Evaluating primary care research networks: a review of currently available tools

Joan M. C. Bleeker
Wim A. B. Stalman
Henriëtte E. van der Horst

Published: J Am Board Fam Med 2010;23:465–475.

(Reproduced by permission of the American Board of Family Medicine)
Abstract

Background
An increasing number of primary care research networks (PCRN) are being developed around the world. Despite the fact that they have existed for a long time in some countries, little is known about what they have actually achieved. There is an ongoing debate in the literature about the appropriate framework for the evaluation of PCRN. Here, we aim to provide an overview of the tools that are currently available for measuring the performance of PCRN and practices involved in PCRN or research.

Methods
We performed electronic searches in bibliographic databases and several additional searches. We composed a checklist to evaluate the design, content, and methodological quality of the tools.

Results
We identified four tools for the evaluation of PCRN or the measurement of primary care practices involved in PCRN or research.

Conclusions
The results of our study showed that various methods, areas of interest, dimensions, and indicators for the evaluation of PCRN have been proposed. However, no generic and validated tool that enables meaningful comparison between different network models has been developed. It is, therefore, time to reflect on the appropriateness and effectiveness of PCRN and determine the desired outcomes (ends) of PCRN and how we can best achieve them in the future (means). To open up the “black box” of the effectiveness of the PCRN, it may be relevant to observe the effects of network and research participation on those involved in networks.

Introduction
Increasing numbers of primary care research networks (PCRN) are being developed around the world. In some countries their number is steadily increasing. In the United States, approximately 100 PCRN are currently active whereas a decade earlier only 28 PCRN could be identified. PCRNs initially were developed to provide an infrastructure for primary care research, to strengthen and develop the research base of primary care, and to increase research capacity and research activity among primary care providers. Current networks operate in different ways, depending on aspects such as contexts, baseline of experience, and resources. Therefore, PCRN represent a great diversity of network models, objectives, and configurations. In general, PCRN aim to conduct research close to practice, identify and respond to current primary care research needs, and to engage more primary care professionals with research. In addition to these research-based objectives, it has been suggested that PCRN can also play an important role in the translation and diffusion of research findings into practice, and other activities that are relevant in the field of primary care, like quality improvement, training, and education. Internationally, both policy advisers and researchers
have acknowledged the importance of PCRNs and have recognized the challenging opportunities they provide for the improvement of primary care. Yet, despite their long existence in some countries like The Netherlands and the United Kingdom, little is known about the effectiveness of PCRNs or their added value compared with other research capacity-building initiatives in primary care. Another issue that has not yet been studied is the impact of network and research involvement in daily practice in terms of increased use of research evidence and quality improvement.

There seems to be a fundamental lack of evaluative studies about the progress, impact, and output of PCRNs, which prevents a validated comparison of different network models and thus insight into what type of model is optimal for achieving specific goals. Valid and agreed methods of evaluation are urgently needed to provide stakeholders and policy advisers with comprehensive and meaningful data about these issues. Therefore, we reviewed the literature on PCRNs to identify and describe currently available frameworks and tools for evaluating PCRNs and practices involved in PCRNs or research. The design, content, and methodological quality of the identified frameworks and tools were also reviewed to provide more detailed information about their usefulness.

**Methods**

**Search strategy**

We used the search strategy described in Table 1 to search for relevant publications in the computerized bibliographic database of Medline (PubMed). Subsequently, we searched Embase and Picarta for additional publications. We also screened the references of relevant articles, contacted experts and network organizations, and searched the Internet (Google) for additional publications. Initially, we conducted a manual search in the Dutch scientific journal of general practice (*Huisarts en Wetenschap*) to identify additional publications concerning Dutch PCRNs.

**Table 1. Key words and search strategy**

| Search strategy for Medline¹ | (network¹(tl) AND (health care(tl) OR practice¹(tl) OR research(tl) OR primary care(tl) OR primary health(tl)) OR pbrn(tw) OR (“Community Networks” (MeSH) OR “Group Structure”(MeSH)) AND (_Primary Health Care_(MeSH) OR “Family Practice”(MeSH) OR “Physicians, Family”(MeSH) OR gp(tw) OR (_general OR family_ AND _physician ¹⁾ OR practi¹⁾ OR doctor ¹⁾_ OR “primary care” OR “primary health care”)
| English key words² | “network” combined with “primary care” or “research” or “practice-based” or “general practice” or “pcrn”
| Dutch key words³ | “academisering” or “netwerk” or “werkplaats” combined with “academisch” or “huisarts”

¹ Adding the word ‘academic’ to the search did not yield more relevant publications, but hugely increased the amount of irrelevant publications.
² Used for Embase and Google.
³ Used for Picarta and Google.
Inclusion criteria
Two authors (JB and HvdH) independently screened the titles and abstracts of all citations and selected articles for full-text reading. Each author selected articles for the preliminary selection by using the following inclusion criteria: 1) publications that discussed the evaluation of PCRNs; 2) publications that presented or discussed a (conceptual) framework or tool for measuring the performance of PCRN or practices involved in PCRN or research; and 3) studies that reported the results of a (comparative) PCRN evaluation study. This preliminary selection enabled us to review any method, framework, or tool that had been previously proposed, developed, or used. In the final selection we included only those articles that provided information regarding a (conceptual) framework or tool used for measuring the performance of PCRN or practices involved in PCRN or research.

After each selection both reviewers met to discuss the differences in their article selection and establish consensus. We excluded editorials, letters, and comments.

Data extraction and quality assessment
We composed a provisional checklist (see Table 2) to review the identified frameworks and tools and to detect shortcomings and gaps in knowledge of their methodological quality. To compose this checklist we contacted experts in psychometric research who had been previously involved in the development of a checklist for assessing the methodological quality of studies about measurement properties of health status measurement instruments. Two reviewers (JB and HvdH) independently rated the information that was provided for each framework and tool based on the checklist. Disagreements between the reviewers were discussed and resolved during a consensus meeting.

Description of the checklist dimensions and items

Content validity
When a new tool is constructed to measure a particular concept, such as the “productivity” or “effectiveness” of PCRN, it is important that its measurement areas and items capture all relevant and meaningful aspects of that concept. In psychometric research this is called “establishing content validity.” Content validity can be established by appropriate literature support, involvement of experts, and agreement among subject-matter experts about the relevance of the proposed areas and items to be measured. To judge content validity we determined whether the authors had reviewed relevant literature, involved experts in the development and item selection/reduction process, and/or reported outcomes of a content validity study, such as the level of agreement between subject-matter-experts.

We considered content validity as one of the most important aspects of methodological quality. Only if the content validity of a tool is adequate will one consider using the tool, and evaluation of the other aspects of methodological quality is useful.

Reliability
A prerequisite for the establishment of reliability is that the tool itself and the measurement methods are adequately described and standardized. Reliability can further be established by evaluating the degree
of agreement between different observers (inter-observer reliability) or the agreement between observations made by the same observer/rater on two different occasions (intra-observer reliability). To assess reliability we determined whether the tools and measurement methods were described in sufficient detail to permit replication. For the tools that were tested or applied, we determined whether the authors reported on inter-observer and/or intra-observer reliability.

**Feasibility**
If a new tool is considered to measure all meaningful aspects of the concept of interest, the feasibility of the tool has to be established. We therefore determined whether the tools were tested in a pilot study or had been otherwise applied, for instance for a (comparative) PCRN evaluation survey.

**Practice investment**
Because the usefulness of a tool is also determined by practical and financial aspects, we also evaluated the information provided concerning time investment and costs.

**Description of the tools**
Descriptive data extracted from the publications included the level of measurement (network/practice), the concept of measurement, and the content of the tools in terms of domains of interest/proposed measurement areas and the number of proposed items.

---

**Table 2. Checklist to evaluate methodological quality of the identified tools**

<table>
<thead>
<tr>
<th>Content validity</th>
<th>1. Do authors report that they reviewed relevant literature to identify previous research that have been done in the area and/or previous tools?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Do authors report the involvement of experts/peer groups in the development of the tool and/or item-selection/item-reduction process?</td>
</tr>
<tr>
<td></td>
<td>3. Do authors report the outcomes of the level of agreement among experts reflecting the content relevance of the proposed measurement areas/items?</td>
</tr>
<tr>
<td>Reliability</td>
<td>4. Do the authors describe the tool and measurement method in sufficient detail to permit replication?</td>
</tr>
<tr>
<td></td>
<td>5. Do authors report the results of intra and/or inter-reliability tests?</td>
</tr>
<tr>
<td>Feasibility</td>
<td>6. Do authors report that the acceptability and measurability of the measurement items and feasibility of the tool were tested in a pilot study?</td>
</tr>
<tr>
<td></td>
<td>7. If the authors did not perform a pilot study, has the tool been applied once otherwise?</td>
</tr>
<tr>
<td></td>
<td>8. Do authors report other (empirical) evidence that the measurement quality of the tool has been examined?</td>
</tr>
<tr>
<td>Practice investment</td>
<td>9. Do authors provide sufficient information to evaluate the time-investment?</td>
</tr>
<tr>
<td></td>
<td>10. Do authors provide information about the costs of an assessment/evaluation using this tool?</td>
</tr>
</tbody>
</table>

All questions were answered as yes or no.
Results

We performed two searches in Medline. The first search, which was conducted in July 2005, resulted in 631 hits. We selected 69 articles for full-text reading (62 unanimous, 7 based on consensus); five articles from additional searches were included. For the preliminary selection we included 12 articles (8 unanimous, 4 based on consensus) and for the final selection five articles (unanimous decision). An update of our search in Medline was conducted in May 2009, which resulted in one additional article for the final selection. From the six articles we included in the final selection, two described a tool kit for the evaluation of PCRN, developed by Fenton et al and one article concerned a framework for the evaluation of PCRN, developed by Clement et al. Furthermore, we identified two articles that described a tool, developed by Doorn et al, for measuring practice performance/output in academic general practice networks, and one article that described an accreditation scheme for primary care practices involved in research, developed by Carter et al.

Identified frameworks/tools for evaluating PCRN

1. In 2001, Clements et al developed a framework to evaluate the effectiveness of PCRN. They considered PCRN to be effective if they were capable of achieving predefined and agreed network objectives. They proposed seven relevant and agreed-upon objectives (see Table 3) for the evaluation of PCRN, and for each objective a set of example indicators for measurement; four objectives in the framework referred to the research function of PCRN, one to a better use of research findings, and two to network building (see Table 3).

2. In 2007, Fenton et al presented a tool kit to evaluate the organizational design, management, and appropriateness of PCRN. The tool kit was generated from the evaluation of five PCRN in the United Kingdom. Fenton et al identified eight organizational dimensions and 28 subdimensions (see Table 3) that they considered to be relevant for PCRN and their potential productivity (defined as "creating ideas and knowledge, or intellectual capital"). The focus of the tool kit is on the structures and processes within PCRN and not on measurement of output. In the evaluation study the (sub)dimensions were rated on a 5-point scale, varying from 0 = not present (the dimension is not applicable) to 4 = very strong (the dimension is very strongly present).

Identified frameworks/tools for practices Involved in PCRN or research

1. In 1999, Doorn et al developed a tool with which to measure the extent of academic activity and the quality of health care in general practices involved in a university-linked general practice network: Huisartsgeneeskundige Academiserings Lineaal Maastricht (HALMA). They identified three practice performance/output areas that they considered to be relevant for the effectiveness of PCRN, and proposed 70 expert-validated indicators for evaluating practice performances/output in networks (see Table 4). The tool includes indicators for measuring practice involvement in and contributions to research and registration, the quality of care provided by participating practices,
Evaluating primary care research networks: a review of currently available tools and their involvement in and contributions to academic medical teaching and education.

2. In 2002, the Royal College of General Practitioners (RCGP) in the United Kingdom introduced the Primary Care Research Team Assessment scheme (PCRTA). This scheme provides both primary care teams and stakeholders (such as funding bodies) with a mechanism for assessing the research infrastructure within a practice against professionally developed and tested standards. The scheme presents seven areas of interest for assessing the quality of the research infrastructure in practices involved in research (see Table 4) and distinguishes two levels of accreditation: collaborator research practices and investigator-led research practices. For both levels the scheme describes essential and desirable quality indicators. In addition to focusing on aspects such as strategic planning, research resources, and infrastructure, the scheme also addresses aspects such as education of other primary care professionals and the dissemination of research findings.

A further description of the identified tools, in terms of concept of interest, proposed dimensions, and number of items, is provided in Table 3 and 4.

**Methodological assessment of the identified frameworks/tools**

*Content validity*

All authors reported that they had reviewed PCRN literature to identify earlier research or theoretical concepts. However, both Clement and Fenton had already stated literature to support the development of methods for the evaluation of PCRNs, and their content validity is scarce. Clement’s literature search failed to identify any relevant studies focusing on the evaluation of PCRNs. The set of objectives in Clements’ framework was partially based on the findings of a national survey among 22 network coordinators in the United Kingdom. Harvey, Fenton, and Sturt performed a review of the organizational science literature on networks to identify those network dimensions that create social and intellectual capital for networks.

The involvement of professionals/stakeholders in the development of the tools was reported by Carter (PCRTA), Doorn (HALMA), and Fenton (tool kit). National key figures and stakeholders were involved in the PCRTA development and evaluation process. General practitioners who were participating in a Dutch academic general practice network validated the proposed HALMA indicators. For the latter, Doorn et al also published verifiable outcomes of the level of agreement for each item. Participants in research networks were involved in the development and refining of Fenton’s tool kit.

*Reliability*

For two tools only - the PCRTA (Carter) and the tool kit (Fenton) - the items and measurement methods were described in sufficient detail to permit replication, and these two tools met our requirements for standardization. Carter and colleagues reported that the reliability of the PCRTA assessment was (qualitatively) evaluated by an independent researcher after the pilot study. Outcomes of intra-reliability or inter-reliability tests were not reported.
### Table 3. Description of the identified tools for the evaluation of Primary Care Research Networks

<table>
<thead>
<tr>
<th>Name/Author</th>
<th>Concept of interest</th>
<th>Proposed domains of interest for PCRN evaluation</th>
<th>Nr. of indicators</th>
</tr>
</thead>
</table>
| Objective-based framework/ Clement et al. (2000) | PCRN effectiveness in terms of realising predefined and agreed objectives | **Predefined and agreed objectives for PCRNs:**  
1. To develop a network infrastructure  
2. To develop research capacity of primary care participants (PCPs)  
3. To increase the number/quality of research projects led by PCPs  
4. To increase the use of research findings by PCPs  
5. To increase the number/quality of research projects in which PCPs collaborate  
6. To increase the number/quality of research projects in which PCPs participate  
7. To provide a network that PCPs find acceptable | 45 |
| Evaluation tool kit/ Harvey et al. (2000)  | PCRN suitable for purpose and potential productivity in terms of producing knowledge and creating ideas and intellectual capital | **Organizational and management dimensions which create social and intellectual capital for PCRNs:**  
1. Strategic emphasis  
   a. winner focus capacity-building  
   b. practitioner capacity-building  
2. Policy  
   a. composition of executive  
   b. direction of decision-making  
3. Network structure  
   a. strength of centre  
   b. hierarchy (direction of power & resources)  
4. Research infrastructure  
   a. governance of research  
   b. mechanisms for identifying research needs  
   c. mechanisms for encouraging research ideas  
   d. route of research ideas into projects  
   e. mechanisms for supporting research  
   f. mechanisms for evaluating research  
5. Network processes  
   a. leadership style  
   b. cultural cohesion  
   c. trust relationships  
6. Process facilitation  
   a. ease of joining the network  
   b. mechanisms for enabling interrelationships  
   c. multidisciplinarity  
   d. mechanisms for publicizing the network  
   e. mechanisms for allowing feedback  
   f. ease of contacting support  
7. Network boundaries  
   a. relationships with partner organizations  
   b. relationships with other networks  
8. Self-evaluation by network  
   a. data-collection | 44 |
### Table 4. Description of the Identified Tools for Practices Involved in Primary Care Research Networks or Research

<table>
<thead>
<tr>
<th>Name/Author</th>
<th>Concept of interest</th>
<th>Proposed domains of interest for PCRNs evaluation</th>
<th>Nr. of indicators</th>
</tr>
</thead>
</table>
| Huisartsgeneeskundige Academiserings Lineaal Maastricht (HALMA)/Doorn et al. (1999) | Output of academic general practices in university-linked general practice networks | Practice performance/output (qualitative and quantitative) in:  
1. Research and registration/data-collection  
a. regular research activities  
b. research development  
c. registration/data-collection (for research database)  
2. Quality of health care  
3. Academic medical teaching/education  
a. education/training in practice  
b. faculty teaching/medical education | 70 |
| PCRTA/Carter et al. (2002)                                                 | Research infrastructure of (general) practices involved in research (two levels of accreditation) | Does research infrastructure in practice meet quality standards in:  
1. Practice organization:  
a. practice profile  
b. records and register  
2. Strategic planning  
a. practice research infrastructure  
3. Practice as a learning organization:  
a. individual development  
b. team development  
c. teaching others  
4. Research resources and infrastructure  
a. research resources  
b. computerization and data-handling  
c. links with other organisations  
5. Project funding and management  
a. project funding  
b. project management  
6. Involvement of patients  
a. consumer participation  
b. ethical issues  
7. Dissemination of research  
a. dissemination strategy | 78 |
Feasibility

Two tools were tested in a pilot study (PCRTA and HALMA), and one tool (Fenton’s tool kit) was used once for a comparative case study of five PCRNs in the United Kingdom. One tool (Clements’ framework) has neither been tested nor applied. Although Doorn et al performed a pilot study the information they provided (an unpublished article about the results of this feasibility study) was not sufficient to enable us to complete our assessment of the methodological quality of HALMA.

Practice investment

We could only evaluate time investment and costs for the PCRTA. A sample timetable for the assessment visit was presented whereas time investment in the preparation of documentary evidence for leading general practitioners and administrative support were evaluated for practices that participated in the pilot study. The costs for accreditation could be derived from the website of the RCGP in the United Kingdom. A summary of our qualitative assessment of the checklist items is provided in Table 5.

Discussion

We identified four tools for the evaluation of PCRN or measurement of primary care practices involved in PCRN or research. For evaluating the “overall” performance of PCRN we identified two tools: Clement’s objective-based framework and Fenton et al PCRN tool kit. The content validity of the proposed objectives and indicators in Clement’s objective-based tool seems to be questionable when set against the criteria of our checklist, and because the tool has not been tested or otherwise applied, the feasibility of the indicators is also questionable. However, incorporating the objectives of the network in the framework is a strength of Clement’s tool. Information about objectives is needed to identify appropriate dimensions and input and output indicators for measurement and evaluation. The second tool we identified for the evaluation of PCRN, Fenton’s tool kit, was generated from a contextualized case study of five PCRNs in the United Kingdom and subsequently refined. The data were collected at horizontal (time) and vertical levels (individual, group, organization) of analysis. Fenton et al also incorporated exogenous factors in the evaluation (like institutional and geographical data). This case study provides insight into the early stages of network development because Fenton generated a basic set of generic input indicators of network configuration and management. This set enables us to make meaningful comparisons between different network models from the perspective of congruency between objectives, organization, and management inputs. The underlying assumption is that if these three are well honed, productive outputs will follow. However, the data collection and the analyses seem to be rather time consuming. To evaluate the practice investment, more information on the costs and time investment is needed.

For the measurement of the development and performance of practices involved in PCRN or research we identified two tools: HALMA (Doorn) and the PCRTA (Carter). In addition to providing research-based indicators, both tools also provide a set of indicators to measure practice outputs in the education of professionals (PCRTA) and academic
Table 5. Qualitative assessment of information provided for the checklist items

<table>
<thead>
<tr>
<th>Framework/tool</th>
<th>Literature reviewed</th>
<th>Expert involvement</th>
<th>Outcomes of content validity (such as level of agreement)</th>
<th>Described in sufficient detail to permit replication?</th>
<th>Inter-reliability/ intrareliability</th>
<th>Pilot study</th>
<th>If not, was the tool otherwise applied</th>
<th>Other (empirical) evidence of measurement quality?</th>
<th>Time-investment</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jonathan</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Doorn et al.</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Objective-based</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Framework</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tool kit,</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Harvey et al. (2000)</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Carter et al. (2002)</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

1) Primary care research network level.
2) Practice level.
3) The Tool kit was used once in a comparative case study of 5 primary care research networks in the United Kingdom.
4) Carter et al reported that reliability and validity were qualitatively evaluated after the Primary Care Research Team Assessment (PCRTA) pilot study by an independent researcher, but reported no outcomes of inter-reliability or intrareliability.
medical education/teaching (HALMA).\textsuperscript{26,31} HALMA also provides a set of indicators to measure practice improvement and quality of care over time. For the PCRTA \textsuperscript{33} we were able to make the most complete qualitative assessment of methodologic quality. The information provided about the development process was sufficient to evaluate the content validity of the tool, the description of the scheme and the measurement methods met our criteria for standardization, a pilot study was performed and reported, and we could also evaluate the practice investment in terms of time investment and current costs of accreditation. The scheme provides a dataset of qualitative and quantitative indicators that can be used by PCRN\textregistered s to observe the (possible) effects of network participation on practice/professional development and research performance over time. It can also be used as a benchmark for the quality of the research infrastructure of participating practices. The RCGP in the United Kingdom has acknowledged the scheme and introduced it in 2001. However, despite the low costs of accreditation for primary care practices in England and Wales, the number of practices to date, that have gained accreditation as a collaborator practice (nine practices) or an investigator research practice (20 practices) is still relatively small.\textsuperscript{37} This may indicate that the accreditation process is more time consuming or difficult for practices than the results of the pilot study suggested, or, alternatively, that the benefits gained from accreditation are too small.

For HALMA we could only assess its content validity. Doorn et al \textsuperscript{31} did not provide sufficient information to make it possible to evaluate other aspects of its methodologic quality. We found the number of indicators for measuring the quality of care rather low, and most of these indicators differ from the indicators that are currently used to measure the quality of care (in The Netherlands). HALMA \textsuperscript{26,31} provides a basic set of expert-validated indicators for assessing the extent of research and academic medical teaching/education in PCRN\textregistered s. The Dutch PCRN\textregistered s explicitly aim to map research and teaching activity in everyday practice. However, for PCRN\textregistered s that do not have this focus, it might also be interesting to assess the involvement in education/teaching of practices/professionals involved in PCRN\textregistered s. An earlier study, conducted by Gray et al \textsuperscript{38}, showed that general practices in the United Kingdom with a leading role in research (as can be expected from practices involved in PCRN\textregistered s) are more involved in teaching than in teaching general practices involved in research. PCRN\textregistered s might well also prove to be an effective tool for increasing teaching capacity in primary care.

In the literature about PCRN\textregistered s there is an ongoing debate about the appropriate framework to measure the effectiveness and added value of PCRN\textregistered s. Evaluation involves assessing the success of an intervention against a set of indicators or a body of criteria based on the targets and underlying principles in relation to the initiative. Some of the targets and underlying principles of PCRN\textregistered s are reflected in the design of the identified frameworks and tools. However, most of the identified frameworks and tools were proposed and developed in a relative early state of network development, and ideas about the effectiveness, targets, and principles of PCRN\textregistered s may have evolved over time. We think it is necessary to again discuss the goals, such as desired impact and outcomes, of PCRN\textregistered s and the best ways to
achieve these goals in the future; and 2) reflect on the appropriateness, effectiveness, and added value of PCRNs. The questions are: Why should we invest in PCRNs? In what ways are PCRNs more effective and efficient than other research capacity building initiatives? Can PCRNs address certain issues that are relevant for the future of primary care better than through other approaches? Any framework or tools for internal or comparative evaluation should also consider this perspective.

Our goal was to provide an overview of the tools that are currently available for the evaluation of PCRNs and practices involved in PCRNs or research. We studied both the methodologic quality of the tools and their areas of interest. The results of our study show that various methods, areas of interest, structures, processes, output dimensions, and indicators for the evaluation of PCRNs have been proposed and identified. However, no generic and validated tool that enables meaningful comparison between different network models has yet been developed. In our opinion, Fenton’s tool kit and contextualized case study of PCRNs might provide a basis for a generic framework/tool. To open the “black box” of the effectiveness of PCRNs it may also be relevant to monitor and evaluate the development and performances of practices and professionals over time, as suggested by Doorn 26. The PCRTA and HALMA might provide a basis for a generic framework/tool to observe the effects of network and research participation on those involved in networks.

Strengths and limitations of our study
We conducted a very extensive search of several databases, contacted experts in the relevant fields, and also screened the so-called “grey literature.” We are confident that we did not miss any relevant frameworks or tools for the evaluation of PCRNs or practices involved in PCRNs or research. Regrettably, there is no accepted checklist that can be used to review the methodological quality of such complex tools. Therefore, we contacted experts in this domain to discuss relevant quality criteria to review the identified frameworks/tools and to detect shortcomings and gaps in knowledge of their methodologic quality. We have provided comprehensive information about all quality items so that readers can make their own assessment of the methodologic quality and state of development of the identified frameworks and tools. We acknowledge that our checklist might not be comprehensive and some of the quality criteria (like inter-reliability) might be arbitrary for this type of tools. Further discussion about the relevant quality criteria for evaluating the methodological quality of this type of tools is needed.

Acknowledgements
We like to thank our colleagues at the EMGO Institute and specifically Henrica C.W. de Vet and Sandra D.M. Bot for their time and support in the development of the provisional checklist we used to review the identified framework and tools.
References
