

1 Development of the Tool for Advancing Practice Performance, a practice-level
2 survey to assess primary care structures and processes

3
4 Lorraine Kwok, MPH^{1*}/Margaret M. Paul, PhD, MS^{2†}, Stephanie L. Albert, PhD¹, Daphna
5 Harel, PhD³, Saul B. Blecker, MD, MS¹, Bryan J. Weiner, PhD⁴, Laura J. Damschroder, MS,
6 MPH⁵, Deborah J. Cohen, PhD⁶, Ann M. Nguyen, PhD, MPH⁷, Donna R. Shelley, MD, MPH⁸,
7 Carolyn A. Berry, PhD¹

8 ¹Department of Population Health, New York University Grossman School of Medicine, New
9 York, New York, United States of America

10 ²Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic,
11 Phoenix, Arizona, United States of America

12 ³Department of Applied Statistics, Social Science, and Humanities, New York University
13 Steinhardt School of Culture, Education, and Human Development, New York, New York,
14 United States of America

15 ⁴Department of Global Health, Department of Health Systems and Population Health, University
16 of Washington, Seattle, Washington, United States of America

17 ⁵Implementation Pathways, LLC, Chelsea, Michigan, United States of America

18 ⁶Department of Family Medicine, Oregon Health & Science University, Portland, Oregon,
19 United States of America

20 ⁷Rutgers Center for State Health Policy, New Brunswick, New Jersey, United States of America

21 ⁸Department of Public Health Policy and Management, New York University School of Global
22 Public Health, New York, New York, United States of America

23

24 * Corresponding author

25 Email: Lorraine.Kwok@nyulangone.org (LK)

26

27 [†]LK and MMP are Joint First Authors.

28

29 **Abstract**

30 Despite progress to define primary care practice transformation models, there remains gaps in
31 translating evidence-based guidelines into routine clinical care. Primary care providers (MD,
32 DO, NP, PA) and researchers need tools to assess modifiable factors that improve practice
33 performance to inform practice transformation efforts. We aimed to develop a pragmatic tool for
34 assessing practice-level primary care structures and processes that are associated with better care
35 quality and clinical outcomes. We generated 314 candidate items for the Tool for Advancing
36 Practice Performance (TAPP) using data from a comprehensive literature review, Delphi study,
37 and qualitative interviews with high-performing practices. We used empirical criteria and expert
38 review to eliminate redundancy and improve clarity via removing and retaining items. The
39 retained items were formatted into a survey tool, and we further revised the tool based on
40 feedback elicited from cognitive interviews and pilot testing with primary care providers and
41 staff. The final candidate pool comprised 126 items after refinement and expert review. For the
42 survey tool, we adapted and developed survey questions for each of the 126 items. Eight
43 cognitive interview participants reviewed the tool and provided feedback on its content and
44 language. Based on this feedback, we eliminated 13 items because they were poorly or
45 incorrectly understood by participants, resulting in a 113-item tool. Fifteen participants pilot
46 tested the tool and no additional items were eliminated. The TAPP is a novel, low-burden tool
47 that researchers and primary care providers can use to identify areas for improvement at the
48 practice-level. Practices and health systems could use the TAPP to assess their own performance
49 and identify gaps in their structures and processes, and practice networks and health systems can
50 use the tool to assess structures and processes at individual clinics, track this information over
51 time, and evaluate its relationship to care quality and clinical outcomes.

52 **Introduction**

53 Health care delivered according to evidence-based guidelines has been shown to lead to
54 improvements in quality and clinical outcomes (1, 2). Over the past 15 years, new models of care
55 such as the Chronic Care Model and the Patient-Centered Medical Home (PCMH) principles
56 have been widely adopted throughout the US with the hope of transforming primary care
57 delivery to improve care quality and outcomes (3, 4). Further, the articulation and dissemination
58 of the PCMH principles by the American Academy of Family Physicians and other agencies
59 sparked development of new tools designed for PCMH recognition and quality improvement
60 processes, as well as the evaluation and accreditation of the level of “medical homeness” and its
61 correlation with the triple aim: improving patients’ experience with care, improving population
62 health, and reducing healthcare costs (5-7). There are a number of practice-level tools designed
63 to measure overall primary care performance more generally, such as the Assessment of Chronic
64 Illness Care (ACIC) (8) and the PCMHA (9), as well as tools that measure specific aspects of
65 primary care believed to be predictive of performance including aspects such as teamwork (10,
66 11), organizational communication climate (12), and care coordination (13). There are several
67 patient-reported measures used to assess the overall quality of primary care (14-23); however,
68 patient-reported measures are labor intensive due to the burden of extensive primary data
69 collection and analysis and therefore may be impractical for providers to conduct.

70

71 The underlying assumption behind the development and use of these tools is that the degree of
72 fidelity to the model or desired state leads to a better score, and that the scores on the tools are
73 directly associated with improved outcomes related to the various aspects of the triple aim.

74 However, a notable and significant gap exists: most of the tools used to assess such performance
75 have been neither rigorously developed nor assessed for reliability and validity, and none have
76 established predictive validity with respect to care quality or patient outcomes (6, 7, 24-29).
77 Therefore, despite progress made to define broad practice transformation theories and models
78 and to develop tools to assess practice-level characteristics, there remains a gap in the translation
79 of evidence-based guidelines into routine practice in primary care settings (30). Providers and
80 researchers need tools to assess modifiable factors that improve primary care performance and
81 inform efforts to transform care (2, 31).

82
83 Our overarching study aims to develop and validate a low-burden, practice-level tool that
84 primary care providers and staff can use to assess the factors associated with better care quality
85 outcomes for their practices. Our approach to data collection to inform tool development was
86 guided by the Donabedian framework for assessing health services and quality of care with a
87 focus on structures, or the context in which care is delivered, and processes, the methods by
88 which health care is delivered, as the main drivers of these outcomes (32). The aim of this paper
89 is to describe our comprehensive and data-driven approach to developing the Tool for Advancing
90 Practice Performance (TAPP), including our initial data collection of evidence for specific
91 structures and processes and the subsequent survey development, and cognitive interviews and
92 pilot testing to refine the instrument.

93

94 **Methods**

95 As part of our larger study, we first generated candidate items for the tool from three sources: 1)
96 a comprehensive literature review of US-based research studies linking practice-level primary
97 care structures and processes to quality and clinical outcomes; 2) a Delphi study in which 29
98 experts in the fields of primary care, health care management, and health services research rated
99 the items generated from the literature review in terms of importance to chronic disease
100 management and preventative care and feasibility of implementation (33); and 3) qualitative
101 interviews of providers and quality improvement specialists from a national sample of 20 high
102 performing primary care practices (34, 35). We engaged in a concept mapping exercise with 36
103 primary care providers, health system leaders, health services researchers, and implementation
104 scientists to develop an empirically-derived structure to organize the items (36). The resulting
105 structure consisted of 8 domains related to primary care: 1) Address Social Factors and
106 Encourage Patient Engagement, 2) Reduce Clinical Risk Factors, 3) Provide Enhanced Care, 4)
107 Expand Access to Care, 5) Provide Ancillary Services, 6) Establish Care Team Processes and
108 Workflows, 7) Use Clinical Information Systems, and 8) Use Data and Evidence.

109
110 Following item generation and concept mapping, the development of the TAPP proceeded in
111 four phases: 1) selection and refinement of items generated by our prior research, 2) survey tool
112 development, 3) cognitive interviews to further refine the tool, and 4) pilot testing (see Fig 1).
113 Throughout the study, the team met regularly with an expert panel that included implementation
114 science researchers, health services researchers, and health system stakeholders; this expert panel
115 was involved in key decision making during each phase of the research described in this paper.
116 The study was approved by the New York University (NYU) Grossman School of Medicine
117 Institutional Review Board.

118 **Fig 1. Flowchart of the development and content validation of the Tool for Advancing**
119 **Practice Performance (TAPP).**

120

121 **Generation of items**

122 The primary purpose of the item generation stage was to empirically derive the content for our
123 survey tool and to establish its content validity.

124

125 **Literature review**

126 In consultation with the Medical Library at NYU, we developed and refined our search strategy.
127 We searched PubMed/MEDLINE, Embase, and Cochrane CENTRAL, and Web of Science for
128 US-based research studies published in English between January 1, 2010 and the time the study
129 began: December 31, 2018. We included any published study, commentary, or editorial that
130 mentioned practice-level primary care structures and/or processes with the exception of articles
131 published in languages other than English, studies based entirely outside the US, conference
132 abstracts, and dissertations. We reviewed articles and collected data from our literature search in
133 four stages: 1) title and abstract screening, 2) full-text article review, 3) full-text extraction, and
134 4) assessment of empirical evidence. Two members of the study team independently reviewed
135 the content of each article at every stage and conflicts in opinion between pairs of reviewers were
136 discussed as a group and resolved by consensus. During the title and abstract screening, the team
137 screened articles and eliminated those that did not meet the inclusion criteria. In the second
138 stage, the team reviewed the full text of articles to ensure retention of only those that met the

139 inclusion criteria. We also reviewed the full text of articles in the review articles identified in the
140 literature search and retained those that met the inclusion criteria.

141

142 **Delphi study**

143 Our Delphi study was designed to evaluate the impact and feasibility of primary care structures
144 and processes identified through our literature review and input from our expert panel. We
145 developed this study using the Conducting and Reporting Delphi Studies (CREDES) guidelines.
146 The sample of participants included 29 primary care providers, health system leaders, and health
147 services researchers in the U.S. In the first round of the study, participants rated the importance
148 of each item with respect to both chronic disease management and screening and prevention. In
149 the second round of data collection, participants reviewed aggregate results from the first round
150 and were asked to come to consensus around the findings. Participants were also given the
151 opportunity to contribute new items that were not included in our list of structures and processes,
152 which led to the generation of 3 new items for consideration in our tool. Scores from the Delphi
153 study were assigned to each structure and process in our list and used as evidence to retain or
154 discard each item from the final survey set. The full Delphi study is further described in Albert et
155 al (33).

156

157 **Practice interviews**

158 We conducted qualitative key informant interviews with providers and staff familiar with quality
159 improvement activities. Our sample included 44 interviewees from 22 high performing practices

160 located throughout the US. High performance was defined as meeting at least 2 of the following
161 in 2018: among eligible patients, 70% prescribed aspirin use, 70% with controlled blood pressure
162 (<140/90), 70% prescribed statin therapy, and 90% with controlled diabetes (HBA1c<9). We
163 conducted semi-structured interviews, recorded our discussions, and coded the data to assess
164 perspectives on structures and processes that contributed to their success. As with the Delphi
165 scoring data, we incorporated data on the frequency with which structures and processes were
166 discussed during these interviews as drivers of high performance on our outcomes of interest.
167 This information contributed to our decision to retain or discard items for the tool.

168

169 **Selection and refinement of items**

170 Data from three empirical data sources yielded 314 unique structures and processes, including
171 258 structures and processes identified from an initial broad review of abstracts in peer-reviewed
172 journals and experience of the investigative team and expert panel and subsequently assessed in
173 the Delphi study, and 56 additional items (18 from the final, full-text phase of the literature
174 review, 3 suggested by experts from the Delphi study, 35 from the qualitative interviews). All
175 314 candidate items mapped onto eight domains related to primary care, which were derived
176 from the concept mapping exercise that enabled development of an empirically-derived
177 organizational structure for the 258 items included in the Delphi study.

178

179 We compiled all 314 items and the corresponding data in a matrix managed in Microsoft Excel.
180 The reliance on empirical data from the literature review, Delphi study, and qualitative
181 interviews rigorously established content validity of the tool. In order to refine the pool of

182 structures and processes, we developed empirical criteria for removing and retaining items that
183 relied on information from each mode of item generation. The guiding principle for inclusion in
184 the TAPP was that each item had to have solid evidence, from at least one of our three sources of
185 expert data, of a relationship to quality of care or patient outcomes.

186

187 Our Delphi process generated ratings on the perceived impact each item would have separately
188 on chronic disease management or preventative care outcomes. We retained items if either the
189 chronic disease management average or the screening and prevention average from the Delphi
190 study was 3.5 out of 5 or above, an indication of group consensus on the item being of relatively
191 high importance. We removed items if the Delphi score was below 3.5 for both chronic disease
192 management and the screening and preventative means. Eliminated items could be reintroduced
193 into the pool if four or more high performing practices mentioned them in the qualitative
194 interviews or if there was at least one positive finding from the literature (i.e., the item was
195 positively correlated with at least one relevant quality of care or clinical outcome variable).

196

197 We continued refining the set of items with regular meetings and input from our expert panel,
198 until consensus was reached that all items were unique, clear, and specific. Items that were
199 deemed redundant, lacked clarity, lacked specificity, or were considered universal or non-
200 discriminatory were dropped from the list of candidate items.

201

202 **Survey tool development**

203 We formatted a list of candidate items to create a low-burden survey tool (37), one that is
204 parsimonious and user-friendly. Our goal in terms of burden was a tool that could be completed
205 within 30 minutes and where items were unambiguous and easy to answer with accuracy. All but
206 four of the survey items had a 3-point response set: yes, no, don't know/unsure. Specific wording
207 of questions was informed by an existing instrument (29) and developed de novo by the study
208 team. We refined the tool through an iterative process with our expert panel and study team
209 members to finalize formatting, organization, and phrasing of questions and response options.
210 Our objective was to ensure that survey questions accurately reflected the structures and
211 processes, and that each item was sufficiently brief and clear for primary care providers and
212 staff.

213

214 **Cognitive interviews**

215 We conducted eight cognitive interviews using videoconferencing from August to October 2021
216 to elicit feedback and to inform any further needed refinements. Participants included practice
217 managers, office managers, providers, and quality improvement leaders from primary care
218 practices that are part of two practice-based research networks in New York City and had
219 previously participated in the qualitative interviews: New York City Department of Health and
220 Mental Hygiene's Bureau of Equitable Health Systems (BEHS), a bureau that supports New
221 York City primary care practices in implementing patient-centered strategies for chronic disease
222 prevention and management (38), and NYU Langone Health Faculty Group Practices, a network
223 of three large multidisciplinary ambulatory care locations and more than 30 single and multi-
224 specialty ambulatory care sites (39). We used concurrent think-aloud cognitive interviews where

225 participants shared their thought process as they answered survey questions (37, 40). To facilitate
226 this, participants shared the survey on their computer screens with the interviewers who followed
227 along as the participants read each survey question aloud and narrated their thought processes in
228 interpreting and answering each question. The interviewers probed for more explanation as
229 needed. The use of a virtual meeting platform did not appear to compromise the quality of the
230 feedback received. Interviews lasted for about one hour and participants received a \$150
231 honorarium for their time and effort. We reviewed interview notes and through group consensus,
232 determined survey revisions based on participants' experiences.

233

234 **Pilot testing**

235 Lastly, we piloted the final version of the TAPP with respondents from 15 practices to assess the
236 mechanics of online survey administration, discern the time needed to complete the survey,
237 clarify survey instructions, identify or confirm the roles most appropriate to engage in the survey,
238 address any technical issues, and understand the overall experience taking the survey. The
239 practices were recruited from three practice-based research networks: BEHS; DARTNet, a
240 nonprofit organization that fosters research and supports collaboration with health care
241 organizations, academic medical centers, and providers to extract data from a range of electronic
242 health record (EHR) systems and use those data for quality improvement and research (41); and
243 OCHIN, a nonprofit organization that provides an EHR system to over 500 federally qualitative
244 health centers and community health centers nationwide (42). In addition to providing written
245 feedback at the end of the survey, participants had the option to provide additional feedback via a

246 scheduled call with study team members. Participants received a \$150 honorarium for their time
247 and effort.

248

249 **Results**

250 **Selection and refinement of items**

251 After applying criteria for removing and retaining items to the potential pool of items (n=314),
252 188 items remained in the candidate pool of items, which included 125 items with a Delphi score
253 of 3.5 or higher for chronic disease management or screening and preventative care and 63 items
254 that did not meet criteria in the Delphi process but were reintroduced because four or more
255 practices mentioned the item in the qualitative interviews and/or there was at least one positive
256 finding from the literature. Of the 63 items that were reintroduced, 43 items were mentioned by 4
257 or more practices in the qualitative interviews; 12 items had at least one positive finding from the
258 literature; 5 items were mentioned in the qualitative interviews and had at least one positive
259 finding from the literature; and 3 new items were suggested by experts in the Delphi study.

260

261 During the expert review of the 188 candidate items, we eliminated 70 items including 23 items
262 due to redundancy; 8 for lack of clarity/specificity; 10 for being universal/non-discriminatory; 1
263 with low feasibility ranking from the Delphi study; 1 that reflected one of our outcomes of
264 interest rather than a structure or process (i.e., documentation of patients' smoking status); and 27
265 conceptually similar items were combined to form new item. This reduction process resulted in a
266 refined pool of 118 items.

267

268 Further, we added eight items that were not previously included in our original list of structures
269 and processes. Through expert review and consensus, we included five new items that were split
270 off from existing items that were bundled (e.g., ‘Practice routinely screens patients for alcohol,
271 drug, and tobacco use’ was split into three separate items to individually capture alcohol, drug,
272 and tobacco use screening); two new items that capture additional staff that are part of a care
273 team; and one new item that captures the use of a visual dashboard to manage preventative care
274 for patients.

275

276 After distilling and refining the items, the final candidate pool comprised 126 items.

277

278 **Survey tool development**

279 For the survey tool, we adapted and developed survey questions for each of the 126 items
280 retained in our item pool. We improved the wording of the questions, response options, and
281 organization of the tool based on input from our expert panel and meetings with study team
282 members. To enhance the ease of completion, we used a 3-point response scale: yes, no, and
283 don’t know/unsure. This scale requires less cognitive burden compared to estimates of degree or
284 frequency of implementation. Questions were organized within the eight domains identified in
285 our concept mapping exercise.

286

287 **Cognitive interviews**

288 Our eight cognitive interview participants reviewed the 126-item tool and provided feedback on
289 the content, item structure, and language as they reviewed each question. We reviewed and
290 discussed the respondents' feedback in real time and edited the survey as we progressed through
291 the interviews. Ultimately, we eliminated 13 items from the tool because they were poorly or
292 incorrectly understood by respondents in the interviews, which resulted in a tool with 113 items.
293 Additionally, language for 23 items was modified for clarity based on feedback from the
294 cognitive interviews (e.g., "In this practice, how often do patients see the same PCP?" was
295 modified to "In this practice, how often do patients see the same PCP (other than vacation
296 time)?).

297

298 **Pilot testing**

299 Fifteen participants including executive directors, practice managers, office managers, providers,
300 and quality improvement leaders piloted the 113-item tool. On average, the participants
301 completed the TAPP survey in 30 minutes, meeting our threshold for low burden. The time it
302 took to complete the TAPP ranged from 20 to 90 minutes. In addition to providing written
303 feedback at the end of the pilot survey, two participants opted to provide additional feedback via
304 a scheduled call with the study team. No additional items were eliminated beyond those
305 eliminated based on feedback from the cognitive interviews. Survey instructions were modified
306 slightly, and technical issues related to online survey administration were optimized for data
307 collection.

308

309 **Discussion**

310 The TAPP is the first primary care practice assessment tool based on empirical evidence of
311 relationships between its items and clinical quality and outcome measures, thus establishing
312 strong content validity. The tool assesses 113 discrete primary care structures and processes that
313 contribute to high performance and can be completed within 30 minutes by one or multiple
314 primary care practice staff including practice managers, office managers, providers, and quality
315 improvement leaders. The TAPP was developed through multiple phases of research including
316 selection and refinement of items generated from a comprehensive literature review, a Delphi
317 study, and qualitative interviews with high-performing practices; development and refinement of
318 survey items; cognitive interviewing; and pilot testing. Expert review of items from multiple data
319 sources, cognitive interviewing, and pilot testing resulted in a refined set of primary care
320 structures and processes that, if changed, could reasonably be expected to lead to improvements
321 in health care quality and clinical outcomes related to chronic disease management and
322 prevention. We know of no other measures of primary care practice performance that were
323 developed with such systematic and comprehensive content validity. Moreover, the tool is
324 relatively short and easy to complete with majority yes/no/unsure response set and thus
325 represents a low burden on practices. The TAPP is free and publicly available for review and use
326 in research and practice (see S1 File).

327

328 The development of the TAPP underscores the value of drawing upon multiple methods and data
329 sources. While there was overlap between items identified in the literature, by Delphi experts,
330 and providers and quality improvement specialists at high performing practices, we found that

331 each source contributed key items that would have been underrepresented or absent if we had
332 relied on a single source. Three new items emerged only through the Delphi panel study (e.g., the
333 practice screens all patients using a standardized tool to identify past trauma; the practice has
334 providers and staff trained in implicit bias), and 35 new items emerged from the qualitative
335 interviews with practices (e.g., practice partners with onsite or local dental services; providers
336 routinely update problem list in EHR). Additionally, during the Delphi exclusion stage, 37 items
337 were “rescued” because they were mentioned by four or more practices in the qualitative
338 interviews, suggesting that primary care experts and on-the-ground primary care providers and
339 staff may have differing views on what items most affect care quality and chronic disease
340 management. Efforts to develop and validate complex measurement tools in primary care should
341 consider methods to gather data from multiple data sources to ensure comprehensive
342 identification of primary care elements of interest.

343

344 The TAPP captures complex elements of primary care, such as structures and processes thought
345 to be readily changeable at the practice level (i.e., without external support). The tool is a new
346 resource for researchers and primary care providers to use for assessing care structures and
347 processes and identifying areas for improvement at the practice-level. Practices could use the
348 TAPP as a self-assessment of their performance and identify gaps in their structures and
349 processes. Practice networks and health systems could use the TAPP to assess the structures and
350 processes in place at individual clinics, track this information over time, and assess its
351 relationship to quality and clinical outcomes. Additionally, the TAPP has the potential to
352 measure the effectiveness of interventions designed to improve practice performance. Future
353 work will evaluate the scoring and reliability of the TAPP based on data collected from practices

354 across the US as well as validate the tool against care quality and patient outcomes. Further
355 research is also needed to incorporate recent advancements in the use of telehealth in the delivery
356 of routine primary care.

357

358 **Acknowledgements**

359 The authors thank the participants from the cognitive interviews and pilot testing for reviewing
360 and providing feedback on the survey tool.

361 **References**

- 362 1. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic
363 review of rigorous evaluations. *Lancet*. 1993;342(8883):1317-22.
- 364 2. Malouin RA, Starfield B, Sepulveda MJ. Evaluating the tools used to assess the medical
365 home. *Manag Care*. 2009;18(6):44-8.
- 366 3. American Academy of Family Physicians, American Academy of Pediatrics, American
367 College of Physicians, American Osteopathic Association. *Joint Principles of the Patient-*
368 *Centered Medical Home*. 2007.
- 369 4. Ferrante JM, Balasubramanian BA, Hudson SV, Crabtree BF. Principles of the patient-
370 centered medical home and preventive services delivery. *Ann Fam Med*. 2010;8(2):108-16.
- 371 5. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff*
372 (Millwood). 2008;27(3):759-69.
- 373 6. Burton RA, Devers KJ, Berenson RA. *Patient-Centered Medical Home recognition tools:*
374 *A comparison of ten surveys' content and operational details*. Washington, DC: The Urban
375 Institute, Health Policy Center; 2012.
- 376 7. Quigley DD, Predmore ZS, Hays RD. Tools to gauge progress during Patient-Centered
377 Medical Home transformation. *Am J Accountable Care*. 2017;5(4):e8-e18.
- 378 8. Bonomi AE, Wagner EH, Glasgow RE, VonKorff M. Assessment of chronic illness care
379 (ACIC): a practical tool to measure quality improvement. *Health Serv Res*. 2002;37(3):791-820.
- 380 9. Daniel DM, Wagner EH, Coleman K, Schaefer JK, Austin BT, Abrams MK, et al.
381 *Assessing progress toward becoming a patient-centered medical home: an assessment tool for*
382 *practice transformation*. *Health Serv Res*. 2013;48(6 Pt 1):1879-97.

- 383 10. Baker DP, Amodeo AM, Krokos KJ, Slonim A, Herrera H. Assessing teamwork attitudes
384 in healthcare: development of the TeamSTEPPS teamwork attitudes questionnaire. *Qual Saf*
385 *Health Care*. 2010;19(6):e49.
- 386 11. Lurie SJ, Schultz SH, Lamanna G. Assessing teamwork: a reliable five-question survey.
387 *Fam Med*. 2011;43(10):731-4.
- 388 12. Wynia MK, Johnson M, McCoy TP, Griffin LP, Osborn CY. Validation of an
389 organizational communication climate assessment toolkit. *Am J Med Qual*. 2010;25(6):436-43.
- 390 13. Zlateva I, Anderson D, Coman E, Khatri K, Tian T, Fifield J. Development and
391 validation of the Medical Home Care Coordination Survey for assessing care coordination in the
392 primary care setting from the patient and provider perspectives. *BMC Health Serv Res*.
393 2015;15:226.
- 394 14. Scholle SH, Vuong O, Ding L, Fry S, Gallagher P, Brown JA, et al. Development of and
395 field test results for the CAHPS PCMH Survey. *Med Care*. 2012;50 Suppl(Suppl):S2-10.
- 396 15. Hays RD, Berman LJ, Kanter MH, Hugh M, Oglesby RR, Kim CY, et al. Evaluating the
397 psychometric properties of the CAHPS Patient-centered Medical Home survey. *Clin Ther*.
398 2014;36(5):689-96 e1.
- 399 16. Flocke SA. Measuring attributes of primary care: development of a new instrument. *J*
400 *Fam Pract*. 1997;45(1):64-74.
- 401 17. Roland M, Roberts M, Rhenius V, Campbell J. GPAQ-R: development and psychometric
402 properties of a version of the general practice assessment questionnaire for use for revalidation
403 by general practitioners in the UK. *BMC Fam Pract*. 2013;14:160.
- 404 18. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported
405 Outcomes Measurement Information System (PROMIS) developed and tested its first wave of

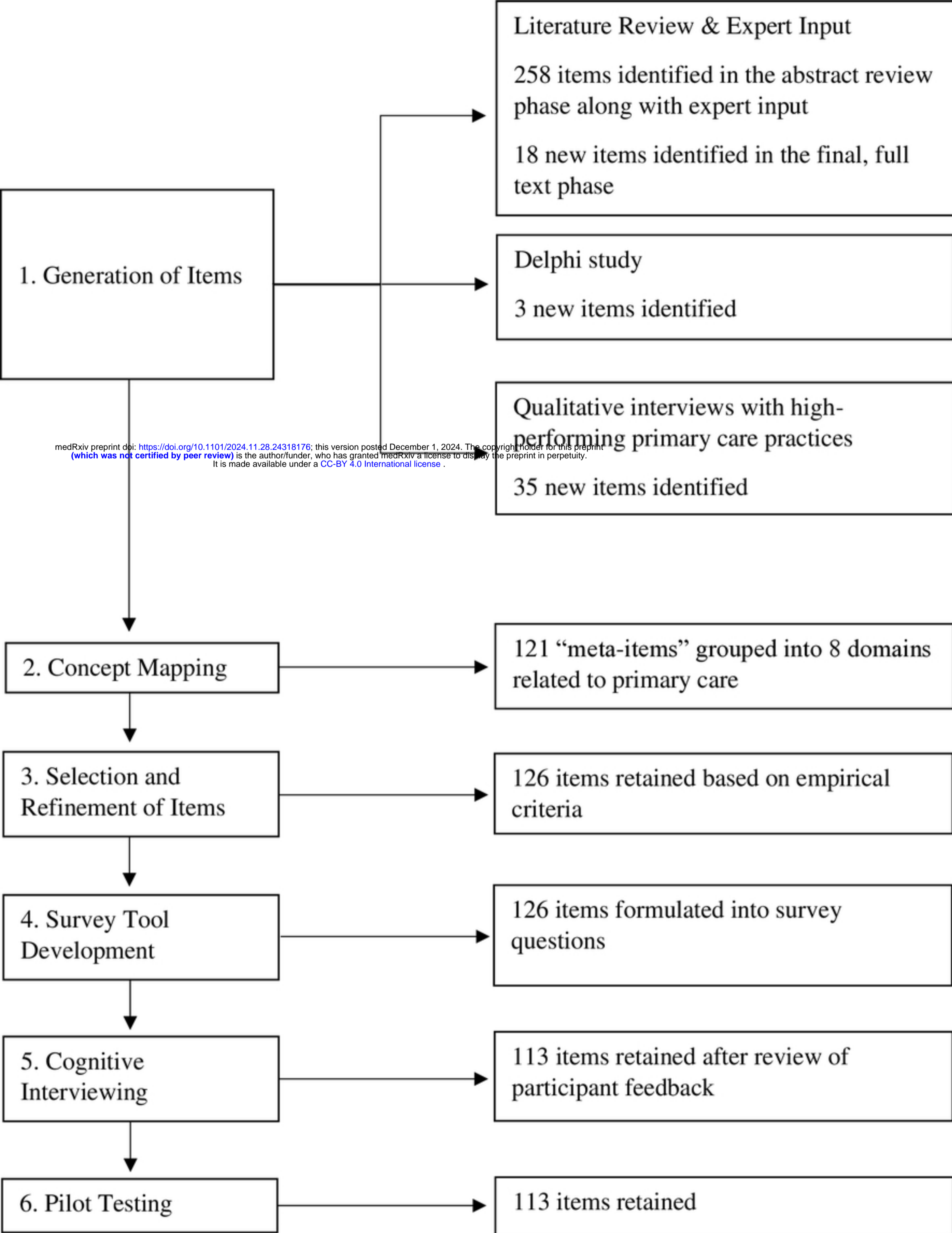
- 406 adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179-
407 94.
- 408 19. Hawthorne G, Sansoni J, Hayes L, Marosszeky N, Sansoni E. Measuring patient
409 satisfaction with health care treatment using the Short Assessment of Patient Satisfaction
410 measure delivered superior and robust satisfaction estimates. *J Clin Epidemiol.* 2014;67(5):527-
411 37.
- 412 20. Wells N, Bronheim S, Zyzanski S, Hoover C. Psychometric evaluation of a Consumer-
413 Developed Family-Centered Care Assessment Tool. *Matern Child Health J.* 2015;19(9):1899-
414 909.
- 415 21. Glasgow RE, Wagner EH, Schaefer J, Mahoney LD, Reid RJ, Greene SM. Development
416 and validation of the Patient Assessment of Chronic Illness Care (PACIC). *Med Care.*
417 2005;43(5):436-44.
- 418 22. Boecker M, Iglesias K, Peytremann-Bridevaux I. Psychometric evaluation of the Patient
419 Assessment of Chronic Illness Care instrument using Item analysis according to the Rasch
420 model. *J Eval Clin Pract.* 2022;28(4):670-9.
- 421 23. Etz RS, Zyzanski SJ, Gonzalez MM, Reves SR, O'Neal JP, Stange KC. A new
422 comprehensive measure of high-value aspects of primary care. *Ann Fam Med.* 2019;17(3):221-
423 30.
- 424 24. Alexander JA, Paustian M, Wise CG, Green LA, Feters MD, Mason M, et al.
425 Assessment and measurement of patient-centered medical home implementation: the BCBSM
426 experience. *Ann Fam Med.* 2013;11 Suppl 1(Suppl 1):S74-81.

- 427 25. Doty MM, Tikkanen R, Shah A, Schneider EC. Primary care physicians' role in
428 coordinating medical and health-related social needs in eleven countries. *Health Aff (Millwood)*.
429 2020;39(1):115-23.
- 430 26. Friedberg MW, Safran DG, Coltin KL, Dresser M, Schneider EC. Readiness for the
431 Patient-Centered Medical Home: structural capabilities of Massachusetts primary care practices.
432 *J Gen Intern Med*. 2009;24(2):162-9.
- 433 27. Friedberg MW, Coltin KL, Safran DG, Dresser M, Zaslavsky AM, Schneider EC.
434 Associations between structural capabilities of primary care practices and performance on
435 selected quality measures. *Ann Intern Med*. 2009;151(7):456-63.
- 436 28. Friedberg MW, Schneider EC, Rosenthal MB, Volpp KG, Werner RM. Association
437 between participation in a multipayer medical home intervention and changes in quality,
438 utilization, and costs of care. *JAMA*. 2014;311(8):815-25.
- 439 29. Berry CA, Mijanovich T, Albert S, Winther CH, Paul MM, Ryan MS, et al. Patient-
440 centered medical home among small urban practices serving low-income and disadvantaged
441 patients. *Ann Fam Med*. 2013;11 Suppl 1(Suppl 1):S82-9.
- 442 30. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of
443 research findings. *Implement Sci*. 2012;7:50.
- 444 31. Coleman K, Wagner E, Schaefer J, Reid R, LeRoy L. Redefining primary care for the
445 21st century. Rockville, MD: Agency for Healthcare Research and Quality; 2016.
- 446 32. Donabedian A. The quality of care. How can it be assessed? *JAMA*. 1988;260(12):1743-
447 8.

- 448 33. Albert SL, Kwok L, Shelley DR, Paul MM, Blecker SB, Nguyen AM, et al. Identifying
449 important and feasible primary care structures and processes in the US healthcare system: a
450 modified Delphi study. *BMJ Open*. 2024;14(11):e082989.
- 451 34. Nguyen AM, Paul MM, Shelley DR, Albert SL, Cohen DJ, Bonsu P, et al. Ten common
452 structures and processes of high-performing primary care practices. *J Public Health Manag Pract*.
453 2022;28(2):E639-E44.
- 454 35. Cohen DJ, Wyte-Lake T, Bonsu P, Albert SL, Kwok L, Paul MM, et al. Organizational
455 factors associated with guideline concordance of chronic disease care and management practices.
456 *J Am Board Fam Med*. 2022;35(6):1128-42.
- 457 36. Paul M, Albert S, Shelley D, Blecker S, Kwok L, Berry C. Use of concept mapping to
458 derive a taxonomy of practice-level structures and processes to improve primary care outcomes.
459 *AcademyHealth Science of Dissemination and Implementation Annual Conference; 2023*;
460 Arlington, VA.
- 461 37. Groves RM, Fowler Jr. FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. *Survey
462 methodology*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2009.
- 463 38. NYC REACH. About us [Available from: <https://nycreach.org/about-us/>].
- 464 39. NYU Langone Health Faculty Group Practices. NYU Langone Health Faculty Group
465 Practices, our story. [Available from: <https://nyulangone.org/our-story>].
- 466 40. Alhadreti O, Mayhew P. To intervene or not to intervene: An investigation of three think-
467 aloud protocols in usability testing. *Journal of Usability Studies*. 2017;12(3):111-32.
- 468 41. DARTNet Institute. About DARTNet Institute [Available from:
469 <http://www.dartnet.info/AboutDI.htm>].
- 470 42. OCHIN. Mission & history [Available from: <https://ochin.org/mission-history>].

471 **Supporting information**

472 **S1 File. The Tool for Advancing Practice Performance (TAPP).**



Figure