Erasmus MC
Universitair Medisch Centrum Rotterdam

Cost-effectiveness of a proactive, integrated primary care approach for community-dwelling frail older persons

Erasmus University Rotterdam

ABSTRACT

Background

The article reports on the cost-effectiveness of the proactive, integrated primary care program *Finding and Follow-up of Frail older persons* (FFF) compared with usual primary care for community-dwelling frail older persons in the Netherlands.

Methods

This study had a matched quasi-experimental design (pretest and posttest). The economic evaluation was performed from a healthcare perspective with a time horizon of 12 months. The target population consisted of community-dwelling frail older persons aged ≥ 75 years in the FFF intervention group (11 general practitioner (GP) practices) and in the control group receiving usual care (4 GP practices). The effectiveness measures for the cost-effectiveness and cost-utility analyses were subjective well-being (Social Production Function Instrument for the Level of well-being short; SPF-ILs) and QALYs (EuroQol; EQ-5D-3L), respectively. Costs were assessed using resource use questionnaires. Differences in mean effectiveness between groups were assessed using univariate, multilevel and propensity score matched analyses, with and without imputation of missing values. Differences in costs were assessed using Mann-Whitney *U*-tests and independent samples *t*-tests. Bootstrapping was performed, and predicted incremental cost-effectiveness ratios (ICERs) and incremental cost-utility ratios (ICURs) were depicted on cost-effectiveness planes.

Results

The various analyses showed slightly different results with respect to differences in estimated costs and effects. Multilevel analyses showed a small but significant difference between the groups for well-being, in favor of the control group. No significant differences between groups in terms of QALYs were found. Imputed data showed that mean total costs were significantly higher in the intervention group at follow-up.

Conclusion

Proactive, integrated care for community-dwelling frail older persons as provided in the FFF program is most likely not a cost-effective initiative, compared with usual primary care in the Netherlands, in terms of well-being and QALYs over a 12-month period.

Erasmus University Rotterdam

BACKGROUND

This article reports on the cost-effectiveness of a proactive, integrated primary care approach compared with usual primary care for community-dwelling frail older persons in the Netherlands. We evaluated the *Finding and Follow-up of Frail older persons* (FFF) approach, which aims to maintain or improve older people's well-being and is implemented by part of the Dutch general practitioners (GPs). The FFF approach consists of proactive identification of frail older persons in the community and subsequent multidisciplinary (including professionals with geriatric expertise) consultations and individualized follow-up coordinated by case managers. Integrated care and support is widely acknowledged to be a key initiative in improving care and support for older persons (Harvey, Dollard, Marshall, & Mittinty, 2018). In addition, integrated care approaches, like the FFF program, may help to maintain community-dwelling frail older persons' well-being (Nieboer & Cramm, 2018). Over the years, a shift has occurred from a disease-oriented care model toward a more proactive and integrated approach (Hopman et al., 2016). Traditional disease-specific care delivery approaches for frail older persons, who often have multiple conditions, do not meet these individuals' comprehensive (healthcare) needs (Boyd et al., 2005; De Lepeleire, Iliffe, Mann, & Degryse, 2009; Guthrie, Payne, Alderson, McMurdo, & Mercer, 2012; Nolte & McKee, 2008; van Weel & Schellevis, 2006). Moreover, frailty has been associated with increased utilization of primary, hospital, and nursing home care (e.g. Ilinca & Calciolari, 2015; Rochat et al., 2010). The provision of high-quality care and support to the growing number of frail older persons poses a challenge (Banerjee, 2015; WHO, 2015), and the comprehensive (healthcare) needs of this population place a burden on healthcare resources (Chatterji, Byles, Cutler, Seeman, & Verdes, 2015). Integrated care initiatives are assumed to improve quality of care and ultimately aim to enhance patient outcomes while making efficient use of healthcare resources (Mattke, Seid, & Ma, 2007; Wagner et al., 2005). Important elements of integrated care are: (i) a proactive approach that is coordinated effectively around a person's health and social care needs; (ii) a patient-centered approach in which a person is involved in decision-making and care processes, and the person's needs are taken into consideration; (iii) an approach in which multiple interventions are delivered (simultaneously); and (iv) a multidisciplinary approach in which professionals from multiple disciplines are involved (Hopman et al., 2016). GPs are considered to be key actors in the implementation of promising initiatives targeting frail older persons (Ilinca & Calciolari, 2015). Many integrated care initiatives have emerged and are implemented in the primary healthcare sector, but evidence of their effectiveness and cost-effectiveness remains mixed (Blom et al., 2018; de Bruin et al., 2012; Eklund & Wilhelmson, 2009; Hopman et al., 2016; Looman, Huijsman, & Fabbricotti, 2018; Low, Yap, & Brodaty, 2011; Smith, Wallace, O'Dowd, & Fortin, 2016). Integrated primary care programs for frail older persons have shown no effect on the majority of outcomes, and evidence for their cost-effectiveness is limited (Looman et al., 2018). Although the FFF approach has been found to have positive effects on the quality of care as perceived by healthcare professionals (Vestjens, Cramm, & Nieboer, 2018b),

and to achieve improvements in older persons' perceived care quality and coproduction of care over time (Vestjens, Cramm, and Nieboer, 2019), its cost-effectiveness has yet to be investigated. Therefore, the aim of the present study was to evaluate the cost-effectiveness of the FFF approach in a population of community-dwelling frail older persons.

METHODS

Design, setting and participants

This longitudinal evaluation study had a matched quasi-experimental design with one pretest and one posttest (12-month follow-up period). The study was conducted in 15 GP practices located in the western part of North Brabant Province, the Netherlands, between 2014 and 2017. The intervention group consisted of community-dwelling frail persons aged 75 years and older who were registered at 11 GP practices that implemented the proactive, integrated primary care approach FFF. The control group consisted of community-dwelling frail older persons ($\geq$ 75 years of age) who were registered at 4 GP practices that delivered usual primary care. Written informed consent to participate in the study was obtained from all participants. All participants in the intervention group were individually matched one-to-one to participants in the control group based on sex (male/female), educational level (low/high), and frailty score. As shown in Figure 1, each group consisted of 232 frail older persons at baseline. At T1, 182 older persons remained in the intervention group and 176 older persons remained in the control group (loss to follow-up rates of 21.6% and 24.1% respectively). The medical research ethics committee of the Erasmus Medical Centre in Rotterdam, the Netherlands, concluded that the rules laid out in the Medical Research Involving Human Subjects Act did not apply (study protocol number MEC-2014-444). More details of the study design have been published elsewhere (Vestjens, Cramm, Birnie, & Nieboer, 2018a).

Usual primary care

Compared with the primary care systems in many countries in Europe, the Dutch primary care system is strongly developed. Many different (healthcare) providers, including GPs, primary care psychologists and physiotherapists, are involved in primary care delivery in the Netherlands (Kroneman et al., 2016). GPs have a central role in the healthcare sector and a strong gatekeeping function (Schäfer et al., 2010), implying that referral is generally necessary to access most hospital and specialist care (Kringos, Boerma, Hutchinson, & Saltman, 2015; Kroneman et al., 2016). Each patient is registered at a GP practice of his or her choice, usually located in the person's neighborhood. GPs are commonly patients' first contact with the healthcare system, and most first encounters take place after the occurrence of a (medical) problem. In general, GPs in the Netherlands are considered to be non-interventionist, resulting in relatively low prescription and referral rates. In comparison with GPs in other European countries, Dutch GPs provide a

broad scope of (healthcare) services to their patients. Collaboration between GPs and (practice) nurses is common (Kroneman et al., 2016). An example of a nurse-led service is the provision of diabetes management in primary care (Kringos et al., 2015). In the care for community-dwelling frail older persons, GPs can consult elderly care physicians with expertise in geriatric medicine (Koopmans, Lavrijsen, Hoek, Went, & Schols, 2010). The primary care system lacks, however, sufficient coordination and continuity (with specialist care) and is reactive and characterized by fragmentation (Boeckxstaens & De Graaf, 2011). Frail older persons in this study's control group received usual care services provided by their GP practices and local health and community organizations. For a detailed description of care delivery and implemented interventions in control and intervention GP practices, see Vestjens, Cramm, and Nieboer (2018b).
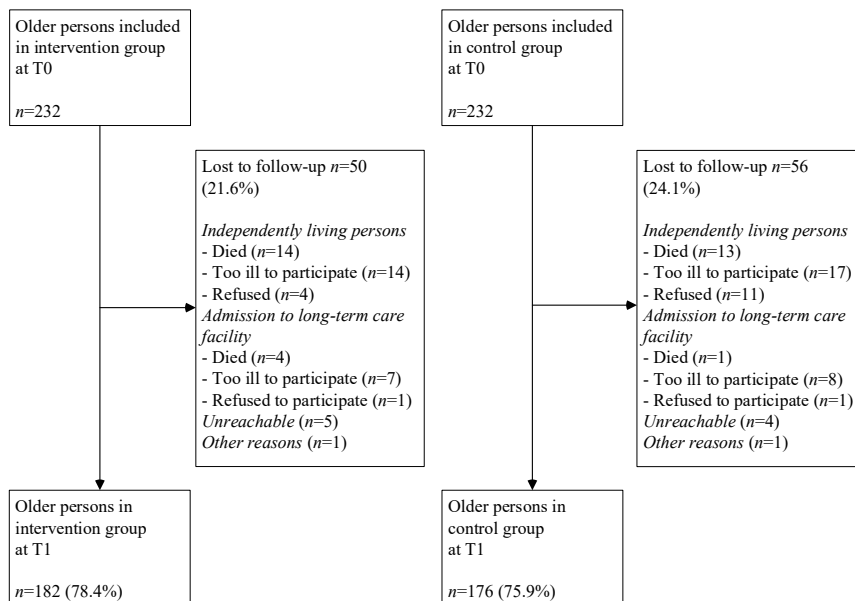

Figure 1 Flowchart of study participation

Intervention

Frail older persons in the intervention group received primary care according to the FFF approach. This approach combines several interrelated components (see Table 1) with the aim of providing high-quality proactive, integrated primary care for frail community-dwelling older persons. A decline of the well-being of community-dwelling frail older persons may be expected over time (Nieboer & Cramm, 2018). Consequently, the aim of the FFF program is to maintain or improve frail older people's well-being and protect against its deterioration. The FFF approach is implemented in GP practices and led by GPs. Community-dwelling older patients registered at the GP practices are screened for frailty using the Tilburg Frailty Indicator (TFI) (Gobbens,

van Assen, Luijkx, Wijnen-Sponselee, & Schols, 2010) during a home visit by the practice nurse, homecare nurse or geriatric nurse. This 15-item questionnaire assesses frailty in the physical, psychological, and social domains. Scores range from 0 to 15, and persons with scores ≥ 5 are identified as frail (Gobbens et al., 2010). Persons with TFI scores < 5 can also be identified as frail based on additional examination by professionals. Problems and needs are reported in multiple domains according to the SFSPC-model, i.e., somatic (e.g., pain, fall risk), functional (e.g., limitations in activities of daily living like problems with eating or household activities), social (e.g., social network), psychological (e.g., fear, coping, depression), and communication (e.g., visual or hearing impairments). Outcomes of this in-home assessment are reported and discussed with the GPs and elderly care physicians, i.e., physicians in primary care with expertise in geriatric medicine (Koopmans et al., 2010). Multidisciplinary primary care teams and collaboration among different disciplines in multiple FFF-related activities are central to the FFF approach. Geriatric expertise is easily accessible by close involvement of elderly care physicians and geriatric nurses. Older persons' (healthcare) needs are discussed in multidisciplinary consultation at least once a year. Individualized care plans include reported problems and (healthcare) needs, tailored (self-management) interventions, plans for multidisciplinary follow-up and evaluation. The care plan is discussed with the older person during a home visit by the practice nurse, homecare nurse or geriatric nurse. The care plan is then tailored to the person's needs and wishes. Follow-up of older patients is arranged by a multidisciplinary team of (healthcare) professionals and an appointed case manager, who coordinates and evaluates the process, and provides support in goal setting and self-management. Older patients' medication use is examined at least annually by GPs, elderly care physicians or pharmacists and discussed with the patients and their informal

Box 1 A case of a frail older person participating in the FFF approach

Mr. Buys is 82 years old and has always lived in his parental home in the countryside near Roosendaal. His wife passed away 2 years ago and his two sons live and work in the capital city, Amsterdam. A diabetes check-up by his practice nurse raised alarm regarding Mr. Buys' physical and social well-being. In response, the practice nurse screened Mr. Buys for frailty during a home visit; his TFI score was 8. In addition, Mr. Buys reported problems in the somatic, functional, and social domains. It became apparent that Mr. Buys misses having people around him and experiences problems in his daily life due to fatigue and difficulty in walking. He explained to the practice nurse that he lacks contact with his social network. After discussion with the GP and elderly care physician, a preliminary individualized care plan was established and Mr. Buys' case was discussed in multidisciplinary consultation. Based on Mr. Buys' reported problems and needs, a physiotherapist, geriatric nurse, and social worker were included in the multidisciplinary team, along with the practice nurse, GP, and elderly care physician. The geriatric nurse was appointed as Mr. Buys' case manager (responsible for, e.g., discussing the (self-management) interventions that were proposed in the multidisciplinary consultation and adjusting the care plan to his wishes). The elderly care physician examined Mr. Buys' medication use and arranged a home visit to evaluate his diabetes management. The physiotherapist visited Mr. Buys regularly to improve his physical functioning and to minimize fall risk through, e.g., walking and balance exercises. The case manager discussed several options to improve Mr. Buys' social contact and independence. He decided to visit a day care center twice a week to be involved in meaningful activities and contact with older persons in his area of residence. The geriatric nurse contacted Mr. Buys (by home visit or telephone) and evaluated his follow-up regularly.

caregivers or relatives. Box 1 illustrates the application of the FFF approach. Further details on the components of the FFF approach have been published elsewhere (Vestjens et al., 2018a; 2018b).

Cost-utility and cost-effectiveness analyses

The economic evaluation of the FFF approach consisted of a cost-utility analysis (CUA) and a cost-effectiveness analysis (CEA) from a healthcare perspective with a time horizon of 12 months. Costs and effects were measured at baseline (T0) and 12 months (T1). Trained interviewers administered questionnaires during in-home interviews to collect data regarding healthcare utilization and outcomes. The incremental cost-utility ratio (ICUR) and the incremental cost-effectiveness ratio (ICER) represent the difference in mean total costs adjusted for baseline costs between the intervention and control groups in the numerator and the difference in mean effectiveness adjusted for baseline effectiveness in the denominator.

Outcomes and measures

The primary outcome of the CUA was quality-adjusted life-years (QALYs). The validated EuroQol (EQ-5D-3L) was used as a preference-based health status measure to estimate utilities in the QALY measure (Brooks, 1996; EuroQol Group, 1990). The descriptive system of the EQ-5D measures health-related quality of life in five health dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) on a three-point scale (1 = no problems, 2 = some or moderate problems, 3 = severe problems), resulting in 243 distinct health states (Brooks, 1996; EuroQol Group, 1990). The EQ-5D health states were transformed into utility scores using the Dutch EQ-5D tariffs. Utilities based on the Dutch tariff range from -0.33 to 1 (<0 = health state considered worse than death, 0 = death or health state regarded to be equivalent to death, 1 = full health) (Lamers, Stalmeier, McDonnell, Krabbe, & van Busschbach, 2005). For the CEA, the primary outcome was subjective well-being, measured with the validated short form of the Social Production Function Instrument for the Level of well-being (SPF-ILs) (Nieboer, Lindenberg, Boomsma, & van Bruggen, 2005). This 15-item instrument assesses whether five instrumental goals (comfort, stimulation, status, behavioral confirmation, and affection) are met in order to optimize universal goals of social and physical well-being (Nieboer & Cramm, 2018; Nieboer et al., 2005; Ormel, Lindenberg, Steverink, & Verbrugge, 1999; Ormel, Lindenberg, Steverink, & Von Korff, 1997). Mean SPF-ILs scores range from 1 to 4, with higher scores indicating greater subjective well-being (Nieboer et al., 2005). EQ-5D and SPF-ILs outcomes were measured among frail older patients in the intervention and control groups at T0 and T1.

Healthcare utilization and costs

Total costs of intervention and control care at T0 and T1 were estimated as the summation of resources used multiplied by prices or valuations. Volumes of healthcare utilization were determined by the administration of resource use questionnaires during in-home interviews at T0 and T1. Frail older persons reported the types and frequencies (e.g., days of hospitalization or

number of visits to the GP) of care they had received. To determine volumes of resource use, the following data were collected: numbers of GP consultations, out-of-hours GP consultations (i.e., at home or in the care clinic on evenings/nights/weekends), admissions to hospital/nursing home/home for the elderly, and visits to the physiotherapist/exercise therapist/psychologist/ psychiatrist/social worker/medical specialist; and types of homecare service (i.e., household activities, personal care and nursing care at home), and elderly daycare or daycare treatment received. We also included costs related to the purchase of assistive aids (e.g., wheelchair) and in-home modifications as patient-related costs. Intervention costs included all costs related to the FFF activities, i.e., selection of patients, proactive screening for frailty, provision of feedback information, multidisciplinary consultation, individualized care plan development, medication review, and follow-up of frail older persons (see Table 1). We estimated the average amount of time spent on intervention related activities per patient by (a combination of) different healthcare professionals involved. Information for this estimation was collected by the FFF project leader and was based on registers of the contact persons for GP practices, minutes from multidisci-plinary consultations, and observations made during frailty screening. The healthcare profes-sionals involved in FFF-related activities differed among GP practices and frail older patients, due to the compositions of the practice teams, (healthcare) disciplines with services accessible in the region, and tailoring of the FFF approach to the wishes and needs of individual patients. FFF follow-up involved healthcare utilization (e.g., consultation with the practice nurse or social worker). To avoid duplicate inclusion of costs, we included such service use in the healthcare costs, and not in the intervention costs. Only consultations with healthcare professionals that were not registered on the resource use questionnaires were included in the intervention costs. Study-related activities and costs were excluded. We did not consider costs related to the training of involved healthcare professionals.

Healthcare resource volumes were valued using the Dutch manual for costing in healthcare (Hakkaart-van Roijen, van der Linden, Bouwmans, Kanters, & Tan, 2015). Volumes of resource use were multiplied by standardized cost prices per unit of resource use (in euros) to estimate costs. Prices were inflated to 2015 (reference year) using the general consumer price index of 0.6% (Statistics Netherlands). When standardized costs per unit of resource use were unavail-able, we estimated costs using true economic costs in the year 2015. To estimate costs related to out-of-hour GP consultations, we used the true economic weighted mean costs for this service in western North Brabant Province provided by the Dutch Healthcare Authority. Average expenditures based on Internet sources and expenditures obtained in previous research using the same resource use questionnaire were used to value purchased assistive aids and in-home modifications (van Dijk, Cramm, Lötters, van Exel, & Nieboer, 2015). Annual depreciation costs were calculated according to the annuity method (Hakkaart-van Roijen et al., 2015). Interven-tion costs were based on the average amount of time invested per FFF element and hourly wages

of the professionals involved, with proportional time investment applied when more than one professional was involved.

Missing data

Results are presented with and without imputation of missing values. Missing values were imputed according to the type of parameter (cost volume, utility, QALYs, SPF-ILs score), time point (T0, T1) and, for T1, reason for drop-out (see Figure 1). Missing cost volumes at T0 were imputed with the mean cost volume of the specific service for the intervention group or control group at T0. The imputation of QALYs at T0 depended on the number of missing EQ-5D domains. When only one EQ-5D domain score was missing, the EQ-5D utility score was imputed using the median utility score of other persons in the same (intervention or control) group who had the same scores on the non-missing EQ-5D domains. When no participant in the same group had the same scores or more than one domain was missing, the missing utility score was replaced with the median utility score for the respective group. The mean SPF-ILs score was calculated when at least 10 of the 15 items were reported. Missing values at T0 were replaced with the mean SPF-ILs score in the respective group at T0. Missing costs, utilities, QALYs and SPF-ILs scores for participants in the intervention and control groups at T1 were imputed the same way as at T0. Missing costs and effects on T1 questionnaires of older persons that were lost to follow-up in the intervention group and control group between T0 and T1 ($n = 50$ and $n = 56$ respectively; see Figure 1) were imputed as follows. Based on registrations of case managers and GPs, we estimated the number of months that a person lived at home, lived in a nursing/elderly home, and the number of months lost due to mortality. For each older person that was lost to follow-up, missing costs at T1 were imputed with the sum of (1) the number of months the person lived at home multiplied by the mean monthly healthcare costs (excluding the costs of nursing/elderly home admission) in the respective group at T1, and when applicable, (2) the number of months the person lived in a nursing/elderly home multiplied by monthly costs of nursing/elderly home admission, and (3) costs were set at zero from the month a participant died during the follow-up period. For persons for whom nothing further was known ($n = 11$), we used the mean healthcare costs in the respective group at T1. Missing QALYs at T1 were replaced with the sum of (1) the number of months a person lived at home multiplied by the median utility score at T1, and (2) the number of months a person lived in a nursing/elderly home multiplied by the utility score of 0.5 (Makai, Brouwer, Koopmanschap, & Nieboer, 2012), and (3) a utility score of 0 was assigned from the month a person died. For persons for whom nothing further was known ($n = 11$), we used the QALYs in the respective group at T1. Finally, missing SPF-ILs values were imputed with the mean SPF-ILs group score at T1 due to the lack of SPF-ILs norm values. Additional file 1: Table S1 outlines the number of participants with missing data.

Statistical analyses

We assessed differences in baseline characteristics between the intervention and control groups using independent samples *t*-tests (for continuous variables with approximately normal distributions) and Chi-squared tests (for categorical variables). Unadjusted differences in mean SPF-ILs scores and QALYs between groups were assessed using independent samples *t*-tests. Unadjusted differences in mean SPF-ILs scores and QALYs over time within each group were assessed using paired sample *t*-tests. Furthermore, these univariate analyses were complemented with multilevel analyses (linear mixed-effects models) to investigate effectiveness of the FFF approach. Multilevel models are considered appropriate for investigating relationships in data sets with continuous dependent variables and a clustered structure of the data (persons within GP practices) (West, Welch, & Gatecki, 2015). A random intercept was used on the level of the individual GP practices. Outcome estimates in the multilevel analyses were adjusted for baseline values of the respective outcome variable, background variables (i.e., age, sex, marital status, educational level, frailty score and multimorbidity) and control/intervention group. We performed the multilevel models (with QALYs and well-being as outcome estimates) using data with and without imputation of missing values.

Volumes of healthcare utilization were presented as means (and corresponding standard deviations; SDs) per service use category. Differences in costs between groups were tested using Mann-Whitney *U*-tests (skewed data) and independent samples *t*-tests (for mean values). Differences in costs over time within each group were assessed using related-samples Wilcoxon signed rank tests and paired sample *t*-tests.

Furthermore, propensity score matching was used to deal with potential different distributions of covariates between the intervention and control groups at baseline (Indurkhya, Mitra, & Schrag, 2006). According to Indurkhya, Mitra, and Schrag (2006), the propensity score is considered the probability that a person is assigned to the intervention group conditional on the person's covariate information. For each individual person, the propensity to be part of the intervention group was estimated using a binary logistic regression model predicting assignment to the intervention group from baseline covariates. The covariates in the first logistic regression model (Model 1) were age, sex, marital status, educational level, frailty score, and multimorbidity. Next to these covariates, we also included baseline SPF-ILs, QALYs and costs in the second logistic regression model (Model 2). We then compared observed outcomes between intervention and control groups conditional on the propensity matched scores (Rosenbaum & Rubin, 1983).

We performed nonparametric bootstrapping (percentile method) to generate 1500 samples from the original sample of 232 matched pairs. Predicted ICERs and ICURs were depicted on cost-effectiveness planes to show uncertainty therein. A statistical significance level of 5% (two-sided) was used in the analyses. All statistical analyses were performed with IBM SPSS Statistics version 24.

Erasmus University Rotterdam

Table 1 Activities related to the FFF integrated primary care approach

FFF related activities	Explanation	Disciplines involved	Mean time per patient
Selection of patients	Selecting patients that are eligible for proactive frailty screening	GP or practice nurse	5 minutes
Proactive frailty screening	Home visit for administering the Tilburg Frailty Indicator (TFI) to assess frailty. Consultation with the patient and reporting needs and problems based on the SFSPC-model, i.e. model for reporting on Somatic, Functional, Social, Psychological, and Communicative indications.	Practice nurse, geriatric nurse, or homecare nurse	90 minutes
Feedback information	Feedback information about the screening (e.g., TFI score) and problem analysis (SFSPC-model) to the GP and elderly care physician. First draft of individualized care plan for the patient.	Practice nurse, geriatric nurse, or homecare nurse	100 minutes
Multidisciplinary consultation	Discussing the older patient in multidisciplinary consultation in the GP practice. Discussion of screening, problems listed according to SFSPC-model, possible (self-management) interventions, and involvement of (healthcare) professionals.	*In general* GP Practice nurse Homecare nurse Elderly care physician Geriatric nurse *Frequently involved* Physiotherapist, occupational therapist and/or social worker	On average patients are discussed once or twice per year for 15 minutes
Individualized care plan	Definitive version of the individualized care plan is established, including (self-management) interventions discussed in multidisciplinary consultation.	Practice nurse, geriatric nurse, or homecare nurse	10 minutes
Medication review	Older persons' medication use is examined in a medication review.	GP, pharmacist, or elderly care physician	10 minutes
Multidisciplinary follow-up	Individual follow-up of patients by a multidisciplinary team of (healthcare) professionals. A case manager is responsible for coordination and evaluation of the follow-up. An elderly care physician and geriatric nurse can provide geriatric expertise.	Involvement of (healthcare) professionals based on the needs and wishes of the patient and can include, but are not limited to, practice nurses, physiotherapists, medical specialists, social workers, and so on.	4 to 10 hours

Erasmus University Rotterdam

RESULTS

Table 2 shows the background characteristics of the study population at baseline. In total, 72.4% of participants were female, 41.8% had a low educational level, and 94.4% were considered to be frail according to the TFI (mean TFI score, 7.38) in both groups. At baseline, compared with participants in the control group, older persons in the intervention group were significantly less often single ($p < 0.05$). No significant difference in mean age or the proportion of older persons with multimorbidity was observed between the groups.

Table 2 Background characteristics of older persons in the two study groups at baseline

	Care as usual ($n = 232$)	FFF approach ($n = 232$)
Characteristics		
Age	82.41 (5.16)	82.45 (5.44)
Sex (female)	168 (72.4%)	168 (72.4%)
Marital status (single)	160 (69.0%)	134 (57.8%)*
Education (low)	97 (41.8%)	97 (41.8%)
Frailty score (TFI)	7.38 (2.39)	7.38 (2.40)
Frail (TFI score ≥ 5)	219 (94.4%)	219 (94.4%)
Multimorbidity (≥ 2 conditions)	208 (89.7%)	215 (92.6%)

Values are presented as mean (SD) or number (%)
TFI: Tilburg Frailty Indicator (range, 0-15)
Independent samples *t*-tests or Chi-squared tests
*$p < 0.05$ (two-tailed)

Table 3 shows the mean QALYs (with utilities based on the EQ-5D) and mean well-being scores (SPF-ILs) at T0 and T1 using the imputed dataset. Independent samples *t*-tests showed no statistically significant differences in QALYs between groups at T0 or T1 (univariate analysis). Paired sample *t*-tests showed a statistically significant improvement in QALYs over time in the control group (Δ0.05; $p < 0.05$), but not in the intervention group (Δ0.04; $p = 0.07$). Without imputation of missing values, the data also showed a significant improvement in terms of QALYs in the intervention group over time (paired sample *t*-test, Δ0.05; $p < 0.05$). Well-being did not differ significantly at T0 or T1 between the control and intervention groups, or over time in either group. Additional file 2: Table S2 displays the mean QALYs and SPF-ILs results of the univariate analyses based on data without imputation of missing values. Analyses based on data of matched participants, i.e., pairs with complete data, yielded comparable findings; independent samples *t*-tests showed no significant differences in mean QALYs and mean well-being scores between the groups at T0 and T1 (see Additional file 3: Tables S4-S7).

Table 3 Well-being and QALYs at baseline (T0) and 12 months (T1)

		Care as usual (*n* = 232)	FFF approach (*n* = 232)
Outcome measures			
Well-being (SPF-ILs)	T0	2.62 (0.50)	2.63 (0.49)
	T1	2.67 (0.49)	2.59 (0.46)
QALYs (utilities based on EQ-5D-3L)	T0	0.66 (0.24)	0.63 (0.26)
	T1	0.71 (0.20)*[a]	0.67 (0.24)

Values are presented as mean (SD)

SPF-ILs: Social Production Function Instrument for the Level of well-being short (range, 1-4); EQ-5D-3L: five-dimensional three-level EuroQol (range for utilities, -0.33-1)

Data from univariate analyses after imputation of missing values

Paired sample *t*-tests or independent samples *t*-tests

*$p < 0.05$ (two-tailed)

[a] Significant improvement in QALYs in the control group over time based on paired data

Multilevel analyses of SPF-ILs scores adjusted for background variables and baseline values showed a small but significant difference between the intervention group and control group for well-being at follow-up, in favor of the control group (-0.09 (with imputation) and -0.10 (without imputation)). No significant differences between the groups in terms of QALYs were observed (-0.03 (with imputation) and -0.02 (without imputation)) (Additional file 4: Tables S8-S11). Regression analyses to investigate multivariable relationships among the variables yielded comparable results as the multilevel analyses (details not shown). The multilevel analyses were redone for the propensity score matched group which showed a significant difference between the groups for well-being in favor of the control group (-0.09 (with imputation) and -0.10 (without imputation)). We found no significant differences in QALYs between the intervention group and control group (-0.03 (with imputation) and -0.02 (without imputation)). For details see Additional file 5: Tables S12-S19.

For the imputed dataset, mean total costs were 7717 euros (SD, 9824 euros) in the control group and 9182 euros (SD, 11,754 euros) in the intervention group at baseline (independent samples *t*-test, *p* = 0.15; Mann-Whitney *U*-test, *U* = 28,618.50, t = 1.18, *p* = 0.24; Table 4). At 12 months, mean total costs were significantly higher in the intervention group (11,659 [SD, 14,600] euros; including intervention costs) than in the control group (8902 [SD, 11,227] euros) (independent samples *t*-test, *p* < 0.05; Table 5). In addition, differences in the median total costs at follow-up were statistically significant (Mann-Whitney *U*-test, *U* = 29,952.00, t = 2.11, *p* < 0.05). The mean total costs increased significantly over time in the intervention group (paired sample *t*-test, *p* < 0.05), but not in the control group (paired sample *t*-test, *p* = 0.14). The difference in median costs between T0 and T1 was significant in the intervention group (related-samples Wilcoxon signed rank test, t = 3.18, *p* < 0.05) and in the control group (related-samples Wilcoxon signed rank test, t = 2.34, *p* < 0.05). Based on the data without imputation of missing values, no statistically significant differences in total costs between the control and intervention groups were found at

Erasmus University Rotterdam

Table 4 Healthcare use and costs (in euros) per patient per year in the intervention and control groups at baseline

	Care as usual (*n* = 232)		FFF approach (*n* = 232)	
	Mean use[a] (SD)	Total costs (SD) in €	Mean use[a] (SD)	Total costs (SD) in €
Healthcare				
Hospital (days)	1.87 (4.90)	893.66 (2327.38)	1.76 (5.53)	845.29 (2634.78)
Consultations with the GP	3.51 (2.93)	116.43 (94.87)	3.80 (3.90)	126.25 (126.85)
Consultations out-of-hours GP	0.23 (0.67)	22.12 (63.01)	0.28 (0.70)	26.74 (65.77)
Professional homecare (hours per week)				
- Household activities (homecare or personal budget)	1.58 (1.77)	1780.95 (1967.73)	1.67 (1.76)	1887.95 (1955.74)
- Household activities (private)	0.52 (1.20)	582.04 (1320.02)	0.59 (1.80)	664.86 (1988.32)
- Personal care	0.88 (1.89)	2315.93 (4766.32)	1.04 (2.22)	2731.71 (5583.19)
- Nursing care	0.28 (1.48)	1085.97 (5512.92)	0.29 (1.57)	1108.73 (5757.76)
Care home (weeks)	0.11 (0.96)	129.16 (1128.74)	0.13 (1.37)	151.08 (1613.16)
Nursing home (weeks)	0.07 (0.71)	77.49 (838.43)	0.14 (1.09)	162.03 (1285.98)
Elderly day care (days per week)	0.03 (0.27)	213.34 (1889.62)	0.09 (0.49)	606.89 (3397.51)
Day care treatment (days per week)	0.00 (0.00)	0.0 (0.0)	0.03 (0.23)	375.02 (3269.36)
Physiotherapist (consultations)	6.77 (15.34)	224.68 (506.64)	6.75 (16.22)	224.06 (537.29)
Exercise therapist (consultations)	0.28 (1.94)	9.41 (65.98)	0.72 (5.77)	24.75 (195.62)
Medical specialist (consultations)	2.71 (4.58)	248.02 (408.44)	2.33 (3.36)	213.07 (301.12)
Social worker (sessions)	0.09 (1.03)	6.05 (66.81)	0.02 (0.21)	1.14 (13.56)
Psychologist or psychiatrist (sessions)	0.06 (0.47)	3.99 (29.69)	0.25 (1.30)	15.96 (83.10)
Assistive aids and in-home modifications				
- Wheelchair	0.03 (0.18)	1.63 (8.65)	0.04 (0.20)	2.04 (9.63)
- Alarm system	0.06 (0.25)	1.61 (6.14)	0.06 (0.23)	1.40 (5.74)
- Wheeled walker	0.09 (0.29)	1.86 (5.91)	0.13 (0.34)	2.66 (6.91)
- Stairlift	0.0 (0.0)	0.0 (0.0)	0.02 (0.15)	6.71 (45.30)
- Adjusted doorsteps	0.004 (0.07)	0.08 (1.23)	0.02 (0.13)	0.32 (2.44)
- Adjusted bathroom	0.09 (0.29)	3.31 (10.24)	0.11 (0.31)	3.76 (10.84)
Mean total costs[b]		7717.72 (9824.92)		9182.42 (11,754.75)

[a]Means (SDs) were calculated including persons without healthcare utilization; [b]Mean total costs calculated after imputation of missing healthcare costs

baseline (independent samples *t*-test, $p = 0.15$; Mann-Whitney *U*-test, $U = 17,165.50$, t = 0.60, *p* = 0.55) and 12 months (independent samples *t*-test, $p = 0.09$; Mann-Whitney *U*-test, $U = 14,375.50$, t = 1.09, *p* = 0.28). For details see Additional file 2: Table S3. In addition, univariate analyses for the propensity score matched group yielded comparable results; for the imputed dataset mean total costs were significantly higher in the intervention group compared with the control group at 12 months. Based on data without imputation of missing values, no significant differences in mean total costs between groups were observed at both time points. Univariate analyses based on data of matched participants, i.e., pairs with complete data, showed no significant difference in mean total costs between the groups at baseline and follow-up. See Additional file 3: Tables S4-S7.

Using the imputed dataset, estimated differences in effectiveness and costs were both in favor of usual care, producing an ICER of -14,788 euros per SPF-ILs point and an ICUR of -126,711 euros per QALY, indicating the FFF approach is inferior in both approaches. In Figure 2 (cost-effectiveness plane for costs versus effects in terms of well-being; SPF-ILs), 0.9% of all boot-strapped ICERs appear in the southeast quadrant (dominance; FFF approach is more effective and less costly), 78.9% appear in the northwest quadrant (inferiority; FFF intervention is more expensive and less effective), 1.5% appear in the northeast quadrant (FFF intervention is more effective, but also more expensive) and 18.7% appear in the southwest quadrant (FFF intervention is less costly, but also less effective). The probability that the FFF approach is cost-effective ranges between 0.9% and 21.1%, depending on the cost-effectiveness ratio a decision maker could apply for policy decisions. In Figure 3 (cost-effectiveness plane for costs versus effects in terms of QALYs), 9.0% of bootstrapped ICURs are located in the southeast quadrant, 54.4% appear in the northwest quadrant, 26.1% are located in the northeast quadrant, and 10.5% are located in the southeast quadrant. The probability that the FFF approach is cost-effective ranges between 9.0% and 45.6%, depending on the cost-effectiveness ratio applied.

Table 5 Healthcare use and costs (in euros) per patient per year in the intervention and control groups at 12 months

	Care as usual (n = 176)		FFF approach (n = 182)	
	Mean use[a] (SD)	Total costs (SD) in €	Mean use[a] (SD)	Total costs (SD) in €
Healthcare				
Hospital (days)	2.66 (6.92)	1275.14 (3305.59)	2.14 (9.55)	1026.89 (4546.80)
Consultations with the GP	3.57 (3.41)	118.61 (111.78)	3.83 (3.44)	127.14 (113.30)
Consultations out-of-hours GP	0.25 (0.88)	24.23 (82.89)	0.22 (0.60)	21.05 (56.49)
Professional homecare (hours per week)				
- Household activities (homecare or personal budget)	1.16 (1.50)	1302.25 (1671.23)	1.58 (1.63)	1772.57 (1816.37)
- Household activities (private)	0.76 (1.37)	850.15 (1520.05)	0.46 (1.21)	520.28 (1344.72)
- Personal care	0.92 (1.91)	2393.81 (4928.75)	1.37 (2.83)	3574.16 (7270.45)
- Nursing care	0.11 (0.50)	432.95 (1871.74)	0.31 (1.33)	1182.55 (4976.57)
Care home (weeks)	0.20 (1.27)	236.61 (1496.48)	0.17 (1.18)	197.18 (1387.62)
Nursing home (weeks)	0.09 (0.81)	102.85 (962.21)	0.28 (2.15)	330.47 (2524.32)
Elderly day care (days per week)	0.05 (0.37)	358.45 (2624.88)	0.05 (0.33)	350.48 (2361.89)
Day care treatment (days per week)	0.02 (0.23)	246.11 (3264.99)	0.04 (0.30)	641.70 (4244.80)
Physiotherapist (consultations)	6.01 (10.89)	199.39 (361.68)	8.64 (18.07)	286.74 (594.85)
Exercise therapist (consultations)	0.31 (3.48)	10.75 (118.69)	1.36 (9.63)	46.36 (327.42)
Medical specialist (consultations)	2.65 (4.07)	242.74 (371.38)	2.25 (3.17)	227.60 (287.96)
Social worker (sessions)	0.06 (0.45)	3.72 (29.43)	0.45 (4.06)	29.59 (263.51)
Psychologist or psychiatrist (sessions)	0.04 (0.38)	2.58 (24.20)	0.22 (1.20)	13.95 (76.65)
Assistive aids and in-home modifications				
- Wheelchair	0.01 (0.11)	0.54 (5.03)	0.03 (0.18)	1.57 (8.47)
- Alarm system	0.06 (0.23)	1.41 (5.78)	0.08 (0.28)	2.06 (6.87)
- Wheeled walker	0.07 (0.25)	1.40 (5.20)	0.09 (0.29)	1.93 (6.00)
- Stairlift	0.01 (0.11)	3.54 (33.09)	0.02 (0.13)	5.16 (39.74)
- Adjusted doorsteps	0.02 (0.15)	0.42 (2.79)	0.0 (0.0)	0.0 (0.0)
- Adjusted bathroom	0.09 (0.28)	2.97 (9.76)	0.03 (0.16)	0.96 (5.72)
Mean total costs[b]		8902.06 (11,227.42)		11,426.21 (14,600.79)
Mean total intervention costs		n/a		233

[a] Means (SDs) were calculated including persons without healthcare utilization; [b] Mean total costs calculated after imputation of missing healthcare costs including persons lost to follow-up between T0 and T1 (n = 50 in the intervention group and n = 56 in the control group)
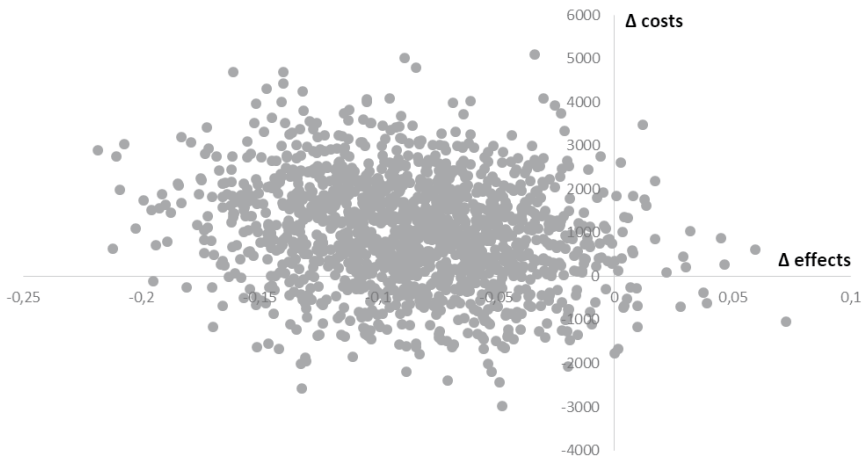
Figure 2 Cost-effectiveness plane for costs (in euros) versus effects (SPF-ILs; range, 1-4) adjusted for baseline differences; data after imputation of missing values
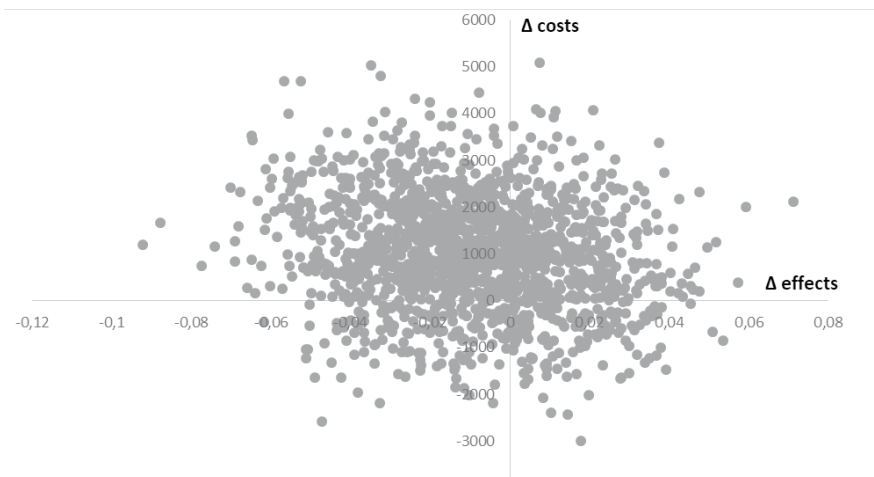
Figure 3 Cost-effectiveness plane for costs (in euros) versus effects (QALYs; range, -0.33-1) adjusted for baseline differences; data after imputation of missing values

Although different analyses (e.g., univariate, multilevel, and propensity score matched analyses, with and without imputation) showed slightly different results with respect to estimated costs and effects, the data suggest that the FFF approach is most likely not cost-effective compared with usual primary care in the Netherlands in terms of well-being and QALYs over a 12 month-period, irrespective of analytical approach and method of handling missing values.

DISCUSSION

The results of our economic evaluation indicate that proactive, integrated care for community-dwelling frail older persons as provided in the FFF program is most likely not a cost-effective initiative compared with usual primary care in the Netherlands, in terms of well-being and QALYs over a 12-month period. Our results are in line with outcomes of other studies investigating the cost-effectiveness of integrated care for frail older persons in the primary care setting in the Netherlands (e.g. Hoogendijk, 2016). The comparability of integrated care programs and evaluation studies is limited due to differences in study populations, interventions and outcomes (Looman et al., 2018).

One explanation for the lack of effect may be the conceivably small difference between the FFF approach and usual primary care services in the Netherlands. There are indications that reforms in the primary care system in the Netherlands resulted in developments in the control GP practices to improve their care delivery. Although these practices did not provide care and support according to the FFF approach, several control GP practices implemented interventions, such as systematic follow-up of older adults and multidisciplinary consultation, during the study period (Vestjens et al., 2018b). In addition, the lack of effectiveness of complex interventions may be partly due to failure to (fully) implement the programs as intended (Øvretveit & Gustafson, 2002). Indeed, we have found suboptimal implementation of intervention components in GP practices organizing care according to the FFF approach (Vestjens et al., 2018b). Most interventions, and especially complex care programs like the FFF approach, require extensive time and effort to achieve full implementation (Faes, Reelick, Esselink, & Rikkert, 2010; van de Wetering, Olde Rikkert, van der Wilt, & Adang, 2014). We noted differences among intervention GP practices with respect to the implementation and execution of the FFF program, including differences in the selection of older persons for proactive screening, the (number of) professionals involved in screening procedures, the organization of multidisciplinary consultations (e.g., frequency, number of patients discussed, (type of) professionals involved), and (the organization) of long-term follow-up of frail older persons (Vestjens et al., 2018b). These differences may obscure the added value of the FFF approach in terms of QALYs and well-being. Analyses based on matched participants of intervention GP practices with a high degree of implementation of (FFF-related) interventions (i.e., practices that implemented more interventions than average) (Vestjens et al., 2018b), showed that the mean SPF-ILs score was higher, indicating greater subjective well-being, compared with participants in other intervention GP practices (Additional file 6: Tables S20-S23). Therefore, the degree of implementation may have an effect on effectiveness of complex interventions like the FFF approach. For a detailed description of implemented (FFF-related) interventions in the GP practices see Vestjens, Cramm and Nieboer (2018b). Even with optimal implementation of such interventions, clinically meaningful improvement in outcomes is not guaranteed in the short term (van de Wetering et al., 2014). The length of the study period,

being 12 months, may have been too short to detect improvements in older persons' outcomes (Eklund & Wilhelmson, 2009). Especially in the short term, variations in costs and effects can be expected (van de Wetering, Woertman, & Adang, 2012). Patterns of healthcare utilization show, for example, a substantial increase in primary and hospital care utilization in frail older persons (Ilinca & Calciolari, 2015). Consequently, the identification of frailty and introduction of interventions to postpone or prevent a decline into worse health states (Lang, Michel, & Zekry, 2009) may result in higher healthcare costs in the short term, but might reduce use of more expensive healthcare services and adverse outcomes in the long term (Ilinca & Calciolari, 2015). Another explanation might be related to the heterogeneity of the population of older persons considered to be frail. No consensus has been reached about the conceptualization and measurement of frailty in older persons. Major approaches include the frailty phenotype, which focuses on physical aspects of frailty (Fried et al., 2001; Fried, Ferrucci, Darer, Williamson, & Anderson, 2004), and a multidimensional approach to frailty including, for example, physical, social, and psychological factors (Gobbens, van Assen, Luijkx, & Schols, 2012). Although we used a multidimensional approach to assess frailty in this study, Looman and colleagues (2018) showed that distinction among domains of frailty does not fully capture its complexity. The TFI (Gobbens et al., 2010), which we used to measure (the degree of) frailty in older persons, does not discern among types of underlying problems in these domains or weigh different domains (Looman et al., 2018). Researchers have suggested that the heterogeneity of frailty should be taken into account in the evaluation of integrated care programs (Looman et al., 2018), especially to better understand how interventions can be optimally aligned with different well-being needs of frail older persons (Nieboer & Cramm, 2018).

Strengths and limitations

One strength of this study is that we measured the subjective well-being of community-dwelling frail older persons along with health-related quality of life. QALY measures in economic evaluations are based predominantly on aspects of health-related quality of life alone. Care programs for older persons may also aim to improve non-health related domains of quality of life. Thus, the sole use of health-related quality of life measures in economic evaluations may not be appropriate, as it may not capture broader benefits of such interventions beyond health (Makai, Brouwer, Koopmanschap, Stolk, & Nieboer, 2014). Consequently, Makai and colleagues (2014) recommended the inclusion of well-being measures with health measures like the EQ-5D in economic evaluations of care programs for older persons. We did so, although the different perspectives did not lead to different recommendations regarding the preference of the FFF intervention. Another strength of our study is the quality of the data gathered. We used dedicated, trained interviewers who collected the data in face-to-face interviews during home visits. All interviewers lived in the western North Brabant Province, assuring a cultural fit, and had backgrounds in healthcare. Moreover, we used a detailed resource use questionnaire covering a wide range of healthcare categories to assess healthcare utilization at the individual level. We

included care disciplines that are frequently not included in studies, such as paramedical (e.g., physiotherapy) and psychological care, which may have increased content validity. Our study also has several potential limitations. First, we used a quasi-experimental design, which is more susceptible to bias due to the absence of randomization (Eccles, Grimshaw, Campbell, & Ramsay, 2003). To increase comparability of the intervention and control groups, we used one-to-one matching based on key covariables. Despite this effort, the control group contained significantly more single persons than did the intervention group. Moreover, we noted indications (based on interviews with healthcare professionals and (project) managers) of a strong motivation to organize care and support for the elderly population in some control GP practices. Professionals in these practices may have perceived that the FFF program would not add value to their usual care practices and were therefore perhaps especially eager to participate in the control group. Second, recall bias might have occurred due to the retrospective assessment of service use in the preceding 12 months. Under-reporting and over-reporting of effects have been found in previous research in which health service utilization was assessed retrospectively (Brusco & Watts, 2015). Unfortunately, we were not able to include administrative or registry data to complement the reported healthcare service use. Nonetheless, given the same data collection procedure in both groups, we have no indication that recall bias varied significantly between the intervention and control groups. Third, mean standard costs of the FFF program were estimated, instead of assessing intervention costs for individual participants. We attempted to avoid duplicate inclusion of costs by including service use related to the follow-up of older patients in the FFF context only in healthcare costs, and not in intervention costs. The implementation and execution of (elements of) the FFF approach differed among intervention GP practices (Vestjens et al., 2018b). However, results of sensitivity analyses in which intervention costs were varied to test the robustness of the estimated ICER and ICUR did not affect the overall recommendation regarding the preference of the FFF program. Fourth, despite recommendations (Hakkaart-van Roijen et al., 2015), we were unable to collect data on informal care due to practical considerations. The impact of informal care costs on the mean total costs in the intervention and control groups remains unknown, although we found no indication (based on interviews with healthcare professionals, (project) managers, and frail older persons) of unequal distribution of informal care costs between groups. In addition, we did not account for medication costs in either group or intervention training and implementation costs in the FFF group. We have noted no indication that medication use differed between groups.

Conclusions

Our study findings add to the current unconvincing body of evidence with respect to the cost-effectiveness of integrated primary care aimed at community-dwelling frail older persons. Future economic evaluations should use sufficiently long follow-up periods to assess durable costs and effects, adopt a societal perspective, and take into account the degree of implementation and the

target population. Continued effort is required to unravel the black box of integrated care and find (cost-)effective (components of) programs for community-dwelling frail older persons.

REFERENCES

Banerjee, S. (2015). Multimorbidity--older adults need health care that can count past one. *Lancet, 385*(9968), 587-589.

Blom, J. W., Van den Hout, W. B., Den Elzen, W. P. J., Drewes, Y. M., Bleijenberg, N., Fabbricotti, I. N., . . . TOPICS-MDS research consortium. (2018). Effectiveness and cost-effectiveness of proactive and multidisciplinary integrated care for older people with complex problems in general practice: An individual participant data meta-analysis. *Age and Ageing, 47*(5), 705-714.

Boeckxstaens, P., & De Graaf, P. (2011). Primary care and care for older persons: Position paper of the european forum for primary care. *Quality in Primary Care, 19*(6), 369-389.

Boyd, C. M., Darer, J., Boult, C., Fried, L. P., Boult, L., & Wu, A. W. (2005). Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: Implications for pay for performance. *Jama, 294*(6), 716-724.

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy, 37*(1), 53-72.

Brusco, N. K., & Watts, J. J. (2015). Empirical evidence of recall bias for primary health care visits. *BMC Health Services Research, 15*, 381-015-1039-1.

Chatterji, S., Byles, J., Cutler, D., Seeman, T., & Verdes, E. (2015). Health, functioning, and disability in older adults--present status and future implications. *Lancet, 385*(9967), 563-575.

de Bruin, S. R., Versnel, N., Lemmens, L. C., Molema, C. C., Schellevis, F. G., Nijpels, G., & Baan, C. A. (2012). Comprehensive care programs for patients with multiple chronic conditions: A systematic literature review. *Health Policy (Amsterdam, Netherlands), 107*(2-3), 108-145.

De Lepeleire, J., Iliffe, S., Mann, E., & Degryse, J. M. (2009). Frailty: An emerging concept for general practice. *The British Journal of General Practice, 59*(562), e177-82.

Eccles, M., Grimshaw, J., Campbell, M., & Ramsay, C. (2003). Research designs for studies evaluating the effectiveness of change and improvement strategies. *Quality & Safety in Health Care, 12*(1), 47-52.

Eklund, K., & Wilhelmson, K. (2009). Outcomes of coordinated and integrated interventions targeting frail elderly people: A systematic review of randomised controlled trials. *Health & Social Care in the Community, 17*(5), 447-458.

EuroQol Group. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199-208.

Faes, M. C., Reelick, M. F., Esselink, R. A., & Rikkert, M. G. (2010). Developing and evaluating complex healthcare interventions in geriatrics: The use of the medical research council framework exemplified on a complex fall prevention intervention. *Journal of the American Geriatrics Society, 58*(11), 2212-2221.

Fried, L. P., Ferrucci, L., Darer, J., Williamson, J. D., & Anderson, G. (2004). Untangling the concepts of disability, frailty, and comorbidity: Implications for improved targeting and care. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 59*(3), 255-263.

Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., . . . McBurnie, M. A. (2001). Frailty in older adults: Evidence for a phenotype. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 56*(3), M146-156.

Gobbens, R. J., van Assen, M. A., Luijkx, K. G., & Schols, J. M. (2012). Testing an integral conceptual model of frailty. *Journal of Advanced Nursing, 68*(9), 2047-2060.

Gobbens, R. J., van Assen, M. A. L. M., Luijkx, K. G., Wijnen-Sponselee, M. T., & Schols, J. M. G. A. (2010). The tilburg frailty indicator: Psychometric properties. *Journal of the American Medical Directors Association, 11*(5), 344-355.

Guthrie, B., Payne, K., Alderson, P., McMurdo, M. E., & Mercer, S. W. (2012). Adapting clinical guidelines to take account of multimorbidity. *BMJ, 345*, e6341.

Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T., & Tan, S. S. (2015). *Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg.* Rotterdam: Erasmus University Rotterdam.

Harvey, G., Dollard, J., Marshall, A., & Mittinty, M. M. (2018). Achieving integrated care for older people: Shuffling the deckchairs or making the system watertight for the future? *International Journal of Health Policy and Management, 7*(4), 290-293.

Hoogendijk, E. O. (2016). How effective is integrated care for community-dwelling frail older people? The case of the Netherlands. *Age and Ageing, 45*(5), 585-588.

Hopman, P., de Bruin, S. R., Forjaz, M. J., Rodriguez-Blazquez, C., Tonnara, G., Lemmens, L. C., . . . Rijken, M. (2016). Effectiveness of comprehensive care programs for patients with multiple chronic conditions or frailty: A systematic literature review. *Health Policy, 120*(7), 818-832.

Ilinca, S., & Calciolari, S. (2015). The patterns of health care utilization by elderly Europeans: Frailty and its implications for health systems. *Health Services Research, 50*(1), 305-320.

Indurkhya, A., Mitra, N., & Schrag, D. (2006). Using propensity scores to estimate the cost-effectiveness of medical therapies. *Statistics in Medicine, 25*(9), 1561-1576.

Koopmans, R. T. C. M., Lavrijsen, J. C. M., Hoek, J. F., Went, P. B. M., & Schols, J. M. G. A. (2010). Dutch elderly care physician: A new generation of nursing home physician specialists. *Journal of the American Geriatrics Society, 58*(9), 1807-1809.

Kringos, D., Boerma, W., Hutchinson, A., & Saltman, R. (2015). *Building primary care in a changing Europe.* Copenhagen: World Health Organization.

Kroneman, M., Boerma, W., van den Berg, M., Groenewegen, P., de Jong, J., & van Ginneken, E. (2016). The Netherlands: Health system review. *Health Systems in Transition, 18*(2), 1-240.

Lamers, L. M., Stalmeier, P. F. M., McDonnell, J., Krabbe, P. F. M., & van Busschbach, J. J. (2005). Measuring the quality of life in economic evaluations: The dutch EQ-5D tariff. [Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D-tarief] *Nederlands Tijdschrift voor Geneeskunde, 149*(28), 1574-1578.

Lang, P. O., Michel, J. P., & Zekry, D. (2009). Frailty syndrome: A transitional state in a dynamic process. *Gerontology, 55*(5), 539-549.

Looman, W. M., Fabbricotti, I. N., Blom, J. W., Jansen, A. P. D., Lutomski, J. E., Metzelthin, S. F., . . . TOPICS-MDS Research Consortium. (2018). The frail older person does not exist: Development of frailty profiles with latent class analysis. *BMC Geriatrics, 18*(1), 84-018-0776-5.

Looman, W. M., Huijsman, R., & Fabbricotti, I. N. (2018). The (cost-)effectiveness of preventive, integrated care for community-dwelling frail older people: A systematic review. *Health & Social Care in the Community, 27*(1), 1-30.

Low, L. F., Yap, M., & Brodaty, H. (2011). A systematic review of different models of home and community care services for older persons. *BMC Health Services Research, 11*, 93.

Makai, P., Brouwer, W. B., Koopmanschap, M. A., & Nieboer, A. P. (2012). Capabilities and quality of life in dutch psycho-geriatric nursing homes: An exploratory study using a proxy version of the ICECAP-O. *Quality of Life Research, 21*(5), 801-812.

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 & Medicine, 102*, 83-93.

Mattke, S., Seid, M., & Ma, S. (2007). Evidence for the effect of disease management: Is $1 billion a year a good investment? *The American Journal of Managed Care, 13*(12), 670-676.

Nieboer, A. P., & Cramm, J. M. (2018). How do older people achieve well-being? validation of the social production function instrument for the level of well-being-short (SPF-ILs). *Social Science & Medicine, 211*, 304-313.

Nieboer, A. P., Lindenberg, S., Boomsma, A., & van Bruggen, A. C. (2005). Dimensions of well-being and their measurement: The SPF-IL scale. *Social Indicators Research, 73*(3), 313-353.

Nolte, E., & McKee, M. (2008). *Caring for people with chronic conditions: A health system perspective*. Maidenhead: Open University Press.

Ormel, J., Lindenberg, S., Steverink, N., & Verbrugge, L. M. (1999). Subjective well-being and social production functions. *Social Indicators Research, 46*(1), 61-90.

Ormel, J., Lindenberg, S., Steverink, N., & Vonkorff, M. (1997). Quality of life and social production functions: A framework for understanding health effects. *Social Science & Medicine, 45*(7), 1051-1063.

Øvretveit, J., & Gustafson, D. (2002). Evaluation of quality improvement programmes. *Quality & Safety in Health Care, 11*(3), 270-275.

Rochat, S., Cumming, R. G., Blyth, F., Creasey, H., Handelsman, D., Le Couteur, D. G., . . . Waite, L. (2010). Frailty and use of health and community services by community-dwelling older men: The concord health and ageing in men project. *Age and Ageing, 39*(2), 228-233.

Rosenbaum, P. R., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Schäfer, W., Kroneman, M., Boerma, W., van den Berg, M., Westert, G., Devillé, W., & van Ginneken, E. (2010). The Netherlands: Health system review. *Health Systems in Transition, 12*(1), 1-229.

Smith, S. M., Wallace, E., O'Dowd, T., & Fortin, M. (2016). Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. *Cochrane Database of Systematic Reviews.*

van de Wetering, G., Olde Rikkert, M., van der Wilt, G. J., & Adang, E. (2014). Understanding and anticipating lag-time bias in cost-effectiveness studies: The role of time in cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care, 30*(6), 608-611.

van de Wetering, G., Woertman, W. H., & Adang, E. M. (2012). Time to incorporate time in cost-effectiveness analysis. *The European Journal of Health Economics, 13*(3), 223-226.

van Dijk, H. M., Cramm, J. M., Lötters, F. J. B., van Exel, J., & Nieboer, A. P. (2015). *Even buurten: De complexiteit van een wijkgerichte aanpak*. Rotterdam: Erasmus Universiteit Rotterdam.

van Weel, C., & Schellevis, F. G. (2006). Comorbidity and guidelines: Conflicting interests. *Lancet, 367*(9510), 550-551.

Vestjens, L., Cramm, J. M., Birnie, E., & Nieboer, A. P. (2018a). Evaluating an integrated primary care approach to improve well-being among frail community-living older people: A theory-guided study protocol. *BMC Geriatrics, 18*(1), 173.

Vestjens, L., Cramm, J. M., & Nieboer, A. P. (2018b). An integrated primary care approach for frail community-dwelling older persons: A step forward in improving the quality of care. *BMC Health Services Research, 18*(1), 28.

Vestjens, L., Cramm, J. M., Nieboer, A. P. (2019). Quality of primary care delivery and productive interactions among community-living frail older persons and their general practitioners and practice nurses. *BMC Health Services Research, 19*(1), 496.

Wagner, E. H., Bennett, S. M., Austin, B. T., Greene, S. M., Schaefer, J. K., & Vonkorff, M. (2005). Finding common ground: Patient-centeredness and evidence-based chronic illness care. *Journal of Alternative and Complementary Medicine, 11*, S7-15.

West, B. T., Welch, K. B., & Gatecki, A. T. (2015). *Linear mixed models. A practical guide using statistical software* (Second edition ed.). Boca Raton: CRC press/Taylor & Francis Group.

WHO (2015). *WHO global strategy on people-centred and integrated health services. Interim report.* Geneva: World Health Organization.

APPENDIX

Additional file 1. Overview of missing data

Table S1 Number of participants with missing data on the EQ-5D-3L, SPF-ILs and resource use questionnaire at T0 and T1 (total n = 464)

	T0			T1		
	EQ-5D-3L	SPF-ILs	Resource use	EQ-5D-3L	SPF-ILs	Resource use
Data completely missing[a]	3	4	0	108[c]	108[c]	106[c]
Data partially missing[b]	1	4	100	0	1	30

SPF-ILs: Social Production Function Instrument for the Level of well-being short; EQ-5D-3L: five-dimensional three-level EuroQol. [a]Data completely missing = none of the items in the questionnaires are answered. [b]Data partially missing EQ-5D-3L = 1-4 missings of the 5 questions on the EQ-5D-3L; Data partially missing SPF-ILs = 6-14 missings of the 15 questions on the SPF-ILs; Data partially missing resource use = 1-21 missings of the 22 questions on the resource use instrument. [c]Including persons lost to follow-up between T0 and T1 (total n = 106, see Fig. 1)

Additional file 2. Outcomes and cost estimates without imputation of missing values

Table S2 Well-being and QALYs at baseline (T0) and 12 months (T1) *without* data imputation

		Care as usual	n	FFF approach	n
Outcome measures					
Well-being (SPF-ILs)	T0	2.62 (0.50)	230	2.63 (0.50)	226
	T1	2.71 (0.53)	176	2.60 (0.50)	179
QALYs	T0	0.66 (0.25)	230	0.63 (0.26)	230
(utilities based on EQ-5D-3L)	T1	0.72 (0.21)*[a]	176	0.70 (0.25)*[a]	180

Values are presented as mean (SD)
SPF-ILs: Social Production Function Instrument for the Level of well-being short (range, 1-4); EQ-5D-3L: five-dimensional three-level EuroQol (range for utilities, -0.33-1)
Data from univariate analyses *without* imputation of missing values
*$p < 0.05$ (two-tailed)
Paired sample t-tests or independent samples t-tests
[a] Significant improvement in QALYs in the control group and intervention group over time based on paired data

Table S3 Healthcare costs (in euros) at baseline (T0) and 12 months (T1) *without* data imputation

	Care as usual	n	FFF approach	n
Healthcare costs				
Mean total costs at T0[a]	7023.18 (9499.67)	180	8635.90 (11,735.34)	184
Mean total costs at T1[a]	7798.13 (10,442.08)	160	10,209.36 (14,598.36)	168

[a]Means (SDs) were calculated including persons without healthcare utilization

Additional file 3. Analyses based on matched participants

Table S4 Analyses based on matched participants, i.e., pairs with complete data on EQ-5D-3L and resource use, at T0 (*n* = 146 pairs) and at T1 (*n* = 111 pairs)

		Care as usual	FFF approach
QALYs (utilities based on EQ-5D-3L)	T0	0.66 (0.24)	0.63 (0.24)
	T1	0.74 (0.20)	0.71 (0.23)
Mean total costs[a] (resource use questionnaire)	T0	6742.44 (9958.21)	8829.61 (11,594.37)
	T1	8330.41 (11,368.95)	9029.57 (12,942.87)

Values are presented as mean (SD); [a]Means (SDs) were calculated including persons without healthcare utilization
EQ-5D-3L: five-dimensional three-level EuroQol (range for utilities, -0.33-1)
*p < 0.05 (two-tailed)
Independent samples *t*-tests or Mann-Whitney *U*-tests

Table S5 Analyses based on matched participants, i.e., pairs with complete data on SPF-ILs and resource use, at T0 (*n* = 145 pairs) and T1 (*n* = 111 pairs)

		Care as usual	FFF approach
Well-being (SPF-ILs)	T0	2.64 (0.49)	2.64 (0.47)
	T1	2.75 (0.53)	2.62 (0.53)
Mean total costs[a] (resource use questionnaire)	T0	6878.86 (10,038.26)	8651.38 (11,287.80)
	T1	8330.41 (11,368.95)	9029.57 (12,942.87)

Values are presented as mean (SD); [a]Means (SDs) were calculated including persons without healthcare utilization
SPF-ILs: Social Production Function Instrument for the Level of well-being short (range, 1-4)
*p < 0.05 (two-tailed)
Independent samples *t*-tests or Mann-Whitney *U*-tests

Table S6 Analyses based on matched participants, i.e., pairs with complete data on EQ-5D-3L and resource use on ***both*** T0 and T1 (*n* = 71 pairs)

		Care as usual	FFF approach
QALYs (utilities based on EQ-5D-3L)	T0	0.68 (0.24)	0.68 (0.22)
	T1	0.74 (0.21)	0.73 (0.21)
Mean total costs[a] (resource use questionnaire)	T0	6071.65 (9243.57)	7435.85 (9709.72)
	T1	7848.30 (11,386.60)	9038.16 (13,728.47)

Values are presented as mean (SD); [a]Means (SDs) were calculated including persons without healthcare utilization
EQ-5D-3L: five-dimensional three-level EuroQol (range for utilities, -0.33-1)
*p < 0.05 (two-tailed)
Independent samples *t*-tests or Mann-Whitney *U*-tests

Table S7 Analyses based on matched participants, i.e., pairs with complete data on SPF-ILs and resource use on **both** T0 and T1 (*n* = 70 pairs)

		Care as usual	FFF approach
Well-being (SPF-ILs)	T0	2.74 (0.49)	2.68 (0.48)
	T1	2.78 (0.52)	2.65 (0.53)
Mean total costs[a]	T0	6107.91 (9305.23)	7508.36 (9761.75)
(resource use questionnaire)	T1	7952.83 (11,434.45)	9135.14 (13,803.08)

Values are presented as mean (SD); [a]Means (SDs) were calculated including persons without healthcare utilization
SPF-ILs: Social Production Function Instrument for the Level of well-being short (range, 1-4)
*$p < 0.05$ (two-tailed)
Independent samples *t*-tests or Mann-Whitney *U*-tests

Additional file 4. Multilevel analyses

Table S8 Multilevel analysis of well-being (SPF-ILs) at T1, after data imputation (*n* = 464)

	B	SE
Constant	2.20***	0.31
Intervention group	-0.09**	0.04
Age	-0.01	0.003
Sex (female)	0.03	0.04
Educational level (low)	-0.03	0.04
Marital status (single)	-0.07	0.04
Frailty score	-0.03***	0.01
Multimorbidity	-0.02	0.05
Well-being at T0	0.48***	0.04

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
Analysis shows the impact of intervention group, adjusted for age, sex, educational level, marital status, frailty score, multimorbidity, well-being at T0 and GP practice (random effect)

Table S9 Multilevel analysis of well-being (SPF-ILs) at T1, **without** data imputation (*n* = 354)

	B	SE
Constant	2.28***	0.44
Intervention group	-0.10*	0.05
Age	-0.01	0.01
Sex (female)	0.04	0.06
Educational level (low)	-0.03	0.05
Marital status (single)	-0.08	0.05
Frailty score	-0.04***	0.01
Multimorbidity	-0.03	0.08
Well-being at T0	0.47***	0.05

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
Analysis shows the impact of intervention group, adjusted for age, sex, educational level, marital status, frailty score, multimorbidity, well-being at T0 and GP practice (random effect)

Table S10 Multilevel analysis of QALYs at T1, after data imputation (*n* = 464)

	B	SE
Constant	1.07***	0.16
Intervention group	-0.03	0.02
Age	-0.01**	0.002
Sex (female)	-0.04	0.02
Educational level (low)	0.04*	0.02
Marital status (single)	0.02	0.02
Frailty score	-0.01*	0.01
Multimorbidity	-0.02	0.03
QALYs at T0	0.27***	0.04

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
Analysis shows the impact of intervention group, adjusted for age, sex, educational level, marital status, frailty score, multimorbidity, QALYs at T0 and GP practice (random effect)

Table S11 Multilevel analysis of QALYs at T1, ***without*** data imputation (*n* = 355)

	B	SE
Constant	0.95***	0.19
Intervention group	-0.02	0.02
Age	-0.004	0.002
Sex (female)	-0.06*	0.03
Educational level (low)	0.04	0.02
Marital status (single)	0.03	0.03
Frailty score	-0.01*	0.01
Multimorbidity	-0.04	0.04
QALYs at T0	0.35***	0.05

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
Analysis shows the impact of intervention group, adjusted for age, sex, educational level, marital status, frailty score, multimorbidity, QALYs at T0 and GP practice (random effect)

Additional file 5. Propensity score matching

Table S12 Multilevel analysis well-being, using propensity score matching (Model 1)[a] and data imputation ($n = 463$)

	B	SE
Constant	2.19***	0.31
Intervention group	-0.09*	0.04
Age	-0.01	0.004
Sex (female)	0.03	0.04
Educational level (low)	-0.03	0.04
Marital status (single)	-0.07	0.04
Frailty score	-0.03***	0.01
Multimorbidity	-0.02	0.05
Well-being at T0 (SPF-ILs)	0.48***	0.04

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
[a]Logistic regression model 1 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity

Table S13 Multilevel analysis well-being, using propensity score matching (Model 2)[a] and data imputation ($n = 459$)

	B	SE
Constant	2.16***	0.32
Intervention group	-0.09*	0.04
Age	-0.01	0.004
Sex (female)	0.04	0.04
Educational level (low)	-0.03	0.04
Marital status (single)	-0.08	0.04
Frailty score	-0.03***	0.01
Multimorbidity	-0.02	0.05
Well-being at T0 (SPF-ILs)	0.49***	0.04

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
[a]Logistic regression model 2 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity, and baseline SPF-ILs, QALYs and costs

Table S14 Multilevel analysis well-being, using propensity score matching (Model 1)[a] and ***without*** data imputation (*n* = 353)

	B	SE
Constant	2.27***	0.44
Intervention group	-0.10*	0.05
Age	-0.01	0.01
Sex (female)	0.04	0.06
Educational level (low)	-0.03	0.05
Marital status (single)	-0.08	0.05
Frailty score	-0.04***	0.01
Multimorbidity	-0.03	0.08
Well-being at T0 (SPF-ILs)	0.47***	0.05

SE, standard error; *p < 0.05 (two-tailed); **p < 0.01 (two-tailed); ***p < 0.001 (two-tailed)
[a]Logistic regression model 1 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity

Table S15 Multilevel analysis well-being, using propensity score matching (Model 2)[a] and ***without*** data imputation (*n* = 349)

	B	SE
Constant	2.22***	0.44
Intervention group	-0.10*	0.05
Age	-0.01	0.01
Sex (female)	0.05	0.06
Educational level (low)	-0.03	0.05
Marital status (single)	-0.09	0.05
Frailty score	-0.04***	0.01
Multimorbidity	-0.03	0.08
Well-being at T0 (SPF-ILs)	0.49***	0.05

SE, standard error; *p < 0.05 (two-tailed); **p < 0.01 (two-tailed); ***p < 0.001 (two-tailed)
[a]Logistic regression model 2 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity, and baseline SPF-ILs, QALYs and costs

Table S16 Multilevel analysis QALYs, using propensity score matching (Model 1)[a] and data imputation (*n* = 463)

	B	SE
Constant	1.07***	0.16
Intervention group	-0.03	0.02
Age	-0.01**	0.002
Sex (female)	-0.04	0.02
Educational level (low)	0.04*	0.02
Marital status (single)	0.02	0.02
Frailty score	-0.01*	0.01
Multimorbidity	-0.02	0.03
QALYs at T0 (utilities based on EQ-5D-3L)	0.26***	0.04

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
[a]Logistic regression model 1 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity

Table S17 Multilevel analysis QALYs, using propensity score matching (Model 2)[a] and data imputation (*n* = 459)

	B	SE
Constant	1.05***	0.16
Intervention group	-0.03	0.02
Age	-0.01**	0.002
Sex (female)	-0.05*	0.02
Educational level (low)	0.04*	0.02
Marital status (single)	0.02	0.02
Frailty score	-0.01*	0.004
Multimorbidity	-0.02	0.03
QALYs at T0 (utilities based on EQ-5D-3L)	0.25***	0.04

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
[a]Logistic regression model 2 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity, and baseline SPF-ILs, QALYs and costs

Table S18 Multilevel analysis QALYs, using propensity score matching (Model 1)[a] and *without* data imputation (*n* = 354)

	B	SE
Constant	0.95***	0.19
Intervention group	-0.02	0.02
Age	-0.004	0.002
Sex (female)	-0.06*	0.03
Educational level (low)	0.04	0.02
Marital status (single)	0.03	0.03
Frailty score	-0.01*	0.01
Multimorbidity	-0.04	0.04
QALYs at T0 (utilities based on EQ-5D-3L)	0.35***	0.05

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
[a]Logistic regression model 1 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity

Table S19 Multilevel analysis QALYs, using propensity score matching (Model 2)[a] and *without* data imputation (*n* = 351)

	B	SE
Constant	0.91***	0.19
Intervention group	-0.02	0.02
Age	-0.004	0.002
Sex (female)	-0.06*	0.03
Educational level (low)	0.04	0.02
Marital status (single)	0.03	0.03
Frailty score	-0.009	0.01
Multimorbidity	-0.04	0.04
QALYs at T0 (utilities based on EQ-5D-3L)	0.34***	0.05

SE, standard error; *$p < 0.05$ (two-tailed); **$p < 0.01$ (two-tailed); ***$p < 0.001$ (two-tailed)
[a]Logistic regression model 2 with covariates: age, sex, marital status, educational level, frailty score, and multimorbidity, and baseline SPF-ILs, QALYs and costs

Additional file 6. Analyses based on subgroup of intervention GP practices with high degree of implementation

Table S20 Analyses of participants of intervention GP practices with high degree of implementation and matched controls *without* data imputation

		Care as usual	*n*	FFF approach	*n*
Outcome measures					
Well-being (SPF-ILs)	T0	2.62 (0.45)	92	2.65 (0.48)	91
	T1	2.70 (0.53)	68	2.70 (0.53)	79
QALYs	T0	0.66 (0.24)	92	0.64 (0.28)	93
(utilities based on EQ-5D-3L)	T1	0.70 (0.23)	68	0.70 (0.25)	80

Values are presented as mean (SD)
SPF-ILs: Social Production Function Instrument for the Level of well-being short (range, 1-4); EQ-5D-3L: five-dimensional three-level EuroQol (range for utilities, -0.33-1)
*$p < 0.05$ (two-tailed)
Independent samples *t*-tests

Table S21 Healthcare costs (in euros) of participants of intervention GP practices with high degree of implementation and matched controls *without* data imputation

	Care as usual	*n*	FFF approach	*n*
Healthcare costs				
Mean total costs at T0[a]	8085.43 (12,226.46)	70	7736.96 (10,915.11)	80
Mean total costs at T1[a]	8989.07 (12,264.61)	65	9942.37 (14,985.58)	76

[a]Means (SDs) were calculated including persons without healthcare utilization

Table S22 Analyses of participants of intervention GP practices with high degree of implementation and matched controls, after data imputation

		Care as usual	*n*	FFF approach	*n*
Outcome measures					
Well-being (SPF-ILs)	T0	2.62 (0.45)	93	2.65 (0.48)	93
	T1	2.67 (0.48)	93	2.67 (0.51)	93
QALYs	T0	0.66 (0.23)	93	0.64 (0.28)	93
(utilities based on EQ-5D-3L)	T1	0.70 (0.21)	93	0.68 (0.25)	93

Values are presented as mean (SD)
SPF-ILs: Social Production Function Instrument for the Level of well-being short (range, 1-4); EQ-5D-3L: five-dimensional three-level EuroQol (range for utilities, -0.33-1)
*$p < 0.05$ (two-tailed)
Independent samples *t*-tests

Table S23 Healthcare costs (in euros) of participants of intervention GP practices with high degree of implementation and matched controls, after data imputation

	Care as usual	*n*	FFF approach	*n*
Healthcare costs				
Mean total costs at T0[a]	8940.79 (11,902.79)	93	8804.80 (12,128.38)	93
Mean total costs at T1[a]	10,337.77 (12,419.17)	93	10,060.00 (13,858.17)	93

[a]Means (SDs) were calculated including persons without healthcare utilization