A prognostic model predicted deterioration in health-related quality of life in older patients with multimorbidity and polypharmacy

González-González, Ana I; Meid, Andreas D; Dinh, Truc S; Blom, Jeanet W; van den Akker, Marjan; Elders, Petra Jm; Thiem, Ulrich; De Gaudry, Daniela Küllenberg; Ma Swart, Karin; Rudolf, Henrik; Bosch-Lenders, Donna; Trampisch, Hans-Joachim; Meerpohl, Joerg J; Gerlach, Ferdinand M; Flaig, Benno; Kom, Ghainsom; Snell, Kym Ie; Perera, Rafael; Haefeli, Walter E; Glasziou, Paul P; Muth, Christiane

Published in:
Journal of Clinical Epidemiology

DOI:
10.1016/j.jclinepi.2020.10.006

TITLE

A prognostic model predicted deterioration in health-related quality of life in older patients with multimorbidity and polypharmacy

AUTHOR NAMES

*‡Ana I González-González[1,2] (0000-0002-1707-0596), *Andreas D. Meid[3] (0000-0003-3537-3205), Truc S Dinh[1] (0000-0002-9774-6751), Jeanet W Blom[4] (0000-0001-9989-4937), Marjan van den Akker[1,5,15] (0000-0002-1022-8637), Petra JM Elders[6] (0000-0002-5907-7219), Ulrich Thiem[7], Daniela Küllenberg De Gaudry[8] (0000-0003-4289-5853), Karin MA Swart[6] (0000-0003-2521-5780), Henrik Rudolf[10] (0000-0001-9114-3805), Donna Bosch-Lenders[5], Hans-Joachim Trampisch[10] (0000-0001-9136-1079), Joerg J Meerpohl[8,9] (0000-0002-1333-5403), Ferdinand M Gerlach[1], Benno Flaig[1], Ghainsom Kom[11], Kym IE Snell[11] (0000-0001-9373-6591), Rafael Perera[13] (0000-0003-24182091), Walter E Haefeli[3] (0000-0003-0672-6876), Paul P Glasziou[14] (0000-0001-7564-073X), Christiane Muth[1] (0000-0001-8987-182X)

AFFILIATIONS

[1] Institute of General Practice, Goethe University, 60590 Frankfurt am Main, Germany

[2] Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Madrid, Spain.

[3] Department of Clinical Pharmacology and Pharmacoepidemiology, University Hospital, 69120 Heidelberg, Germany

[4] Department of Public Health and Primary Care, Leiden University Medical Center, 2300RC Leiden, The Netherlands

[5] School of CAPHRI, Department of Family Medicine, Maastricht University, 6200 MD Maastricht, The Netherlands

[6] Department of General Practice and Elderly Care Medicine, Amsterdam UMC, Vrije Universiteit, Amsterdam Public Health Research Institute, 1007 MB Amsterdam, The Netherlands

[7] Chair of Geriatrics and Gerontology, University Clinic Eppendorf, 20246 Hamburg, Germany

[8] Institute for Evidence in Medicine (for Cochrane Germany Foundation), Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Breisacher Strasse 153, 79110 Freiburg, Germany

[9] Cochrane Germany, Cochrane Germany Foundation, Breisacher Strasse 153, 79110 Freiburg, Germany

[10] AMIB, Ruhr University, 44780 Bochum, Germany

[11] Techniker Krankenkasse (TK), 22765 Hamburg, Germany

[12] Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire ST5 5BG, United Kingdom

[13] Nuffield Department of Primary Care, University of Oxford, Oxford OX2 6GG, United Kingdom

[14] Centre for Research in Evidence-Based Practice, Bond University, Robina QLD 4226, Australia

[15] Academic Centre for General Practice / Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

* both authors contributed equally

‡**CORRESPONDING AUTHOR**

Ana I. González-González, Institute of General Practice, Johann Wolfgang Goethe University, Theodor-Stern-Kai 7, 60590 Frankfurt / Main, GERMANY, Tel.: ++49-(0)69-6301-4149/-5687

email: gonzalez-gonzalez@allgemeinmedizin.uni-frankfurt.de

ABSTRACT

Objective - To develop and validate a prognostic model to predict deterioration in health-related quality of life (dHRQoL) in older general practice patients with at least one chronic condition and one chronic prescription.

Study design and setting - We used individual participant data from five cluster-randomized trials conducted in the Netherlands and Germany to predict dHRQoL, defined as a decrease in EQ-5D-3L index score of ≥5 % after six-month follow-up in logistic regression models with stratified intercepts to account for between-study heterogeneity. The model was validated internally, and by using internal-external cross-validation (IECV).

Results – In 3,582 patients with complete data, of whom 1,046 (29.2 %) showed deterioration in HRQoL, 12/87 variables were selected that were related to single (chronic) conditions, inappropriate medication, medication underuse, functional status, well-being and HRQoL. Bootstrap internal validation showed a C-statistic of 0.71 (0.69 to 0.72), and a calibration slope of 0.88 (0.78 to 0.98). In the IECV loop, the model provided a pooled C-statistic of 0.68 (0.65 to 0.70) and calibration-in-the-large of 0 (-0.13 to 0.13). HRQoL/functionality had the strongest prognostic value.

Conclusion – The model performed well in terms of discrimination, calibration, and generalizability and might help clinicians identify older patients at high-risk of dHRQoL.

Registration - PROSPERO ID: CRD42018088129

Keywords –

Multimorbidity, polypharmacy, elderly, patient-centred care, quality of life, functional status, prognostic model.

Running title

A prognostic model to predict deterioration in health-related quality of life in older patients with multimorbidity and polypharmacy.

Word count

74 3671

WHAT IS NEW

Key findings

The PROPERmed prognostic model of future deterioration in health-related quality of life in older patients with multiple conditions and medications performed well in discrimination, calibration, and showed promising generalizability.

The strongest predictors in the model were health-related quality of life and functional status at baseline.

What does this add to what is already known?

PROPERmed-dHRQoL is the first prognostic model to predict deterioration in health-related quality of life in older patients with multiple conditions and medications that is based on an individual participant data meta-analysis.

What is the implication, what should change now?

External validation studies should confirm generalizability beyond internal-external cross-validation.

Measures of health-related quality of life and functional status at baseline, which proved to be the two prognostic variables that are of outstanding relative importance in the prognostic model, might help physicians to detect patients with multimorbidity and polypharmacy at risk for a potentially preventable deterioration.

INTRODUCTION

In aging populations, the increased incidence and severity of multiple (chronic) conditions (two or more) leads to deterioration in health-related quality of life (dHRQoL) (1). Patients with multiple conditions usually have several drug prescriptions (five or more), which increases the risk of overuse, underuse and misuse of medications (2). Potential consequences, such as falls, cognitive decline, loss of autonomy, and hospital admissions, are often severe and may contribute to dHRQoL, a key patient-reported outcome and one of the most relevant in older life (3–5).

Complex drug regimens and high treatment burden make the management of multimorbidity a significant challenge for physicians (6). They are also expensive for health care systems worldwide because they lead to an increase of health care utilization and cost (7). However, not all patients with multiple morbidities need complex care (8). As the multimorbid population is heterogeneous, it would be helpful to identify patients at high risk of dHRQoL because those with high baseline risk and/or higher severity of disease may generally be expected to benefit more from (complex) interventions (9). Furthermore, risk stratification may help allocate resources to the high-risk patients that are expected to benefit most from targeted interventions (10–12).

Prognostic models are generally considered to be important tools to help target interventions and improve clinical and economic outcomes (13). When focusing on dHRQoL, it is of fundamental importance to hinder as far as possible the natural slow decline in longitudinal trajectories of HRQoL punctuated by episodes of serious exacerbations that lead to hospital admissions (14,15), or, in other words, to provide ‘upstream’ preventive care to patients in need before ‘downstream’ morbidity and expenditures occur (13). High-performance prognostic models may be used to detect patients in need of supportive care (e.g. geriatric assessment and medication review) (10–12,16).

To the best of our knowledge, no dHRQoL prognostic model for older patients with multiple chronic conditions and polypharmacy exists. We therefore aimed to develop and validate a model to predict dHRQoL after six months of follow-up in older patients with at least one chronic condition and one chronic prescription, based on an individual participant data meta-analysis (IPD-MA). We used the IPD from a previously harmonized database that contains comprehensive patient-related data on socio-demographics, morbidity, medication, functional status, and well-being from five recent cluster-randomized trials conducted in German and Dutch general practices. We chose a prognostic modelling approach based on IPD-MA because it offers both statistical and clinical advantages over other modelling techniques by permitting the assessment of generalizability. Furthermore, the increased sample size and case-mix variability it provides may reduce overfitting and thus improve external performance (17).

MATERIALS AND METHODS

Source of data

We harmonized individual participant data (IPD) from five cluster-randomized trials that were conducted in the Netherlands and Germany between 2009 and 2012 to optimize pharmacological treatment in older chronically ill patients **(Supplemental Table 1)**. Although conducted in different health care systems, the included trials, namely ISCOPE (18), Opti-Med (19,20), PIL (Nederlands Trial Register, NTR2154) (21), PRIMUM (8,22), and RIME (Deutsches Register Klinischer Studien-ID, DRKS00003610), resemble each other in terms of key study characteristics. Four trials (PRIMUM, Opti-med, PIL, and RIME) compared a structured medication review consisting of several intervention components (i.e., complex interventions) with usual care, while ISCOPE used a functional geriatric approach to compare usual care with a proactive and integrated care plan. Details of the origin and preparation of the source data for the PROPERmed database (PRIMUM, Opti-Med, PIL, ISCOPE, RIME) will be published elsewhere.

[About here link to: Supplemental Table 1 on Main characteristics of the included trials]

Participants

At baseline, we included general practice patients aged 60 years or older with at least one chronic condition and one chronic prescription. We defined chronic conditions in accordance with O´Halloran´s list (23), and chronic prescriptions in the same way as the included trials (two weeks duration in PRIMUM, two months in ISCOPE, and three months in Opti-Med, PIL and RIME).

Outcome

We defined dHRQoL as a decrease of at least five percent from baseline to six-month follow-up in the 5 dimensions 3 level version of EuroQoL (EQ-5D-3L), operationalized using a Likert score. We considered this cut-off as clinically relevant because it corresponds to several studies' estimates of patients' perceptions of minimal important difference (MID) (24–26). In two of the Dutch trials (ISCOPE and PIL), the question relating to the item "mobility" was slightly modified from the original instrument, as it was frequently a missing value in older Dutch populations due to misinterpretation (27).

Prognostic variables

For candidates at baseline, 87 prognostic variables relating to socio-demographics, lifestyle, morbidity, medication, functional status, and well-being were considered for inclusion in the modelling process. The allocation of patients to control and intervention groups was also considered.

Socio-demographics and lifestyle

We collected IPD on age, sex, living situation, and educational level (28) from the trials. Information on smoking status was provided in three (PRIMUM, PIL, and RIME) of the five trials.

Morbidity

We used the second version of the International Classification of Primary Care (ICPC-2) (29) to describe a common list of individual chronic conditions across trials (patient-reported in RIME; in all others we used physician-reported information) and used a modified version of the Diederichs list for morbidity count, which included 15 of the 17 conditions identified in a systematic review (i.e. dementia, kidney and peripheral artery disease were not provided in two of the five trials) (30). The Charlson comorbidity index (31) was provided in two of the trials (PRIMUM and RIME), but could not be calculated for the other trials (e.g. because no information was provided on condition severity).

Medication

Potentially inappropriate prescriptions and medication underuse were mainly assessed using patient-reported medication data (except from ISCOPE which provided physician-reported information) by applying the criteria used in the EU-PIM list (32), STOPP-START criteria (33), the high-risk prescribing criteria applied by Dreischulte et al. (34), the Anticholinergic Drug Scale (ADS) (35,36), the Drug Burden Index (DBI) as a count variable (as the dosage that would have allowed the calculation of the index score was not available in the majority of IPD (37–39)), and Anticholinergic Drug Burden (ADB) (40).

Functional status and well-being

Trials used various instruments to measure functional status such as the Katz-15 (combination of KATZ-6 and Lawton IADL) questionnaire (41), the 13-item vulnerable elderly survey (VES-13) (42), and the Geriatric Giants VAS (GGV) scale (0-10) (43) developed *ad hoc* by one of the trials (Opti-Med). To standardize the metrics used in the scales of the instruments employed in the different trials, numerical values were subtracted from their overall mean (i.e., centred) and subsequently divided by their standard deviations (i.e., scaled) to obtain comparable values that would, however, require back-transformation for clinical interpretability.

The trials assessed the presence of depressive symptoms using different questionnaires (the 15-item Geriatric Depression Scale (GDS) (44,45), GDS-5 (46), SF-12 (47,48), and SF-36 (49). We

considered the standardized mean differences of the various instruments for the modelling approach. The presence of depressive symptoms was used as a binary variable for descriptive purposes and derived from the cut-offs of the original questionnaires used in the various trials. The presence of pain was defined as a binary variable using the categorical classification (no pain or any pain regardless of intensity) from the von Korff index (50), the SF-12 (47,48), the SF-36 (49), and the self-developed VAS scales or single questions used in two of the trials (i.e. Opti-Med, ISCOPE).

Regarding HRQoL at baseline, we used the above described EQ-5D-3L index score (51). In addition, we considered the two independent subscales from the HRQoL Comorbidity Index (52–54) as prognostic variables **(Supplemental Table 2)**.

[About here link to: Supplemental Table 2 on Prognostic variables and their definitions]

Sample size

The sample size reflected the number of available observations in the included trials. In order to calculate achievable performance based on the available sample size, we applied the formulae for minimum sample sizes (55). As we applied the calculation retrospectively, the sample size calculation only has exploratory character. This was part of the process of developing multivariable prediction models to obtain estimates for the heuristic shrinkage factor caused by the number of candidate predictors (55). Based on the sample size of our complete-case analysis and the use of empirical estimates of C-statistics and event frequencies to approximate the prediction model Cox-Snell R-squared's apparent performance (Cox-Snell R^2 of 0.12), we would expect a heuristic shrinkage factor of 0.84, which we considered acceptable.

Missing data

In addition to the core analysis of complete cases, we conducted sensitivity analyses using the missing-indicator method (MIM) (56,57) and multiple imputation. For the latter, we conducted six multiple imputations (MI) in five iterations (58), and pooled them according to Rubin's

Rules (59). For the original trials, stratification was used to graphically explore missing data patterns (60,61). This revealed the various contributions of sporadically and systematically missing values (variable not recorded in the trials). We performed multi-level multiple imputation to adjust for within-trial and between-trial variability (62).

When values were missing systematically, we did not consider the associated candidate prognostic variables in any of the trials (i.e. smoking status, Charlson comorbidity index).

Statistical analysis methods

Modelling framework to deal with within-study correlation and between-study heterogeneity in the IPD

Prognostic model development and validation relied on an established framework for developing and evaluating clinical prediction models in an IPD-MA (17). By virtue of their origins in different independent trials, the clustered data structure first had to be addressed. A stratified intercept model was fitted, which provided a different baseline risk for each trial. This approach was selected over a random intercept model because the validity of the normality assumption for the random intercept in differing random effects models cannot be checked and is open to doubt when five trials are conducted in different health care systems. A generalized linear model was therefore chosen using the logit link function (i.e., logistic model). To improve interpretability, we used effect coding rather than dummy coding in order to estimate trial-specific baseline risks (63). This produces a global intercept (overall average) and shows the deviation from the average for each trial. While in a one-stage meta-analysis for model development and internal validation, the study indicators account for the origin of the data, each study serves as a validation sample in an applied internal-external cross validation (IECV) (17,64).

Model development and variable selection

When developing the model, we defined it structurally by selecting variables using the so-called *Least Absolute Shrinkage and Selection Operator* (LASSO) (65). Age (assumed, like the

other continuous variables, to be linearly associated with outcome), sex, and the effect-coded indicators reflecting the trials' baseline risk, were not regularized. In order to obtain sparser models, we moved away from the default setting, which would have meant choosing the tuning parameter lambda as the value with the minimum mean cross-validated error ("optimal penalty"). In preference, we decided to be stricter and chose the most regularized model, meaning that the error was within one standard error of the minimum ("1-se rule" (66)). Variable importance was derived from the ranks of the absolute values of the final (standardized) coefficients (65). For subsequent cases, the model formula obtained using the LASSO technique was applied to models that were refitted using unpenalized maximum likelihood. We additionally calculated a uniform shrinkage factor from bootstrap internal validation; the uniform shrinkage factor corresponds to one minus the average of all calibration slopes of each bootstrap model applied to the original IPD.

Performance metrics

Predictive performance was assessed by simultaneously using 250 bootstrap samples internally (67), and employing IECV to assess generalizability (17,64). Model performance in terms of discriminatory ability to differentiate patients with dHRQoL from the rest was quantified using the C-statistic (equivalent to the area under the receiver-operating characteristic curve, ROC). Performance metrics for model calibration to assess agreement between observed event frequencies and predicted probabilities were based on the slope of the calibration curve and calibration-in-the-large (CITL), and additionally inspected visually by means of calibration plots (68).

Model validation

With regard to internal bootstrap validation, the prediction model was developed *de-novo* for each of the 250 bootstrap samples, thus maintaining the proportions of the original trial data in the IPD. Performance metrics were calculated from models fitted to the bootstrap samples that were subsequently applied to the original IPD. The mean difference across all bootstrap

samples was the estimated optimism, while the optimism-corrected performance metric was obtained by subtracting estimated optimism from the original apparent performance metric. In IECV loops in particular, CITL was used to reflect overall calibration. Mimicking the application in a new population, the IECV loop repeatedly selects variables and thus fits a prediction model in all but one of the IPD trials (i.e. training set), while also checking predictive performance in the omitted study (i.e. test set). We chose the conservative option of the average intercept of the IECV training set. As they are of special importance for external validation, we extracted the C-statistic and CITL estimate for each omitted study at each stage of the IECV loop (69). Based on the within-study correlation between the C-statistic and CITL obtained using a non-parametric bootstrap (70), the respective estimates were pooled using multivariate random-effects meta-analysis (71). Taking a Bayesian approach with an uninformative prior distribution, a multivariate *t*-distribution (of the pooled means and covariance matrix from the multivariate meta-analysis) was used as an approximate posterior distribution to assess the model's combined discrimination and average calibration performance. Requiring at least modest discriminatory ability of 0.65 and a CITL between -0.1 and 0.1, the proportion of samples from the posterior distributions that achieved this allowed us to calculate the probability of satisfying these requirements (70).

Technical implementation and reporting

All analyses were conducted using the R software environment in version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) with the key packages of *glmnet* (65), *metaphor* (71), *caret* (72)*, mice* (58), and pmsampsize (55).

This research study was reported in accordance with the TRIPOD statement **(Supplemental table 3)** (73).

[About here link to: Supplemental Table 3 on TRIPOD Checklist: Prediction Model Development and Validation]

RESULTS

Of all eligible 4,561 patients from the PROPERmed database for whom multiple imputation datasets were available, 3,582 patients with full data for all candidate prognostic variables were included in the complete-case population (**Figure 1**). In this subset, the HRQoL of 1,046 (29.2 %) patients deteriorated by at least five percent according to the EQ-5D-3L index at six-month follow-up: 105 (27.6 %) patients from PRIMUM, 94 (24.4 %) from Opti-Med, 131 (29.2 %) from PIL, 442 (32.8 %) from ISCOPE and 274 (26.9%) from RIME.

The mean age of the complete-case population was 78 (SD 7) years; 58 % were women, 96 % lived at home, and 88 % had a low/medium level of education. The population had an average of 3 (SD 2) chronic conditions (multimorbidity) and 8 (SD 4) chronic prescriptions (polypharmacy). Seventy-eight percent of patients were taking three or more medications. Sixty-seven percent suffered from pain and 20 % had depressive symptoms.

Table 1 and Supplemental table 4 show the prognostic variables both overall and stratified according to observed dHRQoL status in the complete-case population. In Supplemental Table 5, prognostic variables are shown both overall and stratified according to the interventional status of the original trials in the complete-case population. Supplemental figures 1 and 2 show the baseline HRQoL distribution across countries and study arms.

[About here Figure 1 on Flow chart and schematic course of action]

[About here Table 1 on Prognostic variables and statistically significant univariable associations with dHRQoL]

[About here link to: Supplemental table 4 on Prognostic variables and univariable associations with dHRQoL]

[About here link to: Supplemental table 5 on Candidate prognostic variables and outcome of the five randomized controlled trial stratified by interventional status]

[About here link to: Supplemental figure 1 on Baseline HRQoL distribution across countries]

[About here link to: Supplemental figure 2 on Baseline HRQoL distribution in study arms]

When developing the prognostic model for dHRQoL using the candidates' prognostic variables, variable selection using LASSO yielded a structural model with the items listed in **Table 2**. Refitting the LASSO-derived model formula to CC, MIM, and MI datasets yielded nearly identical performance metrics in terms of model discrimination (**Figure 2A**) and model calibration (**Figure 2B**). Variable importance metrics illustrated the predictive value of the individual prognostic variables (**Table 2**). Baseline quality of life and functional status showed the greatest prognostic relevance, with a relative contribution to the model's performance of 62% and 31% respectively (**Figure 2C**). Bootstrap internal validation from **Table 2** yielded an optimism-corrected C-statistic of 0.71 (95 % confidence interval: 0.69 to 0.72) which was close to the C-statistic of 0.72 and indicated good discrimination. An optimism-corrected calibration slope of 0.88 (0.78 to 0.98) indicated moderate calibration. In an explorative analysis, we grouped the prognostic variables according to clinical origin; this process consistently revealed the considerable significance of functional status and well-being to discriminatory performance (**Figure 2D**), while the model derived using variable selection was comparable to full models in internal validation metrics. Between-study heterogeneity was clearly visible in the stratified trial intercepts (**Table 2**). The model performed well for all trials used as validation datasets in the IECV loop, with a pooled C-statistic of 0.68 (0.65 to 0.70), a CITL of 0 (-0.13 to 0.13) (**Figure 3**) and between-study heterogeneity I^2 of 24.6 % and 78.6 % respectively. We also obtained a joint probability of 75 % of achieving a C-statistic of 0.65 and CITL between -0.1 and 0.1 in an independent but similar population.

[About here: Table 2 on Final multivariable analysis of dHRQoL at six-month follow-up]

[About here: Figure 2 on model development and validation]

[About here: Figure 3 on meta-analytical summary of IECV loop]

DISCUSSION

This is the first IPD-based prognostic model for dHRQoL in a population of older patients with multiple conditions (two or more) and polypharmacy (five or more prescriptions) in general practice. While the prognostic model discriminated well and demonstrated reasonable generalizability in the IECV, intercept recalibration to consider further populations of interest would nevertheless be necessary before implementation. Our model included a wide selection of prognostic variables related to demographics, prescribed medication, potentially inappropriate medication and omissions, functional status, and well-being, which all significantly contributed to the prediction of dHRQoL. Among them, baseline HRQoL (high face validity) was the most important, followed by functional status (well known to be associated with dHRQoL (74)). Simple counts of multimorbidity (30) and polypharmacy did not indicate that patients were at risk per se with regard to dHRQoL, contrary to what is found in the literature (7,75).

Using an IPD-MA to create a model based on primary research data provided a suitable and comprehensive source of information that covered all relevant dimensions that are required in a prognostic model of dHRQoL. The case-mix variability of this database, which includes patients from two different health care contexts and involves a reasonable time frame to avoid limiting external validity, helped us achieve good model performance and promising generalizability. Thus, the IPD framework allowed the generalizability of the prediction model to be estimated, as well as the probability of adequate performance in an independent population. However, the IPD-MA-based modelling approach also entailed the loss of some information (e.g., the smoking status variable was systematically missing, and consideration of common chronic conditions was limited) and made it difficult to clinically interpret some prognostic variables (e.g., standardization of functional status measures). Furthermore, the exclusion criteria of a short life expectancy and dementia limit the generalizability of the findings.

To the best of our knowledge, our dHRQoL prognostic model for older patients with chronic conditions and polypharmacy in general practice is the only one of its kind. Existing risk stratification tools that have been developed and validated to predict negative outcomes in older patients with multiple morbidities have focused mainly on predicting hospital (re-) admissions (76). The C-statistics of these tools varied between 0.5 and 0.85, with the highest C-statistics found in models that included functional status as an outcome (76). Two studies (77,78) that evaluated four risk tools with the aim of identifying people with multiple conditions that were at risk of reduced HRQoL were recently assessed in a NICE guideline review (79). All of these tools demonstrated poor discrimination and calibration in predicting dHRQoL, and their certainty of evidence according to GRADE (80) ranged from low to very low. To date and as far as we are aware, no relevant studies exist that predict dHRQoL in older populations based on polypharmacy or any other medication-related information.

According to the results of the PROPERmed prognostic model, assessment of health-related quality of life and functional status might help physicians to detect patients with multimorbidity and polypharmacy at risk for a potentially preventable deterioration. However, for use in our model, the latter would have to be standardized to take into account mean values and deviation in the target population. Additionally, we recommend using shrunken estimates to multiply the effects of our prognostic variables with the uniform shrinkage factor obtained from internal bootstrap validation. It is also important to consider how best to choose the baseline risk for dHRQoL (intercept) in the new population. While for the original trials an average intercept appeared reasonable for IECV (between-study heterogeneity I^2 of 78.6 % in CITL), implementation in a completely new setting may require adjustments to account for outcome frequencies, or even complete re-estimation (17). Therefore, implementation of the PROPERmed dHRQoL model in a completely new setting will require taking the intermediate steps mentioned above, especially as data from the target population is likely to differ from our own. Furthermore, the PROPERmed dHRQoL model should undergo

an impact assessment, whereby it is particularly important to evaluate its ability as a prognostic tool to prioritize (complex) interventions in general practice, and thus to determine whether it could actually help optimize medication regimens.

CONCLUSION

The first IPD-based prognostic model of dHRQoL in older patients with multiple chronic conditions and medication in general practice performed well in calibration, discrimination and might thus effectively assist in the identification of high-risk patients.

ACKNOWLEDGMENTS

The authors would like to thank all participating local data managers (Sandra Rauck, Masha Twellaar, Karin Aretz, Antonio Fenoy and Kiran Chapidi). We would also like to thank Phillip Elliott for editing the manuscript.

FUNDING

This work was supported by the German Innovation Funds in accordance with § 92a (2) Volume V of the Social Insurance Code (§ 92a Abs. 2, SGB V - Fünftes Buch Sozialgesetzbuch), grant number: 01VSF16018. The funding body did not play any role in the design of the study, the collection, analysis, and interpretation of data, and in writing the manuscript. Andreas D. Meid is funded by the Physician-Scientist Programme of Heidelberg University, Faculty of Medicine. Rafael Perera receives funding from the NIHR Oxford Biomedical Research Council (BRC), the NIHR Oxford Medtech and In-Vitro Diagnostics Co- operative (MIC), the NIHR Applied Research Collaboration (ARC) Oxford and Thames Valley, and the Oxford Martin School. Kym Snell is funded by the National Institute for Health Research School for Primary Care (NIHR SPCR Launching Fellowship). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

AUTHOR CONTRIBUTIONS

JWB, MvdA, UT, WEH, HJT, PJME, GK, JJM, DKdG, RP, PPG, ADM and CM contributed to the design of the PROPERmed study. CM is the guarantor. AIGG and ADM wrote the first draft of the manuscript. JWB, MvdA, DBL, UT, HR, HJT, PJME and CM represent the five included trials and provided all data needed for the IPD-MA. AIGG and TSD developed the harmonized PROPERmed database; KMAS, HR and BF supported. ADM performed the statistical analysis with the support of RP, KIES and HR. All authors contributed to the manuscript and agreed on its publication. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria were omitted.

COMPETING INTEREST

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any additional organizations for the submitted work; no financial relationships over the past three years with any organizations that might have an interest in the submitted work; no other relationships or activities that could have influenced the submitted work.

ETHICAL APPROVAL

The Ethics Commission of the Medical Faculty of the Johann Wolfgang Goethe University, Frankfurt / Main confirmed that no extra vote was necessary for the anonymous use of data from the PROPERmed IPD-MA (13/07/2017). All included trials were separately approved by the relevant ethics commissions as follows:

ISCOPE: The Medical Ethical Committee of Leiden University Medical Center approved the study (date: 30.06.2009, reference: P09.096).

Opti-Med: The Medical Ethics Committee of the VU University Medical Centre Amsterdam approved the study (date: 12.01.2012, reference: 2011/408).

PIL: The Medical Ethics Review Board Atrium-Orbis-Zuyd approved the study (date: 15.12.2009, reference: 09-T-72 NL3037.096.09).

PRIMUM: The Ethics Commission of the Medical Faculty of Johann Wolfgang Goethe University, Frankfurt / Main approved the study (date: 20/05/2010, reference: E 46/10). RIME: The Ethics Commission of Witten University / Herdecke also approved the study (date: 28.02.2012, reference: 147/2011).

REFERENCES

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. Lancet. 2012;380(9836):37–43.
2. Haefeli WE, Meid AD. Pill-count and the arithmetic of risk: Evidence that polypharmacy is a health status marker rather than a predictive surrogate for the risk of adverse drug events. Int J Clin Pharmacol Ther. 2018 Dec;56(12):572–6.
3. Rankin A, Cadogan CA, In Ryan C, Clyne B, Smith SM, Hughes CM. Core outcome set for trials aimed at improving the appropriateness of polypharmacy in older people in primary care. J Am Geriatr Soc. 2018 Jul;66(6):1206–12.
4. Beuscart J-B, Knol W, Cullinan S, Schneider C, Dalleur O, Boland B, et al. International core outcome set for clinical trials of medication review in multi-morbid older patients with polypharmacy. BMC Med. 2018 Dec 13;16(1):21.
5. Saqlain M, Ali H, Kamran S, Munir MU, Jahan S, Mazhar F. Potentially inappropriate medications use and its association with health-related quality of life among elderly cardiac patients. Qual Life Res. 2020 May 20;
6. Wallace E, Hinchey T, Dimitrov BD, Bennett K, Fahey T, Smith SM. A systematic review of the probability of repeated admission score in community-dwelling adults. J Am Geriatr Soc. 2013 Mar;61(3):357–64.
7. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev. 2011

Sep;10(4):430–9.

8. Muth C, Uhlmann L, Haefeli WE, Rochon J, van den Akker M, Perera R, et al. Effectiveness of a complex intervention on Prioritising Multimedication in Multimorbidity (PRIMUM) in primary care: results of a pragmatic cluster randomised controlled trial. BMJ Open. 2018;8(2):e017740.

9. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ. 1995 Nov 18;311(7016):1356–9.

10. Romskaug R, Skovlund E, Straand J, Molden E, Kersten H, Pitkala KH, et al. Effect of Clinical Geriatric Assessments and Collaborative Medication Reviews by Geriatrician and Family Physician for Improving Health-Related Quality of Life in Home-Dwelling Older Patients Receiving Polypharmacy. JAMA Intern Med. 2020 Feb 1;180(2):181.

11. Rankin A, Cadogan CA, Patterson SM, Kerse N, Cardwell CR, Bradley MC, et al. Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Database Syst Rev. 2018 Sep 3;9.

12. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. In: Smith SM, editor. Cochrane Database of Systematic Reviews [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2016. Available from: http://doi.wiley.com/10.1002/14651858.CD006560.pub3

13. Meid AD, Groll A, Schieborr U, Walker J, Haefeli WE. How can we define and analyse drug exposure more precisely to improve the prediction of hospitalizations in longitudinal (claims) data? Eur J Clin Pharmacol. 2017 Mar 24;73(3):373–80.

14. Meid AD, Quinzler R, Groll A, Wild B, Saum K-U, Schöttker B, et al. Longitudinal evaluation of medication underuse in older outpatients and its association with quality of life. Eur J Clin Pharmacol. 2016 Jul 29;72(7):877–85.

15. Lynn J. "Living long in fragile health: the new demographics shape end of life care",

Improving end of life care: why has it been so difficult? In: Hastings Center Report Special Report 35, no6. 2005. p. S14–8.

16. Smith SM, Soubhi H, Fortin M, Hudon C, O'Dowd T. Managing patients with multimorbidity: systematic review of interventions in primary care and community settings. BMJ. 2012 Sep 3;345(sep03 1):e5205–e5205.

17. Debray TPA, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. Stat Med. 2013;32(18):3158–80.

18. Blom J, Elzen W Den, Houwelingen AHV, Heijmans M, Stijnen T, Van Den Hout W, et al. Effectiveness and cost-effectiveness of a proactive, goal-oriented, integrated care model in general practice for older people. A cluster randomised controlled trial: Integrated systematic care for older people-the ISCOPE study. Age Ageing. 2016;45(1):30–41.

19. Willeboordse F, Hugtenburg JG, van Dijk L, Bosmans JE, de Vries OJ, Schellevis FG, et al. Opti-Med: the effectiveness of optimised clinical medication reviews in older people with 'geriatric giants' in general practice; study protocol of a cluster randomised controlled trial. BMC Geriatr. 2014;14(1):116.

20. Willeboordse F, Schellevis FG, Chau SH, Hugtenburg JG, Elders PJM. The effectiveness of optimised clinical medication reviews for geriatric patients: Opti-Med a cluster randomised controlled trial. Fam Pract. 2017;34(4):437–45.

21. Bosch-Lenders D, Maessen DWHA, Stoffers HEJH (Jelle), Knottnerus JA, Winkens B, van den Akker M. Factors associated with appropriate knowledge of the indications for prescribed drugs among community-dwelling older patients with polypharmacy. Age Ageing. 2016 May;45(3):402–8.

22. Muth C, Harder S, Uhlmann L, Rochon J, Fullerton B, Güthlin C, et al. Pilot study to test the feasibility of a trial design and complex intervention on PRIoritising

MUltimedication in Multimorbidity in general practices (PRIMUM pilot). BMJ Open. 2016 Jul 25;6(7):e011613.

23. O'Halloran J, Miller GC, Britt H. Defining chronic conditions for primary care with ICPC-2. Fam Pract. 2004 Aug;21(4):381–6.

24. Kaplan RM. The minimally clinically important difference in generic utility-based measures. COPD. 2005 Mar;2(1):91–7.

25. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005 Aug;14(6):1523–32.

26. Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review. Expert Rev Pharmacoecon Outcomes Res. 2014 Apr;14(2):221–33.

27. Lutomski JE, Baars MAE, Schalk BWM, Boter H, Buurman BM, den Elzen WPJ, et al. The development of The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): A Large-Scale Data Sharing Initiative. Bayer A, editor. PLoS One. 2013 Dec 4;8(12):e81673.

28. Eurostat Statistics Explained. International Standard Classification of Education (ISCED) [Internet]. 2016 [cited 2017 Nov 6]. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_(ISCED)

29. WONCA. ICPC-2. International Classification of Primary Care. Oxford: Oxford University Press; 1998.

30. Diederichs C, Berger K, Bartels DB. The Measurement of Multiple Chronic Diseases--A Systematic Review on Existing Multimorbidity Indices. J Gerontol - Ser A Biol Sci Med Sci. 2011 Mar 1;66A(3):301–11.

31. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis.

1987;40(5):373–83.

32. Renom-Guiteras A, Meyer G, Thürmann P a. The EU(7)-PIM list: a list of potentially inappropriate medications for older people consented by experts from seven European countries. Eur J Clin Pharmacol. 2015;71(7):861–75.

33. O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: Version 2. Age Ageing. 2015;44(2):213–8.

34. Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer prescribing — a trial of education, informatics, and financial incentives. N Engl J Med. 2016;374(11):1053–64.

35. Carnahan RM, Lund BC, Perry PJ, Pollock BG, Culp KR. The Anticholinergic Drug Scale as a measure of drug-related anticholinergic burden: associations with serum anticholinergic activity. J Clin Pharmacol. 2006;46:1481–6.

36. Carnahan RM, Lund BC, Perry PJ, Culp KR, Pollock BG. The relationship of an anticholinergic rating scale with serum anticholinergic activity in elderly nursing home residents. Psychopharmacol Bull. 2002;36(4):14–9.

37. Hilmer SN, Mager DE, Simonsick EM, Ling SM, Windham BG, Harris TB, et al. Drug burden index score and functional decline in older people. Am J Med. 2009 Dec;122(12):1142-1149.e1-2.

38. Cao Y-J, Mager DE, Simonsick EM, Hilmer SN, Ling SM, Windham BG, et al. Physical and cognitive performance and burden of anticholinergics, sedatives, and ACE inhibitors in older women. Clin Pharmacol Ther. 2008 Mar;83(3):422–9.

39. Hilmer SN. A drug burden index to define the functional burden of medications in older people. Arch Intern Med. 2007 Apr 23;167(8):781.

40. Durán CE, Azermai M, Vander Stichele RH. Systematic review of anticholinergic risk scales in older adults. Eur J Clin Pharmacol. 2013 Jul;69(7):1485–96.

585 41. Palmer M, Harley D. Models and measurement in disability: an international review.
586 Health Policy Plan. 2012 Aug;27(5):357–64.
587 42. Saliba D, Elliott M, Rubenstein LZ, Solomon DH, Young RT, Kamberg CJ, et al. The
588 Vulnerable Elders Survey: a tool for identifying vulnerable older people in the
589 community. J Am Geriatr Soc. 2001 Dec;49(12):1691–9.
590 43. Isaacs B. An Introduction to Geriatrics. London: Bailliere, Tindall & Cassell; 1965.
591 44. Sheikh JI, Yesavage JA, Brooks JO, Friedman L, Gratzinger P, Hill RD, et al. Proposed
592 factor structure of the Geriatric Depression Scale. Int psychogeriatrics. 1991;3(1):23–8.
593 45. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and
594 validation of a geriatric depression screening scale: a preliminary report. J Psychiatr
595 Res. 1982;17(1):37–49.
596 46. Hoyl MT, Alessi CA, Harker JO, Josephson KR, Pietruszka FM, Koelfgen M, et al.
597 Development and testing of a five-item version of the Geriatric Depression Scale. J Am
598 Geriatr Soc. 1999 Jul;47(7):873–8.
599 47. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-
600 validation of item selection and scoring for the SF-12 Health Survey in nine countries:
601 results from the IQOLA Project. International Quality of Life Assessment. J Clin
602 Epidemiol. 1998 Nov;51(11):1171–8.
603 48. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of
604 scales and preliminary tests of reliability and validity. Med Care. 1996 Mar;34(3):220–
605 33.
606 49. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al.
607 Translation, validation, and norming of the Dutch language version of the SF-36 Health
608 Survey in community and chronic disease populations. J Clin Epidemiol. 1998
609 Nov;51(11):1055–68.
610 50. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain.

1992 Aug;50(2):133–49.

51. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New York). 1990 Dec;16(3):199–208.

52. Mukherjee B, Ou H-T, Wang F, Erickson SR. A new comorbidity index: the health-related quality of life comorbidity index. J Clin Epidemiol [Internet]. 2011 Mar;64(3):309–19. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0895435610003471

53. Ou H-T, Mukherjee B, Erickson SR, Piette JD, Bagozzi RP, Balkrishnan R. Comparative Performance of Comorbidity Indices in Predicting Health Care-Related Behaviors and Outcomes among Medicaid Enrollees with Type 2 Diabetes. Popul Health Manag. 2012 Aug;15(4):220–9.

54. Cheng L, Cumber S, Dumas C, Winter R, Nguyen KM, Nieman LZ. Health related quality of life in pregeriatric patients with chronic diseases at urban, public supported clinics. Health Qual Life Outcomes. 2003 Oct 31;1(63):63.

55. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019 Mar 30;38(7):1276–96.

56. van der Heijden GJMG, T. Donders AR, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. J Clin Epidemiol. 2006 Oct;59(10):1102–9.

57. Groenwold RHH, White IR, Donders ART, Carpenter JR, Altman DG, Moons KGM. Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. CMAJ. 2012 Aug 7;184(11):1265–9.

58. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;1(3).

59. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.

60. Zhang Z. Missing data exploration: highlighting graphical presentation of missing pattern. Ann Transl Med. 2015 Dec;3(22):356.
61. Kowarik A, Templ M. Imputation with the R Package VIM. J Stat Softw. 2016;74(7).
62. Jolani S, Debray TPA, Koffijberg H, van Buuren S, Moons KGM. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. Stat Med. 2015 May 20;34(11):1841–63.
63. Te Grotenhuis M, Pelzer B, Eisinga R, Nieuwenhuis R, Schmidt-Catran A, Konig R. When size matters: advantages of weighted effect coding in observational studies. Int J Public Health. 2017 Jan;62(1):163–7.
64. Royston P, Parmar MKB, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. Stat Med. 2004;23(6):907–26.
65. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.
66. Thao LTP, Geskus R. A comparison of model selection methods for prediction in the presence of multiply imputed data. Biom J. 2019 Mar;61(2):343–56.
67. Efron B, Tibshirani R. An introduction to the bootstrap. CRC Boca Raton London New York Washington, D.C.: Chapman & Hall; 1993.
68. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Obuchowski N, Pencina MJ, et al. Assessing the performance of prediction models: A framework for some traditional and novel measures. Epidemiology. 2010;21(1):128–38.
69. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019 Dec 16;17(1):230.
70. Snell KIE, Hua H, Debray TPA, Ensor J, Look MP, Moons KGM, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. J Clin Epidemiol. 2016 Jan;69:40–50.

71. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3).
72. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5).
73. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015 Jan 6;162(1):W1.
74. Calderón-Larrañaga A, Vetrano DL, Ferrucci L, Mercer SW, Marengoni A, Onder G, et al. Multimorbidity and functional impairment: bidirectional interplay, synergistic effects and common pathways. J Intern Med. 2018 Oct 24;
75. Brilleman SL, Salisbury C. Comparing measures of multimorbidity to predict outcomes in primary care: a cross sectional study. Fam Pract. 2013 Apr 1;30(2):172–8.
76. Alonso-Morán E, Nuño-Solinis R, Onder G, Tonnara G. Multimorbidity in risk stratification tools to predict negative outcomes in adult population. Eur J Intern Med. 2015;26(3):182–9.
77. Fortin M, Hudon C, Dubois M-F, Almirall J, Lapointe L, Soubhi H. Comparative assessment of three different indices of multimorbidity for studies on health-related quality of life. Health Qual Life Outcomes. 2005 Nov 23;3:74.
78. Beaton K, Grimmer K, Milanese S, Atlas A. Additional measures do not improve the diagnostic accuracy of the Hospital Admission Risk Profile for detecting downstream quality of life in community-dwelling older people presenting to a hospital emergency department. Clin Interv Aging. 2014 Jan;233.
79. NICE Clinical Guideline. Multimorbidity: clinical assessment and management [Internet]. 2016. Available from: https://www.nice.org.uk/guidance/ng56
80. GRADE Working Group. GRADE [Internet]. [cited 2016 Jan 27]. Available from:

http://www.gradeworkinggroup.org

Figure legends

Figure 1: Flow chart and schematic course of action

CC = Complete Cases; IPD = Individual Participant Data; LASSO = Least Absolute Shrinkage and Selection Operator; MI = Multiply Imputed; MIM = Missing-Indicator Method; dHRQoL = deterioration in Health-Related Quality of Life.

Figure 2: **Model development and validation**. (A) By yielding receiver-operating characteristic (ROC) curves, the model's estimates of sensitivity and specificity for calculated risks discriminate between patients with and without dHRQoL. ROC curves are visualized for the following study populations: complete cases (CC), one multiply imputed dataset (MI), and data added using the missing-indicator method (MIM). The added lines mark the median risk cut-off of 0.41, with a sensitivity of 72% and specificity of 59%. (B) Similarly, calibration curves are generated by plotting predicted event probabilities against (cumulative) event frequencies. (C) Scrutinizing the impact of model parameters, a variable importance plot highlights their relative contribution to model performance, adjusted in relation to the most important prognostic variable. (D) Exploring the influence of variable origin, we fitted models composed of variables that are sociodemographic and lifestyle-related alone (α), or combinations of α and morbidity-related (β), medication-related (γ) predictors, and / or predictors related to functional status and well-being (δ) in accordance with **Table 1.** Resulting estimates of C-statistics are presented for bootstrap internal validation and internal-external cross-validation (IECV) if all available variables were included into the model (i.e., full model – grey circles) or only those having actually been selected during model development (black circles).

Figure 3: Meta-analytic summary of model generalizability.

A bivariate random-effects meta-analysis was conducted to determine the pooled performance metrics of C-statistics and calibration-in-the-large (CITL) from internal-external cross-validation (IECV), with the respective trial serving as the validation set for the model that

was refitted in the remaining trials. The Forest plot visualizes trial-specific estimates and their pooled results.

Supplemental figure 1: Baseline HRQoL distribution across countries.

Boxplot of HRQoL measurements at baseline on the respective EQ5D scale of the country the original study came from. Distinct values from the original studies are superimposed to highlight between-study variability.

Supplemental figure 2: Baseline HRQoL distribution in study arms.

Boxplot of HRQoL measurements at baseline on the respective EQ5D scale of the country the original study came from, according to interventional status. Distinct values from the original studies are superimposed to highlight between-study variability.

Table 1. Candidate prognostic variables and statistically significant univariable associations with dHRQoL

Candidate prognostic variable	dHRQoL (complete-case population)		Descriptive univariable p-value
	No n = 2,536	Yes n = 1,046	
Sociodemographic and lifestyle-related			
Age - Mean (SD)	77.2 (6.8)	78.3 (6.9)	< 0.001
Sex (female) - Frequency (%)	1,449 (57.1)	627 (59.9)	0.122
Living situation (Institutionalized living) - Frequency (%)	87 (3.4)	59 (5.6)	0.003
Educational level - Frequency (%)			
- Low	1,018 (40.1)	472 (45.1)	
- Medium	1,206 (47.6)	469 (44.8)	0.024
- High	312 (12.3)	105 (10.0)	0.011
Morbidity-related			
Coronary heart disease - Frequency (%)	817 (32.2)	393 (37.6)	0.002
Medication-related			
Drugs for acid-related disorders - Frequency (%)	950 (68.3)	441 (31.7)	0.009
Systemic corticosteroids instead of inhaled corticosteroids for maintenance therapy in moderate-severe COPD - STOPP G2 - Frequency (%)	15 (0.6)	15 (1.4)	0.015
START criteria* – Median (IQR)	1 (2)	1 (2)	0.002
START criteria* (modified) - Frequency (%)	1,425 (56.2)	634 (60.6)	0.015
Heart failure and/or documented coronary artery disease and NO ACE inhibitor - START A6 - Frequency (%)	255 (10.1)	160 (15.3)	< 0.001
Ischemic heart disease and NO beta-blocker - START A7 - Frequency (%)	203 (8.0)	117 (11.2)	0.003
Diabetes and NO ACE inhibitor or ARB - START F1 - Frequency (%)	150 (5.9)	95 (9.1)	0.001
Functional status and well-being-related			
Functional status – Mean (SD)	-0.123 (0.92)	0.044 (0.99)	< 0.001
Depression ** – Frequency (%)	485 (19.1)	201 (19.2)	0.95
Pain – Frequency (%)	1,728 (68.1)	675 (64.5)	0.037
Health-related quality of life comorbidity index, mental *** – Median (IQR)	1 (1)	1 (1)	0.044

Quality of life: EQ-5D, version 3L, Index value (baseline) – Mean (SD)	0.70 (0.26)	0.81 (0.19)	< 0.001

This table shows candidate prognostic variables stratified according to observed dHRQoL status and univariable associations.

ACE = angiotensin-converting enzyme; ARB = angiotensin II receptor blockers; ATC = anatomical therapeutic chemical; COPD = chronic obstructive pulmonary disease; IQR = interquartile range; SD = standard deviation; dHRQoL= deterioration in health-related quality of life.

* Fifteen START criteria were considered.

**Depression considered possible in case of a positive score on either of the two provided scales (GDS/SF).

*** Score calculated considering a maximum count of 6 conditions/13 points.

Table 2. Final multivariable analysis for dHRQoL at six-month follow-up

Selected prognostic factor	System of measurement	Estimate*	standard error	p value
(Intercept)**		-4.457	0.581	0.000
Age	Years	0.000	0.007	0.969
Sex (male)		-0.175	0.084	0.037
Coronary heart disease (Myocardial infarction and/or angina pectoris) - ICPC-2 codes K74, K75, K76	ICPC-2 codes K74, K75, K76	0.216	0.094	0.022
Drugs for acid-related disorders	ATC code A02	0.274	0.082	0.001
Systemic corticosteroids rather than inhaled corticosteroids for maintenance therapy in moderate-severe COPD - STOPP criteria G2	(ATC codes H02AB OR H02BX) AND (ICPC-2 codes R79, R95 OR R96) NOT (ATC codes R03BA OR R03AK)	1.108	0.432	0.010
START criteria count	15 START criteria were included	-0.003	0.036	0.934
ACE inhibitor with heart failure and/or documented coronary artery disease - START criteria A6	(ICPC-2 codes K74, K75, K76, K77) NOT (ATC codes C09A OR C09B OR C09C OR C09D)	0.212	0.141	0.133
ACE inhibitor or ARB (if intolerant of ACE inhibitor) in diabetes with evidence of renal disease i.e. dipstick proteinuria or microalbuminuria - START criteria F1	(ICPC-2 codes T89 OR T90) NOT (ATC codes C09A OR C09B OR C09C OR C09D)	0.386	0.159	0.015
Functional status	Standardized values taken from the VES-13, Katz-15 and GG mobility instruments used in the original studies	0.557	0.053	0.000
Depression	Cut-offs for diagnosis of depression taken from the GDS 15/5 or SF12/36 instruments	0.363	0.112	0.001

Mental Component Summary score from health-related quality of life comorbidity index	Score calculated according to the modified instrument: maximum count 6 conditions, 13 points	0.072	0.032	0.026
Quality of life: EQ-5D, version 3L, Index value (baseline)	Time Trade-Off values for EQ-5D-3L in German and Dutch populations	4.175	0.263	0.000

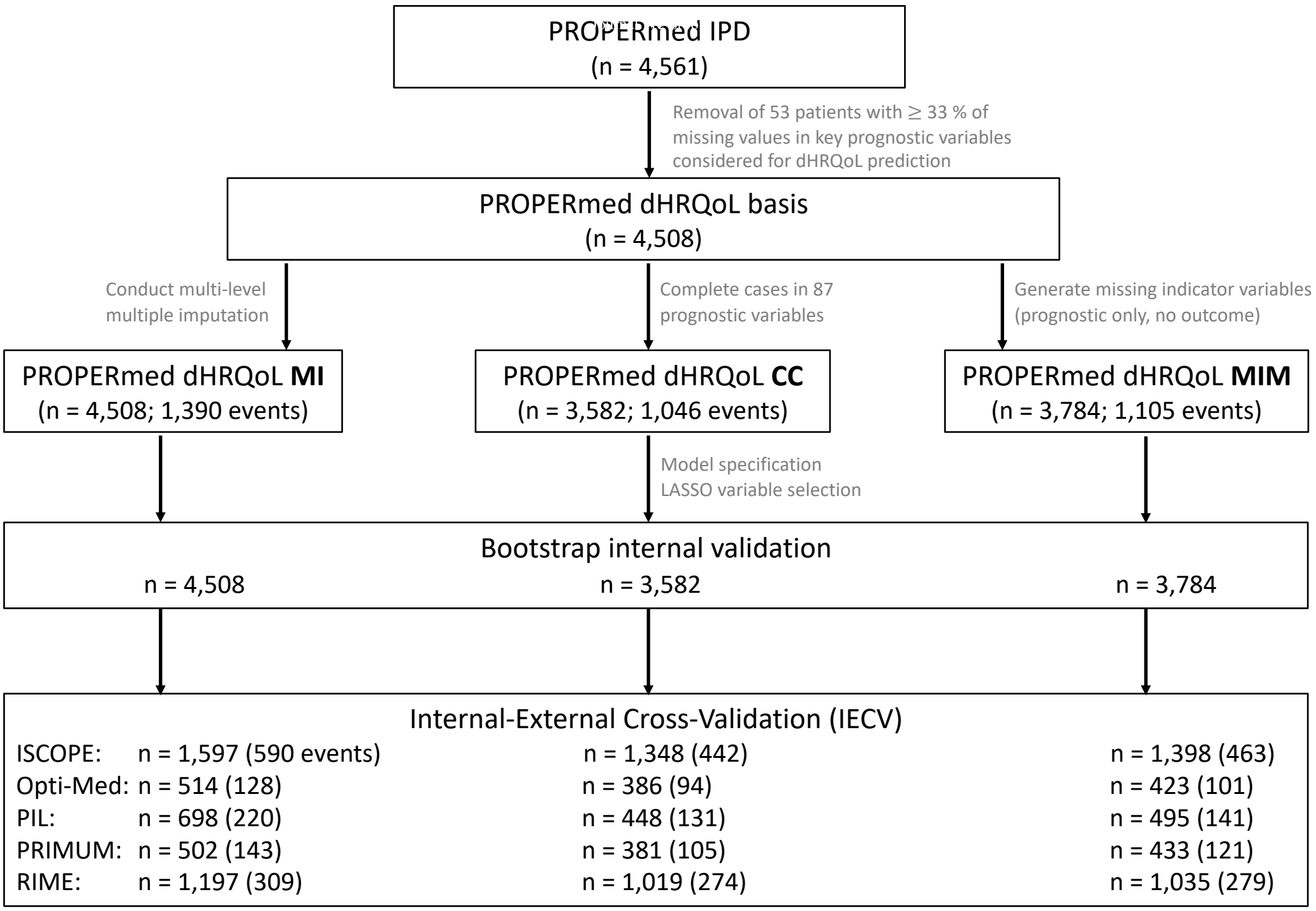
Baseline risks of studies (estimates): RIME -0.136, Opti-Med -0.175, PRIMUM -0.165, PIL 0.000 and ISCOPE 0.476.

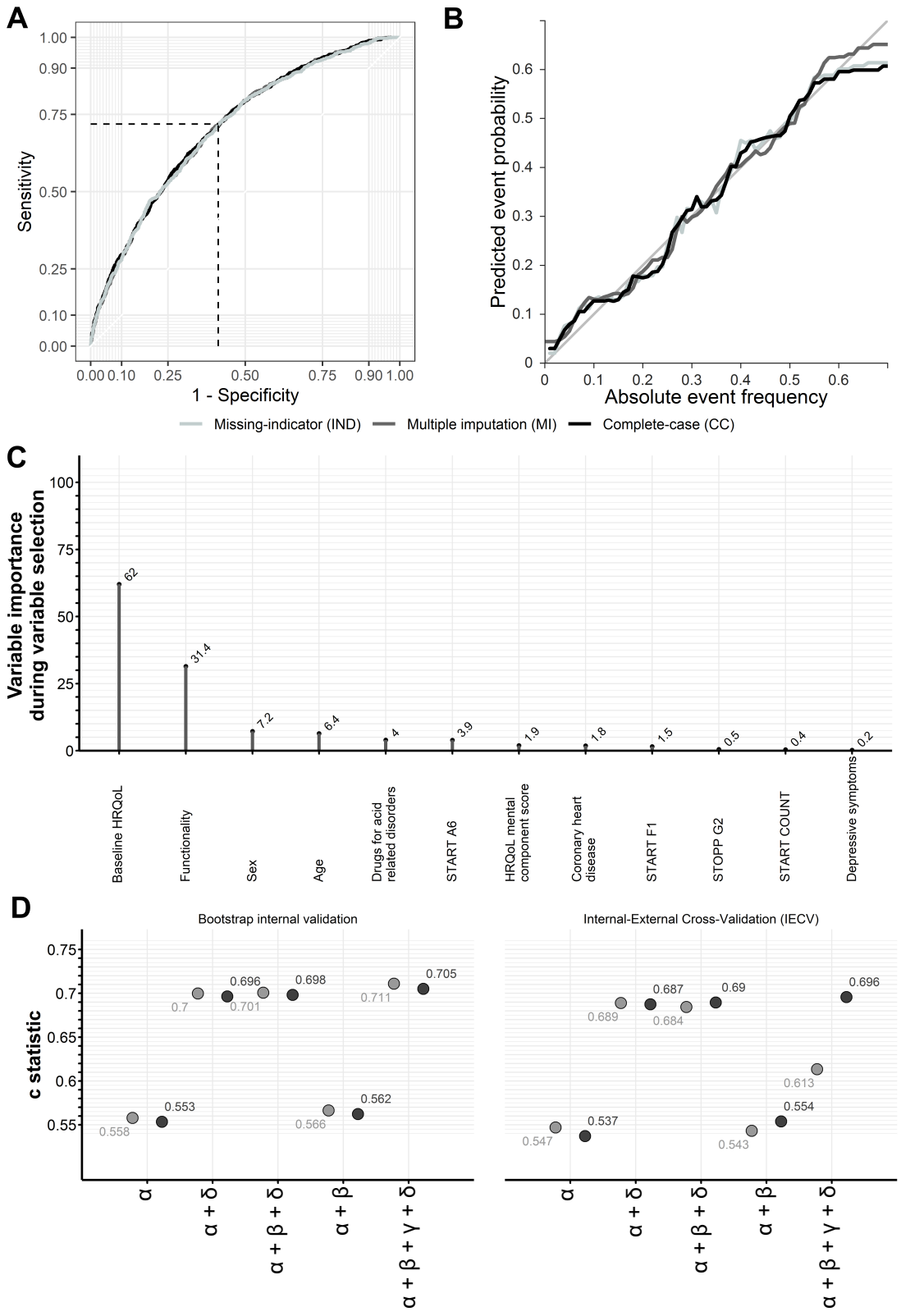
ACE = angiotensin converting enzyme; ARB = angiotensin II receptor blockers; ATC = anatomical therapeutic chemical; COPD = chronic obstructive pulmonary disease; GDS = geriatric depression scale; GG = geriatric giant; Katz-15; ICPC = international classification of primary care; MCS = Modified health-related quality of life comorbidity index, mental; SF = short form survey; TTO = time trade-off; VES = vulnerable elders survey; dHRQoL= deterioration in health-related quality of life.

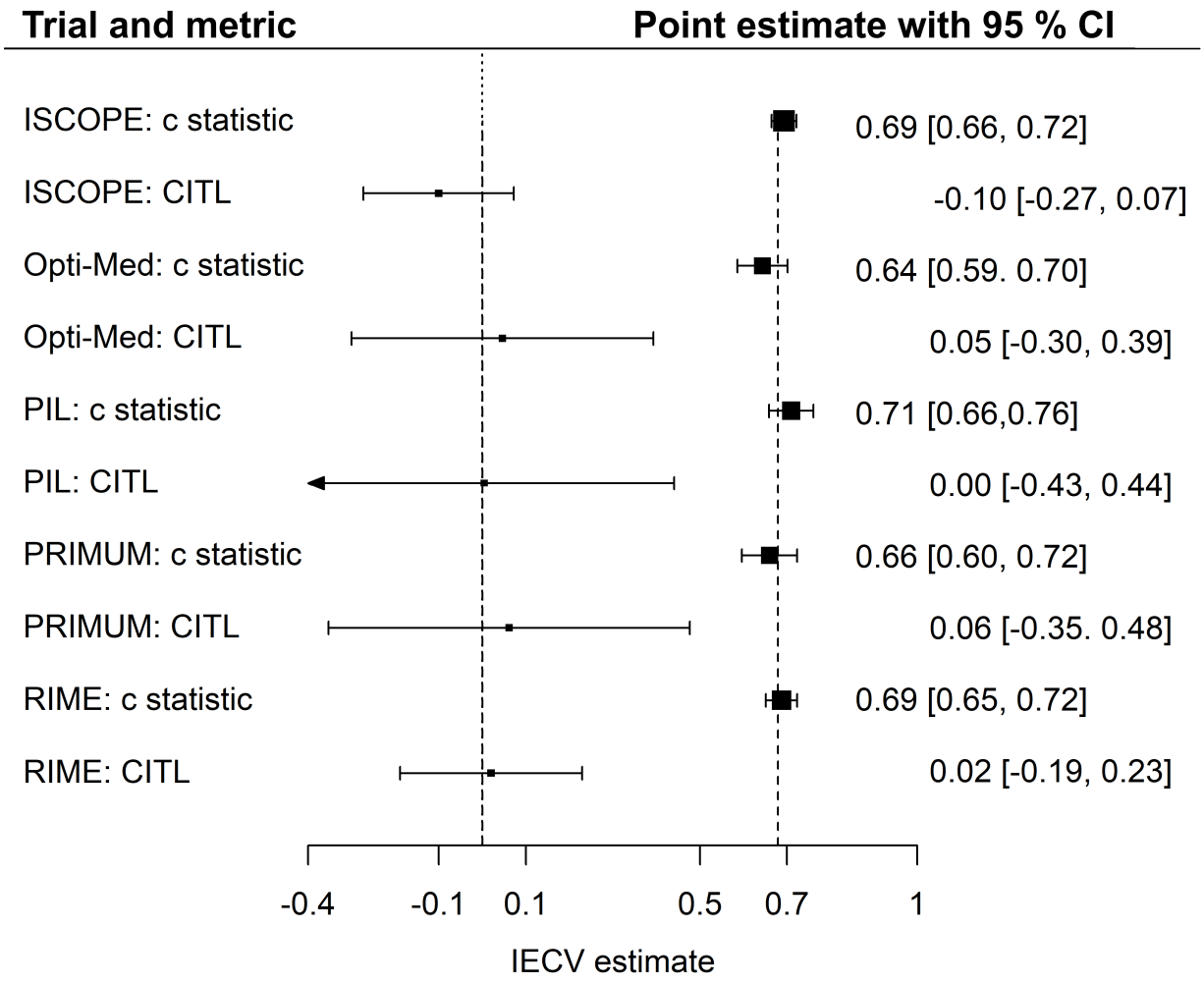
*Estimate = Parameter estimate of the maximum-likelihood fitted logistic regression model (possibly to be multiplied with the uniform shrinkage factor of 0.88).

**Intercept = Overall baseline risk for dHRQoL.

***Depression considered possible in case of a positive score on either of the two provided scales (GDS/SF).


PROPERmed IPD
(n = 4,561)
Removal of 53 patients with ≥ 33 % of missing values in key prognostic variables considered for dHRQoL prediction
PROPERmed dHRQoL basis
(n = 4,508)
Conduct multi-level multiple imputation
Complete cases in 87 prognostic variables
Generate missing indicator variables (prognostic only, no outcome)
PROPERmed dHRQoL MI
(n = 4,508; 1,390 events)
PROPERmed dHRQoL CC
(n = 3,582; 1,046 events)
PROPERmed dHRQoL MIM
(n = 3,784; 1,105 events)
Model specification
LASSO variable selection
Bootstrap internal validation
n = 4,508
n = 3,582
n = 3,784
Internal-External Cross-Validation (IECV)
ISCOPE: n = 1,597 (590 events) n = 1,348 (442) n = 1,398 (463)
Opti-Med: n = 514 (128) n = 386 (94) n = 423 (101)
PIL: n = 698 (220) n = 448 (131) n = 495 (141)
PRIMUM: n = 502 (143) n = 381 (105) n = 433 (121)
RIME: n = 1,197 (309) n = 1,019 (274) n = 1,035 (279)


A
Sensitivity
1 - Specificity
B
Predicted event probability
Absolute event frequency
Missing-indicator (IND)
Multiple imputation (MI)
Complete-case (CC)
C
Variable importance during variable selection
62
31.4
7.2
6.4
4
3.9
1.9
1.8
1.5
0.5
0.4
0.2
Baseline HRQoL
Functionality
Sex
Age
Drugs for acid related disorders
START A6
HRQoL mental component score
Coronary heart disease
START F1
STOPP G2
START COUNT
Depressive symptoms
D
Bootstrap internal validation
Internal-External Cross-Validation (IECV)
c statistic
α
α + δ
α + β + δ
α + β
α + β + γ + δ
0.7
0.696
0.701
0.698
0.711
0.705
0.558
0.553
0.566
0.562
0.689
0.687
0.684
0.69
0.696
0.613
0.547
0.537
0.543
0.554


Trial and metric
Point estimate with 95 % CI
ISCOPE: c statistic
0.69 [0.66, 0.72]
ISCOPE: CITL
-0.10 [-0.27, 0.07]
Opti-Med: c statistic
0.64 [0.59. 0.70]
Opti-Med: CITL
0.05 [-0.30, 0.39]
PIL: c statistic
0.71 [0.66,0.76]
PIL: CITL
0.00 [-0.43, 0.44]
PRIMUM: c statistic
0.66 [0.60, 0.72]
PRIMUM: CITL
0.06 [-0.35. 0.48]
RIME: c statistic
0.69 [0.65, 0.72]
RIME: CITL
0.02 [-0.19, 0.23]
-0.4
-0.1
0.1
0.5
0.7
1
IECV estimate

HIGHLIGHTS

- Multimorbidity and polypharmacy increase the risk of deterioration in quality of life.
- First IPD-based quality-of-life prognostic model for older multimorbid patients.
- Model performed well in terms of discrimination, calibration and generalizability.
- Baseline quality of life and functional status have the strongest prognostic power.
- Quality of life/functionality appraisal might help identify high-risk patients.

AUTHOR CONTRIBUTIONS

JWB, MvdA, UT, WEH, HJT, PJME, GK, JJM, DKdG, RP, PPG, ADM and CM contributed to the design of the PROPERmed study. CM is the guarantor. AIGG and ADM wrote the first draft of the manuscript. JWB, MvdA, DBL, UT, HR, HJT, PJME and CM represent the five included trials and provided all data needed for the IPD-MA. AIGG and TSD developed the harmonized PROPERmed database; KMAS, HR and BF supported. ADM performed the statistical analysis with the support of RP, KIES and HR. All authors contributed to the manuscript and agreed on its publication. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria were omitted.

WHAT IS NEW

Key findings

The PROPERmed prognostic model of future deterioration in health-related quality of life in older patients with multiple conditions and medications performed well in discrimination, calibration, and showed promising generalizability.

The strongest predictors in the model were health-related quality of life and functional status at baseline.

What does this add to what is already known?

PROPERmed-dHRQoL is the first prognostic model to predict deterioration in health-related quality of life in older patients with multiple conditions and medications that is based on an individual participant data meta-analysis.

What is the implication, what should change now?

External validation studies should confirm generalizability beyond internal-external cross-validation.

Measures of health-related quality of life and functional status at baseline, which proved to be the two prognostic variables that are of outstanding relative importance in the prognostic model, may help physicians to detect patients with multimorbidity and polypharmacy at risk for a potentially preventable deterioration.