UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Zentrum für Psychosoziale Medizin
Institut für Gesundheitsökonomie und Versorgungsforschung

Direktor: Prof. Dr. med. Hans-Helmut König, MPH

Health economic evaluation of interventions aimed at improving care for the older population [Gesundheitsökonomische Evaluation von Interventionen zur Versorgungsverbesserung der älteren Bevölkerung]

Kumulative Dissertation

zur Erlangung des Doktorgrades Dr. rer. biol. hum.
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Sophie Gottschalk
aus Neuruppin

Hamburg 2022

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 20.04.2023

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Hans-Helmut König

Prüfungsausschuss, zweite/r Gutachter/in: PD Dr. Christine Blome

Contents

List of tables and figures

List of abbreviations

BMBF – Bundesministerium für Bildung und Forschung (German Ministry for Education and Research)

BREQ-3 – Behavioural Regulations in Exercise Questionnaire

ΔC – incremental costs

CEA – cost-effectiveness analysis

CEAC – cost-effectiveness acceptability curve

CI – confidence interval

COSMIN – Consensus-Based Standards for the Selection of Health Measurement Instrument

CUA – cost-utility analysis

ΔE – incremental effect

EQ-HWB – EQ Health and Wellbeing

FIMA – Fragebogen zur Erhebung von Gesundheitsleistungen im Alter (questionnaire for the use of medical and nonmedical services in old age)

FoF – fear of falling

gLiFE – group version of the Lifestyle-integrated Functional Exercise (LiFE) program

GLM – generalized linear models

HrQoL – health-related quality of life

ICECAP-O – ICEpop CAPability measure for Older people

ICER – incremental cost-effectiveness ratio

IQWiG – Institut für Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care)

LiFE – Lifestyle-integrated Functional Exercise program

LLFDI – Late Life Function and Disability instrument

MAR – missing at random

MCAR – missing completely at random

MICE – multiple imputation by chained equations

NMB – net monetary benefit

NMB_i – individual net monetary benefit

PA – physical activity

PROMIS – Patient-Reported Outcomes Measurement Information System®

QALY – quality-adjusted life year

QoL – quality of life

QOL-ACC – Quality of Life - Aged Care Consumers

RCT – randomized controlled trial

SE – standard error

Short FES-I – Short Falls Efficacy Scale International

WTP – willingness to pay

List of publications and manuscripts

Chapter 3 was published as:

Gottschalk S[1], König HH[1], Schwenk M[2], Nerz C[3], Becker C[3], Klenk J[3,4,5], Jansen CP[2], Dams J[1]. Mediating factors on the association between fear of falling and health-related quality of life in community-dwelling German older people: a cross-sectional study. BMC Geriatr. 2020 Oct 14;20(1):401. DOI: 10.1186/s12877-020-01802-6

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Network Aging Research, Heidelberg University, Heidelberg, Germany.

[3]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany.

[4]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.

[5]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany.

Chapter 4 was published as:

Gottschalk S[1], König HH[1], Schwenk M[2], Nerz C[3], Becker C[3], Klenk J[3,4,5], Jansen CP[2,3], Dams J[1]. Cost-Effectiveness of a Group vs Individually Delivered Exercise Program in Community-Dwelling Persons Aged ≥70 Years. J Am Med Dir Assoc. 2022 May;23(5):736-42.e6. DOI: 10.1016/j.jamda.2021.08.041

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Network Aging Research, Heidelberg University, Heidelberg, Germany.

[3]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany.

[4]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.

[5]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany.

Chapter 5, unpublished manuscript, accepted for publication in *Age and Ageing* as:

Jansen CP[1,2*†], Gottschalk S[3†], Nerz C[1], Labudek S[4], Kramer-Gmeiner F[4], Klenk J[1,5,6], Clemson L[7], Todd C[8,9,10], Dams J[3], König HH[3], Becker C[1], Schwenk M[4,11]. Comparison of falls and cost-effectiveness of the group vs. individually delivered Lifestyle-integrated Functional Exercise (LiFE) program: final results from the LiFE-is-LiFE non-inferiority trial. Age Ageing. [accepted].

[1]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany
[2]Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany
[3]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[4]Network Aging Research, Heidelberg University, Heidelberg, Germany
[5]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany
[6]IB University of Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany
[7]Sydney School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia
[8]School of Health Sciences, Faculty of Biology, Medicine & Health, The University of Manchester, Manchester M13 9PL, UK
[9]Manchester University NHS Foundation Trust, Manchester, UK
[10]Manchester Academic Health Science Centre, Manchester, UK
[11]Human Performance Research Centre, Department of Sport Science, University of Konstanz, Konstanz, Germany
[†]equal contributors

Chapter 6 was published as:

Gottschalk S[1], König HH[1], Schwenk M[2,3], Nerz C[4], Becker C[4], Klenk J[4,5,6], Jansen CP[4,7], Dams J[1]. Willingness to pay for a group and an individual version of the Lifestyle-integrated Functional Exercise program from a participant perspective. BMC Public Health. 2022 Oct 18;22(1):1934. DOI: 10.1186/s12889-022-14322-2

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2]Network Aging Research, Heidelberg University, Heidelberg, Germany.
[3]Human Performance Research Centre, Department of Sport Science, University of Konstanz, Konstanz, Germany.
[4]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany.
[5]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.
[6]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany.
[7]Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany.

Chapter 7 was published as:

Gottschalk S[1], König HH[1], Nejad M[1], Dams J[1]. Psychometric Properties of the EQ-5D for the Assessment of Health-Related Quality of Life in the Population of Middle-Old and Oldest-Old

Persons: Study Protocol for a Systematic Review. Front Public Health. 2020;8:578073. DOI: 10.3389/fpubh.2020.578073

[1]*Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

Chapter 8 was published as:

Gottschalk S[1], König HH[1], Nejad M[1], Dams J[1]. Measurement properties of the EQ-5D in populations with a mean age of ≥ 75 years: a systematic review. Qual Life Res. 2022 Aug 1 [Epub ahead of print]. DOI: 10.1007/s11136-022-03185-0

[1]*Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

1 Synopsis

This thesis deals with health economic evaluations of interventions aimed at improving care for the older population, which is relevant in light of an aging society that challenges the sustainability of healthcare systems. Specifically, an activity promotion and fall prevention program was economically evaluated (publication 2-4; chapter 4-6), complemented by an analysis of mediating factors of fear of falling and health-related quality of life (publication 1; chapter 3). In addition, the measurement properties of the EQ-5D as the most widely used instrument to measure effectiveness of interventions in economic evaluations were systematically reviewed (publications 5-6; chapter 7-8), thereby contributing to a discussion of methods of economic evaluations of interventions targeting an older population.

This chapter provides relevant background information on the implications of an aging population for the healthcare system, the economic relevance of promoting healthy aging and the specific role of preventing falls and physical inactivity, as well as the state of economic research in the field of exercise-based falls prevention. Principles of economic evaluations relevant to this thesis are summarized, the contributions of the thesis are highlighted, followed by the presentation of the specific objectives of each included study. Subsequently, materials and methods are introduced, before the results of each study are briefly summarized and discussed. The chapter closes with a general and overarching discussion of the studies, including implications for future research as well as for decision-makers and practice.

1.1 Background

1.1.1 The implications of an aging society for the healthcare system

In most countries of the world, but especially in western, industrialized countries such as Germany, the proportion of older people has risen and this increase is projected to continue in the future, leading to an aging of the population [1, 2]. In the Summary of the World Population Prospect 2022 from the United Nations it is stated that "Countries with ageing populations should take steps to adapt public programmes to the growing proportion of older persons, including by improving the sustainability of social security and pension systems and by establishing universal health care and long-term care systems" [2]. In Germany, the social health insurance system is mainly financed by income-dependent contributions. Thus, a shrinking working population (higher contributions, lower healthcare resource consumption) in relation to a growing population of older, retired people (lower contributions, higher consumption) will possibly challenge the sustainability of the social health insurance systems. Due to higher (multi)morbidity in older age, healthcare costs are higher in older than in younger age groups [3, 4]. However, to date it is not entirely clear whether and how a shift in the age structure of the population due to low birth rates combined with rising life expectancy affects health expenditure [5]. Recent findings indicate that the famous "red herring" hypothesis [6] – according to which increasing life expectancy does not increase per capita healthcare expenditure since the highest costs are incurred at the very end of life and are compressed to older ages – may not hold (completely) true. For example, Kollerup et al. found that even after accounting for the time to death (which indeed has an impact on healthcare expenditure), age itself predicts individual healthcare expenditure and, in addition, individual-level healthcare expenditure increases faster in older people, driven by the age group 75+ [7]. These findings emphasize the need to promote healthy aging – "the process of developing and maintaining the functional ability that enables wellbeing in older age" [8] – to postpone high end of life costs. Moreover, several authors stress the importance of disease prevention and designing and providing cost-effective interventions to tackle the challenges an aging population poses to the sustainability of public health insurance systems [5, 7].

Promoting health into old age and optimizing healthcare for the older population is, of course, not only economically beneficial, but is also inherent in fundamental ethical principles and our understanding of justice and dignity. The right to health, regardless of age, is enshrined in the Declaration of Human Rights [9] and the Sustainable Development Goals [10], and in Germany healthy aging has been declared as national health goal [11]. Recognition of the societal and economic challenges of an aging population has led to an increase in the development and research of interventions to improve health and care of this population. Economic evaluations of these interventions can aid in deciding which interventions should be implemented into regular care – ideally those that provide value for money (e.g., are cost-effective) and thus contribute to the sustainability of the health and social insurance system.

This thesis focused on the economic evaluation of interventions that aim to prevent unintentional falls and promote physical activity (PA) in older people, two factors that play an important role in the promotion of healthy aging and are both sub goals of the national health goal “healthy aging” [11]. Furthermore, this thesis dealt with the measurement of intervention effects in economic evaluations of programs targeting an older population.

1.1.2 The role of falls and physical inactivity for healthy aging – interplay and economic relevance

Around 30% in the population aged 65 and older falls each year [12], and in Germany the annual incidence of fall-related injuries in the population aged 70 and older is around 15% [13], resulting in considerable disease burden in terms of years lived with disability, high healthcare costs, and a reduction of quality of life (QoL) [14-17].

Physical inactivity is a global pandemic [18] and, as a leading risk factor for the development of widespread non-communicable diseases [19], poses a considerable economic burden on societies and healthcare systems [20]. The costs of physical inactivity are particularly high in the age group 65+, highlighting the importance of promoting PA in the older population [21]. Higher levels of PA are associated with lower mortality and contribute to healthy aging by protecting against chronic diseases and risk factors, against a decline in physical and mental functional status, against limitations in activities of daily living and QoL, and by leading to better psychological well-being and social outcomes [22-26]. Furthermore, people with a higher PA have a lower utilization of healthcare resources and thus incur less costs, both in the general and in the older population [27].

However, PA, especially walking, does not come without risks in older people. For example, Okubo et al. found that increasing PA in people with more than two fall risk factors (poor balance, mobility limitations, knee pain, depressive symptoms, assistive device, polypharmacy, history of falling) was associated with an increase in the fall incidence [28]. Furthermore, Lu et al. described a u-shaped relationship between PA and falls in older people, meaning that both low and very high levels of PA were associated with higher risk of falling [29]. A psychological aspect of falls that is interlinked with PA and falls as well is fear of falling (FoF). It is not only a consequence of falling, but is, e.g., associated with (lower) QoL, independent of an actual fall experience [30] and can be a barrier to being physically active [31]. While a reduction of PA may initially reduces the risk of falling (simply by avoiding exposure), the beneficial effect of PA on health and function is forgone, which “promotes” (further) deterioration of functional capacity, and thereby increases the risk of falling in less advanced activities (e.g. activities of daily living) [32]. This interlink of falls, FoF, and PA strengthens the importance of promoting safe PA in older people, meaning to combine fall prevention strategies in activity promotion programs and vice versa [29], and stresses the importance of measuring the effectiveness of fall prevention programs by adjusting fall outcomes for the level of PA [32, 33]. Moreover, since healthy aging, by definition, is not restricted to physical health but expands to broader concepts of well-being and QoL, FoF is a relevant aspect that can or should be addressed by interventions.

Overall, falls, FoF, and PA are modifiable factors that are related to health and QoL and concern a large proportion of the population. Therefore, (cost-)effective approaches to prevent falls, reduce FoF, and improve PA have the potential to contribute to healthy aging and lower the economic burden associated with falls and physical inactivity.

1.1.3 Exercise-based falls prevention in older people – state of economic research

From a public health and economic perspective, preventing falls and improving PA is attractive for the previously mentioned (long-term) consequences of falls and physical inactivity on health, as well as healthcare and societal costs. Therefore, especially in the last decade, research interest in fall prevention in older people and also the economic evaluation of different strategies of fall prevention have increased. Fall prevention strategies can range from home modification and medication adjustment (elimination of "external" risk factors) to structured exercise interventions. Systematic reviews found strong evidence that exercise-based interventions are highly effective [34, 35] and are therefore the first-line recommendation for older adults at intermediate risk of falls. [36]. This population can be assumed to have a high long-term benefit from exercise programs, as they often still live independently (community-dwelling) and have the functional capacity to perform the exercises safely. Thereby, adverse and costly consequences such as deterioration in physical status, injurious falls, and institutionalization may be avoided in the first place or postponed. However, despite the increasing literature of economic analyses of fall prevention strategies, the number of economic analyses of exercise-based interventions in community-dwelling older people is rather limited [37]. The existing studies show a high heterogeneity in cost-effectiveness results due to heterogeneity in the content, duration, and intensity of the interventions, target populations (e.g., low or high risk of falling; age and disease groups), but also in methodology of the cost-effectiveness analyses [37-39]. Moreover, characteristics of different healthcare and social systems differ between countries, which additionally hampers generalizability of the results (e.g., to Germany). For Germany, only one modelling study that economically evaluated an exercise-based falls prevention program targeting community-dwelling older people has been conducted so far [40]. Besides focusing only on the cost of preventing hip fractures, this study used a pooled estimate of the effectiveness of a hypothetical program taken from a meta-analysis of different exercise programs [41] rather than controlled trial data on the effectiveness of a specific program, which challenges the internal validity of the results.

Considering the usually higher costs for the implementation of exercise-based fall prevention programs compared to, e.g., home assessment and medication adjustment [39], which require less human resources, lowering the implementation costs while maintaining the effectiveness of an intervention is desirable. Therefore, the development of group versions of programs that have been shown to be effective when delivered individually may be an economically attractive alternative, especially as the evidence currently available does not indicate that group programs are inferior to individually delivered exercise programs regarding their cost-effectiveness [42].

1.1.4 Introduction into health economic evaluations

(Health) economic evaluation is an umbrella term for different study types that aim to inform about the optimal allocation of (scarce) resources [43]. Although resources are generally not unlimited, the relevance of health economic evaluations in the authorization or reimbursement of pharmacological and non-pharmacological interventions differs between countries. In the UK, for example, where the National Health Service is a state institution that has to operate on a certain budget, economic evaluations are more established and inform reimbursement decisions [44]. In Germany, with its less centralized health system, economic evaluations have historically been less relevant, although the efficiency of healthcare services is explicitly stipulated by law in the German Social Code, Book Five ("Fünftes Sozialgesetzbuch", SGB V). For example, § 12 SGB V states that services that are not necessary or uneconomical cannot be claimed by insured persons, may not be provided by the service providers, and may not be approved by the health insurance funds. Moreover, § 135 SGB V emphasizes not only medical benefit but also efficiency as a prerequisite for the reimbursement of new interventions, also in comparison to interventions that are already covered by the health insurance. In view of the pressing challenges, e.g. an aging society, considering efficiency in decision-making has become more relevant and may become inevitable in the future, especially if the intention is not to raise the contribution rates to the social insurance (e.g., keep the contributions to the health insurance at around 15%).

Economic evaluations can be defined as “the comparative analysis of alternative courses of action in terms of both their costs and consequences” [45] and thereby allow making allocative decisions (e.g., Should a new intervention be preferred over “usual care” or a competing intervention?). The types of analysis that are mainly conducted to this end and were also conducted in this thesis are cost-effectiveness analyses (CEAs) and cost-utility analyses (CUAs).

CUAs can be seen as sub-type of CEAs as the main idea behind both study types is to evaluate whether (additional) costs with delivering the intervention are justified by the consequences, and therefore can be considered cost-effective. In CEAs, the difference in costs (ΔC) as well as the difference in effects (ΔE) between the intervention of interest and the comparator(s) is calculated, and the incremental cost-effectiveness ratio (ICER) is built.

$$\mathrm{ICER} = \Delta \mathrm{C}/\Delta E$$

The difference between a CEA and CUA lies in the way effectiveness is measured. In a CEA, the ICER can be expressed as costs per any measurement unit of interest in a specific context (e.g., costs per fall prevented or per hospital admission avoided). In a CUA, the ICER refers to the costs per quality-adjusted life year (QALY), typically based on generic QoL instruments, and thus is assumed to allow comparability across studies [45] (more on this in chapter 1.1.4.1).

Moreover, CEAs and CUAs have in common that the perspective that is adopted for the analysis determines which costs and consequences are considered in the analysis. In turn, the perspective adopted depends on whom the analysis is intended to inform [45]. When examining the cost-effectiveness of an intervention that is potentially paid for by a healthcare payer (e.g., the National Health Service or health insurances), the costs and consequences of the intervention for the particular payer are of relevance. In Germany, where the majority of healthcare costs is covered by the health or long-term care insurances, the healthcare payer perspective is often adopted and thus all costs incurred for the healthcare payer should be included in the analysis (e.g., outpatient and inpatient treatment, or medication). However, some interventions may have costs and consequences in areas outside the healthcare sector, so that the true impact of an intervention can only be revealed when the costs or consequences that are relevant from a broader societal perspective are included in the analysis, e.g. costs from lost productivity or informal (unpaid) care.

Economic evaluations can further be distinguished between trial-based and model-based approaches. Trial-based approaches are mostly conducted alongside randomized controlled trials (RCTs), which offer high internal validity due to the randomization process and collected individual patient-level data. However, the generalizability of results may be limited because certain eligibility criteria or other selection effects (people who consent to participate in a study may have certain characteristics) may result in randomized participants no longer being representative of the target population [45]. The informative value of a trial-based economic evaluation is often further limited by the time horizon and sample size of the main trial which is usually determined based on a primary outcome; and this primary outcome is not cost-effectiveness in most cases. Decision-analytic modelling can potentially overcome the limitations of trial-based economic evaluations, but the validity of the results heavily relies on the available input data for the model [45]. This could be especially challenging if there is no solid evidence on how short-term or intermediate costs and consequences will develop in the long-term. Despite its (potential) limitations, trial-based economic evaluations are still deemed important to inform decision-makers if certain methodological standards are followed [46]. These include, among others, appropriately handling missing data and the assessment of uncertainty under consideration of appropriate statistical methods to deal with distributive characteristics of the data, potential baseline imbalances between groups, and the correlation between costs and effects [46-48]. The following paragraphs describe how these standards were implemented in this thesis.

1.1.4.1 Dealing with missing data

When using trial data, missing data on item-level or due to early drop-out are often inevitable, but basing the analysis only on complete cases would increase the risk of bias and unnecessarily reduce the power of the analysis, especially if more than five percent of the observations have missing data [46]. In trial-based economic evaluations, multiple imputation is the recommended approach for dealing with missing data [46, 47]. In multiple imputation, missing values are predicted based on observed data using regression methods. This procedure is repeated several times resulting in m independent datasets (e.g., m=10) which are then analyzed independently. Finally, the results are pooled into an overall estimate by Rubin's rule [49]. In this thesis, missing data was imputed using multiple imputation by chained equations (MICE) with predictive mean matching as imputation method [50, 51]. In MICE, missing values are predicted in a cycle of imputation models: For each variable with missing values, unless specified otherwise, these missing values are predicted by all other variables (often referred to as fully conditional specification). To generate one imputed dataset, several cycles are typically run to stabilize the results. Since each variable is imputed based on its own imputation model, MICE has the advantage that different types of variables or scale levels can be imputed. Predictive mean matching ensures that only observed values are imputed, thus the true distribution is reflected by the imputed values. This is of relevance when data is highly skewed, as is often the case with, e.g., cost data. Since multiple imputation methods are based on the assumption that missing data are either unrelated to observed or unobserved factors (missing completely at random, MCAR) or only related to observed factors (missing at random, MAR), sensitivity analyses should be performed to investigate the impact of possible departures from the MAR assumption on the conclusions [48].

1.1.4.2 Characterizing uncertainty

An important standard is to assess the uncertainty around the ICER, which is only a point estimate. This point estimate alone is often not sufficient to inform decision-makers about the cost-effectiveness of an intervention because it does not take uncertainty around this estimate into account and can be misleading, as typically neither costs nor effects follow a normal distribution. If the differences between groups in costs or effects are examined *separately*, regression methods can be used where ΔC or ΔE are estimated along with confidence intervals representing uncertainty. Furthermore, regression methods have the advantage that they allow for the consideration of covariates that potentially bias the point estimate (e.g., baseline imbalances between groups). The skewed distributions may be addressed by calculating bootstrapped confidence intervals (CIs) for the estimates from a linear regression or by using generalized linear models (GLMs) that are more suitable for specific distributive characteristics [46]. However, cost-effectiveness analyses are concerned with the *joint* distribution of ΔC and ΔE, which in most cases does not follow a normal distribution. Therefore, using the separately modeled (adjusted) estimates of cost and effect differences is not recommended.

Instead, the uncertainty around the ICER can be explored by non-parametric bootstrap [52], where the ICER is repeatedly calculated (e.g., 1,000 times) and these ICER replications can then be plotted on a cost-effectiveness plane to visualize the joint density of cost and effect differences (Figure 1).

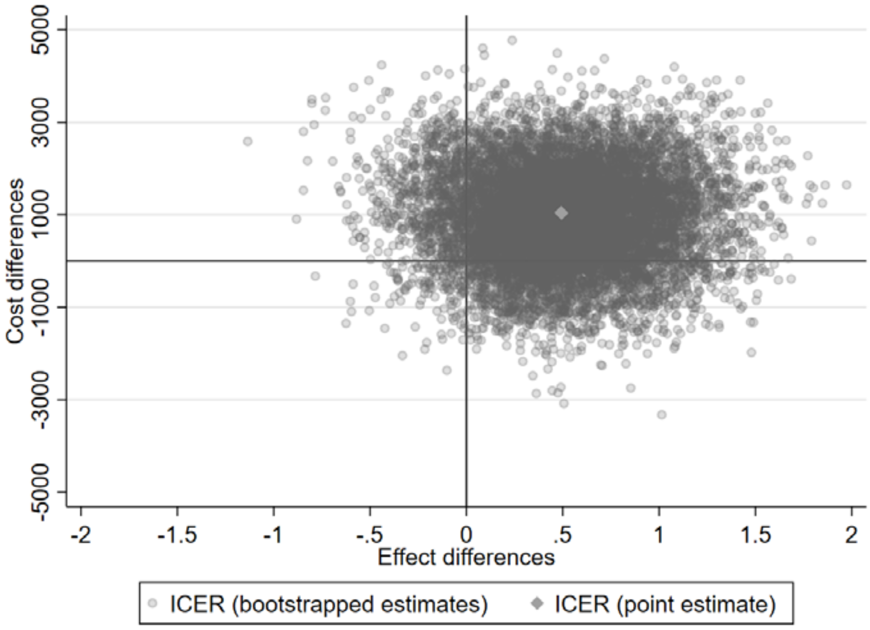


Figure 1 Example for a cost-effectiveness plane with bootstrapped ICERs

ICERs located in the north-western (less effective, more costly) or the south-western quadrant (less effective, less costly) of the cost-effectiveness plane are not desirable, and hence an intervention where the majority of the bootstrapped ICERs are located in these quadrants would not be considered cost-effective. While ICERs in the south-eastern quadrant (more effective, less costly) clearly favor the intervention of interest over the comparator, the decision in the north-eastern quadrant (more effective, more costly) is less clear. There, the cost-effectiveness of an intervention depends on whether the required additional investment can be justified by the additional effect. Therefore, the willingness to pay (WTP, λ) for an additional unit of effect is central in determining the cost-effectiveness of an intervention. In the net-benefit framework [53, 54], an intervention can be considered cost-effective if the ICER is less than the maximum WTP or, rearranged as linear expression, the net monetary benefit (NMB) is greater than zero:

$$NMB = \lambda \cdot \Delta E - \Delta C.$$

Assuming different values of λ, the proportion of the bootstrapped ICERs that have a positive NMB can be determined and visualized as cost-effectiveness acceptability curve (CEAC, Figure 2), indicating the probability of the intervention being cost-effective depending on the WTP. Since the NMB is a linear expression, it is more likely to follow a normal distribution than the ICER [52]. Instead of using non-parametric bootstrapping of the ICER, CEACs can also be derived directly from net-benefit regression models [55], which are linear models for a subjects individual net monetary benefit (NMB_i) and can be formed as

$$NMB_i = \alpha + \delta_{ti} + \varepsilon_i$$

where α is the intercept, δ_{ti} is the coefficient for a treatment dummy t (intervention = 1, comparator = 0), and ε is an error term. The model can be expanded by additional explanatory variables, thereby allowing for covariate adjustment (e.g., baseline imbalances between groups) or subgroup analyses, which is an important advantage over non-parametric approaches. The probability of cost-effectiveness at certain WTP values can be derived from the p-value corresponding to the coefficient of the treatment variable $(1 - p/2\ if\ \delta > 0;\ p/2\ if\ \delta < 0)$ [56]. Given the advantages of the net-benefit regression framework, this method was employed for the CEAs and CUAs in this thesis.

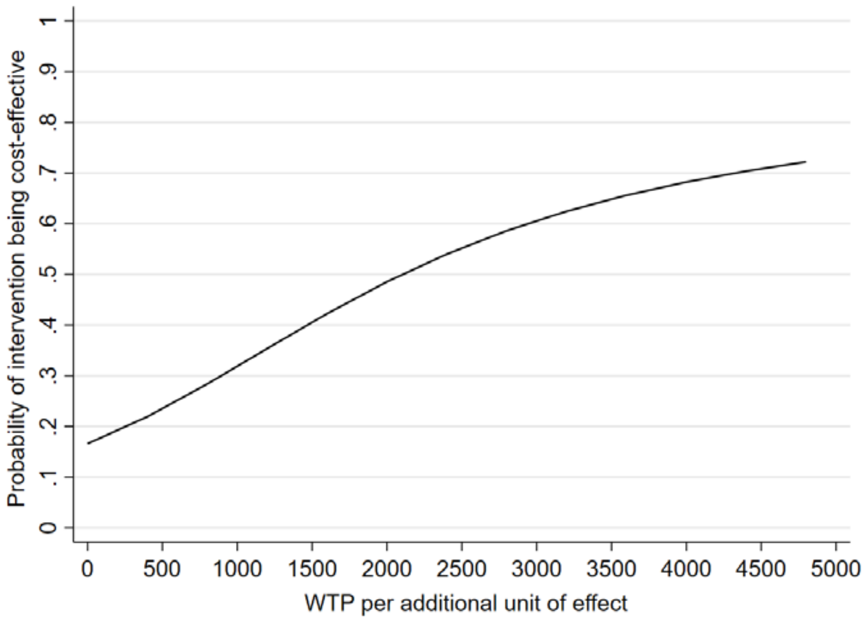


Figure 2 Example for a cost-effectiveness acceptability curve (CEAC)

1.1.4.3 Measuring intervention effects in economic evaluations

Besides measuring the consequences of interventions in terms of costs, an important area of research in health economics is how to measure effectiveness in economic evaluations. For the ultimate aim of economic evaluations, that is to inform about the optimal allocation of (scarce) resources [43], preference-based generic measures of QoL are usually considered most appropriate, as they allow for comparability across diseases or indications [45, 52]. The most widely used instrument is the EQ-5D, a health-related quality of life (HrQoL) questionnaire covering the dimensions “mobility”, “self-care”, “daily activities”, “pain/discomfort”, and “anxiety/depression” [57]. Depending on the version, respondents rate their health in each dimension on a three- (EQ-5D-3L) or five-level scale (EQ-5D-5L). By weighting these answers using preference-based value sets [58, 59], the answers can be transformed to an index with 1 indicating “perfect health”, 0 indicating “death”, and negative values indicating health states considered worse than death. This index, in turn, can be used to adjust the length of life or an observed time period for QoL. Hence, cost-effectiveness results are often expressed as costs per QALY, with one QALY being equivalent to one year in “perfect health”. The preferences used for weighting the severity levels are typically those of the general population, as they are often those that indirectly pay for healthcare interventions via tax or insurance contributions [52]. In Germany, however, the Institute for Quality and Efficiency in Health Care (IQWiG) deviates from international standards in several respects. For example, it only recommends QALYs as primary outcome measure in economic evaluations if preference weights are derived from the (patient) population of interest, and should be complemented by disease-specific QoL instruments [60, 61]. This is reasoned, e.g., by the German social legislation and weaknesses of the QALY concept itself (e.g., being discriminatory against people with disability or older populations).

Furthermore, an important aspect are the measurement properties (e.g., validity, reliability, responsiveness) of generic HrQoL instruments in the population of interest of an economic evaluation. Measurement properties can be broadly categorized as validity (the instrument measures what it intends to measure), reliability (the measurement is accurate and reproducible under the same conditions), and responsiveness (ability to detect important changes in the construct of interest over time) [62]. Insufficient measurement properties result in a potential underestimation of (cost-)effectiveness of an intervention and thus, false conclusions regarding the efficient allocation of resources may be drawn.

As described earlier, in principle, any measurement unit can be used for reporting cost-effectiveness results. However, using clinical outcomes (e.g., prevented falls, increased PA) in economic evaluations to measure the effect of an intervention has also disadvantages, such as the limited comparability to competing interventions from other (disease) areas or the intermediate nature of many outcomes, which are only useful if adequate predictions of long-term or final end points is possible (e.g., how does a certain increase of PA impact QoL or survival time in the long-term) [45]. In addition, the lack of defined threshold values – the maximum willingness to pay for one additional unit of effect – limits the interpretability of the results when an intervention is more effective but also more costly than the comparator.

Thus, the question of the acceptable WTP for, e.g., avoiding a fall or increasing PA by a certain level arises, which is crucial for determining whether an intervention can be considered cost-effective or not. However, the question of an adequate cost-effectiveness threshold does not only concern clinical endpoints but also QALYs, as only few countries formally defined such a threshold (e.g., the UK with £20,000 to £30,000 per QALY gained [63]).

1.1.4.4 Expanding cost-effectiveness frameworks – Willingness to pay as a tool to elicit preferences beyond conventional measures of effectiveness

Conventional measures of effectiveness in economic evaluations, such as QALYs based on the EQ-5D, mainly focus on benefits of an intervention in specific dimensions of health. It can be argued that solely considering health benefits in implementation decisions ignores the potential non-health benefits of interventions, which could lead to suboptimal allocation of resources [64]. This could be of particular relevance in healthcare systems where decisions are to be aligned with the preferences of the "users" or which aim to maximize overall well-being rather than just health in a narrow sense [64]. Assessing the WTP of an individual for one or multiple competing interventions is a tool for measuring the value or the strength of preference respondents assign to certain interventions. This includes aspects beyond health, which could also be process attributes (e.g., how an intervention is delivered) [45, 65]. Whereas cost-effectiveness results expressed as costs per QALY describe the amount of (public) resources that needs to be invested to improve health, WTP describes the amount of money an individual is willing to give up for its own consumption of the resource (e.g., a fall prevention program) [45]. Especially in the field of health prevention (e.g. exercise-based fall prevention programs), in countries such as Germany, many program costs are not completely covered by the health insurance, but only subsidized. Thus, assessing the WTP could also be used by decision-makers to set or increase co-payments, which of course has implications for equitable access to health(care) and could contradict the principle of allocating resources based on need rather than ability to pay [45, 65].

Although it is acknowledged by Health Technology Assessment Agencies, e.g. in the UK [63], that beyond-health aspects or so-called "process utility" are important, there is currently no consensus on how to integrate patient preferences in terms of WTP into cost-effectiveness frameworks [45, 64]. Challenges include (1) the risk of "double counting", as benefits measured by generic preference-based instruments and benefits beyond health are not always easy to disentangle, and (2) the lack of harmonized methods to assess WTP (or "process utility" in a broader sense), as different methods lead to different results [45, 64, 65].

One frequently used method to elicit WTP are payment cards [66], where the respondents are presented with a range of values (prices) and asked to indicate how much they would definitely be willing to pay and how much they would definitely not be willing to pay for a certain intervention. In this thesis, the payment card method was used in the context of an RCT comparing different treatment modalities/versions of an intervention that have similar (cost-)effectiveness where it can give an idea of "what do participants prefer?" and "how valuable do they perceive the intervention?".

1.1.5 Contribution of the thesis to the field of research

This thesis addressed the topic of promoting healthy aging through cost-effective strategies using the economic analysis of an exercise-based activity promotion and fall prevention program as an example. Thereby, it contributes to the field of economic evaluations of exercise-based fall prevention programs for older, community-dwelling people, overall but particularly in the German context where research so far is very limited. In addition, a special feature is that two interventions are compared, that have the same content but differ in delivery mode (group versus individual) and costs of implementation.

The economic evaluation was complemented by an analysis of the association between FoF and HrQoL (an important endpoint in economic evaluations) as well as potential mediators of this association. While the existence of the association between FoF and (Hr)QoL is well-established [30], only few studies examined factors explaining this association. These studies found physical function (an important risk factor for FoF [67]), PA, and the self-concept of health and physical independence being mediators of the association between FoF and QoL [68, 69]. Knowledge of explanatory factors may help addressing FoF in interventions. This, in turn, could be important for improving the (cost-)effectiveness of activity promotion and fall prevention programs since FoF is linked to HrQoL and PA [30, 31].

Beyond this, the thesis contributes to the discussion of appropriate effect measures in health economic evaluations (especially for older age groups and activity promotion and fall prevention programs) in several ways: Using the health economic analysis of the LiFE-is-LiFE study as an example, the influence of the choice of effect measure on the interpretation of cost-effectiveness was explored. Furthermore, the preferences of the participants were examined in more detail and the potential usefulness of WTP as complementary information to conventional CEAs in implementation decisions was discussed. Finally, a systematic review on the measurement properties of the EQ-5D, the most widely used instrument in economic evaluations, in a middle-old to oldest-old population was conducted.

1.2 Objectives

The overall aim of this thesis was to investigate the cost-effectiveness of interventions aimed at improving care for the older population using the example of an activity promotion and falls prevention program for older people. The results should support the decision-making process regarding a possible implementation of the program.

This aim was addressed by three different sub-projects, the results of which were presented in six publications included in this cumulative thesis. The specific objectives of the individual sub-projects were as follows:

In *project 1* (publication 1), the objectives were (1) to examine the association between FoF and HrQoL (EQ-5D and EQ-VAS) and (2) to analyze mediating factors of the association between FoF and HrQoL.

Project 2 comprised the trial-based economic analysis of the LiFE-is-LiFE project. The objectives were (1) to examine the cost-effectiveness of a group version of an activity promotion and falls prevention program compared to the individually delivered version (Publication 2 & 3), and (2) to explore the participants' WTP for the group and the individual program, to examine factors influencing WTP, and to examine whether the perceived benefits – operationalized as WTP – exceed the costs associated with conducting the intervention(s) (publication 4).

In *project 3,* the objective was to synthesize and critically appraise studies assessing the measurement properties – reliability, validity, or responsiveness – of the EQ-5D in a population of middle-old and oldest-old people (publication 5 & 6).

1.3 Material and methods

1.3.1 Empirical studies

1.3.1.1 Framework

Project 1 and 2 (publications 1-4) were conducted alongside the LiFE-is-LiFE study, a multi-center, single-blinded, randomized non-inferiority trial funded by the German Ministry for Education and Research (BMBF, grant number 01GL1705A-D) [70]. The aim of the study was to evaluate a group version of the *Lifestyle-integrated Functional Exercise* program (LiFE) regarding its non-inferiority compared to the original, individually delivered program version. To this end, community-dwelling older people (aged ≥70 years) at risk of falling but still able to execute the exercises taught in the intervention were recruited between 2018 and 2019 from two study sites (Stuttgart and Heidelberg, Germany). Overall 309 participants were randomized into the group (gLiFE, n=153) and individual program version (LiFE, n=156) and followed up over a 12-month time horizon. Data were collected in face-to-face interviews at baseline, 6- and 12 months. The intervention consisted of strength activities, balance activities, and general PA that can be integrated into everyday life tasks. In gLiFE, these activities were taught by two trainers in seven group sessions, whereas in LiFE the activities were taught in a 1:1 setting in form of seven home visits. Detailed inclusion and exclusion criteria as well as description of the intervention(s) can be found in the study protocol [70] and other related publications [71, 72]. In the primary analysis of the study, non-inferiority was determined as activity-adjusted fall rate over the 12-month time horizon. Although, compared to baseline, an increase of PA (operationalized as mean steps per day) and a reduction in the falls incidence was found in both groups, non-inferiority remained inconclusive with the upper level of the CI crossing the 20% non-inferiority margin [73] [74].

1.3.1.2 Materials/Variables/Instruments

HrQoL was assessed using the EQ-5D-5L [75]. The German value set [76] was used to derive the EQ-5D index from participants' answers on the five dimensions. For the economic evaluation, QALYs were calculated for the respective observation period (6 or 12 months) by linearly interpolating the EQ-5D indices of the assessment time points. In addition, participants rated their overall current health status on a visual analogue scale (EQ-VAS) ranging from 0 (worst imaginable health state) to 100 (best imaginable health state).

PA data was collected for seven full days at each assessment time point using the "activPAL4 micro" accelerometer (PAL Technologies Ltd, Glasgow, Scotland). For the purpose of this study, PA was operationalized as the mean number of steps per day.

The *number of falls* was self-reported by the participants in form of falls calendars, which were returned to the respective study center each month. A fall was defined as "an unexpected event in which the participants come to rest on the ground, floor, or lower level" [77].

FoF was assessed using the Short Falls Efficacy Scale International (Short FES-I) [78], which consists of seven items, each with four response levels, combined into a total score, with higher scores indicating higher FoF.

Subjective functional capacity was measured with the Late Life Function and Disability instrument (LLFDI) [79, 80], a questionnaire consisting of a function (ability to perform discrete actions) and a disability component (limitations in performing specific life tasks). Individual-item responses were summed and transformed to a scale between 0 (no limitations) and 100 (high limitations) for each component.

Resource use and costs: Information on health-related resource use (medications; outpatient services; inpatient stays in general hospitals, rehabilitation and psychiatric clinics; formal and informal care) was collected retrospectively from participants at all assessment time points using a questionnaire for the use of medical and nonmedical services in old age (FIMA) [81]. This information was monetized based on standardized unit costs and pharmacy retail prices [82, 83]. All costs were reported in 2018 Euros (€). Given that all participants were ≥ 70 years old, indirect costs (e.g., productivity losses) were not considered.

For the calculation of *intervention costs,* personnel and material costs, travel expenses, and room rent were taken into account. The information was derived from study documentation. Furthermore, an additional intervention cost scenario ("real world") was calculated assuming lower trainer wages and larger gLiFE group sizes to more realistically reflect costs under implementation conditions. Detailed assumptions underlying the calculation of intervention costs can be found in the supplementary material of publication 2.

Participants' WTP was assessed at the 12-month assessment by payment cards [66], where participants stated the amount of money they would surely be willing to pay as well as the amount they would definitely not be willing to pay for one training session. The midpoint between these two values was multiplied by seven (the number of training sessions).

Other variables used for adjustment purposes in the different analyses were sociodemographic characteristics (age, sex, educational status, income, marital status, living situation, health insurance status), the number of chronic conditions (diabetes type 1 and 2, hypertension, acute cardiovascular disease, history of heart attacks, cardiac defect, auricular fibrillation or other cardiac arrhythmias, history of stroke, rheumatoid arthritis, cancer, asthma or chronic obstructive pulmonary disease, osteoporosis, depression), functional mobility (Timed Up-and-Go Test), problems being physically active due to the COVID-19 pandemic, and variables related to participants' experience with the intervention, such as motivation to exercise (BREQ-3 [84]), satisfaction with the program, and training frequency (number of LiFE activities performed per week).

1.3.1.3 Statistical analyses

In *project 1* (publication 1), baseline data from the LiFE-is-LiFE trial was used to conduct a cross-sectional study examining the association between FoF and HrQoL as well as factors mediating this association. Linear regression models were used to examine the association between FoF and the EQ-5D index and the EQ-VAS. Bootstrapped standard errors (SE) and CIs were calculated to deal with the non-normality of the outcomes. Models were adjusted for sociodemographic characteristics (Model 1) and additionally for the number of chronic conditions, number of falls, functional mobility, and subjective functional capacity (LLFDI function and disability) (Model 2). The mediating effect of subjective functional capacity was examined by path models [85] and significance tests for the indirect effects were examined using the Sobel test [86].

In *project 2*, the cost-effectiveness of gLiFE compared to LiFE was examined from a societal and payer's perspective and using QALYs, PA, and number of falls as effect measures (publication 2 & 3). In the analyses from a payer's perspective, costs of informal care and gLiFE participant's travel expenses to the training sessions were excluded. The base case analysis was based on the intention-to-treat population, meaning all 309 randomized participants were included in the analysis. As described earlier, missing data was assumed to be MAR and was imputed on disaggregated level (item-level) by MICE with predictive mean matching as imputation method [50, 51]. Due to the COVID-19 pandemic, the 12-month assessments were partially delayed, so that in order to eliminate possible pandemic-related biases, it was decided to analyze the primary study endpoints (activity-adjusted falls incidence and cost-effectiveness) already after 6 months. However, as no relevant pandemic effects on the main parameters of interests (costs, HrQoL, PA, falls) could be observed, it was deemed justified to conduct the 12-month CEA as originally planned using the ITT sample.

At both follow-up time points (6- and 12-month), mean total costs and effects were compared between gLiFE and LiFE and the ICERs, expressed as costs per QALY, costs per 1,000 additional steps per day, and costs per fall prevented, were calculated. Since the WTP for an additional unit of effect was unknown, and in order to visualize uncertainty around the ICER estimates and the dependency of the cost-effectiveness from the WTP, CEACs were constructed based on the net-benefit approach [53, 54]. These were derived from net-benefit regressions [55] in which group differences in the NMB_i were estimated, adjusted for baseline characteristics (HrQoL, PA, number of falls, costs, chronic conditions) and for problems being physically active due to the COVID-19 pandemic at 12 months. Sensitivity analyses were conducted restricting the analyses to a per-protocol sample (attendance of ≥5 training sessions, n=280) or including the "real world" intervention costs.

The analysis of the WTP from the participants' perspective (publication 4) was based on participants who filled out the payment card at 12 months (n=237). A linear regression model was estimated with WTP as dependent variable and the group variable (gLiFE/LiFE) as well as sex, age, income, number of chronic conditions, healthcare costs, and motivation to exercise as potential determinants. These determinants were selected based on descriptive and bivariate analyses. Non-normality of the data was addressed by estimating the models and respective CIs based on 1,000 bootstrapped samples. The WTP was contrasted with different "real world" intervention cost scenarios (assumption of different, hypothetical subsidies [€0, €50, €75] to the intervention costs by, e.g., a health insurer) by subtracting the intervention costs from the WTP (= net benefit).

1.3.2 Systematic review

In *project 3*, a systematic review on the measurement properties of the EQ-5D in a middle-old to oldest-old population was conducted under the Consensus-Based Standards for the Selection of Health Measurement Instrument (COSMIN) framework [87] (publication 5 & 6). Relevant literature was systematically searched in the databases PubMed, Web of Science, Cochrane Library, Embase, and EconLit. Articles were included if they provided evidence of test-retest reliability, construct validity (convergent/known-groups), and/or responsiveness, reported results for a population with a mean age ≥75 years, and were published in German or English. Articles other than original studies, qualitative studies, those where the EQ-5D was not self-reported by the population of interest (e.g., proxy-reported), and those that only assessed inter-rater reliability were excluded. Selected studies were evaluated for methodological quality based on the COSMIN Risk of Bias Checklist [88]. For the evaluation of measurement properties as “sufficient”, “insufficient” or “inconsistent”, the criteria for good measurement properties were applied [89]. To this end, hypotheses regarding the direction and strength of the association between the EQ-5D and the comparator instrument were formulated in advance for construct validity and responsiveness. The measurement property was rated as “sufficient” or “insufficient” if ≥75% of the individual studies’ results fulfilled or did not fulfill the criterion. Finally, the quality of the evidence was graded as “high”, “moderate”, “low”, or “very low” based on methodological quality, inconsistency of results, and overall sample size [90, 91].

Study selection, rating of methodological quality, and evaluation of measurement properties was undertaken by two independent reviewers. Extraction of relevant data from the individual studies was done by one person and cross-checked by a second person. Any disagreements were resolved by consulting a third person.

Project 3 was conducted independently from project 1 and 2 and received no specific funding.

1.4 Results and discussion of individual studies

1.4.1 Mediating factors of the association between FoF and HrQoL

Results. The association between FoF and HrQoL, as frequently reported in the literature [30], could be replicated in *project 1* (publication 1): A significant negative association between FoF and the EQ-5D index ($\beta = -0.02, p < 0.001$) as well as between FoF and the EQ-VAS ($\beta = -1.54, p < 0.001$) could be found in the model adjusted for sociodemographic variables (Model 1). In the fully adjusted model (Model 2), the association remained significant for the EQ-5D index ($\beta = -0.01, p < 0.01$), but not for the EQ-VAS ($\beta = -0.36, p > 0.05$). Mediation analyses suggested a partial (FoF and EQ-5D via function/disability, FoF and EQ-VAS via disability) or complete mediation (FoF and EQ-VAS via function) of the association between FoF and HrQoL by subjective functional capacity, indicated by a decrease in strength of the association after inclusion of the mediator in the respective model.

Discussion. The results suggest that the association between FoF and HrQoL may be explained by subjective functional capacity, thereby extending the current literature, which has rarely looked at factors explaining the association between FoF and HrQoL. This mediating effect may be inherent in the way FoF was measured in this study, as the FES-I asks respondents to rate their FoF in typical everyday activities. The cross-sectional nature of the study limits drawing causal inferences, but a potential explanation could be that FoF causes avoidance behavior, which can lead to a deterioration in functional capacity, which in turn is associated with lower (Hr)QoL and an increased risk of falling [32, 92-95].

The multidimensionality of HrQoL and the rather small explanatory effects of FoF and subjective functional capacity found in this study indicate that focusing on reducing FoF alone may not lead to clinically important changes in HrQoL and hence, strategies to reduce FoF should be integrated into broader (exercise-based) fall prevention interventions. Nevertheless, the results stress the importance of daily and social life activities in older people for maintaining HrQoL, which implies that strategies should be taught on how to perform daily activities safely despite FoF. Therefore, interventions such as the LiFE program, which integrates exercises into everyday life tasks, have the potential to tackle this postulated causal pathway and thus maintain or improve older peoples' HrQoL.

Due to the design of the study, reverse causality cannot be excluded (e.g., functional limitations cause FoF and both are associated with HrQoL). Furthermore, the simple structure of the path models is unlikely to reflect the true complexity of the association. These limitations may be addressed in future studies by analyzing more complex path models or using longitudinal data. Given that the analyses were based on a RCT population where, e.g., certain individuals were deliberately excluded, generalizability to the broader population of older people may be limited.

1.4.2 Economic analysis LiFE-is-LiFE

1.4.2.1 Cost-effectiveness analysis

Results. The 6-month and 12-month CEAs accompanying the LiFE-is-LiFE trial are reported in publication 2 & 3, respectively. Despite lower intervention costs of gLiFE (−€121 [study conditions] and −€212 [real world]), the mean total costs from a societal perspective were non-significantly higher than in LiFE at both time points (+€77, 95% CI [−€743, €896] and +€1,099, 95% CI [−€960, €3,158], respectively). On the effect side, mean differences between gLiFE and LiFE were marginal and/or not significant for all effect measures and at both time points, except for PA at 6 months (+799 steps per day, 95% CI [207, 1,391]). Using QALYs or the number of falls as effect measure, the probability of gLiFE being cost-effective compared to LiFE remained uncertain (50%-94%) or unlikely (<50%) at all WTP threshold values assumed in the CEACs. Regarding PA as effect measure, gLiFE had a ≥95% probability of being cost-effective at 6 months at a WTP per additional 1,000 steps per day of ≥€1,600 from a societal perspective or ≥€600 from a payer's perspective). After 12 months, the probability of gLiFE being cost-effective to increase PA remained unlikely to uncertain – only at relatively high WTP levels probabilities ≥50% could be attained (e.g., at a WTP ≥€1,200 from a societal or ≥€2400 from a payer's perspective). Replicating the analyses with the per-protocol sample or assuming "real world" intervention costs yielded slightly higher probabilities of cost-effectiveness in the CEACs, but overall confirmed the results of the base case analysis.

Discussion. Despite lower intervention costs of gLiFE, no program version was shown to be clearly superior in terms of cost-effectiveness. According to the non-inferiority design of the LiFE-is-LiFE trial, no difference on the effectiveness level was hypothesized in advance, but it was expected that cost-effectiveness of gLiFE compared to LiFE would be demonstrated solely based on the difference in intervention costs. Instead, higher mean costs from health-related resource use were observed in gLiFE compared to LiFE. It remained unclear whether this is attributable to the intervention or rather a random result due to the sample's lack of power to detect differences in costs, which is a common limitation of trial-based economic evaluations [45]. In addition, it remained unclear how gLiFE/LiFE perform in terms of cost-effectiveness compared to a usual care control group, which constitutes a potential future research direction. Given that both program versions lead to clinical improvements in terms of reduced falls and increased PA over the course of the study [73, 74], it can be assumed, that gLiFE/LiFE also have economic benefits over "usual care". Furthermore, future studies could consider a longer time horizon, as the intervention's effect on PA may only unfold at cost and HrQoL level in the long run (e.g., lower healthcare costs and better HrQoL due to prevention or postponed onset of chronic diseases). Determining the long-term economic benefit of small PA increases such as 1,000 additional steps per day in future studies would contribute to a better understanding of what might be a justified WTP for such an increase. Operationalizing PA as steps per day may not sufficiently capture the complexity of PA, but represents an easily interpretable unit that is associated with mortality [96, 97] and can be used to formulate public health recommendations. Another challenge that might be addressed by future studies is how a combined measure of effectiveness, such as activity-adjusted fall rate, could be incorporated in a cost-effectiveness framework (e.g., ICER), which itself represents a ratio.

1.4.2.2 Participants' WTP

Results. The participants' mean WTP for the program was €196 (95% CI [€172, €221]) in gLiFE and €228 (95% CI [€204, €251]) in LiFE (publication 4). Besides the group variable, income was significantly associated with WTP in the regression model. Comparing the WTP with the "real world" intervention costs showed that the mean WTP for gLiFE not only covers but also exceeds the intervention costs (+83€, 95% CI, [59, 107]), while for LiFE it was lower than the intervention costs. When analyzing the distribution of this net benefit for gLiFE and LiFE according to different hypothetical cost subsidies, the majority of gLiFE participants had a positive net benefit (68% [€0 subsidy], 86% [€50 subsidy] and 95% [€75 subsidy]), while this was only the case for 25% (€0 subsidy), 29% (€50 subsidy) and 40% (€75 subsidy) of LiFE participants.

Discussion. Participants' WTP for an intervention represents a measure of value of the intervention that goes beyond health benefits and societal preferences. Since the LiFE-is-LiFE trial compared two different administration modes of an intervention with otherwise similar content (and presumably similar clinical effectiveness), this measure could be of considerable usefulness. The high WTP consolidates the value of both program versions from the participant perspective, which is relevant in the German implementation context, where the costs for many prevention programs are not completely covered but require a co-payment. This study's results indicate that for gLiFE in particular, the necessary co-payments seemed to be in a range that participants consider justified or a good investment – even without a subsidy, the WTP exceeded the intervention costs in the majority of the sample, but also for LiFE there seemed to be a number of people who are willing to pay the co-payments. Future studies could explore what determines the choice between gLiFE and LiFE (e.g., in a discrete choice experiment), as this remained open in this study owing to the fact that participants were only asked about their WTP for the program version they received. This knowledge could in turn be used to identify specific target groups for the intervention(s). In addition, future studies could further explore the factors that determine WTP for gLiFE/LiFE, as the potential determinants investigated in this study had low explanatory power. Other limitations are related to the sample and the context in which the WTP was obtained from the participants, which may inhibit transferability of results to other contexts (e.g., participants with certain characteristics self-selected into the study and in the study context participants did not actually have to pay for the intervention). The results are still useful as a complement to the CEA because they provide information on the value of the program(s) from a participant perspective that is not bound to predefined dimensions or response options.

1.4.3 Systematic review

Results. The results of the systematic review on the measurement properties of the EQ-5D in populations with a mean age ≥75 years (publication 5 & 6) were mixed depending on the particular measurement property and version of the EQ-5D (3-level or 5-level). Overall, the EQ-5D seemed to have sufficient convergent validity, with the majority of results being in accordance with the hypotheses regarding the strength of the association with comparison instruments. Regarding the ability to differentiate between known-groups (known-groups validity), results were inconsistent for the EQ-5D-3L but sufficient for the EQ-5D-5L. Notably, the EQ-5D seemed to have questionable reliability and responsiveness to change, even though the 5-level version seemed to perform better, but the evidence base to date is weak (responsiveness) or entirely lacking (reliability).

Discussion. Based on the sufficient construct validity of the EQ-5D, it seems suitable for describing HrQoL (e.g., in cross-sectional studies). However, if the EQ-5D is to be used to measure changes in HrQoL (e.g., in economic evaluations), caution is warranted, as at least the EQ-5D-3L was insufficiently responsive and the results were inconsistent in terms of reliability. The problems with responsiveness may be related to the content of the EQ-5D that does not capture all relevant aspects of older people's (Hr)QoL [98-101]. Another potential explanation could lie in the valuation of the individual dimensions' response levels, meaning that the preferences of older people may not be sufficiently represented in the general population's preferences, which are normally used for the valuation [102]. Overall, more evidence is needed, especially for the EQ-5D-5L, and generalizability of the results of this review may be limited by the fact that only few studies exclusively included people aged ≥75 years. There are a number of alternative instruments that have been or are being developed that aim to capture (Hr)QoL or well-being more broadly (e.g., PROMIS [103, 104], EQ-HWB [105]) or that focus specifically on older people (e.g., QOL-ACC [106], ICECAP-O [98]). Their applicability in economic evaluations is so far limited by the partial lack of country- or population-specific tariffs, which have yet to be developed. In addition, their different focus limits the comparability of results between different measures and population groups. Therefore, the EQ-5D may still be used for economic evaluations in populations aged ≥75 years, but should be complemented by age- or disease-specific instruments that can be used to interpret the cost-effectiveness results based on the EQ-5D.

1.5 General discussion

In the context of the pressing challenges of an aging society to the sustainability of healthcare systems and the resulting political goal of promoting healthy aging through cost-effective interventions, this thesis dealt with the economic evaluation of such interventions using the example of an activity-promotion and fall prevention program.

The existing state of research on the association between FoF and HrQoL – the latter being an important outcome used in economic evaluations – could be extended by an analysis of factors mediating this association. The results confirmed the negative association between FoF and HrQoL previously reported by other studies [30] and showed that this association can (partly) be explained by subjective functional capacity (functional limitations and disability in performing specific life tasks). However, it should be noted that the magnitude of the association between FoF (via subjective functional capacity) and HrQoL was rather weak, which can be explained by the multidimensional nature of (Hr)QoL. Hence, focusing on alleviating FoF and reducing limitations in functional capacity in interventions may not be sufficient to generate clinically relevant changes in HrQoL.

The economic evaluation accompanying the LiFE-is-LiFE non-inferiority trial contributed to the literature on economic evaluations of exercise-based fall prevention programs for community-dwelling older people [37], which has been scarce, especially in the German context. It was tested whether gLiFE (the group-delivered version) represents a cost-effective alternative to LiFE (the individually delivered version), meaning that it is similarly effective but at lower cost. Indeed, the "pure" intervention costs for gLiFE were lower than for LiFE, but after taking the costs from resource use into account, superiority of gLiFE over LiFE in terms of cost-effectiveness could not be shown. These results underline the importance of not focusing solely on intervention costs in an economic evaluation, even when non-inferiority in terms of effectiveness and thus no difference at the level of healthcare costs is assumed. Even if small effect losses compared to the comparison group are accepted (e.g., the non-inferiority margin was set to 20% in LiFE-is-LiFE), potential negative consequences of these effect losses should be examined at the level of healthcare costs, as they may outweigh the initial cost savings (e.g., lower intervention costs in gLiFE). The necessity of conducting a full CEA in non-inferiority trials is further supported by the fact that, despite the overall conclusion that no program version dominated the other in terms of cost-effectiveness, the results were sensitive to the choice of effect measure. On the one hand, PA was unexpectedly higher in gLiFE than in LiFE, leading to high probabilities of cost-effectiveness of gLiFE at certain WTP levels, especially at 6 months, where the difference was more pronounced than at 12 months. On the other hand, the CEAC pointed towards dominance of LiFE when QALYs were used as effect measure at 12 months (although the QALY difference was marginal and not significant).

The CEA was complemented by an analysis of the participants' WTP for the program, which consolidated the perceived value of both versions of the intervention from a participant perspective that was, e.g., not captured on HrQoL level. LiFE participants' WTP was on average somewhat higher than the WTP of gLiFE participants (+€32), but assuming that the intervention costs are at least partly borne by the participants themselves (as frequently the case with prevention programs in Germany), only in gLiFE the majority had a WTP high enough to cover the intervention costs. However, other studies from the LiFE-is-LiFE trial showed that both versions are well-accepted and comparable with respect to content evaluation [107, 108], and based on the WTP analysis a small proportion of people may still be willing to pay enough to cover the intervention costs of LiFE. Therefore, no clear recommendation for implementing the one or the other version was derived. Nevertheless, identifying participants' WTP can be useful in informing implementation decisions in contexts where two competing interventions have similar effectiveness and similar (intervention) costs.

The thesis also contributed to the discussion on adequate endpoints in economic evaluations by a systematic review on the measurement properties of the EQ-5D, the most widely used HrQoL instrument in economic evaluations [109]. The results highlighted weaknesses of the measurement properties of the EQ-5D in a middle-old to oldest-old population, especially with regard to responsiveness to change and reliability, but with the caveat that the evidence base for the EQ-5D-5L is still weak. The results have implications for the results of the other studies in this thesis. For example, the EQ-5D may not reflect all relevant aspects of (Hr)QoL that are impacted by FoF, or the mediating effect of function and disability on the association may be explained solely by the strong focus of the EQ-5D on functional dimensions of health [100]. Therefore, the "true" strength of the association between FoF and HrQoL may differ from the size of the coefficients found in the study (e.g., be stronger or not as strong). For the CEA, insufficient responsiveness of the EQ-5D could mean that the intervention effect (and possibly also the cost-effectiveness) were underestimated. For example, the clinical changes in terms of an increased PA and a reduced rate of falls [73, 74] may not be reflected at the level of HrQoL, although an impact of these changes on HrQoL can be assumed [16, 25].

The results of this thesis can also be placed in the general discussion on the adequacy of focusing on HrQoL or a too narrow definition of health in economic evaluations [100, 110, 111], which has a special relevance in older populations where factors other than health (e.g., ability to perform usual activities or social contact) become more important for (Hr)QoL [101]. To address these short-comings, several other instruments have been developed which potentially better capture (Hr)QoL or intervention benefits and thus provide an alternative for measuring the quality-component of a QALY (e.g., EQ-HWB [105] or PROMIS [103, 112]). Using HrQoL/QALYs to measure effectiveness in economic evaluations has been held on to so far due to the generic nature of this outcome, which allows for comparability of interventions across indications [45, 52]. The need for this comparability is argued with the ultimate goal of economic evaluations to inform on the optimal allocation of resources [43]. Having a variety of instruments to choose from in economic evaluations can limit this comparability as the choice of the HrQoL instrument is likely to lead to different QALY and cost-effectiveness estimates, and hence the assumption "a QALY is a QALY is a QALY" [113] no longer holds. This also limits the comparability of interventions against a single cost-effectiveness threshold for a QALY gain. However, in Germany, for example, this comparability across indications is not a prerequisite for deciding on the implementation or reimbursement of a new intervention. Due to the IQWiG's hesitation to recommend QALYs as preferred outcome measure [60, 61] and the generally lower relevance of health economic aspects in implementation and reimbursement decisions in Germany, there is also no fixed cost-effectiveness threshold. However, not knowing what is considered a justified investment for a certain unit-increase in effect impedes drawing conclusions or making judgements regarding whether an intervention can be interpreted as "cost-effective" in cases where a new intervention is more effective but also (initially) more costly than the comparator. A potential solution could be to determine cost-effectiveness thresholds for alternative outcomes to the QALY based on the (predicted) long-term economic consequences of clinical outcomes. Keeping with the example from the LiFE-is-LiFE trial: knowing how a certain increase in PA (e.g., by 1,000 steps per day) affects healthcare costs in the long run (e.g., by preventing or postponing chronic diseases) would give an idea of what a justifiable investment aka cost-effectiveness threshold is for the respective unit of effect.

1.5.1 Limitations and implications for future research

Based on cross-sectional data and using simple path models, publication 1 showed that the association between FoF and HrQoL was mediated by subjective functional capacity. Future studies could build on these findings and further explore the association of FoF and (Hr)QoL as well as factors explaining this association in more detail by using alternative (Hr)QoL instruments, longitudinal data, and/or structural equation models that examine more complex causal pathways. These results, in turn, may then be considered in the design of activity promotion and fall prevention strategies, where explanatory factors could be addressed and thereby may intensify the interventions' effect on HrQoL.

The limitations of the economic evaluation of the LiFE-is-LiFE program (project 2) are reflected in the typical challenges of economic evaluations of disease prevention, health promotion, or behavior change interventions that relate to attribution of effects, measurement and valuation of outcomes, identification of intersectoral costs and consequences, and equity considerations [114-117]. Due to the non-inferiority design of the LiFE-is-LiFE trial, the economic analysis was restricted to the 12-month time horizon of the trial and to the comparison between gLiFE and LiFE. Analyzing the cost-effectiveness of gLiFE and LiFE compared to a "usual care" control group and over a longer time horizon would contribute to the understanding of the economic value of the interventions. Since conducting RCTs over a very long time horizon is often not feasible, different study designs may be used to establish relationships between intermediate outcomes (e.g., an increase in PA) and long-term outcomes (e.g., development of diseases or HrQoL), which in turn could be used to model cost-effectiveness beyond a trial's time horizon. Related to this, determining the long-term consequences of intermediate outcomes on the cost level could be used to determine a cost-effectiveness threshold for intermediate outcomes, such as a certain increase in PA [118, 119]. Besides examining which alternative instruments of (Hr)QoL are more suitable than the EQ-5D for economic evaluations in an older population, a further research perspective could be to find ways to incorporate benefits beyond the individual into cost-effectiveness frameworks. This could be especially relevant in older people, as e.g., maintaining functional independence through exercise-based falls prevention has also effects on the wider network (e.g., family or caregivers).

Furthermore, it should be examined how the cost-effectiveness of gLiFE/LiFE differs by person characteristics as this could help in targeting the intervention to sub groups with a higher benefit from the intervention (e.g., people with a high risk of falling). Net benefit regression models are well suited for such analyses [55].

1.5.2 Implications for decision-makers and practice

Despite the limitations and challenges, the results of this thesis have implications for decision-makers and practice. The results of project 1 indicated that ensuring the maintenance of daily activities and participation in social life contributes to HrQoL and partially explains the association between FoF and (Hr)QoL. This implies that older people should be equipped with strategies that enable them to manage their daily and social life despite their FoF, e.g., through exercise-based fall prevention strategies integrated into daily life, as done in the LiFE program.

From project 2 it can be concluded that participants of both program versions perceive the intervention as valuable, which is reflected in a high WTP and is also supported by studies that evaluated the content and acceptability of gLiFE/LiFE [107, 108]. This subjectively perceived value is also underscored by the clinical effectiveness in terms of a reduction in the rate of falls and an increase in PA in both groups, even though non-inferiority of gLiFE remained inconclusive by statistical definition (which does not automatically mean that gLiFE is inferior) [73, 74]. Thus, both program versions can potentially contribute to healthy aging. When only considering the costs of delivering the intervention, gLiFE was indeed less resource intensive than LiFE. However, when also taking the cost of health-related resource use into account, gLiFE overall tended to incur higher costs over 12 months, and cost-effectiveness was uncertain compared to LiFE (but this does not mean that LiFE was superior to gLiFE). It should be noted that long-term consequences (e.g., in terms of health-related resource use and costs) remained unclear, especially regarding the performance of gLiFE/LiFE compared to no intervention ("usual care"). An analysis that modelled the 5-year budget impact of implementing gLiFE/LiFE from a German health insurance perspective showed that cost savings compared to "usual care" are only possible if the intervention effect on the rate of falls was increased, the intervention costs were reduced, or the intervention was offered only to people with a high risk of falling [120]. However, the budget impact model considered only the economic consequences of falls. Given that gLiFE and LiFE also had an effect on PA, and depending on the sustainability of this effect, the implementation of gLiFE/LiFE could have a lower budget impact or even save costs. Factors that further determine the actual budgetary consequences are the way the intervention is delivered (by whom, in what setting, at what costs?) and how it is financed (full reimbursement by a health insurer versus co-payments). Therefore, reducing the cost of interventions that have been proven to be effective remains important. Developing group versions that require fewer human and financial resources while maintaining comparable effectiveness is therefore an attractive option, also in light of the health workforce shortage. Assuming that, as common practice in Germany, the costs for (certified) prevention courses are not fully covered by e.g. the health insurance anyway and co-payments have to be made by the participants, decision-makers could also consider offering both program versions so that participants can choose between gLiFE and LiFE according to their preferences.

Overall, it can be concluded that both program versions can potentially contribute to the national health goal “healthy aging”, but based on the results of the economic evaluation, no clear recommendation can be made for implementing either gLiFE or LiFE. Due to the lack of a “usual care” control group, it cannot be excluded that the benefits of gLiFE/LiFE can only be achieved at an additional financial costs. This means that investments may be required for which, based on the results of this thesis, it is unclear whether they will be offset by cost savings at the health care cost level in the long run.

2 References

1. Statistisches Bundesamt. Bevölkerung Deutschlands bis 2060: Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung – Hauptvarianten 1 bis 9. 2019. Date accessed: 11.08.2022. Available from: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsvorausberechnung/Publikationen/Downloads-Vorausberechnung/bevoelkerung-deutschland-2060-5124202199014.pdf?__blob=publicationFile.
2. United Nations, Department of Economic and Social Affairs, Population Division. World population prospects 2022. Summary of results. New York: United Nations; 2022.
3. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020. Available from: https://vizhub.healthdata.org/gbd-results/.
4. Statistisches Bundesamt (Destatis). Krankheitskosten, Krankheitskosten je Einwohner: Deutschland, Jahre, Geschlecht, Altersgruppen. Date accessed: 27.11.2022. Available from: https://www-genesis.destatis.de/genesis/online?operation=previous&levelindex=2&levelid=1669563764832&levelid=1669563655198&step=1#abreadcrumb.

5. Breyer F, Lorenz N. The “red herring” after 20 years: ageing and health care expenditures. Eur J Health Econ. 2021 2021/07/01;22(5):661-7.
6. Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: a red herring? Health Econ. 1999 Sep;8(6):485-96.
7. Kollerup A, Kjellberg J, Ibsen R. Ageing and health care expenditures: the importance of age per se, steepening of the individual-level expenditure curve, and the role of morbidity. Eur J Health Econ. 2022 Jan 17.
8. Beard JR, Officer A, de Carvalho IA, Sadana R, Pot AM, Michel JP, et al. The World report on ageing and health: a policy framework for healthy ageing. Lancet. 2016 May 21;387(10033):2145-54.
9. United Nations General Assembly. The Universal Declaration of Human Rights (UDHR). New York: United Nations General Assembly; 1948.
10. United Nations General Assembly. Transforming our world: the 2030 Agenda for Sustainable Development, 21 October 2015, A/RES/70/12015.
11. Bundesministerium für Gesundheit. Nationales Gesundheitsziel - Gesund älter werden. Berlin: Bundesministerium für Gesundheit; 2012.
12. Bergen G, Stevens MR, Burns ER. Falls and Fall Injuries Among Adults Aged ≥65 Years - United States, 2014. MMWR Morb Mortal Wkly Rep. 2016 Sep 23;65(37):993-8.
13. Haagsma JA, Olij BF, Majdan M, van Beeck EF, Vos T, Castle CD, et al. Falls in older aged adults in 22 European countries: incidence, mortality and burden of disease from 1990 to 2017. Inj Prev. 2020;26(Suppl 2):i67.
14. James SL, Lucchesi LR, Bisignano C, Castle CD, Dingels ZV, Fox JT, et al. The global burden of falls: global, regional and national estimates of morbidity and mortality from the Global Burden of Disease Study 2017. Inj Prev. 2020 Oct;26(Supp 1):i3-i11.
15. Heinrich S, Rapp K, Rissmann U, Becker C, König HH. Cost of falls in old age: a systematic review. Osteoporos Int. 2010 Jun;21(6):891-902.
16. Hartholt KA, van Beeck EF, Polinder S, van der Velde N, van Lieshout EM, Panneman MJ, et al. Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life. J Trauma. 2011 Sep;71(3):748-53.
17. Florence CS, Bergen G, Atherly A, Burns E, Stevens J, Drake C. Medical Costs of Fatal and Nonfatal Falls in Older Adults. J Am Geriatr Soc. 2018 Apr;66(4):693-8.
18. Kohl HW, 3rd, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, et al. The pandemic of physical inactivity: global action for public health. Lancet. 2012 Jul 21;380(9838):294-305.

19. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012 Jul 21;380(9838):219-29.
20. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet. 2016 2016/09/24/;388(10051):1311-24.
21. Dallmeyer S, Wicker P, Breuer C. How an aging society affects the economic costs of inactivity in Germany: empirical evidence and projections. Eur Rev Aging Phys Act. 2017 2017/10/17;14(1):18.
22. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ. 2019 Aug 21;366:l4570.
23. Marquez DX, Aguiñaga S, Vásquez PM, Conroy DE, Erickson KI, Hillman C, et al. A systematic review of physical activity and quality of life and well-being. Transl Behav Med. 2020;10(5):1098-109.
24. Daskalopoulou C, Stubbs B, Kralj C, Koukounari A, Prince M, Prina AM. Physical activity and healthy ageing: A systematic review and meta-analysis of longitudinal cohort studies. Ageing Res Rev. 2017 2017/09/01/;38:6-17.
25. Cunningham C, R OS, Caserotti P, Tully MA. Consequences of physical inactivity in older adults: A systematic review of reviews and meta-analyses. Scand J Med Sci Sports. 2020 May;30(5):816-27.
26. Bauman A, Merom D, Bull FC, Buchner DM, Fiatarone Singh MA. Updating the Evidence for Physical Activity: Summative Reviews of the Epidemiological Evidence, Prevalence, and Interventions to Promote "Active Aging". Gerontologist. 2016 Apr;56 Suppl 2:S268-80.
27. Ku PW, Steptoe A, Chen YH, Chen LJ, Lin CH. Prospective association between late-life physical activity and hospital care utilisation: a 7-year nationwide follow-up study. Age Ageing. 2017 May 1;46(3):452-9.
28. Okubo Y, Seino S, Yabushita N, Osuka Y, Jung S, Nemoto M, et al. Longitudinal association between habitual walking and fall occurrences among community-dwelling older adults: Analyzing the different risks of falling. Arch Gerontol Geriatr. 2015 2015/01/01/;60(1):45-51.

29. Lu Z, Lam FMH, Leung JCS, Kwok TCY. The U-Shaped Relationship Between Levels of Bouted Activity and Fall Incidence in Community-Dwelling Older Adults: A Prospective Cohort Study. J Gerontol A Biol Sci Med Sci. 2020 Sep 25;75(10):e145-e51.
30. Schoene D, Heller C, Aung YN, Sieber CC, Kemmler W, Freiberger E. A systematic review on the influence of fear of falling on quality of life in older people: is there a role for falls? Clin Interv Aging. 2019;14:701-19.
31. Jefferis BJ, Iliffe S, Kendrick D, Kerse N, Trost S, Lennon LT, et al. How are falls and fear of falling associated with objectively measured physical activity in a cohort of community-dwelling older men? BMC Geriatr. 2014 Oct 27;14:114.
32. Wijlhuizen GJ, de Jong R, Hopman-Rock M. Older persons afraid of falling reduce physical activity to prevent outdoor falls. Prev Med. 2007 Mar;44(3):260-4.
33. Del Din S, Galna B, Lord S, Nieuwboer A, Bekkers EMJ, Pelosin E, et al. Falls Risk in Relation to Activity Exposure in High-Risk Older Adults. J Gerontol A Biol Sci Med Sci. 2020 May 22;75(6):1198-205.
34. Tricco AC, Thomas SM, Veroniki AA, Hamid JS, Cogo E, Strifler L, et al. Comparisons of Interventions for Preventing Falls in Older Adults: A Systematic Review and Meta-analysis. JAMA. 2017 Nov 7;318(17):1687-99.
35. Sherrington C, Fairhall N, Kwok W, Wallbank G, Tiedemann A, Michaleff ZA, et al. Evidence on physical activity and falls prevention for people aged 65+ years: systematic review to inform the WHO guidelines on physical activity and sedentary behaviour. Int J Behav Nutr Phys Act. 2020 Nov 26;17(1):144.
36. Montero-Odasso M, van der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing. 2022;51(9).
37. Subías-Perié J, Navarrete-Villanueva D, Gómez-Cabello A, Vicente-Rodríguez G, Casajús JA. Health economic evaluation of exercise interventions in people over 60 years old: A systematic review. Exp Gerontol. 2022 May;161:111713.
38. Kwon J, Squires H, Franklin M, Lee Y, Young T. Economic evaluation of community-based falls prevention interventions for older populations: a systematic methodological overview of systematic reviews. BMC Health Serv Res. 2022 Mar 26;22(1):401.
39. Olij BF, Ophuis RH, Polinder S, van Beeck EF, Burdorf A, Panneman MJM, et al. Economic Evaluations of Falls Prevention Programs for Older Adults: A Systematic Review. J Am Geriatr Soc. 2018 Nov;66(11):2197-204.

40. Scheckel B, Stock S, Müller D. Cost-effectiveness of group-based exercise to prevent falls in elderly community-dwelling people. BMC Geriatr. 2021 Jul 26;21(1):440.
41. Sherrington C, Fairhall NJ, Wallbank GK, Tiedemann A, Michaleff ZA, Howard K, et al. Exercise for preventing falls in older people living in the community. Cochrane Database Syst Rev. 2019 Jan 31;1(1):Cd012424.
42. Winser SJ, Chan HTF, Ho L, Chung LS, Ching LT, Felix TKL, et al. Dosage for cost-effective exercise-based falls prevention programs for older people: A systematic review of economic evaluations. Ann Phys Rehabil Med. 2020 Jan;63(1):69-80.
43. Schöffski O, von der Schulenburg J-M. Gesundheitsökonomische Evaluationen. 4 ed. Heidelberg: Springer.
44. Stock S, Lauterbach KW, Sauerland S. Gesundheitsökonomie. 4 ed. Bern: Hogrefe; 2021.
45. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4 ed. Oxford: Oxford University Press; 2015.
46. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. Value Health. 2015 Mar;18(2):161-72.
47. Mutubuki EN, El Alili M, Bosmans JE, Oosterhuis T, F JS, Ostelo R, et al. The statistical approach in trial-based economic evaluations matters: get your statistics together! BMC Health Serv Res. 2021 May 19;21(1):475.
48. El Alili M, van Dongen JM, Esser JL, Heymans MW, van Tulder MW, Bosmans JE. A scoping review of statistical methods for trial-based economic evaluations: The current state of play. Health Econ. 2022 Sep 11.
49. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
50. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011 Feb 20;30(4):377-99.
51. StatCorp. Stata multiple-imputation reference manual. Release 16. College Station, TX: StataCorp LLC; 2019.
52. Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. Applied methods of cost-effectiveness analysis in healthcare. New York: Oxford University Press; 2011.
53. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998 Apr-Jun;18(2 Suppl):S68-80.

54. Zethraeus N, Johannesson M, Jönsson B, Löthgren M, Tambour M. Advantages of using the net-benefit approach for analysing uncertainty in economic evaluation studies. Pharmacoeconomics. 2003;21(1):39-48.
55. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ. 2002 Jul;11(5):415-30.
56. Hoch JS, Rockx MA, Krahn AD. Using the net benefit regression framework to construct cost-effectiveness acceptability curves: an example using data from a trial of external loop recorders versus Holter monitoring for ambulatory monitoring of "community acquired" syncope. BMC Health Serv Res. 2006 Jun 6;6:68.
57. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990 Dec;16(3):199-208.
58. Szende A, Oppe M, Devlin N. EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Dordrecht: Springer; 2007.
59. Devlin N, Roudijk B, Ludwig K. Value Sets for EQ-5D-5L. A Compendium, Comparative Review & User Guide. Cham: Springer; 2022.
60. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. IQWiG Allgemeine Methoden. 6.1 ed. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen,; 2022.
61. Schwalm A, Danner M, Seidl A, Volz F, Dintsios CM, Gerber A. [IQWiG's methods for the cost-benefit assessment : Comparison with an international reference scenario]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2010 Jun;53(6):615-22.
62. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul;63(7):737-45.
63. National Institute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual. 2022. Available from: https://www.nice.org.uk/process/pmg36.
64. Higgins A, Barnett J, Meads C, Singh J, Longworth L. Does convenience matter in health care delivery? A systematic review of convenience-based aspects of process utility. Value Health. 2014 Dec;17(8):877-87.

65. Donaldson C, Mason H, Shackley P. Contingent Valuation in Health Care. The Elgar Companion to Health Economics. 2 ed: Edward Elgar Publishing; 2012. p. 425-34.
66. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, et al. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess. 2001;5(5):1-186.
67. Denkinger MD, Lukas A, Nikolaus T, Hauer K. Factors Associated with Fear of Falling and Associated Activity Restriction in Community-Dwelling Older Adults: A Systematic Review. Am J Geriatr Psychiatry. 2015 2015/01/01/;23(1):72-86.
68. Steckhan GMA, Fleig L, Schwarzer R, Warner LM. Perceived Physical Functioning and Gait Speed as Mediators in the Association Between Fear of Falling and Quality of Life in Old Age. J Appl Gerontol. 2022 Feb;41(2):421-9.
69. Hsu Y, Alfermann D, Lu FJ, Lin LL. Pathways from fear of falling to quality of life: the mediating effect of the self-concept of health and physical independence. Aging Ment Health. 2013;17(7):816-22.
70. Jansen CP, Nerz C, Kramer F, Labudek S, Klenk J, Dams J, et al. Comparison of a group-delivered and individually delivered lifestyle-integrated functional exercise (LiFE) program in older persons: a randomized noninferiority trial. BMC Geriatr. 2018 Nov 6;18(1):267.
71. Clemson L, Munro J, Fiatarone Singh MA. Lifestyle-integrated Functional Exercise (LiFE) program to prevent falls. Trainer's manual. Sydney: Sydney University Press; 2014.
72. Kramer F, Labudek S, Jansen CP, Nerz C, Fleig L, Clemson L, et al. Development of a conceptual framework for a group-based format of the Lifestyle-integrated Functional Exercise (gLiFE) programme and its initial feasibility testing. Pilot Feasibility Stud. 2020;6:6.
73. Jansen CP, Nerz C, Labudek S, Gottschalk S, Kramer-Gmeiner F, Klenk J, et al. Lifestyle-integrated functional exercise to prevent falls and promote physical activity: Results from the LiFE-is-LiFE randomized non-inferiority trial. Int J Behav Nutr Phys Act. 2021 Sep 3;18(1):115.
74. Jansen CP, Gottschalk S, Nerz C, Labudek S, Kramer-Gmeiner F, Klenk J, et al. Comparison of falls and cost-effectiveness of the group vs. individually delivered Lifestyle-integrated Functional Exercise (LiFE) program: final results from the LiFE-is-LiFE non-inferiority trial. Age Ageing. [accepted].

75. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011 Dec;20(10):1727-36.
76. Ludwig K, Graf von der Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. Pharmacoeconomics. 2018 Jun;36(6):663-74.
77. Lamb SE, Jørstad-Stein EC, Hauer K, Becker C. Development of a common outcome data set for fall injury prevention trials: the Prevention of Falls Network Europe consensus. J Am Geriatr Soc. 2005 Sep;53(9):1618-22.
78. Kempen GI, Yardley L, van Haastregt JC, Zijlstra GA, Beyer N, Hauer K, et al. The Short FES-I: a shortened version of the falls efficacy scale-international to assess fear of falling. Age Ageing. 2008 Jan;37(1):45-50.
79. Jette AM, Haley SM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, et al. Late life function and disability instrument: I. Development and evaluation of the disability component. J Gerontol A Biol Sci Med Sci. 2002 Apr;57(4):M209-16.
80. Haley SM, Jette AM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, et al. Late Life Function and Disability Instrument: II. Development and evaluation of the function component. J Gerontol A Biol Sci Med Sci. 2002 Apr;57(4):M217-22.
81. Seidl H, Bowles D, Bock JO, Brettschneider C, Greiner W, König HH, et al. [FIMA--questionnaire for health-related resource use in an elderly population: development and pilot study]. Gesundheitswesen. 2015 Jan;77(1):46-52.
82. Bock JO, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen. 2015 Jan;77(1):53-61.
83. Rote Liste Service GmbH. ROTE LISTE 2018: Pharmaceutical directory for Germany (including EU approvals and certain medical devices). Frankfurt/Main: Rote Liste Service GmbH; 2018.
84. Markland D, Tobin V. A Modification to the Behavioural Regulation in Exercise Questionnaire to Include an Assessment of Amotivation. J Sport and Exerc Psychol. 2004 01 Jun. 2004;26(2):191-6.
85. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986 Dec;51(6):1173-82.
86. Sobel ME. Direct and Indirect Effects in Linear Structural Equation Models. Sociological Methods & Research. 1987;16(1):155-76.

87. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018 May;27(5):1147-57.
88. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018 May;27(5):1171-9.
89. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34-42.
90. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. 2013 [cited 2021 18 Feb]; Available from: https://gdt.gradepro.org/app/handbook/handbook.html.
91. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual2018.
92. Delbaere K, Crombez G, Vanderstraeten G, Willems T, Cambier D. Fear-related avoidance of activities, falls and physical frailty. A prospective community-based cohort study. Age Ageing. 2004 Jul;33(4):368-73.
93. Deshpande N, Metter EJ, Lauretani F, Bandinelli S, Guralnik J, Ferrucci L. Activity restriction induced by fear of falling and objective and subjective measures of physical function: a prospective cohort study. J Am Geriatr Soc. 2008 Apr;56(4):615-20.
94. Pereira C, Bravo J, Raimundo A, Tomas-Carus P, Mendes F, Baptista F. Risk for physical dependence in community-dwelling older adults: The role of fear of falling, falls and fall-related injuries. Int J Older People Nurs. 2020 Sep;15(3):e12310.
95. Patil R, Uusi-Rasi K, Kannus P, Karinkanta S, Sievänen H. Concern about falling in older women with a history of falls: associations with health, functional ability, physical activity and quality of life. Gerontology. 2014;60(1):22-30.
96. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health. 2022 2022/03/01/;7(3):e219-e28.
97. Jayedi A, Gohari A, Shab-Bidar S. Daily Step Count and All-Cause Mortality: A Dose–Response Meta-analysis of Prospective Cohort Studies. Sports Med. 2022 2022/01/01;52(1):89-99.

98. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? Soc Sci Med. 2006 2006/04/01/;62(8):1891-901.
99. Milte CM, Walker R, Luszcz MA, Lancsar E, Kaambwa B, Ratcliffe J. How Important Is Health Status in Defining Quality of Life for Older People? An Exploratory Study of the Views of Older South Australians. Appl Health Econ Health Policy. 2014 2014/02/01;12(1):73-84.
100. Olsen JA, Misajon R. A conceptual map of health-related quality of life dimensions: key lessons for a new instrument. Qual Life Res. 2020 2020/03/01;29(3):733-43.
101. Penton H, Dayson C, Hulme C, Young T. A Qualitative Investigation of Older Adults' Conceptualization of Quality of Life and a Think-Aloud Content Validation of the EQ-5D-5L, SF-12v2, Warwick Edinburgh Mental Wellbeing Scale, and Office of National Statistics-4. Value Health. 2022 2022/06/25/.
102. Ratcliffe J, Lancsar E, Flint T, Kaambwa B, Walker R, Lewin G, et al. Does one size fit all? Assessing the preferences of older and younger people for attributes of quality of life. Qual Life Research. 2017 2017/02/01;26(2):299-309.
103. Hanmer J, Cella D, Feeny D, Fischhoff B, Hays RD, Hess R, et al. Selection of key health domains from PROMIS® for a generic preference-based scoring system. Qual Life Res. 2017 2017/12/01;26(12):3377-85.
104. Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, Hess R, et al. Estimation of a Preference-Based Summary Score for the Patient-Reported Outcomes Measurement Information System: The PROMIS(®)-Preference (PROPr) Scoring System. Med Decis Making. 2018 Aug;38(6):683-98.
105. Brazier J, Peasgood T, Mukuria C, Marten O, Kreimeier S, Luo N, et al. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing and Key Results. Value Health. 2022 2022/04/01/;25(4):482-91.
106. Cleland J, Hutchinson C, McBain C, Walker R, Milte R, Khadka J, et al. Developing dimensions for a new preference-based quality of life instrument for older people receiving aged care services in the community. Qual Life Res. 2021 Feb;30(2):555-65.
107. Reicherzer L, Kramer-Gmeiner F, Labudek S, Jansen CP, Nerz C, Nystrand MJ, et al. Group or individual lifestyle-integrated functional exercise (LiFE)? A qualitative analysis of acceptability. BMC Geriatr. 2021 Feb 1;21(1):93.

108. Nerz C, Kramer-Gmeiner F, Jansen CP, Labudek S, Klenk J, Becker C, et al. Group-Based and Individually Delivered LiFE: Content Evaluation and Predictors of Training Response - A Dose-Response Analysis. Clin Interv Aging. 2022;17:637-52.
109. Kennedy-Martin M, Slaap B, Herdman M, van Reenen M, Kennedy-Martin T, Greiner W, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. Eur J Health Econ. 2020 Nov;21(8):1245-57.
110. Brazier J, Tsuchiya A. Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY. Appl Health Econ Health Policy. 2015 Dec;13(6):557-65.
111. Brazier JE, Rowen D, Lloyd A, Karimi M. Future Directions in Valuing Benefits for Estimating QALYs: Is Time Up for the EQ-5D? Value Health. 2019 Jan;22(1):62-8.
112. Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS® Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. Value Health. 2019 2019/05/01/;22(5):537-44.
113. Weinstein MC. A QALY is a QALY--or is it? J Health Econ. 1988 Sep;7(3):289-90.
114. Weatherly H, Drummond M, Claxton K, Cookson R, Ferguson B, Godfrey C, et al. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. Health Policy. 2009 Dec;93(2-3):85-92.
115. Alayli-Goebbels AF, Evers SM, Alexeeva D, Ament AJ, de Vries NK, Tilly JC, et al. A review of economic evaluations of behavior change interventions: setting an agenda for research methods and practice. J Public Health (Oxf). 2014 Jun;36(2):336-44.
116. Cochrane M, Watson PM, Timpson H, Haycox A, Collins B, Jones L, et al. Systematic review of the methods used in economic evaluations of targeted physical activity and sedentary behaviour interventions. Soc Sci Med. 2019 Jul;232:156-67.
117. Seleznova Y, Alayli A, Stock S, Müller D. Methodological issues in economic evaluations of disease prevention and health promotion: an overview of systematic and scoping reviews. BMC Public Health. 2021 Nov 20;21(1):2130.
118. Wu S, Cohen D, Shi Y, Pearson M, Sturm R. Economic analysis of physical activity interventions. Am J Prev Med. 2011 Feb;40(2):149-58.
119. Mattli R, Farcher R, Syleouni ME, Wieser S, Probst-Hensch N, Schmidt-Trucksäss A, et al. Physical Activity Interventions for Primary Prevention in Adults: A Systematic Review of Randomized Controlled Trial-Based Economic Evaluations. Sports Med. 2020 Apr;50(4):731-50.

120. Dams J, Gottschalk S, Schwenk M, Nerz C, Becker C, Klenk J, et al. Budget impact analysis of a Lifestyle-integrated Functional Exercise (LiFE) program for older people in Germany: A Markov model based on data from the LiFE-is-LiFE trial. [unpublished manuscript].

3 Publication 1 – Mediating factors on the association between fear of falling and health-related quality of life in community-dwelling German older people: a cross-sectional study

Published in *BMC Geriatrics* as:

Gottschalk S[1], König HH[1], Schwenk M[2], Nerz C[3], Becker C[3], Klenk J[3,4,5], Jansen CP[2], Dams J[1]. Mediating factors on the association between fear of falling and health-related quality of life in community-dwelling German older people: a cross-sectional study. BMC Geriatr. 2020 Oct 14;20(1):401. DOI: 10.1186/s12877-020-01802-6. Copyright © The Author(s) 2020

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Network Aging Research, Heidelberg University, Heidelberg, Germany.

[3]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany.

[4]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.

[5]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany.

Gottschalk et al. BMC Geriatrics (2020) 20:401
https://doi.org/10.1186/s12877-020-01802-6

BMC Geriatrics

RESEARCH ARTICLE | Open Access

Mediating factors on the association between fear of falling and health-related quality of life in community-dwelling German older people: a cross-sectional study

Sophie Gottschalk[1*], Hans-Helmut König[1], Michael Schwenk[2], Carl-Philipp Jansen[2], Corinna Nerz[3], Clemens Becker[3], Jochen Klenk[3,4,5] and Judith Dams[1]

Abstract

Background: Previous research has shown that not only falls, but also fear of falling (FoF) influences health-related quality of life (HrQoL) negatively. The EQ-5D (consisting of an index and a visual analogue scale [EQ-VAS]) is a frequently used instrument to determine HrQoL in clinical studies and economic evaluations, but no previous study compared the association between FoF and the EQ-5D index with the association between FoF and the EQ-VAS. Moreover, factors that influence the association between FoF and HrQoL are rarely examined. Thus, this study aimed to examine the association between FoF and HrQoL and to examine factors that mediate the association.

Methods: FoF (Short Falls Efficacy Scale International) and HrQoL (EQ-5D descriptive system, EQ-5D index, and EQ-VAS) were assessed in a sample of community-dwelling older persons (≥70 years) participating in the baseline assessment of a randomized controlled trial ($N = 309$). Linear and logistic regression analyses were performed, adjusting for sociodemographic variables, frequency of falls, number of chronic conditions, functional mobility (Timed up-and-go test), and subjective functional capacity (LLFDI function and disability scales). Multiple regression models were used to test the mediating effects.

Results: Moderate or high FoF was prevalent in 66% of the sample. After adjusting for covariates, FoF was negatively associated with the EQ-5D index, but not with the descriptive system or the EQ-VAS. Subjective functional capacity partly mediated the association between FoF and the EQ-5D index and completely mediated the association between FoF and the EQ-VAS.

Conclusion: FoF was negatively associated with the EQ-5D index. As subjective functional capacity mediated the association between FoF and HrQoL, future interventions should account for subjective functional capacity in their design.

Keywords: Fear of falling, EQ-5D, Health-related quality of life, Falls efficacy, Older persons

* Correspondence: s.gottschalk@uke.de
[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg Center for Health Economics, Martinistraße 52, 20246 Hamburg, Germany
Full list of author information is available at the end of the article

Background

Due to the demographic change, the older population is projected to increase [1]. As this population typically has a higher level of (multi-)morbidity, an increase will probably pose challenges to health care systems in the future. The reasons for higher morbidity in older age are manifold. However, especially falls are a frequent health-deteriorating event in older people. One third of the population aged 65 years and above experiences a fall within a year [2–6], which often leads to severe consequences like injuries, or activity limitations, and consequently, to a decline in health-related quality of life (HrQoL) [7–9].

Since many health systems move beyond the idea of mere survival but focus on maintaining the best possible health status, overarching concepts like HrQoL have become more important in describing the impact of health conditions or the effects of interventions. HrQoL is subjective and depends on a variety of physical, emotional, and social-cultural factors [10, 11]. It is therefore necessary to take the individual valuation of the health status into account. Several measurements of HrQoL have been developed. The EQ-5D [12, 13] is a generic instrument which is widely used in clinical studies and economic evaluations. It comprises a descriptive system and a visual analogue scale (EQ-VAS). The descriptive system of the EQ-5D can be transformed to an index based on societal preference values, whereas the EQ-VAS quantifies the overall current health status based on a respondent's individual preferences [14].

As a recent systematic review by Schoene et al. [15] confirmed, not only falls but also fall-related risk factors like fear of falling (FoF) influence HrQoL negatively. In the past, FoF was considered as consequence of falls, but nowadays FoF is considered as an independent predictor of disability or HrQoL, independent of a prior fall experience [15, 16]. The prevalence of FoF in the population aged 65 years and above varies widely, with the majority of studies reporting a prevalence between 20 and 85% depending on sample characteristics and the measurement used to assess FoF [15, 17]. The prevalence of FoF tends to be higher in females, older persons, as well as in those having a history of falls, being physically impaired, or reporting poor self-rated health [17–19]. Moreover, psychological factors, such as depressive symptoms, loneliness, optimism, or self-esteem, are related to FoF [20, 21]. Consequences of FoF are a decline in cognitive and physical function, higher physical dependence, an increased risk of falling, the avoidance of activities, and restrictions in participation in social activities [17, 22–30].

The review concluded that the association between FoF and HrQoL was consistent, regardless of the instruments used to assess FoF and HrQoL, with the majority of studies using generic multidimensional instruments of HrQoL, like the EQ-5D or the SF-36, and validated instruments of FoF [15]. But these studies on the association between FoF and HrQoL mainly examined FoF as independent predictor [31–34], whereas the factors influencing the association between FoF and HrQoL were hardly addressed. However, identifying these factors is crucial as they might be modifiable [35] and could therefore be considered in the development of interventions. As the risks and consequences of FoF themselves predict HrQoL [33, 36, 37], it is reasonable to assume that they mediate the association between FoF and HrQoL. To our knowledge, only one study explored mediating effects. Using samples of community-dwelling older persons from Germany ($n = 182$) and Taiwan ($n = 193$), Hsu et al. [38] found that the association between FOF and HrQoL, measured using the SF-12, was significantly mediated by the self-concept of health and physical activity.

The EQ-5D is the most frequently used instrument to determine HrQoL in clinical studies and economic evaluations, but no previous study on the association between FoF and HrQoL compared the association between FoF and the EQ-5D index with the association between FoF and the EQ-VAS [15]. Therefore, the current study aimed to close this gap. In addition, the current study focused on factors that mediate the association between FoF and HrQoL in order to better understand the mechanisms underlying this association, which may serve as a basis for new approaches in the design of interventions.

Methods

Sample description/characteristics

Baseline data was taken from a multi-centre, two armed, single-blinded, randomized fall prevention trial (LiFE-is-LiFE) evaluating a group-based version of the 'Lifestyle-integrated Functional Exercise' Program (LiFE) [39] for its non-inferiority compared to the original face-to-face approach [40].

The LiFE-is-LiFE trial included community-dwelling, German-speaking people aged ≥70 years with a history or risk of falling (> 2 falls or 1 injurious fall within the last 12 month or limited balance [Timed Up-and-Go time ≥ 12 s]), who were able to ambulate 200 m without personal assistance. Participants were excluded if they exceeded a certain physical activity level (structured exercise > 1 time per week or self-reported activity level above 150 min of moderate to vigorous physical activity per week in past 3 months), were unavailable for home visits during the intervention time or for completion of the follow-up assessments, if they participated in another scientific trial, or had certain medical conditions that affect the ability to perform the activities taught in the program (e.g., Parkinson's disease or moderate to severe cognitive impairment). A detailed description of the LiFE-is-LiFE project and its inclusion and exclusion criteria can be found elsewhere [40].

Health-related quality of life

HrQoL was assessed using the EQ-5D-5L questionnaire [12, 41]. The EQ-5D-5L descriptive system comprises the five dimensions mobility, self-care, usual activities, pain/discomfort and anxiety/depression. In each dimension, study participants were asked to rate their health problems on an ordinal five level scale with "no problems (1)", "slight problems (2)", "moderate problems (3)", "severe problems (4)" or "extreme problems (5)". By combining the answers, an individual health state out of 3125 (5^5) possible health states was obtained for each participant, with "11111" and "55555" representing the best and worst health state, respectively. Health states were transformed to an index value based on preference-based value sets from the German general population [42]. Since there are health states of the reference population being predicted to be < 0 [42], the EQ-5D index can take values between – 0.662 representing the worst possible HrQoL, 0 representing death, and 1 representing the best possible HrQoL. Generally, a value < 0 is assumed to present a health state which is valued worse than death.

In addition to the descriptive system and the EQ-5D index, HrQoL was assessed on a visual analogue scale (EQ-VAS). Participants were asked to rate their overall current health between 0 (worst) and 100 (best) [12].

Fear of falling

FoF was assessed with the German version of the Short Falls Efficacy Scale International (Short FES-I) [43]. Participants were asked to rate their concerns about falling regarding the execution of seven everyday tasks on a 4-level Likert scale reaching from "not at all concerned" (1), "somewhat concerned" (2), "fairly concerned" (3), to "very concerned" (4). A Short FES-I sum score was calculated by adding up the answers. This score ranged from 7 ("no concern about falling") to 28 ("severe concern about falling") with low, moderate and high concern represented by a score between 7 and 8, 9–13, and 14–28, respectively [44].

Further measurements

The frequency of falls was assessed by the self-reported number of injurious or non-injurious falls in the previous 6 months.

The number of chronic conditions was assessed by a sum score of the following chronic conditions: diabetes type 1 and 2, hypertension, acute cardiovascular disease, a history of heart attacks, a cardiac defect, auricular fibrillation or other cardiac arrhythmias, a history of stroke (more than 6 month ago) or transient ischemic attacks, arthrosis, rheumatoid arthritis, cancer (not on active treatment), asthma or chronic obstructive pulmonary disease (Gold class < III), osteoporosis, or depression.

Functional mobility was assessed via the Timed Up-and-Go Test (TUG) measuring the time a person needs to get up from a chair, walk three meters at a comfortable and safe pace, return, and sit down again [45].

Subjective functional capacity was measured using the Late Life Function and Disability Instrument (LLFDI) [46, 47], an instrument designed to assess physical functioning in older adults based on a theoretical or conceptual model that characterizes physical functioning within a socio-medical model of disability. It measures two distinct outcomes: function and disability. In the 32-item LLFDI function component, participants rate their ability to perform discrete actions or activities on a 5-level Likert scale ("no", "slight", "moderate", "heavy", or "total limitations"). In the 16-item LLFDI disability component, the participants' limitations in performing specific life tasks within a typical sociocultural and physical environment are assessed on a 5-level Likert scale ("not at all", "a little", "somewhat", "a lot", or "completely"). In the current study, the second LLFDI disability dimension focusing on frequency of performance was skipped. For both components (function and disability), a sum score was calculated and transformed to a scale between 0 and 100, with lower scores indicating a higher level of functional limitations or disability.

Sociodemographic variables comprised age, sex, educational status, marital status (married or living in a partnership/widowed/divorced/permanently living separated/single) and living situation (living alone/living with others). Educational status was measured by the highest school leaving qualification achieved. Since the information was assessed based on qualification levels, which are specific for the German educational system, the information was grouped into "low" (9 years of school education), "intermediate" (10 years of school education), and "high" (qualifies to enter university) level of education.

Statistical analysis

In addition to descriptive statistics, the association between FoF and HrQoL measured using the EQ-5D index, the EQ-5D descriptive system and the EQ-VAS was examined using Spearman's rank correlation coefficients. According to Cohen, correlation coefficients between 0.10–0.19, 0.30–0.49 and 0.50–1.00 were interpreted as weak, moderate, and strong, respectively [48]. Furthermore, linear regression models were performed with the EQ-5D index or the EQ-VAS as dependent variables and FoF as independent variable. Neither the EQ-5D index nor the EQ-VAS was distributed normally, thus bootstrapped standard errors and confidence intervals for the regression coefficients from 10,000 resampled data sets were estimated. To examine the association between FoF and the EQ-5D descriptive system, logistic regression models were performed by dichotomizing answers of each EQ-5D dimension, with 0 representing no problems and 1 representing any problems. For each outcome, two models were calculated: the first model (Model 1) included FoF and sociodemographic variables, whereas the second model (Model 2) additionally included the number of chronic conditions, the number of falls, functional mobility (TUG), and subjective functional capacity (LLFDI function and disability scales). Additionally, path models were performed to estimate the mediating effects of function and disability on the association between FoF and EQ-5D-index and EQ-VAS following the Baron and Kenny approach [49]. The indirect effects were tested for significance using the Sobel test [50].

Statistical analyses were conducted using STATA/SE 16.0 [StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC]. For all analyses, the significance level was set to 0.05.

Results

Descriptive statistics

Sociodemographic and clinical characteristics are presented in Table 1. At baseline $N = 309$ participants were included with a mean age of 78.68 (standard deviation [SD] 5.31) years. The majority was female (73.46%) and had an intermediate or high educational level (67.96%). Approximately half of the sample was married or was living in a partnership (45.31%), whereas one third was widowed (35.92%) and one tenth was divorced (11.97%). Only a small proportion was single (5.83%) or permanently living separated (0.97%). On average, participants reported 2.52 (SD 1.56) chronic conditions. The mean scores on the LLDFI function and disability scales were 57.34 (SD 7.94) and 70.66 (SD 11.98), respectively. In the previous 6 months, 40.78% of the sample experienced at least one fall. Among those who fell, the average number of falls was 1.61 (SD 1.21). Low FoF was reported by 33.98%, whereas 52.75% reported moderate FOF and 13.27% reported high FoF. The mean EQ-5D index was 0.84 (SD 0.15) and the mean EQ-VAS was 70.91 (SD 16.46). Furthermore, differences between people experiencing at least one fall and those without falls were not significant (data not shown).

Correlation coefficients

In bivariate analyses (Table 2), Spearman's rank correlations between FoF and the EQ-5D index, EQ-5D descriptive system or EQ-VAS were weak to moderate, with absolute correlation coefficients between $r_S = 0.17$ ($p < 0.05$) and 0.43 ($p < 0.05$). Furthermore, associations

Table 1 Sample characteristics

***N* = 309**			
Female	n (%)	227	73.46
Age	Mean (SD)	78.67	5.31
Educational status	n (%)		
Low		94	30.42
Intermediate		92	29.77
High		118	38.19
Other		5	1.62
Marital status	n (%)		
Married/living in a partnership		140	45.31
Widowed		111	35.92
Divorced		37	11.97
Permanently living separated		3	0.97
Single		18	5.83
Living alone	n (%)	166	53.72
Chronic conditions	Mean (SD)	2.52	1.56
LLFDI function[a]	Mean (SD)	57.34	7.94
LLFDI disability[b]	Mean (SD)	70.66	11.98
TUG (time in seconds)	Mean (SD)	13.29	3.86
Prevalence of fallers	n (%)	126	40.78
Number of falls among fallers	Mean (SD)	1.61	1.21
Fear of falling	n (%)	10.36	3.03
Low concern		105	33.98
Moderate concern		163	52.75
High concern		41	13.27
EQ-5D index	Mean (SD)	0.84	0.15
EQ-VAS	Mean (SD)	70.91	16.46

LLFDI Late Life Function and Disability Instrument, *TUG* Timed up-and-go test
[a]Higher score indicates lower limitations
[b]Higher score indicates lower disability

Table 2 Correlation coefficients between fear of falling and variables of health, functional status, and sociodemographic characteristics

Variables	FES-I
EQ-5D index	**−0.43***
EQ mobility	**0.29***
EQ self-care	**0.35***
EQ usual activities	**0.34***
EQ pain/discomfort	**0.17***
EQ anxiety/depression	**0.25***
EQ VAS	**−0.28***

*$p < 0.05$

between FoF and the living situation and functional mobility were weak ($r_S = 0.15$ and 0.24, $p < 0.05$). Moreover, FoF correlated strongly with function ($r_S = 0.56$, $p < 0.05$) and moderately with disability ($r_S = 0.43$, $p < 0.05$). No significant correlation was found between FoF and age, gender, educational level, the number of chronic conditions, or the number of falls ($p > 0.05$).

Multivariate regressions

Association between FoF and sub-dimensions of HrQoL (EQ-5D descriptive system)

Results of the logistic regression models for the association between FoF and the dimensions of the EQ-5D descriptive system are presented in Table 3. After adjusting for sociodemographic variables (Model 1), FoF was significantly associated with problems in each dimension (odds ratios [OR] between 1.14 and 1.35, $p < 0.05$). These associations became non-significant after adjusting for chronic conditions, functional mobility, and subjective functional capacity (Model 2).

Association between FoF and EQ-5D index

After adjusting for sociodemographic variables (Model 1), linear regression revealed a significant negative association between FoF and the EQ-5D index ($\beta = -0.02$, $p < 0.001$; Table 4). This relationship remained significant after adjusting for chronic conditions, functional mobility, and subjective functional capacity (ß = -0.01, $p < 0.01$; Model 2).

Association between FoF and EQ-VAS

In Model 1, higher FoF was significantly associated with a lower EQ-VAS score ($\beta = -1.54$, $p < 0.001$; Table 4). After adjusting for chronic conditions, functional mobility, and subjective functional capacity (Model 2), FoF did no longer significantly predict the EQ-VAS score ($\beta = -0.36$, $p > 0.05$), whereas the number of comorbidities ($\beta = -1.76$, $p < 0.01$), and the levels of function ($\beta = 0.53$, $p < 0.001$) and disability ($\beta = 0.21$, $p < 0.01$) significantly predicted the EQ-VAS.

Mediating effects of function and disability

Figures 1 and 2 show the mediation results of self-reported function and disability on the relationship between FoF and HrQoL. Separate mediation models were calculated for function and disability. Function and disability partially mediated the association between FoF and the EQ-5D index. The coefficient of FoF increased from − 0.023 to − 0.012 after controlling for function (Sobel test Z = − 3.08, $p < 0.01$) and to − 0.018 after controlling for disability (Sobel test Z = − 5.31, $p < 0.001$). The association between FoF and the EQ-VAS was completely mediated by function as the coefficient of FoF increased from − 1.589 to a non-significant effect of − 0.489 (Sobel test Z = − 1.25, $p > 0.05$) after controlling for function. After controlling for disability, the coefficient of FoF increased from − 1.589 to − 0.989 (Sobel test Z = -2.69, $p < 0.01$), indicating a partial mediating effect of disability on the association between FoF and the EQ-VAS.

Discussion

In this sample of community-dwelling older persons from Germany, 66% had moderate or high FoF. After adjustment for sociodemographic characteristics, chronic conditions, functional mobility, and subjective functional capacity, FoF was significantly associated with HrQoL measured by the EQ-5D index. This confirmed previous findings [15]. The current study did not only examine the influence of FoF on the EQ-5D index or on the EQ-5D descriptive system, but also the influence of FoF on the overall current health status (EQ-VAS). When accounting for sociodemographic characteristics, chronic conditions, functional mobility, and subjective functional capacity, FoF was not associated with the EQ-VAS. Overall, FoF seemed to be better captured by the specific EQ-5D dimensions than by the unspecific assessment of the EQ-VAS. This may be due to the different concepts underlying the EQ-5D index and the EQ-VAS. The EQ-5D index is based on subjective evaluations of health in five specific dimensions. However, these subjective ratings in the respective dimensions were transformed into an index based on societal preference values. These values were obtained using a representative sample of the general population and thus reflect the societal weighting of restrictions in the respective dimensions of the EQ-5D descriptive system [42]. In contrast, the EQ-VAS is subject to a valuation of health based on individual preferences. Thus, by asking how healthy participants felt today on the EQ-VAS without giving predefined dimensions like in the EQ-5D descriptive system, aspects other than FoF may play a greater role for participants in assessing their overall current health.

As already found in another study [31], higher age was associated with better EQ-5D-rated HrQoL. This can probably be explained by a selection bias. When the distribution of HrQoL and age were visually assessed, participants aged 87 and older exclusively reported EQ-5D index values above 0.7, whereas in younger participants, the EQ-5D index values of some individuals were also distributed at lower levels. When excluding participants aged 87 and older in additional analyses, age was no longer significantly associated with HrQoL.

Contrary to previous studies, where falls and FoF were associated and both had a significant relationship with HrQoL [15, 34], no significant association between the number of previous falls and HrQoL was found in the

Table 3 Logistic regression models of the association between fear of falling and the EQ-5D descriptive system

N = 309	EQ-5D Dimensions									
	Mobility		Self-care		Usual activities		Pain/Discomfort		Anxiety/Depression	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Fear of falling	**1.30***** (1.17–1.45)	1.05 (0.93–1.20)	**1.35***** (1.22–1.50)	1.12 (0.99–1.26)	**1.35***** (1.22–1.50)	1.12 (0.99–1.26)	**1.14*** (1.03–1.27)	0.93 (0.81–1.06)	**1.18***** (1.08–1.28)	1.11 (1.00–1.23)
Age	1.00 (0.96–1.05)	10.97 (0.92–1.02)	1.02 (0.97–1.07)	0.99 (0.93–1.04)	1.02 (0.97–1.07)	0.99 (0.93–1.04)	0.96 (0.91–1.00)	**0.93*** (0.88–0.98)	0.95 (0.91–1.00)	**0.92**** (0.87–0.97)
Female	0.82 (0.45–1.48)	0.52 (0.26–1.03)	0.98 (0.55–1.76)	0.69 (0.35–1.34)	0.98 (0.55–1.76)	0.69 (0.35–1.34)	1.40 (0.74–2.67)	0.90 (0.45–1.82)	1.31 (0.71–2.42)	1.39 (0.72–2.69)
Education										
High	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Low	**1.86*** (1.01–3.40)	1.55 (0.79–3.04)	1.39 (0.77–2.52)	1.20 (0.62–2.33)	1.39 (0.77–2.52)	1.20 (0.62–2.33)	**2.64**** (1.33–5.26)	**2.36*** (1.13–4.94)	**1.92*** (1.05–3.53)	**2.09*** (1.10–4.00)
Intermediate	1.31 (0.72–2.38)	1.08 (0.57–2.06)	1.53 (0.84–2.78)	1.20 (0.63–2.31)	1.53 (0.84–2.78)	1.20 (0.63–2.31)	1.49 (0.79–2.81)	1.39 (0.70–2.73)	1.83 (0.99–3.38)	1.47 (0.78–2.79)
Other	0.73 (0.10–5.39)	0.68 (0.05–8.79)	3.58 (0.33–39.0)	6.58 (0.37–116)	3.58 (0.33–39.0)	6.58 (0.37–116)	1.94 (0.19–19.9)	2.34 (0.18–31.0)	3.61 (0.52–24.9)	2.90 (0.34–24.9)
Shared living	0.75 (0.44–1.26)	0.88 (0.50–1.56)	0.95 (0.57–1.61)	1.13 (0.64–2.01)	0.95 (0.57–1.61)	1.13 (0.64–2.01)	1.69 (0.93–3.04)	**2.38**** (1.24–4.55)	0.92 (0.54–1.56)	0.93 (0.53–1.62)
Chronic conditions		1.15 (0.96–1.40)		0.95 (0.80–1.14)		0.95 (0.80–1.14)		1.02 (0.83–1.25)		0.99 (0.83–1.17)
Number of falls		1.09 (0.82–1.44)		1.07 (0.82–1.39)		1.07 (0.82–1.39)		0.96 (0.71–1.30)		0.99 (0.78–1.25)
Function		**0.89***** (0.84–0.95)		**0.92**** (0.87–0.97)		**0.92**** (0.87–0.97)		**0.87***** (0.82–0.93)		1.03 (0.98–1.08)
Disability		**0.97*** (0.94–0.99)		**0.94***** (0.92–0.97)		**0.94***** (0.92–0.97)		0.99 (0.96–1.02)		**0.95***** (0.92–0.98)
Functional mobility		1.02 (0.93–1.13)		1.02 (0.93–1.12)		1.02 (0.93–1.12)		0.98 (0.90–1.08)		1.09* (1.00–1.19)

Odds ratios with 95% CI in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; fear of falling was assessed with the Short Falls-efficacy Scale-International (Short-FES-I), function and disability with the Late-life Function and Disability Instrument (LLFDI), and functional mobility with the Timed up-and-go test

Table 4 Linear regression models of the association between fear of falling and EQ-5D-rated and EQ-VAS-rated HrQoL

N = 309	**EQ-5D Index**				**EQ VAS**			
	Model 1		**Model 2**		**Model 1**		**Model 2**	
	β	**SE**	**β**	**SE**	**β**	**SE**	**β**	**SE**
Fear of falling	**−0.023** ***	(0.004)	**−0.010** **	(0.004)	**−1.535** ***	(0.354)	−0.361	(0.396)
Age	0.001	(0.001)	**0.004** *	(0.001)	0.049	(0.169)	0.228	(0.162)
Female	−0.014	(0.016)	0.007	(0.014)	−0.370	(2.169)	1.527	(2.135)
Education								
High	ref.		ref.		ref.		ref.	
Low	**−0.050** **	(0.018)	−0.033	(0.017)	−2.342	(2.128)	−0.571	(2.018)
Intermediate	**−0.039** *	(0.018)	−0.022	(0.016)	− 2.126	(2.235)	−0.403	(2.209)
Other	−0.036	(0.025)	−0.027	(0.053)	−0.471	(5.543)	0.055	(3.549)
Shared living	−0.002	(0.016)	−0.013	(0.014)	1.511	(1.893)	0.575	(1.779)
Chronic conditions			**−0.014** **	(0.005)			**−1.761** **	(0.573)
Number of falls			0.002	(0.006)			0.370	(0.974)
Function			**0.005** ***	(0.001)			**0.533** ***	(0.154)
Disability			**0.002** **	(0.001)			**0.208** **	(0.080)
Functional mobility			−0.003	(0.003)			0.065	(0.297)
Adj. R-Squared	0.233		0.379		0.072		0.188	

Fear of falling was assessed with the Short Falls-efficacy Scale-International (Short-FES-I), function and disability with the Late-life Function and Disability Instrument (LLFDI), and functional mobility with the Timed up-and-go test
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

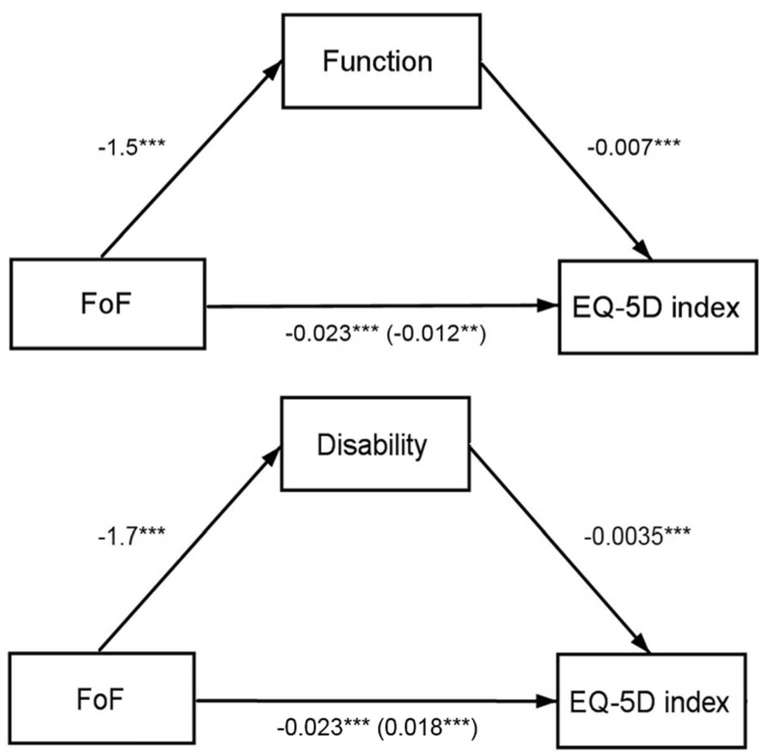


Fig. 1 Mediating effects of function and disability on the association between FoF and HrQoL (EQ-5D index). Note: Path diagrams indicate that function and disability partially mediated the association between fear of falling (FoF) and EQ-5D-rated health-related quality of life. Numbers outside the parentheses denote the path coefficients between variables, whereas numbers in the parentheses indicate the path coefficients after including the mediator (direct effect). Function and disability were assessed with the Late-Life Function and Disability Instrument (LLFDI). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

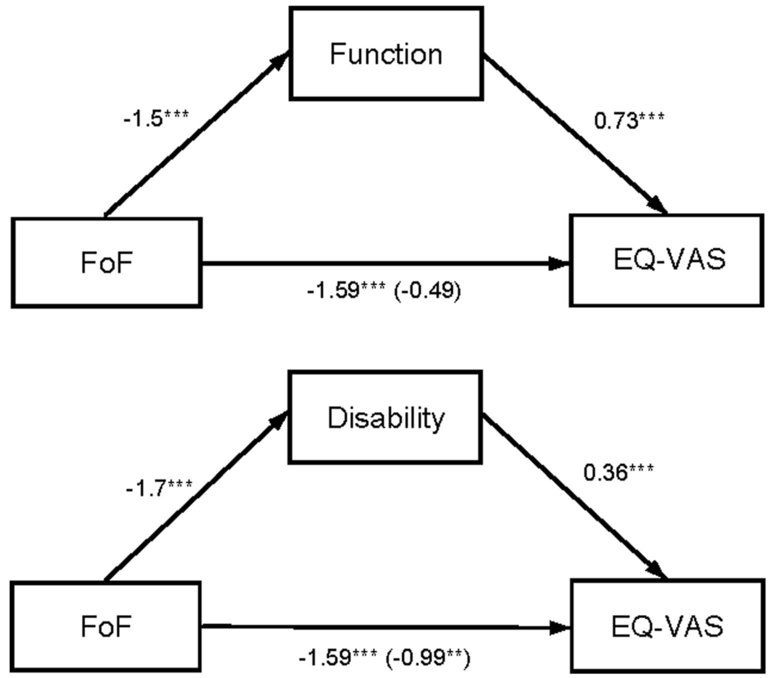


Fig. 2 Mediating effects of function and disability on the association between FoF and HrQoL (EQ-VAS). Note: Path diagrams indicate that function completely mediated and disability partially mediated the association between fear of falling (FoF) and EQ-VAS-rated health-related quality of life. Numbers outside the parentheses denote the path coefficients between variables, whereas numbers in the parentheses indicate the path coefficients after including the mediator (direct effect). Function and disability were assessed with the Late-Life Function and Disability Instrument (LLFDI). $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

current study. However, the information on the history of falls was based solely on retrospective self-reports and may therefore be biased. Instead, factors like chronic conditions and activity restrictions (function and disability) seemed to be more important than falls in the association between FoF and HrQoL. Mediation analyses showed that function and disability partially mediated the association between FoF and the EQ-5D index. With regard to the EQ-VAS, the effect of FoF was partially mediated by disability and completely mediated by function. These findings are not surprising as previous research suggested that FoF is linked to disability and deteriorating function [51, 52]. Moreover, the FES-I measures FoF in the context of typical everyday activities. It therefore seems obvious that functional limitations and disability are reflected in FoF. Although the current study was of cross-sectional nature and therefore no causal inferences can be drawn from the results of the mediation analyses, possible interpretations of these findings can be hypothesized. A certain degree of FoF may protect individuals from an actual fall, because they are more attentive or careful [53]. However, when FoF leads to the avoidance of certain activities, it becomes a vicious circle. Avoiding activities leads to a deterioration in physical functioning, which in turn leads to an increased risk and fear of falling [2, 23, 25, 27, 28, 54–56]. Actual falls again lead to a further deterioration in physical health status [9, 30, 57]. This reduction in physical health status and social activities ultimately results in a higher level of dependence and poorer HrQoL [9, 30, 58, 59]. That function completely mediated the association between FoF and the EQ-VAS in this study may indicate that limitations in doing discrete actions or activities (function) play a greater role in the evaluation of overall current health (EQ-VAS) than the capability of performing less discrete, socially defined life tasks (disability). Furthermore, it may reflect a strong link between FoF and functional limitations. High FoF may prevent people from doing certain activities but may not prevent them from finding solutions to adapt to their FoF and functional limitations which enables them to perform socially defined life tasks despite FoF.

The mediating effect of function and disability in the current study emphasizes the importance to maintain daily and social life activities in older people. Thus, addressing these factors in interventions may lead to a reduction of FoF and an improvement in HrQoL. A randomized controlled trial from the Netherlands evaluated a home-based cognitive behavioural program to encourage older persons in performing activities of daily living [60]. The intervention focused on the identification and restructuring of misconceptions about falls, as well as on the uptake of new or previously avoided daily life activities and their safe execution. Thereby, the home-based cognitive behavioural program was effective in reducing FoF and disability. However, since the effect sizes of FoF, function, and disability on HrQoL were rather small in the current study, considering these factors alone in interventions may not lead to clinically important changes in HrQoL, which is related to the multidimensionality of factors influencing HrQoL.

Strengths and limitations

Even though different studies have already reached consensus regarding the association between FoF and HrQoL [15], to our knowledge, no previous study compared the association between FoF and the EQ-5D index (a multidimensional measure of HrQoL) with the association between FoF and the EQ-VAS (a single-item measure of overall current health). Moreover, previous findings were extended by examining the mediating effects of function and disability on the association between FoF and HrQoL.

This study has some limitations. As the current study used cross-sectional data, mediation models may not reflect the true direction of influence. Instead of being a consequence of FoF, functional limitations may lead to FoF, or even both directions of influence exist. Due to the conceptual overlap between function and disability, separate mediation models for function and disability were calculated, which precludes investigating their independent contributions. However, the mediation results of this study may serve as basis for future studies, which could, for example, investigate the causal relationship between FoF and HrQoL more closely using longitudinal data or by calculating more complex path models. In addition, the sample size of $N = 309$ was relatively small, thus results may not be generalizable to the older population at risk of falling in Germany. Furthermore, the selected sample reached better EQ-5D index values compared with normative values for the general German population of the respective age group [61]. This is most likely due to the exclusion of individuals who were cognitively impaired and had certain chronic conditions. Nevertheless, the prevalence of moderate or high FoF was high (66%), which may be explained by the fact that individuals who participate in a fall prevention project tend to be more sensitive to (fear of) falling. In addition, potential limitations by using the EQ-5D as measure of HrQoL should be noted. The EQ-5D excludes aspects of quality of life beyond health that may also be affected by fear of falling. Even some health-related aspects may not be sufficiently captured in the five dimensions of the EQ-5D. Although the introduction of the 5-level version of the EQ-5D has improved the ability to differentiate between health conditions, ceiling effects remain a problem [61, 62]. The results of this study should therefore be tested in future studies using different measures of (health-related) quality of life. Finally, the transferability of the results to other populations may be limited, because preference-based value sets for the German general population were used to calculate the EQ-5D index and country-specific cultural factors are known to influence the subjective assessment of health.

Conclusion

FoF was a significant negative predictor of the EQ-5D index, whereas FoF did not predict HrQoL measured by the EQ-VAS. This is probably attributable to the different concepts underlying the EQ-5D index and the EQ-VAS. Furthermore, function and disability were shown to mediate the association between FoF and HrQoL. Therefore, future interventions should account for function and disability in their design.

Abbreviations

EQ-5D: EuroQoL-5D: an instrument for measuring health-related quality of life; EQ-5D-5L: 5-level version of the EQ-5D; EQ-VAS: Visual analogue scale of the EQ-5D; FES-I: Falls Efficacy Scale International; FoF: Fear of falling; HrQoL: Health-related quality of life; LiFE: Lifestyle-integrated functional exercise; LLFDI: Late life function and disability instrument; SD: Standard deviation; TUG: Timed up-and-go test

Acknowledgements

We would like to thank all participants for their participation. We would also like to thank all colleagues involved in the conduct of this trial, including the assessors, database managers, and members of the advisory board.

Authors' contributions

MS and CB developed the grant proposal for the LiFE-is-LiFE trial approved for funding. SG and JD developed the methodological approach of this study with substantial contribution from HHK. SG performed the data analysis and produced the first draft of the manuscript. JD, HHK, MS, CPJ, CN, CB, and JK critically revised the manuscript for important intellectual content. The final version of the manuscript was reviewed, edited and approved by all authors.

Funding

This study is funded by the German Federal Ministry of Education and Research (grant: 01GL1705D). The funding source takes no part in the collection, analysis and interpretation of data, in the writing of the manuscript or in the decision to submit the manuscript for publication. Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to ethical and confidentiality concerns but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Ethical approval for the project was obtained from the Ethic Review Board of the Faculty of Behavioral and Cultural Studies at Heidelberg University (document number Schwe2017 2/1–1), and from the Ethic Review Board of the University Hospital and Faculty of Medicine in Tübingen (document number 723/2017BO2). The study is conforming to the respective policy and mandates of the Declaration of Helsinki. Participants' written informed consent was obtained from assessors at their first screening visit at the study site.

Consent for publication

Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg Center for Health Economics, Martinistraße 52, 20246 Hamburg, Germany. [2]Network Aging Research, Heidelberg University, Heidelberg, Germany. [3]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany. [4]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany. [5]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany.

Received: 24 June 2020 Accepted: 28 September 2020
Published online: 14 October 2020

References

1. United Nations, Department of Economic and Social Affairs, Population Division. World population prospects 2019, volume II. New York: Demographic Profiles; 2019.
2. Lord SR, Ward JA, Williams P, Anstey KJ. An epidemiological study of falls in older community-dwelling women: the Randwick falls and fractures study. Aust J Public Health. 1993;17(3):240–5.
3. Bergen G, Stevens MR, Burns ER. Falls and fall injuries among adults aged >/=65 years - United States, 2014. MMWR Morb Mortal Wkly Rep. 2016; 65(37):993–8.
4. Olsson Moller U, Midlov P, Kristensson J, Ekdahl C, Berglund J, Jakobsson U. Prevalence and predictors of falls and dizziness in people younger and older than 80 years of age--a longitudinal cohort study. Arch Gerontol Geriatr. 2013;56(1):160–8.
5. Skalska A, Wizner B, Piotrowicz K, Klich-Raczka A, Klimek E, Mossakowska M, et al. The prevalence of falls and their relation to visual and hearing impairments among a nation-wide cohort of older poles. Exp Gerontol. 2013;48(2):140–6.
6. Schumacher J, Pientka L, Trampisch U, Moschny A, Hinrichs T, Thiem U. The prevalence of falls in adults aged 40 years or older in an urban, German population. Results from a telephone survey. Z Gerontol Geriatr. 2014;47(2): 141–6.
7. Hartholt KA, van Beeck EF, Polinder S, van der Velde N, van Lieshout EMM, Panneman MJM, et al. Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life. J Trauma. 2011;71(3):748–53.
8. Tinetti ME, Williams CS. Falls, injuries due to falls, and the risk of admission to a nursing home. N Engl J Med. 1997;337(18):1279–84.
9. Stel VS, Smit JH, Pluijm SM, Lips P. Consequences of falling in older men and women and risk factors for health service use and functional decline. Age Ageing. 2004;33(1):58–65.
10. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. JAMA. 1995;273(1):59–65.
11. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual model of health-related quality of life. J Nurs Scholarsh. 2005;37(4):336–42.
12. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199–208.
13. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.
14. EuroQol Research Foundation. EQ-5D-5L user guide. 2019.
15. Schoene D, Heller C, Aung YN, Sieber CC, Kemmler W, Freiberger E. A systematic review on the influence of fear of falling on quality of life in older people: is there a role for falls? Clin Interv Aging. 2019;14:701–19.
16. Makino K, Makizako H, Doi T, Tsutsumimoto K, Hotta R, Nakakubo S, et al. Impact of fear of falling and fall history on disability incidence among older adults: prospective cohort study. Int J Geriatr Psychiatry. 2018;33(4):658–62.
17. Scheffer AC, Schuurmans MJ, van Dijk N, van der Hooft T, de Rooij SE. Fear of falling: measurement strategy, prevalence, risk factors and consequences among older persons. Age Ageing. 2008;37(1):19–24.
18. Curcio CL, Wu YY, Vafaei A, Barbosa JFS, Guerra R, Guralnik J, et al. A regression tree for identifying risk factors for fear of falling: the International Mobility in Aging Study (IMIAS). J Gerontol A Biol Sci Med Sci. 2020;75(1): 181–8.
19. Denkinger MD, Lukas A, Nikolaus T, Hauer K. Factors associated with fear of falling and associated activity restriction in community-dwelling older adults: a systematic review. Am J Geriatr Psychiatry. 2015;23(1):72–86.
20. Hajek A, Bock JO, König HH. Psychological correlates of fear of falling: findings from the German Aging Survey. Geriatr Gerontol Int. 2018;18(3): 396–406.
21. Rivasi G, Kenny RA, Ungar A, Romero-Ortuno R. Predictors of incident fear of falling in community-dwelling older adults. J Am Med Dir Assoc. 2019;21(5): 615–20.
22. Hughes CC, Kneebone II, Jones F, Brady B. A theoretical and empirical review of psychological factors associated with falls-related psychological concerns in community-dwelling older people. Int Psychogeriatr. 2015;27(7):1071–87.
23. Delbaere K, Crombez G, Vanderstraeten G, Willems T, Cambier D. Fear-related avoidance of activities, falls and physical frailty. A prospective community-based cohort study. Age Ageing. 2004;33(4):368–73.
24. Wijlhuizen GJ, de Jong R, Hopman-Rock M. Older persons afraid of falling reduce physical activity to prevent outdoor falls. Prev Med. 2007;44(3):260–4.
25. Yardley L, Smith H. A prospective study of the relationship between feared consequences of falling and avoidance of activity in community-living older people. Gerontologist. 2002;42(1):17–23.
26. Peeters G, Feeney J, Carey D, Kennelly S, Kenny RA. Fear of falling: a manifestation of executive dysfunction? Int J Geriatr Psychiatry. 2019;34(8): 1275–82.
27. Murphy SL, Williams CS, Gill TM. Characteristics associated with fear of falling and activity restriction in community-living older persons. J Am Geriatr Soc. 2002;50(3):516–20.
28. Deshpande N, Metter EJ, Lauretani F, Bandinelli S, Guralnik J, Ferrucci L. Activity restriction induced by fear of falling and objective and subjective measures of physical function: a prospective cohort study. J Am Geriatr Soc. 2008;56(4):615–20.
29. Sawa R, Asai T, Doi T, Misu S, Murata S, Ono R. The associations between physical activity, including physical activity intensity, and fear of falling differs by fear severity in older adults living in the community. J Gerontol B Psychol Sci Soc Sci. 2018;75(5):953–60.
30. Pereira C, Bravo J, Raimundo A, Tomas-Carus P, Mendes F, Baptista F. Risk for physical dependence in community-dwelling older adults: the role of fear of falling, falls and fall-related injuries. Int J Older People Nursing. 2020; 15(3):e12310.
31. Bjerk M, Brovold T, Skelton DA, Bergland A. Associations between health-related quality of life, physical function and fear of falling in older fallers receiving home care. BMC Geriatr. 2018;18(1):253.
32. Chang N-T, Chi L-Y, Yang N-P, Chou P. The impact of falls and fear of falling on health-related quality of life in Taiwanese elderly. J Community Health Nurs. 2010;27(2):84–95.
33. Davis JC, Marra CA, Liu-Ambrose TY. Falls-related self-efficacy is independently associated with quality-adjusted life years in older women. Age Ageing. 2011;40(3):340–6.
34. Thiem U, Klaaßen-Mielke R, Trampisch U, Moschny A, Pientka L, Hinrichs T. Falls and EQ-5D rated quality of life in community-dwelling seniors with concurrent chronic diseases: a cross-sectional study. Health Qual Life Outcomes. 2014;12(1):2.
35. Jung D. Fear of falling in older adults: comprehensive review. Asian Nurs Res. 2008;2(4):214–22.
36. Netuveli G, Wiggins RD, Hildon Z, Montgomery SM, Blane D. Quality of life at older ages: evidence from the English longitudinal study of aging (wave 1). J Epidemiol Community Health. 2006;60(4):357–63.
37. Sions JM, Hicks GE. Fear-avoidance beliefs are associated with disability in older American adults with low back pain. Phys Ther. 2011;91(4):525–34.
38. Hsu Y, Alfermann D, Lu FJ, Lin LL. Pathways from fear of falling to quality of life: the mediating effect of the self-concept of health and physical independence. Aging Ment Health. 2013;17(7):816–22.
39. Clemson L, Singh MF, Bundy A, Cumming RG, Weissel E, Munro J, et al. LiFE pilot study: a randomised trial of balance and strength training embedded in daily life activity to reduce falls in older adults. Aust Occup Ther J. 2010; 57(1):42–50.
40. Jansen CP, Nerz C, Kramer F, Labudek S, Klenk J, Dams J, et al. Comparison of a group-delivered and individually delivered lifestyle-integrated functional exercise (LiFE) program in older persons: a randomized noninferiority trial. BMC Geriatr. 2018;18(1):267.
41. Graf JM, Claes C, Greiner W, Uber A. Die deutsche version des EuroQol-Fragebogens. Z Gesundh Wiss. 1998;6(1):3–20.

42. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. PharmacoEconomics. 2018;36(6):663–74.
43. Kempen GIJM, Yardley L, Van Haastregt JCM, Zijlstra GAR, Beyer N, Hauer K, et al. The short FES-I: a shortened version of the falls efficacy scale-international to assess fear of falling. Age Ageing. 2008;37(1):45–50.
44. Delbaere K, Close JCT, Mikolaizak AS, Sachdev PS, Brodaty H, Lord SR. The falls efficacy scale international (FES-I). A comprehensive longitudinal validation study. Age Ageing. 2010;39(2):210–6.
45. Podsiadlo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142–8.
46. Haley SM, Jette AM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, et al. Late life function and disability instrument: II. Development and evaluation of the function component. J Gerontol A Biol Sci Med Sci. 2002;57(4):M217–22.
47. Jette AM, Haley SM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, et al. Late life function and disability instrument: I. development and evaluation of the disability component. J Gerontol A Biol Sci Med Sci. 2002;57(4):M209–16.
48. Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Routledge; 2003.
49. Baron R, Kenny D. The moderator-mediator variable distinction in social psychological research. J Pers Soc Psychol. 1987;51:1173–82.
50. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol Methodol. 1982;13:290–312.
51. Cumming RG, Salkeld G, Thomas M, Szonyi G. Prospective study of the impact of fear of falling on activities of daily living, SF-36 scores, and nursing home admission. J Gerontol A Biol Sci Med Sci. 2000;55(5):M299–305.
52. Auais M, French S, Alvarado B, Pirkle C, Belanger E, Guralnik J. Fear of falling predicts incidence of functional disability 2 years later: a perspective from an international cohort study. J Gerontol A Biol Sci Med Sci. 2018;73(9):1212–5.
53. Delbaere K, Sturnieks DL, Crombez G, Lord SR. Concern about falls elicits changes in gait parameters in conditions of postural threat in older people. J Gerontol A Biol Sci Med Sci. 2009;64(2):237–42.
54. Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis. Epidemiology. 2010;21(5):658–68.
55. Li F, Fisher KJ, Harmer P, McAuley E, Wilson NL. Fear of falling in elderly persons: association with falls, functional ability, and quality of life. J Gerontol B Psychol Sci Soc Sci. 2003;58(5):P283–90.
56. Patil R, Uusi-Rasi K, Kannus P, Karinkanta S, Sievänen H. Concern about falling in older women with a history of falls: associations with health, functional ability, physical activity and quality of life. Gerontology. 2014;60(1):22–30.
57. Stenhagen M, Ekström H, Nordell E, Elmståhl S. Accidental falls, health-related quality of life and life satisfaction: a prospective study of the general elderly population. Arch Gerontol Geriatr. 2014;58(1):95–100.
58. Merchant RA, Chen MZ, Tan LWL, Lim MY, Ho HK, van Dam RM. Singapore Healthy Older People Everyday (HOPE) study: prevalence of frailty and associated factors in older adults. J Am Med Dir Assoc. 2017;18(8):734.e714–39.
59. Kojima G. Frailty as a predictor of disabilities among community-dwelling older people: a systematic review and meta-analysis. Disabil Rehabil. 2017;39(19):1897–908.
60. Dorresteijn TAC, Zijlstra GAR, Ambergen AW, Delbaere K, Vlaeyen JWS, Kempen GIJM. Effectiveness of a home-based cognitive behavioral program to manage concerns about falls in community-dwelling, frail older people: results of a randomized controlled trial. BMC Geriatr. 2016;16(1):2.
61. Grochtdreis T, Dams J, König H-H, Konnopka A. Health-related quality of life measured with the EQ-5D-5L: estimation of normative index values based on a representative German population sample and value set. Eur J Health Econ. 2019;20(6):933–44.
62. Huber MB, Felix J, Vogelmann M, Leidl R. Health-related quality of life of the general German population in 2015: results from the EQ-5D-5L. Int J Environ Res Public Health. 2017;14(4):426.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

4 Publication 2 – Cost-effectiveness of a group vs individually delivered exercise program in community-dwelling persons aged ≥70 years

Published in the *Journal of the American Medical Directors Association* as:

Gottschalk S[1], König HH[1], Schwenk M[2], Nerz C[3], Becker C[3], Klenk J[3,4,5], Jansen CP[2,3], Dams J[1].
Cost-Effectiveness of a Group vs Individually Delivered Exercise Program in Community-Dwelling Persons Aged ≥70 Years. J Am Med Dir Assoc. 2022 May;23(5):736-42.e6. DOI: 10.1016/j.jamda.2021.08.041.

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Network Aging Research, Heidelberg University, Heidelberg, Germany.

[3]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany.

[4]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.

[5]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany.

JAMDA xxx (2021) 1–7

JAMDA

journal homepage: www.jamda.com

Original Study

Cost-Effectiveness of a Group vs Individually Delivered Exercise Program in Community-Dwelling Persons Aged ≥70 Years

Sophie Gottschalk MSc [a,*], Hans-Helmut König MD, MPH [a], Michael Schwenk PhD [b], Corinna Nerz PhD [c], Clemens Becker MD [c], Jochen Klenk PhD [c,d,e], Carl-Philipp Jansen PhD [b,c], Judith Dams PhD [a]

[a] Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg Center for Health Economics, Hamburg, Germany
[b] Network Aging Research, Heidelberg University, Heidelberg, Germany
[c] Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany
[d] Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany
[e] IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany

Keywords:
Cost-effectiveness
economic evaluation
falls
physical activity
older adults

ABSTRACT

Objectives: Interventions aimed at reducing falls and physical inactivity could alleviate the economic burden attributable to these factors. The study aimed to analyze the cost-effectiveness of a group-delivered version of the Lifestyle-integrated Functional Exercise Program compared with an individually delivered program version.
Design: An economic evaluation conducted alongside the LiFE-is-LiFE randomized non-inferiority trial.
Interventions: Group and individually delivered version of a program consisting of strength and balance exercises integrated into everyday activities to prevent falls.
Setting and participants: 309 community-dwelling older adults (aged ≥70 years) at risk of falling recruited around Heidelberg and Stuttgart (Germany).
Methods: Cost-effectiveness of the group program was assessed over 6 months using different effect measures [quality-adjusted life years (QALYs, EQ-5D-5L), physical activity (mean number of steps/day), and falls] and cost perspectives (societal and payer's). Incremental cost-effectiveness ratios were determined, and cost-effectiveness acceptability curves were constructed.
Results: From a societal perspective, mean costs, the number of falls, and the number of steps/day were somewhat higher in the group program, whereas QALYs were almost identical between the 2 interventions. From the payer's perspective, the incremental cost-effectiveness ratio for the group compared to the individual program were €56,733 per QALY and €4755 per fall prevented. Based on the cost-effectiveness acceptability curves, the cost-effectiveness of the group program had to be rated as uncertain for both effect measures and perspectives. In contrast, it demonstrated cost-effectiveness for increasing physical activity at willingness-to-pay values per additional 1000 steps/day of €1600 (societal perspective) or €600 (payer's perspective).
Conclusions and Implications: Compared to the individual program, the group program might be cost-effective for increasing physical activity in older adults but was unlikely to be cost-effective with regard to QALY or for preventing falls. The cost-effectiveness should be evaluated long-term and compared to a regular care group.

This work was supported by the German Federal Ministry of Education and Research, Germany (grant: 01GL1705A-D). The funding source takes no part in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.
The authors declare no conflicts of interest.

* Address correspondence to Sophie Gottschalk, Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany.
E-mail address: s.gottschalk@uke.de (S. Gottschalk).

https://doi.org/10.1016/j.jamda.2021.08.041
1525-8610/

A major challenge in aging societies is the promotion of healthy aging, which has been declared as national health goal in Germany.[1] Physical activity is regarded as a main contributor to healthy aging as it is lowering the risk of aging-associated morbidity and loss of quality of life.[2–4] Conversely, the most common consequences of physical inactivity (eg, ischemic heart disease, stroke, cancers, type 2 diabetes) are associated with high costs, which are projected to increase in Germany, driven by the aging society.[5] Another driver of morbidity in the older population are falls, with an annual incidence higher than 30% in those aged ≥65 years.[6,7] Consequences of falls, such as injuries and a decline in health-related quality of life (HrQoL),[8,9] do not only impact the life of those affected but are also associated with high health care costs.[8,10]

Physical activity and falls are related, such as there is evidence that a higher activity level is associated with a lower risk of falling in the older population.[11] Conversely, a high activity level was found to be associated with an increased risk of falling in people with limited functional capacity.[12,13] Accordingly, promotion of physical activity in older adults should be accompanied by fall prevention strategies.[14,15] Thereby, such interventions may prevent negative fall-related repercussions, a decline in HrQoL, and potentially save health care costs.

This is becoming increasingly important in the context of an increasing older population as the available budget to cover health care costs and health care resources are limited. Economic evaluations can aid in the efficient allocation of resources by comparing costs and effects of interventions with each other.[16]

Several fall prevention programs have been developed and economically evaluated.[17,18] Evidence from systematic reviews suggests that diverse exercise-based interventions, especially those focusing on strength and balance training, are cost-effective strategies to prevent falls.[19,20] However, economic evaluations of those programs in a German setting are scarce. Because cultures, policies, and resources vary between countries and different health systems, the results of cost-effectiveness analyses are hardly transferable to other countries,[18,20] thus emphasizing the need for such evaluations. Moreover, programs that address both physical activity and falls have rarely been economically evaluated.

In the Lifestyle-integrated Functional Exercise Program (LiFE),[21] balance and strength training are integrated into everyday routines in order to achieve long-term sustainability of the LiFE activities.[21,22] It has demonstrated clinical effectiveness in terms of improving physical function and activity, while simultaneously reducing functional disability and falls in an Australian setting.[23] Thus, LiFE is a highly promising intervention concept specifically tailored to older adults' daily life.

However, implementing the individually delivered version would entail high costs, whereas a group program might reduce intervention costs. Because evidence from international literature found no indication that group-delivered exercise programs are inferior to individually delivered programs,[20] it is assumed that the group program constitutes a cost-saving alternative without having to sacrifice on effectiveness.

Therefore, the LiFE-is-LiFE trial aimed to evaluate the group program for its non-inferiority compared with the individual program.[24] As part of this trial, the current study aimed to investigate the cost-effectiveness of the group program compared to the individual program.

Methods

Study Design and Participants

Data were taken from the LiFE-is-LiFE multicenter, 2-armed, single-blinded, randomized trial (registered at clinicaltrials.gov; identifier: NCT03462654).[24] Recruitment took place in Heidelberg and Stuttgart (Germany) between April 2018 and July 2019. Community-dwelling, German-speaking people aged ≥70 years at risk of falling, who were able to ambulate 200 m without personal assistance were included. Participants were excluded if they exceeded a certain activity level (eg, exercising >1 time per week in the past 3 months), knew they would be unavailable during the intervention time or for the follow-up assessments, participated in another scientific trial, or had certain medical conditions that affect the ability to perform the LiFE activities (eligibility criteria are described in detail elsewhere[24]). Participants were randomized to either the individual or the group LiFE program. Data was obtained at baseline prior to randomization (T0) and at 6-month follow-up (T1). The study was approved by the Ethics Review Boards of the Faculty of Behavioral and Cultural Studies at Heidelberg University and the University Hospital and Faculty of Medicine Tübingen, and all subjects gave written informed consent to participate in the study.

Interventions

LiFE consists of 7 home visits (≈1-hour) over 11 weeks during which a trainer presents activities for balance, strength, and general physical activity and gives instructions on how to independently execute and implement these activities in a participant's individual daily routine. In the group version, the program was taught by 2 trainers in 7 sessions (≈2 hours) to groups of 8 to 12 participants. The activities were introduced in a predetermined order and the sessions followed a detailed curriculum, whereas the individual program could be adapted more flexibly to the individual participant's needs. The development of the conceptual framework of the group program is described elsewhere.[25] In both interventions, participants received 2 additional "booster phone calls" 4 and 10 weeks after the last intervention session.

Health Service Use and Costs

Intervention costs

Intervention costs for both program versions were calculated based on the personnel and material costs, travel expenses, and room rent that was incurred for conducting the trainer workshop, the training sessions, and/or phone calls (the calculation is described in detail elsewhere[26]). Because study conditions deviate from real-world implementation conditions (eg, regarding the trainers' salary level or group size), a second, real-world scenario of intervention costs was calculated. Data and assumptions underlying the calculations of each scenario are summarized in Supplementary Table 1.

Costs of health-related resource utilization

Health-related resource utilization was assessed 6 months retrospectively with an adapted version of the questionnaire for the use of medical and nonmedical services in old age (FIMA).[27] The FIMA covers outpatient (physician and nonphysician) and inpatient services (stays in general hospitals, rehabilitation clinics, and psychiatric clinics), and formal and informal care (ambulatory nursing service, paid household help, help from friends and family, utilization of a day care facility). Resource utilization from a societal perspective was monetarily valued in Euro (€) based on standardized unit costs,[28] inflated to the year 2018 according to the consumer price index.[29] From the payer's perspective, informal care costs and participants' travel expenses associated with the group sessions were excluded. Furthermore, data on medication use were collected and monetarily valued with the pharmacy retail price of the German official pharmaceutical index.[30] No indirect costs due to absenteeism or productivity loss were considered because participants were not active labor market participants anymore. As the time horizon of this study was 6 months, costs were not discounted.

Effects

Three different effect measures were used in this study: quality-adjusted life years (QALYs), physical activity, and the number of falls within 6 months.

QALYs were calculated based on the EQ-5D-5L,[31] which consists of the dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Participants rated their health on a scale from no problems (1) to extreme problems (5). By combining the answers, an individual health state was obtained for each participant, with 11111 and 55555 representing the best and worst health state, respectively. Health states were transformed to the EQ-5D index based on preference-based value sets from the German general population,[32] with 0 representing death and 1 representing perfect health. QALYs for the follow-up period of 6 months were calculated using linear interpolation between the index scores at T0 and T1.

Physical activity was quantified as the number of steps/day assessed using an accelerometer (activPAL4 micro; PAL Technologies Ltd, Glasgow, Scotland), which the participants wore centrally on one thigh for 7 full days at T0 and T1.

The number of falls was based on retrospective self-report (T0) and self-reported data from fall calendars that were sent back monthly to the respective study center (T1). Falls were defined based on common recommendations.[33]

Statistical Analyses

Because the analyses were planned as intention-to-treat analyses, missing data were imputed using multiple imputation by chained equations (MICE) with predictive mean matching as the imputation method.[34,35] Missing data resulted from loss to follow-up (n = 44; 14%) and from occasional missing information on resource use. The percentage of missing values varied between 0% and 17% across different variables. In total, 10 data sets were created.

Differences in costs and effects between the group program and the individual program at T1 were calculated using linear regression models adjusting for age, sex, the number of comorbidities [diabetes type 1 and 2, hypertension, acute cardiovascular disease, cardiac defect, cardiac arrhythmias, arthrosis, rheumatoid arthritis, cancer (not on active treatment), asthma or chronic obstructive pulmonary disease (Gold class < III), osteoporosis, depression, or history of heart attacks or stroke], and the respective baseline costs or effects. Bootstrapped standard errors (SEs; 1000 replications) were calculated to account for the skewed distributions.

For all effect measures, unadjusted incremental cost-effectiveness ratios from a societal and payer's perspective were calculated as the ratio of the difference in mean costs and the difference in mean effects between the group and the individual program. The incremental cost-effectiveness ratio is a point estimate and does not account for uncertainty around this estimate. To determine the probability of cost-effectiveness at different levels of willingness to pay (WTP), cost-effectiveness acceptability curves were constructed based on the net benefit approach.[36,37] A series of linear regression models with bootstrapped SEs (1000 replications) was performed, using the individual net monetary benefit ($NMB_i = WTP * E_i - C_i$; E_i and C_i denote the observed effect and costs of patient i at T1) as dependent variable and the group variable as independent variable. Baseline characteristics (age, sex, comorbidities, number of falls and/or steps/day, costs, and health utility) were included as covariates in the models.

As a robustness check, the analyses were additionally performed for a per-protocol sample (n = 280). The per-protocol criterion was defined as having attended at least 5 training sessions. Furthermore, an additional analysis including real-world intervention costs and QALYs as effect measure was carried out.

All statistical analyses were conducted using Stata/SE 16.0 (StataCorp, College Station, TX). The significance level was set to .05.

Results

Baseline characteristics are displayed in Table 1. Participants (N = 309) had a mean age of 78.7 (SE = 0.3). The majority were female (73.5%) and had an intermediate or high educational level (67.0%). Overall, 45.3% were married or living in a partnership, 35.9% were widowed, 12.0% were divorced, and 6.8% were single or permanently living separated. On average, participants reported 2.5 (SE = 0.1) chronic conditions. In addition, 40.8% had experienced at least 1 fall in the previous 6 months, and the mean EQ-5D index of the sample was 0.84 (SE = 0.01). No significant differences between the program versions could be observed.

Cost-effectiveness

Table 2 displays the unadjusted mean 6-month costs and effects at T1. The change in mean HrQoL, physical activity, and the number of

Table 1
Baseline Characteristics of Community-Dwelling Older People at Risk of Falling Participating in the LiFE-is-LiFE Trial

Characteristics	Total Sample (N = 309)	Group Program (n = 153)	Individual Program (n = 156)	*P* Value*
Female, n (%)	227 (73.5)	112 (73.2)	115 (73.7)	>.05
Age	78.7 (0.3)	78.7 (0.4)	78.8 (0.4)	>.05
Educational degree†, n (%)				
Low	94 (30.4)	44 (28.8)	50 (32.1)	>.05
Intermediate	92 (29.8)	41 (26.8)	51 (32.7)	>.05
High	115 (37.2)	62 (40.5)	53 (34.0)	>.05
Other	5 (1.6)	4 (2.6)	1 (0.6)	>.05
No degree	3 (1.0)	2 (1.3)	1 (0.6)	>.05
Marital status, n (%)				
Married/living in a partnership	140 (45.3)	74 (48.4)	66 (42.3)	>.05
Widowed	111 (35.9)	50 (32.7)	61 (39.1)	>.05
Divorced	37 (12.0)	20 (13.1)	17 (10.9)	>.05
Permanently living separated	3 (1.0)	1 (0.7)	2 (1.3)	>.05
Single	18 (5.8)	8 (5.2)	10 (6.4)	>.05
Living alone, n (%)	166 (53.7)	76 (49.7)	90 (57.7)	>.05
Chronic conditions	2.5 (0.1)	2.5 (0.1)	2.5 (0.1)	>.05
Prevalence of fallers, n (%)	126 (40.8)	63 (41.2)	63 (40.4)	>.05
Number of falls among fallers	1.6 (0.1)	1.6 (0.2)	1.6 (0.1)	>.05
EQ-5D index	0.84 (0.01)	0.83 (0.01)	0.85 (0.01)	>.05

Unless otherwise noted, values are mean (SE).

*Based on *F* test.

†Low (9 years of school education), intermediate (10 years of school education), high (qualifies to enter university).

Table 2
Unadjusted Costs* and Effects After 6 Months (T1) of Community-dwelling Older People at Risk of Falling Participating in the LiFE-is-LiFE Trial

Category	Group Program, Mean (SE)	Individual Program, Mean (SE)
Costs		
Outpatient services	386.7 (22.0)	383.0 (28.2)
Inpatient services	616.9 (179.9)	984.0 (329.9)
Informal and formal services	914.0 (228.8)	502.3 (145.4)
Formal services	24.9 (14.1)	14.0 (10.4)
Medication	515.0 (89.9)	395.4 (34.9)
Intervention		
Study conditions	228.9	350.1
Without travel costs for participants	211.0	350.1
“Real world”	120.7	332.1
Total costs: societal perspective	2661.6 (327.3)	2614.8 (378.6)
Total costs: payer’s perspective†	1754.5 (211.1)	2126.5 (338.9)
Effects		
QALYs	0.415 (0.006)	0.421 (0.005)
Steps per day	6847 (258)	6242 (255)
Number of falls‡	0.40 (0.06)	0.32 (0.05)

*Costs are based on 2018 Euros (€).
†Excluded cost categories: informal care and travel costs for participants.
‡Over a 6-month observation period.

falls between baseline and T1 is shown in Figure 1. Mean total costs from the societal perspective were €2662 for the group program and €2615 for the individual program (+€47, $P > .05$ for the group program). From the payer’s perspective, mean total costs were €1755 for the group program and €2127 for the individual program (−€372, $P > .05$). Looking at the unadjusted effects, the group program had almost identical QALYs (−0.007, $P > .05$), a higher mean number of steps/day (+605, $P > .05$), and a higher mean number of falls within 6 months (+0.08, $P > .05$) compared with the individual program. Thus, incremental cost-effectiveness ratios for the group program were €56,733 per QALY and €4755 per fall prevented from the payer’s perspective. In terms of physical activity, the group program dominated the individual program from the payer’s perspective.

None of the unadjusted cost or effect differences reached statistical significance. After adjusting for baseline characteristics (Supplementary Table 2), the differences remained nonsignificant, except for physical activity, where a significant difference of +799 steps/day ($P < .05$) in favor of the group program was found.

The cost-effectiveness acceptability curves based on adjusted total costs and QALYs indicated that the cost-effectiveness of the group program was uncertain for WTP values between €0 and €150,000 per QALY gained (Figure 2A). The cost-effectiveness acceptability curves using steps/day as effect measure indicated that the group program was cost-effective at WTP ≥€1600 and ≥€600 from a societal and payer’s perspective, respectively (Figure 2B). For the prevention of falls, the adjusted cost-effectiveness acceptability curves indicated that the group program was not cost-effective (Figure 2C).

Supplemental Analyses

Overall, the results of the intention-to-treat analyses were confirmed by the per-protocol analyses (Supplementary Tables 3 and 4, Supplementary Figures 1-3). However, the cost-effectiveness acceptability curves indicated a slightly higher probability of cost-effectiveness compared to the intention-to-treat analysis. Similarly, the probability of cost-effectiveness was slightly higher taking “real world” intervention costs into account (Supplementary Figure 4).

Discussion

The current study compared, for the first time, the cost-effectiveness of a group version of the LiFE program with an individually delivered version in a community-dwelling older population at risk of falling in Germany. The results were mixed, depending on the effect measure and cost perspective. The cost-effectiveness with regard to QALYs or prevented falls had to be rated as uncertain, although the probability of cost-effectiveness was considerably higher from the payer’s than from the societal perspective. In contrast, the group program demonstrated possible cost-effectiveness for increasing physical activity: it was cost-effective at a WTP for an additional 1000 steps/day of €600 (payer’s perspective) and €1600 (societal perspective).

HrQoL did not differ between the group and the individual program, and no significant change in HrQoL could be observed between T0 and T1 for either group. Similar results were obtained by other authors who analyzed the cost-effectiveness of a strength and balance training program and did not find a clinically important change in HrQoL after 6 months.[38] These results contrast with previous studies that found an effect of strength and balance exercises on HrQoL.[39,40] However, the baseline EQ-5D index values of the current study’s sample were higher than normative values for the respective age group in the general German population[41] and the EQ-5D was found to be less sensitive to change in people with high baseline utilities,[42] meaning that people whose HrQoL is already high are less likely to show any further improvements.

To account for the possible insufficient responsiveness of the EQ-5D as well as for the fact that the interventions did not primarily aim to improve HrQoL, the cost-effectiveness of the group program was also examined using clinical effect measures.

It was found that the group program might be a cost-effective strategy to increase physical activity in terms of the number of steps/day. Walking is an activity that requires relatively low effort, no equipment, and can easily be integrated into everyday life, and therefore lends itself well for formulating simple and attainable public health recommendations. These could be translated into more interpretable units, such as 1000 steps at an age-typical step length of 60 to 70 cm correspond to 600 to 700 m,[43] which in turn could correspond to a short walk around the neighborhood or to the supermarket. As walking an additional 1000 steps/day was associated with a lower risk of all-cause mortality and cardiovascular disease morbidity,[44] the group program might be able to contribute to the promotion of physical activity and healthy aging. As lowering the burden of morbidity may in the long run translate into cost-savings for the health care system, promoting activity may reduce the economic burden attributable to physical inactivity.[45] It was shown that even a 5% reduction in physical inactivity can result in substantial cost

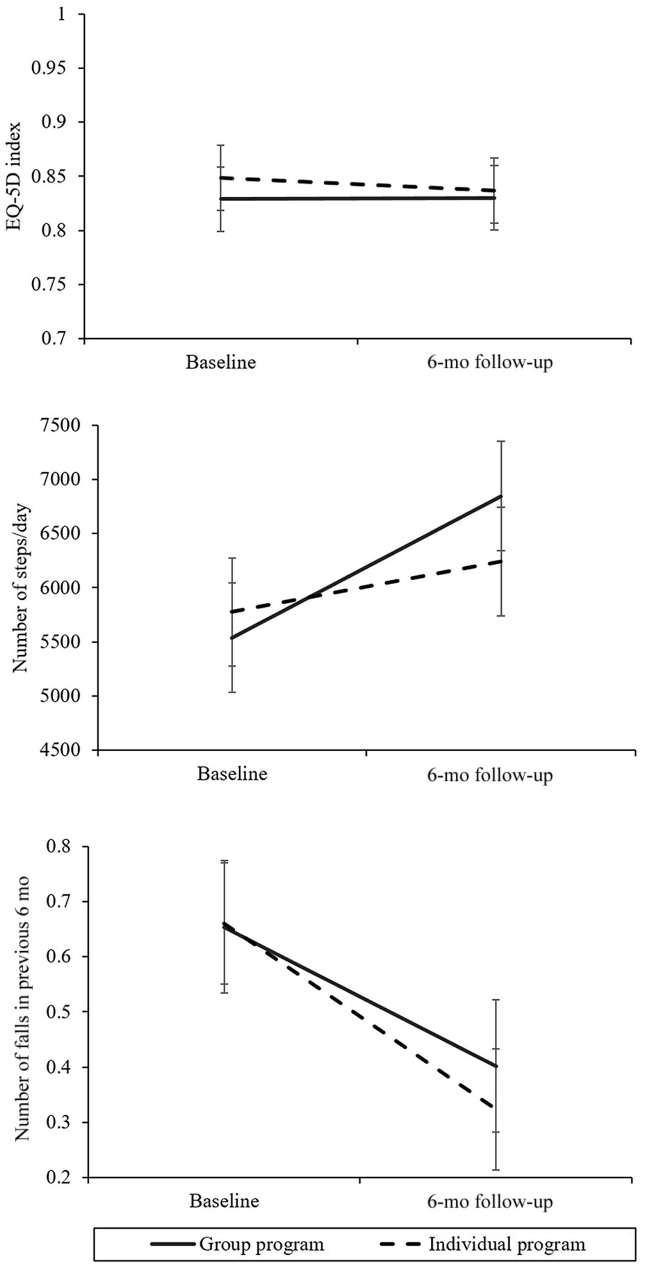


Fig. 1. Change in mean effects (95% CIs) between baseline and 6-month follow-up by intervention group; community-dwelling older people at risk of falling participating in the LiFE-is-LiFE trial (results on effectiveness of the programs are reported in detail elsewhere[26]).

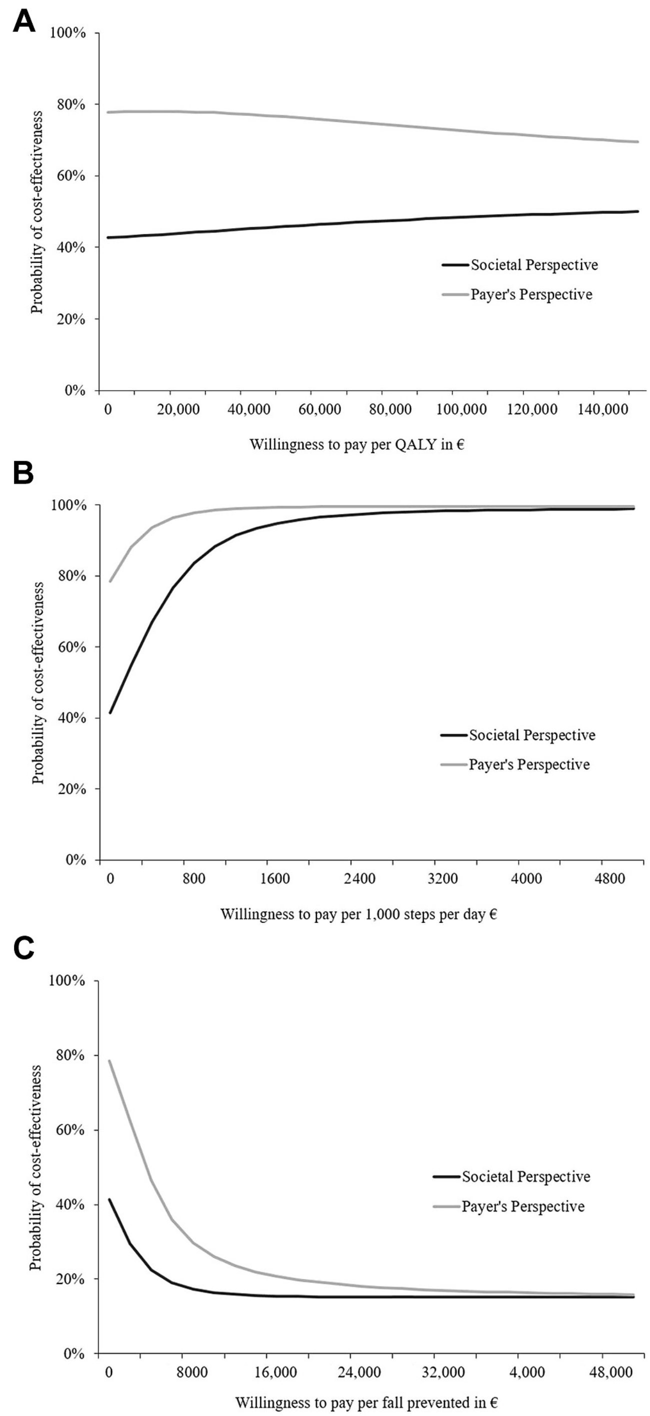


Fig. 2. (A) Cost-effectiveness acceptability curves from a societal and payer's cost perspective; probability of cost-effectiveness for the group program compared to the individual program for different WTP values per QALY; community-dwelling older people at risk of falling participating in the LiFE-is-LiFE trial. (B) Cost-effectiveness acceptability curves from a societal and payer's cost perspective; probability of cost-effectiveness for the group program compared to the individual program for different WTP values per 1000 steps; community-dwelling older people at risk of falling participating in the LiFE-is-LiFE trial. (C) Cost-effectiveness acceptability curves from a societal and payer's cost perspective; probability of cost-effectiveness for the group program compared to the individual program for different WTP values per fall prevented; community-dwelling older people at risk of falling participating in the LiFE-is-LiFE trial.

savings.[5] However, because consequences of physical inactivity, such as noncommunicable diseases, are typically long-term consequences, the benefits of increasing activity on the level of health care costs or HrQoL may only be observed in the long run as well. Thus, the observation period of 6 months may not have been sufficient to find clinically important effects on HrQoL. It further remains unclear how long the effect on physical activity lasts. Hence, a long-term evaluation of the costs and effects of the group and individual program is needed. It would be of particular interest whether the social interaction component (eg, peer support) of the group program fosters long-term adherence. Moreover, it remains an open question whether increasing the number of steps/day by 1000 steps yields enough value for money that needs to be invested (eg, €1600 from a societal perspective).

In this study, the group program was not cost-effective for the prevention of falls compared to the individual program as the participants in the group program had a nonsignificantly higher number of falls at T1. However, between T0 and T1, the incidence of falls was reduced in both groups (37% group program, 55% individual program).[26] Hence, future studies should evaluate the cost-effectiveness of the programs compared with a regular care control group. Fall research is also increasingly moving toward examining falls in relation to exposure time.[46–48] Participants in the group program fell on average slightly more often than in the individual program, but at the same time their exposure time in terms of steps/day increased. In the long run, combining the LiFE activities with increases in physical activity may counteract the frailty process and thereby may also decrease the risk of falling as a direct manifestation of the frailty process.[49] A future research direction may be to include a combined endpoint of falls relative to physical activity as effect measure in cost-effectiveness analyses. However, interpreting the difference between groups of such a combined endpoint might be challenging, especially because the WTP for an additional unit of effect is unknown.

Based on the cost-effectiveness results of this study, no clear recommendation can be made regarding which program version should be implemented. Aspects such as how or by whom the intervention is financed also play a role in the decision-making process. For example, assuming that intervention costs were borne primarily by participants and subsidized by health insurers at a certain amount regardless of the program version, the choice between the group or individual program would be based on individual preferences and the WTP of the participants.

Strengths and Limitations

The current study is the first that economically evaluated a group and an individually delivered version of the LiFE program in terms of costs and cost-effectiveness. Thereby, it complements the studies on clinical effectiveness[23,26] and contributes to the decision-making process regarding a possible implementation of the program. The analyses were based on a relatively large sample, and few missing values were observed.

However, this study has some limitations. As common for clinical trial data, the LiFE-is-LiFE trial is likely to be underpowered for cost-effectiveness analyses and determined only short-term effects over a limited time period. Nevertheless, a recall bias in the retrospective assessment of health-related resource utilization cannot be excluded. Despite randomization, differences in mean health care costs at T0 and T1 could be observed for some cost categories. Although these differences were large, they were not significant and are unlikely to be attributable to the intervention. Moreover, the number of steps/day represents only 1 way of measuring activity. Physical activity is complex and involves activities beyond walking, therefore making it hard to predict the long-term effects of increasing the number of steps/day on health and economic outcomes. Finally, the generalizability of the results to other health care systems or populations outside the eligibility criteria of this study may be limited.

Conclusions and Implications

The current study is the first that economically evaluated an exercise-based intervention in community-dwelling older adults at risk of falling in Germany. It was shown that a group-delivered version of the LiFE program might be cost-effective for improving physical activity in terms of the number of steps/day at certain WTP values, but was unlikely to be cost-effective with regard to QALYs or for preventing falls compared to an individually delivered version of the program. To be able to make definite recommendations to decision makers, the cost-effectiveness should be evaluated in the long term and compared to a regular care group.

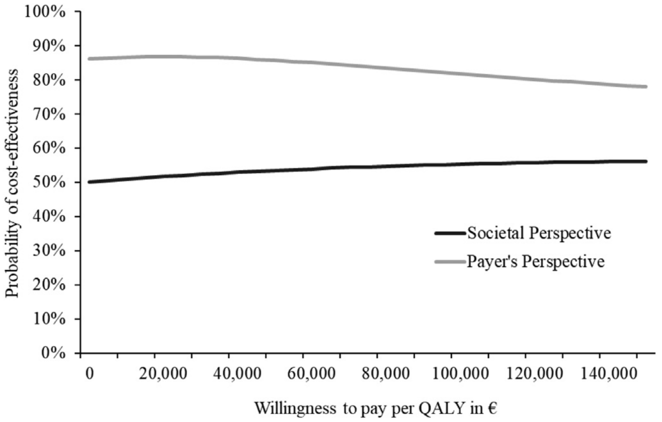
Acknowledgments

We thank all participants for their participation. We also thank all colleagues involved in the conduct of this trial, including the trainers and assessors, database managers, and members of the advisory board.

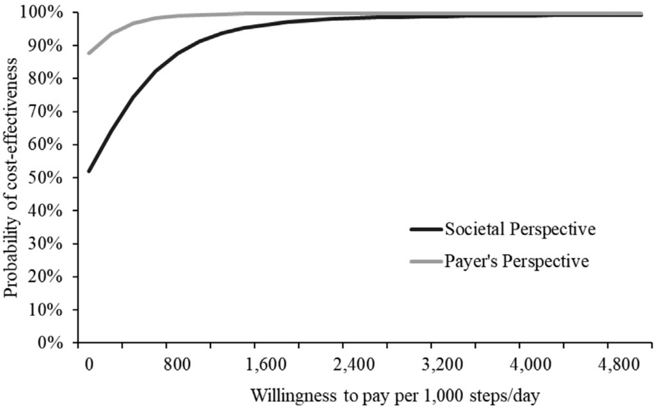
References

1. Federal Ministry of Health. National Health Goal - Healthy Ageing. Berlin: Federal Ministry of Health; 2012 [in German].
2. Daskalopoulou C, Stubbs B, Kralj C, et al. Physical activity and healthy ageing: A systematic review and meta-analysis of longitudinal cohort studies. Ageing Res Rev 2017;38:6–17.
3. Cunningham C, O' Sullivan R, Caserotti P, et al. Consequences of physical inactivity in older adults: A systematic review of reviews and meta-analyses. Scand J Med Sci Sports 2020;30:816–827.
4. Marquez DX, Aguiñaga S, Vásquez PM, et al. A systematic review of physical activity and quality of life and well-being. Transl Behav Med 2020;10: 1098–1109.
5. Dallmeyer S, Wicker P, Breuer C. How an aging society affects the economic costs of inactivity in Germany: Empirical evidence and projections. Eur Rev Aging Phys Act 2017;14:18.
6. Bergen G, Stevens MR, Burns ER. Falls and fall injuries among adults aged >/=65 years - United States, 2014. MMWR Morb Mortal Wkly Rep 2016;65: 993–998.
7. Schumacher J, Pientka L, Trampisch U, et al. The prevalence of falls in adults aged 40 years or older in an urban, German population. Results from a telephone survey. Z Gerontol Geriatr 2014;47:141–146.
8. Hartholt KA, van Beeck EF, Polinder S, et al. Societal consequences of falls in the older population: Injuries, healthcare costs, and long-term reduced quality of life. J Trauma 2011;71:748–753.
9. Tinetti ME, Williams CS. Falls, injuries due to falls, and the risk of admission to a nursing home. N Engl J Med 1997;337:1279–1284.
10. Heinrich S, Rapp K, Rissmann U, et al. Cost of falls in old age: A systematic review. Osteoporos Int 2010;21:891–902.
11. Lee J. The association between physical activity and risk of falling in older adults: A systematic review and meta-analysis of prospective cohort studies. Geriatr Nurs 2020;41:747–753.
12. Jefferis BJ, Merom D, Sartini C, et al. Physical activity and falls in older men: The critical role of mobility limitations. Med Sci Sports Exerc 2015;47:2119–2128.
13. Lewis ZH, Markides KS, Ottenbacher KJ, et al. The role of physical activity and physical function on the risk of falls in older Mexican Americans. J Aging Phys Act 2016;24:342–349.
14. Bauman A, Merom D, Bull FC, et al. Updating the evidence for physical activity: Summative reviews of the epidemiological evidence, prevalence, and interventions to promote "active aging.". Gerontologist 2016;56:S268–S280.
15. Lu Z, Lam FMH, Leung JCS, et al. The U-shaped relationship between levels of bouted activity and fall incidence in community-dwelling older adults: A prospective cohort study. J Gerontol A Biol Sci Med Sci 2020;75: e145–e151.
16. Edlin R, McCabe C, Hulme C, et al., editors. Cost Effectiveness Modelling for Health Technology Assessment. A Practical Guide. Cham, Switzerland: ADIS, Springer; 2015.
17. Davis JC, Robertson MC, Ashe MC, et al. Does a home-based strength and balance programme in people aged ≥80 years provide the best value for money to prevent falls? A systematic review of economic evaluations of falls prevention interventions. Br J Sports Med 2010;44:80–89.
18. Balzer K, Bremer M, Schramm S, et al. Falls prevention for the elderly. GMS Health Technol Assess 2012;8:Doc01.
19. Dubas-Jakóbczyk K, Kocot E, Kissimova-Skarbek K, et al. Economic evaluation of health promotion and primary prevention actions for older people—a systematic review. Eur J Public Health 2017;27:670–679.
20. Winser SJ, Chan HTF, Ho L, et al. Dosage for cost-effective exercise-based falls prevention programs for older people: A systematic review of economic evaluations. Ann Phys Rehabil Med 2020;63:69–80.
21. Clemson L, Munro J, Fiatarone Singh MA. Lifestyle-Integrated Functional Exercise (LiFE) Program to Prevent Falls. Trainer's Manual. Sydney: Sydney University Press; 2014.
22. Clemson L, Singh MF, Bundy A, et al. LiFE Pilot Study: A Randomised Trial of Balance and Strength Training Embedded in Daily Life Activity to Reduce Falls in Older Adults. Aust Occup Ther J 2010;57:42–50.
23. Clemson L, Fiatarone Singh MA, Bundy A, et al. Integration of balance and strength training into daily life activity to reduce rate of falls in older people (the LiFE study): Randomised parallel trial. BMJ 2012;345:e4547.

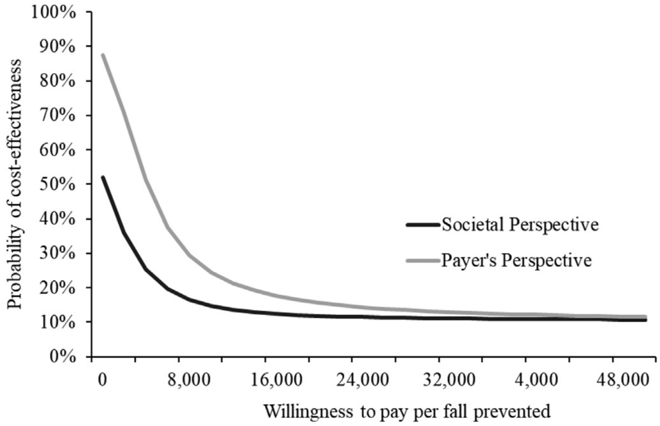
24. Jansen CP, Nerz C, Kramer F, et al. Comparison of a group-delivered and individually delivered lifestyle-integrated functional exercise (LiFE) program in older persons: a randomized noninferiority trial. BMC Geriatr 2018;18:267.
25. Kramer F, Labudek S, Jansen CP, et al. Development of a conceptual framework for a group-based format of the Lifestyle-integrated Functional Exercise (gLiFE) programme and its initial feasibility testing. Pilot Feasibility Stud 2020;6:6.
26. Jansen CP, Nerz C, Labudek S, et al. Lifestyle-integrated functional exercise to prevent falls and promote physical activity: Results from the LiFE-is-LiFE randomized non-inferiority trial. Int J Behav Nutr Phys Act. 18, 115.
27. Seidl H, Bowles D, Bock JO, et al. [FIMA—questionnaire for health-related resource use in an elderly population: development and pilot study]. Gesundheitswesen 2015;77:46—52 [in German].
28. Bock JO, Brettschneider C, Seidl H, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen 2015;77:53—61 [in German].
29. Organization for Economic Cooperation and Development. Inflation (CPI) (indicator) 2020.
30. Rote Liste Service GmbH. ROTE LISTE 2018: Pharmaceutical directory for Germany (including EU approvals and certain medical devices). Frankfurt/Main: Rote Liste Service GmbH; 2018 [in German].
31. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20:1727—1736.
32. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. Pharmacoeconomics 2018;36:663—674.
33. Lamb SE, Jørstad-Stein EC, Hauer K, et al. Development of a common outcome data set for fall injury prevention trials: The Prevention of Falls Network Europe consensus. J Am Geriatr Soc 2005;53:1618—1622.
34. StataCorp. Stata Multiple-Imputation Reference Manual. Release 16. College Station, TX: StataCorp LLC; 2019.
35. Azur MJ, Stuart EA, Frangakis C, et al. Multiple imputation by chained equations: What is it and how does it work? Int J Methods Psychiatr Res 2011;20:40—49.
36. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: A framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ 2002;11:415—430.
37. Briggs AH, O'Brien BJ, Blackhouse G. Thinking outside the box: Recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. Annu Rev Public Health 2002;23:377—401.
38. Davis JC, Khan KM, Hsu CL, et al. Action Seniors! Cost-effectiveness analysis of a secondary falls prevention strategy among community-dwelling older fallers. J Am Geriatr Soc 2020;68:1988—1997.
39. Bjerk M, Brovold T, Davis JC, et al. Health-related quality of life in home care recipients after a falls prevention intervention: A 6-month follow-up. Eur J Public Health 2020;30:64—69.
40. Bjerk M, Brovold T, Davis JC, et al. Evaluating a falls prevention intervention in older home care recipients: A comparison of SF-6D and EQ-5D. Qual Life Res 2019;28:3187—3195.
41. Grochtdreis T, Dams J, König HH, et al. Health-related quality of life measured with the EQ-5D-5L: Estimation of normative index values based on a representative German population sample and value set. Eur J Health Econ 2019;20:933—944.
42. Joore M, Brunenberg D, Nelemans P, et al. The impact of differences in EQ-5D and SF-6D utility scores on the acceptability of cost-utility ratios: results across five trial-based cost-utility studies. Value Health 2010;13:222—229.
43. Hollman JH, McDade EM, Petersen RC. Normative spatiotemporal gait parameters in older adults. Gait Posture 2011;34:111—118.
44. Hall KS, Hyde ET, Bassett DR, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020;17:78.
45. Ding D, Lawson KD, Kolbe-Alexander TL, et al. The economic burden of physical inactivity: A global analysis of major non-communicable diseases. Lancet 2016;388:1311—1324.
46. Klenk J, Kerse N, Rapp K, et al. Physical activity and different concepts of fall risk estimation in older people—Results of the ActiFE-Ulm Study. PLoS One 2015;10:e0129098.
47. Wijlhuizen GJ, Chorus AMJ, Hopman-Rock M. The FARE: A new way to express FAlls Risk among older persons including physical activity as a measure of Exposure. Prev Med 2010;50:143—147.
48. Del Din S, Galna B, Lord S, et al. Falls risk in relation to activity exposure in high-risk older adults. J Gerontol A Biol Sci Med Sci 2020;75:1198—1205.
49. McPhee JS, French DP, Jackson D, et al. Physical activity in older age: Perspectives for healthy ageing and frailty. Biogerontology 2016;17:567—580.



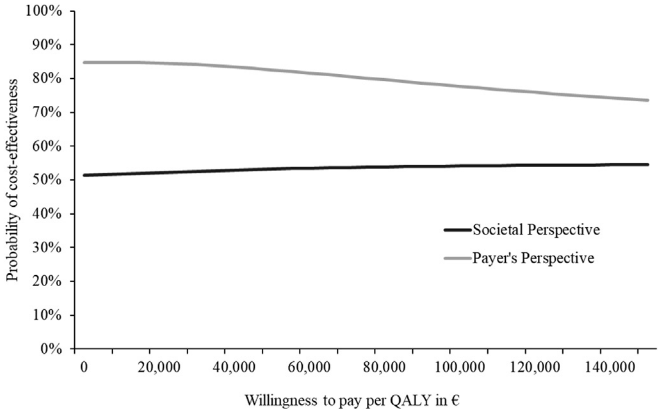
Supplementary Fig. 1. Cost-effectiveness acceptability curves (CEACs) from a societal and payer's cost perspective; probability of cost-effectiveness for the group program compared to the individual program for different WTP values per QALY; per protocol analysis. (QALY, quality-adjusted life year; WTP, willingness to pay.).



Supplementary Fig. 2. Cost-effectiveness acceptability curves (CEACs) from a societal and payer's cost perspective; probability of cost-effectiveness for the group program compared to the individual program for different WTP values per 1000 steps; per protocol analysis. (WTP, willingness to pay.).



Supplementary Fig. 3. Cost-effectiveness acceptability curves (CEACs) from a societal and payer's cost perspective; probability of cost-effectiveness for the group program compared to the individual program for different WTP values per fall prevented; per protocol analysis. (WTP, willingness to pay.).



Supplementary Fig. 4. Cost-effectiveness acceptability curves (CEACs) from a societal and payer's cost perspective; intervention costs based on "real world" assumptions; probability of cost-effectiveness for the group program compared to the individual program for different WTP values per QALY. (QALY, quality-adjusted life year; WTP, willingness to pay.).

Supplementary Table 1
Data and Assumptions for the Calculation of Intervention Costs by Scenario

	Study Conditions	"Real World"
Trainer workshop		
Duration, d	3	3
Salary group coaches conducting the workshop	TVöD* E13	TVöD E13
Number of trainers	8	20
Salary level of the trainers	TVöD E10 or E13	TVöD E8
Sessions: individual program		
Duration (including preparation and traveling time), min, mean	111	120
Number of sessions per participant, mean	6.49	7
"Booster" phone calls		
Number, n, mean	1.66	2.00
Duration, min, mean	27	30
Number of material sets per trainer	1	1
Sessions: group program		
Duration (including preparation and traveling time) †, min, mean	180	150
Initial number of participants per group, mean	10.2‡	12.0§
"Booster" phone calls		
Number, n, mean	1.65	2.00
Duration, min, mean	29	30
Room rent per session, €	50‖	0
Number of material sets per trainer pair	1	1
Other assumptions		
Number of sessions per trainer (pair) per week	Not relevant: 8 trainers taught the program to a similar number of LiFE/gLiFE participants (n = 156 and n = 153) over the study period	LiFE: 15 (30 h); gLiFE: 12 (30 h)

Study conditions = calculation of intervention costs as incurred during the study; "Real world" = calculation of intervention costs based on modified assumptions reflecting more realistic conditions in case the program is implemented than study conditions.

*TVöD: civil service collective agreement.

†In one study center, trainers had traveling time to an external room where the group sessions were held; in the "real world" scenario, no traveling time is taken into account for gLiFE (assumption: a suitable room is available at the workplace).

‡Average number of participants per session due to study dropout and occasional nonparticipation: 7.9.

§Assumption: sessions have to be paid regardless of participation.

‖For 8 of 15 groups (study center: Stuttgart).

Supplementary Table 2
Adjusted* Differences in Costs and Effects of the Group Program Compared to the Individual Program After 6 Months (T1): Community-Dwelling Older People at Risk of Falling Participating in the LiFE-is-LiFE trial—ITT Sample

Category	Mean	SE	95% Confidence Interval		*P* Value†
			Lower Limit	Upper Limit	
Outpatient services (€)	11.4	29.9	−47	70	.70
Inpatient services (€)	−274.5	319.9	−902	353	.39
Informal and formal care services (€)	311.6	219.9	−119	743	.16
Formal services (€)	1.9	12.1	−22	26	.87
Medication (€)	61.3	47.7	−32	155	.20
Intervention‡ (€)					
Study conditions	−121.2				
Without travel costs for participants	−139.1				
“Real world”	−211.5				
Total costs: societal perspective (€)	76.8	418.2	−743	896	.85
Total costs: payer's perspective§ (€)	−267.5	350.5	−954	419	.45
QALYs	0.001	0.004	−0.008	0.009	.90
Steps per day‖	799	302	207	1391	.01
Number of falls**	0.07	0.08	−0.088	0.222	.40

QALY, quality-adjusted life-year; SE, standard error.
*Adjusted for age, sex, baseline health utility (EQ-5D), baseline number of falls, baseline number of comorbidities, and baseline costs in the respective cost category.
†Based on *F* test.
‡Unadjusted.
§Excluded cost categories: informal care and travel costs for participants.
‖Adjusted for baseline steps per day instead of baseline health utility.
**Observation period >6 months.

Supplementary Table 3
Unadjusted Costs and Effects After 6 Months (T1)—Per Protocol Sample

Category	gLiFE		iLiFE	
	Mean	SE	Mean	SE
Outpatient services (€)	381	20	373	25
Inpatient services (€)	538	151	978	380
Informal and formal services (€)	884	237	424	115
Formal services (€)	27	16	13	10
Medication (€)	454	77	398	35
Intervention (€)				
Study conditions	229		350	
Without travel costs for participants	211		350	
Total costs: societal perspective (€)	2486	294	2523	408
Total costs: payer perspective* (€)	1611	176	2112	386
QALYs	0.417	0.006	0.421	0.005
Steps per day	6797	258	6266	251
Number of falls†	0.41	0.06	0.31	0.05

*Excluded cost categories: informal care and travel costs for participants.
†Over 6 months' observation period.

Supplementary Table 4
Adjusted* Differences in Costs and Effects After 6 Months (T1)—Per Protocol Sample

Category	Mean	SE	95% Confidence Interval		*P* Value†
			Lower Limit	Upper Limit	
Outpatient services (€)	12	28	−44	67	.68
Inpatient services (€)	−356	340	−1023	310	.29
Informal and formal care services (€)	382	198	−7	770	.05
Formal services (€)	6	13			.64
Medication (€)	39	35	−29	107	.26
Intervention‡ (€)					
Study conditions	−121				
Without travel costs for participants	−139				
Total costs: societal perspective (€)	−2	434	−852	848	>.99
Total costs: payer perspective§ (€)	−387	357	−1085	312	.28
QALYs	0.001	0.004	−0.008	0.009	.85
Steps per day‖	787	289	221	1353	.01
Number of falls**	0.10	0.08	−0.06	0.26	.24

*Adjusted for age, sex, baseline health utility (EQ-5D), baseline number of falls, baseline number of comorbidities, and baseline costs in the respective cost category.
†Based on *F* test.
‡Unadjusted.
§Excluded cost categories: informal care and travel costs for participants.
‖Adjusted for baseline steps per day instead of baseline health utility.
**Observation period >6 months.

5 Publication 3 – Comparison of falls and cost-effectiveness of the group vs. individually delivered Lifestyle-integrated Functional Exercise (LiFE) program: final results from the LiFE-is-LiFE non-inferiority trial

Accepted for publication in *Age and Ageing* as:

Jansen CP[1,2*†], Gottschalk S[3†], Nerz C[1], Labudek S[4], Kramer-Gmeiner F[4], Klenk J[1,5,6], Clemson L[7], Todd C[8,9,10], Dams J[3], König HH[3], Becker C[1], Schwenk M[4,11]. Comparison of falls and cost-effectiveness of the group vs. individually delivered Lifestyle-integrated Functional Exercise (LiFE) program: final results from the LiFE-is-LiFE non-inferiority trial. Age Ageing. [accepted].

[1]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany

[2]Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany

[3]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[4]Network Aging Research, Heidelberg University, Heidelberg, Germany

[5]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany

[6]IB University of Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany

[7]Sydney School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia

[8]School of Health Sciences, Faculty of Biology, Medicine & Health, The University of Manchester, Manchester M13 9PL, UK

[9]Manchester University NHS Foundation Trust, Manchester, UK

[10]Manchester Academic Health Science Centre, Manchester, UK

[11]Human Performance Research Centre, Department of Sport Science, University of Konstanz, Konstanz, Germany

[†]equal contributors

Comparison of falls and cost-effectiveness of the group vs. individually delivered Lifestyle-integrated Functional Exercise (LiFE) program: final results from the LiFE-is-LiFE non-inferiority trial

Abstract

Background: The individually delivered Lifestyle-integrated Functional Exercise (LiFE) was shown to improve physical activity (PA) and reduce fall incidence, however being rather resource-consuming due to one-to-one delivery. A potentially less resource-intensive group format (gLiFE) was developed and compared against the original program, considering higher risk of falling due to possible PA enhancement.

Objective: To investigate non-inferiority in terms of PA-adjusted fall risk and cost-effectiveness of gLiFE at 12-month follow-up.

Design: Single-blinded, randomized, multi-center non-inferiority trial.

Setting: Community.

Subjects: 309 adults aged 70+ years at risk of or with history of falling; n=153 in gLiFE, n=156 in LiFE.

Methods: LiFE was delivered one-to-one at the participants' homes, gLiFE in a group. PA-adjusted fall risk was analyzed using negative binomial regression to compare incidence rate ratios (IRR). Cost-effectiveness was presented by incremental cost-effectiveness ratios and cost-effectiveness acceptability curves, considering quality-adjusted life years, PA, and falls as effect measures. Secondary analyses included PA (steps/day) and fall outcomes.

Results: Non-inferiority was inconclusive (IRR 0.96; 95% CI 0.67; 1.37); intervention costs were lower for gLiFE, but cost-effectiveness was uncertain. gLiFE participants significantly increased PA (+1,090 steps/day; 95% CI 345; 1.835) vs. insignificant increase in LiFE (+569, 95% CI -31; 1,168). Number of falls and fallers were reduced in both formats.

Conclusion: Non-inferiority of gLiFE compared to LiFE was inconclusive after 12 months. Increases in PA were clinically relevant in both groups, although nearly twice as high in gLiFE. Despite lower intervention costs of gLiFE, it was not clearly superior in terms of cost-effectiveness.

Keywords: fall prevention; non-inferiority trial; economic evaluation; fall risk; physical activity promotion

Key points: (1) Non-inferiority of the group format compared to the individual LiFE format was inconclusive. (2) Despite lower intervention costs of gLiFE, no program version was clearly superior in terms of cost-effectiveness. (3) Both the group and the individual LiFE format showed meaningful improvements in daily physical activity and fall incidence. (4) The group format came with significant improvements of more than +1,000 steps/day.

Introduction

There is a large body of evidence highlighting the importance of physical activity (PA) for healthy aging [1]. However, there are certain conditions in which the enhancement of PA may come with some risk. Especially in older adults with a history of falling, more transitions from sit to stand or more steps/day may increase the risk of falling [2-5], although beneficial effects of PA are generally considered to be worth the risk. Hence, interventions that are capable of increasing PA to sufficient levels while at the same time reducing the risk of falls in older adults have the potential to alleviate the economic burden of falls and physical inactivity [6-8]. The use of a combined endpoint of fall risk and PA along these lines is the optimal approach to reflect this [9-11].

The 'Lifestyle-integrated Functional Exercise' (LiFE) randomised controlled trial by Clemson et al. [12] was a pivotal study, having induced exercise into older Australians' daily routines while reducing fall incidence. As a home-based intervention individually delivered by therapists, LiFE comes with considerable human resources and costs. Therefore, a group version (gLiFE) was developed and evaluated for its non-inferiority compared to the original LiFE program (LiFE) [13, 14]. Preliminary results from a previous analysis comparing gLiFE and LiFE at 6-month follow-up show that LiFE participants had a lower fall incidence rate ratio (IRR) than gLiFE participants but gLiFE had a considerable advantage over LiFE in terms of intervention costs [15].

However, from a societal or payer's perspective, not only costs associated with the implementation of a program are of interest, but also costs associated with health-related resource utilization. When considering both cost types in the 6-month cost-effectiveness analysis, results were mixed depending on the effect measure: cost-effectiveness of gLiFE compared to LiFE was uncertain regarding quality-adjusted life years (QALY) or the prevention of falls, but highly probable for increasing PA at certain willingness to pay (WTP) levels for an additional 1,000 steps/day [16]. However, the LiFE program aims for the development and long-term maintenance of LiFE exercise-infused routines that last beyond trainer support, which leads to the main question of the study: How do PA, fall incidence, and cost-effectiveness evolve over a longer period of time, especially after maintenance of LiFE activities had been put in the participants' own responsibility and trainers' support was terminated?

Against this background, this study presents the central analysis of the primary outcomes of the LiFE-is-LiFE trial as pre-specified in the study protocol [13]. The primary aim of the present work is twofold: 1) to evaluate gLiFE's non-inferiority in terms of PA-adjusted fall risk and 2) to examine the cost-effectiveness of gLiFE compared to LiFE at 12-month follow-up. As secondary outcomes, PA in terms of steps/day and fall-related outcomes were compared.

Methods

Study design

This study ("LiFE-is-LiFE") was a multi-center, single-blinded, randomized non-inferiority trial conducted in Heidelberg and Stuttgart, Germany (clinicaltrials.gov identifier: NCT03462654; registered March 12th 2018). The study protocol is available elsewhere [13]. Reporting in this article is aligned with the CONSORT extension in non-inferiority trials [17]. Data was obtained at baseline (T0), 6 months (T1), and 12 months (T2) after intervention start (±2 weeks; reference was the date of the first (g)LiFE session). Ethical approval was obtained for both sites, Heidelberg [Schwe2017 2/1–1]; Stuttgart [723/2017BO2]. All participants gave written informed consent.

Participants and eligibility criteria

Adults aged ≥70 years were drawn from municipality registries in Heidelberg and Stuttgart; participant flow is shown in Figure 1. Further information on recruitment and screening can be found in the study protocol [13]. To be eligible, participants had either a) experienced at least one injurious or multiple non-injurious falls in the year prior to study participation, or b) were designated as having high risk of falls when indicating balance decline in the past 12 months and needing ≥12 seconds for the "Timed Up-and-Go" (TUG) [18]. Participants who already reached WHO PA recommendation levels of 150 minutes of moderate to vigorous PA per week or exercised more than once per week were excluded. A detailed list of further exclusion criteria is provided elsewhere [13].

Randomization and blinding

Participants were randomized after baseline assessment into gLiFE or LiFE through block-randomization; assessors were blinded to group allocation. Further details on the randomization procedure are provided elsewhere [15].

Intervention programs

gLiFE was developed based on the original LiFE program. [12]. Highest possible fidelity was ensured by following specific manuals developed for the LiFE program [19, 20]. Methods and didactics were slightly adapted to accommodate the group setting. A detailed description of the intervention programs including a TIDieR checklist are part of the study protocol [13]; details on intervention development are presented in a publication on pilot results [14]. LiFE (in the participant's home) and gLiFE (in a group) were delivered in seven sessions during eleven weeks, plus two booster phone calls four and 10 weeks after the last intervention session. A total of seven balance activities, seven strength activities, and two PA promoting activities were presented to the participants. They were explained how to independently select, execute, and adapt the intensity of activities, including identification of daily situations into which activities can be integrated. gLiFE sessions took about two hours and were led by two trainers (physio or occupational therapists, sports scientists, health psychologists) with a maximum of 12 participants; LiFE sessions took about one hour and were led by one trainer.

Primary outcomes

To assess the combined endpoint, "activPAL4™ micro" accelerometers (PAL Technologies Ltd., Glasgow, Scotland) attached to participants' central front thigh measured daily PA for seven days (7 times 24 hours). Reliability and validity of the device have been established [21]. Data were used if at least two weekdays and the Sunday were fully captured [22]. With walking activity being denominated as the most hazardous PA in terms of fall risk [23, 24], PA was operationalised as mean steps/day and used as offset variable, that is, steps/day were introduced as denominator of exposure variable used for the analysis.

Falls were defined as "unexpected event in which the participant comes to rest on the ground, floor, or lower level" [25]. To collect data on falls in the past 12 months before baseline, falls were recorded using a retrospective question; at follow-up, information on falls (location, date, time, injuries, subsequent treatment, and fall circumstances) was collected using monthly fall calendars and followed-up via telephone calls [26].

For the cost-effectiveness analysis, three different effect measures (quality-adjusted life years [QALYs], PA, and falls) as well as intervention and resource utilization costs from a societal and healthcare payer's perspective were considered.

QALYs were calculated based on the EQ-5D-5L [27]. Participants rated their health on 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) on a scale from “no problems” (1) to “extreme problems” (5). The obtained EQ-5D health states were transformed to an index based on the German general population value set [28] (0 representing death and 1 representing perfect health). For the calculation of QALYs over 12 months, the index scores at T0, T1 and T2 were linearly interpolated.

Intervention costs for gLiFE and LiFE were calculated based on personnel and material costs, travel expenses, and room rent (described in detail elsewhere [15]).

Data on resource utilization was collected with an adapted version of the questionnaire for the use of medical and non-medical services in old age (FIMA), which covers outpatient and inpatient services and formal and informal care [29]. Resource use was monetary valued in Euro (€) based on standardized unit costs [30] and inflated to the year 2018 according to the consumer price index [31]. For the payer’s perspective, informal care costs and participants’ travel expenses associated with gLiFE sessions were excluded. Furthermore, data on medication use was collected and monetary valued based on the pharmacy retail price of the German official pharmaceutical index [32]. Indirect costs (e.g., absenteeism or productivity loss) were not considered since participants were no active labor market participants anymore. Given the 12-month time horizon, costs were not discounted.

Secondary outcomes

Mean steps/day were used to compare PA trajectories between groups. Fall outcomes were number of falls in the past year, fall rate per person year, number of fallers, and number of frequent fallers (>1 fall in the past year).

Participant characteristics

Participant characteristics included age, sex, body-mass index, number of medications, number of comorbidities, and cognitive status (Montreal Cognitive Assessment) [33].

Sample size and non-inferiority margin

Sample size for the non-inferiority analysis was calculated based on 12-month data from the LiFE trial [12]; further information on this calculation and the non-inferiority margin (Δ of 20%) is provided in the study protocol [13].

Statistical analyses

Analyses were carried out according to the intention-to-treat principle (ITT); all randomized participants were included. As a robustness check, primary analyses were additionally performed for a per protocol sample (n=280) of participants who attended ≥5 training sessions. The overall percentage of missing values across different variables varied between 0% (e.g., baseline variables) and 24% (some T2 variables). Missing data was imputed using multiple imputation by chained equations (MICE) with predictive mean matching [34]. In total, 10 datasets were created based on data from T0, T1, and T2 assessments.

Negative binomial regression was used to compare IRRs of falls between gLiFE and LiFE participants. For the combined endpoint, mean steps/day were log-transformed and used as exposure variable (offset) in the regression model. In this way, observations (number of falls) were modified from a count into a rate (per steps/day). Non-inferiority was determined if the upper limit of the two-sided 95% confidence interval (CI) for gLiFE remains below the predefined non-inferiority margin (upper CI of IRR <1.20). CIs for comparing changes between T0 and T2 in PA (steps/day) were calculated with a generalized linear model with repeated measures.

Cost-effectiveness analysis

Differences in costs and effects (QALYs, PA, falls) between gLiFE and LiFE over the 12-month observation period were calculated using linear regression models adjusted for age, sex, number of comorbidities, self-reported problems of being physically active due to the COVID-19 pandemic (yes/no), and the respective baseline costs, EQ-5D, PA, or falls. For all effect measures, unadjusted incremental cost-effectiveness ratios (ICER) from a societal and payer's perspective were calculated as the ratio of incremental mean costs and incremental mean effects between gLiFE and LiFE. To account for uncertainty around the ICER and to determine the probability of cost-effectiveness at different WTP levels, cost-effectiveness acceptability curves (CEAC) were constructed based on the net benefit approach [35], performing a series of linear regression models adjusted for the covariates mentioned above. Bootstrapped standard errors (1,000 replications) were calculated to account for skewed distributions.

Analyses were performed using SPSS (Version 27.0. Armonk, NY: IBM Corp) and STATA/SE 16.0 (StataCorp. 2019. Release 16. College Station, TX: StataCorp LLC).

Results

Participant flow and baseline characteristics

309 persons were randomized from June 2018 to July 2019 into gLiFE (n=153) and LiFE (n=156); 15 persons dropped out before the intervention started. At T2, 61 participants

(19.7%) were lost to follow-up, n=29 in gLiFE (19.0%) and n=32 in LiFE (20.5%) (see Figure 1). Of the envisaged sensor-based 7-day PA measurement, at least 6 full days were completed by 99.0% of the participants at T0 and 96.6% at T2 (of those who attended). The mean observation duration until the first missing calendar was 321 days (SE=5.2) of 360 possible days. 62% of the participants had completed all calendars until end of the study, 80% had 330 days or more. Similarity of groups at T0 indicates successful randomization (Table 1); no significant differences were observed (data not shown). The majority was female, cognitively intact, moderately active, had mediocre motor function, and rather low fear of falling. No study-associated serious adverse events were reported.

Table 1 Participant characteristics at baseline according to ITT analyses

N (mean ± SE)	**All** N=309	**LiFE** N=156	**gLiFE** N=153
Age, years	78.7 ± 0.3	78.8 ± 0.4	78.7 ± 0.4
Sex, n (%) female	227 (73.5)	115 (73.7)	112 (73.2)
BMI [kg/m^2]	27.2 ± 0.3	27.7 ± 0.4	26.8 ± 0.4
No. of medications	4.9 ± 0.2	5.0 ± 0.3	4.8 ± 0.3
No. of comorbidities	2.5 ± 0.1	2.5 ± 0.1	2.5 ± 0.1
EQ-5D index	0.84 ± 0.01	0.83 ± 0.01	0.85 ± 0.01
MoCA Score	26.0 ± 0.1	26.1 ± 0.2	25.9 ± 0.2
No. of steps/day	5,675 ± 173	5,795 ± 245	5,538 ± 2,828
No. of falls p.p. in past 12 months*	0.94 ± 1.1	1.01 ± 1.2	0.87 ± 1.1
Fallers in past 12 months, n (%)	157 (50.8)	79 (50.6)	78 (51.0)
LLFDI Function	57.3 ± 0.5	57.4 ± 0.6	57.3 ± 0.6
LLFDI Frequency	49.4 ± 0.2	49.3 ± 0.3	49.5 ± 0.4
LLFDI Disability	70.7 ± 0.7	71.7 ± 1.0	69.6 ± 0.9
Gait speed comfortable [m/s]	1.03 ± 0.0	1.03 ± 0.0	1.03 ± 0.0
Gait speed fast [m/s]	1.40 ± 0.0	1.37 ± 0.0	1.43 ± 0.0
30 sec Chair Stand	9.1 ± 0.2	9.2 ± 0.3	9.0 ± 0.3
8 Level Balance Scale	4.3 ± 0.8	4.2 ± 0.1	4.4 ± 0.1
Short FES-I	10.4 ± 0.2	10.4 ± 0.2	10.3 ± 0.2
ABC Scale	75.2 ± 1.0	75.0 ± 1.4	75.5 ± 1.4

Note. *: SD displayed; ABC Scale: Activities-specific Balance Confidence Scale; BMI: body mass index; CI: confidence interval; EQ-5D: health-related quality of life instrument; FES-I: Falls Efficacy Scale International; LLFDI: Late Life Function and Disability Instrument; max: maximal; MoCA: Montreal Cognitive Assessment; No.: Number; p.p.: per person; SD: standard deviation; SE: standard error; TUG: Timed Up-and-Go

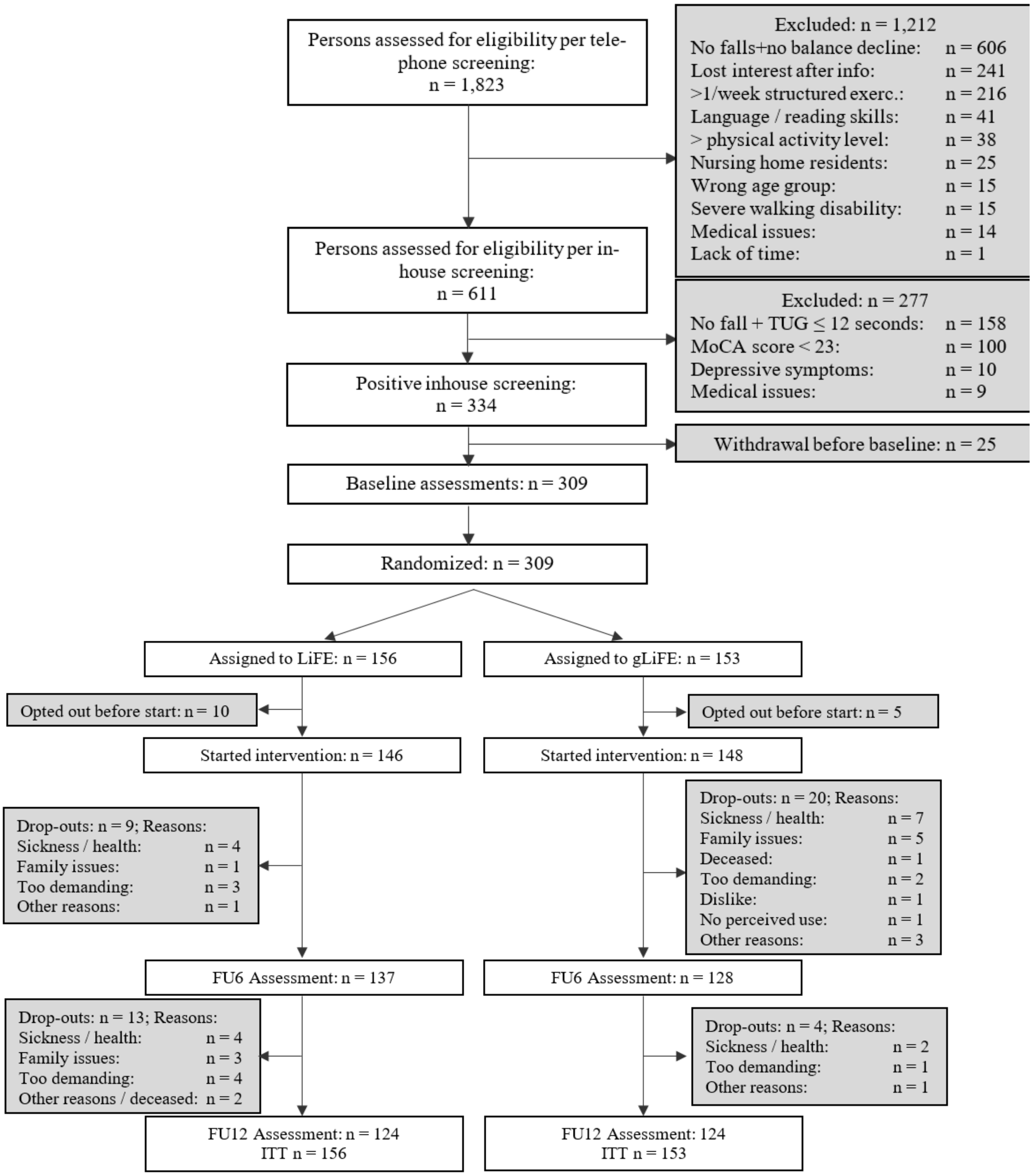


Figure 1 Participant flow. FU6: 6-month follow-up; FU12: 12-month follow-up; ITT: intention-to-treat; MoCA: Montreal Cog-nitive Assessment; TUG: Timed Up-and-Go;

Non-inferiority of PA-adjusted fall incidence

Intention-to-treat non-inferiority between gLiFE and LiFE was inconclusive. IRR indicates an insignificant, 4% lower PA-adjusted risk of falling for gLiFE participants, with the upper CI crossing the non-inferiority margin (IRR 0.96; 95% CI 0.67; 1.37) (Figure 2). The additional per protocol analysis (IRR 0.98; 95% CI 0.67; 1.32) confirmed ITT results.

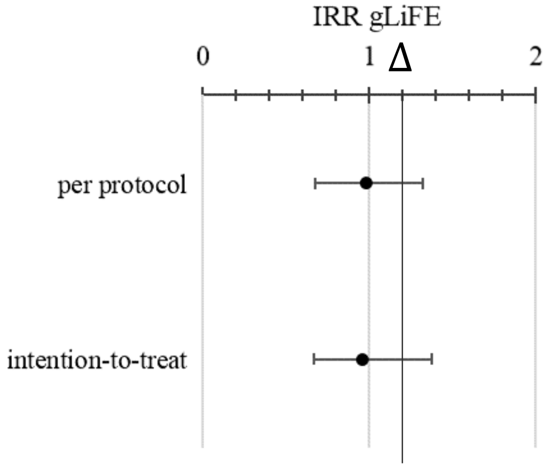


Figure 2 Observed treatment differences between LiFE (reference) and gLiFE in incidence rate ratio (IRR) of PA-adjusted risk of falling according to intention-to-treat and per protocol analysis at 12-month follow-up; non-inferiority Δ set to 20%

Costs, effects, and cost-effectiveness of gLiFE vs. LiFE

gLiFE had non-significantly higher unadjusted mean costs from a societal (+€1,094, SE=€1,184) and payer's perspective (+€513, SE=€798) (Table 2). Regarding the unadjusted effects, gLiFE resulted in non-significantly lower QALYs (-0.02, SE=0.02) and thus in negative ICERs (less effective, more costly), indicating that gLiFE was dominated by LiFE. Conversely, gLiFE participants had a non-significantly higher number of steps/day (+281, SE=449), and a lower number of falls (-0.03, SE=0.13). From a societal perspective, the ICERs were €3,895 per additional 1,000 steps/day and €39,420 per fall prevented (€1,828 and €18,503 from a payer's perspective). Adjusted cost and effect differences remained non-significant.

The CEACs (Figure 3A-C) indicated that cost-effectiveness of gLiFE versus LiFE was unlikely (QALY) or uncertain (falls and PA), independent of the WTP and cost perspective. In the per protocol analyses (Supplementary Table S1 & Figure S1), the probabilities of gLiFE being cost-effective compared to LiFE were slightly higher than in the ITT analyses, but remained below 95%.

Table 2 Costs and effects after 12 months

Category	gLiFE	LiFE	Difference, unadjusted	Difference, adjusted[b]
	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE
Costs[a]				
Outpatient services	882.1 ± 51.0	838.8 ± 55.8	43.3 ± 75.6	52.6 ± 65.6
Inpatient services	2,664.6 ± 583.1	2,491.3 ± 488.9	173.2 ± 675.8	226.3 ± 647.9
Informal & formal services	2,854.7 ± 614.0	2,067.5 ± 461.7	787.2 ± 736.3	611.5 ± 638.8
Formal services	366.4 ± 188.0	141.5 ± 57.1	224.9 ± 197.2	165.3 ± 158.8
Medication	1,107.6 ± 168.3	896.6 ± 72.6	211.0 ± 179.7	101.8 ± 91.9
Intervention (SP)	228.9	350.1	-121.2	
Intervention (PP)	211.0	350.1	-139.1	
Total costs (SP)	7,737.9 ± 994.1	6,644.4 ± 791.1	1,093.5 ± 1,183.8	1,099.1 ± 1,050.5
Total costs (PP)[c]	5,231.7 ± 688.9	4,718.4 ± 552.4	513.3 ± 798.2	591.4 ± 752.5
Effects				
QALY	0.820 ± 0.012	0.841 ± 0.010	-0.022 ± 0.015	-0.011 ± 0.011
Steps/day	6,644 ± 364	6,363 ± 290	281 ± 449	493 ± 413
Number of falls[d]	0.69 ± 0.09	0.72 ± 0.09	-0.03 ± 0.13	-0.03 ± 0.13

Note. SE: standard error, PP: payer's perspective, SP: societal perspective, QALY: quality-adjusted life years; Informal services include help from family/friends/neighbors; formal services include mobile home care services, payed domestic help, daycare, and inpatient care.

[a] Costs are based on 2018 Euros (€).

[b] Adjusted for age, sex, number of comorbidities (diabetes type 1 and 2, hypertension, acute cardiovascular disease, cardiac defect, cardiac arrhythmias, arthrosis, rheumatoid arthritis, cancer [not on active treatment], asthma or chronic obstructive pulmonary disease [Gold class < III], osteoporosis, depression, history of heart attacks or stroke), problems being physically active related to the COVID-19 pandemic, and the respective baseline costs, EQ-5D index, PA, or falls.

[c] excluded cost categories: informal care and travel costs for participants

[d] over 12 months observation period

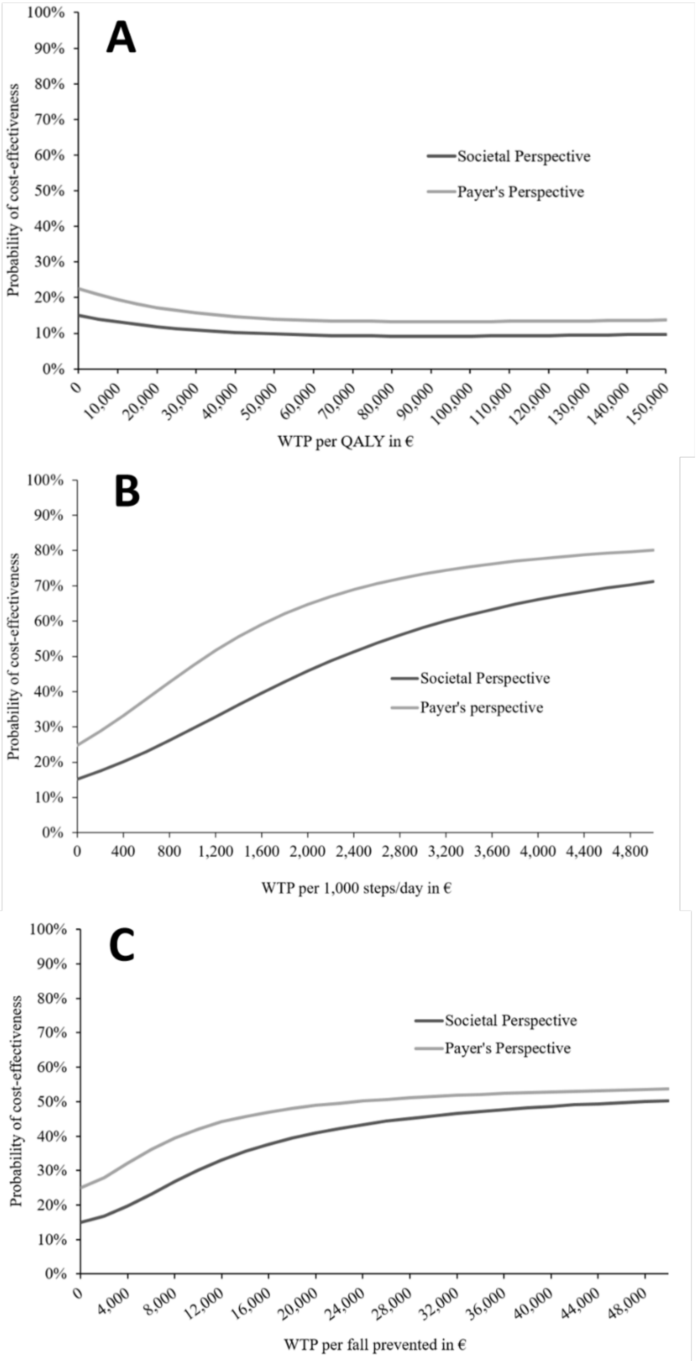


Figure 3 Cost-effectiveness acceptability curves from a societal and payer's cost perspective; probability of cost-effectiveness for gLiFE compared to LiFE for different willingness to pay (WTP) values per (A) quality-adjusted life years (QALY) gained, (B) 1,000 additional steps/day, or (C) fall prevented

Secondary analyses: physical activity

gLiFE participants increased their steps/day from 5,554 (SE=243) to 6,644 (SE=366), meaning a significant increase of 1,090 steps (95% CI 345; 1,835). In LiFE, the increase in steps/day from 5,795 (SE=245) to 6,363 (SE=293) was not significant (+569, 95% CI -31; 1,168), as was the difference between changes in both formats (+521, 95% CI -348; 1,391).

Secondary analyses: falls

According to ITT analyses, 291 falls in the past year were recorded at T0, 158 in LiFE and 133 in gLiFE; at T2, there were 218 falls (LiFE: 112; gLiFE: 106). At T0, 45 persons fell more than once in LiFE and 31 in gLiFE; at T2, these numbers were lower (LiFE: 26; gLiFE: 27). Overall, there were 79 fallers in LiFE and 78 in gLiFE at T0 and 64 fallers in both groups at T2. At T0, the incidence of falls per person year was 1.01 (95% CI 0.82; 1.01) in LiFE and 0.87 (95% CI 0.70; 1.04) in gLiFE; at T2, the incidence was reduced to 0.72 (95% CI 0.54; 0.89) in LiFE and 0.69 (95% CI 0.52; 0.85) in gLiFE. The median time to the first fall was identical in both groups (360 days). Comparing the number of falls per person year at T0 and T2, the incidence decreased about 21% (0.87 to 0.69) in gLiFE and 29% in LiFE (1.01 to 0.72).

Discussion

In the present study, we analyzed whether gLiFE was non-inferior to the original LiFE program in terms of PA-adjusted fall incidence and examined cost-effectiveness of gLiFE versus LiFE at 12-month follow-up. We found that non-inferiority by definition [17] remained inconclusive. One main reason for this finding is the rather wide confidence interval in the primary outcome; the upper confidence interval clearly crossed the non-inferiority margin. As both formats were designed in a similar way, the reason for inconclusiveness can be related to characteristics inherent in the different forms of delivery (group vs. one-to-one). For example, LiFE participants were visited by a trainer who accompanied implementation at home as opposed to gLiFE participants who had to implement LiFE activities at home on their own. This also led to a higher attendance rate on the side of the LiFE participants [15]. In line with this finding, it was shown that LiFE participants perceived the execution of the LiFE activities as slightly easier than gLiFE participants, despite both groups carrying out the same amount of LiFE activities [36, 37]. On the other hand, being without trainer support at home from the start may have led to more independence in gLiFE participants.

Looking at PA trajectories, the large enhancement of gLiFE participants' steps/day of >1,000 can be seen as a meaningful increase that was kept up in the long term, indicating the potential of gLiFE in terms of sustainable PA promotion. In LiFE, the rise of more than 500 steps per day was about half of that in gLiFE, but still meant an improvement of about 10%. Through these improvements, the magnitude of the mean steps/day exceeded 6,000 steps in both groups – a threshold that has just recently been found to be associated with lower risk of mortality in persons >60 years [38], which consolidates the value of the LiFE program in this regard.

Falls were reduced in both formats, which corresponds with findings of the Australian LiFE trial [12]. However, in the present sample, fall incidence was less than half compared to the Australian sample, indicating a much lower fall risk. Still, the intervention was able to reduce number of falls, highlighting its potential across various populations (Australia vs. Germany) and different fall risk magnitudes (higher vs. lower risk).

In clinical practice, our findings point out that the LiFE program is a safe means to increase PA in older persons at risk of falling, and that higher PA levels are sustained for at least half a year beyond the cessation of program delivery. Hence, especially persons at risk of further functional decline and experiencing detrimental effects of sedentary behavior can benefit from executing the LiFE program in the mid-term. Further research is required to establish whether this can be upheld in the long-term, that is, over a year and longer. In addition, PA trajectories of LiFE participants should be compared to controls in order to evaluate the "true" PA enhancement induced by LiFE. Aspects related to higher PA effects and adherence in gLiFE could be inherent within the group format. Participants' motivation could have been enhanced by social comparison, identification of role models, and peer-to-peer support, as was shown in a previous qualitative focus group study with participants from both formats [37].

From an economic point of view, it was expected that resource use would not differ between formats, but that gLiFE, overall, would be less costly due to the lower intervention costs and thus would also be cost-effective compared to LiFE. Intervention costs were indeed lower for gLiFE (€229 vs. €350), but were offset by higher resource utilization costs. The sample size and the short time horizon may inhibit drawing robust conclusions about the differences in resource utilization costs, which were at no point significant. Especially the increase in PA is probably not reflected in resource use in this study since healthcare costs attributable to PA are predominantly associated with treating long-term consequences such as ischemic heart disease, cancer, or type 2 diabetes [8]. Thus, the effects of increasing PA on the cost side may only be observed over a longer time horizon and should be examined in future studies. These long-term cost-savings from increasing PA by a certain level could also inform decision makers about what level of investment is justifiable for increasing PA by, e.g., 1,000 additional steps/day.

Limitations

IRR confidence intervals were wider than expected. This could be related to the present study sample being too fit or having had fewer falls, which may have been due to the rather inclusive approach in terms of performance-based estimation of fall risk by using a TUG cut-off of 12 seconds as opposed to higher values (e.g., 13 seconds) [39].. Another weakness was that baseline number of falls was collected retrospectively and thus with lower psychometric quality than the "prospective" calendar-based falls assessment at follow-up.

The T2 assessment was interrupted by the onset of the COVID-19 pandemic; N=70 participants completed their assessments parallel to COVID-19 regulations being in effect, and some of them with a delay of up to two months, which might have impacted study results. However, most of the observation period and the complete intervention itself were unaffected and no relevant influence of the pandemic on the main variables of interest in this study (PA, number falls, QALY, or costs) was observed; no bias between groups was detected.

Conclusions

This study extends current findings on ways of delivering effective interventions to prevent falls. Although non-inferiority of gLiFE compared to LiFE remained inconclusive, a reduction in absolute number of falls was observed and increases in walking activity were clinically relevant and sustainable beyond intervention end in both groups. Despite lower intervention costs of gLiFE, no program version was clearly superior in terms of cost-effectiveness.

Conflict of interest

None

Trial registration

The study was preregistered under clinicaltrials.gov (identifier: NCT03462654) on March 12th 2018; https://clinicaltrials.gov/ct2/show/NCT03462654

Funding

This work was supported by the German Federal Ministry of Education and Research [grant number 01GL1705A-D]. The content of this paper is the responsibility of the authors. The funders did not take any part in this work.

Acknowledgements

We thank Christoph Endress, Carolin Barz, Rebekka Leonhardt, Julia Gugenhan, Anna Kroog, Birgit Och, Malte Liebl-Wachsmuth, Mona Bär, Dr. Martin Bongartz, and Annette Lohmann for their invaluable support in carrying out this study.

Reference list

1. Daskalopoulou C, Stubbs B, Kralj C, Koukounari A, Prince M, Prina AM. Physical activity and healthy ageing: A systematic review and meta-analysis of longitudinal cohort studies. Ageing Res Rev. 2017 Sep;38:6-17.
2. Mertz KJ, Lee DC, Sui X, Powell KE, Blair SN. Falls among adults: the association of cardiorespiratory fitness and physical activity with walking-related falls. American journal of preventive medicine. 2010 Jul;39(1):15-24.
3. Lawton BA, Rose SB, Elley CR, Dowell AC, Fenton A, Moyes SA. Exercise on prescription for women aged 40-74 recruited through primary care: two year randomised controlled trial. BMJ (Clinical research ed). 2008 Dec 11;337:a2509.
4. Jefferis BJ, Merom D, Sartini C, Wannamethee SG, Ash S, Lennon LT, et al. Physical Activity and Falls in Older Men: The Critical Role of Mobility Limitations. Medicine and science in sports and exercise. 2015 Oct;47(10):2119-28.
5. Gregg EW, Pereira MA, Caspersen CJ. Physical activity, falls, and fractures among older adults: a review of the epidemiologic evidence. Journal of the American Geriatrics Society. 2000 Aug;48(8):883-93.
6. Heinrich S, Rapp K, Rissmann U, Becker C, König HH. Cost of falls in old age: a systematic review. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2010 Jun;21(6):891-902.
7. Hartholt KA, van Beeck EF, Polinder S, van der Velde N, van Lieshout EM, Panneman MJ, et al. Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life. J Trauma. 2011 Sep;71(3):748-53.

8. Dallmeyer S, Wicker P, Breuer C. How an aging society affects the economic costs of inactivity in Germany: empirical evidence and projections. Eur Rev Aging Phys Act. 2017;14:18.
9. Klenk J, Kerse N, Rapp K, Nikolaus T, Becker C, Rothenbacher D, et al. Physical Activity and Different Concepts of Fall Risk Estimation in Older People--Results of the ActiFE-Ulm Study. PloS one. 2015;10(6):e0129098.
10. Del Din S, Galna B, Lord S, Nieuwboer A, Bekkers EMJ, Pelosin E, et al. Falls Risk in Relation to Activity Exposure in High-Risk Older Adults. J Gerontol A Biol Sci Med Sci. 2020 May 22;75(6):1198-205.
11. Wijlhuizen GJ, Chorus AM, Hopman-Rock M. The FARE: a new way to express FAlls Risk among older persons including physical activity as a measure of exposure. Preventive medicine. 2010 Mar;50(3):143-7.
12. Clemson L, Fiatarone Singh MA, Bundy A, Cumming RG, Manollaras K, O'Loughlin P, et al. Integration of balance and strength training into daily life activity to reduce rate of falls in older people (the LiFE study): randomised parallel trial. BMJ (Clinical research ed). 2012 Aug 7;345:e4547.
13. Jansen CP, Nerz C, Kramer F, Labudek S, Klenk J, Dams J, et al. Comparison of a group-delivered and individually delivered lifestyle-integrated functional exercise (LiFE) program in older persons: a randomized noninferiority trial. BMC geriatrics. 2018 Nov 6;18(1):267.
14. Kramer F, Labudek S, Jansen CP, Nerz C, Fleig L, Clemson L, et al. Development of a conceptual framework for a group-based format of the Lifestyle-integrated Functional Exercise (gLiFE) programme and its initial feasibility testing. Pilot and feasibility studies. 2020;6:6.
15. Jansen CP, Nerz C, Labudek S, Gottschalk S, Kramer-Gmeiner F, Klenk J, et al. Lifestyle-integrated functional exercise to prevent falls and promote physical activity: Results from the LiFE-is-LiFE randomized non-inferiority trial. Int J Behav Nutr Phys Act. 2021 Sep 3;18(1):115.
16. Gottschalk S, König HH, Schwenk M, Nerz C, Becker C, Klenk J, et al. Cost-Effectiveness of a Group vs Individually Delivered Exercise Program in Community-Dwelling Persons Aged ≥70 Years. Journal of the American Medical Directors Association. 2022 May;23(5):736-42.e6.

17. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. Jama. 2012 Dec 26;308(24):2594-604.
18. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. Journal of the American Geriatrics Society. 1991 Feb;39(2):142-8.
19. Clemson L, Munro J, Fiatarone Singh MA. Lifestyle-integrated Functional Exercise (LiFE) program to prevent falls. Trainer's manual. Sydney: Sydney University Press; 2014.
20. Clemson L, Munro J, Fiatarone Singh MA. Lifestyle-integrated Functional Exercise (LiFE) program to prevent falls. Participant's manual. Sydney: Sydney University Press; 2014.
21. Ryan CG, Grant PM, Tigbe WW, Granat MH. The validity and reliability of a novel activity monitor as a measure of walking. British journal of sports medicine. 2006 Sep;40(9):779-84.
22. Klenk J, Peter RS, Rapp K, Dallmeier D, Rothenbacher D, Denkinger M, et al. Lazy Sundays: role of day of the week and reactivity on objectively measured physical activity in older people. Eur Rev Aging Phys Act. 2019;16:18.
23. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. The New England journal of medicine. 1988 Dec 29;319(26):1701-7.
24. Berg WP, Alessio HM, Mills EM, Tong C. Circumstances and consequences of falls in independent community-dwelling older adults. Age and ageing. 1997 Jul;26(4):261-8.
25. Lamb SE, Jorstad-Stein EC, Hauer K, Becker C. Development of a common outcome data set for fall injury prevention trials: the Prevention of Falls Network Europe consensus. Journal of the American Geriatrics Society. 2005 Sep;53(9):1618-22.
26. Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson LM, et al. Interventions for preventing falls in older people living in the community. The Cochrane database of systematic reviews. 2012 Sep 12(9):Cd007146.
27. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2011 Dec;20(10):1727-36.

28. Ludwig K, Graf von der Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. PharmacoEconomics. 2018 Jun;36(6):663-74.
29. Seidl H, Bowles D, Bock JO, Brettschneider C, Greiner W, König HH, et al. [FIMA--questionnaire for health-related resource use in an elderly population: development and pilot study]. Gesundheitswesen. 2015 2015/01//;77(1):46-52.
30. Bock JO, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). 2015 Jan;77(1):53-61.
31. "Inflation (CPI)" [database on the Internet]. OECD. 2018 [cited 10 Oct 2020]. Available from: https://doi.org/10.1787/eee82e6e-en
32. Rote Liste Service GmbH. ROTE LISTE 2018: Pharmaceutical directory for Germany (including EU approvals and certain medical devices). Frankfurt/Main: Rote Liste Service GmbH; 2018.
33. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society. 2005 Apr;53(4):695-9.
34. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011 Mar;20(1):40-9.
35. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ. 2002 Jul;11(5):415-30.
36. Nerz C, Kramer-Gmeiner F, Jansen CP, Labudek S, Klenk J, Becker C, et al. Group-Based and Individually Delivered LiFE: Content Evaluation and Predictors of Training Response - A Dose-Response Analysis. Clin Interv Aging. 2022;17:637-52.
37. Reicherzer L, Kramer-Gmeiner F, Labudek S, Jansen CP, Nerz C, Nystrand MJ, et al. Group or individual lifestyle-integrated functional exercise (LiFE)? A qualitative analysis of acceptability. BMC geriatrics. 2021 Feb 1;21(1):93.
38. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health. 2022 Mar;7(3):e219-e28.
39. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E, et al. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. J Geriatr Phys Ther. 2017 Jan/Mar;40(1):1-36.

Appendix

Table S 1 Costs[a] and effects after 12 months of community-dwelling older people at risk of falling participating in the LiFE-is-LiFE trial (per protocol sample, N=280)

Category	**gLiFE**	**LiFE**	**Difference, unadjusted**	**Difference, adjusted[b]**
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
Outpatient services	812.7 (46.0)	788.3 (48.6)	24.4 (65.1)	28.6 (58.7)
Inpatient services	1,933.5 (414.5)	2,241.0 (503.7)	-307.5 (623.1)	-266.0 (572.2)
Informal and formal services	2,446.4 (578.7)	1,757.0 (454.5)	689.4 (722.3)	574.7 (608.0)
formal services	328.1 (202.6)	107.3 (51.5)	220.8 (210.0)	161.8 (167.0)
Medication	988.5 (163.1)	878.8 (71.3)	109.7 (177.1)	79.4 (90.9)
Intervention – societal perspective	228.9	350.1	-121.2	- 121.17
Intervention – payer's persp.	211.0	350.1	-139.1	- 139.09
Total costs – societal persp.	6,410.1 (770.8)	6,015.3 (761.7)	394.9 (1,044.6)	458.0 (943.0)
Total costs – payer's persp.[c]	4,273.8 (520.3)	4,365.5 (545.0)	-91.7 (725.0)	-19.3 (660.2)
QALY	0.827 (0.012)	0.843 (0.010)	-0.016 (0.015)	-0.008 (0.011)
Steps/day	6,624 (332)	6,297 (286)	327 (453)	617 (428)
Number of falls[d]	0.71 (0.09)	0.72 (0.09)	-0.002 (0.13)	-0.01 (0.13)

Note: SE = standard error, persp. = perspective, QALY = quality-adjusted life years.

[a] Costs are based on 2018 Euros (€).

[b] Adjusted for age, sex, number of comorbidities, problems being physically active related to the COVID-19 pandemic, and the respective baseline costs or effects.

[c] excluded cost categories: informal care and travel costs for participants

[d] over 12 months observation period

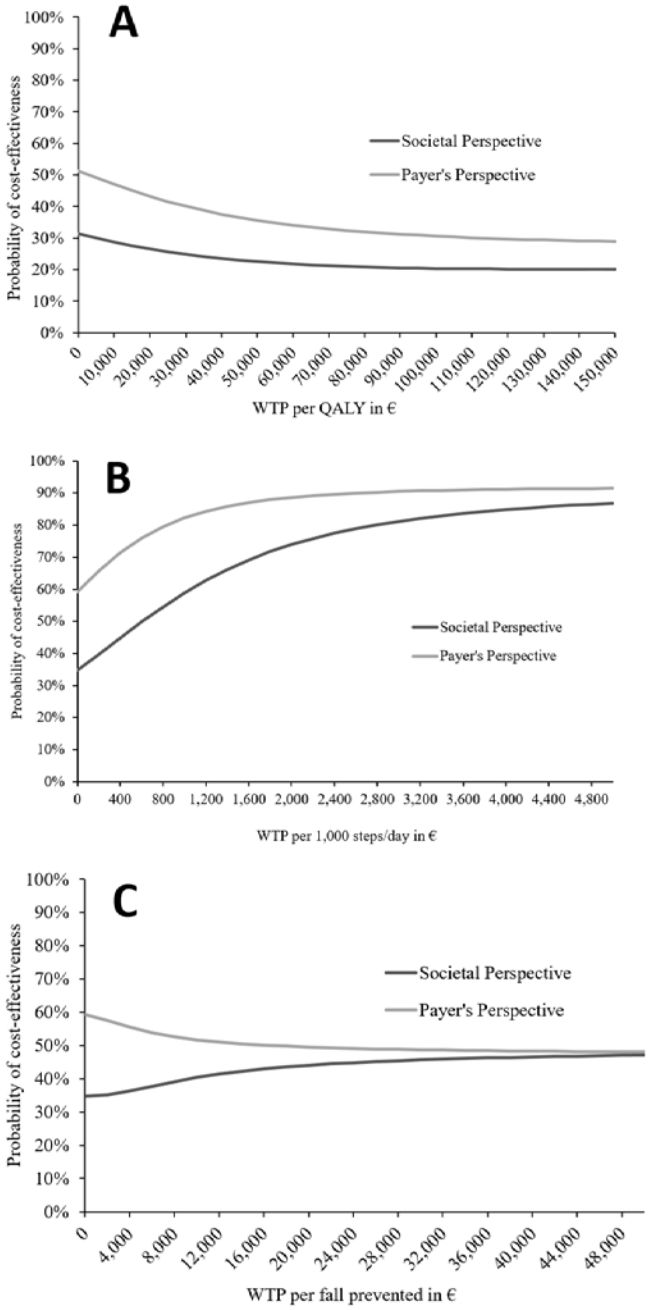


Figure S 2 Cost-effectiveness acceptability curves from a societal and payer's cost perspective; probability of cost-effectiveness for gLiFE compared to LiFE for different willingness to pay (WTP) values per (A) quality-adjusted life years (QALY) gained, (B) 1,000 additional steps/day, or (C) fall prevented; community-dwelling older people at risk of falling participating in the LiFE-is-LiFE trial (per protocol sample).

6 Publication 4 – Willingness to pay for a group and an individual version of the Lifestyle-integrated Functional Exercise Program from a participant perspective.

Published in *BMC Public Health* as:

Gottschalk S[1], König HH[1], Schwenk M[2,3], Nerz C[4], Becker C[4], Klenk J[4,5,6], Jansen CP[4,7], Dams J[1].
Willingness to pay for a group and an individual version of the Lifestyle-integrated Functional Exercise program from a participant perspective. BMC Public Health. 2022 Oct 18;22(1):1934.
DOI: 10.1186/s12889-022-14322-2.

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Network Aging Research, Heidelberg University, Heidelberg, Germany.

[3]Human Performance Research Centre, Department of Sport Science, University of Konstanz, Konstanz, Germany.

[4]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany.

[5]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.

[6]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany.

[7]Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany.

Gottschalk *et al. BMC Public Health* (2022) 22:1934
https://doi.org/10.1186/s12889-022-14322-2

BMC Public Health

RESEARCH

Open Access

Willingness to pay for a group and an individual version of the Lifestyle-integrated Functional Exercise program from a participant perspective

Sophie Gottschalk[1*], Hans-Helmut König[1], Michael Schwenk[2,6], Corinna Nerz[3], Clemens Becker[3], Jochen Klenk[3,4,5], Carl-Philipp Jansen[3,7] and Judith Dams[1]

Abstract

Background Perceived benefits of intervention programs from a participant perspective can be examined by assessing their willingness to pay (WTP). Aiming to support decision-makers in their decision to implement a fall prevention program, this study examined (1) the WTP for a group-based and an individually delivered fall prevention program, (2) which factors influence WTP, and (3) whether the WTP exceeds the intervention costs.

Methods WTP was elicited using Payment Cards from 237 individuals who participated in a randomized non-inferiority trial (LiFE-is-LiFE) comparing a group version of the Lifestyle-integrated Functional Exercise program (gLiFE) with the individually delivered version (LiFE). Linear regression models were used to examine factors associated with WTP. The net benefit for (g)LiFE was calculated as the difference between WTP and intervention costs, assuming different scenarios of intervention costs (varying group sizes of gLiFE) and hypothetical subsidy levels by a payer (€0, €50, or €75).

Results The mean WTP was €196 (95% CI [172, 221]) for gLiFE and €228 (95% CI [204, 251]) for LiFE. In the linear regression model, WTP was significantly associated with delivery format (−€32, 95% CI [−65, −0.2], for gLiFE) and net household income (+68€, 95% CI [23, 113], for ≥€3000 compared to <€2000). The net benefit for gLiFE was positive in most cases. Due to higher intervention costs of LiFE compared to gLiFE (€298 vs. €113), the net benefit for LiFE was negative for the majority of the sample, even at a subsidy of €75.

Conclusion The results provide insight into how valuable the interventions are perceived by the participants and thereby may be used by decision-makers as complement to cost-effectiveness analyses. WTP for both programs was generally high, probably indicating that participants perceived the intervention as quite valuable. However, further research is needed on the WTP and net benefit of fall prevention programs, as results relied on the specific context of the LiFE-is-LiFE trial.

*Correspondence:
Sophie Gottschalk
s.gottschalk@uke.de

Full list of author information is available at the end of the article

Keywords Willingness to pay, Patient preferences, Participant perspective, Contingent valuation method, Fall prevention, Physical activity promotion

Introduction

In the context of demographic change, the development of effective intervention programs to promote health into old age has become a priority in societies with an increasing population of older people. Besides clinical effectiveness, the (widespread) implementation of a program also depends on available resources, on whether the benefits of the program outweigh the costs associated with the implementation, and on who bears the costs of the program. Depending on the perspective, different benefits and costs are of relevance when deciding in favor of or against the implementation of a program. In economic analyses of healthcare programs, a societal or payer's perspective is frequently adopted, assuming that intervention costs are (at least partly) covered by the state or a health insurer [1]. However, health and well-being are also perceived as individual responsibility and hence can be seen as a good that people are willing to invest into. According to welfare economic theory, the benefits of a good or service are reflected in the form of willingness to pay (WTP) – the maximum amount of money an individual is willing to give up for the good or service. Thus, WTP is a concept that can be used to assign a monetary value to a good. Thereby, WTP goes beyond health and, unlike, for example, quality-adjusted life years (QALYs), which are frequently being used as effect measure in economic evaluations, does not restrict participants to express their preferences on pre-specified dimensions [2, 3]. Knowing the strength of a preference, aka how much people would be willing to pay, could also be relevant from a payer's perspective who may opt for a cost subsidy rather than full coverage of an intervention.

WTP for healthcare interventions is frequently captured using stated preference methods, which can be classified into direct (e.g., payment cards) and indirect (e.g., discrete choice experiments) methods [4]. In indirect methods, individuals are typically presented with several intervention options that differ in, e.g., the intervention characteristics, expected effects, and price. Participants are then asked to choose their preferred intervention option. In direct methods, on the other hand, individuals' WTP is determined by directly asking how much individuals would be willing to pay for an intervention. When collecting WTP data from participants of a clinical trial after completing the intervention, the stated WTP is assumed to reflect their individual perception of benefits, which may go beyond clinically visible effects such as improved physical performance or QALYs.

In Germany, the promotion of healthy ageing has been defined as national health goal [5]. Expanding measures to prevent falls is defined as sub goal since falls have a high prevalence with around one third of the population aged 65 years and older experiencing a fall at least once per year [6–8]. Falls can lead to injuries (e.g., hip fractures) which have serious consequences on health, quality of life and the healthcare budget [9]. Effective fall prevention programs could therefore be of high relevance for the promotion of healthy ageing and reducing the economic burden of falls. The LiFE-is-LiFE project compared a group-delivered version of the Lifestyle-integrated Functional Exercise program (gLiFE) with the original, individually delivered version (LiFE) [10]. Both programs consist of strength and balance activities that are integrated into everyday routines. In both programs, falls were reduced and physical activity was improved, while gLiFE was less costly in terms of intervention costs [11, 12]. Moreover, a content evaluation showed that both program versions were similar in terms of perceived safety, intensity of the exercises, integrability, and acceptance [13]. To our knowledge, no study has assessed how much participants are willing to pay for an exercise program aiming to maintain physical function and activity and reduce the risk of falling. When it comes to individual preferences and perceived benefits beyond clinical effectiveness, one could assume that the WTP of such a program differs by mode of administration. For example, gLiFE may be perceived as more valuable since it involves a social component (e.g., increased motivation through peer support) or, on the other hand, LiFE might be preferred as the individual training in the participant's home may be perceived as an advantage for implementing the LiFE activities into daily routines [13, 14].

Therefore, the aim of the current study was to explore WTP for gLiFE and LiFE, to examine factors influencing WTP, and to examine whether the perceived benefits – operationalized as WTP – exceed the costs associated with conducting the intervention(s).

Methods

Study design and sample

Data was taken from the LiFE-is-LiFE study (registered on 12/03/2018 under clinicaltrials.gov, identifier: NCT03462654), a multi-center, two armed, single-blinded, randomized non-inferiority trial, including community-dwelling, German-speaking people aged ≥ 70 years at risk of falling, who were able to ambulate 200 m without personal assistance [10]. Participants were randomized to either LiFE or gLiFE. Data was obtained at three time points (baseline, 6 months, and 12 months). WTP was assessed at 12 months.

Interventions and intervention costs

LiFE consisted of seven home visits (≈ 1 h) where a trainer presented activities for balance, strength, and general physical activity, adapting the performance and uptake of the activities to the needs of the participants. The trainer gave instructions on how to independently execute these activities and helped in implementing these activities in an individual participant's daily routine. In gLiFE, the program was taught by two trainers in seven sessions (≈ 2 h) to groups of 8 to 12 participants. The intervention sessions followed a detailed curriculum as trainers were not able to adapt flexibly to each individual's preferences. In both intervention arms, the participants received 2 additional 'booster phone calls' 4 and 10 weeks after the last intervention session. A detailed description of the interventions (including a TIDieR checklist) can be found in the study protocol [10]. The development of the conceptual gLiFE framework and a content analysis as well as a qualitative analysis of the acceptance of the two program versions were published separately [13–15].

Intervention costs for gLiFE and LiFE which incurred for the training sessions and booster phone calls were calculated as costs per participant based on personnel and material costs and travel expenses, assuming group sizes of 12 (scenario 1, base case), 10 (scenario 2), or 8 participants (scenario 3) in gLiFE. Assumptions underlying the calculation of different scenarios are presented in Table A1 (Additional file 1). For each scenario, the amount of costs from the participant perspective was derived by subtracting different hypothetical levels of subsidy (e.g., by a health insurer) of €0, €50, and €75.

Willingness to pay

Participants' WTP was elicited using Payment Cards, which are commonly used for assessing WTP for healthcare interventions [16]. Using response categories from €0, €5, €10, €20 to 'more than €100', participants receiving LiFE or gLiFE were asked about the amount of money they would surely be willing to pay as well as the amount they would definitely not be willing to pay for one training session of the respective program. The WTP for one training session was determined as the mean between these two values, which was then multiplied by the number of training sessions to obtain the total WTP for the intervention.

Explanatory variables

The following sample characteristics were considered in the analyses: intervention group (gLiFE/LiFE), age, sex, marital status, net household income, health insurance status (statutory vs. private), number of chronic conditions, healthcare costs, baseline fall status (non-faller vs. faller in the previous 6 months), motivation to exercise, satisfaction with the program, and training frequency (number of LiFE activities performed per week) at 12-month follow-up.

For the calculation of healthcare costs, costs from inpatient and outpatient service utilization, as well as medication and formal care use in the previous 6 months before the baseline assessment were considered. Resource utilization was monetarily valued in Euro (€) based on standardized unit costs [17] and inflated to the year 2018 [18].

Motivation to exercise was measured based on the autonomous motivation score of the Behavioral Regulation in Exercise Questionnaire (BREQ-3) [19], ranging from 0 to 4, with higher scores indicating higher motivation.

Satisfaction with the program was measured on a 5-point Likert scale (higher scores indicate higher satisfaction) and by a German school grade system using response categories from 1 (best grade) to 6 (worst grade).

Statistical analysis

The WTP was descriptively analyzed for persons with different sample characteristics for the total sample as well as for gLiFE and LiFE separately. Potential determinants of WTP were examined by linear regression models including the group variable (gLiFE/LiFE), sex, age, income, number of chronic conditions, healthcare costs, and motivation to exercise as independent variables. The mean net benefit from the participant perspective was calculated for different intervention scenarios (varying group sizes in gLiFE) and levels of subsidy by subtracting intervention costs from the WTP. The incremental net benefit of gLiFE over LiFE was determined by linear regression models adjusted for the potential determinants mentioned above.

Skewness of data was taken into account using a bootstrapped sample with n = 1,000 replicates. All analyses were conducted using STATA/SE 16.0 [StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC]. The significance level was set to 0.05.

Results

Sample characteristics

Sample characteristics are displayed in Table 1. Two hundred and thirty seven participants of the LiFE-is-LiFE trial completed the payment card at 12-month follow-up. Of those, the majority were female (74%), had an intermediate or high education (67%), were married/living in a partnership (45%) or widowed (37%), and insured by statutory health insurance (74%). The mean age was 79 years and 41% had fallen at least once in the previous 6

Table 1 Sample characteristics

		Total (n = 237)		gLiFE (n = 117)		LiFE (n = 120)	
Female	n (%)	176	(74.26)	87	(74.36)	89	(74.17)
Age	mean (SE)	78.66	(0.36)	78.52	(0.52)	78.80	(0.49)
Educational degree[1]	n (%)						
low		71	(29.96)	33	(28.21)	38	(31.67)
intermediate		67	(28.27)	31	(26.50)	36	(30.00)
high		93	(39.24)	48	(41.03)	45	(37.50)
other/no degree		6	(2.53)	5	(4.27)	1	(0.83)
Marital status	n (%)						
married/living in a partnership		107	(45.15)	57	(48.72)	50	(41.67)
widowed		87	(36.71)	38	(32.48)	49	(40.83)
divorced		28	(11.81)	17	(14.53)	11	(9.17)
permanently living separated		3	(1.27)	1	(0.85)	2	(1.67)
single		12	(5.06)	4	(3.42)	8	(6.67)
Net household income	n (%)						
€500 to <€750		5	(1.98)	2	(2.14)	2	(1.83)
€750 to <€1000		5	(2.24)	2	(1.79)	3	(2.67)
€1000 to <€1500		27	(11.27)	15	(12.65)	12	(9.92)
€1500 to <€2000		52	(22.03)	22	(18.89)	30	(25.08)
€2000 to <€3000		70	(29.45)	33	(28.55)	36	(30.33)
€3000 to <€5000		58	(24.51)	32	(27.01)	26	(22.08)
€5000+		20	(8.52)	11	(8.97)	10	(8.08)
Health insurance status	n (%)						
statutory		175	(73.84)	82	(70.09)	93	(77.50)
private		62	(26.16)	35	(29.91)	27	(22.50)
Number of chronic conditions	mean (SE)	2.44	(0.10)	2.43	(0.14)	2.45	(0.13)
Healthcare costs in €	mean (SE)	1585.60	(171.03)	1375.36	(179.37)	1790.57	(288.57)
Prevalence of fallers	n (%)	98	(41.35)	52	(44.44)	46	(38.33)
Number of falls among fallers	mean (SE)	1.65	(0.13)	1.62	(0.20)	1.70	(0.18)
Motivation to exercise (range 0–4)[2,3]	mean (SE)	2.96	(0.05)	3.00	(0.07)	2.92	(0.08)
Satisfaction with the program (max. = 5)[2]	mean (SE)	4.69	(0.06)	4.61	(0.08)	4.77	(0.08)
"School grade"[4]	mean (SE)	1.60	(0.04)	1.62	(0.06)	1.58	(0.05)
Training frequency[5]	mean (SE)	53.46	(1.50)	53.20	(2.26)	53.71	(2.00)

[1] low (9 years of school education), intermediate (10 years of school education), high (qualifies to enter university)

[2] higher scores indicate higher motivation/satisfaction

[3] BREQ-3 autonomous motivation score at FU12

[4] „school grade": 1 (A) ="sehr gut", 2 (B) = "gut", 3 (C) ="befriedigend", 4 (D) = "ausreichend", 5 (E) = „mangelhaft", 6 (F)= „ungenügend"

[5] training frequency = number of LiFE activities performed per week

months before the baseline assessment. On average, participants performed 53 LiFE activities per week and were overall satisfied with the program (mean satisfaction = 4.7 [maximum 5]; mean "school grade" = 1.6). Sample characteristics were similar for gLiFE and LiFE and between drop-outs and completers (results not shown).

Willingness to pay

The mean WTP stratified by groups of sample characteristics is displayed in Table 2. In the total sample, gLiFE participants had a lower mean WTP than LiFE participants (€196, 95% CI [172, 221] vs. €228, 95% CI [204, 251]) and participants with an income of €2000 to <€3000 or ≥€3000 had a higher mean WTP than those with an income of €500 to <€2000 (€218, 95% CI [187, 248] / €250, 95% CI [215, 284] vs. €175, 95% CI [152, 199]). Moreover, WTP was higher in males (€245, 95% CI [207, 283] vs. €201, 95% CI [183, 219], privately insured participants (€249, 95% CI [213, 285] vs. €200, 95% CI [181, 218]), those with higher healthcare costs (tertile 3: €235 95% CI [202, 268]; tertile 1: €185, 95% CI [156, 215]), and those with lower motivation to exercise (score ≤ 3: €238, 95% CI [210, 265]; score > 3: €190, 95% CI [169, 210]).

In the linear regression model identifying the determinants of WTP (Table 3), gLiFE was associated with a significantly lower WTP (–€32, 95% CI [−65, −0.2]) compared to LiFE. Among the other potential determinants in the model, only income was significantly associated with WTP, with the highest income group having a €68

Table 2 Older adults' willingness to pay (€) for the gLiFE/LiFE intervention by sample characteristics

	Total (n = 237)		gLiFE (n = 117)		LiFE (n = 120)	
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)
Group						
gLiFE	196	(172, 221)				
LiFE	228	(204, 251)				
Age						
< 80	219	(198, 241)	208	(175, 241)	230	(200, 259)
80+	202	(176, 228)	180	(143, 217)	225	(188, 262)
Sex						
Male	245	(207, 283)	226	(171, 280)	264	(211, 316)
female	201	(183, 219)	186	(160, 213)	216	(191, 240)
Marital status						
married/living in partnership	224	(197, 251)	211	(172, 250)	238	(197, 280)
widowed	197	(172, 222)	161	(131, 191)	225	(188, 262)
divorced	231	(181, 280)	248	(172, 325)	204	(171, 236)
permanently living separated	257	(165, 349)	175	(175, 175)	298	(218, 377)
single	169	(108, 231)	109	(43, 176)	199	(112, 286)
Net household income						
€500 to <**€**2000	175	(152, 199)	160	(124, 196)	189	(159, 219)
€2000 to <**€**3000	218	(187, 248)	203	(162, 245)	231	(189, 273)
€3000+	250	(215, 284)	227	(181, 273)	276	(226, 326)
Health insurance status						
statutory	200	(181, 218)	184	(156, 212)	213	(189, 238)
private	249	(213, 285)	226	(178, 273)	279	(224, 333)
Number of chronic conditions						
0–2 (ref.)	212	(190, 235)	194	(162, 226)	233	(202, 263)
3	195	(154, 235)	151	(95, 207)	217	(167, 266)
4–7	229	(195, 263)	228	(184, 271)	231	(180, 281)
Healthcare costs						
tertile 1 (≤€570)	185	(156, 215)	168	(120, 216)	203	(166, 239)
tertile 2 (>€570 to ≤€1,132)	217	(193, 241)	199	(173, 225)	234	(198, 271)
tertile 3 (>€1,132)	235	(202, 268)	222	(174, 270)	247	(203, 292)
Fall status						
non-faller	209	(189, 229)	188	(161, 215)	228	(199, 256)
Faller	217	(188, 246)	207	(163, 250)	229	(190, 267)
Motivation to excercise[1]						
lower medium (score ≤ 3)	238	(210, 265)	222	(183, 260)	252	(215, 288)
upper medium (score > 3)	190	(169, 210)	176	(145, 206)	205	(178, 232)
Satisfaction with the program						
(rather) unsatisfied	208	(145, 270)	159	(73, 246)	263	(180, 345)
rather satisfied	223	(188, 258)	212	(164, 260)	235	(188, 282)
(very) satisfied	207	(187, 227)	193	(163, 223)	220	(195, 246)
"School grade"[2]						
D/C (ref.)	201	(131, 271)	165	(111, 218)	273	(105, 441)
B	217	(192, 242)	209	(170, 247)	224	(192, 256)
A	209	(184, 233)	191	(154, 228)	227	(196, 258)
Training frequency[3]						
lower 3rd (0–42)	198	(170, 225)	197	(159, 235)	198	(162, 235)
middle 3rd (43–63)	217	(189, 245)	190	(142, 239)	237	(203, 271)
upper 3rd (64–112)	221	(189, 254)	201	(157, 244)	245	(198, 291)

[1]BREQ-3 autonomous motivation score at FU12

[2]„school grade": 1 (A) ="sehr gut", 2 (B) = "gut", 3 (C) ="befriedigend", 4 (D) = "ausreichend", 5 (E) = „mangelhaft" (not reported), 6 (F)= „ungenügend" (not reported)

[3]number of LiFE activities performed per week

Table 3 Determinants of older adults' willingness to pay for the (g)LiFE intervention

	Beta	SE	95% CI	p-value
gLiFE (ref. LiFE)	−32	16	(−64.61, −0.16)	0.049
Female (ref. male)	−19	23	(−63.33, 25.73)	0.408
Age	3	2	(−0.16, 5.98)	0.063
Net household income (ref. <€2000)				
€2000-€3000	32	20	(−8.02, 72.23)	0.117
€3000+	68	23	(23.21, 112.96)	0.003
Number of chronic conditions	2	5	(−8.78, 12.69)	0.721
Healthcare costs	0	0	(−0.01, 0.01)	0.727
Motivation to exercise[2]	−22	11	(−44.31, 0.63)	0.057
Intercept	40	121	(−196.46, 275.91)	0.742
Adjusted R-Squared	*0.071*			

[2] BREQ-3 autonomous motivation score at FU12

(95% CI [23, 113]) higher WTP compared to the lowest income group.

Net benefit

The intervention costs per participant for LiFE were €298. For gLiFE, intervention costs varied depending on the group size with €113, €123, and €138 for 12, 10, and 8 participants, respectively (Table A1, Additional file 1). In the base case scenario, gLiFE had a significant positive mean net benefit between €83 (95% CI, [59, 107]) at €0 subsidy and €158 (95% CI, [134, 182]) at €75 subsidy (Table 4). When lower group sizes were assumed (Scenario 2 and 3), the mean net benefit was somewhat lower, but remained positive for each subsidy level. For LiFE, the intervention costs exceeded the WTP, resulting into negative mean net benefits, except for the case that €75 were subsidized (€5, 95% CI [−19, 28]).

When the distributions of the net benefit for gLiFE and LiFE were graphically examined for scenario 1 by different subsidy levels (Fig. 1), it could be observed that the majority of gLiFE participants (68% [€0 subsidy], 86% [€50 subsidy], and 95% [€75 subsidy]) had a positive net benefit, whereas this applied to only 25%

Table 4 Mean net benefit by intervention groups, scenarios of intervention costs, and subsidy schemes

		gLiFE		LiFE		Difference	
	Subsidy	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)
Scenario 1	€0	83	(59, 107)	−70	(−94, −47)	153	(119, 188)
	€50	133	(109, 157)	−20	(−44, 3)		
	€75	158	(134, 182)	5	(−19, 28)		
Scenario 2	€0	73	(49, 98)	-70	(−94, −47)	143	(109, 178)
	€50	123	(99, 148)	-20	(−44, 3)		
	€75	148	(124, 173)	5	(−19, 28)		
Scenario 3	€0	58	(34, 83)	-70	(−94, −47)	129	(94, 163)
	€50	108	(84, 133)	-20	(−44, 3)		
	€75	133	(109, 158)	5	(−19, 28)		

Notes: Scenarios 1–3 differ by group size for gLiFE which influenced the intervention costs: scenario 1 (base case, 12 participants, €113), scenario 2 (10 participants, €123), scenario 3 (8 participants, €138). Intervention costs for LiFE were €298.

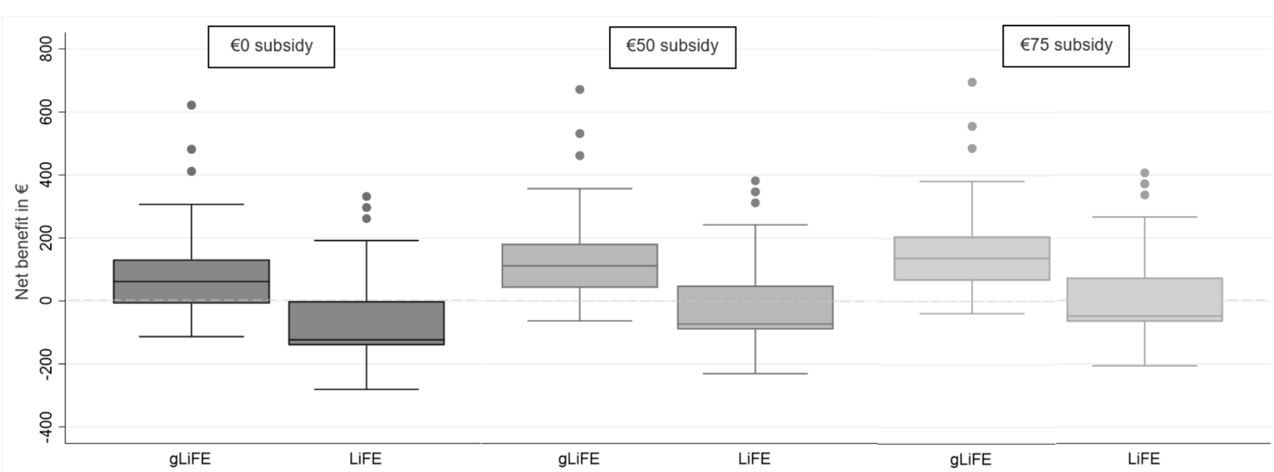


Fig. 1 Distribution of the mean net benefit for gLiFE/LiFE by different subsidy levels of intervention costs. Intervention costs based on Scenario 1 (gLiFE: €113; LiFE: €298)

(€0 subsidy), 29% (€50 subsidy), and 40% (€75 subsidy) of LiFE participants

The unadjusted incremental net benefit for gLiFE compared to LiFE was €153 (95% CI [119, 188],Table 4). Adjusting the incremental net benefit did not change the estimate relevantly (€154, 95% CI [122, 186]; not displayed in table)

Discussion

This study is the first that assessed the WTP for a group-based and an individually delivered version of a fall prevention and activity promotion program (LiFE) in a sample of community-dwelling German older adults at risk of falling. WTP for both programs was generally high, probably indicating that participants perceived the intervention as quite valuable and thus possibly reflecting a demand for such interventions. WTP was determined by delivery format and income, with LiFE participants on average reporting €32 higher WTP than gLiFE participants and higher income groups reporting higher WTP. For gLiFE, benefits in terms of WTP exceeded intervention costs in most cases, while LiFE had considerably higher intervention costs than gLiFE (€298 vs. €113), and thus the WTP was lower than the intervention costs in the majority of the sample (60–75%, depending on hypothetical subsidy level). Hence, the difference in WTP between gLiFE and LiFE did not compensate for the higher intervention costs (+€185), even when subsidized by up to €75.

Asking participants of an intervention study who have actual experience with the intervention of interest about their willingness to pay (rather than reporting their WTP for hypothetical intervention scenarios) has not been done frequently, especially in the field of physical activity interventions for community-dwelling older people. However, this approach might be an attractive complement to the evaluation of (cost-)effectiveness of competing interventions – WTP constitutes a measure of the perceived benefits or value of the intervention from a participant perspective that is not restricted to predefined dimensions on which benefits can be expressed (e.g., in patient-reported outcome measures). As WTP may be based on factors other than effectiveness alone [20], knowing the preferences (WTP) of the target population may be particularly useful when the effectiveness of different program formats based on conventional measures (e.g. reduction of falls [11, 12]) is indifferent or similar. Furthermore, WTP extends conventional (cost-)effectiveness frameworks, as it may reveal additional benefits perceived by the participants that may otherwise have been overlooked or not captured. For example, in RCTs that evaluated exercise interventions for older people, only marginal (and probably not clinically important) differences in QALYs between the intervention and control group are found, at least over time horizons of six months to two years [21]. Overall, determining the WTP may aid decision-makers in deciding which intervention should be preferred for implementation [2].

Beyond the level of willingness to pay, it is also interesting to know which factors influence willingness to pay, as this information may then be used to adapt interventions according to the preferences of the target group. In the current study, only income (besides program version) determined WTP, and overall only 7% of the variance was explained by the potential determinants in the multivariate regression model, indicating that other (unobserved) factors determine the WTP to a large extent. Other factors that can be hypothesized to determine WTP could be the individually perceived relevance (e.g. individually perceived risk of falling) and perceived effectiveness of the intervention, the presence of other health conditions whose treatment may be given a higher priority, or the relationship with the trainer [14]. It is also not clear whether participants factored the cost of providing the intervention, and thus the additional effort required for home visits in LiFE, into their willingness to pay, which may explain the difference in WTP between program versions [20].

That WTP is associated with income is not surprising as it is inherently limited by wealth [22]. This carries a danger of self-selection of only higher-income populations into participating in the program which poses a threat to the idea of equal health opportunities, for example, making prevention accessible to everyone independent of socio-economic position [23]. In Germany, prevention programs can be certified, which qualifies them for subsidies of the intervention costs by the health insurances. These subsidies lower the intervention costs and thereby make interventions more accessible to people that are economically less well of, while at the same time alleviating the burden on health insurers' budgets. Assuming a subsidy of €75, the WTP of almost all gLiFE participants (95%) covered (or even exceeded) the intervention costs, providing a strong argument for the implementation of gLiFE over LiFE. However, it does not seem reasonable to give recommendations for not offering and/or subsidizing LiFE – there may be still demand for LiFE as between 25% and 40% of the LiFE participants were willing to pay enough to cover the intervention costs. Those people, based on individual preferences, may still opt for the individual program despite being more costly. For example, some people may prefer individual supervision and learning the program in their own home where the activities could be adapted to the individual conditions and are therefore more easy to integrate into everyday life, whereas for others the social aspects of a group program (e.g., motivation through peer support) may be more important [13, 14]. Moreover, the individual approach in LiFE could be more suitable for people for whom participation in group programs would be difficult, for example, because of physical and transport limitations.

Despite the reduction of access barriers through the subsidies, the uptake of the intervention still depends to a certain extent on the financial resources of the individual. This selection effect could be reduced if a payer fully reimburses the intervention costs, which may still be economically attractive for a payer if the effects of an intervention are expected to spill over to cost savings (e.g., lower health-related resource use). Therefore, assessing participants' preferences in terms of their WTP does not substitute cost-effectiveness analyses, but may be used as a complement, especially when there are two competing interventions where neither is clearly superior to the other in terms of cost-effectiveness, as was the case with gLiFE and LiFE [12, 24].

Limitations

Limitations arise from the fact that the elicited WTP is tied to the specific survey context and therefore does probably not reflect 'real world' behavior [22]. The results are based on a selective population of participants who can be assumed to have a special interest in fall prevention and may not be representative of the target population in terms of their socio-economic status, which influences WTP. Given that all participants valued their WTP based on their personal experience of the program, the WTP may not represent the WTP of individuals who do not have actual experiences with the program to draw on, and comparing the strength of the preference between gLiFE and LiFE based on this study may be limited since the participants were only asked about their WTP for the version they received (either gLiFE or LiFE). Moreover, the time point of assessing the WTP (at 12-month follow-up) could have influenced the valuation. Test-retest reliability of the WTP was not examined in this study and no information on the WTP of study drop-outs (n=64) was available. Future studies may use a discrete choice experiment to find out which attributes determine the level of WTP or the decision for one or the other program version.

Conclusion

This study explored WTP for a group-delivered and an individually delivered activity promotion and fall prevention program. The results are useful for decision-makers or potential payers as they provide insight into how valuable the interventions are perceived by the participants and thereby complement the cost-effectiveness studies of the LiFE program with a participant perspective. The high WTP for both programs suggests that they are perceived as valuable by the participants, thereby supporting previous results on the acceptability and content of the (g)LiFE program. WTP was associated with income and was higher for LiFE than for gLiFE, but this difference did not compensate the higher intervention costs in LiFE. gLiFE was likely to yield a positive net benefit, especially when the intervention is subsidized by a potential payer, while for the individual program the WTP was less likely to exceed the costs. Payers may still consider subsidizing both versions, as there may be a number of individuals who prefer the individual program despite its higher cost.

List of abbreviations

BREQ-3 Behavioral Regulation in Exercise Questionnaire.
gLiFE Group-based Lifestyle-integrated Functional Exercise.
LiFE Lifestyle-integrated Functional Exercise.
QALY Quality-adjusted life year.
WTP Willingness to pay.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-022-14322-2.

Supplementary Material 1

Acknowledgements

We thank all participants for their participation. We also thank all colleagues involved in the conduct of this trial, including the trainers and assessors, database managers, and members of the advisory board.

Authors' contributions

MS and CB developed the grant proposal for the LiFE-is-LiFE trial approved for funding. SG, JD, and HHK developed the methodological approach of this study. SG performed the data analysis and produced the first draft of the manuscript. JD, HHK, MS, CPJ, CN, CB, and JK critically revised the manuscript for important intellectual content. The final version of the manuscript was reviewed, edited, and approved by all authors.

Funding

This work was supported by the German Federal Ministry of Education and Research (grant: 01GL1705A-D). The funding source takes no part in the collection, analysis and interpretation of data, in the writing of the manuscript or in the decision to submit the manuscript for publication.
Open Access funding enabled and organized by Projekt DEAL.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to ethical and confidentiality concerns but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Ethical approval for the project was obtained from the Ethic Review Board of the Faculty of Behavioral and Cultural Studies at Heidelberg University (for the study site Heidelberg; document number Schwe2017 2/1–1), and from the Ethic Review Board of the University Hospital and Faculty of Medicine in Tübingen (for the study site Stuttgart; document number 723/2017BO2). All participants gave written informed consent prior to participation The study is conforming to the respective policy and mandates of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Health Economics and Health Services Research, Hamburg Center for Health Economics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2]Network Aging Research, Heidelberg University, Heidelberg, Germany
[3]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany
[4]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany
[5]IB University of Applied Health and Social Sciences, Study Centre Stuttgart, Stuttgart, Germany
[6]Human Performance Research Centre, Department of Sport Science, University of Konstanz, Konstanz, Germany
[7]Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany

Received: 14 June 2022 / Accepted: 5 October 2022
Published online: 18 October 2022

References

1. Gray AM, Clarke P, Wolstenholme J, Wordsworth S: Defining, measuring, and valuing costs. In: *Applied methods of cost-effectiveness analysis in healthcare*. New York: Oxford University Press; 2011: 119–141.
2. Olsen JA, Smith RD. Theory versus practice: a review of 'willingness-to-pay' in health and health care. Health Econ. 2001;10:39–52.
3. Gerard K, Mooney G. QALY league tables: handle with care. Health Econ. 1993;2:59–64.
4. Breidert C, Hahsler M, Reutterer T. A Review of Methods for Measuring Willingness-to-Pay., 2(4). Innovative Marketing. 2006;2.
5. Bundesministerium für Gesundheit. Nationales Gesundheitsziel - Gesund älter werden. Berlin: Bundesministerium für Gesundheit; 2012.
6. Bergen G, Stevens MR, Burns ER. Falls and Fall Injuries Among Adults Aged >/=65 Years - United States, 2014. MMWR Morb Mortal Wkly Rep. 2016;65:993–8.
7. Olsson Moller U, Midlov P, Kristensson J, Ekdahl C, Berglund J, Jakobsson U. Prevalence and predictors of falls and dizziness in people younger and older than 80 years of age–a longitudinal cohort study. Arch Gerontol Geriatr. 2013;56:160–8.
8. Schumacher J, Pientka L, Trampisch U, Moschny A, Hinrichs T, Thiem U. The prevalence of falls in adults aged 40 years or older in an urban, German population. Results from a telephone survey. Z Gerontol Geriatr. 2014;47:141–6.
9. Hartholt KA, van Beeck EF, Polinder S, van der Velde N, van Lieshout EMM, Panneman MJM, van der Cammen TJM, Patka P. Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life. J Trauma. 2011;71:748–53.
10. Jansen CP, Nerz C, Kramer F, Labudek S, Klenk J, Dams J, Konig HH, Clemson L, Becker C, Schwenk M. Comparison of a group-delivered and individually delivered lifestyle-integrated functional exercise (LiFE) program in older persons: a randomized noninferiority trial. BMC Geriatr. 2018;18:267.
11. Jansen CP, Nerz C, Labudek S, Gottschalk S, Kramer-Gmeiner F, Klenk J, Dams J, König HH, Clemson L, Becker C, et al. Lifestyle-integrated functional exercise to prevent falls and promote physical activity: Results from the LiFE-is-LiFE randomized non-inferiority trial. Int J Behav Nutr Phys Act. 2021;18:115.
12. Jansen CP, Gottschalk S, Nerz C, Labudek S, Kramer-Gmeiner F, Klenk J, Clemson L, Todd C, Dams J, König HH, Becker C, et al. Comparison of falls and cost-effectiveness of the group vs. individually delivered Lifestyle-integrated Functional Exercise (LiFE) program: final results from the LiFE-is-LiFE non-inferiority trial. Age Ageing. [under review].
13. Nerz C, Kramer-Gmeiner F, Jansen CP, Labudek S, Klenk J, Becker C, Schwenk M. Group-Based and Individually Delivered LiFE: Content Evaluation and Predictors of Training Response - A Dose-Response Analysis. Clin Interv Aging. 2022;17:637–52.
14. Reicherzer L, Kramer-Gmeiner F, Labudek S, Jansen C-P, Nerz C, Nystrand MJ, Becker C, Clemson L, Schwenk M. Group or individual lifestyle-integrated functional exercise (LiFE)? A qualitative analysis of acceptability. BMC Geriatr. 2021;21:93.
15. Kramer F, Labudek S, Jansen C-P, Nerz C, Fleig L, Clemson L, Becker C, Schwenk M. Development of a conceptual framework for a group-based format of the Lifestyle-integrated Functional Exercise (gLiFE) programme and its initial feasibility testing. Pilot and Feasibility Studies. 2020;6:6.
16. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess. 2001;5:1–186.
17. Bock JO, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, König HH. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen. 2015;77:53–61.
18. Inflation (CPI) (indicator) OECD. 2020. Accessed 14 January 2020.
19. Markland D, Tobin V. A Modification to the Behavioural Regulation in Exercise Questionnaire to Include an Assessment of Amotivation. J Sport Exerc Psychol. 2004;26:191–6.
20. Haines TP, McPhail S. Patient preference for falls prevention in hospitals revealed through willingness-to-pay, contingent valuation survey. J Eval Clin Pract. 2011;17:304–10.
21. Subías-Perié J, Navarrete-Villanueva D, Gómez-Cabello A, Vicente-Rodríguez G, Casajús JA. Health economic evaluation of exercise interventions in people over 60 years old: A systematic review. Exp Gerontol. 2022;161:111713.
22. Carson RT, Flores NE, Meade NF. Contingent Valuation: Controversies and Evidence. Environ Resource Econ. 2001;19:173–210.
23. gesundheitsziele.de, editor: Leitfragen zur Stärkung der Querschnittsanforderung ‚Gesundheitliche Chancengleichheit' erarbeitet von der UAG des Evaluationsbeirates zur Stärkung der Querschnittsanforderung ‚Gesundheitliche Chancengleichheit'. Köln: GVG; 2014.
24. Gottschalk S, König HH, Schwenk M, Nerz C, Becker C, Klenk J, Jansen CP, Dams J. Cost-Effectiveness of a Group vs Individually Delivered Exercise Program in Community-Dwelling Persons Aged ≥ 70 Years. J Am Med Dir Assoc. 2022;23:736-742.e6.

Publisher's Note

Additional file 1

Table A1 Data and assumptions for the calculation of intervention costs by scenario

LiFE sessions			
Average duration[1]		120 min	
gLiFE sessions			
Average duration[1]		150 min	
(Average) number of participants per group[2]		12 (scenario 1) 10 (scenario 2) 8 (scenario 3)	
Average travel expenses per participant		€ 17.92	
Other assumptions			
Salary level of the trainers		TVöD[3] E8	
Number of sessions per participant[2]		7	
Number and average duration of "booster phone calls"		2 x 30 min	
Costs material set for each participant		€ 28.65	
Total intervention costs	Scenario 1	Scenario 2	Scenario 3
gLiFE	€ 113.32	€ 123.26	€ 138.18
LiFE	€ 298.30	€ 298.30	€ 298.30

[1] including preparation and travelling time

[2] assumption: sessions have to be paid regardless of participation

[3] TVöD: civil service collective agreement; 2018 wages

7 Publication 5 – Psychometric Properties of the EQ-5D for the Assessment of Health-Related Quality of Life in the Population of Middle-Old and Oldest-Old Persons: Study Protocol for a Systematic Review

Published in *Frontiers in Public Health* as:

Gottschalk S[1], König HH[1], Nejad M[1], Dams J[1]. Psychometric Properties of the EQ-5D for the Assessment of Health-Related Quality of Life in the Population of Middle-Old and Oldest-Old Persons: Study Protocol for a Systematic Review. Front Public Health. 2020;8:578073. DOI: 10.3389/fpubh.2020.578073.

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

frontiers in Public Health

STUDY PROTOCOL
published: 30 October 2020
doi: 10.3389/fpubh.2020.578073

Psychometric Properties of the EQ-5D for the Assessment of Health-Related Quality of Life in the Population of Middle-Old and Oldest-Old Persons: Study Protocol for a Systematic Review

Sophie Gottschalk, Hans-Helmut König, Mona Nejad and Judith Dams*

Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

OPEN ACCESS

Edited by:
Roza Adany,
University of Debrecen, Hungary

Reviewed by:
Tissa Wijeratne,
The University of Melbourne, Australia
Simon Grima,
University of Malta, Malta

Correspondence:
Sophie Gottschalk
s.gottschalk@uke.de

Specialty section:
This article was submitted to Health Economics, a section of the journal Frontiers in Public Health

Received: *30 June 2020*
Accepted: *28 September 2020*
Published: *30 October 2020*

Citation:
Gottschalk S, König H-H, Nejad M and Dams J (2020) Psychometric Properties of the EQ-5D for the Assessment of Health-Related Quality of Life in the Population of Middle-Old and Oldest-Old Persons: Study Protocol for a Systematic Review. Front. Public Health 8:578073. doi: 10.3389/fpubh.2020.578073

Introduction: Health care interventions for middle-old and oldest-old individuals (75 years or older) are often economically evaluated using the EuroQol questionnaire (EQ-5D) to measure health-related quality of life. However, the psychometric performance of the EQ-5D in this population has been questioned, as it probably does not adequately capture relevant aspects of quality of life in the older population. Because the results of economic evaluations using the EQ-5D often guide decision-makers, it is important to know whether the EQ-5D has satisfactory psychometric properties in the middle-old and oldest-old population. Therefore, studies assessing the psychometric properties of the EQ-5D in this population should be synthesized by a systematic review.

Methods and Analysis: A systematic review of studies providing empirical evidence of reliability, validity, and/or responsiveness of the EQ-5D in a sample with a mean age ≥75 years will be conducted. The databases PubMed, Web of Science, and EconLit will be searched. In addition, reference lists of included studies will be hand-searched. Two independent reviewers will select studies and assess their risk of bias with the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) Risk of Bias checklist. Relevant data will be extracted by one reviewer and cross-checked by a second reviewer. Potential disagreements in any phase will be resolved through discussion with a third person. The guidelines for systematic reviews of measurement properties proposed by the COSMIN group, including criteria of good measurement properties, will guide the synthesis and interpretation of the results.

Discussion: The review's results could facilitate the making of recommendations for the use of the EQ-5D in a population of middle-old and oldest-old people and thereby being of interest for decision-makers or for researchers designing new intervention studies for older people. Heterogeneity of individual studies regarding the population under study could limit the possibility of making a synthesized statement on the appropriateness of the EQ-5D for the middle-old to oldest-old population.

Keywords: systematic review, psychometric properties, oldest-old population, older population, EQ-5D

INTRODUCTION

In the context of demographic change, the older population, especially the population of middle-old and oldest-old (75 years or older), is increasing (1). As this population has typically a high number of (co)morbidities, a range of health care interventions aiming to improve the health and quality of life (QoL) of older persons has been developed. However, given the scarcity of resources, economic evaluations of new interventions are crucial for decision-making regarding their implementation as they provide information on the efficient allocation of resources. In economic evaluations, effectiveness of interventions is often measured by health-related QoL (HrQoL). In order to make effects comparable across interventions, generic instruments of HrQoL are used.

The most frequently used generic instrument of HrQoL in economic evaluations is the EuroQol 5 dimensions questionnaire (EQ-5D) (2). The advantage of the EQ-5D is its brevity and easy administration by consisting of only five questions covering the dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Depending on the version of the EQ-5D, each dimension has three (EQ-5D-3L) or five (EQ-5D-5L) severity levels. Despite its brevity, it is important that the EQ-5D is psychometrically sound in the population it is used, meaning that it measures what it intended to measure (validity) in an accurate and reproducible way (reliability) and is able to detect small but important changes over time (responsiveness). In the absence of sufficient psychometric properties, the results of economic evaluations of interventions fail in measuring the true effect of interventions on HrQoL and thus are not suitable as basis for decision-making regarding their implementation.

The approach to primarily focus on HrQoL in the form of health utility gains in economic evaluations has been criticized as it excludes aspects of QoL beyond health. As people's needs and desires change with age, significant intervention effects beyond the health status may not be sufficiently captured for the middle-old and oldest-old. Therefore, other instruments than the EQ-5D with a different theoretical approach have been developed. One example are the ICECAP instruments (3), which were developed based on the capability theory (4). Contrary to HrQoL, capability focuses on the *ability* of a person to function and not on *functioning*. With the ICECAP-O, an instrument has been developed especially for the assessment of QoL in older people (5). The development was based on in-depth interviews with the aim of identifying *attributes* of QoL instead of only *influences* on QoL. In this context, health was seen as an influence on attributes rather than as an attribute on its own (6). Especially at the end of life, it was shown that it is not appropriate to apply an exclusively health-focused perspective in economic evaluations, because aspects that go beyond health (e.g., choice/having a say, being with people who care, dignity, and preparation) become more important (7).

Nevertheless, the EQ-5D is still the most widely used instrument for economic evaluations, and as it aims to measure HrQoL, its validity cannot be judged by the fact that it does not capture factors beyond health. Previous reviews have been conducted regarding the psychometric performance of the EQ-5D focusing on different population groups. The EQ-5D was found appropriate for depression and personality disorders (8, 9), urinary incontinence (10), some skin diseases (11), and in people 60 years or older (12). However, the EQ-5D lacked psychometric performance in populations with anxiety, schizophrenia, and bipolar disorders, as well as in those with multiple sclerosis (8, 9, 13). Moreover, Tordrup et al. (14) evaluated the responsiveness of the EQ-5D in various disorders and concluded that the instrument is not sensitive to change in a range of disorders. Regarding the use of the EQ-5D in dementia, the validity was found problematic as there are significant disagreements between patient and proxy ratings and as the EQ-5D does not capture aspects that are particularly important for people with dementia (15, 16). Similarly, other authors conclude that the EQ-5D may not be appropriate in other conditions prevalent in the older population, such as hearing impairments, visual disorders, and some cancers (17, 18).

These findings, together with the literature on the capability approach that shifts the focus away from a mere health-utility perspective, raise questions regarding the appropriateness of the EQ-5D in a population of middle-old and oldest-old people. Because the results of economic evaluations using the EQ-5D as effect measure are often considered when deciding on the implementation of interventions targeting the middle-old and oldest-old population, it is important to know whether the EQ-5D has satisfactory psychometric properties in this population.

Objective

This article provides the protocol for a systematic review that aims to synthesize and critically appraise studies assessing the psychometric properties of the EQ-5D in a population of middle-old and oldest-old people. Of interest are all studies reporting on reliability, validity, or responsiveness of the EQ-5D in a study population with a mean age of at least 75 years.

METHODS AND ANALYSIS

This protocol was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P) (19) and will be registered in PROSPERO (registration not yet completed/currently being assessed).

Eligibility Criteria

Cross-sectional or observational studies providing empirical evidence of reliability, validity, and/or responsiveness of the EQ-5D in a sample with a mean age of at least 75 years will be included. Included studies shall be published in peer-reviewed journals in German or English languages. Systematic reviews, studies applying a qualitative design, or studies being published in forms other than original articles (e.g., conference abstracts or comments) will be excluded. Furthermore, studies relying on proxy assessments only or those with the single objective of investigating agreement between different modes of administration of the EQ-5D will be excluded (e.g., studies

Abbreviations: HrQoL, health-related quality of life; COSMIN, COnsensus-based Standards for the selection of health Measurement Instruments project.

TABLE 1 | Search strategy in PubMed using an adapted version of the patient-reported outcome measurement filter available on the COSMIN website (20).

1#	(instrumentation[MeSH Subheading] OR "reproducibility of results"[MeSH Terms] OR reproducib*[Title/Abstract] OR "psychometrics"[MeSH] OR psychometr*[Title/Abstract] OR "discriminant analysis"[MeSH] OR reliab*[Title/Abstract] OR valid*[Title/Abstract] OR "internal consistency"[Title/Abstract] OR (cronbach*[Title/Abstract] AND (alpha[Title/Abstract] OR alphas[Title/Abstract])) OR "item correlation"[Title/Abstract] OR "item correlations"[Title/Abstract] OR agreement[Text Word] OR test–retest [Title/Abstract] OR (test[Title/Abstract] AND retest[Title/Abstract]) OR (reliab*[Title/Abstract] AND (test[Title/Abstract] OR retest[Title/Abstract])) OR intra-rater[Title/Abstract] OR intratester[Title/Abstract] OR intra-tester[Title/Abstract] OR OR intraobserver[Title/Abstract] OR intra-observer[Title/Abstract] OR intraindividual[Title/Abstract] OR intra-individual[Title/Abstract] OR intraparticipant[Title/Abstract] OR intra-participant[Title/Abstract] OR kappa[Title/Abstract] OR kappa's[Title/Abstract] OR kappa's[Title/Abstract] OR "coefficient of variation"[Title/Abstract] OR repeatable*[Text Word] OR ((replica*[Text Word] OR repeated[Text Word]) AND (measure[Text Word] OR measures[Text Word] OR findings[Text Word] OR result[Text Word] OR results[Text Word] OR test[Text Word] OR tests[Text Word])) OR concordance[Title/Abstract] OR (infraclass[Title/Abstract] AND correlation*[Title/Abstract]) OR discriminative[Title/Abstract] OR "known group" [Title/Abstract] OR "factor analysis"[Title/Abstract] OR "factor analyses"[Title/Abstract] OR "factor structure"[Title/Abstract] OR "factor structures"[Title/Abstract] OR dimensionality[Title/Abstract] OR subscale*[Title/Abstract] OR "item discriminant"[Title/Abstract]OR "interstate correlation"[Title/Abstract] OR "interstate correlations"[Title/Abstract] OR "individual variability"[Title/Abstract] OR "standard error of measurement"[Title/Abstract] OR sensitive*[Title/Abstract] OR responsive*[Title/Abstract] OR "minimal detectable concentration"[Title/Abstract] OR (small*[Title/Abstract] AND (real[Title/Abstract] OR detectable[Title/Abstract]) AND (change[Title/Abstract] OR difference[Title/Abstract])) OR "meaningful change"[Title/Abstract] OR "minimal important change"[Title/Abstract] OR "minimal important difference"[Title/Abstract] OR "minimally important change"[Title/Abstract] OR "minimally important difference"[Title/Abstract] OR "minimal detectable change"[Title/Abstract] OR "minimal detectable difference"[Title/Abstract] OR "minimally detectable change"[Title/Abstract] OR "minimally detectable difference"[Title/Abstract] OR "minimal real change"[Title/Abstract] OR "minimal real difference"[Title/Abstract] OR "minimally real change"[Title/Abstract] OR "minimally real difference"[Title/Abstract] OR "Item response model"[Title/Abstract] OR IRT[Title/Abstract] OR Rash[Title/Abstract] OR "Differential item functioning"[Title/Abstract] OR DIF[Title/Abstract])
#2	(EQ-5D) OR (EQ5D) OR (EuroQoL)
#3	(aged, 80 and over[MeSH Terms]) OR (aged[MeSH Terms]) OR (elderly[MeSH Terms]) OR (aged[Title/Abstract]) OR (elderly*[Title/Abstract]) OR (older*[Title/Abstract]) OR (geriatric*[Title/Abstract])
#4	#1 AND #2 AND #3
#5	("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses" [Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nigh"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])
#6	#4 NOT #5

only examining inter-rater agreement between the older person and a proxy). The question of inter-rater agreement between the patient and a proxy often concerns people with dementia and has been the subject of previous reviews (15, 16). There will be no restrictions relating to interventions, comorbidities/health conditions, publication date, or the version of the EQ-5D (three-level or five-level version).

Information Sources and Search Strategy

PubMed, Web of Science, and EconLit will be searched electronically in August 2020 using predefined search terms, including *EQ-5D*, *EuroQoL*, *aged*, *elder**, *old**, *geriatric**, *ag(e)ing*, and an adapted search filter for finding studies on measurement properties (20). This filter was developed to account for the large variation in terminology for measurement properties and unreliable indexing of studies under specific index terms, making it difficult to find all relevant studies under a small set of search terms (20). Because, for example, studies focusing on proxy assessments or interrater agreement only will be excluded from the planned review, search terms covering nonrelevant measurement properties (e.g., inter-rater reliability) were removed from the search filter. Where possible, search terms will be used as keywords in the title/abstract or Medical Subject Headings (MeSH terms). An example for the search strategy in PubMed is displayed in **Table 1**. Depending on the specific requirements of each database, the search terms will be modified. Additionally, the reference lists of included studies and previous reviews on HrQoL for middle-old and oldest-old people will be hand-searched.

Study Records (Data Management, Selection, and Collection)

Search results from all databases will be combined in a shared data repository and managed with the software Endnote X8. After removing duplicates, two independent reviewers (SG and MN) will screen the titles and abstracts for eligibility. Next, full texts of the selected abstracts will be assessed for eligibility by SG and MN independently. In case of disagreement or uncertainty, a third person (JD) will be consulted.

Using a standardized data extraction sheet, relevant data from the eligible studies will be extracted by one reviewer (SG) and cross-checked by the second reviewer (MN). Data extracted from the individual studies will include setting/country, population characteristics [sample size, distribution of age and sex, information on comorbidities (e.g., people with dementia)], instrument administration, type and method of validity, reliability and responsiveness assessment, and results of psychometric tests. The study selection process will be visualized in the form of a PRISMA flowchart.

Data Items

The review's main outcomes of interest will be the results regarding validity, reliability, and responsiveness of the EQ-5D reported by the individual studies. Regarding the outcomes, we adhere to the taxonomy and definitions from the COnsensus-based Standards for the selection of health Measurement Instruments project (COSMIN) (21). According to the COSMIN group, reliability refers to the degree to which the measurement is free from measurement error and can be differentiated between internal consistency, reliability, and measurement error. Validity is referred to as the degree to which an instrument measures the construct(s) it purports to measure and consists of the subtypes content validity, construct validity, and criterion validity. Responsiveness is defined as the ability of an instrument to detect change over time in the construct to be measured.

Assessment of Study Quality/Risk of Bias

Methodological quality of included studies will be assessed by the COSMIN Risk of Bias checklist, which has specifically been developed for use in systematic reviews of patient-reported outcome measures (22). It consists of 10 boxes, each referring to a particular measurement property and containing a different number of sub-questions. Each item is rated on a four-point scale, reaching from "very good" to "inadequate" (a "not applicable" option is also included). For each measurement property, an overall score will be determined by taking the lowest rating of any standard in the box ("worst score counts" principle). The checklist will be filled out by SG and MN independently. Any disagreements will again be resolved through discussion with a third person (JD).

Data Synthesis

Based on criteria of good measurement properties (23, 24), the results of the individual studies will be rated as either "sufficient" (+), "insufficient" (–), or "indeterminate" (?). The individual studies' results will then be summarized, and an overall rating of the measurement property will be assigned. The results will be presented in a thematic order by structuring the results section in the following sub-sections: validity, reliability, and responsiveness. Each sub-section will be further divided into sections on different types of reliability or validity (e.g., content validity, construct validity, criterion validity). If necessary, e.g., in case of inconsistencies between different study populations, the results will be presented separately for different population groups (e.g., validity in people with dementia) or versions of the EQ-5D (EQ-5D-3L and EQ-5D-5L). The guidelines for systematic reviews of measurement properties proposed by the COSMIN group (25) will guide the synthesis and interpretation of the results.

DISCUSSION

This review aims to provide a summary statement on the appropriateness of using the EQ-5D in the middle-old and oldest-old population by summarizing the evidence regarding the validity, reliability, and responsiveness. Although previous reviews had a similar aim, they either focused on the psychometric properties of the EQ-5D in people with dementia only (15, 16) or are outdated and were not specifically focusing on the population of middle-old and oldest-old (12). The planned review could identify gaps in research that should be addressed by future studies. Furthermore, recommendations for the use of the EQ-5D in a population of middle-old and oldest-old people could be made based on the results of the review. For example, the review may conclude that, in addition to the EQ-5D, age- or disease-specific instruments should be used to better capture the specific needs and experiences of older people or specific subgroups of older people. Thereby, the results may be of interest not only for decision-makers, but also for researchers planning or designing new intervention studies for older people.

Potential limitations may arise because of the heterogeneity of the individual studies regarding the population under study (e.g., people with dementia, people with femoral fractures), which may limit the possibility of making a synthesized statement on the appropriateness of the EQ-5D for the middle-old to oldest-old population. The expected heterogeneity in study design, measurements used, and populations further precludes the possibility of performing a meta-analysis. Moreover, it may not be possible to make a statement exclusively for the population 75 years or older as there seems to be a lack of studies focusing exclusively on this population. Therefore, the inclusion criteria have been adapted to a mean age of the sample of at least 75 years, which may lead to the inclusion of a number of persons younger than 75 years.

AUTHOR CONTRIBUTIONS

The study concept was developed by SG, JD, and H-HK. The manuscript of the protocol was drafted by SG and critically revised by JD, H-HK, and MN. The search strategy was developed by SG and JD. Study selection, data extraction, and quality assessment will be performed by SG and MN, with JD as a third party in case of disagreements. All authors contributed to the article and approved the submitted version.

REFERENCES

1. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects 2019, Volume II: Demographic Profiles.* New York, NY (2019).
2. EuroQol G. EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy.* (1990) 16:199–208. doi: 10.1016/0168-8510(90)90421-9
3. *ICECAP Capability Measures.* Available online at: https://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/HE/ICECAP/index.aspx (accessed June 06, 2020).
4. Sen A. *Choice, Welfare and Measurement.* Cambridge, MA: Harvard University Press (1982).
5. Coast J, Flynn T, Grewal I, Lewis J, Natarajan L, Sproston K, et al. Developing an index of capability for older people: a new form of measure for public health interventions? In: Dawson S, Morris ZS, editors. *Future Public Health.* London: Palgrave Macmillan (2009). p. 193–206.
6. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? *Soc Sci Med.* (2006) 62:1891–901. doi: 10.1016/j.socscimed.2005.08.023

7. Sutton EJ, Coast J. Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. *Palliat Med.* (2014) 28:151–7. doi: 10.1177/0269216313489368
8. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess.* (2014) 18:vii–viii, xiii–xxv:1–188. doi: 10.3310/hta18340
9. Mulhern B, Mukuria C, Barkham M, Knapp M, Byford S, Soeteman Dr, et al. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. *Br J Psychiatry.* (2014) 205:236–43. doi: 10.1192/bjp.bp.112.122283
10. Davis S, Wailoo A. A review of the psychometric performance of the EQ-5D in people with urinary incontinence. *Health Qual Life Outcomes.* (2013) 11:20. doi: 10.1186/1477-7525-11-20
11. Yang Y, Brazier J, Longworth L. EQ-5D in skin conditions: an assessment of validity and responsiveness. *Eur J Health Econ.* (2015) 16:927–39. doi: 10.1007/s10198-014-0638-9
12. Haywood KL, Garratt AM, Fitzpatrick R. Quality of life in older people: a structured review of generic self-assessed health instruments. *Qual Life Res.* (2005) 14:1651–68. doi: 10.1007/s11136-005-1743-0
13. Kuspinar A, Mayo NE. A review of the psychometric properties of generic utility measures in multiple sclerosis. *Pharmacoeconomics.* (2014) 32:759–73. doi: 10.1007/s40273-014-0167-5
14. Tordrup D, Mossman J, Kanavos P. Responsiveness of the EQ-5D to clinical change: is the patient experience adequately represented? *Int J Technol Assess Health Care.* (2014) 30:10–9. doi: 10.1017/S0266462313000640
15. Hounsome N, Orrell M, Edwards RT. EQ-5D as a quality of life measure in people with dementia and their carers: evidence and key issues. *Value Health.* (2011) 14:390–9. doi: 10.1016/j.jval.2010.08.002
16. Li L, Nguyen KH, Comans T, Scuffham P. Utility-based instruments for people with dementia: a systematic review and meta-regression analysis. *Value Health.* (2018) 21:471–81. doi: 10.1016/j.jval.2017.09.005
17. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ.* (2018) 19:557–70. doi: 10.1007/s10198-017-0902-x
18. Payakachat N, Ali MM, Tilford JM. Can the EQ-5D detect meaningful change? A systematic review. *Pharmacoeconomics.* (2015) 33:1137–54. doi: 10.1007/s40273-015-0295-6
19. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* (2015) 4:1. doi: 10.1186/2046-4053-4-1
20. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* (2009) 18:1115–23. doi: 10.1007/s11136-009-9528-5
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* (2010) 63:737–45. doi: 10.1016/j.jclinepi.2010.02.006
22. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res.* (2018) 27:1171–9. doi: 10.1007/s11136-017-1765-4
23. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* (2007) 60:34–42. doi: 10.1016/j.jclinepi.2006.03.012
24. Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials.* (2016) 17:449. doi: 10.1186/s13063-016-1555-2
25. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* (2018) 27:1147–1157. doi: 10.1007/s11136-018-1798-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8 Publication 6 – Measurement properties of the EQ-5D in populations with a mean age of ≥ 75 years: a systematic review

Published in *Quality of Life Research* as:

Gottschalk S[1], König HH[1], Nejad M[1], Dams J[1]. Measurement properties of the EQ-5D in populations with a mean age of ≥ 75 years: a systematic review. Qual Life Res. 2022 Aug 1 [Epub ahead of print]. DOI: 10.1007/s11136-022-03185-0.

[1]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Quality of Life Research
https://doi.org/10.1007/s11136-022-03185-0

REVIEW

Measurement properties of the EQ-5D in populations with a mean age of ≥ 75 years: a systematic review

Sophie Gottschalk[1] · Hans-Helmut König[1] · Mona Nejad[1] · Judith Dams[1]

Accepted: 21 June 2022

Abstract

Purpose Healthcare interventions for middle-old and oldest-old individuals are often (economically) evaluated using the EQ-5D to measure health-related quality of life (HrQoL). This requires sufficient measurement properties of the EQ-5D. Therefore, the current study aimed to systematically review studies assessing the measurement properties of the EQ-5D in this population.

Methods The databases PubMed, Cochrane library, Web of Science, Embase, and EconLit were searched for studies providing empirical evidence of reliability, validity, and/or responsiveness of the EQ-5D-3L and EQ-5D-5L in samples with a mean age ≥ 75 years. Studies were selected by two independent reviewers, and the methodological quality was assessed using the COSMIN Risk of Bias checklist. Results were rated against updated criteria for good measurement properties (sufficient, insufficient, inconsistent, indeterminate). The evidence was summarized, and the quality of evidence was graded using a modified GRADE approach.

Results For both EQ-5D versions, high-quality evidence for sufficient convergent validity was found. Known-groups validity was sufficient for the EQ-5D-5L (high-quality evidence), whereas the results were inconsistent for the EQ-5D-3L. Results regarding the reliability were inconsistent (EQ-5D-3L) or entirely lacking (EQ-5D-5L). Responsiveness based on correlations of change scores with instruments measuring related/similar constructs was insufficient for the EQ-5D-3L (high-quality evidence). For the EQ-5D-5L, the available evidence on responsiveness to change in (Hr)QoL instruments was limited.

Conclusion Since the responsiveness of the EQ-5D in a population of middle-old and oldest-old individuals was questionable, either using additional instruments or considering the use of an alternative, more comprehensive instrument of (Hr)QoL might be advisable, especially for economic evaluations.

Keywords EQ-5D · Older population · Oldest-old population · Psychometric properties · Systematic review

Abbreviations

ADL	Activities of daily living
ASCOT	Adult Social Care Outcomes Toolkit
AQoL	Assessment of Quality of Life
BBS	Berg Balance Scale
BADL	Bristol Activities of Daily Living Scale
CCCQ	Client-centered Care Questionnaire
CDR	Clinical Dementia Rating
CMAI	Cohen-Mansfield Agitation Inventory
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
DEMQOL	Dementia Quality of Life instrument
ESM	Electronic supplementary material
EQ-HWB	EQ Health and Wellbeing instrument
EQ-VAS	EQ-Visual Analogue Scale
FAST	Functional Assessment Staging Tool
HrQoL	Health-related quality of life
HUI3	Health Utilities Index
IADL	Instrumental activities of daily living
ICC	Intraclass correlation coefficient

✉ Sophie Gottschalk
s.gottschalk@uke.de

Hans-Helmut König
h.koenig@uke.de

Mona Nejad
m.nejad@uke.de

Judith Dams
j.dams@uke.de

[1] Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany

Published online: 01 August 2022

ICECAP-O	ICEpop CAPability measure for Older people
MBI	Modified Barthel Index
MCID	Minimally clinically important difference
MeSH	Medical subject headings
NHP	Nottingham Health Profile
NOSGER	Nurses' Observation Scale for Geriatric Patients
OHS	Oxford Hip Score
OPQOL-Brief	Older People's Quality of Life questionnaire, short version
PPA	Physiological Profile Assessment
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analysis
PROMIS	Patient-Reported Outcomes Measurement Information System
QALY	Quality-adjusted life years
QoL	Quality of life
QoL-AD	Quality of Life in Alzheimer's Disease scale
QOL-AD-NH	Quality of Life in Alzheimer's Disease in Nursing Homes
QWB	Quality of Well-Being scale
SF-36	36-item Short-Form health survey
SF-12	12-item Short-Form health survey
SF-6D	Short Form 6 Dimensions
SPPB	Short Physical Performance Battery
SPVU-5D	5-Dimensional Sheffield Preference-based Venous Ulcer questionnaire
UK	United Kingdom
US	United States
WHOQOL-OLD	World Health Organization Quality of Life - Older Adults
30 s STS	30-second Sit-To-Stand test

Introduction

Maintaining health of an increasing number of middle-old and oldest-old people is a major challenge for aging societies [1]. Population norms of health-related quality of life (HrQoL) suggest that HrQoL decreases with age and drops considerably beyond the age of 75 [2, 3]. Numerous interventions targeting this population are, therefore, being developed. In the face of scarce resources, new interventions should be economically evaluated before being implemented in the healthcare system, as such information can assist in the efficient allocation of resources.

To make effects comparable across interventions, economic evaluations often measure effectiveness in terms of quality-adjusted life years (QALY), where the 'Q' is measured using generic HrQoL instruments. The most frequently used instrument, in general but also for evaluation of interventions targeting the older population, is the EQ-5D [4–6], which is the officially required standard measurement in some countries (e.g., UK [7]). It consists of five questions covering the dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Depending on the version of the EQ-5D, each dimension has three (EQ-5D-3L) or five (EQ-5D-5L) severity levels ("no problems" to "extreme problems"). The combined answers can be transformed to an index with 0 representing death and 1 representing the best possible HrQoL. It is important that the EQ-5D is psychometrically sound in the population it is used, meaning that it measures what it intended to measure (validity) in an accurate and reproducible way (reliability) and is able to detect important changes over time (responsiveness). In the absence of sufficient measurement properties, the results of economic evaluations fail in measuring the true effect of interventions and, thus, are not suitable as basis for decision making regarding their implementation.

Previous reviews examined the psychometric performance of the EQ-5D in different population groups. It was found appropriate for depression and personality disorders [8, 9], urinary incontinence [10], some skin diseases [11], and in people aged 60 or older [12]. However, its psychometric performance was lacking in populations with anxiety, schizophrenia, bipolar disorders, or multiple sclerosis [8, 9, 13]. Moreover, it was found insufficiently sensitive to change in a range of disorders [14]. Regarding its use in dementia, the validity was found problematic as there are significant disagreements between patient and proxy ratings and aspects being important for people with dementia are not adequately reflected [15, 16]. Similarly, other authors conclude that the EQ-5D may not be appropriate in other conditions prevalent in the older population, such as hearing impairments, visual disorders, and some cancers [17, 18]. A common problem seems to be that the EQ-5D has limited ability to differentiate between healthier individuals [19]. Although this ceiling effect could be reduced for the EQ-5D-5L, it still exists [20]. Moreover, the EQ-5D has been criticized for its narrow focus of health, which may fall short on or excludes important aspects of health (e.g., social aspects) [21]. As people's needs and desires change with age, it can be assumed that, especially in old age or at the end of life, such aspects become more important [22–24].

These findings raise questions regarding the measurement properties of the EQ-5D in middle-old and oldest-old people. To our knowledge, there has been no systematic summary of the measurement properties of the EQ-5D in this population. In a review that is more than a decade old, Haywood et al. [12] evaluated the measurement and practical properties of generic health instruments in older people and found evidence for the validity of the EQ-5D. In terms of responsiveness, the EQ-5D appeared to perform well in people with substantial changes in health; however, responsiveness in terms of correlation of change scores between the EQ-5D and other (clinical) measures was rarely addressed until then. In addition to being outdated and hence including only studies using the EQ-5D-3L, this review did not specifically focus on middle-old and oldest-old people. More recent reviews concluded that the EQ-5D has good feasibility properties in an older population [25], but due to its sole focus on health status, may not be appropriate for measuring outcomes in economic evaluation within aged care, especially in interventions that have effects beyond health status [6, 26, 27]. However, the authors focused exclusively on dependent older people and/or did not systematically summarize the measurement properties of the EQ-5D. Therefore, the aim of the current study was to extend the existing literature by synthesizing and critically appraising studies assessing the measurement properties—reliability, validity, or responsiveness—of the EQ-5D in a population of middle-old and oldest-old people (mean age ≥ 75 years).

Materials and methods

This review was conducted in adherence with the Consensus-Based Standards for the Selection of Health Measurement Instrument (COSMIN) Methodology for Systematic Reviews of Measurement Properties of PROMs [28]. It has been registered with PROSPERO (Registration Number: CRD42020196070), and a study protocol has been published [29]. The manuscript was prepared based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist (electronic supplementary material [ESM] 1) [30].

Eligibility criteria

Cross-sectional or observational studies providing empirical evidence of reliability, validity, and/or responsiveness of the EQ-5D in a sample with a mean age of ≥ 75 years were included. Studies had to be published in peer-reviewed journals in German or English languages. Systematic reviews, studies applying a qualitative design, or not being original research articles (e.g., conference abstracts or comments) were excluded. Furthermore, studies relying on proxy assessments only or those with the single objective of investigating agreement between different modes of administration of the EQ-5D were excluded. The question of inter-rater agreement between the patient and a proxy often concerns people with dementia and has been addressed in previous reviews [15, 16]. No restrictions relating to interventions, health conditions, publication date, or the version of the EQ-5D (3-level or 5-level) were made.

Data sources and search strategy

PubMed, Web of Science, Cochrane Library, Embase, and EconLit were searched electronically on March 10, 2021 using predefined search terms, including *quality of life*, *health-related quality of life*, *EQ-5D*, *EuroQoL*, *aged*, *elder**, *old**, *geriatric**, and *ag*(*e*)*ing* and an adapted search filter for finding studies on measurement properties [31]. Search terms covering non-relevant measurement properties were removed from the search filter (e.g., inter-rater reliability or cross-cultural validity). Where possible, search terms were used as keywords in the title/abstract or Medical Subject Headings (MeSH). An example for the search strategy in PubMed is displayed in Table S1 (ESM 1). Additionally, reference lists of included studies were hand searched.

Selection of studies and data extraction

Search results from all databases were combined in a shared data repository and managed with Endnote X8. After removing duplicates, two independent reviewers (SG and MN) screened the titles and abstracts and assessed the full texts of the selected abstracts for eligibility. In case of disagreement or uncertainty, a third person (JD) was consulted. Using a standardized data extraction sheet, relevant data from the eligible studies were extracted by one reviewer (SG) and cross-checked by the second reviewer (MN). Data extracted from the individual studies included setting/country, population characteristics, type and method of validity, reliability and responsiveness assessment, and results for each measurement property.

Assessment of study quality

Methodological quality of included studies was assessed by two reviewers (MN and SG) using the COSMIN Risk of Bias checklist, which was developed specifically for the use in systematic reviews of patient-reported outcome measures [32]. It consists of 10 boxes, each referring to a particular measurement property and containing a different number of sub-questions. Each item is rated on a four-point scale ("very good" to "inadequate"). Any disagreements were resolved through discussion with a third person (JD). Risk of bias rating for each study and measurement property are provided in ESM 2.

Evaluation of measurement properties

Updated criteria for good measurement properties were applied to rate the individual studies' results as "sufficient" (+), "insufficient" (−), or "indeterminate" (?) [33]. Reliability was considered "sufficient" if the intraclass correlation coefficient (ICC) was ≥ 0.70. Construct validity and

Table 1 Generic/general hypotheses for construct validity and responsiveness (adapted from Prinsen et al., [28])

H1	Correlations with (changes in) instruments measuring similar constructs should be high (≥ 0.5)
H2	Correlations with (changes in) instruments measuring related, but dissimilar constructs should be at least moderate (≥ 0.3)
H3	Correlations with (changes in) instruments measuring weakly related constructs should be at least weak (≥ 0.1)
H4	Correlations with (changes in) instruments measuring unrelated constructs should be negligible (< 0.1)
H5	Meaningful changes between relevant (sub) groups. MCID of the EQ-5D: 0.074 (EQ-5D-3L)[35] or 0.063 (EQ-5D-5L) [36]
H6	For responsiveness (criterion approach), AUC should be ≥ 0.7
H7	HrQoL may decreases with age, but not necessarily, given the circumstances that this review focusses only on middle-old to oldest-old people
H8	Higher education level/social class might be associated with higher HrQoL, but not necessarily, since the differences may no longer be present in this age group (in later life, lifestyle factors such as physical activity become more important [103])
H9	Lower cognitive status is hypothesized to be associated with lower HrQoL in institutionalized people and/or people with severe dementia, whereas this association may not be visible in people with mild to moderate dementia or non-institutionalized people [102]

AUC area under the curve, *MCID* minimal clinically important differences

responsiveness were rated "sufficient" if the result was in accordance with predefined hypotheses. The hypotheses were formulated by the review team in advance and where partly (but not necessarily) adopted from the authors of the individual studies. Generic hypotheses applied in this study are presented in Table 1. A detailed overview of specific hypotheses for each individual study is provided in Table S2, ESM 1. The hypotheses regarding the discriminative ability of the EQ-5D between relevant subgroups (e.g., known-groups validity or responsiveness) were accepted if the difference between subgroups was clinically relevant, which was considered more important than whether the difference is statistically significant [34]. For the EQ-5D-3L index, a minimally clinically important difference (MCID) of 0.074 was applied, which was identified as the mean MCID across different patient groups [35]. The studies reporting on known-groups validity or responsiveness of the EQ-5D-5L index were either conducted in the UK or used UK value sets. Therefore, an MCID of 0.063 was applied, which was identified as MCID for England [36].

Summary and grading of the quality of evidence

Criteria for good measurement properties were applied to the summarized results from the individual studies on each measurement property by rating each property as "sufficient" (+), "insufficient" (−), "inconsistent" (±), or "indeterminate" (?) [33, 37]. For construct validity and responsiveness, the measurement property was rated "sufficient" when ≥ 75% of the individual studies' results were in accordance with predefined hypotheses. The results were qualitatively summarized by providing, e.g., a range of correlation coefficients for convergent validity and the percentage of hypotheses accepted. The evidence synthesis was performed separately for the EQ-5D-3L and EQ-5D-5L. If the results were inconsistent, reasons for inconsistency were explored (e.g., different results for different subgroups). If no reason for inconsistency could be identified, the result was rated "inconsistent" and the quality of evidence was not further explored. Due to heterogeneity of the populations included in the individual studies, quantitative pooling of results was not performed.

The quality of evidence was graded as "high," "moderate," "low," or "very low" using a modified GRADE approach [38]. Starting with the assumption of "high quality," it was downgraded if there was a risk of bias (up to − 3 levels), (unexplained) inconsistency (up to − 2 levels), imprecision (e.g., small sample size; up to − 2 levels), or indirect results. Indirectness was not applied in this study since studies examining the measurement properties in other populations than the population of interest were excluded. Specific criteria for downgrading are described in the COSMIN manual [34].

Results

Search results

The search strategy resulted in 4346 records (duplicates removed). After screening of title and abstract, 4107 records were excluded, leaving 239 records of which full texts were assessed for eligibility. Finally, 38 records were included for the qualitative synthesis (Fig. 1). No further relevant studies were identified through reference screening. The majority of studies ($n = 30$) evaluated the measurement properties of the EQ-5D-3L [39–68], whereas 9 studies evaluated the EQ-5D-5L [41, 69–76]. One study evaluated both EQ-5D versions [41].

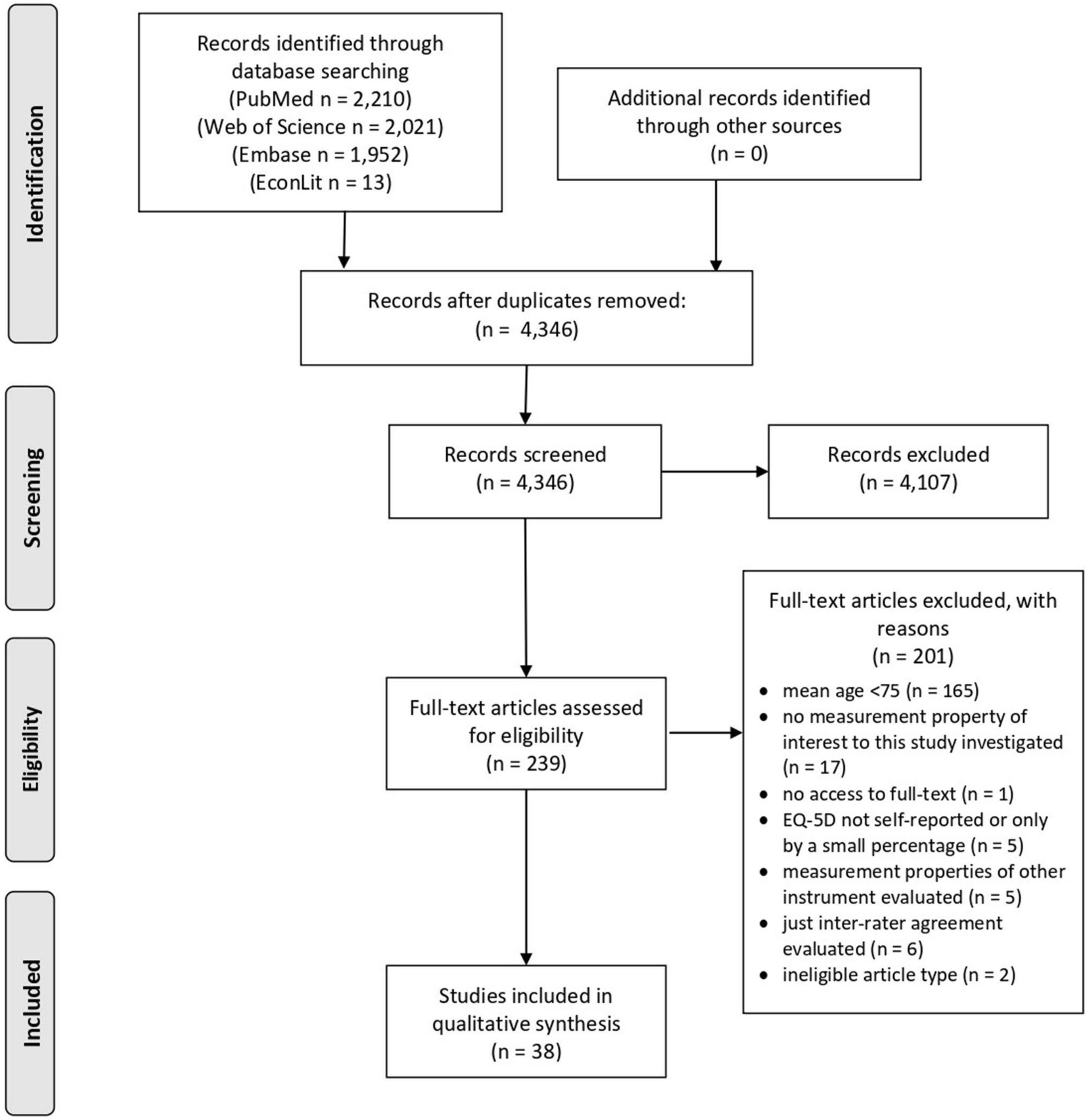


Fig. 1 Selection process of included studies

General characteristics of the articles

Characteristics of the included studies are described in Table 2. Studies covered a variety of (disease) populations, such as people with dementia or cognitive impairment ($n = 13$) [39, 50, 52, 54, 57, 58, 60, 62–64, 69, 72–74], people with different kinds of fractures ($n = 7$) [43, 46, 59, 61, 65, 66, 76], people who were frail or had a history of falling ($n = 4$) [44, 45, 67, 70], or people with venous leg ulcers ($n = 2$) [68, 71]. The studies were conducted in the UK ($n = 12$) [40, 42, 43, 47, 49, 60, 61, 68, 69, 73–75], Sweden ($n = 3$) [59, 65, 66], Spain ($n = 2$) [62, 63], Norway ($n = 2$) [46, 70], Finland ($n = 1$) [48], France ($n = 1$) [39], Germany ($n = 2$) [54, 57], Korea ($n = 1$) [53], the Netherlands ($n = 2$) [55, 67], Australia ($n = 4$) [51, 71, 72, 76], Canada ($n = 3$) [44, 45, 58], the USA ($n = 2$) [52, 56], Mexico ($n = 1$) [64], Sweden/Denmark/Finland/Norway ($n = 1$) [50], or Belgium/Ireland/Netherlands/Switzerland ($n = 1$) [41]. Participants

Table 2 Characteristics of the included studies

Ref	Population			Disease/population characteristics			Instrumental administration	
	N*	Age Mean (SD, range)	% female	Disease/other characteristics of the study population/recruited from	Disease duration	Disease severity	Interview administration mode	Country
EQ-5D-3L								
Ankri et al. [39]	142	82.9 (8.3) (60–99)	79.6%	PwD; hospitalized, institutionalized, or outpatients; recruited from geriatric hospital centers	N.R	47.1% moderate, 27.9% severe	Assisted interviews	France
Barton et al. [40]	392	N.R.[1]	N.R	Registered in one general practice	N.A	N.A	N.R	UK
Brazier et al. [42]	377	80.1 (4.5)	100%	Older women, 86.5% long-standing illness or disability; recruited from four general practices into a RCT of clodronate	N.A	N.A	N.R	UK
Coast et al. [43]	214	79[2] (74–84)[3]	70%	Elderly acute care patients being suitable for rehabilitation in their own home (mainly fractured neck of the femur, elective hip and knee replacements, other fractures, and stroke)	N.A	N.A	N.R	UK
Davis et al. [44]	215	79.3 (6.2)	71.6%	Older adults at risk of mobility impairment and a fall history; visiting the Vancouver falls prevention clinic	N.A	N.A	N.R	Canada
Davis et al. [45]	356	82 (6.5)	63%	Older adults at risk of mobility impairment and a fall history; visiting the Vancouver falls prevention clinic	N.A	N.A	N.R	Canada

Table 2 (continued)

Ref	Population			Disease/population characteristics			Instrumental administration	
	N*	Age Mean (SD, range)	% female	Disease/other characteristics of the study population/recruited from	Disease duration	Disease severity	Interview administration mode	Country
Frihagen et al. [46]	222 (complete cases at follow-up: $n=79$)	82.8 (7.48)	74%	Patients with displaced femoral neck fracture	N.R	$n=23$ complication group, $n=56$ non-complication group	Outpatient clinic and home visits	Norway
Hazell et al. [47]	721	N.R.[1]	N.R	Registered in two general practices	N.A	N.A	Postal questionnaire	UK
Heiskanen et al. [48]	36	N.R.[1]	N.R	Patients admitted for CABG operation	N.A	> 60% Canadian cardiovascular society class 3 or 4 (indicating lower functional status)	As part of the pre-operative hospital admission process (baseline) and via postal survey (follow-up)	Finland
Holland et al. [49]	145	84.7	57%	People taking ≥ 2 medications; admitted to hospital as an emergency; to be discharged from hospital and returning home/to a warden control accommodation	N.A	Median of 6 daily drugs	Assisted (baseline) interviews	UK
Jönsson et al. [50]	272	75.9	62.3	PwAD, recruited among patients attending regular visits at memory clinics; community dwelling or in residential care	Diagnosed on average 1.6 years prior to inclusion	MMSE > 25 to MMSE < 10	N.R	Sweden, Denmark, Finland, Norway
Kaambwa et al. [51]	87	80 (65–93)	66%	Receiving community aged care services, but cognitively intact	N.A	N.A	Group setting in central venues (research team just there for clarification of questions)	Australia

Table 2 (continued)

Ref	Population			Disease/population characteristics			Instrumental administration	
	N*	Age Mean (SD, range)	% female	Disease/other characteristics of the study population/recruited from	Disease duration	Disease severity	Interview administration mode	Country
Karlawish et al. [52]	93	76.8 (2.7) (55–91)	45%	PwAD, not living in care homes, receiving CG assistance; recruited from geriatric medicine practice or memory clinic	N.R	Mild to moderate AD (71% very mild)	Assisted interviews at the participant's/CG's home or other convenient location	USA
Kim et al. [53]	2826	N.R.[1]	0%	General male adult population sample, participating in Korean community health survey	N.A	57% moderate to severe lower urinary tract symptoms	Face-to-face interviews	
Kunz [54]	390	80.2 (6.7) (65–100)	68%	PwD, living at home and supported by a family caregiver; recruited via general practitioners	N.R	Mild to moderate	Assessed at GP practices by trained GP and medical secretary	Germany
Lutomski et al. [55]	25,637	78 (6)	58.3%	Community-dwelling older persons aged 65+; recruited from primary care centers, hospitals, or the general population	N.A	73% with hearing issues, joint damage, urinary incontinence, and/or dizziness with falls	N.R	Netherlands
Malkin et al. [56]	77	77[2] (27–98)	71%	Low-vision patients; presenting for low-vision rehabilitation at private outpatient centers	N.R	55% macular disorders	Telephone interview before 1st visit at low vision rehabilitation site	USA
Michalowsky et al. [57]	560	79.03 (8.5)	59.1%	PwD, living in the community; supported by informal CG	N.R	N.R	Face-to-face interview	Germany
Naglie et al. [58]	57	78.6 (53.8–93.7)	61.7%	PwAD, supported by a family CG; recruited from dementia clinics & geriatric practices	N.R	Mild to moderate	Assisted interview at participant's home/referring clinic	Canada

Table 2 (continued)

Ref	Population			Disease/population characteristics			Instrumental administration	
	N*	Age Mean (SD, range)	% female	Disease/other characteristics of the study population/recruited from	Disease duration	Disease severity	Interview administration mode	Country
Olerud et al. [59]	145	74.7 (9.6) (55–93)	84%	Patients with acute proximal humeral fracture; living non-institutionalized; no severe cognitive dysfunction	N.A	2-part to 4-part fractures	N.R	Sweden
Orgeta et al. [60]	478	75.5 (7.3)	49.6%	PwD living in the community; supported by carers assisting with ADL	N.R	Mild (74.6%) to moderate (25.4%) dementia	Assisted interview	UK
Parsons et al. [61]	225 (sample 1) 249 (sample 2)	83.1 (7.94) 83.6 (7.77)	71% 75%	Patients with hip fracture	N.A	32% and 41% PwD	Face to face at BL; telephone interview at 4 week, 4 month & 12 month FU	UK
Pérez-Ros and Marínez-Arnau [62]	251	84.6 (9.22) (70–104)	76.9%	Nursing home residents with cognitive impairment	N.A	Mean MMSE score: 15.6 (5.23)	Face-to-face interview	Spain
Pérez-Ros et al. [63]	188	79.19 (5.18) (70–95)	64.9%	Community-dwelling older people with cognitive impairment	N.A	MMSE scores 10–24	Face-to-face interview	Spain
Sanchez-Arenas et al. [64]	109	78.5 (7.09)	64.2%	PwD; community dwelling	N.A	N.R	In-home face-to-face interview	Mexico
Tidermark et al. [65]	110	80.0 (70–96)	79%	People with acute displaced femoral neck fractures; no severe cognitive impairment	Fractures ≤ 24 h. old	51% total hip replacement, 49% internal fixation	Structured interview at initial hospital stay and at 4 months	Sweden
Tidermark & Bergström [66]	59	82.9 (5.4) (70–92)	100%	Women with acute femoral neck fractures; no severe cognitive impairment	Fractures ≤ 24 h. old	N.A	Structured interview at initial hospital stay and at 6 months	Sweden
van Leeuwen et al. [67]	190	82.4 (7.7)	71.6%	Frail older adults living at home; recruited from general practices	N.A	N.A	Interview at participants' homes	Netherlands

Table 2 (continued)

Ref	Population			Disease/population characteristics			Instrumental administration	
	N*	Age Mean (SD, range)	% female	Disease/other characteristics of the study population/recruited from	Disease duration	Disease severity	Interview administration mode	Country
Walters et al. [68]	233	75[2]	66.5%	People with venous leg ulcers	present for ≥ 3 months	N.R	Community clinic setting	UK
EQ-5D-5L								
Aguirre et al. [69]	272	82.6 (8.1) (52–100)	61%	PwD; recruited from community (58%) and care homes (42%)	N.R	Mild to moderate	N.R	UK
Bjerk et al. [70]	155	82.7 (6.7)	79.3%	Older home-care recipients; experienced ≥ 1 fall in previous 12 months, not cognitively impaired; recruited from lists of people receiving professional home care	N.A	N.A	Interviewed in participants' homes by trained research assistants	Norway
Cheng et al. [71]	80	75 (13.88) (30–95)	59%	People with venous leg ulcers; recruited from 2 communities, 1 specialist & 1 hospital outpatient wound clinic	0–369 months	Venous insufficiency (41%), reduced mobility (81%)	N.R	Australia
Easton et al. [72]	143	85.7 (8.8) (49–99)	72%	Residential care setting	N.A	45% mild or moderate cognitive impairment, 25% dementia	N.R	Australia
Griffiths et al. [73]	726 (377 completed self-report measures)	85.6 (7.64) (57–102)	73.8%	PwD, living in care homes	N.R.	N.R	N.R	UK
Martin et al. [74]	1004	85.5 (58–103)	73.2%	PwD, living in residential care	N.A	N.A	Recruited from care homes	UK
Nikolova et al. [75]	1038	N.R (75% 75–84, 25% 85+)	52.7%	Community-dwelling older people; recruited from general practices	N.A	20.2% fit, 51.4% pre-frail, 28.4% frail	Face-to-face interview	UK

Table 2 (continued)

Ref	Population			Disease/population characteristics			Instrumental administration	
	N*	Age Mean (SD, range)	% female	Disease/other characteristics of the study population/recruited from	Disease duration	Disease severity	Interview administration mode	Country
Ratcliffe et al. [76]	240 (EQ-5D self-completed: $n=82$)	88.6 (5.6)	74.2%	Patients with hip fracture, living in residential aged care; recruited from acute orthopedic wards	N.A	93% moderate/severe dementia	N.R	Australia
EQ-5D-3L & EQ-5D-5L								
Bhadhuri et al. [41]	224	77.5 (5.35)	43.8%	People with multimorbidity and polypharmacy participating in a structured medication review RCT	N.A	N.A	Telephone interview	Belgium, Ireland, Netherlands, Switzerland

N.R not reported, *N.A.* not applicable, *ADL* activities of daily living, *BL* baseline, *CG* caregiver, *FU* follow-up, *GP* general practice/practitioner, *MMSE* mini mental state examination, *Pw(A)D* people with (Alzheimer's) dementia, *RCT* randomized controlled trial

*Sample size may differ for specific analyses, [1]only results for age group ≥ 75 years (or ≥ 80 years [53]) were extracted and included in this review, [2]median, [3]inter-quartile range were recruited from different settings, e.g., residential care homes, home-care registries, general practices, falls prevention clinics, or the general population.

Evidence synthesis (Measurement properties)

The summarized results are presented in Table 3 (EQ-5D-3L) and Table 4 (EQ-5D-5L).

Reliability

In total, five studies assessed the reliability of the EQ-5D-3L index, with three reporting sufficient [39, 58, 67] and two reporting insufficient reliability [42, 52]. In one of the two studies of insufficient reliability [42], the time interval between measurements (6 months) was inappropriate (doubtful methodological quality). However, for the other study with insufficient reliability [52], no possible explanation could be found (similar population and/or time interval like in other studies reporting sufficient reliability [39, 58]). Thus, the overall rating of reliability of the EQ-5D-3L was inconsistent. Very low-quality evidence regarding the reliability of the individual dimensions of the EQ-5D-3L was available from one study [39], which found insufficient reliability based on Kappa coefficients between 0.34 and 0.59.

No study regarding the reliability of the EQ-5D-5L could be identified.

Convergent validity

Overall, convergent validity for both EQ-5D versions was supported by multiple studies, with the majority of hypotheses being supported at moderate to high quality of evidence.

As hypothesized, strong correlations between the EQ-5D-3L index and other instruments of HrQoL (SF-12, SF-6D, SF-36, HUI3) were found [40, 58, 65, 67]. At least moderate correlations were found with instruments of QoL (ICECAP-O, OPQOL-Brief, ASCOT, AQOL, QWB, QoL-AD) [44, 49–51, 57, 58, 61, 67], activities of daily living (ADL) (Barthel, Katz, BADL) [54, 58, 62, 64, 67], or single-scale instruments of general health or QoL [39, 50, 55, 57, 58, 63, 67]. Moreover, at least weak correlations with instruments of instrumental activities of daily living (IADL) (e.g., Lawton-Brody, NOSGER) [44, 54, 58, 62, 64] and comorbidities [58, 64] were found in the majority of studies. Results were inconsistent regarding the convergent validity of the EQ-5D-3L index with measures of depression/anxiety, which were hypothesized to be at least weakly correlated [57, 58, 60, 62].

Similarly, the EQ-5D-5L index was strongly correlated with the SF-6D as measure of HrQoL [70, 75]. At least moderate associations were found with QoL instruments (DEMQOL, DEMQOL-U, QOL-AD, SPVU-5D) [69, 71–73] (with the exception of the QoL-AD-NH [74]), as well as with a single-scale instrument for general health (EQ-VAS) [71] or a measure of ADL (MBI) [72, 76]. Results were inconsistent for associations with measures of cognitive status (Hypothesis 9, Table 1) [72, 74, 76], where one study found a positive correlation, although an association in the opposite direction was hypothesized [72].

Several studies [39, 41, 43, 44, 50, 51, 55, 56, 62–64, 68, 70–72, 75] also assessed convergent validity by correlating the EQ-5D index with the individual dimensions of the comparator instrument, the EQ-5D dimensions with a comparator instrument's summary score, or the EQ-5D dimensions with the comparator's dimensions (Tables S3 & S4, ESM 1). For both EQ-5D versions, the majority of results were in accordance with the hypotheses, thus, supporting the overall rating of convergent validity as sufficient.

Known-groups validity

Twelve studies assessed known-groups validity of the EQ-5D-3L index in a variety of populations [39, 42, 43, 47, 49, 51–55, 57, 68]. Overall, known-groups validity was inconsistent as < 75% of the results (67%) were in accordance with the hypotheses.

For the EQ-5D-5L index, known-groups validity was assessed in three studies [71, 72, 76]. The overall result was rated sufficient (78% of the hypotheses supported) and the quality of evidence was rated high.

Detailed information about the groups that the EQ-5D-3L and EQ-5D-5L were able to discriminate between can be found in Tables 3 & 4.

Responsiveness

Eight studies assessed responsiveness of the EQ-5D-3L index by examining the associations of change scores with other instruments [48, 49, 54, 56, 59, 65–67]. With one exception (AQoL) [49], the correlations with changes in instruments of HrQoL (SF-36, SF-12, NHP, 15D) [48, 65–67], QoL (ICECAP-O, ASCOT) [67], single-scale instruments of general health or QoL [67], ADL (Barthel, Katz) [54, 67], and IADL (NOSGER) [54] were weaker than hypothesized. Thus, responsiveness based on the comparison with other instruments was rated insufficient, and the summarized quality of evidence was rated high.

Ten studies assessed responsiveness of the EQ-5D-3L index based on comparisons between subgroups [41, 43, 45, 46, 54, 59, 61, 65, 66, 68]. These studies were primarily conducted on specific patient populations and assessed, e.g., the ability of the EQ-5D to differentiate between different outcomes after fractures or venous leg ulcers. Overall,

Table 3 Summary of findings—EQ-5D-3L

Measurement property	Summary	Overall rating	Quality of evidence
Reliability	Sub-dimensions [39]: Kappa: 0.34–0.59 (*n = 45*)	–	**very low**
	Index: *ICC* = 0.58–0.79 [39, 52, 58, 67], *r* = 0.67 [42] (*n = 439*)	±	**N/A**
Construct validity			
Convergent validity		**+ (91%)**	**high**
HrQoL instruments (*Hypothesis*: $r \geq 0.5$)	SF-6D [40], SF-36 [65], HUI3 [58], 15D [48]: **0.44** [48]–0.74; SF-12 MCS [67]: 0.36[a]; SF-12 PCS [67]: 0.60 (*n = 633 or higher* (*n.c.r.* [58]))	+ (83%)	high
QoL instruments (*Hypothesis*: $r \geq 0.3$)	ICECAP-O [44, 61, 67], OPQOL-Brief [51], ASCOT [51, 67], AQoL [49], QWB scale [58], QoL-AD [50, 57]: 0.34–0.73 (*n≈1,588* (*n.c.r.* [50, 57, 58, 61]))	+ (100%)	high
General health/QoL (single-scale) (*Hypothesis*: $r \geq 0.3$)	Health GRS [67], EQ-VAS [39, 50, 58, 63], QoL GRS [67], Cantril's Self-Anchoring ladder [55], SF-36 general health [58], other [57] (3-pt ordinal scale): 0.34–0.52 (*n≈27,978* (*n.c.r.* [50]))	+ (100%)	high
ADL (*Hypothesis*: $r \geq 0.3$)	Barthel [54, 62], Katz [58, 64, 67]: **0.25** [58]–0.71; Bristol Activities of Daily Living Scale [60]: $\beta = -0.257$ (*n = 1356 or higher* (*n.c.r.* [58]))	+ (86%)	moderate
IADL (*Hypothesis*: $r \geq 0.1$)	Lawton & Brody [44, 58, 62], other [64], NOSGER [54]: **0.03** [44], 0.22–0.62 (*n = 904 or higher* (*n.c.r* [58]))	+ (80%)	moderate
Comorbidities (*Hypothesis*: $r \geq 0.1$)	Charlson [64], other [58] (0, 1, ≥ 2): 0.30–0.36 (*n = 102 or higher* (*n.c.r.* [58]))	+ (100%)	high
Cognitive status/dementia severity (*Hypothesis*: $r < 0.3$)	MMSE [44, 50, 54, 58, 64]: 0.07–0.20 (*n≈1,000* (*n.c.r.* [50, 58]))	+ (100%)	moderate
Depression/anxiety (*Hypothesis*: $r \geq 0.1$)	GDS [57, 58, 62]: **0.042** [62], 0.21–0.55; CSDD [60]: $\boldsymbol{\beta = -0.065}$ ($p > 0.05$); RAID [60]: $\beta = -0.168$ (*n≈1,280* (*n.c.r.* [57, 58]))	± (60%)	N/A
Other instruments	*n≈770* (*n.c.r.* [61])	+ (100%)	high
(*Hypothesis*: $r \geq 0.3$)	OHS [61]: 0.70–0.77		
(*Hypothesis*: $r \geq 0.1$)	Pearlin Mastery Scale [67], Tinetti [62], VAS Pain [62]: 0.17–0.33		
(*Hypothesis*: $r < 0.3$)	CCCQ [67], PPA [44], SPPB [44]: 0.01–0.06		
Known-groups validity	*n≈31,176* (*n.c.r.* [49, 53, 54, 57])	**± (67%)**	**N/A**
Supported for groups of…	Age[b] [42, 43, 49, 55, 68], sex [49, 51], social class[b] [49], education level[b] [51, 55], general health [51, 52, 57], mental & physical functioning (SF-12) [52], QoL-AD Score/Whole/Memory [52], IADL impairment (Lawton & Brody) [52, 57], disability severity [42], walking ability [68], number of medications [49], lower urinary tract symptom severity [53], obstructive airways disease (y/n) [47], depression (GDS) [52, 57], hospital stay (y/n) [42], multimorbidity [55], longstanding illness (y/n) [42], cognition (MMSE)[b] [52], confusion (mental test score)[b] [49], memory problems (GDS Memory)[b] [52],		
Rejected for groups of…	Age[b] [51], sex [55], living situation (alone vs. not alone/other arrangement) [49, 51, 55], informal care support (y/n) [51], marital status [55], GP visit (y/n) [42], outpatient attendance (y/n) [42], accident/emergency department attendance (y/n) [42], ADL impairment (higher vs. lower, Lawton-Brody) [52], only dementia vs. dementia + additional comorbidity [54], leg ulcer size and duration [68], functional impairment due to dementia [57], QoL-AD Life [52]		
Responsiveness			
Construct approach		**− (22%)**	**high**
HrQoL instruments (*Hypothesis*: $r \geq 0.5$)	SF-36 [65], NHP [66], SF-12 PCS [67], 15D[c] [48]: **0.23–0.39;** SF-12 MCS[a] [67]: **0.02** (*n = 430*)	− (0%)	high
QoL instruments (*Hypothesis*: $r \geq 0.3$)	ICECAP-O [67], ASCOT [67]: **0.01–0.09;** AQoL [49]: 0.48 (*n≈219* (*n.c.r.* [49]))	± (33%)	high
General health/QoL (single-scale) (*Hypothesis*: $r \geq 0.3$)	Health GRS [67], QoL GRS [67]: 0.12–0.14 (*n = 149*)	− (0%)	high

Table 3 (continued)

Measurement property	Summary	Overall rating	Quality of evidence
ADL (*Hypothesis*: $r \geq 0.3$)	Barthel [54], Katz [67]: **0.04–0.19** (*n = 484*)	− (0%)	moderate
IADL (*Hypothesis*: $r \geq 0.1$)	NOSGER [54]: **0.01** (*n = 336*)	− (0%)	high
Cognitive status/dementia severity (*Hypothesis*: $r < 0.3$)	MMSE [54]: 0.00 (*n = 369*)	+ (100%)	low
Other instruments	*n = 371*	± (50%)	N/A
(*Hypothesis*: $r \geq 0.3$)	DASH [59]: 0.47		
(*Hypothesis*: $r \geq 0.1$)	Pearlin Mastery Scale [67], Activity inventory [56]: **0.02–0.06**		
(*Hypothesis*: $r < 0.3$)	CCCQ [67]: 0.09		
Comparison between subgroups	*n≈1,711* (*n.c.r.* [54])	**+ (79%)**	**moderate**
Supported for groups of…	Improvement/worsening on the Barthel index [41], knee replacement vs. femur fracture [43], femur fracture vs. stroke [43], fallers vs. non-fallers [45], complication vs. non-complication after femoral neck fracture [46], deterioration in health status (CGI-I) [54], less good vs. good outcome after femoral neck fracture (pain and/or needing walking aids) [65], perceived health change and healing status in people with venous leg ulcers [68], complications/non-complications after femoral neck fracture [46], improvement/deterioration status (DASH) after proximal humeral fracture [59], death/non-death after hip fracture [61], displaced/undisplaced femoral neck fractures [66]		
Rejected for groups of…	Improvement/worsening on the EQ-VAS [41], hip replacement vs. femur fracture [43]; healed vs. non-healed leg ulcers at 3 months follow-up [68], revision after hip fracture [61]		
Before and after intervention			
Supported for…	Deterioration/improvement of HrQoL over time after hip or proximal humeral fracture [59, 61] (*n = 340*)	+ (100%)	high
Rejected for …	Low-vision rehabilitation [56] (*n = 77*)	– (0%)	moderate

Unless otherwise indicated, reported numbers refer to absolute correlation coefficients, correlation coefficients printed in **bold** indicate results for which the hypotheses were rejected

y/n yes/no, *n.s.* not significant, *N/A* not applicable, *N/R* not reported, *r* correlation coefficient, *β* regression coefficient, *n* sample size, *n.c.r.* not clearly reported, *ADL* activities of daily living, *ASCOT* adult social care outcomes toolkit, *AQoL* assessment of quality of life, *CCCQ* client-centered care questionnaire, *CGI-I* clinical global impression of improvement, *CSDD* Cornell Scale for depression in dementia, *DASH* disabilities of arm, shoulder, and hand, *EQ-VAS* Visual Analogue Scale, *GDS* Geriatric Depression Scale, *GRS* Global Rating Scale, *HrQoL* health-related quality of life, *HUI3* Health Utilities Index, *IADL* instrumental activities of daily living, *ICC* intraclass correlation coefficient, *ICECAP-O* ICEpop CAPability measure for older people, *MCS* mental health component summary, *MMSE* mini-mental state examination, *NHP* Nottingham Health Profile, *NOSGER* Nurses' Observation Scale for Geriatric Patients, *OHS* oxford hip score, *OPQOL-Brief* older people's quality-of-life brief questionnaire, *PCS* physical health component summary, *PPA* physiological profile assessment, *QoL* quality of life, *QoL-AD* qualityof life in Alzheimer's diseases, *QoL GRS* Quality-of-Life Global Rating Scale, *QWB* quality of well-being, *RAID* Rating of Anxiety in Dementia Scale, *SF-36* 36-item short-form health survey, *SF-12* 12-item short-form health survey, *SF-6D* six-dimensional short form, *SPPB* short physical performance battery, *VAS Pain* visual analogue scale for pain

[a]deviating hypothesis: $r \geq 0.1$

[b]no relevant difference between groups hypothesized

[c]no calculation of correlation, instead comparison of EQ-5D & 15D in terms of proportions of changes stratified according to the minimally important difference values

Table 4 Summary of findings—EQ-5D-5L

Measurement property	Summary or pooled results	Overall rating	Quality of evidence
Reliability	N/R		
Construct validity			
Convergent validity		**+(84%)**	**High**
HrQoL instruments (*Hypothesis*: $r \geq 0.5$)	SF-6D: 0.71 [70], *ICC* = 0.61 [75] (*n≈1193* (*n.c.r.* [75]))	+(100%)	High
QoL instruments (*Hypothesis*: $r \geq 0.3$)	DEMQOL [69], DEMQOL-U [72], QOL-AD [69, 73]: 0.30–0.48 QOL-AD-NH [74]: **0.28;** SPVU-5D [71]: *ICC* = 0.55 (*n≈1417* (*n.c.r.* [71]))	+(83%)	High
General health/QoL (single scale) (*Hypothesis*: $r \geq 0.3$)	EQ-VAS [71]: 0.39 (*n≈75* (*n.c.r.*))	+(100%)	Moderate
ADL (*Hypothesis*: $r \geq 0.3$)	MBI [72, 76]: 0.46–0.49 (*n* = *225*)	+(100%)	High
Cognitive status (*Hypothesis*: $r < 0.3$)	Pas-Cog* [72]: **0.24**; MMSE [76]: 0.22; CDR [74]: 0.025 (*n* = *1116*)	±(67%)	N/A
Other instruments	*n* = *1113*	+(80%)	High
(*Hypothesis*: $r \geq 0.1$)	CSDD [76], PainAd [76]: 0.33–0.45; FAST [74]: **0.049**		
(*Hypothesis*: $r < 0.3$)	CMAI [74], NPI-Q [72]: 0.1		
Known-groups validity	*n≈306* (*n.c.r.* [71])	**+(78%)**	**High**
Supported for Groups of…	Age[a] [71], general health (EQ-VAS) [71], leg ulcer healing status [71], physical functioning/ADL (MBI) [72, 76], pain (PainAd) [76], depression (CSSD) [76]		
Rejected for groups of…	cognitive impairment (PAS-Cog)* [72], ulcer duration [71]		
Responsiveness			
Construct approach		**+(75%)**	**High**
QoL instruments (*Hypothesis*: $r \geq 0.3$)	QOL-AD-NH [74]: $\boldsymbol{\beta \approx 0.007}$ ($p < 0.05$) (*n≈261*(*n.c.r.*))	− (0%)	Moderate
Cognitive status (*Hypothesis*: $r < 0.3$)	CDR [74]: β = n.s (*n≈261*(*n.c.r.*))	+(100%)	High
Other instruments	*n≈396* (*n.c.r.* [74])	+(83%)	High
(*Hypothesis*: $r \geq 0.3$)	BBS [70]: Elasticity = 0.54		
(*Hypothesis*: $r \geq 0.1$)	30 s STS [70], 4 m walk test [70], FES-I [70]: Elasticity = 0.09–0.24; FAST [74]: $\boldsymbol{\beta}$ **= n.s**		
(*Hypothesis*: $r < 0.3$)	CMAI [74]: β = n.s		
Comparison between subgroups	*n* = *269*	**+(75%)**	**High**
Supported for groups of…	Improvement/worsening on the Barthel index [41], healing status and duration of venous leg ulcers [71]		
Rejected for groups of…	Improvement/worsening on the EQ-VAS [41]		
Before and after intervention	N/R		

+ sufficient, − insufficient, ± inconsistent, *y/n* yes/no, *n.s.* not significant, *r* correlation coefficient, *β* regression coefficient, *n* sample size, *n.c.r.* not clearly reported, *N/R* not reported, *BBS* Berg Balance Scale, *CDR* clinical dementia rating, *CMAI* Cohen-Mansfield Agitation Inventory, *CSDD* Cornell Scale for Depression in Dementia, *DEMQOL* dementia quality of life, *EQ-VAS* Visual Analog Scale, *FAST* functional assessment staging, *FES-I* Falls Efficacy Scale International, *ICC* intraclass correlation coefficient, *MBI* Modified Barthel Index, *MMSE* mini-mental state examination, *PainAd* Pain Assessment in Advanced Dementia Scale, *PAS-Cog* Psychogeriatric Assessment Scales-Cognitive Impairment Scale, *QoL* quality of life, *QoL-AD* quality of life in Alzheimer's disease, *QOL-AD-NH* quality of life in Alzheimer's disease nursing home version, *SF-6D* six-dimensional short-form health survey, *30 s STS* 30-second sit-to-stand test, *SPVU-5D* five-dimensional sheffield-preference-based venous ulcer questionnaire

*result in the opposite of the hypothesized direction (H9)

[a]no relevant difference between groups hypothesized

moderate-quality evidence for sufficient responsiveness of the EQ-5D-3L based on comparisons between subgroups was found, as 79% of the hypotheses were supported.

Three studies [56, 59, 61] examined responsiveness by testing hypotheses regarding change in the EQ-5D-3L index in response to an intervention. Two hypotheses regarding the improvement or deterioration of HrQoL after fracture were supported, whereas, opposed to the hypothesis, low vision rehabilitation did not change HrQoL.

For the EQ-5D-5L index, two studies [70, 74] assessed responsiveness based on comparisons with other instruments. 75% of the results were in accordance with the hypotheses and, thus, were rated as sufficient at high quality of evidence. The correlations of change scores were as high (or low) as hypothesized between the EQ-5D-5L and measures of cognitive status or agitation (CDR, CMAI) [74], measures of physical function (BBS, 30 s STS, 4 m walk test) [70] but were lower than hypothesized between the EQ-5D-5L and a QoL instrument (QOL-AD-NH) [74] or a measure of functional symptoms in dementia (FAST) [74].

Two studies examined responsiveness of the EQ-5D-5L index in terms of subgroup comparisons [41, 71]. 75% of the hypotheses were supported and, thus, the overall result was sufficient. The quality of evidence was rated high.

Results not included in the qualitative synthesis

Some results were not included in the qualitative synthesis as no specific results (e.g., correlation coefficients) were reported. Regarding convergent validity, Michalowsky et al. [57] found a poor association (not further specified) between the EQ-5D-3L index and IADL. Other authors examined the association between the EQ-5D dimensions with ADL and found significant associations between several dimensions but did not provide information about the strength of the association [39, 43]. Moreover, the authors assessed known-groups validity and found, e.g., that women were more anxious than men [39] and that people with disability had lower HrQoL than people with no disability [43]. However, it could not be evaluated whether the differences were clinically important because the mean EQ-5D of each group was not reported.

Discussion

The current study synthesized reliability, validity, and responsiveness of the EQ-5D in a population of middle-old and oldest-old people. Regarding reliability, results were inconsistent for the EQ-5D-3L, and for the EQ-5D-5L, studies were entirely lacking. This may pose a problem in contexts where the EQ-5D is used at different time points to quantify a ‘true’ difference or change in HrQoL, such as in economic evaluations. Previous reviews report mixed results on the reliability of the EQ-5D in people with dementia (moderate to strong) [16] and sufficient reliability in people with diabetes or stroke [77, 78]. Another review further suggests sufficient reliability of the EQ-5D-5L in various patient groups (e.g., osteoarthritis, diabetes and cancer patients, cardiovascular and liver diseases) and general population samples [79]. However, so far, the evidence on reliability for both the EQ-5D-3L and EQ-5D-5L is relatively limited and entirely lacking for certain patient groups.

For both EQ-5D versions, high-quality evidence of sufficient convergent validity was found. It should be noted that high correlations with other generic instruments (e.g., SF-36/-12, SF-6D, HUI3) do not necessarily support the use of the EQ-5D in middle-old to oldest-old people, as it does not preclude that both instruments do not capture aspects that are important to the population of interest. In some cases, convergent validity was assessed by correlations with instruments which were collected only in a single, specific study (e.g., OHS, Pearlin Mastery Scale). These results summarized as “other instruments” despite measuring different constructs in Table 3 and 4, may not be generally relevant for the population aged 75+ but were mostly in accordance with the hypotheses.

Known-groups validity of the EQ-5D-3L was inconsistent. One potential explanation could be a ceiling effect of the EQ-5D-3L, which may have compromised its ability to discriminate between known groups. Moreover, it can be questioned whether the groups for evaluating known-groups validity are relevant (e.g., marital status, living alone vs. not alone). Similarly, it could be questioned whether it is reasonable to examine, e.g., convergent validity of the EQ-5D with instruments measuring constructs which are hardly related to HrQoL (e.g., CCCQ, PPA, SPPB). The evaluation of measurement properties should be theory driven and not exploratory by using all available variables from studies that were initially designed for a different purpose. More precise preliminary hypotheses of associations between measures in studies analyzing an instrument’s measurement properties would, therefore, be desirable. In addition, rather “soft” hypotheses regarding the strength of the association between two instruments were defined in this review, e.g., by not setting an upper limit for correlations between instruments measuring related but dissimilar constructs ($r \geq 0.3$) or weakly related constructs ($r \geq 0.1$). This was done to avoid “penalizing” relatively strong correlations between instruments that were assumed to be not necessarily but potentially highly correlated (e.g., EQ-5D and ADL instruments). Since, according to the COSMIN methodology, the synthesized evaluation of a measurement property is based on a majority principle (≥ 75% of the hypotheses supported), these aspects could have influenced the (synthesized) results. For the EQ-5D-5L, high-quality evidence of sufficient known-groups validity was found. There, the selection of groups that the EQ-5D was expected to differentiate between seemed to be less arbitrary, but overall, the results were based on only three studies. The COSMIN methodology recommends judging an instrument’s ability to discriminate between relevant groups based on clinically important rather than statistically significant differences [34]. While being aware that there is no single MCID for EQ-5D index values since it varies by population characteristics [80], in the absence of specific MCIDs for each country-specific tariff and disease group of the individual studies included in this review, MCIDs commonly used in previous literature were nevertheless used but could have influenced the results regarding known-groups validity.

Responsiveness was insufficient (high-quality evidence) for the EQ-5D-3L when correlated with instruments being hypothesized to be related (e.g., other (Hr)QoL instruments). However, it seemed to be responsive to outcomes after fracture or healing status of leg ulcers [43, 46, 59, 61, 65, 66, 68]. These are conditions with substantial changes in health, where the EQ-5D has previously been shown to be more likely to be responsive (in an older population) [12, 18]. Although responsiveness of the EQ-5D-5L (construct approach) was found sufficient according to the majority principle of the COSMIN methodology, the evidence was limited as it was based on only two studies which used very study-specific instruments to evaluate responsiveness (e.g., 30 s STS) [70, 74]. These instruments were hypothesized to be only weakly associated with the EQ-5D and were, therefore, not responsive to changes in HrQoL.

Overall, the results regarding the responsiveness of the EQ-5D suggest that at least the EQ-5D-3L is hardly able to adequately reflect clinical changes over time. In turn, clinically relevant changes may remain undetected; thus, intervention effects may be underestimated based on the EQ-5D. For example, economic evaluations of fall prevention programs showed that clinical effects could not be found on HrQoL [81–83]. This does not seem to be an exclusive problem of the EQ-5D but also of other generic HrQoL instruments, such as the SF-36 or SF-12 [82, 83]. So far, the evidence on responsiveness of the EQ-5D is mainly based on studies using the EQ-5D-3L. The sparse evidence on the responsiveness of the EQ-5D-5L is not limited to the population of middle old to oldest old but is also found in general for other populations [79]. Moreover, the majority of the included studies reported substantial ceiling effects, which may limit the ability to capture small changes at the upper end of HrQoL. Ceiling effects were found to be particularly common among people with dementia [15], who make up a large proportion in the current study. Generally, the EQ-5D-5L was found to reduce this ceiling effect [84, 85]. However, it persists in general population studies but also in some patient populations [79]. Further studies are needed, which evaluate the responsiveness of the EQ-5D-5L to change in, e.g., other (age or disease specific) (Hr)QoL instruments. It would be of particular interest to examine whether the EQ-5D-5L is more responsive than the EQ-5D-3L which was insufficiently responsive in this respect.

The approach to primarily focus on HrQoL in the form of health utility gains in economic evaluations has been criticized for excluding aspects of QoL beyond health [23, 86]. Furthermore, HrQoL instruments such as the EQ-5D or the SF-12/SF-36 are mainly functioning oriented and, thus, do not reflect the breadth of the concept of health as stated in the WHO definition [21], e.g., social aspects of health fall short or are not assessed differentiated enough. This seems to be especially relevant to older people as it was found that not only health but also social domains are important to their overall QoL [23, 87]. Therefore, other instruments were and are currently being developed, which may provide an alternative or complement to measure (Hr)QoL based on a broader or more comprehensive framework of health or well-being in the future. Some age- or disease-specific QoL instruments exist, and the current study showed that although being moderately to strongly associated with the EQ-5D when assessed at a single time point (sufficient convergent validity), changes on these instruments are not reflected on the EQ-5D (insufficient responsiveness). This suggests that the EQ-5D is not able to capture changes in (Hr)QoL that are important to older people. However, the existing age- or disease-specific instruments differ in domains of (Hr)QoL that are captured [6] and, thus, pose a problem for the comparability of intervention effects across diseases and populations. Moreover, the lack of preference-based value sets for some of these instruments (e.g., for the WHOQOL-OLD, an older people-specific QoL instrument [87]) or value sets being only available for the population in the country where the instruments were developed, impedes their use in economic evaluations. Another recently developed instrument is the PROMIS-29, a health profile measure from the Patient-Reported Outcomes Measurement Information System® (PROMIS®) [88–90] that captures health in a broader sense than the EQ-5D. Although value sets are available for the PROMIS-29 [89–92], they are so far only available for the US. Moreover, the 'Extending the QALY' research project is currently developing the EQ-HWB, a broad measure of QoL for use in economic evaluations across health and social care (https://scharr.dept.shef.ac.uk/e-qaly/), and thus, could be a potential alternative to the EQ-5D in the future. However, these age-unspecific instruments carry the risk that scoring algorithms used to derive the utility index are based on the preferences of the general adult populations, whose preferences for health may differ from those of older people [6, 24]. Another research group is seeking to address this issue and is currently developing an instrument for quality assessment and economic evaluation that adequately captures the aspects of quality of life that are important to older people, using a person-centered approach [93, 94]. Consequently, as long as there is no single preference-based generic instrument that comprehensively captures relevant aspects of (Hr)QoL in middle-old and oldest-old people or its use is limited in certain situations (e.g., lack of country/population-specific tariffs), age- or disease-specific instruments should be used as complement to the EQ-5D and help interpreting the results of (cost-)effectiveness analyses (e.g., whether the effects of an intervention are likely to be underestimated).

Beyond these alternative instruments, several "bolt-on" dimensions to the EQ-5D have been proposed and a wide variety of methods have been applied to identify or select relevant bolt-on dimensions [95]. Finch, Brazier, Mukuria, and Bjorner [96] identified hearing, sleep, cognition, energy, and relationships as potentially relevant bolt-on dimensions, and some studies have shown that higher severity levels in the bolt-on dimensions impact the health state values or preferences for the health state [97–99]. Recently, Chen and Olsen [100] proposed vitality, sleep, social relationships, and community connectedness as bolt-on dimensions. They argue that adding these four dimensions would provide a solution to assess HrQoL in a single, brief instrument, but still include all key dimensions of the conceptual map of HrQoL by Olsen and Misajon, [21] and, thus, capture health and well-being more broadly than current EQ-5D instruments. However, to use the additional information from the bolt-on dimensions in economic evaluations, the bolt-on dimension scores would need to be incorporated into the utility index, which would require new valuation studies. Moreover, extensive testing on whether the bolt-on dimensions improve psychometric performance of the EQ-5D would be needed, in general, but also particularly in middle-old and oldest-old people.

A large number of the included studies ($n = 13$) assessed the measurement properties of the EQ-5D in people with dementia or cognitive impairment. As part of the validation, the association between (change in) cognitive status and (change in) the EQ-5D was examined [44, 49, 52, 54, 58, 64, 72, 74, 76]. However, the relationship between cognition and (Hr)QoL seems to be complex [101, 102], which made it difficult to formulate (generic) hypotheses regarding the direction and strength of the association in this study.

This review deliberately did not focus on the comparison of self- and proxy-rated EQ-5D scores and did not consider correlations between the self-rated EQ-5D and proxy-rated other (Hr)QoL instruments in the synthesis. (Hr)QoL is a subjective concept; therefore, it is not surprising that different people evaluate it differently, especially when self-perception is impaired by a condition such as dementia, where proxies typically rate the HrQoL of a person with dementia lower than the person him/herself [15, 16]. It is not possible to determine whose rating is more "correct." However, it is important to be aware of these variations and to select the administration mode depending on the perspective from which the benefits of an intervention are to be measured.

This study applied the updated COSMIN methodology to systematically review the measurement properties of the EQ-5D in a middle-old and oldest-old population. However, several limitations must be acknowledged. First, only studies which directly aimed to examine the measurement properties of the EQ-5D were included, whereas studies providing indirect evidence on measurement properties (e.g., by correlating the EQ-5D with instruments being hypothetically related) were not included. Second, the generalizability of the results may be limited: although this study was deliberately not restricted to specific populations such as disease groups, it is not clear, whether the results apply to the general population of middle-old to oldest-old adults as, e.g., a large share of the included studies included only people with dementia. Moreover, the results do not exclusively apply to the population aged 75+ as a number of persons < 75 years are also included in some of the studies. To date, there have been few studies focusing exclusively on the population aged 75 years and older, representing a gap in research. Such studies could allow a comparison between the measurement properties of the EQ-5D between younger-old (e.g., aged 60+) and middle-old to oldest-old people, which was not directly possible based on the current data. Finally, the evidence stems exclusively from western, industrialized countries and, therefore, may not be transferable to other countries or regions.

Conclusion

The results of this systematic review are relevant as improving the care and maintaining the health and QoL of an older population is a political goal in many countries. Thereby, the results may be of interest to decision makers, but also to researchers planning, designing, or evaluating interventions for older people.

Based on the findings of this study, both EQ-5D versions seem to have sufficient convergent validity and may, therefore, be used in cross-sectional studies to assess HrQoL. However, caution is advised when using the EQ-5D to assess change in HrQoL, as the EQ-5D-3L was found to be insufficiently responsive to change (except for conditions with substantial changes in health) and results regarding the reliability were inconsistent. As specifically for the EQ-5D-5L little evidence on reliability and responsiveness is available so far, further research might be needed in this regard. If responsiveness cannot be demonstrated, either using additional disease- or age-specific instruments or considering the use of an alternative, more comprehensive instrument of (Hr)QoL might be advisable, especially for economic evaluations. Promising research is currently underway to develop new, more comprehensive instruments that will better capture the aspects of QoL that are important to older people. However, there is still a long way to go to verify their measurement properties, generate population- and country-specific value sets, and thus, be broadly applicable to economic evaluations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11136-022-03185-0.

Author contributions The study concept was developed by SG, JD, and HHK. The search strategy was developed by SG and JD. Study selection, data extraction, and quality assessment were performed by SG and MN, with JD as a third party in case of disagreements. The manuscript was drafted by SG and critically revised by JD, HHK, and MN. All authors have approved the final version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability All data generated or analyzed during this study are included in this published article and its supplementary information files.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. United Nations, Department of Economic and Social Affairs, & Population Division. (2019). *World population prospects 2019* (Vol. 2). Demographic Profiles.
2. Janssen, M. F., Szende, A., Cabases, J., Ramos-Goñi, J. M., Vilagut, G., & König, H. H. (2019). Population norms for the EQ-5D-3L: A cross-country analysis of population surveys for 20 countries. *The European Journal of Health Economics, 20*(2), 205–216. https://doi.org/10.1007/s10198-018-0955-5
3. Marten, O., & Greiner, W. (2021). EQ-5D-5L reference values for the German general elderly population. *Health and Quality of Life Outcomes, 19*(1), 76. https://doi.org/10.1186/s12955-021-01719-7
4. EuroQol Group. (1990). EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199–208. https://doi.org/10.1016/0168-8510(90)90421-9
5. Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., Busschbach, J., & Boye, K. S. (2020). Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *The European Journal of Health Economics, 21*(8), 1245–1257. https://doi.org/10.1007/s10198-020-01195-8
6. Cleland, J., Hutchinson, C., Khadka, J., Milte, R., & Ratcliffe, J. (2019). A review of the development and application of generic preference-based instruments with the older population. *Applied Health Economics and Health Policy, 17*(6), 781–801. https://doi.org/10.1007/s40258-019-00512-4
7. National Institute for Health and Care Excellence. (2013). Guide to the methods of technology appraisal 2013 [Internet]. *Process and Methods Guides No9*. National Institute for Health and Care Excellence.
8. Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., Jones, M. L., Paisley, S., O'Cathain, A., Barkham, M., Knapp, M., Byford, S., Gilbody, S., & Parry, G. (2014). A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technology Assessment, 18*(34), 1–188. https://doi.org/10.3310/hta18340
9. Mulhern, B., Mukuria, C., Barkham, M., Knapp, M., Byford, S., Soeteman, D. R., & Brazier, J. (2014). Using generic preference-based measures in mental health: Psychometric validity of the EQ-5D and SF-6D. *British Journal of Psychiatry, 205*(3), 236–243. https://doi.org/10.1192/bjp.bp.112.122283
10. Davis, S., & Wailoo, A. (2013). A review of the psychometric performance of the EQ-5D in people with urinary incontinence. *Health and Quality of Life Outcomes, 11*(1), 20. https://doi.org/10.1186/1477-7525-11-20
11. Yang, Y., Brazier, J., & Longworth, L. (2015). EQ-5D in skin conditions: An assessment of validity and responsiveness. *The European Journal of Health Economics, 16*(9), 927–939. https://doi.org/10.1007/s10198-014-0638-9
12. Haywood, K. L., Garratt, A. M., & Fitzpatrick, R. (2005). Quality of life in older people: A structured review of generic self-assessed health instruments. *Quality of Life Research, 14*(7), 1651–1668. https://doi.org/10.1007/s11136-005-1743-0
13. Kuspinar, A., & Mayo, N. E. (2014). A review of the psychometric properties of generic utility measures in multiple sclerosis. *PharmacoEconomics, 32*(8), 759–773. https://doi.org/10.1007/s40273-014-0167-5
14. Tordrup, D., Mossman, J., & Kanavos, P. (2014). Responsiveness of the EQ-5D to clinical change: Is the patient experience adequately represented? *International Journal of Technology Assessment in Health Care, 30*(1), 10–19. https://doi.org/10.1017/s0266462313000640
15. Hounsome, N., Orrell, M., & Edwards, R. T. (2011). EQ-5D as a quality of life measure in people with dementia and their carers: Evidence and key issues. *Value in Health, 14*(2), 390–399. https://doi.org/10.1016/j.jval.2010.08.002
16. Li, L., Nguyen, K. H., Comans, T., & Scuffham, P. (2018). Utility-based instruments for people with dementia: A systematic review and meta-regression analysis. *Value in Health, 21*(4), 471–481. https://doi.org/10.1016/j.jval.2017.09.005
17. Finch, A. P., Brazier, J. E., & Mukuria, C. (2018). What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *The European Journal of Health Economics, 19*(4), 557–570. https://doi.org/10.1007/s10198-017-0902-x

18. Payakachat, N., Ali, M. M., & Tilford, J. M. (2015). Can the EQ-5D detect meaningful change? A systematic review. *PharmacoEconomics, 33*(11), 1137–1154. https://doi.org/10.1007/s40273-015-0295-6
19. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics, 13*(9), 873–884. https://doi.org/10.1002/hec.866
20. Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn, P., & Busschbach, J. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research, 22*(7), 1717–1727. https://doi.org/10.1007/s11136-012-0322-4
21. Olsen, J. A., & Misajon, R. (2020). A conceptual map of health-related quality of life dimensions: Key lessons for a new instrument. *Quality of Life Research, 29*(3), 733–743. https://doi.org/10.1007/s11136-019-02341-3
22. Sutton, E. J., & Coast, J. (2014). Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. *Palliative Medicine, 28*(2), 151–157. https://doi.org/10.1177/0269216313489368
23. Milte, C. M., Walker, R., Luszcz, M. A., Lancsar, E., Kaambwa, B., & Ratcliffe, J. (2014). How important is health status in defining quality of life for older people? An exploratory study of the views of older South Australians. *Applied Health Economics and Health Policy, 12*(1), 73–84. https://doi.org/10.1007/s40258-013-0068-3
24. Ratcliffe, J., Lancsar, E., Flint, T., Kaambwa, B., Walker, R., Lewin, G., Luszcz, M., & Cameron, I. D. (2017). Does one size fit all? Assessing the preferences of older and younger people for attributes of quality of life. *Quality of Life Research, 26*(2), 299–309. https://doi.org/10.1007/s11136-016-1391-6
25. Marten, O., Brand, L., & Greiner, W. (2021). Feasibility of the EQ-5D in the elderly population: A systematic review of the literature. *Quality of Life Research*. https://doi.org/10.1007/s11136-021-03007-9
26. Bulamu, N. B., Kaambwa, B., & Ratcliffe, J. (2015). A systematic review of instruments for measuring outcomes in economic evaluation within aged care. *Health and Quality of Life Outcomes, 13*(1), 179. https://doi.org/10.1186/s12955-015-0372-8
27. Makai, P., Brouwer, W. B. F., Koopmanschap, M. A., Stolk, E. A., & Nieboer, A. P. (2014). Quality of life instruments for economic evaluations in health and social care for older people: A systematic review. *Social Science and Medicine, 102*, 83–93. https://doi.org/10.1016/j.socscimed.2013.11.050
28. Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3
29. Gottschalk, S., König, H. H., Nejad, M., & Dams, J. (2020). Psychometric properties of the EQ-5D for the assessment of health-related quality of life in the population of middle-old and oldest-old persons: Study protocol for a systematic review. *Frontiers in Public Health, 8*, 578073. https://doi.org/10.3389/fpubh.2020.578073
30. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The, P. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
31. Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research, 18*(8), 1115–1123. https://doi.org/10.1007/s11136-009-9528-5
32. Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1171–1179. https://doi.org/10.1007/s11136-017-1765-4
33. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012
34. Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2018). COSMIN methodology for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147–1157.
35. Walters, S. J., & Brazier, J. E. (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research, 14*(6), 1523–1532. https://doi.org/10.1007/s11136-004-7713-0
36. McClure, N. S., Sayah, F. A., Xie, F., Luo, N., & Johnson, J. A. (2017). Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value in Health, 20*(4), 644–650. https://doi.org/10.1016/j.jval.2016.11.015
37. Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set"—a practical guideline. *Trials, 17*(1), 449. https://doi.org/10.1186/s13063-016-1555-2
38. GRADE Handbook (2013). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Retrieved February18, 2021, from https://gdt.gradepro.org/app/handbook/handbook.html
39. Ankri, J., Beaufils, B., Novella, J. L., Morrone, I., Guillemin, F., Jolly, D., Ploton, L., & Blanchard, F. (2003). Use of the EQ-5D among patients suffering from dementia. *Journal of clinical epidemiology, 56*(11), 1055–1063. https://doi.org/10.1016/s0895-4356(03)00175-6
40. Barton, G. R., Sach, T. H., Avery, A. J., Jenkinson, C., Doherty, M., Whynes, D. K., & Muir, K. R. (2008). A comparison of the performance of the EQ-5D and SF-6D for individuals aged > or = 45 years. *Health Economics, 17*(7), 815–832. https://doi.org/10.1002/hec.1298
41. Bhadhuri, A., Kind, P., Salari, P., Jungo, K. T., Boland, B., Byrne, S., Hossmann, S., Dalleur, O., Knol, W., Moutzouri, E., O'Mahony, D., Murphy, K. D., Wisselink, L., Rodondi, N., & Schwenkglenks, M. (2020). Measurement properties of EQ-5D-3L and EQ-5D-5L in recording self-reported health status in older patients with substantial multimorbidity and polypharmacy. *Health and Quality of Life Outcomes, 18*(1), 317. https://doi.org/10.1186/s12955-020-01564-0
42. Brazier, J. E., Walters, S. J., Nicholl, J. P., & Kohler, B. (1996). Using the SF-36 and EuroQol on an elderly population. *Quality of Life Research, 5*(2), 195–204. https://doi.org/10.1007/bf00434741
43. Coast, J., Peters, T. J., Richards, S. H., & Gunnell, D. J. (1998). Use of the EuroQoL among elderly acute care patients. *Quality of Life Research, 7*(1), 1–10. https://doi.org/10.1023/a:1008857203434
44. Davis, J. C., Bryan, S., McLeod, R., Rogers, J., Khan, K., & Liu-Ambrose, T. (2012). Exploration of the association between quality of life, assessed by the EQ-5D and ICECAP-O, and falls risk, cognitive function and daily function, in older adults with mobility impairments. *BMC Geriatrics, 12*, 65. https://doi.org/10.1186/1471-2318-12-65
45. Davis, J. C., Best, J. R., Dian, L., Khan, K. M., Hsu, C. L., Chan, W., Cheung, W., & Liu-Ambrose, T. (2017). Are the EQ-5D-3L and the ICECAP-O responsive among older adults with impaired

mobility? Evidence from the Vancouver falls prevention cohort study. *Quality of Life Research, 26*(3), 737–747. https://doi.org/10.1007/s11136-016-1487-z
46. Frihagen, F., Grotle, M., Madsen, J. E., Wyller, T. B., Mowinckel, P., & Nordsletten, L. (2008). Outcome after femoral neck fractures: A comparison of Harris hip score, Eq-5d and Barthel index. *Injury, 39*(10), 1147–1156. https://doi.org/10.1016/j.injury.2008.03.027
47. Hazell, M., Frank, T., & Frank, P. (2003). Health related quality of life in individuals with asthma related symptoms. *Respiratory Medicine, 97*(11), 1211–1218. https://doi.org/10.1016/S0954-6111(03)00249-X
48. Heiskanen, J., Tolppanen, A.-M., Roine, R. P., Hartikainen, J., Hippeläinen, M., Miettinen, H., & Martikainen, J. (2016). Comparison of EQ-5D and 15D instruments for assessing the health-related quality of life in cardiac surgery patients. *European Heart Journal, 2*(3), 193–200. https://doi.org/10.1093/ehjqcco/qcw002
49. Holland, R., Smith, R. D., Harvey, I., Swift, L., & Lenaghan, E. (2004). Assessing quality of life in the elderly: A direct comparison of the EQ-5D and AQoL. *Health Economics, 13*(8), 793–805. https://doi.org/10.1002/hec.858
50. Jönsson, L., Andreasen, N., Kilander, L., Soininen, H., Waldemar, G., Nygaard, H., Winblad, B., Jönhagen, M. E., Hallikainen, M., & Wimo, A. (2006). Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL. *Alzheimer Disease and Associated Disorders, 20*(1), 49–55. https://doi.org/10.1097/01.wad.0000201851.52707.c9
51. Kaambwa, B., Gill, L., McCaffrey, N., Lancsar, E., Cameron, I. D., Crotty, M., Gray, L., & Ratcliffe, J. (2015). An empirical comparison of the OPQoL-Brief, EQ-5D-3 L and ASCOT in a community dwelling population of older people. *Health and Quality of Life Outcomes, 13*(1), 164. https://doi.org/10.1186/s12955-015-0357-7
52. Karlawish, J. H., Zbrozek, A., Kinosian, B., Gregory, A., Ferguson, A., & Glick, H. A. (2008). Preference-based quality of life in patients with Alzheimer's disease. *Alzheimer's & Dementia, 4*(3), 193–202. https://doi.org/10.1016/j.jalz.2007.11.019
53. Kim, S.-K., Kim, K.-H., Kim, S.-H., Yoo, S.-J., & Jeong, Y.-W. (2019). Health-related quality of life in adult males with lower urinary tract symptoms. *Quality of Life Research, 28*(9), 2419–2428. https://doi.org/10.1007/s11136-019-02205-w
54. Kunz, S. (2010). Psychometric properties of the EQ-5D in a study of people with mild to moderate dementia. *Quality of Life Research, 19*(3), 425–434. https://doi.org/10.1007/s11136-010-9600-1
55. Lutomski, J. E., Krabbe, P. F., Bleijenberg, N., Blom, J., Kempen, G. I., MacNeil-Vroomen, J., Muntinga, M. E., Steyerburg, E., Olde-Rikkert, M. G., & Melis, R. J. (2017). Measurement properties of the EQ-5D across four major geriatric conditions: Findings from TOPICS-MDS. *Health and Quality of Life Outcomes, 15*(1), 45. https://doi.org/10.1186/s12955-017-0616-x
56. Malkin, A. G., Goldstein, J. E., Perlmutter, M. S., & Massof, R. W. (2013). Responsiveness of the EQ-5D to the effects of low vision rehabilitation. *Optometry and vision science, 90*(8), 799–805. https://doi.org/10.1097/opx.0000000000000005
57. Michalowsky, B., Xie, F., Kohlmann, T., Gräske, J., Wübbeler, M., Thyrian, J. R., & Hoffmann, W. (2020). Acceptability and validity of the EQ-5D in patients living with dementia. *Value in Health, 23*(6), 760–767. https://doi.org/10.1016/j.jval.2020.01.022
58. Naglie, G., Tomlinson, G., Tansey, C., Irvine, J., Ritvo, P., Black, S. E., Freedman, M., Silberfeld, M., & Krahn, M. (2006). Utility-based quality of life measures in Alzheimer's disease. *Quality of Life Research, 15*(4), 631–643. https://doi.org/10.1007/s11136-005-4364-8
59. Olerud, P., Tidermark, J., Ponzer, S., Ahrengart, L., & Bergström, G. (2011). Responsiveness of the EQ-5D in patients with proximal humeral fractures. *Journal of Shoulder and Elbow Surgery, 20*(8), 1200–1206. https://doi.org/10.1016/j.jse.2011.06.010
60. Orgeta, V., Edwards, R. T., Hounsome, B., Orrell, M., & Woods, B. (2015). The use of the EQ-5D as a measure of health-related quality of life in people with dementia and their carers. *Quality of Life Research, 24*(2), 315–324. https://doi.org/10.1007/s11136-014-0770-0
61. Parsons, N., Griffin, X. L., Achten, J., & Costa, M. L. (2014). Outcome assessment after hip fracture: Is EQ-5D the answer? *Bone Joint Res, 3*(3), 69–75. https://doi.org/10.1302/2046-3758.33.2000250
62. Pérez-Ros, P., & Martínez-Arnau, F. M. (2020). EQ-5D-3L for assessing quality of life in older nursing home residents with cognitive impairment. *Life, 10*(7), 100.
63. Pérez-Ros, P., Vila-Candel, R., Martin-Utrilla, S., & Martínez-Arnau, F. M. (2020). Health-related quality of life in community-dwelling older people with cognitive impairment: EQ-5D-3L measurement properties. *Journal of Alzheimer's Disease, 77*(4), 1523–1532. https://doi.org/10.3233/jad-200806
64. Sanchez-Arenas, R., Vargas-Alarcon, G., Sanchez-Garcia, S., Garcia-Peña, C., Gutierrez-Gutierrez, L., Grijalva, I., Garcia-Dominguez, A., & Juárez-Cedillo, T. (2014). Value of EQ-5D in Mexican city older population with and without dementia (SADEM study). *International Journal of Geriatric Psychiatry, 29*(5), 478–488. https://doi.org/10.1002/gps.4030
65. Tidermark, J., Bergström, G., Svensson, O., Törnkvist, H., & Ponzer, S. (2003). Responsiveness of the EuroQol (EQ 5-D) and the SF-36 in elderly patients with displaced femoral neck fractures. *Quality of Life Research, 12*(8), 1069–1079. https://doi.org/10.1023/a:1026193812514
66. Tidermark, J., & Bergström, G. (2007). Responsiveness of the EuroQol (EQ-5D) and the Nottingham health profile (NHP) in elderly patients with femoral neck fractures. *Quality of Life Research, 16*(2), 321–330. https://doi.org/10.1007/s11136-006-9004-4
67. van Leeuwen, K. M., Bosmans, J. E., Jansen, A. P., Hoogendijk, E. O., van Tulder, M. W., van der Horst, H. E., & Ostelo, R. W. (2015). Comparing measurement properties of the EQ-5D-3L, ICECAP-O, and ASCOT in frail older adults. *Value in Health, 18*(1), 35–43. https://doi.org/10.1016/j.jval.2014.09.006
68. Walters, S. J., Morrell, C. J., & Dixon, S. (1999). Measuring health-related quality of life in patients with venous leg ulcers. *Quality of Life Research, 8*(4), 327–336. https://doi.org/10.1023/a:1008992006845
69. Aguirre, E., Kang, S., Hoare, Z., Edwards, R. T., & Orrell, M. (2016). How does the EQ-5D perform when measuring quality of life in dementia against two other dementia-specific outcome measures? *Quality of Life Research, 25*(1), 45–49. https://doi.org/10.1007/s11136-015-1065-9
70. Bjerk, M., Brovold, T., Davis, J. C., & Bergland, A. (2019). Evaluating a falls prevention intervention in older home care recipients: A comparison of SF-6D and EQ-5D. *Quality of Life Research, 28*(12), 3187–3195. https://doi.org/10.1007/s11136-019-02258-x
71. Cheng, Q., Kularatna, S., Lee, X. J., Graves, N., & Pacella, R. E. (2019). Comparison of EQ-5D-5L and SPVU-5D for measuring quality of life in patients with venous leg ulcers in an Australian setting. *Quality of Life Research, 28*(7), 1903–1911. https://doi.org/10.1007/s11136-019-02128-6
72. Easton, T., Milte, R., Crotty, M., & Ratcliffe, J. (2018). An empirical comparison of the measurement properties of the EQ-5D-5L, DEMQOL-U and DEMQOL-Proxy-U for older people in residential care. *Quality of Life Research, 27*(5), 1283–1294. https://doi.org/10.1007/s11136-017-1777-0

73. Griffiths, A. W., Smith, S. J., Martin, A., Meads, D., Kelley, R., & Surr, C. A. (2020). Exploring self-report and proxy-report quality-of-life measures for people living with dementia in care homes. *Quality of Life Research, 29*(2), 463–472. https://doi.org/10.1007/s11136-019-02333-3
74. Martin, A., Meads, D., Griffiths, A. W., & Surr, C. A. (2019). How should we capture health state utility in dementia? Comparisons of DEMQOL-proxy-U and of self- and proxy-completed EQ-5D-5L. *Value in Health, 22*(12), 1417–1426. https://doi.org/10.1016/j.jval.2019.07.002
75. Nikolova, S., Hulme, C., West, R., Pendleton, N., Heaven, A., Bower, P., Humphrey, S., Farrin, A., Cundill, B., Hawkins, R., & Clegg, A. (2020). Normative estimates and agreement between 2 measures of health-related quality of life in older people with frailty: Findings from the community ageing research 75+ cohort. *Value in Health, 23*(8), 1056–1062. https://doi.org/10.1016/j.jval.2020.04.1830
76. Ratcliffe, J., Flint, T., Easton, T., Killington, M., Cameron, I., Davies, O., Whitehead, C., Kurrle, S., Miller, M., Liu, E., & Crotty, M. (2017). An Empirical comparison of the EQ-5D-5L, DEMQOL-U and DEMQOL-proxy-U in a post-hospitalisation population of frail older people living in residential aged care. *Applied Health Economics and Health Policy, 15*(3), 399–412. https://doi.org/10.1007/s40258-016-0293-7
77. Janssen, M. F., Lubetkin, E. I., Sekhobo, J. P., & Pickard, A. S. (2011). The use of the EQ-5D preference-based health status measure in adults with type 2 diabetes mellitus. *Diabetic Medicine, 28*(4), 395–413. https://doi.org/10.1111/j.1464-5491.2010.03136.x
78. Cameron, L. J., Wales, K., Casey, A., Pike, S., Jolliffe, L., Schneider, E. J., Christie, L. J., Ratcliffe, J., & Lannin, N. A. (2021). Self-reported quality of life following stroke: A systematic review of instruments with a focus on their psychometric properties. *Quality of Life Research*. https://doi.org/10.1007/s11136-021-02944-9
79. Feng, Y.-S., Kohlmann, T., Janssen, M. F., & Buchholz, I. (2021). Psychometric properties of the EQ-5D-5L: A systematic review of the literature. *Quality of Life Research, 30*(3), 647–673. https://doi.org/10.1007/s11136-020-02688-y
80. Devlin, N., Parkin, D., & Janssen, B. (2020). Advanced topics. *Methods for analysing and reporting EQ-5D data* (pp. 87–98). Springer International Publishing.
81. Davis, J. C., Khan, K. M., Hsu, C. L., Chan, P., Cook, W. L., Dian, L., & Liu-Ambrose, T. (2020). Action seniors! cost-effectiveness analysis of a secondary falls prevention strategy among community-dwelling older fallers. *Journal of the American Geriatrics Society, 68*(9), 1988–1997. https://doi.org/10.1111/jgs.16476
82. Hewitt, J., Saing, S., Goodall, S., Henwood, T., Clemson, L., & Refshauge, K. (2019). An economic evaluation of the SUNBEAM programme: A falls-prevention randomized controlled trial in residential aged care. *Clinical Rehabilitation, 33*(3), 524–534. https://doi.org/10.1177/0269215518808051
83. Robertson, M. C., Campbell, A. J., Gardner, M. M., & Devlin, N. (2002). Preventing injuries in older people by preventing falls: A meta-analysis of individual-level data. *Journal of the American Geriatrics Society, 50*(5), 905–911. https://doi.org/10.1046/j.1532-5415.2002.50218.x
84. Buchholz, I., Janssen, M. F., Kohlmann, T., & Feng, Y.-S. (2018). A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *PharmacoEconomics, 36*(6), 645–661. https://doi.org/10.1007/s40273-018-0642-5
85. Janssen, M. F., Bonsel, G. J., & Luo, N. (2018). Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *PharmacoEconomics, 36*(6), 675–697. https://doi.org/10.1007/s40273-018-0623-8
86. Grewal, I., Lewis, J., Flynn, T., Brown, J., Bond, J., & Coast, J. (2006). Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? *Social Science and Medicine, 62*(8), 1891–1901. https://doi.org/10.1016/j.socscimed.2005.08.023
87. Power, M., Quinn, K., & Schmidt, S. (2005). Development of the WHOQOL-old module. *Quality of Life Research, 14*(10), 2197–2214. https://doi.org/10.1007/s11136-005-7380-9
88. Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS® adult health profiles: Efficient short-form measures of seven health domains. *Value in Health, 22*(5), 537–544. https://doi.org/10.1016/j.jval.2019.02.004
89. Dewitt, B., Feeny, D., Fischhoff, B., Cella, D., Hays, R. D., Hess, R., Pilkonis, P. A., Revicki, D. A., Roberts, M. S., Tsevat, J., Yu, L., & Hanmer, J. (2018). Estimation of a preference-based summary score for the patient-reported outcomes measurement information system: The PROMIS(®)-preference (PROPr) scoring system. *Medical Decision Making, 38*(6), 683–698. https://doi.org/10.1177/0272989x18776637
90. Hanmer, J., Cella, D., Feeny, D., Fischhoff, B., Hays, R. D., Hess, R., Pilkonis, P. A., Revicki, D., Roberts, M., Tsevat, J., & Yu, L. (2017). Selection of key health domains from PROMIS® for a generic preference-based scoring system. *Quality of Life Research, 26*(12), 3377–3385. https://doi.org/10.1007/s11136-017-1686-2
91. Himmler, S., van Exel, J., & Brouwer, W. (2020). Happy with your capabilities? Valuing ICECAP-O and ICECAP-A states based on experienced utility using subjective well-being data. *Medical Decision Making, 40*(4), 498–510. https://doi.org/10.1177/0272989x20923015
92. Coast, J., Flynn, T. N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J. J., & Peters, T. J. (2008). Valuing the ICECAP capability index for older people. *Social Science and Medicine, 67*(5), 874–882. https://doi.org/10.1016/j.socscimed.2008.05.015
93. Cleland, J., Hutchinson, C., McBain, C., Walker, R., Milte, R., Khadka, J., & Ratcliffe, J. (2021). Developing dimensions for a new preference-based quality of life instrument for older people receiving aged care services in the community. *Quality of Life Research, 30*(2), 555–565. https://doi.org/10.1007/s11136-020-02649-5
94. Ratcliffe, J., Cameron, I., Lancsar, E., Walker, R., Milte, R., Hutchinson, C. L., Swaffer, K., & Parker, S. (2019). Developing a new quality of life instrument with older people for economic evaluation in aged care: Study protocol. *British Medical Journal Open, 9*(5), e028647. https://doi.org/10.1136/bmjopen-2018-028647
95. Geraerds, A. J. L. M., Bonsel, G. J., Janssen, M. F., Finch, A. P., Polinder, S., & Haagsma, J. A. (2021). Methods used to identify, test, and assess impact on preferences of bolt-ons: A systematic review. *Value in Health, 24*(6), 901–916. https://doi.org/10.1016/j.jval.2020.12.011
96. Finch, A. P., Brazier, J. E., Mukuria, C., & Bjorner, J. B. (2017). An exploratory study on using principal-component analysis and confirmatory factor analysis to identify bolt-on dimensions: The EQ-5D case study. *Value in Health, 20*(10), 1362–1375. https://doi.org/10.1016/j.jval.2017.06.002
97. Yang, Y., Rowen, D., Brazier, J., Tsuchiya, A., Young, T., & Longworth, L. (2015). An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. *Value in Health, 18*(1), 52–60. https://doi.org/10.1016/j.jval.2014.09.004

98. Finch, A. P., Brazier, J., & Mukuria, C. (2021). Selecting bolt-on dimensions for the EQ-5D: Testing the impact of hearing, sleep, cognition, energy, and relationships on preferences using pairwise choices. *Medical Decision Making, 41*(1), 89–99. https://doi.org/10.1177/0272989x20969686
99. Finch, A. P., Brazier, J. E., & Mukuria, C. (2019). Selecting bolt-on dimensions for the EQ-5D: Examining their contribution to health-related quality of life. *Value in Health, 22*(1), 50–61. https://doi.org/10.1016/j.jval.2018.07.001
100. Chen, G., & Olsen, J. A. (2020). Filling the psycho-social gap in the EQ-5D: The empirical support for four bolt-on dimensions. *Quality of Life Research, 29*(11), 3119–3129. https://doi.org/10.1007/s11136-020-02576-5
101. Beerens, H. C., Zwakhalen, S. M., Verbeek, H., Ruwaard, D., & Hamers, J. P. (2013). Factors associated with quality of life of people with dementia in long-term care facilities: A systematic review. *International Journal of Nursing Studies, 50*(9), 1259–1270. https://doi.org/10.1016/j.ijnurstu.2013.02.005
102. Jing, W., Willis, R., & Feng, Z. (2016). Factors influencing quality of life of elderly people with dementia and care implications: A systematic review. *Archives of Gerontology and Geriatrics, 66*, 23–41. https://doi.org/10.1016/j.archger.2016.04.009
103. Asakawa, K., Senthilselvan, A., Feeny, D., Johnson, J., & Rolfson, D. (2012). Trajectories of health-related quality of life differ by age among adults: Results from an eight-year longitudinal study. *Journal of Health Economics, 31*(1), 207–218. https://doi.org/10.1016/j.jhealeco.2011.10.002

9 Summary in English and German

9.1 Summary

The sustainability of healthcare systems is challenged by an aging population, making cost-effective interventions to promote healthy aging desirable. The overall aim of this cumulative thesis was to economically evaluate interventions to improve care for the older population, using the example of fall prevention and activity promotion programs that potentially contribute to healthy aging. The economic analysis was complemented by an analysis of mediating factors of fear of falling (a factor linked to physical activity and falls) and health-related quality of life (HrQoL). In addition, the measurement properties of the EQ-5D as the most widely used HrQoL instrument to measure effectiveness in economic evaluations were systematically reviewed in a middle-old to oldest-old population.

The results of project 1 confirmed the negative association found in previous studies between fear of falling and HrQoL and showed that this association is mediated by subjective functional capacity. Project 2 presents the economic analysis of the LiFE-is-LiFE randomized controlled non-inferiority trial (n=309). To this end, a cost-effectiveness analysis comparing the group version (gLiFE) of the *Lifestyle-integrated Functional Exercise* program with the individually delivered version (LiFE) was conducted. It can be concluded that, despite lower intervention costs, gLiFE was not clearly superior to LiFE in terms of cost-effectiveness. However, the results constituted a high degree of uncertainty and varied depending on the outcome used to quantify effectiveness (quality-adjusted life years [QALYs], physical activity, number of falls). An analysis of the participants' willingness to pay (WTP) for the intervention indicated that both program versions are perceived as valuable by the participants. The WTP was on average higher in LiFE than in gLiFE participants, but given the lower intervention costs of gLiFE, only in this group the majority had a WTP high enough to cover the intervention costs, whereas this was only the case in a small proportion of LiFE participants. In project 3, the results of the systematic review highlighted weaknesses of the EQ-5D, especially with regard to responsiveness and reliability, which has implications for the results of the other studies, such as a potential underestimation of effectiveness in the cost-effectiveness analysis.

Overall the thesis contributes to the field of economic evaluations of exercise-based fall prevention programs for older, community-dwelling people in the German context where research so far is limited. Beyond, it contributes to the discussion of appropriate effect measures in economic evaluations of interventions targeting older people.

9.2 Zusammenfassung

Die alternde Bevölkerung stellt eine Herausforderung für die Nachhaltigkeit von Gesundheitssystemen dar, wodurch kosteneffektive Interventionen zur Förderung gesunden Alterns an Bedeutung gewinnen. Diese kumulative Thesis untersuchte am Beispiel von Programmen zur Aktivitätsförderung und Sturzprävention die Kosteneffektivität von Interventionen zur Versorgungsverbesserung der älteren Bevölkerung. Diese ökonomische Evaluation wurde durch eine Analyse der Faktoren ergänzt, die den Zusammenhang zwischen Sturzangst (die sowohl mit physischer Aktivität als auch mit Stürzen zusammenhängt) und gesundheitsbezogener Lebensqualität (HrQoL) erklären. Darüber hinaus wurde eine systematische Übersichtsarbeit zu den Messeigenschaften des EQ-5D (dem meist verbreitetsten Instrument zur Messung von HrQoL) bei hochaltrigen Personen angefertigt.

In Projekt 1 der Thesis konnte ein negativer Zusammenhang zwischen Sturzangst und gesundheitsbezogener Lebensqualität gefunden werden, der (teilweise) durch subjektive funktionale Kapazität erklärt wird. Projekt 2 umfasst die ökonomische Evaluation der randomisierten Nicht-Unterlegenheitsstudie *LiFE-is-LiFE* (N=309). Dazu wurde eine Kosteneffektivitätsanalyse durchgeführt, die eine Gruppenvariante (gLiFE) des *Lifestyle-integrated Functional Exercise*-Programms mit der individuell vermittelten Variante (LiFE) vergleicht. Trotz deutlich geringerer Interventionskosten von gLiFE konnte die Kosteneffektivität nicht eindeutig gezeigt werden. Die Ergebnisse wiesen einen hohen Grad an (statistischer) Unsicherheit auf und variierten je nach verwendetem Effektmaß (qualitätsadjustierte Lebensjahre, physische Aktivität, Stürze). Eine Analyse der Zahlungsbereitschaft der Teilnehmenden zeigte, dass diese zwar im Durchschnitt höher für LiFE als für gLiFE war, aber nur bei gLiFE in der Mehrheit ausreichte, um die Interventionskosten zu decken. Hingegen war das nur bei wenigen LiFE-Teilnehmenden der Fall. Die Ergebnisse von Projekt 3 deuteten auf Schwächen des EQ-5D hinsichtlich der Veränderungssensitivität und Reliabilität hin, was Implikationen für die Ergebnisse der anderen Studien dieser Thesis hat (z.B. eine mögliche Unterschätzung des Interventionseffektes in der Kosteneffektivitätsanalyse).

Insgesamt leistet die Thesis einen Beitrag zum Forschungsfeld der ökonomischen Evaluationen von übungsbasierten Sturzpräventionsprogrammen für ältere, nicht-institutionalisierte Personen, wozu es besonders in Deutschland bisher wenig Forschung gibt. Darüber leistet sie einen Beitrag zur Diskussion geeigneter Effektmaße in ökonomischen Evaluationen von Intervention zur Versorgungsverbesserung der älteren Bevölkerung.

10 Erklärung des Eigenanteils an den Publikationen

Die in diese Thesis eingeschlossenen Publikationen wurden von mir, Sophie Gottschalk, in alleiniger Erstautorenschaft (Publikationen 1, 2, 4, 5 und 6) bzw. geteilter Erstautorenschaft (Publikation 3) verfasst.

Für die Publikationen 1, 2 und 4 wurden die Daten durch mich aufbereitet und in regelmäßigem Austausch mit insbesondere PD Dr. Judith Dams und Prof. Dr. Hans-Helmut König, aber auch den weiteren Co-Autor:innen, analysiert und interpretiert.

Die Idee für die systematische Übersichtsarbeit (Publikationen 5 und 6) kam von Prof. Dr. Hans-Helmut König. Das konkrete Thema und Ziel der Arbeit wurden aber von mir weiter spezifiziert. Zudem wurde die Recherche in den Datenbanken (einschließlich der Anpassung des durch die COSMIN-Gruppe veröffentlichten Suchalgorithmus) durch mich durchgeführt. Bei der Identifikation geeigneter Studien zum Einschluss in das Review, der Extraktion relevanter Daten sowie bei der Bewertung der methodischen Qualität eingeschlossener Arbeiten wurde ich durch Mona Nejad unterstützt.

Die ersten Entwürfe der Manuskripte in alleiniger Erstautorenschaft wurden von mir angefertigt und in Abstimmung mit den Co-Autor:innen, insbesondere PD Dr. Judith Dams und Prof. Dr. Hans-Helmut König, überarbeitet. Zudem wurden diese Artikel durch mich bei entsprechenden Zeitschriften eingereicht und sämtliche Überarbeitungen im Rahmen des Peer-Review-Verfahrens vorgenommen.

Für Publikation 3 lag mein Eigenanteil in der Durchführung der multiplen Imputation für die gesamte Studie, der Durchführung und Interpretation der Kosteneffektivitätsanalyse (inklusive entsprechender Datenaufbereitung), im Entwurf der Textabschnitte zu den Interventionskosten und der Kosteneffektivitätsanalyse, sowie der Unterstützung von Dr. Carl-Philipp Jansen bei der Entwicklung des Entwurfs für das Gesamtmanuskript. Zudem wurden Überarbeitungen im Rahmen des Peer-Review-Verfahrens an den die ökonomische Analyse betreffenden Abschnitten durch mich vorgenommen.

11 Danksagung

Es gibt eine Reihe von Menschen, ohne deren Unterstützung die Umsetzung meines Promotionsprojektes nicht möglich gewesen wäre.

Ich möchte mich herzlich bei meinem Doktorvater Prof. Dr. Hans-Helmut König bedanken, der mich durch seine fachliche Expertise in allen Teilprojekten unterstützt und mir sehr viel Wertschätzung für meine Arbeit entgegengebracht hat, die ich als sehr motivierend empfunden habe.

Mein besonderer Dank gilt außerdem meiner Arbeitsgruppenleiterin PD Dr. Judith Dams für die intensive Beratung und Betreuung während der Zeit.

Daneben danke ich auch den Mitgliedern meines Thesis-Komitees, PD Dr. Christine Blome und Prof. Dr. Holger Schulz, für die wertvollen Anmerkungen zu den Thesis-Projekten.

Außerdem bedanke ich mich beim gesamten Projektteam der LiFE-is-LiFE-Studie – insbesondere meinen Co-Autor:innen – für die herzliche Aufnahme in das Projekt und die gute interprofessionelle Zusammenarbeit.

Zu guter Letzt bedanke ich mich bei meinen Kolleg:innen, Freund:innen und meiner Familie, die mir stets zur Seite standen und mich auch immer wieder daran erinnert haben, dass Promotion und Arbeit nicht den Mittelpunkt der Welt darstellen.

12 Lebenslauf

Der Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.

13 Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: ....