programme no date. *Evaluation of Engage Programme*, Engage.

EPHPP 2009. Quality Assessment Tool for Quantitative Studies, *Effective Public Health Practice Project*, viewed 13 August 2015, <www.ephpp.ca/tools.html>.

Fopp, R 2015. 'Problematising Aspects of Evidence-Based Policy: An Analysis Illustrated by an Australian Homelessness Policy 1985-2008', *Australian Journal of Social Issues*, 50(3): 297-317.

Foyer Foundation 2015. *Foyer Foundation*, viewed 7 June 2015, <foyer.org.au/>.

Foyer Health Programme no date. *Evaluation Report on the Foyer Health Programme 2007–2010*.

Funnell, SC and Rogers, PJ 2011. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*, Jossey-Bass, San Francisco.

Grace, M, Keys, D, Hart, A and Keys, B 2011. *Achieving (Extra)ordinary Aspirations: A Research Project Exploring the Role that Step Ahead Program Has Played in the Lives of Young People Affected by Homelessness*, Melbourne City Mission and Victoria University, Melbourne.

Grayson, L 2002. *Evidence Based Policy and the Quality of Evidence: Rethinking Peer Review*, ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary, University of London.

Gronda, H 2009. *Evidence to Inform NSW Homelessness Action Priorities 2009–2010*, AHURI.

Gronda, H, Ware, V-A and Vitis, L 2011. *What Makes a Difference? Building a Foundation for Nationally Consistent Outcome Measures*. National Homelessness Research Agenda 2009–2013 Australian Housing Research Institute Melbourne.

Hanover, 2012. Youth Foyer Project – Information Sheet, viewed 3 December 2015, <hanover.org.au/wp-content/uploads/2012/07/foyer_infosheet-final_26042012.pdf>.

Hwang, SW, Tolomiczenko, G, Kouyoumdjian, FG & Garner, RE 2005. 'Interventions to Improve the Health of the Homeless: A Systematic Review', *American Journal of Preventive Medicine*, 29(4): 311-9.

Hopewell, S, Clarke, M and Mallett, S 2005. Grey literature and systematic reviews, in HR Rothstein, AJ Sutton and M Bornstein (eds), *Publication Bias in Meta-Analysis Prevention, Assessment and Adjustments*, John Wiley and Sons LTD, Chichester.

Johnson, G, Daniel Kuehnle, Parkinson, S, Sesa, S & Tseng, Y-P 2015. *Sustaining Exits from Long-Term Homelessness: A Randomised Controlled Trial Examining the 48 Month Social and Economic Outcomes from the Journey to Social Inclusion Pilot Program*, RMIT University, University of Melbourne, Sacred Heart Mission, Melbourne.

Kaplan, J 2014. Cost Effectiveness Analysis in *Better Evaluation*, viewed 28 January 2016, <betterevaluation.org/evaluation-options/CostEffectivenessAnalysis>.

Lawrence, A 2014. *Grey Literature Review Code,* Australian Policy Online, Melbourne, viewed 7 August 2015, <apo.org.au/tools/review-code>.

Lawrence, A, Houghton, J, Thomas, J and Weldon, P 2014. *Where is the Evidence? Realising the Value of Grey Literature for Public Policy and Practice*, Swinburne Institute for Social Research, Melbourne, <apo.org.au/research/where-evidence-realising-value-grey-literature-public-policy-and-practice>.

Leigh, A 2009. What evidence should social policymakers use?, *Economic Roundup*, 9(1): 27–43.

Livingstone, C, Rintoul, A and Francis, L 2014. What is the evidence for harm minimisation measures in gambling venues?, *Evidence Base*, 2: 1-24.

Lovatt, R and Whitehead, C 2006. Launch Pad for Life: An Assessment of the Role of Foyers in Housing Association Provision, Report commissioned by the Housing Corporation, University of Cambridge.

Mallett, S, Rosenthal, D, Keys, D and Averill, R 2010. *Moving Out, Moving On: Young People's Pathways in and through Homelessness*, Taylor and Francis, London.

Mallett, S, James, S, McTiernan, N and Buick, J 2014. *Education First Youth Foyer Practice Framework*, Hanover Welfare Services and Brotherhood of St Laurence, Melbourne.

Morris, A and Jamieson, M, Patulny, R 2012. Is social mixing of tenures a solution for public housing estates?, *Evidence Base* 1: 1–21.

Parsell, C, Fitzpatrick, S & Busch-Geertsema, V 2014. 'Common Ground in Australia: An Object Lesson in Evidence Hierarchies and Policy Transfer ', *Housing Studies*, 29(1): 69-87.

Pawson, R and Tilley, N 1997. *Realistic Evaluation*, SAGE, London.

Pleace, N 2013. *Evaluating Homelessness Services and Strategies: A Review*, HABITACT, viewed 7 December 2015, <www.habitact.eu/files/activity/actionresearch/_evaluationreview_habitact.pdf>.

Popay, J, Rogers, A and Williams, G 1998. Rationale and standards for the systematic review of qualitative literature in health services research, *Qualitative Health Research*, 8(3): 341-51.

Quilgars, D 2001. Dispersed Foyers: *A New Approach? An Evaluation of the Shortlife Plus Project*, Centre for Housing Policy, University of York.

Quilgars, D, Johnsen, S and Pleace, N 2008. *Youth Homelessness in the UK: A Decade of Progress?*, Joseph Rowntree Foundation, York.

Ralphs, R 2004. *Beyond Transitions: Problematizing the Experience of Young People in Contemporary Society*, PhD Thesis, Department of Sociology, The Manchester Metropolitan University, Manchester.

Randolph, B and Wood, H 2005. *An Interim Evaluation of the Miller Live 'N' Learn Campus*, Australian Housing and Urban Research Institute, Sydney.

Rogers, PJ 2010. Learning from the evidence about evidence-based policy, in Productivity Commission

(ed.), *Strengthening Evidence Based Policy in the Australian Federation, Volume 1: Proceedings*, Productivity Commission, Roundtable Proceedings, Canberra.

Ronicle, J 2013. *Evaluation of Connect Yourself: Final Report*, York Consulting LLP, Leeds.

Schorr, BL and Farrow, F 2011. *Expanding the Evidence Universe: Doing Better by Knowing More*, Center for the Study of Social Policy, Washington.

Scutella, R, Bevitt, A, Chigavazira, A, Herault, N, Johnson, G, Moschion, J, Tseng, Y-P, Wooden, M & Kalb, G 2015. *Journeys Home Research Report No. 6: Complete Findings from Waves 1 to 6*, Melbourne Institute, Melbourne.

Smith, J, Browne, O, Newton, V and O'Sullivan, A 2007. *What Happened Next? A Report on Ex-Residents of Foyers*, Centre for Housing and Community Research, Cities Institute, London Metropolitan University.

Steen, A and Mackenzie, D 2013. *Financial Analysis of Foyer and Foyer-like Youth Housing Models*, Swinburne University, Homelessness Research Collaboration, Melbourne.

Taylor, H, Stuttaford, M, Broad, B and Vostanis, P 2005. *Evaluation of the Strong Minded Mental Health Service for Young Homeless People in Foyers*, University of Leicester.

Taylor, H, Stuttaford, M and Vostanis, P 2006. A UK survey on how homeless shelters respond to the mental health needs of homeless young people, *Housing, Care and Support*, 9(2): 13–8.

Tyndall, J 2008. *How Low Can We Go? Towards a Hierarchy of Grey Literature*, presented to the Dreaming 08 – Australian Library and Information Association Biennial Conference, Alice Springs, NT Australia.

University of Melbourne 2015. *Grey Literature for the Health Sciences*, University Library, viewed 7 August 2015, <www.unimelb.libguides.com/content.php?pid=616476&sid=5105765>.

Wilson, DB 2009. Missing a critical piece of the pie: Simple document search strategies inadequate for systematic reviews, *Journal of Experimental Criminology,* 5: 429–40.

Worley and Smith, J 2001. *Moving Out, Moving On: From Foyer Accommodation to Independent Living*, YMCA, London.

Appendix

Table 1 List of categories in extraction spreadsheet

Group	Variable	Example
Study characteristics	Paper reference	Roncile (2013)
	Material type	Evaluation report
	Who delivered the program	Foyer Federation, Changemakers, Youth at Risk (YaR)
	Who conducted the study	York Consulting
	Who commissioned the study	Connect Yourself Partnership (Foyer Federation, Youth at Risk (YaR) and Changemakers)
Program characteristics	Program name	Connect Yourself Programme
	Program's rationale/aims/purpose	The program aimed to enable young people to connect themselves to opportunities that identify, develop and promote their talents and achieve their future goals
	Types of main services provided by program	Young people attended a four-day intensive workshop, six follow up sessions and were assigned a life coach with whom they worked with for six months. There were projects to improve young people's employability outcomes by developing their skills, assets and talents for the world of work.
	Criteria for admission into program	Mostly foyer residents (223/ 390) and the rest from various services that seem similar.
	Maximum duration of program	18 months
Sample characteristics at baseline	Total sample at different time points	390
	Gender	43% females (n=168), 53% males (n=207), 4% no response (n=16)
	Age range	17 or above (max age not specified)
	Ethnicity	Not reported
	Education levels completed at baseline	N=87 (22%) not in employment, education or training, and 29 young people from schools had previously been excluded or suspended from school (29/390 = 7.4%). In foyers: 73/223 = 32.7% not in education or employment. Excluded/suspended from schools: 12/223 = 5.4%
	Employment at baseline	N=87 (22%) not in employment, education or training. In foyers: 73/223 = 32.7% NEET
	Housing at baseline	Almost half (183, 47%) were either homeless or living in sheltered accommodation. In foyers: 154/223 = 69.1%
Study background	Study location (country)	UK
	Main research questions or aims	How has the programme impacted on young people's emotional characteristics and behaviour?
	Has the study been set in a theoretical context/literature? If yes, elaborate	No, only brief intro to the program

Study methodology	Research design and data collection	Cohort pre-post
	Does the study have a rationale for the design? If yes, elaborate	Not reported
	Years of data collection	2011–2013
	Does the study use a sampling procedure? If yes, detail	No, aims to follow a cohort
	Methods used	Case study visits to all four areas, interviewing project staff, partner agencies and community members, case study young people interviews, focus groups
	Response rates at each time point	100% at baseline, 21% at exit
	Method/s of data analysis	Descriptive stats
	Methodological limitations stated in study	Difficulties in gathering completion data. Also sample is not large enough to be representative or statistically significant and may therefore not reflect the impact across the whole cohort
	Additional methodological limitations (not stated)	Lack of a clear methods section. Lack of qualitative analysis to identify nuances. No precise data of anything, only baseline.
Study results – outcomes/outputs	Housing quantity results	No data except 'Young people interviewed ... moved into more secure accommodation'
	Employment quantity results	'Many young people interviewed reported how the programme had either helped them progress into Employment, Education or Training (EET), or at least move closer towards it'.
	Education quantity results	'Many young people interviewed reported how the programme had either helped them progress into Employment, Education or Training (EET), or at least move closer towards it'.
	Other quantity results	Reduction in substance abuse, and equipment of new skills
Study conclusions	Main conclusions about program effectiveness	'The programme had a positive impact on some, though not all, young people involved' (p. ii)
	Main policy implications relating to program effectiveness	The 3 delivery strands need to be more connected, the hub approach is the optimal model, and the order of delivery of the strands should have been different.
	Write your own critical assessment of findings/conclusions/claims as made in the study and their significance	Seems adequate despite major problems with methods. Interpretation seems critical enough as it recognises the study's limitations.

Table 2 Qualitative and quantitative quality appraisal criteria

Qualitative quality appraisal (Popay et al. 1998)

1. Does the research illuminate subjective meaning, contexts and actions of those being investigated?
   Does the report show how behaviours are understood from within the culture, social setting or group being studied?
   Do the researchers find ways of giving lay knowledge equal worth to other forms of knowing?
   Does the research study how people act, employ knowledge and experience and understand the phenomenon of interest?
2. Is there evidence of responsiveness to the social context and flexibility of design?
   - Is there evidence that the research design was flexible and the researcher was responsive and adapted to the social circumstances of the study (as needed)?
3. Evidence of adequate description - this is the richness of the findings and interpretation of data that the research produces
   - Is the description provided detailed enough to allow the researcher/reader to interpret the meaning and context of what is being described?
   - To what extent does the text show evidence of or support the description or interpretation the researcher provides?
4. Evidence of data quality
   - How are different sources of knowledge about the same issues compared and contrasted (triangulation)? It is not expected that accounts will coincide, but that multiple data sources will help illuminate different facets of the reality under investigation.
   - Are participants' subjective experiences and local knowledge treated as knowledge in its own right?
   - To what extent has the researcher rendered transparent the processes by which data were collected, analysed and presented (audit trail)? This includes describing the time, extent and natures of the researcher's involvement in the social situation (reflexivity).
5. Evidence of theoretical and conceptual adequacy
   - Interpretive validity - how does the researcher move from a description of the data to quotation of examples, to analysis and interpretation of the meaning of data? What kind of method is used (e.g. constant comparative method)?
6. Potential for assessing typicality
   - What claims are being made for the generalisability of the findings to either other bodies of knowledge or to other populations or groups?
   - Reports should provide enough background information to make judgments of typicality with regard to the features of a case or a sample

Quantitative quality appraisal (EPHPP 2010)

1. Are the individuals selected to participate in the study likely to be representative of the target population? Elaborate and answer also:
   1. Very likely; 2. Somewhat likely; 3. Not likely; 4. Can't tell
2. What percentage of selected individuals agreed to participate? Elaborate and answer also:
   1. 80–100% agreement; 2. 60–79% agreement; 3. Less than 60% agreement; 4. Not applicable; 5. Can't tell
3. Indicate the study design: Elaborate and answer also:
   1. Randomized controlled trial; 2. Controlled clinical trial; 3. Cohort analytic (two group pre + post); 4. Case-control; 5. Cohort (one group pre + post (before and after)); 6. Interrupted time series; 7. Other specify; 8. Can't tell
4. Were withdrawals and drop-outs reported in terms of numbers and/or reasons per group? Elaborate and answer also:
   1. Yes; 2. No; 3. Can't tell; 4. Not Applicable (i.e. one time surveys or interviews)
5. Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest). Elaborate and answer also:
   1. 80–100%; 2. 60–79%; 3. less than 60%; 4. Can't tell; 5. Not Applicable (i.e. Retrospective case–control)
6. Are the statistical methods appropriate for the study design? Elaborate and answer also:
   1. Yes; 2. No; 3. Can't tell

Mixed methods quality

Describe the integration between different methods and data types

Table 3 Characteristics of final selected studies (in chronological order within countries)

Study	Commissioned/funded by	Form of publication	Foyer(s) location	Program evaluated	Study methods
Studies from the UK					
Ronicle (2013)	Connect Yourself Partnership (Foyer Federation, Youth at Risk (YaR) and Changemakers	Evaluation report, not peer-reviewed	UK	Connect Yourself Programme	Qualitative
Foyer Health Programme 2007–2010 (no date)	Funded by Big Lottery Fund	Evaluation report, not peer-reviewed	UK	Foyer Health Programme	Mixed methods
Smith et al. (2007)	In association with the Foyer Federation, funded by the Housing Corporation, Diageo and the Countryside Agency	Evaluation report, not peer-reviewed	UK	Ten foyers	Mixed methods
Taylor et al. (2006)	Independent research	Journal article, peer-reviewed	UK	132 foyers	Quantitative
Taylor et al. (2005)	Not reported, possibly Foyer Federation	Evaluation report, not peer-reviewed	UK	Strong Minded	Two separate studies, one quant and one qual
Ralphs (2004)	Independent research	PhD thesis, peer-reviewed	UK	A foyer in an inner-city location of a large city	Qualitative
Quilgars (2001)	Centre for Housing Policy, University of York	Evaluation report, not peer-reviewed	UK	Shortlife Plus Project	Mixed methods
Worley and Smith (2001)	Unclear from report	Evaluation report, not peer-reviewed	UK	West London YMCA Foyer	Qualitative
Anderson and Quilgars (1995)	Independent evaluation. Funded by the Joseph Rowntree Foundation	Evaluation report, not peer-reviewed	UK	Foyer Pilot Project	Mixed methods
Evaluation of Engage Programme (no date)	European Union EQUAL Programme	Evaluation report, not peer-reviewed	UK (Northern Ireland)	Engage Programme	Qualitative
Studies from Australia					
Steen and Mackenzie (2013)	Swinburne University, supported by the Australian Government	Evaluation report, not peer-reviewed	Australia and UK	Various foyers	Quantitative
EJD Consulting and Associates (2013)	Commissioned by Housing NSW, Department of Family and Community Services (FACS)	Evaluation report, not peer-reviewed	Australia	NSW Homelessness Action Plan (HAP)	Mixed methods
Grace et al. (2011)	Funded by Melbourne City Mission and Victoria University	Evaluation report, not peer-reviewed	Australia	Step Ahead program	Mixed methods
Randolph and Wood (2005)	NSW Department of Housing	Evaluation report, not peer-reviewed	Australia	The Miller Live 'N' Learn Campus	Qualitative

Study from the US					
Common Ground Community (2009)	Published by Common Ground Community and Good Shepherd Services and supported by the Annie E. Casey Foundation	Evaluation report, not peer-reviewed	US	The Chelsea Foyer at the Christopher	Quantitative

Table 4 Coding and analysis of studies (in chronological order within countries)

Study	Study claims program is effective	Study meets most quality appraisal criteria	Validation of claims about effectiveness			Limitations of research design and methodology		
			Study differentiates btwn outcomes and outputs	Study documents program	Study documents methods	Study has a comparison group	Study has a post-intervention follow-up	Methodological limitations (identified by study's author/s)
Studies from the UK								
Ronicle (2013)	Somewhat	Somewhat	Yes	Yes	No	No	Yes	Researchers were unable to report on completion rates due to difficulties in gathering completion data.
Foyer Health Programme 2007–2010 (no date)	Yes	No	Yes	No	No	No	No	Proper cost-benefit or Return on Investment calculation of the impact of the training program did not prove possible.
Smith et al. (2007)	Somewhat	Somewhat	Yes	Yes	Somewhat	No	Yes	Comparing outcomes for young people from different ethnic groups was difficult, both because of the diversity of the ethnic origin and also because of the diversity of nationality.
Taylor et al. (2006)	No	Yes	Yes	Somewhat	Yes	No	No	Respondents were not mental health practitioners and it was expected that their knowledge of mental health would vary.
Taylor et al. (2005)	Somewhat	Somewhat	Yes	Somewhat	Somewhat	No	Yes	None
Ralphs (2004)	No	Yes	Yes	Yes	Yes	No	No	There is a need to employ a longitudinal study in order to evaluate the benefits of interventions such as living in a Youth Foyer.
Quilgars (2001)	Yes	Somewhat	Yes	Yes	Somewhat	No	No	Comparing two data-sets from different periods of time (1992–1994, the pilot study (Anderson and Quilgars 1995)) and this study (1997–2000) made it a difficult task due to changes in contextual factors.
Worley and Smith (2001)	Somewhat	Yes	No	No	Somewhat	No	No	Difficulties in contacting young people after they leave the program. - Absence of a longitudinal research design.

Study							Comments
Anderson and Quilgars (1995)	Yes	Somewhat	No	Yes	Somewhat	No	No	Measurement of outcomes from the Foyer pilot proved problematic because returns were incomplete and lack of detailed operational objectives meant there was no benchmark against which to measure the project's performance.
Evaluation of Engage Programme (no date)	Somewhat	Somewhat	Yes	No	No	No	No	None
Studies from Australia							
Steen and Mackenzie (2013)	No	Yes	Yes	Yes	Yes	Not Applicable (cost-benefit analysis)	Not Applicable (cost-benefit analysis)	Agencies took considerable amount of time to eNotract detailed performance and accounting data. - Comments on performance have relied on what various foyers claimed were their outcomes and the team had no way of critically eNoamining these claims. - Insufficient data for a cost-benefit analysis.
EJD Consulting and Associates (2013)	Somewhat	No	No	Yes	Somewhat	No	No	No control group for the project. - Absence of longitudinal data. - Insufficient data for a cost-benefit analysis.
Grace et al. (2011)	Somewhat	Yes	Somewhat	Yes	Yes	No	No	Only 46% of all previous clients have agreed to participate in the research and it is not clear if this group is representative.
Randolph and Wood (2005)	Somewhat	Yes	Yes	Yes	Yes	No	Yes	Time gaps between first and second interviews were too short, and methods have been more limited than planned.
Study from the US							
Common Ground Community (2009)	Yes	Somewhat	Yes	Yes	Yes	No	Yes	Tracking participants once they leave the program has been difficult. - The program is still in process of optimising its data platform thus developing indicators to measure participants' progress is difficult.