Evaluation of home-based reablement: A systematic review

WORKING PAPER

By:

Tore Bersvendsen1*, Jochen Jungeilges1, 2 and Eirik Abildsnes3

1 Department of Economics and Finance, School of Business and Law
University of Agder, Postbox 422, 4604 Kristiansand, Norway

2 Institute of Mathematics and Natural Sciences
Ural Federal University, Mira 19, Ekaterinburg 620000, Russian Federation

3 Agency for Health Promotion and Innovation, Health and Social Sector
Kristiansand Municipality, Postbox 417, 4604 Kristiansand, Norway

* Correspondence:
Tore Bersvendsen, tore.bersvendsen@uia.no, +4790031820

E-mail addresses:
Jochen Jungeilges: jochen.jungeilges@uia.no
Eirik Abildsnes: eirik.abildsnes@kristiansand.kommune.no

Abstract

Background: Home-based reablement (HBR) aims to restore or increase a patient’s level of functioning, thereby increasing the patient’s self-reliance and consequently decreasing their dependence on healthcare services. To date, the evidence on whether HBR is an efficient method has not been comprehensively reviewed. The aim of this study was to provide a concise summary of relevant existing findings. In addition, we provide a critical constructive assessment of the publications reflecting the extant research.

Method: The relevant literature on this topic was identified through a systematic search of appropriate databases. Thereafter, we screened the studies, first by title, followed by abstract and then by assessing full-text eligibility. A checklist of 15 criteria was developed and used as the basis for the quality assessment.

Results: In total, 11 studies from Australia, New Zealand, the USA and Norway were included in the full-text review. The studies reported estimated cost differences between HBR and usual care after the intervention. All the studies indicated lower costs for HBR, but not all of them reported a significant difference. The same pattern was also found for other measures of physical functioning and quality of life. The assessment revealed one specific common pattern: None of the papers scrutinized provided sufficient information about the data or the statistics employed, and all lacked external validity.

Conclusion: Some promising results have been reported with respect to HBR reducing the need for specialist or residential care. In short, the existing evidence regarding the effects of HBR is still inconclusive. The findings from the quality assessment should motivate a multidisciplinary approach for future research on HBR.

Keywords: Reablement, economic/econometric evaluation, rehabilitation, RCT technology, assessment tool

JEL Classification: I19 C18
1. Background

The western world is facing a significant demographic change in the years to come. These forthcoming developments are expected to lead to a persisting change in the age distribution of the population. As the elderly population grows, the number of individuals facing age-related diseases and multimorbidity will increase (1). Costs of healthcare services increase with age and with the degree of multimorbidity (2).

According to Martins and de la Maisonneuve (3), the costs of long-term care for people over 65 years old are predicted to double or triple by 2050 in countries belonging to the Organization for Economic Co-operation and Development (OECD). Along with these upcoming demographic challenges, the number of participants in the workforce per senior citizen in OECD countries will decrease (4).

The increasing costs, economic and otherwise, resulting from the increased need for care will have to be shared by a decreasing proportion of employed individuals. The upcoming challenges will increase the demand for long-term services as home-based care (5). Since home-based care is more cost-effective, many high-income countries are actively bolstering a shift from residential care to home-based care as a potentially more effective and financially sustainable approach (6). Another incentive for this shift is that older people prefer to ‘age in place’ (7). The forthcoming challenges will force the healthcare industry to focus more on preventive measures, early intervention, increased use of technology, rehabilitation and healthcare services that are less manpower-intensive, and services that empower senior citizens to self-manage chronic diseases (8).

Home-based reablement (HBR), known as restorative care in Australia, the USA and New Zealand, is one fairly new way of providing homecare services. The main goal of HBR is to restore or increase a patient’s level of functioning, thereby increasing the patient’s self-reliance and consequently decreasing their dependence on healthcare services. Even though HBR is not a standardized treatment and the content of HBR varies, all such interventions intend to restore or increase the level of functioning (9, 10). This type of intervention has gained significant prominence internationally in recent years (6). HBR represents an ecological approach, taking into account patients’ preferences and resources. The main features of being time-limited, multidisciplinary, home-based, goal-oriented and person-centred are homogenous across HBR programmes. Patients are mainly senior citizens with or at risk of functional decline (11). Typically, a multidisciplinary team works towards a patient-defined goal that focuses on everyday activities important to the patient (10). When developing an HBR model, Baker, Gottschalk (12) learned that several traditional homecare routines created barriers for functional independence in older patients. They also agreed that involving the patient in goal setting was essential. One might expect that HBR may lead to increased risks of falling, injury or other adverse events. There is evidence that reablement in nursing homes does not have such effects (13). We are not aware of similar evidence concerning HBR. A Danish study concluded that policy-makers mainly motivated by economic considerations were pivotal for the implementation of HBR (14). High-quality care is clearly an essential goal in health care services, but future resources are limited, inevitably leading to priority setting and trade-offs (15). Assessing the efficiency and effects of new interventions, including HBR, is therefore crucial.

To our knowledge, there are few comprehensive and systematic overviews of research relating to the effects of HBR. Interestingly, no HBR studies were included in an overview of systematic reviews on economic evaluations of health-related rehabilitation (16). Five HBR studies were included in a systematic review identifying interventions that aimed to reduce dependency in activities of daily living (ADL) (9). The latter review had two objectives, of which the second was to determine the effect an intervention had on improving a person’s ADL. HBR could have effects on factors other then ADL. The two studies most similar to our paper are those by Tessier, Beaulieu (17) and Legg, Gladman (18), both systematic reviews from 2016. Tessier, Beaulieu (17) examined the effectiveness of HBR and factors that might contribute to successful implementation for Canadian policy makers. They focused on three outcomes, function, health-related quality of life (HRQol) and service utilization, concluding that there is good evidence supporting the effectiveness of HBR, especially regarding HRQol and service utilization. Interestingly, Legg, Gladman (18) studied whether publicly funded HBR affects patient health or use of services. They found no data suitable for evaluating the effects of HBR and concluded that there is no evidence that HBR fulfils its goals. In sum, previous reviews either focus on minor aspects of potential benefits of HBR alone or do not include studies on HBR, as such studies failed to meet the inclusion criteria defined by the respective reviewers. It is therefore the objective of this paper to provide a comprehensive and systematic review of current literature assessing HBR through empirical evaluation. First, we aim to provide a concise summary of relevant existing findings generated in the course of the research process. In addition, we provide a critical constructive assessment of the publications reflecting the extent
research. The application of statistical concepts and models plays a central role in the research efforts we analysed. Consequently, our review adopts a dual perspective: the health-economic angle is augmented by a pronounced statistical/econometric viewpoint.

The remainder of the paper is organized as follows: In Section 2, we outline the methodological basis for the systematic review. The main findings from relevant HBR research are summarized, and the results of our literature assessment are presented in Section 3. Section 4 provides a thorough discussion of the results. A short selection of concluding remarks in Section 5 finalizes the paper.

2. Methods

2.1 Systematic search

With the aim of providing a comprehensive systematic review of the relevant scientific literature on HBR, we designed and implemented a sufficiently sensitive search and selection strategy. To optimize our search design, we relied on the expertise of an experienced librarian. Given the intrinsically multidisciplinary nature of HBR, we needed to extend our search to multiple databases covering the fields of medicine, health care, social work and economics. Thus, the search algorithms were applied in the digital databases Scopus, EBSCOhost, CINAHL Plus (with full text), MEDLINE, Academic Search Complete, SociINDEX, Social Work Abstracts, Business Source Complete and Econlit. The development of the search syntax reflects the terminological uncertainty concerning HBR as well as our goal to allow for the location of publications that assess the economic dimension of the care strategy studied. The search results discussed below are based on the string “(reablement OR re-ablement OR restorative W/3 (home OR care)) AND (economic* OR cost* OR evaluation*)”, where the sub-command “restorative W/3 (home OR care)” indicated that we were looking for instances in which either the term “home” or the term “care” can be found within a three-word-neighbourhood of the term “restorative”. The initial search was performed on 2016-03-17. It resulted in a total of 554 records. Consecutive updates were run on 2016-08-03 and 2017-11-15. Figure 1 shows the main steps of our sequential search and selection process.

2.2 Eligibility criteria

While the first stage of the literature search relied on algorithms, the second stage involved the authors functioning as “human classifiers”. Our work was guided by a predefined list of inclusion and exclusion criteria. A study qualified for inclusion if it (i-1) contained at least a partial evaluation of the economic effects of HBR, i.e., concepts like “effectiveness”, “benefits” and “costs” of the treatment were considered, and (i-2) was published in a peer-reviewed academic journal. To maintain focus on the HBR intervention and to consider only original research satisfying reasonable design standards, two sets of exclusion criteria were defined. While the first set centres around characteristics of the intervention itself (related to particular branches of medicine and institutional/organizational aspects), the second set refers to features of the respective research study. Specifically, we agreed to exclude studies of reablement (e-1) closely linked to dental health or paediatrics or (e-2) provided by and in hospitals or nursing homes. Moreover, an article was excluded if (e-3) it could be classified as a “conceptual article”, “review article” or “research protocol”, or if (e-4) it did not assess well-defined comparator interventions. Titles, abstracts and full texts were checked against the inclusion and exclusion criteria by at least two authors independently.

2.3 Selection and categorization

One reviewer (TB) organized and carried out the initial search and eventually removed duplicates in sporadic coordination with the co-authors. Following this initial stage, a stepwise elimination procedure based on (e1)-(e4) was performed. First, two reviewers (TB, JJ) collaborated to filter records by keywords appearing in the title and the journal name. The keywords used for this purpose were "dental", "dentist", "caries", "children", "oral", and "surgery". For all matches, titles were screened and records removed if required. In the second stage, two reviewers (TB, JJ) independently screened the remaining titles. In almost 80% of those cases, the reviewers came to a unanimous decision. As a rule, a split decision lead to inclusion of the article in question. In the following stage, all reviewers independently screened the abstracts of all remaining records before discussing full-text eligibility. Subsequent searches and elimination exercises, i.e., those referred to as
“updates” in Figure 1, followed an analogous procedure with different roles assigned to the reviewers. During the update(s), one reviewer (TB) performed the filtering process on all new titles. Subsequently, the reviewers (TB, EA) screened the remaining titles for abstract eligibility. While the first update lead to the inclusion of five new records, no additional articles could be identified during the second update. Next, each reviewer independently read and analysed the articles identified in the previous stages to decide on their full-text eligibility. Finally, following a discussion, the team of reviewers reached a consensus on the pool of studies to be included in this review.

We chose to categorize the final included studies. The categorization of Emmert et al. [15] is constructive, since, in addition to providing an overview of the types of studies included in this review, it also generates the structure for the presentation of our results in Section 3. Studies that focus on cost and other consequences regarding economic efficiency were grouped into Category 1. If studies analysed the impact of HBR on both cost and consequences but did not clearly differentiate between the two, they were still included in this category. Studies evaluating health benefits for patients were placed in Category 2. Category 3 includes articles that assess the consequences of HBR on health-service usage. One could argue that reduced service usage is beneficial for the patient and should therefore be a part of Category 2. However, an effect on service usage could also have a direct monetary effect. The latter argument motivates the consideration of an additional category. Studies with multiple outcome measures were categorized by their primary outcome measure.

Figure 1: Flow chart that captures the main steps of our sequential search and selection process
2.4 Quality assessment

A 35-point checklist from Drummond, Jefferson (19) is often used as a quality assessment tool for health-economic evaluations. These guidelines contain 10 sections under three headings: study design, data collection and analysis, and interpretation of results. A similar instrument has been proposed by Drummond, Sculpher (20). In our view, these tools are particularly suited for the assessment of classic economic evaluation studies focusing on cost-effects analysis. Not all of the studies included in our review were designed for this purpose. Therefore, we saw the need to devise an instrument suitable for the assessment of research papers that are related to the complex topic of HBR. The resulting tool (c.f. Table 1) is rooted in the research and publication culture of economics, econometrics, and statistics. Those areas of expertise are represented by two of the reviewers (TB, JJ). The questions in this instrument target four different aspects of a research effort: i) general introduction, ii) data sampling and description, iii) statistics and iv) external validity. By describing the scheme of this instrument to the reader, we will allow our review process to be transparent and comprehensible. Moreover, the instrument serves as the basis of a rough scoring scheme applied to the papers included in this review. Our checklist contains 15 different questions, each checking for a specific attribute of the study scrutinized. Given the binary scaling (yes (1) or no (0)), the total score of a paper will range from 0 to 15 points. The scores reflected in the remainder of the paper were generated in a 2-stage process. After independently scoring the HBR studies, two reviewers (TB, JJ) discussed their decisions and agreed on a final score.

Prior to presenting results, a caveat should be mentioned. The scores described should be interpreted in a sensible manner. A low score is not necessarily indicative of a low quality of the research reported; it may just reflect the fact that the article assessed by means of our instrument came from an area of science in which publication standards differ from those in economics, econometrics and statistics or was published in a journal that does not emphasize certain standards to reach a specific clientele. It should be evident that our tool is not intended to denigrate the valuable work of our colleagues but instead is of a constructive nature, which should be emphasized. Use of this tool is meant to support the transparency of our review process and help to reveal differences in the publication culture across disciplines. In that sense, the tool might produce a rough indication regarding the state of HBR literature.

Table 1: The assessment scheme

<table>
<thead>
<tr>
<th>Areas</th>
<th>No</th>
<th>Question</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>General introduction</td>
<td>1</td>
<td>Does the research question identify the outcome(s) of interest?</td>
<td>A well-formulated question clearly identifies the type of effects the study seeks out to investigate. This clearly informs the reader about the aim of the study.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Does the research question identify the treatment alternatives being compared?</td>
<td>The research question should identify the alternative treatment(s) being compared. It will inform the reader about the relevant sphere of the health services studied.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Are the important stakeholders identified?</td>
<td>Health systems vary across countries. One factor that causes this variation is the incorporation of different stakeholders. The reference to the major stakeholders in the study context makes the reader aware of a key institutional characteristic of the health system under scrutiny. Important stakeholders include patients, health service providers, institutes, insurance companies, municipalities and governments.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Is the study context clearly defined?</td>
<td>The motivation for and the background of the study should be understood by the reader. This includes the motivation for implementing and testing HBR.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Are the alternatives being compared clearly described?</td>
<td>Detailed descriptions of the treatment alternatives will enable the reader to comprehend the typical service profiles provided. The natural baseline for HBR is ‘usual care’ which will vary across individuals, and therefore cannot be described in every case. We expect that the authors provide a clear description of the most common service provided to the reference group.</td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Are methods for evaluating health states and other benefits described?</td>
<td>The readers should be able to understand all methods used for evaluation. Authors should not assume that every reader is familiar with all existing instruments for measuring health. It is therefore essential that a short description of each instrument is provided.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Are the necessary scales for the methods used described?</td>
<td>A description of an evaluation method is incomplete without information about the scales of the methods. Only a reader who knows the scaling will be able to fully comprehend and appreciate the results of the study.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Is the sampling procedure clearly described?</td>
<td>The sampling procedure should be described in detail. If different instruments where used, then the interview setting should be described. Are data extracted from different databases, then the content of each source should be described. Authors should provide details of the dataset used and inform the reader on aspects such as timeframe, frequency, number of observations etc.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Does the paper provide a clear data description?</td>
<td>All empirical economic papers should provide a table of descriptive statistics and describe the data based on the table. Providing a table of baseline descriptive statistics without describing data and findings is not sufficient.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Is the choice of statistical methods used discussed and justified?</td>
<td>Statistical models are based on assumptions, that implies that they have strengths and weaknesses. Most of these models are designed for different settings and types of data. As the choice of statistical methods could have a direct influence on the results, authors should explain their choice of methods based on the sampling procedure and their research setting.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Are assumptions underlying statistical methods used discussed and addressed explicitly?</td>
<td>Typically, the operational characteristics of statistical methods are known conditional on a set of assumptions being fulfilled. A violation of such assumptions might seriously affect statistical validity. Ramifications in the context of the study should be discussed and addressed whenever possible.</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Are alternative statistical estimators discussed?</td>
<td>The researchers should carefully motivate the statistical estimators used. Notably they should reflect the main drawbacks of potential alternative estimators in the specific research setting.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Is the data analytical part of the study replicable?</td>
<td>Replicability is regarded as an important requirement for studies published in the field of economics. Given the dataset and the methodical description in the paper, an independent researcher should be able to replicate the results.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Has the study a large degree of external validity?</td>
<td>External validity is essential for a policy-maker who is considering the implementation of HBR. Several studies are not designed for providing information beyond their study setting. Studies with low degree of external validity should not be the basis of a policy-makers’ decision. Studies lacking external validity may still be interesting in terms of learning about and developing HBR.</td>
<td></td>
</tr>
</tbody>
</table>
3. Results

The search strategy identified 338 potentially relevant studies after discarding duplicates. The full-texts of 17 studies were assessed after screening titles and abstracts. The 11 articles that met our eligibility criteria are presented in Table 2. Three studies were associated with Category 1. Two of these evaluate cost and consequences separately. Four studies were assigned to Category 2. Category 3 included four studies, but three of them had secondary outcome measures that fit Category 2. Of the 11 studies identified, five were conducted in Australia, three in New Zealand, two in Norway and two in the US. All studies included in this review had a maximum intervention period of three months. The Australian HBR model specifically targets patients with low to medium levels of need (22), whereas the HBR interventions in other cases target frailter, older patients on the verge of residential care (23). The other studies included did not have a directly specified target group in terms of needs. In all reviewed studies, the multidisciplinary teams were composed of a physiotherapist, occupational therapist and a nurse. One of the team members functioned as a care manager for each client (24).

3.1 Category 1 – Costs and consequences

Kjerstad and Tuntland (25) carried out a cost-effectiveness analysis (CEA) of HBR using data from the first randomized controlled trial (RCT) on HBR conducted in Europe (26). The recruited sample consisted of 61 participants randomized to HBR (n = 31) or usual care (n = 30), but the CEA was conducted on a sample of 46 participants (HBR = 25 and control = 21). All participants were fully assessed at baseline, 3 months and 9 months. Self-perceived activity performance and satisfaction with performance were chosen as effectiveness measures. Cost data were based on individual registrations of the number of home visits, duration of each visit and profession of service delivered. There was no significant difference in the mean cost per participant during the intervention period (3 months), but the HBR group had, on average, fewer but longer visits compared to the control group. At the 9-month follow-up (6 months post-intervention period), the authors found a significant difference in mean cost per visit in favour of HBR. However, the difference of 1.5 €\textsuperscript{3} (14.7 NOK) was modest. There was no statistically significant difference in mean cost per participant. The mean changes between baseline, 3- and 9-month follow-up for both effectiveness measures were significant. The incremental cost-effectiveness ratios for the intervention period were -89.5 €\textsuperscript{3} (-868.2 NOK) for the activity performance measure and -68.7 €\textsuperscript{3} (-666.3 NOK) in terms of satisfaction with performance.

Using data from an Australian RCT (27), Lewin, Allan (28) examined the use of healthcare services and the associated costs of HBR compared to conventional care. Participants were compared at baseline and after 1- and 2-year follow-ups. Seven hundred fifty participants were included, with 375 in each group for the intention-to-treat (ITT) analysis. For the actual treatment (AT) analysis, 310 participants were included in the HBR group, 395 were in the control group, and 45 participants were excluded. The mean homecare costs per participant were different over the first year and the overall 2-year period in favour of the HBR group. The differences were 959 €\textsuperscript{2} (745 €\textsuperscript{2}) and AU$1,511 (AU$1,174) for the AT (ITT) analysis after the first year and 1,886 €\textsuperscript{2} (1,613 €\textsuperscript{2}) and AU$2,971 (AU$2,541) overall. A significantly lower proportion of HBR participants compared to conventional care patients were approved for residential or equivalent homecare at the end of the study. The HBR group had a 30% reduced risk for emergency department presentation in the AT analysis. Over the 2-year period, the AT (ITT) analysis indicated that total costs per participant for all hospital admissions were 825 €\textsuperscript{2} (194 €\textsuperscript{2}) and AU$1,299 (AU$306) lower for HBR participants than control patients. Additionally, the HBR group had a reduced risk for unplanned hospital admission in the AT analysis. Over the 2-year period,
the mean aggregated cost per participant was lower for the HBR group, and the difference was 1,821 €² (AU$2,869) in the ITT analysis and 2,754 €² (AU$4,338) in the AT analysis. The HBR group was significantly less costly in the first year and over the total 2-year period in the AT analysis only. The results for the second year alone did not show a significant difference. Randomization of participants was compromised, and there was some measurement bias in hours of service.

In a retrospective study, Lewin, Alfonso (29) investigated whether individuals using HBR reduced their need for ongoing services and had lower homecare costs compared to those receiving usual care. By linking several data sources, the authors created a dataset with 10,368 individuals and a time period of 57 months. The individuals received usual care or either of two different HBR versions. In the first HBR version the patients were referred from the community, and in the second version patients were discharged from the hospital. For the second HBR version, the maximum intervention period was 8 weeks and not the standard 12 weeks. The need for ongoing services was measured by a binary yes/no variable and used as an outcome variable in a regression framework. HBR users referred from either the community or a hospital were less likely to use ongoing services over the first 3 years compared to those getting usual care. This effect persisted over the whole time period for HBR users who were referred from the community. Quantile regression was used when making cost comparisons at 3, 12, 24, 36, 48 and 57 months. The costs for both HBR groups were substantially less than that for conventional care over the observation period. The median savings per HBR participant after 57 months amounted to more than 7935 €² (AU$12,500) in both HBR groups.

3.2 Category 2 – Health benefits

A cluster RCT conducted in New Zealand by King, Parsons (30) examined the impact of HBR versus usual care and applied HRQoL as the primary outcome. The following secondary outcomes were included: functional mobility, sense of control and social support network. All outcome data were collected at baseline and at 4- and 7-month follow-ups with structured face-to-face interviews. In total, 186 participants were included at baseline with 93 participants in each group. At the final 7-month assessment, 157 participants remained, 76 in the HBR group and 81 in the control group. HRQoL was measured by the 36-Item Short Form Health Survey (SF36³), an instrument that generates an overall score between 0 and 100, with larger numbers indicating better HRQoL. The instrument also provides separate mental and physical subscores. After adjusting for baseline demographics, the SF36 overall score differences were statistically significant at the 10% level in favour of the HBR group. The mean difference in SF36 score from baseline to 7 months was 3.8. Splitting the SF36 into the two different components indicated significant results for the mental subscore only. This suggests that HBR may improve HRQoL. For all the secondary outcomes, no evidence for significant differences between the groups was found.

The study by Lewin and Vandermeulen (31), which used data collected from 2001 to 2003, is the first Australian evaluation study included in this review. Using a non-randomized design, they investigated whether HBR participants had better personal and service outcomes compared to those receiving usual care. Data were collected manually with standardized outcome measures of functional independence, confidence and well-being. All participants were assessed at baseline, 3 months and 1 year. One hundred participants were included in each group at baseline. At the 1-year follow-up, there were 67 HBR participants remaining and 73 participants in the control group. The HBR group scored significantly worse on all measures at baseline compared with those getting usual care. At both follow-ups, the HBR group showed improvements in all measures, whereas the participants receiving usual care remained mostly the same. These differences were significant when examined by the Mann-Whitney U-test. The regression analysis also confirmed these results for all measures except the Philadelphia Geriatric Morale Scale⁴. HBR participants also had a statistically significant decrease in the probability of needing ongoing services. The latter analysis was adjusted for baseline differences between the two groups. The authors pinpointed three major limitations: some potential selection bias, a lack of independent data to confirm the service outcomes and a lack of clinical information.
<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Design</th>
<th>Year of study</th>
<th>Participant – inclusion criteria</th>
<th>Participant – exclusion criteria</th>
<th>Data source</th>
<th>Category</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kjerstad and Tuntland</td>
<td>Norway</td>
<td>RCT</td>
<td>May 2012 to Feb 2014</td>
<td>Applying or referred for homecare service (18+) and had functional decline in one or more daily activities of living (ADL).</td>
<td>Not able to understand Norwegian, were in need for residential care or rehabilitation, terminally ill, or moderately or severely cognitively reduced</td>
<td>Face-to-face interviews, time registration manually</td>
<td>Cat. 1</td>
<td>7</td>
</tr>
<tr>
<td>Lewin, Allan (28)</td>
<td>Australia</td>
<td>RCT</td>
<td>Jun 2005 to Aug 2007</td>
<td>Assessed and eligible for homecare service (65+) due to ongoing difficulties with ADL and referred for personal care</td>
<td>Not able to communicate in English, require acute or post-acute care, known diagnosis of dementia or terminal illness, complex care need requiring 15+ hours of homecare a week</td>
<td>Linking data from several databases</td>
<td>Cat. 1</td>
<td>5</td>
</tr>
<tr>
<td>Lewin, Alfonso (29)</td>
<td>Australia</td>
<td>RLS</td>
<td>Jan 2004 to Dec 2008</td>
<td>Assessed and eligible for homecare service (65+) due to ongoing difficulties with ADL. Referred from the community or discharged from hospital</td>
<td>Not able to communicate in English, known diagnosis of dementia or receiving palliative care</td>
<td>Linking data from several databases</td>
<td>Cat. 1</td>
<td>7</td>
</tr>
<tr>
<td>King, Parsons (30)</td>
<td>New Zealand</td>
<td>RCT - cluster</td>
<td>Dec 2005 to May 2007</td>
<td>Received homecare assistance (65+) from the agency included in the study</td>
<td>Unable to participate in the interview due to physical and mental health condition.</td>
<td>Face-to-face interviews</td>
<td>Cat. 2</td>
<td>6</td>
</tr>
<tr>
<td>Lewin and Vandermeulen</td>
<td>Australia</td>
<td>PLS</td>
<td>2001 - 2003</td>
<td>Assessed and eligible for homecare service (60+) due to ongoing difficulties with ADL and referred for domestic or personal care</td>
<td>Not able to communicate in English, require acute or post-acute care, known diagnosis of dementia or other progressive neurological disorders</td>
<td>Client database and data collected at home</td>
<td>Cat. 2</td>
<td>6</td>
</tr>
<tr>
<td>Parsons, Sheridan (32)</td>
<td>New Zealand</td>
<td>RCT - cluster</td>
<td>Sep 2007 to May 2008</td>
<td>Community-dwelling, new referrals for homecare (65+)</td>
<td>Severe cognitive impairment and referral for assessment for admission to residential care, care support, or short-term service</td>
<td>Not specified</td>
<td>Cat. 2</td>
<td>2</td>
</tr>
<tr>
<td>Tuntland, Aaslund (26)</td>
<td>Norway</td>
<td>RCT</td>
<td>May 2012 to Feb 2014</td>
<td>Applying or referred for homecare service (18+) and had functional decline in one or more daily activities of living (ADL)</td>
<td>Not able to understand Norwegian, were in need for residential care or rehabilitation, terminally ill, or</td>
<td>Face-to-face interviews in patient’s home</td>
<td>Cat. 2</td>
<td>7</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Study Design</td>
<td>Time Frame</td>
<td>Inclusion Criteria</td>
<td>Exclusion Criteria</td>
<td>Data Collection</td>
<td>Outcome Category</td>
<td>Score</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>--------------</td>
<td>------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------</td>
</tr>
<tr>
<td>Lewin, De San Miguel (27)</td>
<td>Australia</td>
<td>RCT</td>
<td>Jun 2005 to Aug 2007</td>
<td>Assessed and eligible for homecare service (65+) due to ongoing difficulties with ADL, and referred for personal care</td>
<td>Not able to communicate in English, require acute or post-acute care, known diagnosis of dementia or other progressive neurological disorders, or receiving palliative care</td>
<td>Client database and data collected at home</td>
<td>Cat. 3</td>
<td>7</td>
</tr>
<tr>
<td>Senior, Parsons (23)</td>
<td>New Zealand</td>
<td>RCT</td>
<td>Nov 2003 to Jun 2006</td>
<td>Assessed by hospital clinical team or regional geriatric assessment and having a high risk for institutionalization (65+)</td>
<td>Requiring immediate permanent residential care or unable to communicate in English</td>
<td>Data collected at home or residence.</td>
<td>Cat. 3</td>
<td>0</td>
</tr>
<tr>
<td>Tinetti, Baker (33)</td>
<td>USA</td>
<td>PLS</td>
<td>Nov 1998 to Apr 2000</td>
<td>Person at risk of functional decline after acute illness or hospitalization (65+), but with potential for maintaining or improving function, and receiving homecare</td>
<td>Severe cognitive impairment, requiring total assistance with care and not bedridden</td>
<td>Client database</td>
<td>Cat. 3</td>
<td>6</td>
</tr>
<tr>
<td>Tinetti, Charpentier (34)</td>
<td>USA</td>
<td>PLS</td>
<td>Nov 1998 to Apr 2000</td>
<td>Person at risk of functional decline after acute illness or hospitalization (65+), but with potential for maintaining or improving function, and received homecare</td>
<td>Severe cognitive impairment, requiring total assistance with care and not bedridden</td>
<td>Client database</td>
<td>Cat. 3</td>
<td>5</td>
</tr>
</tbody>
</table>

**Abbreviation:**
RCT = Randomized controlled trial
RCS = Retrospective longitudinal study
PLS = Prospective longitudinal study
Parsons, Sheridan (32) used a clustered RCT to determine whether HBR improves physical functioning and social support compared to standard care. The study included 205 participants at baseline, and 197 remained at the 6-month follow-up (106 HBR patients and 91 traditional care patients). The researchers who completed the assessments were experienced and were blinded to group allocation. Physical functioning was measured by the Short Physical Performance Battery (SPPB). The SPPB test contains three elements: standing balance, timed walk and timed rising/sitting from a chair. The results were interpreted conservatively, because the p-values were not corrected for multiple testing. Therefore, a 1% significance level was used in the primary analysis, and all evaluations followed the ITT principle. The HBR group had a significantly greater mean increase in overall SPPB score and in the walk component over time compared to the usual care group. There was no difference between the two groups in the balance or chair-stand components. Social support showed no difference over time between the two groups. In addition, there was no evidence for a significant relationship between allied health referrals and improvement in physical functioning over time. The authors argue that there is considerable ambiguity in determining whether a clinically meaningful change in physical function can be associated with HBR.

Tuntland, Aaslund (26) carried out the first RCT on HBR in Europe. The goal was to evaluate the effect of HBR compared to usual care on self-perceived activity performance and satisfaction with performance. Secondary outcomes were physical functioning and HRQoL. Sixty-one participants were randomized to HBR or usual care, and assessments were done at baseline and at 3- and 9-month follow-ups. At the last follow-up, 25 participants remained in the HBR group and 26 in usual care. The main outcome was measured by the Canadian Occupational Performance Measure (COPM), and all analyses followed the ITT principle and used a significance level of 5%. There was a significant mean difference in COPM-Performance at both the 3- and 9-month follow-ups. For COPM-Satisfaction, there was only a significant mean difference after 9 months. All results were in favour of HBR, and all analyses were adjusted for baseline values. The 9-month mean difference was 1.4 for both performance and satisfaction, which is below the cut-off value of 2. According to the COPM manual, this indicates a clinically relevant change. The authors acknowledge this value but also argue that there is a lack of evidence supporting this cut-off value. All the secondary outcomes were insignificant after 3 and 9 months. The study constraints rendered it inevitable that the same healthcare personnel provided services to both groups.

3.3 Category 3 – Health services usage

An Australian RCT carried out by Lewin, De San Miguel (27) investigated whether individuals receiving HBR had less need for ongoing services compared to those getting usual care. In the follow-up study by Lewin, Allan (28) that is based on the same RCT data, the main outcome was a binary variable (yes/no) representing the need for ongoing personal care services. Data were collected at baseline and at 3 months and 12 months. The study also included secondary outcomes by examining functional status and quality of life (QoL) in a subgroup recruited within the RCT group. For the AT (ITT) analysis, 294 (300) participants were recruited to this subgroup at baseline. At the 12-month follow-up, 192 (198) participants remained, and 100 (88) of these received HBR. Using logistic regression adjusted for baseline covariates, HBR was found to significantly reduce the probability of using ongoing services. These results hold for the ITT and AT analyses in both follow-ups. Regarding functional status, there was a significant difference between the groups at the 12-month follow-up. Functional dependency increased for the usual care group between the 3- and 12-month follow-ups but was maintained in the HBR group. The latter results were adjusted for baseline covariates and were only significant in the AT analysis. QoL showed no significant difference between groups.

Using an RCT design, Senior, Parsons (23) examined whether HBR participants reduced their need for permanent residential care over a 24-month period. The study also included secondary outcomes focusing on functional and social health, which were measured at the 18-month follow-up. Patients received HBR either at home or in a short-term facility, and results were not presented separately for the different locations. A total of 105 participants were recruited, with 52 in the HBR group and 53 that received usual care. Only 17 participants were included in the 18-month follow-up, with 11 in the HBR group and 6 receiving usual care. All patients included were at high risk of residential care placement. Research assistants performed randomization and data collection. Data were collected on a laptop either at the older person’s residence or at the respective short-term care facility. The ITT principle was used in all analyses. A Cox proportional hazard model with
covariates was used for the primary outcome. For the combined primary outcome of death or residential care, there were no statistically significant results. The insignificant result was a 24% reduction in favour of HBR regarding the probability of residential care or death. Additionally, all the secondary outcomes showed no statistically significant differences after 18 months. The authors also argued that the lack of blinding constituted a limitation.

Tinetti, Baker (33) investigated functional status and the likelihood of remaining at home for persons receiving HBR versus usual care in a real-world setting. This controlled clinical trial compared usual care with HBR in areas like functional status, likelihood of remaining at home, duration and intensity of the homecare episode, emergency visits to a physician and emergency department (ED) visits. There were 691 HBR users included, and from a pool of potential control participants, 691 pairs were created using an algorithm based on several covariates. A subset of 382 pairs was created for patients remaining at home after the completion of either HBR or usual care. Data on functional outcomes were only available for this subset. The descriptive and outcome data used were based on patient records. HBR patients were significantly more likely to remain at home after completion of the homecare episode. The study showed no significant difference in the likelihood of visits to a physician’s office. HBR patients were less than half as likely to have an ED visit during the homecare episode. Patients in the HBR group had significantly shorter homecare durations compared to those getting usual care. All these results were adjusted for baseline covariates. Discharge scores for self-care, home management and mobility were all slightly significantly higher for HBR users.

Tinetti, Charpentier (34) aimed to analyse the frequency of hospital readmissions for HBR compared to usual care after an acute hospitalization. Based on data from the original clinical trial study (33), 864 participants were admitted to homecare after discharge from an acute hospital stay and were therefore eligible for this study. In total, 770 participants were included, comprising 341 matched pairs and 88 additional unmatched participants. Outcome variables were hospital readmission and length of homecare episode. The algorithm previously used in Tinetti, Charpentier (34) was applied to generate matched pairs. All descriptive and outcome data came from the patient records of the original study. The main outcome variable, hospital readmission, was measured using a binary (yes/no) scale. The mean length of homecare episodes was significantly different between the two groups. The HBR group mean length was shorter than that of the control group. According to a conditional logistic regression analysis, HBR participants were 32% less likely to be readmitted than participants receiving usual care in the matched pair analysis. For the unmatched analysis, the corresponding result was 29%. The statistical significance was only marginal, with p-values for the matched and unmatched analyses of 0.10 and 0.09, respectively.

### 3.4 Assessment

Next, we discuss the results rendered by the application of our assessment tool (cf. Table 1, in subsection 2.4) to the papers reviewed above. The detailed scores reflecting our assessment are presented in Table A-1 of the Appendix.

For our sample of 11 papers the range of observed total scores was 0 to 7. The largest score observed (7) also constitutes the modal score, which was assigned to 4 papers (25-27, 29). Apparently, even the strongest papers reached only 47% of the maximum possible points. The mean and median total score were estimated as 5.3 and 6.0, respectively, while the standard deviation was 2.3. Approximately 82% of the papers received scores of 5, 6 or 7. Two papers were assigned 2 points or less.
Figure 2: Total scores for studies by category. Y-axis represents our defined study categories, and the X-axis represents total scores from our assessment scheme. Maximum possible score = 15. The triangular symbols represent the total score obtained for different studies, whereas the red multiplication sign indicates category means. Total scores is presented in Appendix.

Figure 2 illustrates the differences in the distribution of scores between the three categories defined in subsection 2.3. The highest average score (6.3) and the highest minimum score (5) were found in Category 1 – Costs and Consequences. In Category 2 – Health benefits and Category 3 – Service Outcomes, the average scores were 5.3 and 4.5, respectively. In those categories, the spread of the overall scores around the mean score exceeded the spread observed for Category 1.

Figure 3: Total scores for studies organized by country. Y-axis are countries of the included studies, and the X-axis represents total scores from our assessment scheme. Maximum possible score = 15. The triangular symbols represent the total score obtained for different studies, whereas the red multiplication sign indicates country means. Total scores is presented in Appendix.
When presenting the total scores of the studies by country of origin and year of publication as done in Figure 3, pronounced differences in means (x) and variances across countries became apparent. Two of the studies from Australia received a score of 7, but since two papers scored less, the Australian studies had an average score of 6.3. The mean score for the two studies carried out in the US was 5.5. Since two of the studies originating in New Zealand constitute the lower end of the range of total scores, the average score for New Zealand was only 2.7. In contrast, both Norwegian studies received a score of 7.

Figure 3 does not provide evidence for a clear time trend. Nonetheless, it should be pointed out that both Norwegian studies represent fairly recent efforts. Thus, the authors of these studies had the chance to learn the shortcomings of earlier publications, and they effectively avoided them in their subsequent work.

Our descriptive analysis of the total scores attests that the existing HBR literature contains some heterogeneity but has an overall mediocre quality level. An inspection of the detailed scores given in Table A-1 (cf. appendix) readily reveals the reason for this assessment. All the papers accumulated scores of 0 on items 10-13 (Statistics) and items 14 and 15 (External validity). According to our assessment, virtually all the studies failed to be informative about key aspects of statistical modelling and lacked external validity. With respect to the former trait, the only exception is the study by Kjerstad and Tuntland (25) who score a 1 on item 10 for their explicit and careful motivation of the “bootstrap” strategy.

A consequence of the previous observations is that all the existing heterogeneity can be attributed to differences in scores for items 1-5 (General introduction) and items 6-9 (Data).

The aggregated scores for the General Introduction (items 1-5) section ranged between 0 and 5. The average score was 3.6, which corresponds to 70% of the maximum possible score. Scaling the standard deviation of 1.5, we obtain a coefficient of variation of 42%. The modal score of ‘4’ was received by 4 of the studies. On items 1 and 2, which relate to the explanation of the research question, 9 of the studies received a score of ‘1’. In contrast, only 6 studies were found to identify and describe important stakeholders in an adequate way. Summing the scores for items 6-9 resulted in scores varying between 0 and 3. The mean score of 1.6, equivalent to only 41% of the maximum score attainable, indicates that one can expect a paper from the HBR literature to show deficits in the data focus area of sampling and description. Note that on item 9, which relates to a clear data description, only the study by King, Parsons (30) received a score of ‘1’. Moreover, only 4 studies were found to describe methods evaluating health states or other benefits (item 6) in a satisfactory way. Scaling the standard deviation of scores for this focus area (0.9) by the mean results in a coefficient of variation of 56%. Moreover, 3 studies (27, 29, 31) with scores on items 1-5 that exceeded their respective mean scores also had scores larger than the overall mean for items 6-9. The correlation coefficient between the two sets of scores was 0.4.

Analysing the outcomes of the assessment process suggests that while a typical HBR paper describes the motivation and all aspects of the research question in a satisfactory manner, the documentation of data-related issues could clearly be improved. The latter issue also seems to contribute slightly more to the heterogeneity in quality.

The most striking outcome of the assessment so far is that the majority of the HBR papers under review failed to be informative about key aspects of the statistical modelling. This is surprising, since due to the nature of our selection process, all papers under review appear to rely on statistical methodology. One can group the techniques implemented into two groups, i) mean comparisons, both parametric and non-parametric, and ii) regression analyses. Table 3 lists the different models and inferential techniques applied in the context of the primary outcomes.
Table 3: Statistical methods for primary outcome in each study

<table>
<thead>
<tr>
<th>Author:</th>
<th>Bootstrap</th>
<th>Wilcoxon signed rank test</th>
<th>Regression:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Linear</td>
<td>Logistic</td>
<td>GLM</td>
<td>Quantile</td>
<td>GLMM</td>
<td>MM</td>
</tr>
<tr>
<td>Kjerstad and Tuntland (25)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lewin, Allan (28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lewin, Alfonso (29)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>King, Parsons (30)</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lewin and Vandermeulen (31)</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parsons, Sheridan (32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Tuntland, Aaslund (26)</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lewin, De San Miguel (27)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior, Parsons (23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tinetti, Baker (33)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tinetti, Charpentier (34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

GLM = Generalized linear model
GLMM = Generalized linear mixed model
MM = Mixed model, also called mixed effects models (35)
Cox-Hazard = Cox proportional hazard model

Apparently, various types of regression models feature prominently in the HBR literature. According to our assessment, it is a prominent feature of the published HBR literature that the choice of such a model is virtually never justified. Alternative modelling approaches are not explicitly discussed. Models are not presented explicitly. Underlying key assumptions are not documented and it is typically not substantiated that they hold in light of the data collected. The ‘path’ leading from the data to the model is not made explicit. This, of course, has negative ramifications for the reader’s ability to critically appraise the results as well as for the replicability of the research documented. To be clear on this point, we do not believe that the authors ignored the stated aspects of statistical modelling in the research process. We simply point out the fact that, for whatever reason, there is not sufficient space allocated to such considerations in the publications under scrutiny.

The homogenous ‘0’ responses to item 14 regarding external validity suggest that the HBR studies existing so far still lack external validity. Three reasons for this drawback were typically identified and discussed: (i) small sample size and short timeframe, (ii) only one service provider or region and (iii) various methodical issues. These methodical issues arose from selection bias and manipulation of the randomization process. Finally, the fact that all studies were assigned a ‘0’ score on item 15 regarding theoretical foundation does not come as a surprise. Here, we see the manifestation of a common trend stated by Deaton (21) (p. 425), “Econometric analysis has changed its focus over the years, away from analysis of models derived from theory toward much looser specifications that are statistical representations of program evaluation.”

4. Discussion

The application of our assessment scheme revealed one specific common pattern among the studies. In our view, none of the papers scrutinized provided sufficient information about the data or the statistics employed. All studies lacked external validity. For questions 9 – 15 (data description, statistics, external validity), only two points were given. We do not believe that this evidence is indicative of the quality of the underlying research.
process. More likely, our findings reflect an established publication standard idiosyncratic to the health and medical journals where these studies were published. Failure to provide full information on data and statistics can create uncertainty in an informed reader. For example, knowledge of the sampling procedure and the process of data generation is essential for choosing an identification strategy. Without this information, the reader will not be able to fully understand the data or the strengths and weaknesses of the study. Nine out of eleven studies used regression models. The models were not presented. None of the studies provided information regarding the estimation technique used or possible adjustments of the standard errors. Not providing this type of essential information leads to a lack of transparency that in turn will reduce the replicability of a study. Thus, our assessment detects a trait of the publication culture that runs counter to two values we believe should be promoted in social sciences in general and in health economics in particular.

Since seven of the included studies are RCTs, i.e., randomized experiments, it is interesting to discuss RCTs more explicitly. The ideal RCT is often considered the “gold standard” approach for establishing causality. In biostatistics, RCTs are also viewed as the only credible approach, while experimental evaluations have traditionally been less common in economics (36). One might, however, argue that a “gold standard” does not exist (37). The primary benefit of an experiment lies in the fact that it solves the selection bias problem, not by removing the bias but by balancing the bias between the experimental groups (38). Experiments also provide a generalizable estimate of the treatment effect for the population when the sample size is large (39). The lack of sufficient sample size in the RCT studies reviewed contributes to their rather low external validity. In the design phase, all studies use power calculations for determining the target sample size. If one compares those targets to the numbers of participants included at the last follow-up, only Parsons, Sheridan (32) meets the number of participants indicated by their respective power calculation. Providing power calculations and meeting the indicated estimates does not necessarily translate into a possibly causal estimated treatment effect. Power calculations are also based on assumptions, and substantial guess work is needed (40). Computing the results of an RCT is fairly straightforward, as it simply involves comparing the group means. However, for statistical inference one needs to estimate the standard errors, which is more complicated (21). There are several alternatives for testing the significance of differences in means, but the workhorse for these computations is regression. As Table 3 indicates, most of the studies included used regression, and six of the seven RCT studies relied on regression.

Freedman (41) points out that it is common practice to adjust data from clinical trials using regression models and the like, which is also confirmed by the observations in this study. The standard way of performing a regression on data from clinical trials is to regress the outcome variable on one assignment variable, including a constant, and often control for baseline covariates. Freedman (41) analyses this model in detail and concludes that this standard way is nothing like a standard regression. He shows that the main issue is the dependence between the assignment variable and the error term, which violates key OLS assumptions. This could bias the estimated treatment effect substantially in small samples. The bias tends to decrease as the number of participants increases, but it is possible that a regression without covariates may render superior results. It is difficult to identify the studies in our analysis that use regression and OLS, but there are clues pointing at two studies (27, 32). Freedman (42) also discuss the issues of using logit/probit regression on experimental data. His key finding is that randomization does not justify the assumptions underlying these models because the outcome value is deterministic given the assignment value. Under a logit model, the outcome variable is interpreted as being random. Consequently, the usual maximum likelihood estimates could be inconsistent. The main problem here is not necessarily that these models have been used, as there are ways to solve the apparent problems, but rather the lack of discussing potential drawbacks. Freedman (41) (p. 13) states this issue quite sharply: “Practitioners will doubtlessly be heard to object that they know all this perfectly well. Perhaps, but then why do they so often fit models without discussing assumptions?” There are some non-technical problems with experiments, and these are more difficult to solve. Randomized experiments in the social setting often rest on unstated assumptions, especially considering the behavioural response of the participants, whose behaviour is often altered due to the randomization (43). Randomization bias, or deviations from assignment, cannot necessarily be treated as random measurement error and can therefore influence the results (21). None of the RCT studies discussed the latter aspects. The RCT technology may constitute a powerful tool in applied situations when the underlying assumptions are met. Often these assumptions are not arguably better than assumptions found in non-experimental econometrics and statistics (43).
One of the objectives of this paper is to provide an overview of economic evaluations of HBR. Previous reviews either found no data evaluating the effects of HBR or only concentrated on a few outcome measures (17, 18). Our review effort differs from earlier attempts, especially in terms of “wider” inclusion criteria with fewer limitations on study type and outcome measures. Each of the eleven studies found to be eligible for our review was assigned to one of three categories. Three studies estimated the cost differences between HBR and usual care after the intervention, and all showed lower costs for HBR participants (25, 28, 29). However, these results are not clearly significant. In the Norwegian study, the mean cost difference was not statistically significant (25), and one study did not report significance (29). For one of the studies, the significant differences in mean cost differed between the ITT and AT analysis, with AT showing a significant difference (28). Like the former, the latter study also includes cost for ED and hospital admissions along with homecare costs. If one focuses on homecare cost alone, then the potential yearly cost reduction for HBR seems to range from approximately 800 – 1,700 € per participant. Here, we have to stress that this is only a rough estimate, and the results of the study vary greatly with time. This might explain the wide range of potential cost savings. It seems that the potential savings increase with the length of the post-intervention period (29).

Table 2 exhibits some general information about the studies included in this review. Scrutinizing columns five and six of the table, one finds similar inclusion and exclusion criteria defining the pool of participants in the various HBR studies. The main exclusion criterion was that participants are not in need of residential care and not significantly cognitively reduced. Most studies applied narrow inclusion criteria requiring that patients eligible for care are older than 65 years. An exception is the Norwegian studies, in which the minimum age was set to 18 years. There are, however, only small variations in the mean age of included participants. For the HBR group, the mean age was between 76 and 82, while for the usual care group it was between 77 and 83. An additional trait common to the studies reviewed is the length of the HBR intervention itself, which was a maximum of 12 weeks. In the New Zealand version, in which participants were referred from the hospital, the length was limited to 8 weeks. Two studies (30, 32), failed to be informative with respect to this aspect. The homogeneity with respect to this feature increases the comparability of the studies. In fact, according to our observations, the length of the intervention itself was hardly ever explicitly explained. Neither the actual amount of HBR administered to the participants nor the possible effects of a variation in treatment duration on treatment outcome was discussed. Studies examining potential health benefits from HBR do not use one standardized instrument. In fact, to establish different types of health benefits, one may require different instruments, but often there are several instruments used for the examination of the same type of benefit. Directly comparing the results then becomes difficult. We will therefore focus the discussion on whether there were some common trends in terms of statistical significance for potential health benefits.

Physical functioning or independence were the potential benefit categories where we found the most studies, and often these focused on ADL. The study by Lewin and Vandermeulen (31) is the first study to use functional gain as the primary outcome. This study produced some promising results. The HBR group scored significantly better on all physical measures after 3- and 12-month follow-ups. These results are consistent with earlier studies examining short-term effects (33). A more recent study also indicated improvements in physical functioning for the HBR group (32). Less clear are the results of an Australian study, in which statistical significance in instrumental ADL could only be established in the context of the AT analysis. The latter study used a 12-month follow-up period. In contrast, three studies showed no statistical significance in either functional mobility or ADL (23, 26, 30). The follow-up periods in these three studies lasted between 7 and 18 months. This is longer than the respective period in the studies reporting positive statistical significance in favour of HBR. A common pattern for all the results is that there were no significant or clear effects on physical functioning. These studies all included physical gain as a secondary outcome. They were not originally designed for detecting any effect on physical functioning. This may influence the results. However, this argument partially also holds for the studies reporting a positive significant effect. The study by Parsons, Sheridan (32) defines physical functioning as the primary outcome, but it relies on data collected from another study (44). There is no clear evidence supporting the notion that HBR significantly increases physical functioning. Encouragingly, in the studies that produced no significant difference, HBR tended to lead to superior results on the selected instruments.

Increased HRQoL or QoL is often used as a measure of increased health benefits, and three studies in our review relied on this measure. However, only one study had change in HRQoL as the primary outcome (30). This study showed a promising result, with the HBR group scoring significantly better at the 7-month follow-up. The mental health component of SF36 was the main driver for the increased score for the HBR group. For the
mean difference between baseline and 7-month follow-up, both the overall score and the mental component had a p-value of 0.05. The two remaining studies looking into HRQoL or QoL reported insignificant differences between HBR and usual care (26, 27). The follow-up periods in these studies were 9 – 12 months. If one examines the result table in Tuntland, Aaslund (26), most of the HRQoL components are in favour of HBR. All these effects are associated with p-values larger than 0.05. To summarize, there is no convincing evidence that HBR increase HRQoL or QoL. Again, the studies that report insignificant differences had HRQoL or QoL as their secondary outcomes. With respect to other self-perceived health benefits, the results are also not definite. Two studies (30, 32) reported no significant difference in social support measured with the Duke Social Support Index (DSSI) (45). The end follow-up periods for these studies only differed by 1 month, and both were conducted in New Zealand, making these studies highly comparable. The p-value associated with the change in the DSSI score between usual care and HBR was found to be 0.09 in the study by Parsons, Sheridan (32). However, the authors argue for the use of a significance level of lower than 0.05 because of the risk of type-II errors. Regression results from assessing the state of psychological well-being of older people also showed no significant difference at the 12-month follow-up (31). Self-perceived activity performance and satisfaction with that performance was analysed in (26). Both the performance and satisfaction measures were significantly better for the HBR group at the 9-month follow-up. However, the treatment effect was below the clinically relevant change cut-off.

Postponing the need for residential care is an obvious potential benefit for a patient and for the policy-maker, as residential care is, in general, costly. In an unadjusted analysis, it was demonstrated that HBR users were significantly less often assessed and approved for a higher level of care in a 2-year perspective (28). Senior, Parsons (23) observed that HBR reduced the probability of death or permanent residential care, but their observations lacked statistical significance. It was also shown that HBR users were less than half as likely to have an ED visit during the home care episode (33). Over a 2-year period, HBR recipients had significantly less ED presentations compared to individuals receiving the baseline treatment, though these results only hold for the AT analyses and were unadjusted (28). The latter findings also hold for the number of hospital admissions. Moreover, an earlier study concluded that HBR participants were less likely to be readmitted to the hospital compared to subjects under usual care, a result that was only significant at a 10% level (34). In addition, HBR is showing some promising results with respect to reducing the need for specialist or residential care. As discussed earlier, HBR seems to reduce homecare costs, and therefore one would expect a decline in the volume of homecare services. In the first study included in this review, it was shown that HBR participants were significantly more likely to remain at home after a homecare episode (33). This effect seems to hold in a 12-month perspective, as it was shown that HBR participants were significantly less likely to need ongoing services (27). There is evidence for the fact that relative to usual care, HBR significantly reduces the number of homecare hours and visits as well as the general duration of homecare episodes in the long-term (25, 28, 33, 34).

5. Conclusion

This review summarizes and assesses the currently available literature on empirical evaluations of the modern care concept of HBR. In short, the existing evidence regarding the effects of HBR is still inconclusive. The results are inconsistent, as some studies report a significant positive effect of HBR versus usual care, while others fail to establish such an effect. However, so far it has not been established that HBR renders negative effects. On one hand, this review is concerned with a concise summary of relevant existing findings generated by research focusing on HBR. On the other hand, it tries to provide a critical, constructive assessment of the associated publication process. Having worked on this project, we understand that HBR is a complex intervention implemented in an equally complex setting. Out of this understanding grows the utmost respect for all current research efforts aimed at estimating the effects of HBR. The research reviewed provides a basis to build on. With complex interventions in social settings, there might also be a need for different “eyes” to capture this complexity. To ensure successful future research efforts, the multidisciplinary spirit of HBR needs to be reflected in the diversity of the research teams taking on the challenge.
Endnotes
1 Exchange rates from 06.02.2018 and collected from the Norwegian national bank, 1 € = 9.7005 NOK
2 Exchange rates from 06.02.2018, and collected from the Reserve bank of Australia, 1 AU$ = 0.6348 €
3 https://www.rand.org/health/surveys_tools/mos/36-item-short-form.html
5 https://www.nia.nih.gov/research/labs/leps/short-physical-performance-battery-sppb
6 http://www.thecopm.ca/

References
## Appendix

### Table A-1: Results of the quality assessment

<table>
<thead>
<tr>
<th>Studies</th>
<th>No 1</th>
<th>No 2</th>
<th>No 3</th>
<th>No 4</th>
<th>No 5</th>
<th>No 6</th>
<th>No 7</th>
<th>No 8</th>
<th>No 9</th>
<th>No 10</th>
<th>No 11</th>
<th>No 12</th>
<th>No 13</th>
<th>No 14</th>
<th>No 15</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kjerstad and Tuntland (25)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Lewin, Allan (28)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Lewin, Alfonso (29)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>King, Parsons (30)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Lewin and Vandermeulen (31)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Parsons, Sheridan (32)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Tuntland, Aaslund (26)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Lewin, De San Miguel (27)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Senior, Parsons (23)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tinetti, Baker (33)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Tinetti, Charpentier (34)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td><strong>0</strong></td>
</tr>
</tbody>
</table>