Received: 8 March 2021 | Accepted: 5 November 2021

DOI: 10.1111/aphw.12322

ORIGINAL ARTICLE

Effectiveness and equity evaluation of an insurance-wide telephone-counseling program for self-management of chronic diseases: The Health Coach Study

Babette Renneberg[1] | **Julian Schulze**[2] | **Sylvia Böhme**[1] | **Stephen G. West**[3] | **Benjamin Schüz**[4]

[1]Department of Clinical Psychology and Psychotherapy, Freie Universität Berlin, Berlin, Germany

[2]Division of Psychological Assessment, Differential and Personality Psychology, Freie Universität Berlin, Berlin, Germany

[3]Psychology Department, Arizona State University, Tempe, Arizona, USA

[4]Institute for Public Health and Nursing Research, University of Bremen, Bremen, Germany

Correspondence
Benjamin Schüz, Institute for Public Health and Nursing Research, University of Bremen, Bremen, Germany.
Email: benjamin.schuez@uni-bremen.de

Funding information
Techniker Krankenkasse

Abstract

Trajectories of chronic illnesses depend on patient socioeconomic status (SES). This study examines main and equity effects (age, gender, education, region of residence) of a brief telephone self-management intervention on self-rated health and depressive symptoms of health insurance clients with chronic illnesses. Randomized invitation design ($n = 2628$) with predominantly male (82%) older individuals (modal age = 65–74) with one or more chronic illnesses. Primary outcomes: Self-rated health and depressive symptoms. Intervention: Brief CBT-based telephone counseling. Propensity score matching was used to equate intervention and control groups ($n = 1314$ pairs). Change score models were used to analyze changes in health-related outcome measures. The intervention resulted in improvements in self-rated health ($d = .37$) and fewer depressive symptoms ($d = .17$) over 4 and 6 months. There were comparable effects across education and regions, but younger and female participants profited more from the intervention compared with older and male participants. A brief telephone-based intervention

led to improved self-rated health and well-being in a large sample of participants with chronic health conditions. This effect was observed over and above regular medical care. The intervention was equitable with respect to education and region, but not age and gender.

KEYWORDS

depression, equity effects, propensity score matching, self-rated health, telephone counseling

INTRODUCTION

The rapid increase of chronic illnesses in industrialized and developing countries is a major public health challenge. For example, in the United States, seven out of the 10 most common causes of death for the entire adult population are chronic diseases (Centers for Disease Control and Prevention, 2020): diseases of the heart, malignant neoplasms, chronic lower respiratory disease, cerebrovascular diseases, and diabetes mellitus. A similar picture emerges globally, with six out of 10 leading causes of death being chronic diseases (GBD 2019 Diseases and Injuries Collaborators, 2020; Vos et al., 2017). In Germany, the setting of the current study, nine out of the 10 most common causes of death, both in women and men, are chronic diseases (Destatis, 2021). The most frequent chronic diseases include hypertension, hyperlipidemia, chronic back pain, obesity, and osteoarthritis (Destatis, 2021). Chronic diseases are the main factors determining illness burden in Germany in terms of reductions of disability-adjusted life years (Plass et al., 2014).

Chronic diseases and the burden from chronic disease, however, are not equally distributed across the population. They are more prevalent in elderly and more disadvantaged individuals. In Germany, individuals with lower education are more than two times (men) or three times (women) more likely to suffer from Type 2 diabetes (Heidemann et al., 2009); similar odds occur for cardiovascular disease and arthritis (Lampert et al., 2017). In addition to differences in the prevalence of chronic disease, there are also differences in more distal health outcomes such as health-related quality of life. Individuals with chronic diseases are more likely to experience worse outcomes if they come from socioeconomically disadvantaged backgrounds (Mielck et al., 2014).

In addition to regular medical treatment, successful control of chronic diseases requires patient self-management, that is, adhering to medication, treatment regimes, and following recommendations regarding specific lifestyle behaviors. For example, for cardiovascular diseases—affecting 8.3% of the German adult population—increases in physical activity are recommended (Lanier et al., 2016). Similarly, for the effective control of Type 2 diabetes, regular self-management in the form of blood sugar monitoring, adherence to medication, dietary, and physical activity are paramount.

Challenges in self-managing health-related behaviors are a contributing factor to the global burden of illness (Bosworth, 2010). Fostering self-management competencies is a key element of the chronic care model (Bodenheimer et al., 2002; Coleman et al., 2009), the current best-

practice approach to treating chronic illness. This implies that developing effective, yet scalable, behavioral interventions that focus on improving patient self-management of chronic disease is a key public health task. Taxonomies such as the PRISMS taxonomy of self-management support (Pearce et al., 2015) identify components that increase the effectiveness of self-management interventions.

In this study, we examine the effects of a telephone self-management support intervention that included the following PRISMS components: (a) education about condition and management; (b) information about available resources, in order to facilitate crucial health literacy skills for chronic disease (Heijmans et al., 2015); (c) provision of specific and personalized action plans (e.g., Ring et al., 2011); (d) training/rehearsal for psychological strategies; (e) social support; and (f) lifestyle advice and support. These components have been shown to be effective in previous reviews of management of long-term chronic conditions (e.g., Fisher et al., 2005), and they can be feasibly delivered via telephone.

A recent umbrella review of previous systematic reviews and meta-analyses (Timpel et al., 2020) on the effects of remotely administered self-management support programs suggested that such programs can produce clinically relevant reductions in glycated hemoglobin levels and lipid levels in patients with diabetes, but not with patients with high blood pressure. Single trials have shown that telephone-based self-management support can reduce coronary events (Lisspers et al., 2005), symptoms and impairment (Böhme et al., 2012), as well as improve self-rated health (SRH) (Bambauer et al., 2005). Telephone-based self-management support has also been shown to lead to improvements in quality of life and reduced depressive symptoms in patients with chronic disease (Mons et al., 2013; Swoboda et al., 2017). Finally, telephone-delivered self-management support programs can be highly cost effective and are usually well tolerated by participants (Oksman et al., 2017).

However, patients with chronic diseases from disadvantaged backgrounds may be less likely to profit from both telephone-based and other self-management programs for a range of reasons (Lehne et al., 2019): lower enrollment and adherence rates (e.g., Kure-Biegel et al., 2016), barriers to participation in the health-care system (e.g., Goodridge et al., 2019), the relative absence of programs that are tailored toward patients from more disadvantaged backgrounds (Van Hecke et al., 2017), and a lack of programs targeted to appropriate levels of health literacy (Berkman et al., 2011). Thus, the extent to which short telephone-based counseling programs for the self-management of chronic diseases work equally well for participants from different socioeconomic backgrounds remains an open question. This is an important issue, as health promotion and prevention strategies, even if successful on average across a population, may unintentionally increase health disparities by benefiting educationally disadvantaged population groups less than more advantaged population groups (equity effects of interventions or intervention-generated inequalities) (Lorenc & Oliver, 2014).

This study therefore evaluated a telephone-based self-management support program for patients with serious chronic health problems (e.g., chronic heart failure [CHF], diabetes, chronic heart disease [CHD]) at high risk for future hospitalization. The aims of the present study were (a) to evaluate the overall effectiveness of a telephone-delivered self-management support program based on those components of the PRISMS taxonomy (Pearce et al., 2015) that can be delivered via the telephone and (b) to investigate the robustness of any intervention

effect across four key (demographic) indicators of social inequality in Germany (gender, age, educational level, and state [former GDR vs. FRG prior to German reunification]). These indicators are a subset of those recommended in the PROGRESS-Plus framework (O'Neill et al., 2014) for examining equity effects of interventions.

METHODS

Design and procedure

This study was implemented within a cohort of clients of one of Germany's largest statutory health insurance companies (Techniker Krankenkasse; tk.de/en). To accommodate standard operating procedures of the insurance company, we used a randomized invitation design (Holland, 1988; West et al., 2014; a.k.a. randomized consent design, Zelen, 1979). We identified the population of all individuals insured by the health insurance company who met the criteria of at least one severe chronic disease (e.g., CHF, diabetes, and CHD) and a high score on the insurance company's proprietary measure of the likelihood of hospitalization during the next year. Individuals diagnosed with dementia or severe mental disorders were excluded. From this population of eligible individuals, a random sample of individuals were contacted by telephone and invited to be in the intervention program (intervention group [IG]). The remaining individuals served as a reservoir of eligible non-invited individuals, which then served as the control group (CG). Participation in the intervention program was voluntary; 1927 of 2977 (64.73%) participants participated in at least two sessions. Participants who provided written informed consent at the beginning of the first session were enrolled into the study. IG participants were matched 1:1 with members of the CG reservoir using propensity scores (see below). The insurance company sent the questionnaires via mail to all eligible individuals in the IG and CG at the same three assessment points[1] (see below). De-identified data were transmitted to the authors and analyzed independently. Figure 1 presents a participant flow chart that illustrates the key features of the design.

The internal review board of the health insurance approved the study protocol, and the study was conducted in accordance with the Declaration of Helsinki. Authors obtained

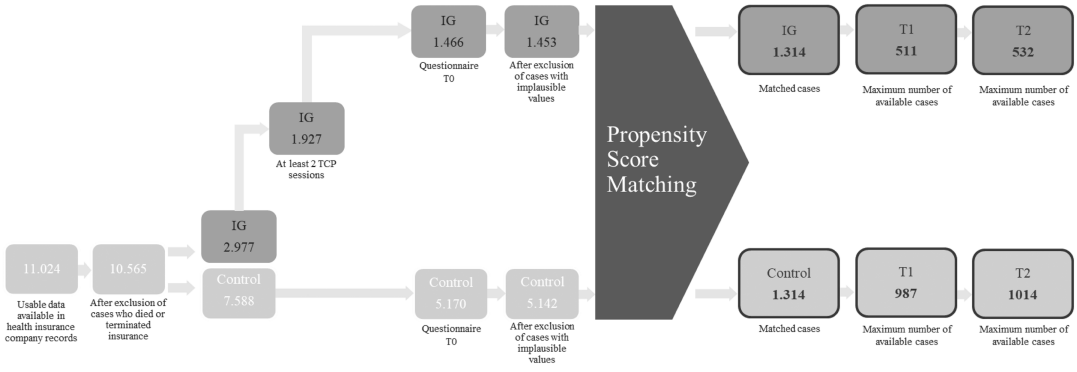

FIGURE 1 Participant flow chart. T1 = immediate posttest (4 months after T0), T2 = longer term posttest. (6 months after T0) IG = intervention group (Health Coach). Sample sizes represent the number of available cases

permission from the health insurance company to analyze the de-identified data. Further ethics approval was not required as per German ethical guidelines at the time. All participants in both the IG and CG received their regular medical care during the study. Thus, IG participants received the telephone-based program plus standard medical care, whereas CG participants received only standard medical care. The trial was registered at DRKS (German register of clinical trials) DRKS00023477.

Brief telephone counseling

Participants in the IG received a brief manualized telephone intervention (Techniker Krankenkasse, 2008). The manual consists of four main modules related to essential behaviors that influence the conditions of chronic diseases and health: (a) physical activity, (b) nutrition, (c) fluid intake, and (d) adherence to medication intake. Each module includes (a) a detailed analysis of the participant's current behavior, (b) education about the relevant disease and specific health behavior, (c) the identification and setting of goals, and (d) the planning of health-enhancing behavior in order to achieve these goals. IG participants were contacted every 2 weeks until both the participant and the coach decided active coaching on relevant modules was completed. IG participants were contacted for one or two booster sessions 4–6 weeks after program completion to facilitate maintenance of behavior change. Counselors were specially trained nurses, psychologists, or other health practitioners with a background in health counseling. They received continuous supervision. In addition to providing information about symptoms and treatment of the addressed diseases, the manual focused on basic self-management strategies in health care (Coulter & Ellins, 2007; Olsen & Nesbitt, 2010) and motivational interviewing strategies (Miller & Rollnick, 2003). The counseling process consisted of three phases: introduction, behavior change, and maintenance. In the initial contact session (introduction), the intervention was outlined, the ability to participate was assured (e.g., no hearing impairment), and the participant's consent to participate in the intervention was obtained. Next, participant's symptoms and medical condition were assessed, and initial planning was undertaken. In subsequent behavior change sessions, each module was approached in the same structured way: First, participants' current knowledge was assessed, current health behavior was analyzed, individual goals were set for this specific behavior, and health behavior barriers were identified. Next, behavior change was planned with respect to the participant's motivational and behavioral barriers. For each module, up to four contacts were scheduled every seven to 14 days.

After 4–6 weeks, participants were again contacted, and their health status and health behavior were assessed. Counselors provided encouragement to continue and offered support regarding potential problems in maintenance. Three to four months after finishing the last module, participants were contacted one last time to consolidate behavior change and assess health behavior status. Follow-up assessments for all participants (IG and CG) were scheduled approximately 4 months after the beginning of the intervention in the IG (Time 1) and approximately 6 months after the beginning of the intervention in the IG (Time 2).

Measures

Diagnosis/comorbidity

The index diagnosis and the number of comorbidities were derived from de-identified insurance data from a list of 11 chronic diseases that were diagnosed by a physician in the previous

24 months before participation (arteriosclerosis, arthrosis, asthma, cancer, cardiac infarction, hypertension, ischemic heart disease, stroke, CHF, COPD, diabetes).

Primary outcomes

Self-rated health

SRH was assessed using a well-established and validated single item measure (DeSalvo et al., 2006). Participants were asked to estimate their SRH on a scale ranging from 0 (*very poor*) to 10 (*very good*). The exact wording was: "If you were to rate your general state of health on a scale from 0 to 10, ('0,' meaning *couldn't be worse* and '10,' meaning *couldn't be better*), how would you rate your current state of health?" Higher scores indicate better self-rated health.

Depressive symptoms

We used the World-Health-Organization-Five Scale (WHO-5; WHO Regional Office for Europe, 1998), a brief, widely used screening instrument for depressive symptoms (Löwe et al., 2004). The WHO-5 has shown excellent internal consistency (Cronbach's $\alpha = .91$) and excellent criterion validity relative to physicians' diagnoses (Löwe et al., 2004). Lower scores indicate more depressive symptoms.

Measures used in propensity score matching

Propensity scores served as a vehicle to equate the IG and CG participants on key background variables that might confound the results of the study. Only participants in the IG had to actively agree to participate in the program, giving rise to potential differential selection into the IG and CG. Overall, 89 variables from the health insurance company records and the baseline questionnaire were used to construct propensity scores, the predicted probability of agreeing to participate in the intervention if invited. Areas assessed included *motivation* for improving one's health, behavioral risk factors for health (e.g., cigarette smoking, lack of exercise, adherence to medication, and diet), medical variables (e.g., blood pressure, number of physician contacts, and proprietary likelihood of hospitalization score), medical diagnoses (e.g., COPD, asthma, and diabetes), psychological variables (e.g., self-efficacy, quality of life, and well-being), and limited sociodemographic data (e.g., gender, age, and education).

ANALYSES

Data integration and propensity score matching

The logic of the current study involves the comparison of those participants whose propensity to receive treatment could be equated in the IG and CG. A total population of 11,024 eligible individuals were available for whom usable data were available in health insurance company records. From this number, 459 cases were excluded because the individual died or terminated the insurance. The resulting initial *potential* sample was $n_{IG} = 2977$ in the IG and $n_{CG} = 7588$ in the CG reservoir. All individuals in the initial potential sample were sent a baseline (Time 0) questionnaire. Individuals who did not return the baseline questionnaire and IG individuals with only one contact (the initial enrolment session) were excluded. These exclusion criteria resulted in $n = 1466$ in IG and $n = 5170$ in the CG reservoir. Finally, we excluded participants

with implausible values on self-reports of medical variables (systolic and diastolic blood pressure, resting heart rate, height)[2] (e.g., resting heart rate below 30 or above 180). This screening led to a total of $n = 41$ individuals being excluded from the outcome analysis. Thus, the final potential sample sizes were $n = 5142$ in the CG reservoir and $n = 1453$ in the IG[3] (see Figure 1). Using logistic regression, we then estimated each participant's propensity score (probability of agreeing to participate in the IG if invited) based on 89 covariates measured before the intervention, 62 binary indicators of missingness, and interaction terms between key variables (see Cham & West, 2016). Participants in the IG and CG reservoir were then matched 1:1 on the basis of their propensity scores (see Imbens & Rubin, 2015; Rosenbaum & Rubin, 1983). This procedure yielded identical sample sizes for the matched groups of IG and CG participants ($n = 1314$ each) and resulted in well-equated propensity scores and covariates (see Figure 2). Technical details of the propensity score estimation and matching, and evaluation of the balance achieved by propensity score matching in the IG and CG, are presented in the online supplement. Given this sample size, following Cohen (1988), a power analysis showed that an intervention effect in the population of $\delta = \frac{\mu_{IG} - \mu_{CG}}{\sigma} = 0.11$ can be detected with desired statistical power = .80 and $\alpha = .05$, two tailed, where μ_{IG} is the mean of the IG, μ_{CG} is the mean of the CG, and σ is the standard deviation in the population. The interaction between the intervention and the four educational or age subgroups given desired power = .80 and $\alpha = .05$ is $f = \sqrt{\frac{\eta^2}{1-\eta^2}} = .065$, where η^2 is the proportion of variance accounted for by the interaction term. Cohen defined population standardized effect sizes of $\delta = .2$ and $f = .1$ as small effects, indicating that the design was adequately powered to detect very small effects.

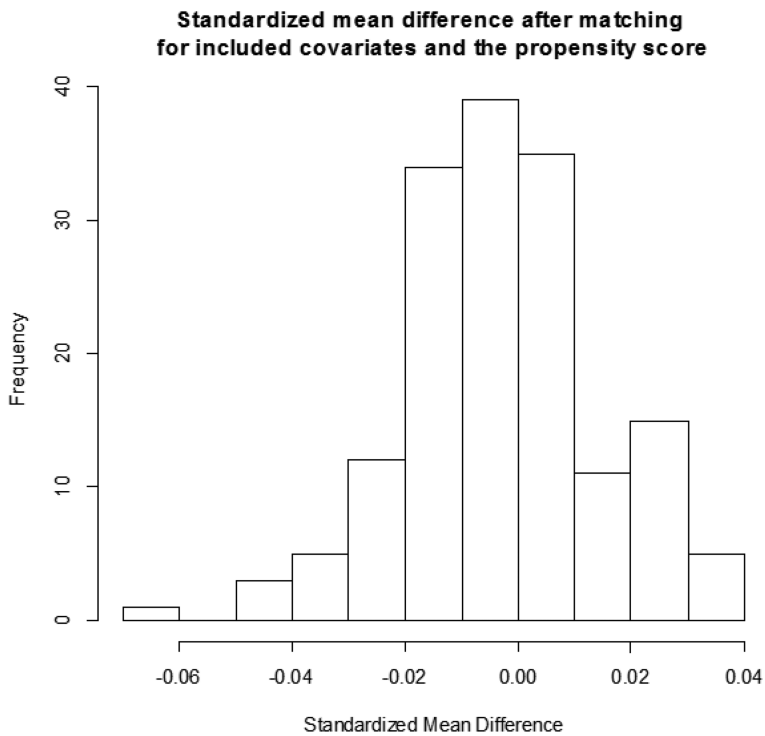

FIGURE 2 Histogram of standardized mean differences for all included covariates after matching

Measurement and scoring of socioeconomic variables

The key outcomes (self-rated health and WHO-5) were measured in both IG and CG participants at baseline (Time 0), shortly after the completion of the initial intervention (Time 1, baseline + approximately 4 months) to assess immediate effects and shortly after the

TABLE 1 Baseline demographic and medical characteristics: Propensity score matched analysis sample

(a) % of participants in demographic and medical diagnosis categories			
Demographic category	**% overall**	**% CG**	**% IG**
Gender (male)	81.6%	81.7%	81.6%
Former East Germany	14.3%	14.3%	14.2%
Medical diagnosis			
Malignant neoplasms	20.1%	20.5%	19.7%
Hypertension	93.3%	93.5%	93.0%
Ischemic heart disease	89.0%	89.5%	88.4%
Atherosclerosis	24.8%	25.0%	24.7%
Myocardial infarction	33.9%	34.3%	33.5%
Heart failure	45.0%	45.2%	44.7%
Stroke	10.1%	10.0%	10.1%
COPD	15.3%	15.7%	14.9%
Asthma	8.8%	8.3%	9.2%
Diabetes	49.0%	50.2%	47.8%
Arthrosis	36.9%	37.1%	36.8%
(b) % of participants in categories of educational level (ISCED) and age			
Educational level	**% in category overall**	**% in category CG**	**% in category IG**
ISCED 2	55.7%	55.9%	55.6%
ISCED 3	8.9%	9.0%	8.9%
ISCED 5	22.3%	22.1%	22.6%
ISCED 7	13.0%	13.1%	12.9%
Age category			
Younger than 55 years	10.2%	9.8%	10.6%
55–64	23.5%	23.9%	23.1%
65–74	50.7%	50.9%	50.5%
75 and older	15.6%	15.4%	15.9%

Note: Total sample size for age = 2628; $n = 91$ individuals have missing data on the education variable (not considered in percentages). Medical diagnosis includes primary diagnosis and all comorbid diagnoses listed in patient's record at baseline and 24 months prior to baseline. ISCED Categories 1 and 2 for educational level were combined after matching because of sparse data in category 1.

Abbreviations: CG, control group; IG, intervention group.

completion of the sustainability sessions (Time 2, baseline + approximately[4] 6 months) to assess longer term effects of the intervention.

Four different indicators of socioeconomic status were available: gender (male vs. female), region of residence (former East Germany, GDR, vs. former West Germany, FRG), education, and age in years. Education was categorized according to the International Standard Classification of Education (ISCED 2011; UNESCO, 2012). The following levels were present in the data: ISCED 2 (some secondary education), ISCED 3 (secondary education degree providing qualification for university entrance), ISCED 5 (college or technical training of at least 2 years), and ISCED 7 (completed university degree). We divided the sample into four age categories (participants aged below 55 years, 55–64 years, 65–74 years, 75 plus years) to represent potential nonlinear relationships with age. Table 1 gives an overview of the indicators of socioeconomic status and medical diagnoses.

Statistical analyses

Three sets of analyses addressed the questions of interest. First, we examined potential baseline socioeconomic differences between the participants *prior* to any intervention. Second, using change score models (Kievit et al., 2018), we compared the mean differences in the IG and CG on (a) change score 1 representing the shorter term (Time 1 − Time 0) and (b) change score 2 representing the longer term (Time 2 − Time 0) effects of intervention. To test the effect of the intervention on the change scores (change score 1: Time 1 vs. Time 0; change score 2: Time 2 vs. Time 0), a saturated model was first fitted to the data, followed by models that separately restricted, each change score to be equal in the IG and CG. Following standard model comparison procedures in the general linear model (Cohen et al., 2003) and structural equation modeling, we compared the increase in prediction from each restricted to the saturated model (hierarchical model tests) to test the intervention effects. The intervention effects were tested separately for change score 1 and change score 2. The R package lavaan (Rosseel, 2012) was used to fit the change score models (see Supporting Information for lavaan code; Figure S2 shows the basic change score model). Third, we examined whether socioeconomic indicators moderated the effects of intervention (e.g., whether the intervention effects differed between males and females). We addressed missing data using full-information maximum likelihood estimation (Graham, 2009; Little & Rubin, 2020). Between-group effect sizes were computed to provide an estimate of the magnitude of the overall intervention effects. Following Cohen (1988), standardized between-group effect sizes (Cohen's *d*) were estimated based on the model based pooled standard deviation (i.e., we divided the difference in change scores between IG and CG by the pooled standard deviation of the change scores of each group).

RESULTS

Baseline differences as a function of socioeconomic indicators

Table 2 reports means and statistical tests of baseline differences in mean SRH and WHO-5 depressive symptoms as a function of the socioeconomic indicators. Males had higher means on the SRH and WHO-5 depressive symptoms compared with female participants indicating better SRH and fewer depressive symptoms. Compared with residents of former West German (BRD)

TABLE 2 Sociodemographic and socioeconomic differences at baseline (model-based estimates)

Outcome	Variable and χ^2-test	Category	Mean T0	Variance T0
SRH	Gender* ($\chi^2(1) = 6.23$, $p = .013$)	F	5.42	3.146
		M	5.65	3.160
	Region of residence ($\chi^2(1) = 0.18$, $p = .669$)	E	5.57	2.503
		W	5.61	3.274
	Education ($\chi^2(3) = 4.02$, $p = .259$)	ISCED 2	5.57	3.217
		ISCED 3	5.52	3.588
		ISCED 5	5.70	2.942
		ISCED 7	5.73	3.038
	Age* ($\chi^2(3) = 19.59$, $p < .001$)	< 55	5.45	3.801
		55–64	5.39	3.628
		65–74	5.76	2.869
		>74	5.57	2.815
WHO-5	Gender* ($\chi^2(1) = 38.70$, $p < .001$)	F	52.83	563.438
		M	60.58	509.353
	Region of residence* ($\chi^2(1) = 9.92$, $p = 0.002$)	E	62.50	432.255
		W	58.66	540.887
	Education* ($\chi^2(3) = 23.02$, $p < .001$)	ISCED 2	57.77	557.436
		ISCED 3	56.33	593.431
		ISCED 5	61.61	460.148
		ISCED 7	62.73	463.123
	Age* ($\chi^2(3) = 118.80$, $p < .001$)	< 55	47.50	542.498
		55–64	54.73	627.243
		65–74	62.89	440.001
		>74	62.06	469.023

Abbreviations: F, female; M, male; SRH, Self-Rated Health; T0, baseline; WHO-5, World-Health-Organization-Five Scale.
*Significant χ^2-test ($p < .05$) for baseline mean differences.

states, residents of the former DDR states had higher means on WHO-5 depressive symptoms, but there was no difference on SRH. On average, more highly educated participants reported fewer depressive symptoms than less educated participants, but the two groups did not differ in SRH. Finally, there were differences between age groups, with the 67–74 age group reporting the highest levels of SRH and the lowest levels of depressive symptoms.

Main effects of the Health Coach intervention at Waves 1 and 2

Self-rated health

We first performed a check on the extent to which propensity score matching led to balance on baseline self-rated health prior to treatment. There were no significant differences between the

IG and the CG at baseline in self-rated health ($\chi^2(1) = 0.42$, $p = .518$, $d = -.03$). The results from the change score models showed that the mean gain for the IG was larger than that for the CG for both the Time 1 vs. Time 0 (baseline) change score, $\chi^2(1) = 50.29$, $p < .001$, $M_{gain} = 0.59$, $d = .37$, and the Time 2 vs. Time 0 change score, $\chi^2(1) = 24.06$, $p < .001$, $M_{\text{gain}} = 0.40$, $d = .25$. The model-estimated means for each group over time are depicted in Figure 3 (top panel). As can be seen, the gains from the initial level were larger for the IG than for the CG for both shorter and longer term change.

WHO-5

The check on the success of the propensity score matching showed no significant differences between the IG and the CG at baseline in the WHO-5 score ($\chi^2(1) = 0.001$, $p = .974$, $d = .00$).

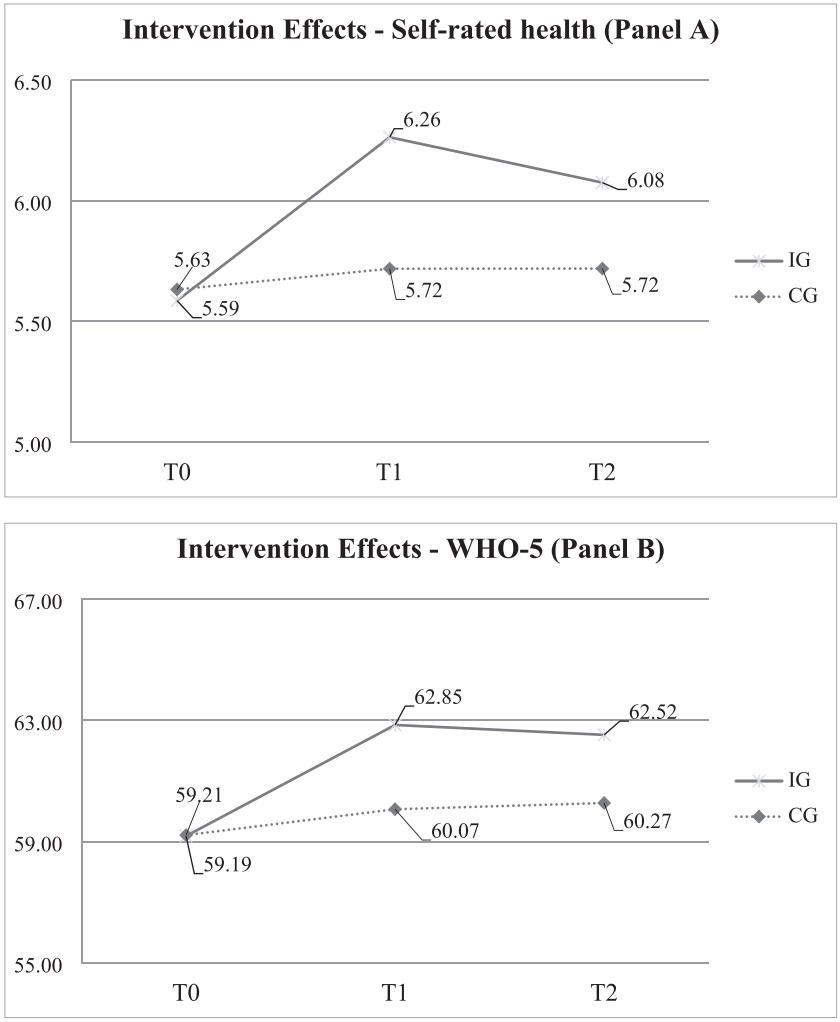

FIGURE 3 Model-implied means for the control and intervention groups at each measurement wave on self-rated health (Panel A; original scale: 0–10) and WHO-5 score (Panel B; original scale: 0–100). T0 = baseline, T1 = immediate posttest, T2 = longer-term posttest. CG = control group, IG = intervention group

The results from the change score models showed that the gain from the intervention was larger than that for the control group for both the Time 1 vs. Time 0 change score, $\chi^2(1) = 10.21$, $p = .001$, $M_{gain} = 2.81$, $d = .17$, and the Time 2 vs. Time 0 change score, $\chi^2(1) = 5.97$, $p = .015$ $M_{gain} = 2.28$, $d = .13$. The model-estimated means for each group across the three waves are depicted in Figure 3 (bottom panel). The results showed larger gains from the intervention than the control relative to the baseline levels.

Equity effects

Self-rated health

There was no significant gender x intervention interaction for change score 1, $\chi^2(2) = 1.16$, $p = .559$, nor for change score 2, $\chi^2(2) = 0.79$, $p = .675$. The test of the moderator effect for region of residence (former East vs. West Germany states) also did not show a region × intervention interaction for change score 1, $\chi^2(2) = 1.37$, $p = 0.503$ or for change score 2, $\chi^2(2) = 1.29$, $p = .525$. The test of the moderator effect for education group was also not significant for change score 1, $\chi^2(6) = 2.67$, $p = .848$, nor for change score 2, $\chi^2(6) = 7.12$, $p = .310$. However, age group moderated the change for both change score 1, $\chi^2(6) = 21.74$, $p = .001$ and change score 2, $\chi^2(6) = 17.13$, $p = .009$. From baseline (Time 0) to Time 1, the change in the IG exceeded the change in the CG for each age category: $M_{\Delta < 55} = +0.43$, $M_{\Delta\ 55–64} = +0.85$, $M_{\Delta\ 65–74} = +0.50$, and $M_{\Delta > 74} = +0.53$. From baseline to Time 2, the longer term change in the IG again exceeded the change from the CG for each age category: Mean difference in change between IG and CG (M_Δ) for each age category was $M_{\Delta < 55} = +0.21$; $M_{\Delta\ 55–64} = +0.59$; $M_{\Delta\ 65–74} = +0.30$; $M_{\Delta > 74} = +0.55$, with the 55–64 age category showing the largest positive change.

WHO-5

There were no significant moderator effects for region of residence or education. Specifically, the test of the interaction effect with intervention for region of residence (former East vs. West Germany states) again did not show a region x intervention interaction for change score 1, $\chi^2(2) = 0.33$, $p = .847$, nor for change score 2, $\chi^2(2) = 0.57$, $p = .750$. The test of the moderator effect for education group was again not significant for change score 1, $\chi^2(6) = 11.62$, $p = .071$, nor for change score 2, $\chi^2(6) = 7.73$, $p = .258$. However, there were significant moderator effects for gender and age. The test of the moderator effect for gender did show a significant gender × intervention interaction for change score 1, $\chi^2(2) = 7.05$, $p = .030$ and for change score 2, $\chi^2(2) = 6.06$, $p = .048$. Relative to the CG, females gained more from the intervention from both Time 0 to Time 1, $M_{\Delta\ female} = +5.39$ than males, $M_{\Delta\ male} = +2.37$ and from Time 0 to Time 2, $M_{\Delta\ female} = +5.62$, $M_{\Delta\ male} = +1.64$. Age group also moderated the intervention effect for both change score 1, $\chi^2(6) = 17.55$, $p = .007$ and change score 2, $\chi^2(6) = 21.85$, $p = .001$. Mean differences in change between the IG and CG (M_Δ) for each age category were $M_{\Delta < 55} = -0.83$, $M_{\Delta\ 55–64} = +5.15$, $M_{\Delta\ 65–74} = +1.94$, and $M_{\Delta > 74} = +4.26$. Change in the WHO-5 score in the IG relative to the CG from baseline to Time 1 varied with age category. Mean difference in change between the IG and CG (M_Δ) for each age category from baseline to Time 2 was $M_{\Delta < 55} = +1.58$; $M_{\Delta\ 55–64} = +4.06$; $M_{\Delta\ 65–74} = +1.43$; $M_{\Delta > 74} = +3.70$. These

results suggested that the longer term gains for the intervention relative to the control group were mostly positive, but variable across the four age categories. The 55–64 age category showed the largest positive intervention effect.

DISCUSSION

This study evaluated the effectiveness of a brief telephone-based counseling program ("Health Coach") for patients with chronic diseases after 4 and 6 months. It also evaluated whether intervention effects differed according to participant-level indicators of social inequality (PROGRESS-Plus [O'Neill et al., 2014] dimensions of place of residence, age, gender, and education). To the extent such moderator effects exist, they may represent potential equity effects, that is, differential effects of the intervention according to indicators of participant socioeconomic status and gender (Lehne et al., 2019; Lorenc & Oliver, 2014). We used a randomized invitation (sometimes also "randomized consent"; Zelen, 1979) design to balance the IG and CG on baseline covariates. Given a possible selection bias into the IG because of the need for consent, we augmented the basic randomized invitation design by matching the IG and CG based on their propensity scores. This additional design feature allowed us to contrast changes in individuals who agreed to participate in the Health Coach program against changes in the initially similar group of nonparticipants who would be expected to accept the intervention if offered. We found that participants in the Health Coach program reported significantly increased levels of SRH and emotional well-being over the course of 4 months and maintained these gains for another 2 months. The standardized effect sizes were between small and moderate in magnitude (Cohen, 1988). Recall that the population studied was selected to be at high risk of hospitalization and that both groups received regular medical care as prescribed by the German national health-care system. Consequently, we regard the additional effect of the Health Coach intervention over and above standard German medical care to be sufficiently large to be of clinical significance.

Our results show that a brief telephone-based counseling program that included components on education about the chronic disease, information on available resources, provision of action plans, rehearsal of self-help strategies, social support, and general lifestyle advice can be effective in improving subjective ratings of health and emotional well-being. Older adults with chronic illnesses are at high risk for deterioration in well-being and self-rated health (Sheridan et al., 2019); such a program, if rolled out on a broader basis, has the potential to increase well-being in this chronic disease population.

Equity effects

These small to medium-sized effects are further supported by an absence of more than small equity effects of the intervention with regard to the four PROGRESS-Plus dimensions that could be examined. Consistent small differences in the effects of the intervention on SRH and depressive symptoms were found with regard to age. Descriptively, the IG had more positive outcomes than the CG in all but one of the four age categories, and females relative to males showed a somewhat greater benefit from the intervention on the WHO-5. This latter effect could be either due to lower WHO-5 scores at baseline—that is, more room for mitigation of depressive symptoms—or due to women profiting more from the attention delivered by the intervention. Note that this effect was only found on depressive symptoms, but not on self-rated health.

The general absence of equity effects for education and place of residence are promising. As described in Section 1, educational attainment is a dimension of socioeconomic status that has tended to be particularly relevant for health outcomes in patients with chronic disease in Germany. Individuals with lower educational attainment are have both a higher prevalence of chronic diseases and suffer more consequences from chronic disease in terms of health and depressive symptoms. Previous studies have suggested that participation and retention in self-management trials is stratified by socioeconomic factors, in particular age and education (e.g., Kure-Biegel et al., 2016), putting participants who are younger and who have lower educational attainment at higher risk. In addition, individuals living in the federal states that constituted the former East Germany compared with those that constituted the former West Germany were also more likely to be affected by chronic disease, possibly due to higher poverty and lower employment rates. As such, our finding that intervention effects were of comparable effect size across education and place of residence is promising, because it suggests that a brief telephone counseling program for self-management in chronic disease could contribute to an overall improvement in SRH and emotional well-being while being unlikely to further exacerbate existing health inequalities.

Of importance, this study is one of the few intervention studies in chronic disease self-management that specifically test differential effects of an intervention in individuals from different socioeconomic backgrounds. Previous studies have mainly *controlled* for the effects of indicators of socioeconomic inequality (e.g., Contant et al., 2019). These studies reported that the effects of interventions change in magnitude when such factors are controlled for, but such tests remain silent about the specific differential effects of interventions according to socioeconomic background. This absence of equity-specific analyses has been highlighted in some previous reviews of self-management interventions (e.g., for chronic osteoarthritis, Kroon et al., 2014). To our knowledge, no previous published study has tested for potential equity effects in self-management for chronic disease. This lack of tests of equity effects is potentially problematic: If self-management interventions were to have differential effects according to participants' socioeconomic background, existing inequalities in chronic disease health outcomes (Mielck et al., 2014) could be further exacerbated (Lorenc & Oliver, 2014).

In this study, we based the equity dimensions on the PROGRESS-Plus framework (O'Neill et al., 2014), which has been specifically formulated for the examination of equity effects of interventions. Although the set of indicators suggested in the PROGRESS-Plus framework is by no means comprehensive, it provides a viable framework for examining equity effects that could be used in future studies. If more studies reported such information, which is often possible using already collected indicators of socioeconomic position, an evidence base for interventions that are likely or unlikely to work in different socioeconomic groups could be built and considered in health services decision-making.

This study did not use the full set of equity dimensions suggested in the PROGRESS-Plus framework. We only considered information on place of residence (P), gender (G), age (Plus), and education (E), as more information was not available in the data set. This implies that other equity effects of the intervention are possible, but could not be explored here.

Design considerations: Strengths and limitations

There are particular strengths and limitations to applying a randomized invitation design. In a conventional randomized clinical trial, participants initially agree to participate in either the IG

or CG with IG participants typically being given incentives for full participation in the intervention to maximize treatment adherence. The effect of the treatment would be evaluated using an intention-to-treat (ITT) analysis. In contrast, in our randomized invitation design, eligible participants in the population were randomly assigned to be *invited* to receive the intervention, whereas the CG reservoir was randomly assigned to be not invited to participate and was given treatment as usual. Thus, our starting point was a randomized invitation design in which participants in the IG and CG were expected to be balanced at baseline on all measured and unmeasured covariates. However, as is common in a randomized invitation design, not all participants accepted the invitation, leading to the possibility of imbalance between IG participants and participants in the CG reservoir. In addition, the insurance company decided to not collect data on participants assigned to the IG who did not agree to receive the intervention program. A number of treatment noncompliance models have been proposed (Sagarin et al., 2014) that permit strong causal inferences in the absence of treatment compliance. These models compare the mean outcome of participants in the treatment group who comply with the intervention with the mean outcome of matched participants in the control group who would be expected to comply if they had been assigned to the treatment group. The effect estimated is the average treatment effect on the treated.

The success of this matching procedure depends on having available a rich set of covariates that are potentially related to both the treatment received and the outcome. Here, we included 89 covariates in the propensity score model: On these measured covariates, we were able to achieve balance similar in degree to the balance that would be achieved in a randomized clinical trial with full treatment compliance. The major limitation of the approach is that there might be other unmeasured covariates on which the two groups were not balanced, potentially confounding the results of the study. However, given the rich set of 89 covariates, any uncontrolled covariate would need to have a unique relationship with both the intervention condition and outcome over and above that of the 89 measured covariates. Previous studies (e.g., Shadish et al., 2008; Steiner et al., 2010) that have compared the effect sizes obtained in non-randomized and randomized studies have pointed to measures of motivation, self-efficacy, and treatment preference as particularly important covariates that should ideally be measured. Both motivation and self-efficacy were included among our covariates,[5] substantially reducing the likelihood that this is a serious limitation in the present study. When a rich set of covariates are available, comparisons of propensity score and other similar procedures for equating treatment and control groups have shown their success in yielding an unbiased estimate of the magnitude of the causal effect (Cook et al., 2008; Shadish et al., 2008; Steiner et al., 2010)., Indeed, it has been argued (Imbens & Rubin, 2015) that, in contrast to the ITT effect that provides an estimate of the average effect of treatment *assignment* on the outcome, treatment noncompliance models provide unbiased estimates of the average treatment on the treated, a more informative measure of treatment effectiveness.[6] In addition, the need to attain participant's permission to receive either the IG or the CG and the use of incentives to maximize compliance in a randomized clinical trial can lead to limits on the generalization of the results of the trial (Diener et al., in press; see Relton et al., 2010 for a comparison of designs). Cartwright and Hardie (2012) have documented the many failed educational and public-health interventions based on prior randomized clinical trials that used participants and enhanced interventions that were not representative of the target populations and treatments of interest.

Another limitation is that there were missing data (approximately 57% complete cases) as is common in designs in which the data are collected by government-related companies rather than researchers. Full information maximum likelihood is expected to produce unbiased

estimates of the intervention effect assuming that data are missing at random (Little & Rubin, 2020). Missing at random implies that missingness can be adequately predicted by other measured variables in the model and that other remaining sources of missingness are random. In contrast, if missingness is predicted by would be values of the outcome variables themselves, then estimates of intervention effects may be biased.

A final limitation lies in the broad study end points of depressive symptoms and SRH. At the same time, measures of SRH have been shown to be valid indicators and predictors of both morbidity and mortality (DeSalvo et al., 2006).

CONCLUSION

Despite these limitations, there are important implications that can be drawn from the current study. A relatively brief telephone counseling program for the self-management of chronic disease resulted in sustained improvements in SRH and depressive symptoms. These effects were of comparable size for groups of participants differing by education and place of residence, with some tentative evidence of differences for gender and age groups. These findings point to the potential of such interventions to be scaled up and to markedly improve broad health outcomes in adults with chronic disease. More research is needed to further explore equity effects of self-management interventions in chronic disease.

ACKNOWLEDGEMENT

The study was supported by the Techniker Krankenkasse (https://www.tk.de/en). Content is the sole responsibility of the authors. Stephen G. West was supported by a research stay supplement to his Forschungspreis from the Alexander von Humboldt Stiftung.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

ETHICS STATEMENT

The protocol for this study was approved by the internal review board of the study funder (Techniker Krankenkasse). The study was conducted in full accordance with the declaration of Helsinki. Authors obtained permission from the health insurance company to analyze the de-identified data. Further ethics approval was not required as per German ethical guidelines at the time.

DATA AVAILABILITY STATEMENT

Truncated R code from the analyses is available in the Supporting Information. Anonymized data are available from the authors on reasonable request.

ORCID

Babette Renneberg https://orcid.org/0000-0003-2177-3566
Benjamin Schüz https://orcid.org/0000-0002-0801-498X

ENDNOTES

[1] The insurance company did *not* send questionnaires at the two follow-up assessments to insured individuals who did not agree to participate in the intervention.

[2] We thank Professor Andreas Schaper, M.D., University of Göttingen Medical School for guidance in establishing plausible values for the self-reports of medical variables.

[3] The implausible self-report values on the self-reports of medical variables only affect the calculation of the propensity scores. If we correlate propensity scores based on the 595 participants available for analysis following screening of participants for implausible self-report values and the propensity scores based on the full sample of 6636 unscreened cases, the correlation of the propensity score is extremely high $r = .976$. Consequently, there are only minor differences between the screened and unscreened participants. We report the results of the screened participants.

[4] Interviews took place in an approximate 1-month window.

[5] Given that IG and CG participants were unaware of the existence of the other treatment condition, no measure of treatment preference could be feasibly included.

[6] According to Sheiner and Rubin (1995), the ITT effect answers the following question: "What are the expected outcomes for the typical patient instructed, in the context of the trial, to take the treatment to which he was assigned?" (p. 9). Given that outcome data were not collected on noncompliers, the ITT effect could not be estimated in the present study.

REFERENCES

Bambauer, K. Z., Aupont, O., Stone, P. H., Locke, S. E., Mullan, M. G., Colagiovanni, J., & McLaughlin, T. J. (2005). The effect of a telephone counseling intervention on self-rated health of cardiac patients. *Psychosomatic Medicine*, *67*(4), 539–545. https://doi.org/10.1097/01.psy.0000171810.37958.61

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, *155*(2), 97–107. https://doi.org/10.7326/0003-4819-155-2-201107190-00005

Bodenheimer, T., Wagner, E. H., & Grumbach, K. (2002). Improving primary care for patients with chronic illness. *Journal of the American Medical Association*, *288*(14), 1775–1779. https://doi.org/10.1001/jama.288.14.1775

Böhme, S., Geiser, C., Mühlenhoff, T., Holtmann, J., & Renneberg, B. (2012). Telephone counseling for patients with chronic heart failure: Results of an evaluation study. *International Journal of Behavioral Medicine*, *19*(3), 288–297. https://doi.org/10.1007/s12529-011-9179-0

Bosworth, H. B. (2010). *Improving patient treatment adherence: A clinician's guide*. Springer.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.

Centers for Disease Control and Prevention. (2020). Mortality in the United States, 2019. NCHS Data Brief No. 395. Accessed July 23, 2021. https://www.cdc.gov/nchs/data/databriefs/db395-H.pdf

Cham, H., & West, S. G. (2016). Propensity score analysis with missing data. *Psychological Methods*, *21*(3), 427–445. https://doi.org/10.1037/met0000076

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple correlation/regression analysis for the social sciences*. Lawrence Erlbaum Associates.

Coleman, K., Austin, B. T., Brach, C., & Wagner, E. H. (2009). Evidence on the chronic care model in the new millennium. *Health Affairs*, *28*(1), 75–85. https://doi.org/10.1377/hlthaff.28.1.75

Contant, É., Loignon, C., Bouhali, T., Almirall, J., & Fortin, M. (2019). A multidisciplinary self-management intervention among patients with multimorbidity and the impact of socioeconomic factors on results. *BMC Family Practice*, *20*(1), 53. https://doi.org/10.1186/s12875-019-0943-6

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750. https://doi.org/10.1002/pam.20375

Coulter, A., & Ellins, J. (2007). Effectiveness of strategies for informing, educating, and involving patients. *British Medical Journal*, *335*(7609), 24–27. https://doi.org/10.1136/bmj.39246.581169.80

DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality prediction with a single general self-rated health question. *Journal of General Internal Medicine*, *21*(3), 267–275. https://doi.org/10.1111/j.1525-1497.2005.00291.x

Destatis. (2021). Health: Causes of death. Accessed July 23, 2021. https://www.destatis.de/EN/Themes/Society-Environment/Health/Causes-Death/_node.html

Diener, E., Northcutt, R., Zyphur, M., & West, S. G. (in press). Beyond experiments. *Perspectives on Psychological Science*.

Fisher, E. B., Brownson, C. A., O'Toole, M. L., Shetty, G., Anwuri, V. V., & Glasgow, R. E. (2005). Ecological approaches to self-management: The case of diabetes. *American Journal of Public Health*, *95*(9), 1523–1535. https://doi.org/10.2105/AJPH.2005.066084

GBD 2019 Diseases and Injuries Collaborators. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *The Lancet*, *396*, 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Goodridge, D., Bandara, T., Marciniuk, D., Hutchinson, S., Crossman, L., Kachur, B., Higgins, D., & Bennett, A. (2019). Promoting chronic disease management in persons with complex social needs: A qualitative descriptive study. *Chronic Respiratory Disease*, *16*, 1–9. https://doi.org/10.1177/1479973119832025

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Heidemann, C., Kroll, L., Icks, A., Lampert, T., & Scheidt-Nave, C. (2009). Prevalence of known diabetes in German adults aged 25–69 years: Results from national health surveys over 15years. *Diabetic Medicine*, *26*(6), 655–658. https://doi.org/10.1111/j.1464-5491.2009.02739.x

Heijmans, M., Waverijn, G., Rademakers, J., van der Vaart, R., & Rijken, M. (2015). Functional, communicative and critical health literacy of chronic disease patients and their importance for self-management. *Patient Education and Counseling*, *98*(1), 41–48. https://doi.org/10.1016/j.pec.2014.10.006

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology*, *18*, 449–484. https://doi.org/10.2307/271055

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A. L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., NSPN Consortiumm, Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, *33*, 99–117. https://doi.org/10.1016/j.dcn.2017.11.007

Kroon, F. P. B., van der Burg, L. R. A., Buchbinder, R., Osborne, R. H., Johnston, R. V., & Pitt, V. (2014). Self-management education programmes for osteoarthritis. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD008963.pub2

Kure-Biegel, N., Schnohr, C. W., Hindhede, A. L., & Diderichsen, F. (2016). Risk factors for not completing health interventions and the potential impact on health inequalities between educational groups – A mixed method study from Denmark. *International Journal for Equity in Health*, *15*(1), 54. https://doi.org/10.1186/s12939-016-0344-6

Lampert, T., Hoebel, J., Kuntz, B., Müters, S., & Kroll, L. E. (2017). Gesundheitliche Ungleichheit in verschiedenen Lebensphasen. https://doi.org/10.17886/RKI-GBE-2017-003

Lanier, J. B., Bury, D. C., & Richardson, S. W. (2016). Diet and physical activity for cardiovascular disease prevention. *American Family Physician*, *93*(11), 919–924.

Lehne, G., Voelcker-Rehage, C., Meyer, J., Bammann, K., Gansefort, D., Brüchert, T., & Bolte, G. (2019). Equity impact assessment of interventions to promote physical activity among older adults: A logic model framework. *International Journal of Environmental Research and Public Health*, *16*(3), 420. https://doi.org/10.3390/ijerph16030420

Lisspers, J., Sundin, O., Ohman, A., Hofman-Bang, C., Ryden, L., & Nygren, A. (2005). Long-term effects of lifestyle behavior change in coronary artery disease: Effects on recurrent coronary events after percutaneous coronary intervention. *Health Psychology*, *24*(1), 41–48. https://doi.org/10.1037/0278-6133.24.1.41

Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd ed.). Wiley.

Lorenc, T., & Oliver, K. (2014). Adverse effects of public health interventions: A conceptual framework. *Journal of Epidemiology and Community Health*, *68*(3), 288–290. https://doi.org/10.1136/jech-2013-203118

Löwe, B., Spitzer, R. L., Gräfe, K., Kroenke, K., Quenter, A., Zipfel, S., Buchholz, C., Witte, S., & Herzog, W. (2004). Comparative validity of three screening questionnaires for DSM-IV depressive disorders and

physicians' diagnoses. *Journal of Affective Disorders*, *78*(2), 131–140. https://doi.org/10.1016/s0165-0327(02)00237-9

Mielck, A., Vogelmann, M., & Leidl, R. (2014). Health-related quality of life and socioeconomic status: Inequalities among adults with a chronic disease. *Health and Quality of Life Outcomes*, *12*(1), 58. https://doi.org/10.1186/1477-7525-12-58

Miller, W. R., & Rollnick, S. (2003). Review of 'motivational interviewing: Preparing people for change'. *American Journal of Forensic Psychology*, *21*(1), 83–85.

Mons, U., Raum, E., Krämer, H. U., Rüter, G., Rothenbacher, D., Rosemann, T., Szecsenyi, J., & Brenner, H. (2013). Effectiveness of a supportive telephone counseling intervention in type 2 diabetes patients: Randomized controlled study. *PLoS ONE*, *8*(10), e77954. https://doi.org/10.1371/journal.pone.0077954

Oksman, E., Linna, M., Hörhammer, I., Lammintakanen, J., & Talja, M. (2017). Cost-effectiveness analysis for a tele-based health coaching program for chronic disease in primary care. *BMC Health Services Research*, *17*(1), 138. https://doi.org/10.1186/s12913-017-2088-4

Olsen, J. M., & Nesbitt, B. J. (2010). Health coaching to improve healthy lifestyle behaviors: An integrative review. *American Journal of Health Promotion*, *25*(1), e1–e12. https://doi.org/10.4278/ajhp.090313-LIT-101

O'Neill, J., Tabish, H., Welch, V., Petticrew, M., Pottie, K., Clarke, M., Evans, T., Pardo Pardo, J., Waters, E., White, H., & Tugwell, P. (2014). Applying an equity lens to interventions: Using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of Clinical Epidemiology*, *67*(1), 56–64. https://doi.org/10.1016/j.jclinepi.2013.08.005

Pearce, G., Parke, H. L., Pinnock, H., Epiphaniou, E., Bourne, C. L. A., Sheikh, A., & Taylor, S. J. C. (2015). The PRISMS taxonomy of self-management support: Derivation of a novel taxonomy and initial testing of its utility. *Journal of Health Services Research & Policy*, *21*(2), 73–82. https://doi.org/10.1177/1355819615602725

Plass, D., Vos, T., Hornberg, C., Scheidt-Nave, C., Zeeb, H., & Kramer, A. (2014). Trends in disease burden in Germany: Results, implications and limitations of the global burden of disease study. *Deutsches Ärzteblatt International*, *111*(38), 629–638. https://doi.org/10.3238/arztebl.2014.0629

Relton, C., Torgerson, D., O'Cathain, A., & Nicholls, J. (2010). Rethinking pragmatic randomised controlled trials: Introducing the "cohort multiple randomised controlled trial" design. *BMJ*, *340*, c1066. https://doi.org/10.1136/bmj.c1066

Ring, N., Jepson, R., Hoskins, G., Wilson, C., Pinnock, H., Sheikh, A., & Wyke, S. (2011). Understanding what helps or hinders asthma action plan use: A systematic review and synthesis of the qualitative literature. *Patient Education and Counseling*, *85*(2), e131–e143. https://doi.org/10.1016/j.pec.2011.01.025

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological Methods*, *19*(3), 317–333. https://doi.org/10.1037/met0000013

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*(484), 1334–1344. https://doi.org/10.1198/016214508000000733

Sheiner, L. B., & Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics*, *57*(1), 6–15. https://doi.org/10.1016/0009-9236(95)90260-0

Sheridan, P. E., Mair, C. A., & Quiñones, A. R. (2019). Associations between prevalent multimorbidity combinations and prospective disability and self-rated health among older adults in Europe. *BMC Geriatrics*, *19*(1), 198. https://doi.org/10.1186/s12877-019-1214-z

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, *15*(3), 250–267. https://doi.org/10.1037/a0018719

Swoboda, C. M., Miller, C. K., & Wills, C. E. (2017). Impact of a goal setting and decision support telephone coaching intervention on diet, psychosocial, and decision outcomes among people with type 2 diabetes. *Patient Education and Counseling*, *100*(7), 1367–1373. https://doi.org/10.1016/j.pec.2017.02.007

Techniker Krankenkasse. (2008). *TK-Gesundheitscoach: Schulungshandbuch*. Techniker Krankenkasse.

Timpel, P., Oswald, S., Schwarz, P., & Harst, L. (2020). Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: An umbrella review of systematic reviews and meta-analyses. *Journal of Medical Internet Research*, *22*(3), e16791. https://doi.org/10.2196/16791

UNESCO. (2012). International standard classification of education ISCED 2011. Montreal: UNESCO Institute for Statistics. http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

Van Hecke, A., Heinen, M., Fernández-Ortega, P., Graue, M., Hendriks, J. M. L., Høy, B., Köpke, S., Lithner, M., & Van Gaal, B. G. I. (2017). Systematic literature review on effectiveness of self-management support interventions in patients with chronic conditions and low socio-economic status. *Journal of Advanced Nursing*, *73*(4), 775–793. https://doi.org/10.1111/jan.13159

Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., Abdulkader, R. S., Abdulle, A. M., Abebo, T. A., Abera, S. F., Aboyans, V., Abu-Raddad, L. J., Ackerman, I. N., Adamu, A. A., Adetokunboh, O., Afarideh, M., Afshin, A., Agarwal, S. K., Aggarwal, R., ... Murray, C. J. L. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: A systematic analysis for the global burden of disease study 2016. *The Lancet*, *390*(10100), 1211–1259. https://doi.org/10.1016/S0140-6736(17)32154-2

West, S. G., Cham, H., Thoemmes, F., Renneberg, B., Schulze, J., & Weiler, M. (2014). Propensity scores as a basis for equating groups: Basic principles and application in clinical treatment outcome research. *Journal of Consulting and Clinical Psychology*, *82*(5), 906–919. https://doi.org/10.1037/a0036387

WHO Regional Office for Europe. (1998). Wellbeing measures in primary health care/the Depcare project. WHO Regional Office for Europe: Copenhagen.

Zelen, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine*, *300*, 1242–1245. https://doi.org/10.1056/NEJM197905313002203

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Renneberg, B., Schulze, J., Böhme, S., West, S. G., & Schüz, B. (2022). Effectiveness and equity evaluation of an insurance-wide telephone-counseling program for self-management of chronic diseases: The Health Coach Study. *Applied Psychology: Health and Well-Being*, *14*(2), 606–625. https://doi.org/10.1111/aphw.12322