COMMUNITY MEDICINE AND EDUCATION JOURNAL
Vol. 7, No. 2, 2026, pp. 136–152
e-ISSN 2774-2962 |
hmpublisher.com/index.php/cmej

SYSTEMATIC REVIEW / META-ANALYSIS

# The Role of Family Physician-Based Models in Integrated Primary Care: A Systematic Review and Meta-Analysis of Patient-Experienced Primary Care Quality

Luh Vida Sasmitha[1*], Putu Aryani[2]

[1] *Master of Public Health Study Programme, Faculty of Medicine, Universitas Udayana, Denpasar, Indonesia*
[2] *Department of Public Health and Preventive Medicine, Faculty of Medicine, Universitas Udayana, Denpasar, Indonesia*

ARTICLE HISTORY
**Received:** *August 20, 2026*
**Revised:** *September 15, 2026*
**Accepted:** *October 2, 2026*
**Available online:** *October 7, 2026*
**Published:** *October 8, 2026*

KEYWORDS
Integrated primary care
Meta-analysis
Physicians, Family
Primary Health Care
Quality of Health Care

CORRESPONDING AUTHOR
Luh Vida Sasmitha
sasmithavida@gmail.com

DOI
https://doi.org/10.37275/cmej.v7i2.952

## ABSTRACT

**Background:** Indonesia's Integrated Primary Care (ILP) reform relies on doctors to lead life-course clusters in community health centres, but the added value of family physician-based care is unknown.

**Objective:** To assess how family physician-based models relate to the primary care quality experienced by users.

**Methods:** PubMed and Scopus were searched from database inception for comparative studies of family physician-based models (contracted family doctors, Family Health Strategy teams, family medicine-trained physicians) versus conventional primary care reporting user-rated Primary Care Assessment Tool or equivalent scores. Hedges *g* was pooled with random-effects models (Hartung–Knapp); certainty was rated with GRADE.

**Results:** Fifteen cross-sectional studies from China and Brazil were included; twelve (13 comparisons, 25,101 users) contributed to the primary analysis. Family physician-based models were associated with higher overall quality ($g$=0.37, 95% CI 0.20 to 0.55; $I^2$=95.3%; prediction interval −0.18 to 0.93), about 0.6 points on the 0–10 PCATool scale. Associations were strongest for longitudinality ($g$=0.42) and community orientation ($g$=0.32) and not significant for coordination ($g$=0.12) or comprehensiveness ($g$=0.17). The direction was consistent across countries and sensitivity analyses, but trim-and-fill attenuated the estimate to $g$=0.19 and certainty was very low.

**Conclusion:** Family physician-based primary care was associated with better experienced continuity and community orientation, but not with better coordination or comprehensiveness. ILP should link registered populations to physician-led teams while strengthening referral systems and service scope.

**Copyright:** © 2026 The Author(s). Published by HM Publisher in collaboration with CMHC Research Center. The authors retain copyright and grant the journal the right of first publication. **Access license:** This is an open-access article under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), https://creativecommons.org/licenses/by-nc-sa/4.0/, which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and derivative works are shared under the same license.

## 1. Introduction

Primary health care is the level of the health system at which most people first seek help, and its strength is judged less by the number of facilities than by how well it performs a small set of core functions: accessible first contact, a continuing relationship between a person and a clinician or team, comprehensive services for common needs, and coordination with the rest of the system. In Indonesia these functions are delivered mainly through community health centres (Puskesmas), their sub-district posts and the village-level integrated health posts. Since 2023 the Ministry of Health has reorganised this network through Integrated Primary Care (Integrasi Layanan Primer, ILP), which replaced programme-based service lines with life-course clusters, extended services to sub-district and village posts, and introduced local-area monitoring dashboards.[1] Early evaluations of ILP found that limited human resources and capacity, fragmented governance and information systems, insufficient financing and incomplete readiness of village health posts and of the nursing workforce constrained implementation.[2–4]

The reform is taking place in a system in which the use of primary care under the national health insurance scheme differs between socioeconomic and demographic groups, and in which potentially preventable hospitalisations vary substantially across districts, indicating uneven performance of first-contact care.[5–7] Chronic disease management illustrates the problem: providers in Jakarta primary care centres reported gaps in referral pathways and in hypertension and diabetes management, fewer than

half of adults with raised blood pressure or glucose at community screening completed follow-up at a Puskesmas, and performance-based capitation produced only modest gains in diabetes care.[8–10] Whether the doctor in the Puskesmas cluster should function as a family physician – a named clinician responsible for a registered population over time – is therefore a practical question for the design of ILP rather than a purely professional one.

Many countries have strengthened primary care by organising it around a family physician, either as a named doctor with a registered list of patients or as a multidisciplinary team led by a doctor with a defined population. China's family doctor contract services register residents with a general practitioner-led team in community health centres; difference-in-differences and cohort studies have associated contracting with more follow-up visits, better continuity and medication adherence among patients with diabetes and hypertension, a shift of first contact towards primary care, and fewer unmet needs.[11–15] Brazil's Family Health Strategy assigns a registered territory to teams composed of a family doctor, a nurse and community health workers, and large cohort and panel studies have linked its expansion to lower mortality from tuberculosis, AIDS and conditions amenable to primary care and to fewer emergency admissions.[16–19] Iran, South Africa, Vietnam and South Asian countries have introduced family physician programmes or specialist family medicine training with variable coverage and results.[20–23] Across settings, the relational continuity that family physicians are expected to provide has been associated with fewer hospital admissions, fewer cardiovascular events, lower mortality and fewer emergency visits for minor problems in large cohorts of older adults and people with chronic disease.[24–27]

Outcome studies of these reforms are persuasive but difficult to transfer to Indonesia, because they measure different end points in health systems with different financing and because a population outcome can improve for many reasons other than the work of the family physician. A more direct question is whether care organised around a family physician actually performs the functions of primary care better, as experienced by the people who use it. The Primary Care Assessment Tool (PCAT) and related attribute-based instruments were designed for this purpose and have been adapted and validated in sub-Saharan Africa, China and Australia, among other settings.[28–30] Their scores are clinically meaningful: in national Japanese cohorts a higher overall score predicted fewer hospitalisations and higher longitudinality, comprehensiveness and community-orientation scores predicted a more stable usual source of care, and among Chinese patients with hypertension perceived primary care quality mediated part of the association between family doctor contracting and preventable hospitalisation.[31–33] Comparative PCAT studies of family physician-based and conventional models have been published in China and Brazil, but most have been small, single-city and inconsistent: some found large differences in favour of the family physician model and others found none.[34–37] Systematic reviews of these programmes have summarised clinical outcomes or implementation evidence,[38,39] and none has produced a pooled estimate of how much a family physician-based organisation changes the experienced quality of primary care or which functions change most. This evidence is directly relevant to ILP, which relies on doctors to lead life-course clusters and in which the competencies and readiness of the primary care workforce for integrated care remain uneven.[4,40,41]

The novelty of this study lies in pooling, for the first time, the comparative evidence on user-experienced primary care quality across the two largest family physician-based reforms in middle-income countries – family doctor contracting in China and the Family Health Strategy in Brazil – together with evidence on family medicine-trained physicians, and in decomposing the association across the individual functions of primary care that Indonesia's Integrated Primary Care reform seeks to strengthen. The aim of this study was to estimate the difference in overall and domain-specific primary care quality, as reported by service users on the Primary Care Assessment Tool or equivalent attribute-based instruments, between family physician-based models and conventional primary care, to rate the certainty of that evidence, and to interpret the findings for the role of, and the need for, family physicians in ILP-based community health centres.

## 2. Methods

### 2.1. Study design

This was a systematic review and meta-analysis of comparative observational studies reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement.[42] The review question, eligibility criteria, outcomes, effect measure and the planned subgroup and sensitivity analyses were specified before data extraction; the protocol was not registered or published in advance. Ethical approval was not required because only published aggregate data were used.

### 2.2. Search strategy

PubMed/MEDLINE and Scopus were searched from database inception without language or date restrictions, and the reference lists of included reports were screened. The search combined two concept blocks: (1) the outcome instrument (Primary Care Assessment Tool, PCATool, PCAT-AE, primary care attributes, Assessment Survey of Primary Care) and (2) family physician-based models of care (family physician, family doctor, family medicine, general

practitioner, Family Health Strategy, family health team, contracted services, care model, traditional model), using free-text terms and the Medical Subject Headings Physicians, Family; Family Practice; General Practitioners; General Practice and National Health Programs. The complete PubMed string was:

("primary care assessment tool"[tiab] OR PCAT[tiab] OR PCATool[tiab] OR "PCA-Tool"[tiab] OR "PCAT-AE"[tiab] OR "PCAT-CE"[tiab] OR KPCAT[tiab] OR "primary care attributes"[tiab] OR "attributes of primary care"[tiab] OR "attributes of primary health care"[tiab] OR "primary health care attributes"[tiab] OR "assessment survey of primary care"[tiab] OR "primary care assessment survey"[tiab] OR "essential attributes"[tiab]) AND ("family physician*"[tiab] OR "family doctor*"[tiab] OR "family medicine"[tiab] OR "family practice*"[tiab] OR "general practitioner*"[tiab] OR "family health strategy"[tiab] OR "family health program*"[tiab] OR "family health team*"[tiab] OR "family health unit*"[tiab] OR "family medicine group*"[tiab] OR "Saude da Familia"[tiab] OR "contract service*"[tiab] OR "contracted service*"[tiab] OR "contracting service*"[tiab] OR "contracted family"[tiab] OR "care model*"[tiab] OR "models of care"[tiab] OR "primary care model*"[tiab] OR "traditional model*"[tiab] OR "traditional unit*"[tiab] OR "traditional primary"[tiab] OR "traditional basic"[tiab] OR "Physicians, Family"[mh] OR "Family Practice"[mh] OR "General Practitioners"[mh] OR "General Practice"[mh] OR "National Health Programs"[mh])

The equivalent Scopus string (TITLE-ABS-KEY) is reported in Supplementary Table S1. Scopus indexes the main Brazilian public-health and nursing journals published in SciELO and the Chinese General Practice journal, but regional databases (LILACS, CNKI, Wanfang) and Web of Science were not searched separately. Portuguese- and Chinese-language reports were translated for eligibility assessment and data extraction.

### 2.3. Eligibility criteria

Studies were eligible if they met all of the following criteria. Population: users of primary care services of any age (adults reporting on their own care, or parents and caregivers reporting on the care of children). Exposure: primary care organised around a family physician, defined by a common element – a continuing, doctor-led relationship with a registered population or list of users – and operationalised as contracted family doctor services, Family Health Strategy teams, general practitioner gatekeeping enrolment (in which users register with a community general practitioner as their point of first contact), or care delivered by family medicine-trained physicians within such teams. Comparator: conventional primary care in the same health system without such a model (users without a contracted family doctor, traditional basic health units, or services without family medicine-trained physicians). Outcome: overall (total, general or essential) primary care quality score reported by users on the PCAT (any validated version) or an equivalent instrument, defined as a validated user questionnaire derived from the same attribute framework that measures at least four of the core functions (first contact, continuity, comprehensiveness, coordination), such as the Assessment Survey of Primary Care; and domain scores for first-contact utilisation, first-contact accessibility, longitudinality (continuity), coordination, comprehensiveness, family centredness and community orientation. Design: comparative cross-sectional, cohort or controlled before-after studies and trials reporting outcome data separately for the two groups.

Studies were excluded if they surveyed only professionals or managers, evaluated a single model without a comparator, compared models that were both family physician-based, reported item-level results only, or did not report an extractable group-level estimate (group means with a standard deviation, standard error or 95% confidence interval, a regression coefficient with its confidence interval and group sizes, or the number of users with a high score in each group). When several reports described the same sample, the report with the most complete data was retained and the others were treated as companion reports. The gatekeeping and family medicine-training variants of the exposure were prespecified and examined in sensitivity analyses.

### 2.4. Study selection and data extraction

Records from both databases were de-duplicated by exact and fuzzy title matching with manual confirmation. One reviewer screened titles and abstracts against the eligibility criteria and assessed the full texts of potentially eligible reports, retrieved from PubMed Central, SciELO, publisher websites and the Chinese General Practice archive; the reason for every full-text exclusion was recorded (Supplementary Table S2). The proportion of records meeting the criteria at each stage is reported in the flow diagram. Data were extracted by one reviewer into a standardised spreadsheet with 23 predefined fields (study identifiers, country, setting and year, design, participants, model definitions, instrument version, group sizes, age and sex, means and dispersion measures for the overall and domain scores, reported effect estimates, analysis and adjustment, and risk-of-bias judgements), and every extracted value was verified against the source table in a complete second pass. The extracted data and the covariates used for adjustment are listed in Supplementary Table S3. Values that were not reported were recorded as not reported and were never imputed.

When a study reported more than one estimate, the following hierarchy was applied: propensity score-matched group means, covariate-adjusted group means, regression-adjusted coefficients, and finally unadjusted group means. Domain scores were mapped to seven common domains (Supplementary Table S4): the utilisation and accessibility subdomains of first contact; longitudinality or ongoing care; coordination, using the integration-of-care subdomain when available and the information-system subdomain otherwise; comprehensiveness, using services provided when available; family centredness; and community orientation. For the Assessment Survey of Primary Care, first visit/first-line care, service accessibility, continuity of the doctor-patient relationship, comprehensive services and coordinated

services were mapped to the corresponding PCAT domains.

### 2.5. Risk-of-bias assessment

Because all included studies were observational, risk of bias was assessed with an adaptation of the Newcastle–Ottawa Scale for cross-sectional studies covering six domains, each rated as low, moderate or high risk using prespecified criteria (Supplementary Table S5). Representativeness (D1) was low risk for probability or census samples of users, moderate for multistage or convenience samples within selected facilities, and high for samples restricted to hospitalised patients or to one or two clinics. Sample size and response (D2) was low risk when the sample was justified and non-response was below about 20% or described, and moderate otherwise. Ascertainment of the exposure (D3) was low risk when the model was assigned from facility or registration records and moderate when it was self-reported. Comparability (D4) was low risk for propensity score matching or multivariable adjustment of the reported estimate, moderate when adjustment was reported only in a separate regression, and high for unadjusted comparisons or comparisons confounded by city. Outcome measurement (D5) was low risk for a full validated instrument and moderate for a reduced module. Statistical analysis and reporting (D6) was low risk when group-level estimates with dispersion were reported for all domains and moderate when reporting was partial. The overall judgement was high if any domain was rated high, low if no domain was high and at most one was moderate, and moderate otherwise.

### 2.6. Effect measures and statistical analysis

Because the included studies used different versions and scorings of the PCAT (0–10 or 0–100 transformations, domain sums and mean item scores) and one used the Assessment Survey of Primary Care, the standardised mean difference with Hedges' small-sample correction (Hedges $g$) was the effect measure; positive values indicate higher quality in the family physician-based model. For each comparison, $g$ was calculated with Equations 1 to 3, in which $M_1$ and $M_2$ are the mean scores, $SD_1$ and $SD_2$ the standard deviations and $n_1$ and $n_2$ the numbers of users in the family physician-based and conventional arms, respectively, and $J$ is the small-sample correction factor.[43]

$$g = J \times \frac{M_1 - M_2}{\mathrm{SD}_{\mathrm{pooled}}} \quad (1)$$

$$\mathrm{SD}_{\mathrm{pooled}} = \sqrt{\frac{(n_1-1)\mathrm{SD}_1^2 + (n_2-1)\mathrm{SD}_2^2}{n_1+n_2-2}} \quad (2)$$

$$J = 1 - \frac{3}{4(n_1+n_2)-9} \quad (3)$$

Missing standard deviations were derived with the standard formulae of the Cochrane Handbook:[43] from the standard error (SE) of an arm mean with Equation 4; from an arm-level 95% confidence interval with upper limit UL and lower limit LL with Equation 5; and, when only the 95% confidence interval of the mean difference ($LL_{MD}$ to $UL_{MD}$) was reported, as the pooled standard deviation implied by that interval with Equation 6.

$$\mathrm{SD} = \mathrm{SE} \times \sqrt{n} \quad (4)$$

$$\mathrm{SD} = \sqrt{n} \times \frac{\mathrm{UL} - \mathrm{LL}}{3.92} \quad (5)$$

$$\mathrm{SD}_{\mathrm{pooled}} = \frac{\mathrm{UL}_{\mathrm{MD}} - \mathrm{LL}_{\mathrm{MD}}}{3.92 \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6)$$

A dichotomised outcome (the proportion of users with a high overall score, ≥6.6 on the 0–10 scale) was converted from the odds ratio (OR) to the standardised mean difference scale with Equation 7, and a regression coefficient ($\beta$) reported with its 95% confidence interval was converted through its $t$ statistic ($t = \beta/\mathrm{SE}_\beta$, with $\mathrm{SE}_\beta$ = (UL − LL)/3.92) with Equation 8; the resulting $d$ was multiplied by $J$ (Equation 3) to give $g$.[43] Independent samples within a study (adults and children surveyed separately) were analysed as separate comparisons.

$$d = \ln(\mathrm{OR}) \times \frac{\sqrt{3}}{\pi} \quad (7)$$

$$d = t \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (8)$$

Two assumptions of these derivations are important. For the Rio de Janeiro survey, the arm-level confidence intervals were design-based, so the derived standard deviations include the design effect and bias the standardised mean difference towards zero; for the Brazilian national survey, the group sizes were approximated from the published weighted proportions (72.45% and 21.94% of 9,677 respondents) and the $t$ statistic was design-based, which is also conservative. The effect of analysing the adult and child samples of the Rio de Janeiro survey as separate comparisons was examined by omitting each in turn.

Effects were pooled with a random-effects model using the DerSimonian–Laird estimator of the between-study variance ($\tau^2$) and Hartung–Knapp confidence intervals, which keep nominal coverage when few studies are pooled and heterogeneity is large; the Hartung–Knapp $t$ statistic and its degrees of freedom are reported for the pooled estimate. Heterogeneity was quantified with Cochran's $Q$, $I^2$ (Equation 9) with its 95% confidence interval, and $\tau^2$. A 95% prediction interval (PI), describing the range of effects expected in a new setting, was calculated with Equation 10, in which $\mu$ is the pooled random-effects estimate, $\mathrm{SE}(\mu)$ its standard error and $t_{k-1}$ the 97.5th percentile of the $t$ distribution with $k$ − 1 degrees of freedom for $k$ comparisons.

$$I^2 = \max\left\{0, \frac{Q - \mathrm{df}}{Q}\right\} \times 100\% \quad (9)$$

$$\mathrm{PI} = \mu \pm t_{k-1} \times \sqrt{\tau^2 + \mathrm{SE}(\mu)^2} \quad (10)$$

Prespecified subgroup analyses compared health system (China, family doctor contracting; Brazil, Family Health Strategy), population (adults; children),

estimate type (adjusted or matched; unadjusted), source of the standard deviation (reported; derived), overall risk of bias (high; low or moderate) and instrument (PCAT; Assessment Survey of Primary Care), using a test of subgroup differences. Sensitivity analyses comprised leave-one-out analysis, restricted maximum-likelihood estimation, Wald-type confidence intervals, a common-effect model and the successive exclusion of studies with derived standard deviations, unadjusted estimates, high risk of bias, a non-contracting exposure (gatekeeping and family medicine training), survey-weighted or *t*-converted estimates, and the dichotomised outcome. Small-study effects were examined with a funnel plot and Egger's regression test, which was applied because at least ten comparisons were available, and with the trim-and-fill method. Because Egger's test can be distorted for standardised mean differences, whose standard error depends on the effect size, and because trim-and-fill performs poorly when heterogeneity is large, both were interpreted as indicators of possible small-study effects rather than as corrections. Each primary care domain was analysed with the same model. Two-sided $p<0.05$ was considered statistically significant. Analyses were performed in R 4.5.3 with the meta (version 8.2–1) and metafor (version 4.8–0) packages.

To aid interpretation, the pooled effect was re-expressed as a difference in points on the 0–10 PCATool scale (Δ) with Equation 11, in which $SD_{median}$ is the median standard deviation observed in the included studies that used that scale and reported, or allowed direct derivation of, simple-random-sample standard deviations.

$$\Delta = g \times SD_{median} \quad (11)$$

### 2.7. Certainty of evidence

The certainty of the evidence for the overall score and for each domain was rated with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for non-randomised evidence.[43] Certainty started at low because all studies were observational and was downgraded for risk of bias, inconsistency (large unexplained heterogeneity or a prediction interval including no effect), indirectness, imprecision (a confidence interval including no effect) or suspected publication bias; upgrading was not applied because no dose-response gradient or large effect was demonstrated.

## 3. Results

### 3.1. Study selection

The database searches identified 395 records (PubMed 149, Scopus 246), and three further records were found through citation searching. After removal of 140 duplicates, 258 records were screened and 220 were excluded on title and abstract. Of the 38 reports sought, 4 could not be retrieved in full, and 18 of the 34 reports assessed were excluded, most often because scores were reported without any measure of dispersion or because the report duplicated a sample already included. The flow of records is illustrated in Figure 1. Fifteen studies (16 reports) were included in the systematic review, of which twelve contributed thirteen independent comparisons to the meta-analysis of the overall score; three further studies contributed only domain-level data.[44–47] Overall, 5.8% of the screened records met the eligibility criteria. Records excluded at the title and abstract stage were mainly evaluations of a single model of care, surveys of professionals or managers, instrument development or validation studies, and studies of hospital, dental or disease-specific services rather than models of primary care; the reports excluded after full-text assessment are listed with reasons in Supplementary Table S2.

### 3.2. Study characteristics

The characteristics of the included studies are detailed in Table 1. Six studies were conducted in China, where the exposure was a contract with a family doctor team or enrolment in general practitioner gatekeeping,[34,35,48–51] and nine in Brazil, where the exposure was the Family Health Strategy or care by family medicine-trained physicians.[36,37,44,45,47,52–55] All were cross-sectional surveys. Sample sizes ranged from 188 to 9,134 users; one study was a national household survey,[53] one a multi-province survey of parents,[50] and one a survey across twelve national primary care pilot zones.[51] Four comparisons concerned children (reported by parents or caregivers), one concerned older adults,[36] one patients with chronic disease,[49] and one children hospitalised for ambulatory care-sensitive conditions.[45] Twelve of the thirteen comparisons used a version of the PCAT and one used the Assessment Survey of Primary Care.[51] Five studies provided adjusted or propensity score-matched estimates.[34,35,48,50,53] Users of family physician-based care were generally older, more often retired or chronically ill and, in some studies, less educated than users of conventional care.[34–36]

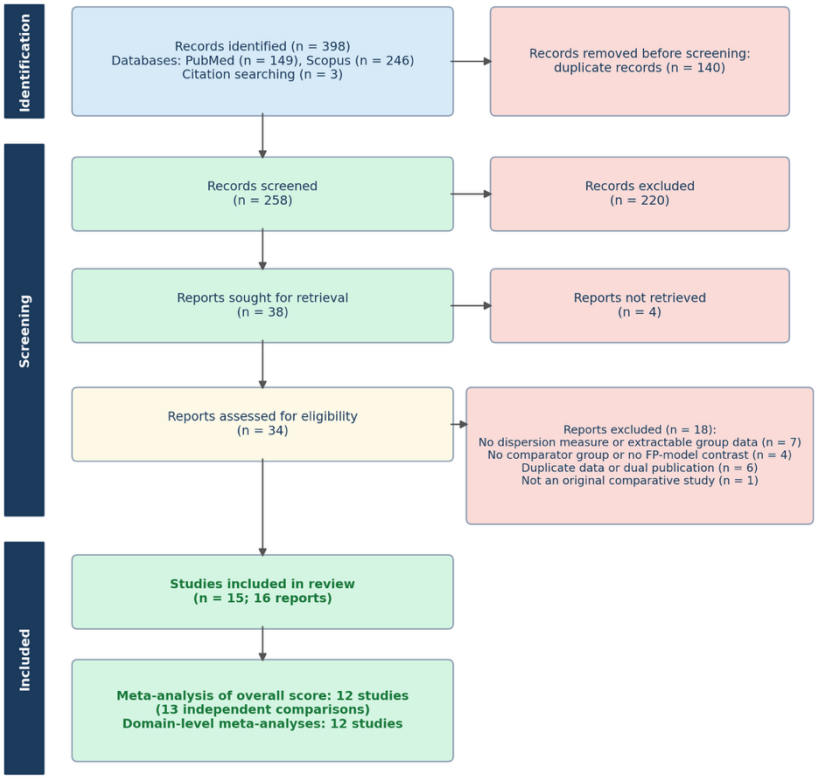



**Figure 1.** PRISMA 2020 flow diagram of the identification, screening and inclusion of studies.

### 3.3. Risk of bias

Risk-of-bias judgements for each domain are illustrated in Figure 2, and the criteria and the justification of every judgement are detailed in Supplementary Table S5. One study was at low overall risk of bias,[35] six at moderate risk and eight at high risk. The most frequent concern was comparability (D4), rated high risk in seven studies that compared unadjusted group means although users of the two models differed in age, education and morbidity, and in one study each model was sampled in a different city.[37] Representativeness was rated high risk in two studies that sampled hospitalised children or only two clinics,[45,54] and exposure ascertainment was rated moderate in all six Chinese studies and in three Brazilian studies, in which contract status, gatekeeping enrolment, household registration or the regular source of care was self-reported. Outcome measurement was at low risk in all but one study, which used a reduced PCAT module.[53]

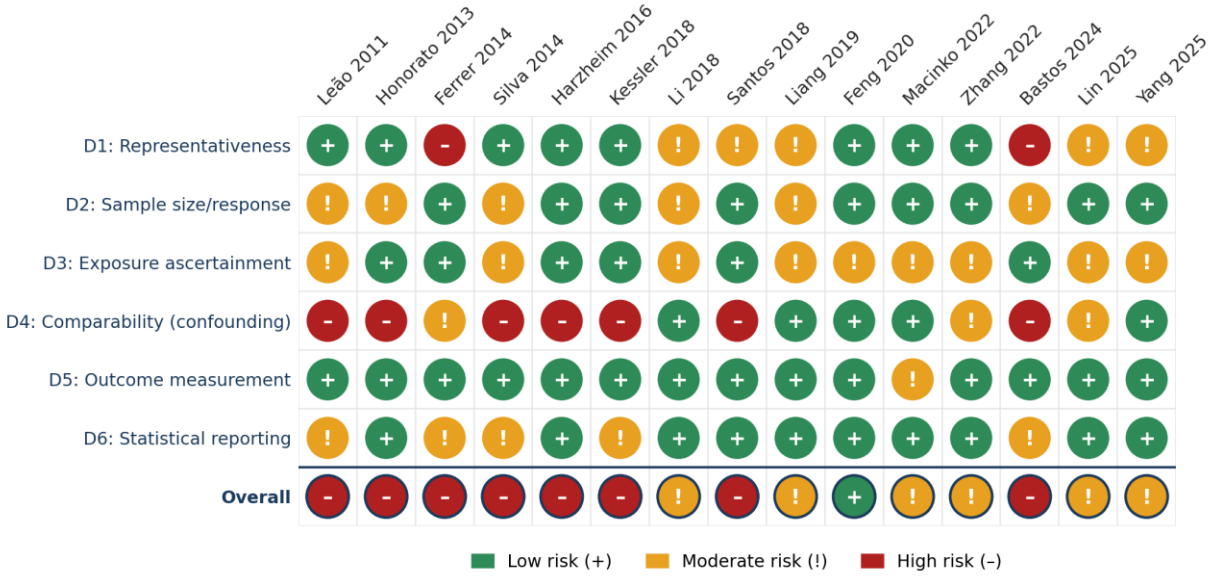



**Figure 2.** Risk-of-bias judgements for the fifteen included studies (adapted Newcastle–Ottawa Scale for cross-sectional studies). D1, representativeness; D2, sample size and response; D3, ascertainment of the exposure; D4, comparability of groups; D5, outcome measurement; D6, statistical analysis and reporting. Overall: high if any domain was high risk; low if no domain was high and at most one was moderate; otherwise moderate.

**Table 1.** Characteristics of the included studies.

| Study | Country, setting; year of data | Population | Family physician-based model (*n*) | Comparator (*n*) | Instrument | Estimate used | Overall RoB |
|---|---|---|---|---|---|---|---|
| Leão 2011[55] | Brazil, Montes Claros, Minas Gerais; 2009 | Caregivers of children <24 months | FHS as regular source of care (*n*=272) | Other services as regular source (*n*=78) | PCATool-Brasil child | Unadjusted (dichotomous high score) | High |
| Honorato 2013[36] | Brazil, Ilhéus, Bahia; 2010–2011 | Adults ≥60 years (mean 72.8) | Family Health Strategy (FHS) units (*n*=193) | Traditional basic health units (*n*=316) | PCATool-Brasil adult | Unadjusted | High |
| Silva 2014[44]‡ | Brazil, Montes Claros, Minas Gerais; NR | Adult users | FHS (*n*=124) | Other primary care services (*n*=249) | PCATool adult (3 domains) | Unadjusted | High |
| Ferrer 2014[45,46]‡ | Brazil, São Paulo (western region); 2011 | Children 0–14 years hospitalised for ACSC (caregivers) | FHS (*n*=197) | Traditional primary care units (*n*=304) | PCATool-Brasil child (longitudinality, first contact) | Unadjusted; PR adjusted for age/diagnosis | High |
| Harzheim 2016[52] | Brazil, Rio de Janeiro, 10 planning areas; 2014 | Adult users and caregivers of child users | Type A units (family clinics staffed only by family health teams) (*n*=2,695 adults; 2,421 children) | Type B units (traditional/mixed health centres) (*n*=790 adults; 695 children) | PCATool-Brasil adult | Design-weighted, unadjusted | High |
| Kessler 2018[47]‡ | Brazil, 32 municipalities, Rio Grande do Sul; 2015 | Adult users | FHS (*n*=450) | Traditional model (*n*=447) | PCATool-Brasil adult (longitudinality) | Unadjusted | High |
| Li 2018[34] | China, Guangzhou, 3 CHCs; 2014 | Adult CHC patients 18–89 years | Contracted GP (family practice contract services) (*n*=94) | No contracted GP (*n*=94) | PCAT-AE (Chinese) | Propensity-score matched | Moderate |
| Santos 2018[37] | Brazil, João Pessoa (FHS) vs Cascavel (traditional); 2012–2013 | Caregivers of children <10 years | Family health units (*n*=344) | Traditional basic health units (*n*=531) | PCATool-Brasil child | Unadjusted (ANOVA) | High |
| Liang 2019[48] | China, Shenzhen and Dongguan, 5 CHCs; NR | Adult CHC patients | GP gatekeeping enrolment (first contact with community GP) (*n*=238) | No gatekeeping enrolment (*n*=238) | PCAT-AE (Chinese) | Propensity-score matched | Moderate |
| Feng 2020[35] | China, 3 pilot cities, Pearl River Delta; 2015 | Adult CHC patients | Family doctor contract (*n*=453) | No contract (*n*=375) | PCAT-AS (Chinese) | Covariate-adjusted means | Low |
| Macinko 2022[53] | Brazil, National Health Survey (PNS); 2019 | Adults with a recent public primary care visit | FHS-registered household (*n*=7,011) | Traditional public health posts (*n*=2,123) | Reduced PCAT (PNS module) | Regression-adjusted (model 3) | Moderate |
| Zhang 2022[49] | China, Chaozhou, Guangdong; 2020 | Patients with chronic disease | Family doctor contract (*n*=189) | No contract (*n*=364) | PCAT-AS (Chinese) | Unadjusted (regression reported) | Moderate |
| Bastos 2024[54] | Brazil, Rio de Janeiro, 2 family clinics; NR | Adult users | Family clinic with Family and Community Medicine residency (FM-trained physicians) (*n*=243) | Family clinic without FCM residency (*n*=245) | PCATool-Brasil adult (short) | Unadjusted | High |
| Lin 2025[51] | China, 12 national comprehensive PHC pilot zones, 36 facilities; 2023 | Adult outpatients | Family doctor contract (*n*=791) | No contract (*n*=366) | ASPC (Chinese) | Unadjusted (regression reported) | Moderate |
| Yang 2025[50] | China, 12 CHCs in Sichuan, Anhui, Fujian, Guangdong; 2022 | Parents of children <16 years | Family physician team contract (*n*=1,971) | No contract (*n*=1,971) | PCAT child (Chinese) | Propensity-score matched | Moderate |

‡Contributed domain-level data only. Harzheim 2016 contributed independent adult and child samples; Macinko 2022 group sizes were approximated from published weighted proportions. ACSC, ambulatory care-sensitive conditions; ANOVA, analysis of variance; ASPC, Assessment Survey of Primary Care; CHC, community health centre; FCM, Family and Community Medicine; FHS, Family Health Strategy; GP, general practitioner; *n*, number of users; NR, not reported; PCAT, Primary Care Assessment Tool (AE, adult expanded; AS, adult short); PHC, primary health care; PNS, National Health Survey; PR, prevalence ratio; RoB, risk of bias.

### 3.4. Primary analysis: overall primary care quality

Across 13 comparisons including 25,101 users, family physician-based models were associated with higher overall primary care quality than conventional care, with a pooled Hedges *g* of 0.37 (95% CI 0.20 to 0.55), as illustrated in Figure 3 ($t$=4.57, df=12, $p$<0.001). Twelve of the thirteen point estimates were above zero and nine had confidence intervals excluding zero; none favoured conventional care. Heterogeneity was

considerable ($Q$=254.01, df=12, $p$<0.001; $I^2$=95.3%, 95% CI 93.4 to 96.6%; $\tau^2$=0.060), and the 95% prediction interval ranged from −0.18 to 0.93, indicating that the effect in a new setting could range from a small negative effect to a large benefit. The largest effects were observed for care by family medicine-trained physicians in Rio de Janeiro ($g$=0.85),[54] for the Family Health Strategy among older adults in Bahia ($g$=0.71)[36] and for family doctor contracts in southern China ($g$=0.65 and 0.65).[35,49] Small, non-significant effects were observed in the Rio de Janeiro comparison of units staffed only by family health teams with mixed units, in the comparison of child care between cities with different models, and for gatekeeping enrolment in Guangdong.[37,48,52]

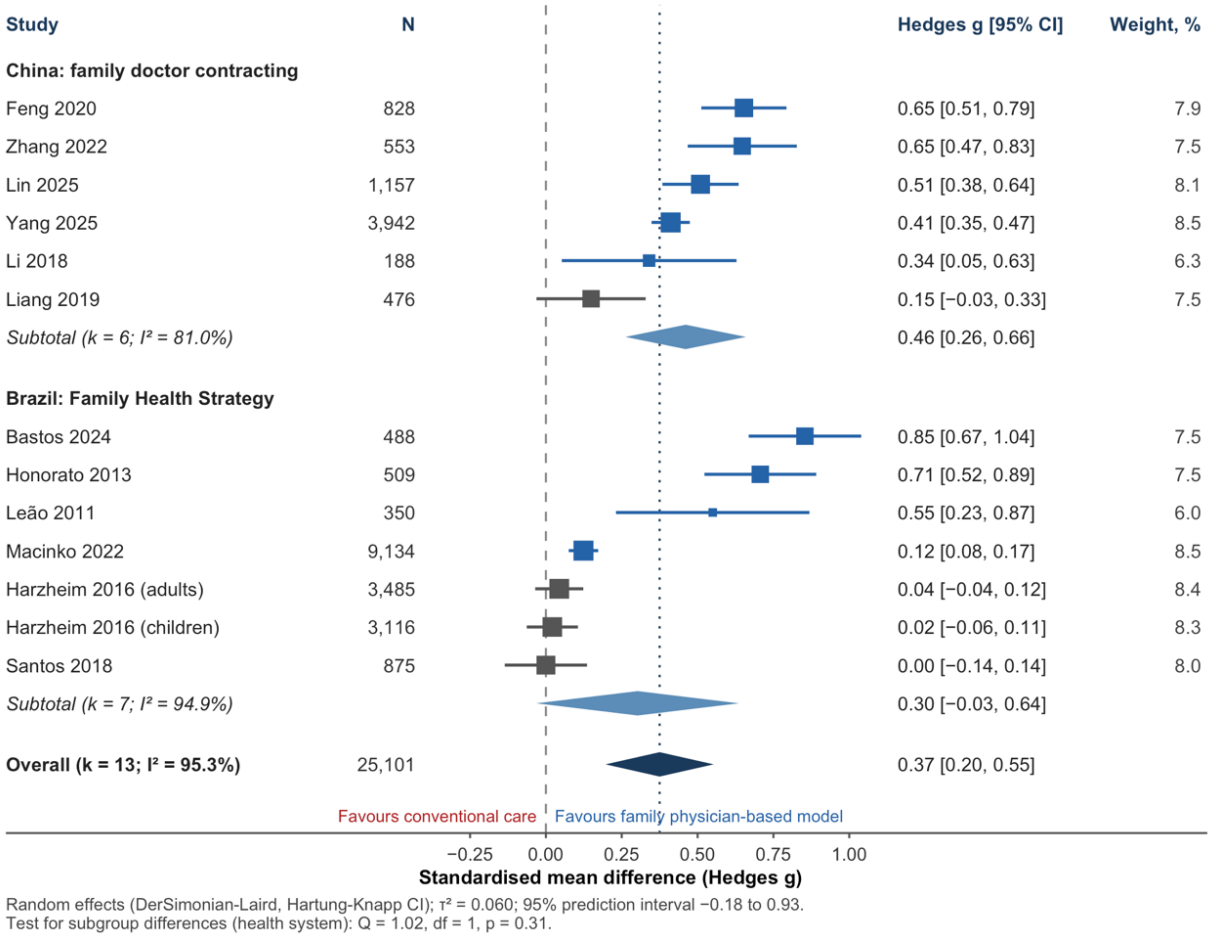



**Figure 3.** Forest plot of the overall primary care quality score comparing family physician-based models with conventional primary care, grouped by health system. Squares are proportional to the random-effects weight; diamonds show subgroup and overall pooled estimates (DerSimonian–Laird with Hartung–Knapp 95% CI); blue estimates have confidence intervals excluding zero. *N*, number of users in the comparison; positive values favour the family physician-based model.

On the 0–10 PCATool scale, whose median standard deviation in the included studies was 1.61, the pooled estimate corresponds (Equation 11) to a difference of about 0.6 points (95% CI 0.3 to 0.9). The mean overall scores of the family physician-based arms that used this scale ranged from 5.0 to 6.4, close to the threshold of 6.6 conventionally used to define a high score, so a difference of this size is large enough to move a typical service across that threshold.

### 3.5. Domain-specific and subgroup analyses

The pooled estimates for the seven primary care domains are illustrated in Figure 4, and the corresponding values are detailed in Table 2. The largest difference was in longitudinality ($g$=0.42, 95% CI 0.21 to 0.63; $k$=13; $I^2$=95.6%), followed by community orientation ($g$=0.32, 95% CI 0.05 to 0.58; $k$=7; $I^2$=93.4%), first-contact utilisation ($g$=0.26, 95% CI 0.04 to 0.47; $k$=11; $I^2$=94.7%), first-contact accessibility ($g$=0.22, 95% CI 0.01 to 0.43; $k$=10; $I^2$=94.5%) and family centredness ($g$=0.18, 95% CI 0.03 to 0.33; $k$=8; $I^2$=88.4%). Differences were not statistically significant for comprehensiveness ($g$=0.17, 95% CI −0.01 to 0.35; $k$=9; $I^2$=92.8%) or coordination ($g$=0.12, 95% CI −0.11 to 0.36; $k$=9; $I^2$=92.2%).

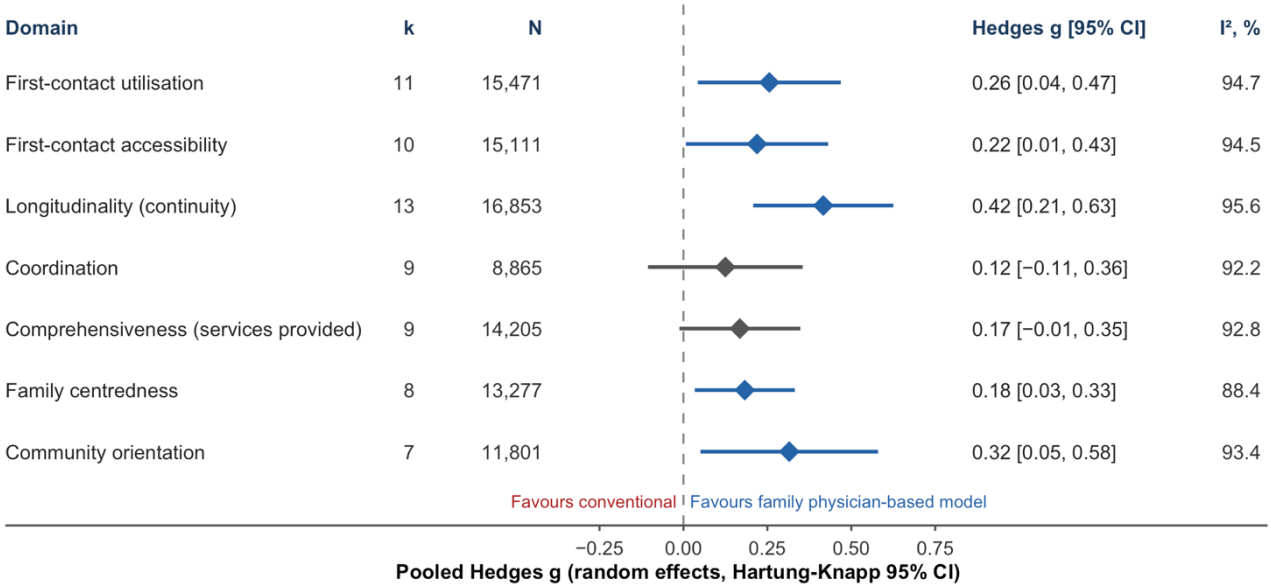



**Figure 4.** Pooled estimates for the seven primary care domains comparing family physician-based models with conventional primary care (random effects, DerSimonian–Laird with Hartung–Knapp 95% CI). Blue estimates have confidence intervals excluding zero. *k*, number of comparisons; *N*, number of users; $I^2$, proportion of variability due to between-study heterogeneity.

In the subgroup analyses of the overall score, which are detailed in Table 2 and displayed graphically in Supplementary Figure S1, the pooled effect was 0.46 (95% CI 0.26 to 0.66) for the six Chinese comparisons ($I^2$=81.0%) and 0.30 (95% CI −0.03 to 0.64) for the seven Brazilian comparisons ($I^2$=94.9%), without evidence of a subgroup difference ($p$=0.313). Effects were 0.44 among adults and 0.23 among children ($p$=0.196), 0.33 for adjusted or matched and 0.41 for unadjusted estimates ($p$=0.647), 0.48 for studies with reported and 0.30 for studies with derived standard deviations ($p$=0.239), 0.40 for studies at low or moderate and 0.35 for studies at high risk of bias ($p$=0.753), and 0.36 for comparisons using the PCAT and 0.51 for the single comparison using the Assessment Survey of Primary Care ($p$=0.179). None of the prespecified factors explained a statistically significant part of the heterogeneity; because heterogeneity remained above 80% in every subgroup with more than one comparison, these tests had little power, and the absence of a subgroup difference should not be read as evidence of equal effects.

### 3.6. Sensitivity analyses

In leave-one-out analysis, the pooled estimate ranged from 0.33 (omitting Bastos 2024) to 0.41 (omitting Santos 2018) and remained statistically significant after the omission of any single comparison; omitting either the adult or the child sample of the Rio de Janeiro survey gave 0.41 and 0.41, respectively, so treating the two samples as independent did not influence the result. Restricted maximum-likelihood estimation gave 0.38 (95% CI 0.22 to 0.54), a Wald-type confidence interval gave 0.37 (95% CI 0.23 to 0.52), and a common-effect model gave 0.24 (95% CI 0.21 to 0.27). Restricting the analysis to studies with reported standard deviations gave 0.48 (95% CI 0.22 to 0.74); $k$=5; to adjusted or matched estimates, 0.33 (95% CI 0.06 to 0.61); $k$=5; and to studies at low or moderate risk of bias, 0.40 (95% CI 0.20 to 0.61); $k$=7. Excluding the survey-weighted and $t$-converted estimates gave 0.48 (95% CI 0.29 to 0.67); $k$=10; excluding the gatekeeping and family medicine-training contrasts gave 0.35 (95% CI 0.17 to 0.54); $k$=11; excluding only the family medicine-training contrast gave 0.33 (95% CI 0.16 to 0.51); $k$=12; and excluding the dichotomised outcome gave 0.36 (95% CI 0.17 to 0.56); $k$=12. All sensitivity estimates are detailed in Table 2, and the complete leave-one-out results are listed in Supplementary Table S6.

**Table 2.** Pooled estimates for the overall score, the primary care domains, subgroups and sensitivity analyses.

| Analysis | *k* | *N* | Hedges *g* (95% CI) | $I^2$ (%) | *p* |
|---|---|---|---|---|---|
| Overall primary care quality score | 13 | 25,101 | 0.37 (0.20 to 0.55) | 95.3 | <0.001 |
| **Primary care domains** | | | | | |
| First-contact utilisation | 11 | 15,471 | 0.26 (0.04 to 0.47) | 94.7 | 0.023 |
| First-contact accessibility | 10 | 15,111 | 0.22 (0.01 to 0.43) | 94.5 | 0.044 |
| Longitudinality (continuity) | 13 | 16,853 | 0.42 (0.21 to 0.63) | 95.6 | <0.001 |

| Analysis | *k* | *N* | Hedges *g* (95% CI) | $I^2$ (%) | *p* |
|---|---|---|---|---|---|
| Coordination | 9 | 8,865 | 0.12 (−0.11 to 0.36) | 92.2 | 0.247 |
| Comprehensiveness (services provided) | 9 | 14,205 | 0.17 (−0.01 to 0.35) | 92.8 | 0.064 |
| Family centredness | 8 | 13,277 | 0.18 (0.03 to 0.33) | 88.4 | 0.023 |
| Community orientation | 7 | 11,801 | 0.32 (0.05 to 0.58) | 93.4 | 0.027 |
| **Subgroups (overall score)** | | | | | |
| China: family doctor contracting | 6 | | 0.46 (0.26 to 0.66) | 81.0 | 0.313[a] |
| Brazil: Family Health Strategy | 7 | | 0.30 (−0.03 to 0.64) | 94.9 | |
| Adults | 9 | | 0.44 (0.22 to 0.67) | 95.7 | 0.196[a] |
| Children | 4 | | 0.23 (−0.20 to 0.66) | 95.7 | |
| Adjusted or matched estimates | 5 | | 0.33 (0.06 to 0.61) | 95.2 | 0.647[a] |
| Unadjusted estimates | 8 | | 0.41 (0.12 to 0.70) | 95.8 | |
| Reported SDs | 5 | | 0.48 (0.22 to 0.74) | 84.1 | 0.239[a] |
| Derived SDs | 8 | | 0.30 (0.03 to 0.58) | 94.8 | |
| Low or moderate risk of bias | 7 | | 0.40 (0.20 to 0.61) | 94.7 | 0.753[a] |
| High risk of bias | 6 | | 0.35 (−0.06 to 0.76) | 95.7 | |
| PCAT | 12 | | 0.36 (0.17 to 0.56) | 95.3 | 0.179[a] |
| ASPC | 1 | | 0.51 (0.38 to 0.64) | – | |
| **Sensitivity analyses (overall score)** | | | | | |
| Leave-one-out analysis[b] | 12 | | 0.33 to 0.41 | | |
| REML estimator (Wald-type CI) | 13 | 25,101 | 0.38 (0.22 to 0.54) | 95.3 | <0.001 |
| DerSimonian–Laird (Wald-type CI) | 13 | 25,101 | 0.37 (0.23 to 0.52) | 95.3 | <0.001 |
| Common-effect model | 13 | 25,101 | 0.24 (0.21 to 0.27) | – | <0.001 |
| Reported SDs only | 5 | 6,637 | 0.48 (0.22 to 0.74) | 84.1 | 0.007 |
| Adjusted or matched estimates only | 5 | 14,568 | 0.33 (0.06 to 0.61) | 95.2 | 0.027 |
| Low or moderate risk of bias only | 7 | 16,278 | 0.40 (0.20 to 0.61) | 94.7 | 0.003 |
| Excluding survey-weighted/*t*-converted estimates | 10 | 9,366 | 0.48 (0.29 to 0.67) | 90.5 | <0.001 |
| Excluding gatekeeping and FM-training contrasts | 11 | 24,137 | 0.35 (0.17 to 0.54) | 95.2 | 0.002 |
| Excluding the FM-training contrast | 12 | 24,613 | 0.33 (0.16 to 0.51) | 94.8 | 0.001 |
| Excluding the dichotomised outcome | 12 | 24,751 | 0.36 (0.17 to 0.56) | 95.6 | 0.002 |
| Trim-and-fill adjusted estimate | 18[c] | | 0.19 (−0.01 to 0.39) | | |

Hedges *g* >0 favours the family physician-based model. Random-effects (DerSimonian–Laird) estimates with Hartung–Knapp 95% CI unless stated otherwise. [a]*p* for subgroup difference. [b]Range of the thirteen estimates obtained by omitting one comparison at a time (each $k$=12). [c]*k* includes the five comparisons imputed by the trim-and-fill method. ASPC, Assessment Survey of Primary Care; CI, confidence interval; FM, family medicine; *k*, number of comparisons; *N*, number of users; PCAT, Primary Care Assessment Tool; REML, restricted maximum likelihood; SD, standard deviation.

### 3.7. Publication bias

The funnel plot of the overall score is illustrated in Figure 5. Egger's regression test did not indicate statistically significant asymmetry ($t$=1.89, df=11, $p$=0.085), but the trim-and-fill method imputed five comparisons on the left of the funnel and reduced the pooled estimate to 0.19 (95% CI −0.01 to 0.39), a pattern consistent with small-study effects or with genuine heterogeneity between the smaller single-city studies and the large population surveys. The trim-and-fill estimate is also reported in Table 2. Because both methods are unreliable when heterogeneity is as large as in this review, the trim-and-fill estimate was treated as a lower plausible bound rather than as a corrected effect.

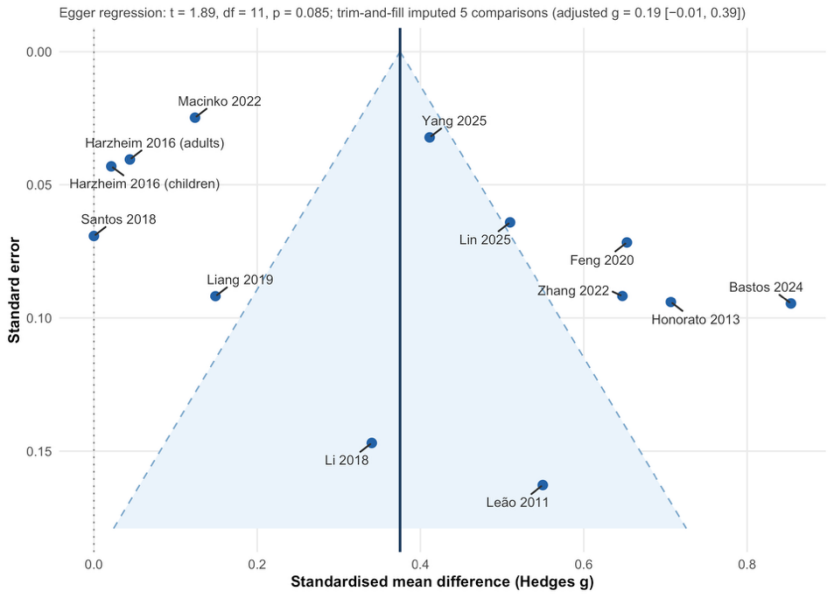



**Figure 5.** Funnel plot of the overall primary care quality score. Each point is one comparison; the solid vertical line marks the pooled random-effects estimate, the dashed lines the pseudo-95% confidence limits and the dotted line no effect. The results of Egger's regression test and of the trim-and-fill method are shown above the plot.

### 3.8. Certainty of evidence

The certainty of the evidence was rated very low for the overall score and for every domain, as detailed in Table 3. All outcomes started at low certainty because the evidence was observational and were downgraded one level for inconsistency, because heterogeneity was large and the prediction intervals included no effect. Risk of bias was not downgraded separately because estimates from studies at high and at low or moderate risk were similar, and indirectness was not downgraded for the review question, although the absence of Indonesian studies limits applicability to ILP. For coordination and comprehensiveness, the confidence intervals also included no effect, and for the overall score the trim-and-fill result raised concern about small-study effects; these outcomes were already at the lowest level of certainty.

**Table 3.** Certainty of the evidence (GRADE) for the overall score and the primary care domains.

| Outcome | *k* | *N* | Hedges *g* (95% CI) | 95% PI | Reasons for downgrading | Certainty |
|---|---|---|---|---|---|---|
| Overall score | 13 | 25,101 | 0.37 (0.20 to 0.55) | −0.18 to 0.93 | Inconsistency ($I^2$=95.3%; PI includes 0); suspected publication bias (trim-and-fill $g$=0.19) | Very low |
| First-contact utilisation | 11 | 15,471 | 0.26 (0.04 to 0.47) | −0.35 to 0.86 | Inconsistency ($I^2$=94.7%; PI includes 0) | Very low |
| First-contact accessibility | 10 | 15,111 | 0.22 (0.01 to 0.43) | −0.37 to 0.80 | Inconsistency ($I^2$=94.5%; PI includes 0) | Very low |
| Longitudinality (continuity) | 13 | 16,853 | 0.42 (0.21 to 0.63) | −0.25 to 1.08 | Inconsistency ($I^2$=95.6%; PI includes 0) | Very low |
| Coordination | 9 | 8,865 | 0.12 (−0.11 to 0.36) | −0.49 to 0.74 | Inconsistency ($I^2$=92.2%; PI includes 0); imprecision (95% CI includes 0) | Very low |
| Comprehensiveness (services provided) | 9 | 14,205 | 0.17 (−0.01 to 0.35) | −0.34 to 0.67 | Inconsistency ($I^2$=92.8%; PI includes 0); imprecision (95% CI includes 0) | Very low |
| Family centredness | 8 | 13,277 | 0.18 (0.03 to 0.33) | −0.21 to 0.58 | Inconsistency ($I^2$=88.4%; PI includes 0) | Very low |
| Community orientation | 7 | 11,801 | 0.32 (0.05 to 0.58) | −0.24 to 0.88 | Inconsistency ($I^2$=93.4%; PI includes 0) | Very low |

All outcomes started at low certainty (observational studies) and were downgraded by one level for inconsistency; further concerns did not lower certainty below very low. Risk of bias and indirectness were judged not serious: estimates from studies at high and at low or moderate risk of bias were similar, and the included populations and exposures matched the review question, although no Indonesian study was available. CI, confidence interval; *k*, number of comparisons; *N*, number of users; PI, 95% prediction interval.

## 4. Discussion

This meta-analysis of thirteen comparisons from twelve studies and more than 25,000 primary care users found that care organised around a family physician was associated with higher primary care quality as experienced by users ($g$=0.37; Figure 3), equivalent to about 0.6 points on the 0–10 PCATool scale (Equation 11). The improvement was concentrated in the functions that define family medicine – a continuing relationship with a known doctor or team ($g$=0.42) and an orientation towards the community served ($g$=0.32) – with smaller gains in first-contact use and access and in family centredness, and no clear improvement in coordination or comprehensiveness, as illustrated in Figure 4. The

direction of effect was consistent across China and Brazil, across adults and children and across adjusted and unadjusted estimates, and it was preserved in every sensitivity analysis (Table 2). Its size, however, was uncertain: heterogeneity was considerable, the prediction interval included no effect, the trim-and-fill correction attenuated the estimate, and the certainty of the evidence was very low (Table 3). The findings should therefore be read as consistent evidence of an association rather than as a precise estimate of the effect of introducing a family physician.

These findings complement the outcome evidence for the same reforms. Family doctor contracting in China has been associated with better hypertension control in a meta-analysis of trials of contracted-service interventions – an outcome of clinical control rather than experienced quality – with more follow-up visits, more continuous and adherent diabetes care and fewer unmet needs in quasi-experimental studies, and with better self-rated health among older adults and in the emotional role functioning of people with diabetes.[11–13,15,38,56,57] Brazil's Family Health Strategy has been associated with lower adult, tuberculosis-related and AIDS-related mortality, fewer ambulatory care-sensitive and emergency admissions, less emergency-service use among adults with multimorbidity, better infant outcomes and a positive long-run net benefit.[16–19,58–61] The present results suggest one pathway linking these reforms to outcomes: in both countries users of family physician-based care reported more continuous, more community-oriented and somewhat more accessible care. This is consistent with cohort evidence that better experienced first contact, continuity and coordination predict fewer hospitalisations and a more stable usual source of care, and that primary care quality mediates part of the association between family doctor contracting and preventable hospitalisation.[31–33] By contrast, the evidence for Iran's family physician programme has been judged insufficient and contradictory, which underlines that the label of a family physician programme does not guarantee that its core functions are delivered.[20,39] Iranian evaluations illustrate why implementation fidelity matters: the programme reduced specialist visits in the short term but was later followed by increased use of general practitioner consultations, diagnostics and hospitalisation, and a health impact assessment found its effectiveness constrained by socioeconomic inequities, cultural barriers and managerial inefficiencies.[20,62]

Pooling different reforms requires a clear statement of what they share. The common element of the included exposures was a continuing, doctor-led relationship with a registered population or list of users; the reforms differed in team composition, financing and service packages. The Brazilian model embeds community health workers in the team, the Chinese contract adds a service package and priority access, gatekeeping enrolment is an insurance rule, and the Rio de Janeiro training contrast isolates the competencies of family medicine-trained physicians within otherwise similar family clinics.[54] The association observed therefore cannot be attributed to the individual physician alone. That the largest single effect was found for family medicine-trained physicians, and that excluding the gatekeeping and training contrasts left the pooled estimate almost unchanged, suggests that both the organisational model and the competencies of the doctor contribute. Comparator contamination worked in the opposite direction: when the comparator itself contained family health teams, as in the Rio de Janeiro survey, or when models were compared across cities, the contrast tested was weaker than the review question and the resulting estimates close to zero are expected to bias the pooled result towards the null.[37,52]

The pattern across domains is plausible. Registration of a household with a named doctor-led team creates the conditions for relational continuity, and the large continuity effects mirror the finding that family doctor registration increased the continuity index of Chinese patients with diabetes[12] and that continuity with a known general practitioner or physician team was associated with fewer hospitalisations, fewer cardiovascular events, lower mortality and fewer emergency visits for minor problems in large cohorts,[24–27] although in one Finnish cohort of people with diabetes the added benefit of very high continuity was small.[63] Community orientation benefits from a defined territory and, in Brazil, from community health workers embedded in the team; the corresponding Indonesian structures are the village-level integrated health posts and cadres that ILP links to Puskesmas clusters.[1,3] Coordination and comprehensiveness, however, depend on features that a family physician cannot supply alone: a functioning referral and counter-referral system, shared information systems, access to diagnostics and medicines, and a service package broad enough to be used. In Brazil, traditional units that employed paediatricians and gynaecologists offered a wider range of services than family health units,[37] and in China contracted and non-contracted residents used the same community health centres; in both settings the family physician model may therefore have had little room to change these domains unless the surrounding system changed as well. This explanation remains a hypothesis. Measurement may also have contributed: the coordination domain combined an integration-of-care subdomain, which is answered only by users who were referred, with an information-system subdomain, and both had wide confidence intervals in the individual studies.

Heterogeneity was considerable and deserves interpretation rather than dismissal. Part of it was methodological: the Rio de Janeiro survey compared units staffed only by family health teams with mixed units that also contained such teams, and its design-

based confidence intervals inflated the derived standard deviations; the national Brazilian survey contributed a regression-based estimate converted through a design-based *t* statistic; and one Brazilian study sampled each model in a different city.[37,52,53] Excluding the two survey-weighted studies reduced $I^2$ and increased the pooled effect (Table 2), suggesting that these conservative conversions diluted rather than inflated the estimate. Part of the heterogeneity is likely to be real: the content of a family doctor contract varies between cities, many contracted residents in China 'sign but do not consult', acceptance of family doctors depends on trust and satisfaction, and among older adults in Beijing enrolment alone did not improve health status.[64–67] The prediction interval, which included zero, indicates that introducing a family physician-based model will not automatically improve experienced quality; the model has to be implemented with enough intensity for users to experience a continuing relationship.

Equity deserves particular attention. In the included studies, users of family physician-based care were older, more often chronically ill and, in some settings, less educated than users of conventional care,[34–36] and outcome studies of the Family Health Strategy found larger reductions in tuberculosis, ambulatory care-sensitive admissions and infant mortality among the poorest and least educated groups and an attenuation of wealth-related inequalities in mortality among older adults.[16,19,59,68] Family physician-based models may therefore concentrate their benefits among the groups that ILP most needs to reach. The present data cannot establish whether they narrow inequalities in experienced quality, because none of the included studies reported domain scores by socioeconomic group; this is an important question for evaluations of ILP.

Transferability to Indonesia depends on how closely ILP resembles the reforms studied. ILP shares with the Brazilian model a defined territory, a link between the clinic and community health workers (cadres at integrated health posts) and a public provider, and with the Chinese model a community health centre base and a focus on chronic disease.[1] It differs in organising Puskesmas services by life-course cluster rather than by a named doctor's list, in combining clinical care with extensive public health programmes, and in financing clinical care through national health insurance capitation with a modest performance component.[10,69] These differences suggest that the continuity benefit observed here will be realised only if ILP clusters make an individual doctor or small team accountable for identifiable households over time, and that the community-orientation benefit depends on how well cluster teams use cadres and village posts, whose readiness and data systems are still incomplete.[2,3] Evidence that follow-up after community screening is poor and that chronic disease referral pathways are weak indicates where a family physician-led cluster could add most value.[8,9] Costs also matter: in Brazil the net monetary benefit of the Family Health Strategy was positive only where coverage was intensive, which suggests that partial implementation of a family physician model may add costs without the corresponding gains.[60]

The review has several strengths. It is, to our knowledge, the first quantitative synthesis of comparative PCAT evidence across family physician-based reforms, it included Portuguese- and Chinese-language reports that are rarely captured by English-language reviews, it decomposed the overall effect into the individual functions of primary care, and it used transparent, prespecified conversions and a broad set of sensitivity analyses, all of which preserved the direction of the main result, and it rated the certainty of the evidence explicitly so that the strength of the conclusions matches the data.

The limitations must also be considered. First, all included studies were cross-sectional, and users who choose a family doctor contract or live in a Family Health Strategy area differ from other users; although adjusted and matched estimates were similar to unadjusted ones, residual confounding and reverse causation – users who value continuity seeking a family doctor – cannot be excluded. Second, eight of the thirteen comparisons required derived standard deviations or effect-size conversions; these followed standard formulae, and the two design-based conversions were conservative, but they add uncertainty. Third, one reviewer performed screening and extraction, with complete verification of every extracted value in a second pass; independent duplicate screening could have identified additional eligible reports. Fourth, only two bibliographic databases were searched and four reports could not be retrieved; Brazilian studies published in journals indexed only in LILACS and Chinese-language studies indexed only in national databases may have been missed, and because such studies are often small, their omission could bias the estimate in either direction.

Fifth, the evidence came from two countries, so its transferability to Indonesia rests on the similarity of the models rather than on direct Indonesian data. Sixth, the different PCAT versions and the Assessment Survey of Primary Care may not measure the domains identically, and the standardised mean difference assumes that the underlying constructs are equivalent across instruments. Seventh, the asymmetry of the funnel plot (Figure 5) and the trim-and-fill correction suggest that small-study effects may have inflated the pooled estimate, even though Egger's test was not significant. Finally, the outcome was experienced quality rather than clinical outcomes; although experienced quality predicts hospitalisation and continuity of care in cohort studies,[31,32] user-reported scores may also reflect expectations, trust and the

design of the questionnaire, as psychometric evaluations of short PCAT versions have shown.[30]

For Indonesia, the findings support the intention of ILP to assign each household to an identifiable Puskesmas team and suggest that the added value of a family physician lies primarily in continuity and in linking clinical care to the community. They also indicate that ILP is unlikely to improve coordination and comprehensiveness through reorganisation alone: referral pathways, digital information systems and service packages need to be strengthened in parallel, a conclusion consistent with the barriers reported in ILP pilot districts and chronic disease programmes.[2,8,9] The need for family physicians follows from this reasoning rather than from a workforce calculation: if continuity and community orientation are the functions most strongly associated with physician-led models, ILP clusters need doctors who are trained and deployed to provide them, and experience in South Africa and Vietnam shows that such coverage and retention do not arise without deliberate training, deployment and incentive policies.[21,22,70] The size of that workforce requirement could not be estimated from the included studies.

Future research should address the gaps identified here. A PCAT version validated for Puskesmas users is needed, building on recent adaptations in sub-Saharan Africa and the rapid assessment version developed in China,[28,29] so that ILP clusters can be monitored against the functions that matter most. Controlled before-after or stepped-wedge evaluations of ILP should compare clusters with and without named physician-led teams, measure experienced quality by socioeconomic group, and link it to outcomes such as potentially preventable hospitalisation and chronic disease control.[7] Studies should also report group-level scores with standard deviations for all domains, which would have allowed more studies to contribute to this synthesis.

## 5. Conclusion

In this systematic review and meta-analysis, primary care organised around a family physician – whether through contracted family doctor teams in China, Family Health Strategy teams in Brazil or family medicine-trained physicians – was associated with higher primary care quality as experienced by users, by about 0.6 points on the 0–10 PCATool scale. The association was strongest for continuity of care and community orientation, smaller for first-contact use, accessibility and family centredness, and not statistically significant for coordination and comprehensiveness. The direction was consistent across health systems, age groups and analytic choices, but the magnitude varied considerably between settings, may be overestimated by small-study effects, and the certainty of the evidence was very low because all studies were cross-sectional and heterogeneity was large.

For Indonesia's Integrated Primary Care reform, the evidence indicates a role for family physicians as the doctors who carry a continuing relationship with a registered population and connect Puskesmas care with the community. It also indicates that family physicians need to be supported by functioning referral and information systems and by an adequate scope of services, without which coordination and comprehensiveness are unlikely to improve. On the basis of this indirect evidence, linking households to named physician-led teams within ILP clusters, investing in family medicine training and deployment, and routinely measuring user-experienced primary care quality are reasonable priorities, and they should be evaluated prospectively as ILP is implemented.

## Declarations

### *Declaration of generative AI and AI-assisted technologies in the writing process*

The authors confirm that no generative artificial intelligence (AI) tools were used at any stage of preparing this manuscript.

### *Author contributions (CRediT)*

L.V.S.: conceptualisation, methodology, data curation, formal analysis, writing – original draft. P.A.: conceptualisation, supervision, validation, writing – review and editing. Both authors approved the final manuscript.

### *Conflict of interest*

The authors declare no conflict of interest.

### *Funding*

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

### *Ethical approval and informed consent*

Not applicable. This study analysed published aggregate data and did not involve human participants directly.

### *Consent for publication*

Not applicable.

### *Data availability*

All data extracted for this review, the search strategies, the risk-of-bias judgements and the leave-one-out results are available from the corresponding author on reasonable request.

### *Acknowledgements*

The authors have no acknowledgements to declare.

## References

1. Kementerian Kesehatan Republik Indonesia. Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/2015/2023 tentang petunjuk teknis integrasi pelayanan kesehatan primer [Technical guidelines for integrated primary health care; in Indonesian]. Jakarta: Kementerian Kesehatan RI; 2023. Available from: https://renbut.kemkes.go.id/regulasi/Buku%20KMK%20

No%202015%20tahun%202023%20tentang%20Petunjuk%20Teknis%20ILP.pdf

2. Widiastuti M, Frans SO, Hafala SN, et al. Barriers and enablers in implementing integrated primary health care: A qualitative assessment in West Sumbawa District, Indonesia. PLOS Glob Public Health. 2026;6(9):e0007238. doi:10.1371/journal.pgph.0007238
3. Setiaasih R, Sunjaya DK, Sofiatin Y, et al. Readiness of health posts for primary health care integration in Indonesia: a mixed-methods study. BMC Public Health. 2025;25(1):1429. doi:10.1186/s12889-025-22520-x
4. Wiarsih W, Triyanto E, Andriany M, et al. Nursing competency readiness for integrated primary health care implementation: a multi-province cross-sectional study in Indonesia. BMC Nurs. 2026. doi:10.1186/s12912-026-05010-2
5. Kusnali A, Nugraheni WP, Mubasyiroh R, et al. Urban-rural disparities in primary health care utilization in Indonesia: a cross-sectional study. Korean J Fam Med. 2026;47(5):426–33. doi:10.4082/kjfm.25.0092
6. Cheng Q, Fattah RA, Susilo D, et al. Determinants of healthcare utilization under the Indonesian national health insurance system – a cross-sectional study. BMC Health Serv Res. 2025;25(1):48. doi:10.1186/s12913-024-11951-8
7. Solikha DA, Butler D, Korda RJ, et al. Geographic variation in potentially preventable hospitalisations in Indonesia: a multilevel analysis using National Health Insurance Sample Data. BMJ Open. 2026;16(3):e112206. doi:10.1136/bmjopen-2025-112206
8. Hustrini NM, Susalit E, van Diepen M, et al. Perceived barriers to management of patients with diabetes mellitus and hypertension in primary care centers in Indonesia. Front Health Serv. 2026;5:1715125. doi:10.3389/frhs.2025.1715125
9. Marthoenis M, Yenni R, Sudharsanan N. Follow-up after Posbindu NCD screening in Indonesia: sociodemographic predictors of primary care engagement. Public Health Nurs. 2026;43(4):1086–9. doi:10.1111/phn.70117
10. Pratama FB, Dartanto T. The impact of performance-based capitation on diabetes care: Evidence from Indonesia's national health insurance program. Int J Health Econ Manag. 2026;26(1):2. doi:10.1007/s10754-025-09407-5
11. Zhang L, Zhang P, Chen W. Can family doctor system improve health service utilization for patients with hypertension and diabetes in China? A difference-in-differences study. BMC Health Serv Res. 2024;24(1):454. doi:10.1186/s12913-024-10903-6
12. Liu X, Zhang L, Chen W. Impact of the family doctor system on the continuity of care for diabetics in urban China: a difference-in-difference analysis. BMJ Open. 2023;13(2):e065612. doi:10.1136/bmjopen-2022-065612
13. Tong X, Zou H, Zhang L, et al. Impact of the family doctor system on medication adherence among type 2 diabetes patients in China: a difference-in-differences analysis. BMC Public Health. 2025;25(1):404. doi:10.1186/s12889-025-21656-0
14. Fu P, Wang Y, Zhao D, et al. Does contracting family doctor promote primary healthcare utilization among older adults? – evidence from a difference-in-differences analysis. BMC Geriatr. 2024;24(1):749. doi:10.1186/s12877-024-05336-z
15. Tan J, Wang J, Xu L, et al. Impact of the family doctor contracting system on unmet healthcare needs in Shandong Province, China. Health Policy Plan. 2026;41(1):22–35. doi:10.1093/heapol/czaf069
16. Jesus GS, Pescarini JM, Silva AF, et al. The effect of primary health care on tuberculosis in a nationwide cohort of 7·3 million Brazilian people: a quasi-experimental study. Lancet Glob Health. 2022;10(3):e390–7. doi:10.1016/S2214-109X(21)00550-7
17. Pinto PFPS, Macinko J, Silva AF, et al. The impact of primary health care on AIDS incidence and mortality: A cohort study of 3.4 million Brazilians. PLoS Med. 2024;21(7):e1004302. doi:10.1371/journal.pmed.1004302
18. Diaz MDM, Teixeira AD, Postali FAS, et al. Assessment of the association between the Brazilian family health strategy and adult mortality. Health Policy Plan. 2022;37(4):461–71. doi:10.1093/heapol/czac011
19. Hone T, Macinko J, Trajman A, et al. Expansion of primary healthcare and emergency hospital admissions among the urban poor in Rio de Janeiro Brazil: A cohort analysis. Lancet Reg Health Am. 2022;15:100363. doi:10.1016/j.lana.2022.100363
20. Safarpoor Z, Lotfi F, Bayati M, et al. The effect of family physician program and health transformation plan on utilization and cost of health services. Cost Eff Resour Alloc. 2025;23(1):30. doi:10.1186/s12962-025-00637-5
21. Rubler TA, Von Pressentin KB, Mash RJ. Family physician deployment in South Africa's district health system: A cross-sectional study. S Afr Fam Pract (2004). 2026;68(1):a6289. doi:10.4102/safp.v68i1.6289
22. Pham CAT, Le NQT, Kim D. What drives family physicians to rural practice? A discrete choice experiment in Vietnam. Health Policy Open. 2026;11:100175. doi:10.1016/j.hpopen.2026.100175
23. Ali FA, Kumar R, Ahmed K, et al. Scope, training, and systemic challenges of family medicine in South Asia: Perspectives of general practitioners and family physicians. Pak J Med Sci. 2026;42(ICON26):S110–6. doi:10.12669/pjms.42.(ICON26).15697
24. Caughey GE, Schwabe J, Pulling BW, et al. Primary health care services and continuity of care are associated with better health outcomes in the older population. J Am Geriatr Soc. 2026;74(7):1950–63. doi:10.1111/jgs.70465
25. Xu W, Yu EYT, Chin WY, et al. Team-based continuity of care for patients with hypertension: a retrospective primary care cohort study in Hong Kong. Br J Gen Pract. 2023;73(736):e807–15. doi:10.3399/BJGP.2023.0150
26. Yun EK, Shim JY, Shin SJ, et al. Continuity of care for patients with hypertension or diabetes mellitus reduced medical utilization, total medical costs, and cardiovascular events based on the Korean NHIS-HEALS cohort. Nutr Metab Cardiovasc Dis. 2026;36(6):104527. doi:10.1016/j.numecd.2025.104527
27. Tiagi R. The association between continuity of care and emergency department utilization among Canadian senior adults. J Am Board Fam Med. 2026;39(1):165727. doi:10.3122/jabfm.2025.250211R1
28. Mash R, Bello K, Besigye IK, et al. Adapting the Primary Care Assessment Tool for sub-Saharan Africa: a validation study. BJGP Open. 2025;9(1):BJGPO.2024.0084. doi:10.3399/BJGPO.2024.0084
29. Zhong C, Huang J, Li L, et al. Development and validation of a rapid assessment version of the Assessment Survey of Primary Care in China. Front Public Health. 2022;10:852730. doi:10.3389/fpubh.2022.852730
30. Bui CM, Batterham MJ, Mullan J, et al. Psychometric properties of the Adult Primary Care Assessment Tool Short form (PCAT-S) among high-risk patients in Australian general practice. PLoS One. 2026;21(2):e0341250. doi:10.1371/journal.pone.0341250
31. Aoki T, Sugiyama Y, Mutai R, et al. Impact of primary care attributes on hospitalization during the COVID-19 pandemic: a nationwide prospective cohort study in Japan. Ann Fam Med. 2023;21(1):27–32. doi:10.1370/afm.2894
32. Aoki T, Zukeran S, Matsushima M. The role of primary care attributes in preventing loss or change of usual source of care: a nationwide cohort study. Fam Pract. 2024;41(5):726–31. doi:10.1093/fampra/cmae006

33. Li W, Hou Y, An J, et al. Impact of family doctor contract services on preventable hospitalizations amongst patients with hypertension in rural China: mediating role of primary healthcare quality. Risk Manag Healthc Policy. 2024;17:2151–60. doi:10.2147/RMHP.S474933
34. Li L, Zhong C, Mei J, et al. Effect of family practice contract services on the quality of primary care in Guangzhou, China: a cross-sectional study using PCAT-AE. BMJ Open. 2018;8(11):e021317. doi:10.1136/bmjopen-2017-021317
35. Feng S, Cheng A, Luo Z, et al. Effect of family doctor contract services on patient perceived quality of primary care in southern China. BMC Fam Pract. 2020;21(1):218. doi:10.1186/s12875-020-01287-7
36. Honorato dos Santos de Carvalho VC, Rossato SL, Fuchs FD, et al. Assessment of primary health care received by the elderly and health related quality of life: a cross-sectional study. BMC Public Health. 2013;13:605. doi:10.1186/1471-2458-13-605
37. Santos NCCB, Vaz EMC, Nogueira JA, et al. Presença e extensão dos atributos de atenção primária à saúde da criança em distintos modelos de cuidado [Presence and extent of primary care characteristics under different models for children's healthcare]. Cad Saude Publica. 2018;34(1):e00014216. doi:10.1590/0102-311X00014216
38. Zheng Y, Liu Y, Xue D, et al. Research on the control rate of hypertension under family physician-contracted service. BMC Prim Care. 2024;25(1):47. doi:10.1186/s12875-024-02280-0
39. Kakemam E, Hajizadeh A, Doshmangir L. Insufficient evidence on the effectiveness of family physician program in Iran: a systematic review of existing literature. BMC Public Health. 2026;26(1):951. doi:10.1186/s12889-026-26636-6
40. Akbar MA, Sahar J, Rekawati E, et al. Development and psychometric properties of the interprofessional collaboration Practice Assessment Scale for primary care (Interprime Scale) in Indonesia. J Interprof Care. 2026 [Epub ahead of print]. doi:10.1080/13561820.2026.2719707
41. Maulina F, Hasanbasri M, Busari JO, et al. Impact of a pilot leadership development programme for rural and remote primary care physicians in Aceh, Indonesia: a Kirkpatrick level 3 evaluation. BMJ Lead. 2025 [Epub ahead of print]. doi:10.1136/leader-2023-000953
42. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71
43. Higgins JPT, Thomas J, Chandler J, et al., editors. Cochrane handbook for systematic reviews of interventions version 6.5 (updated August 2024). London: Cochrane; 2024. Available from: https://training.cochrane.org/handbook
44. Silva CS, Fonseca AD, Souza LP, et al. Integralidade e atenção primária à saúde: avaliação sob a ótica dos usuários [The comprehensive nature of primary health care: evaluation from the user standpoint]. Cien Saude Colet. 2014;19(11):4407–15. doi:10.1590/1413-812320141911.14122013
45. Ferrer AP, Brentani AV, Sucupira AC, et al. The effects of a people-centred model on longitudinality of care and utilization pattern of healthcare services--Brazilian evidence. Health Policy Plan. 2014;29(Suppl 2):ii107–13. doi:10.1093/heapol/czu077
46. Ferrer AP, Grisi SJ. Assessment of access to primary health care among children and adolescents hospitalized due to avoidable conditions. Rev Assoc Med Bras. 2016;62(6):513–23. doi:10.1590/1806-9282.62.06.513
47. Kessler M, Lima SBS, Weiller TH, et al. Longitudinality in primary health care: a comparison between care models. Rev Bras Enferm. 2018;71(3):1063–71. doi:10.1590/0034-7167-2017-0014
48. Liang C, Mei J, Liang Y, et al. The effects of gatekeeping on the quality of primary care in Guangdong Province, China: a cross-sectional study using primary care assessment tool-adult edition. BMC Fam Pract. 2019;20(1):93. doi:10.1186/s12875-019-0982-z
49. Zhang J, Chen X, Mao X, et al. Quality of primary care services: a perspective from chronic disease patients [in Chinese]. Chin Gen Pract. 2022;25(19):2391–8. doi:10.12114/j.issn.1007-9572.2022.0017
50. Yang H, Cui Y, Wang X, et al. Effect of family physician contract service on patients' perceived quality of primary child healthcare in urban China: analysis using propensity score matching. Public Health. 2025;242:124–30. doi:10.1016/j.puhe.2025.02.015
51. Lin C, Li S, Zhang Y, et al. Evaluation and measurement of primary care services based on patients' experience in China. Chin Gen Pract J. 2025;2(1):100047. doi:10.1016/j.cgpj.2025.100047
52. Harzheim E, Pinto LF, Hauser L, et al. Avaliação dos usuários crianças e adultos quanto ao grau de orientação para Atenção Primária à Saúde na cidade do Rio de Janeiro, Brasil [Assessment of child and adult users of the degree of orientation of primary healthcare in the city of Rio de Janeiro, Brazil]. Cien Saude Colet. 2016;21(5):1399–408. doi:10.1590/1413-81232015215.26672015
53. Macinko J, Mullachery PH. Primary care experiences among Brazilian adults: cross-sectional evidence from the 2019 National Health Survey. PLoS One. 2022;17(6):e0269686. doi:10.1371/journal.pone.0269686
54. Bastos EPB, Pinto LF. Evaluation of the longitudinality between users and residency programs in Family and Community Medicine in a region of the municipality of Rio de Janeiro, Brazil. Cien Saude Colet. 2024;29(11):e04002024.
doi:10.1590/1413-812320242911.04002024en
55. Leão CD, Caldeira AP. Avaliação da associação entre qualificação de médicos e enfermeiros em atenção primária em saúde e qualidade da atenção [Assessment of the association between the qualification of physicians and nurses in primary healthcare and the quality of care]. Cien Saude Colet. 2011;16(11):4415–23. doi:10.1590/s1413-81232011001200014
56. Jia C, Wang Z, Zhang L, et al. Strengthening primary care gatekeeping for healthy aging in China: family doctor contract service and multidimensional health outcomes among older adults. Front Public Health. 2026;14:1875855. doi:10.3389/fpubh.2026.1875855
57. Wang L, Liu W. Effects of family doctor contract services on the health-related quality of life among individuals with diabetes in China: evidence from the CHARLS. Front Public Health. 2022;10:865653. doi:10.3389/fpubh.2022.865653
58. Lamonato LCXL, Sarti TD, Almeida APSC. Effect of primary health care on the association between multimorbidity and emergency service utilization: National Health Survey, 2019. Rev Bras Epidemiol. 2024;27:e240062. doi:10.1590/1980-549720240062
59. Hone T, Been JV, Saraceni V, et al. Associations between primary healthcare and infant health outcomes: a cohort analysis of low-income mothers in Rio de Janeiro, Brazil. Lancet Reg Health Am. 2023;22:100519. doi:10.1016/j.lana.2023.100519
60. Ferreira-Batista NN, Teixeira AD, Diaz MDM, et al. Is primary health care worth it in the long run? Evidence from Brazil. Health Econ. 2023;32(7):1504–24. doi:10.1002/hec.4676
61. Carneiro VCCB, Ribeiro de Oliveira PT, Rassy Carneiro S, et al. Impact of expansion of primary care in child health: a population-based panel study in municipalities in the Brazilian Amazon. BMJ Open. 2022;12(3):e048897. doi:10.1136/bmjopen-2021-048897

62. Hajibadali P, Hashemiparast M, Kohneloo AJ, et al. Translating policy into health outcomes: a mixed-methods health impact assessment of the family physician program in Bonab County, Iran. BMC Public Health. 2025;25(1):3483. doi:10.1186/s12889-025-24788-5
63. Mellanen EH, Kauppila T, Kautiainen H, et al. Continuity of care and mortality in patients with type 2 diabetes: a cohort study. BJGP Open. 2025;9(1):BJGPO.2024.0144. doi:10.3399/BJGPO.2024.0144
64. Li X, Shi Y, Li X. Why do they sign but not consult? A study on the effectiveness of China's family doctor contract services from the perspective of patient-doctor interaction. Front Med (Lausanne). 2026;13:1848658.
doi:10.3389/fmed.2026.1848658
65. Feng J, Lei Z, Li X, et al. Acceptance of family doctors among residents in China: a cross-sectional study. Front Med (Lausanne). 2024;11:1435940.
doi:10.3389/fmed.2024.1435940
66. Zhao N, Gu M, Li J, et al. Factors influencing contracting of residents with family doctors in China: a national cross-sectional survey. BMC Health Serv Res. 2024;24(1):213. doi:10.1186/s12913-024-10606-y
67. Lv B, Wang Z, Meng K. Does family doctor contracting improve the health of the elderly: an empirical study from Beijing, China. J Multidiscip Healthc. 2026;19:1–17. doi:10.2147/JMDH.S596260
68. Kessler M, Thumé E, Marmot M, et al. Family Health Strategy, primary health care, and social inequalities in mortality among older adults in Bagé, southern Brazil. Am J Public Health. 2021;111(5):927–36.
doi:10.2105/AJPH.2020.306146
69. Barber SL, Mathauer I, Rosenberg M, et al. Purchasing primary care services for quality chronic care: capitation with performance payments in four countries. Int J Health Plann Manage. 2025;40(4):1007–12.
doi:10.1002/hpm.3929
70. Musonda JM, Moosa S. Exploring stakeholders' experiences on implementing family medicine in urban South Africa. Afr J Prim Health Care Fam Med. 2025;17(1):a4675. doi:10.4102/phcfm.v17i1.4675