Development and implementation of a performance measure tool in an academic pediatric research network

Rachel Stanley a,⁎, Kathleen A. Lillis b, Sally Jo Zuspan c, Richard Lichenstein d, Richard M. Ruddy e, Michael J. Gerardi f, J. Michael Dean c and the Pediatric Emergency Care Applied Research Network (PECARN)

a Department of Emergency Medicine, University of Michigan, Ann Arbor MI, United States
b Department of Pediatrics, State University of New York at Buffalo School of Medicine and Biomedical Sciences, Buffalo NY, United States
c Department of Pediatrics, University of Utah, Salt Lake City, UT, United States
d Department of Pediatrics, University of Maryland, MD, United States
e Department of Pediatrics, Children’s Hospital of Cincinnati, Cincinnati, OH, United States
f Department of Emergency Medicine, Morristown Memorial Hospital, Morristown NJ, United States

Abstract

Background: The Pediatric Emergency Care Applied Research Network (PECARN) is a federally funded multi-center research network. To promote high quality research within the network, it is important to establish evaluation tools to measure performance of the research sites.

Purpose: To describe the collaborative development of a site performance measure tool “report card” in an academic pediatric research network. To display report card template information and discuss the successes and challenges of the report cards.

Development and implementation of the network performance measure tool: The PECARN Quality Assurance Subcommittee and the PECARN data center were responsible for the development and implementation of the report cards. Using a Balanced Scorecard format, four key metrics were identified to align with PECARN’s research goals. Performance indicators were defined for each of these metrics. After two years of development, the final report cards have been implemented annually since 2005. Protocol submission time to the Institutional Review Board (IRB) improved between 2005 and 2007. Mean overall report card scores for site report cards increased during this period with less variance between highest and lowest performing sites indicating overall improvement.

Conclusions: Report cards have helped PECARN sites and investigators focus on performance improvement and may have contributed to improved operations and efficiencies within the network.

1. Background

Measuring site performance in clinical research studies remains a challenge especially for organizations conducting large multi-center trials. The pharmaceutical industry often measures site performance based on frequent, comprehensive site monitoring visits. Academic research networks need similarly effective methods to identify high and low performing sites, but intensive site monitoring is usually not financially feasible.

The challenges of coordinating and monitoring research networks have been described in several publications [1–10]. Youngblood et al. developed performance measures for a phase 3 trial and concluded that setting objective performance measures made it easier to monitor performance goals on a site-by-site and trial-wide basis [9]. Most prior publications

⁎ Corresponding author. Tel.: +1 734 9361724.
E-mail address: stanleyr@umich.edu (R. Stanley).
conclude that monitoring performance goals is an ongoing process that serves to maintain the integrity of a study, to identify local problems and to maintain study cohesiveness and motivation. However, although monitoring performance is important for academic research networks, most publications to date do not give a description of the development or implementation of performance measures [7,9,10].

Academic research networks vary considerably in focus, size, structure and level of funding and therefore standardized performance measures may be of limited value given network diversity [7,9,10]. However academic research networks need to develop and implement performance measures to ensure accountability and sustainability [10].

In this paper we sought to give a detailed description of the process of development and implementation of performance measures in an academic research network.

2. Purpose
The purpose of this paper is to:

A. Describe the collaborative development of a performance measure tool, called a report card, in an academic multi-center pediatric research network (PECARN).
B. Display report card template information and describe aggregate results from the first three years of implementation.
C. Discuss the successes and challenges of implementing a network report card.

3. Description of the research network

The Pediatric Emergency Care Applied Research Network (PECARN) consists of 22 Emergency Department (ED) sites across the US and a data coordinating center [11,12]. The goal of the network is to conduct high priority, high quality multi-center research for the prevention and management of acute illnesses and injuries in children.

The network is organized into four research nodes. Each node has a lead site and four to six affiliated Hospital EDs, some of which are in large academic children’s hospitals, and others that are in general hospitals (Fig. 1). Each node has a cooperative agreement with the federal funding agency and the funding is given to the lead site. Guidance from the federal funding agency suggests that each site be provided with support for a lead site investigator and a full time research coordinator to provide the research infrastructure. Funds for individual research projects are sought on a competitive basis.

PECARN has a data coordinating center that is separately funded and provides training, study materials, site support, data management and statistical analysis. PECARN is currently conducting or has recently completed more than a dozen investigator initiated studies including two large randomized controlled trials [13–15]. PECARN conducts diverse research that includes retrospective studies, prospective observational trials and randomized controlled trials. The network ED sites care for approximately 900,000 acutely ill and injured children per year.

![PECARN Organizational Structure](image)

**Fig. 1.** Diagram of Pediatric Emergency Applied Research Network structure.
4. Need for a network performance measure tool

Early experience with implementation of PECARN multicenter research projects demonstrated the need for a method to measure both network and individual site performance in carrying out multiple network studies. Differences in the quality of performance across sites were evident from the inception of the network; time from availability to submission of the initial protocol to the site's Institutional Review Board (IRB) for approval varied by several months among sites. Initial enrollment numbers met projections at some sites but lagged at others, and based on site monitoring, adherence to and understanding of regulatory requirements varied significantly.

Although some site monitoring is employed in PECARN studies, frequent, comprehensive monitoring visits are not feasible. PECARN lacked a consistent way to recognize high performing sites, and to identify low performing sites that could benefit from a focused intervention. Furthermore, it was important to distinguish whether site-specific structural differences (such as local IRB policies and practices) accounted for the variation in site performance, or if site investigator and research coordinator job performance explained variation. The latter would be more remediable than the former. To address these issues, we designed and implemented an annual performance measure tool called the PECARN report card to assess overall site performance across multiple network studies. The report card has enabled us to identify sites’ performance levels with the ultimate goal of improving the quality of research within the network.

5. Development and implementation of the network performance measure tool

Based on guidance from our federal funding agency, the PECARN Steering Committee charged the Quality Assurance (QA) Subcommittee to develop a performance measure tool that would systematically evaluate all PECARN clinical sites and the data coordinating center. The QA Subcommittee is composed of investigators and coordinators from PECARN clinical sites, the data coordinating center, and staff from the federal funding agency. The subcommittee met in person three times per year for two years with several conference calls between meetings to develop the assessment tool. After each meeting, a summary was presented to the PECARN Steering Committee for input and review of key metrics and site measures. After final Steering Committee approval, the PECARN report cards were implemented in 2005.

5.1. Evaluative tool development

The QA Subcommittee selected the Balanced Scorecard as the evaluative tool [16]. The Balanced Scorecard is a performance management and measurement tool that translates organizational goals into performance measures and monitors the organization's success in achieving these measures. This measurement tool is used in the business world and has been adopted by some health care institutions [16–18].

5.2. Key metrics

Before developing the performance measures, it was important to define key metrics. Key metrics are specific aspects of the clinical site’s performance that were identified as important in achieving PECARN goals. The final key metrics that were selected were considered to be important for daily operations, sustainability, and the strategic growth of PECARN. Four key metrics were selected for the report card: general contributions by sites, intellectual development, project implementation, and project completion. Definable, measurable and customizable performance measures corresponding to each of these key metrics were developed.

5.2.1. General contributions by sites

This metric evaluates activities related to supporting network infrastructure including participating in meetings and fulfilling responsibilities of tasks or positions.

5.2.2. Intellectual development

This metric includes development of original research concepts, refinement of scientific protocols accepted for implementation, and preparation of grant applications to support the costs of protocol implementation. This metric is applied to each of the four research nodes rather than individual sites, because intellectual development requires involvement from multiple sites and investigators. This metric is of critical importance to PECARN’s sustainability, which depends on continuous scientific proposal generation, refinement and implementation.

5.2.3. Project implementation

This metric includes the implementation of PECARN research projects, including timeliness of IRB submission, adherence to Good Clinical Practices (GCP) and other research regulations, and submission of high quality data.

5.2.4. Project completion

This metric assesses activities associated with timely completion of PECARN research projects.

5.3. Clinical site (Hospital ED) performance measures

The QA Subcommittee defined specific performance measures for each of the key metrics. Each performance measure was assigned a percentage score based on its overall importance. For feasibility purposes, it was important to create measures that could be reasonably tracked by the data coordinating center. Objective quantitative measures were defined when possible, but at least one overall subjective qualitative measure was included to describe performance in each metric. Table 1 shows the key metrics, corresponding performance measures, scoring system and whether quantitative or qualitative data were used for clinical site assessment.

5.3.1. General contribution measures

General contributions were measured by a site’s ability to participate in appropriate network studies given their size, resources, and patient population. Meeting attendance,
subcommittee participation, and timely voting were also assessed for all investigators.

5.3.2. Project implementation measures

Project implementation performance was measured by timely IRB submission and approval, evidence of GCP compliance, timeliness and accuracy of data entry, response to site monitors, and response to data center requests.

5.3.3. Project completion measures

Enrollment success was the most important measure of project completion. In order to account for site differences in ED volume and available subjects, enrollment success was measured based on the percentage of available patients enrolled at each site for individual studies. For example, in a randomized controlled trial of dexamethasone in infants with bronchiolitis [16], PECARN registry data were used to establish the number of infants with bronchiolitis seen annually in each site ED. This generated a target enrollment number against which actual enrollment was compared and measured for the report card. In a network observational study, the percentage of eligible patients enrolled was used as a measure of enrollment success.

5.4. Research node performance measures

The QA Subcommittee considered it important to evaluate individual nodal contribution to the network in addition to evaluating individual sites since each node administers its affiliated sites independently. All four of the key metrics were used to evaluate the nodes. Table 2 shows the performance measures, scoring system and whether qualitative or quantitative data were used for research node assessment.

5.4.1. General contributions measures

General contributions were measured by participation in Steering Committee meetings, timely electronic voting and subcommittee contributions by each node.

5.4.2. Intellectual contribution measures

Nodal intellectual contributions were measured by the number of original research concepts submitted to the Steering Committee and successful protocol development, as well as the number of grant proposals submitted and successfully funded.

5.4.3. Project implementation measures

Nodal project implementation was measured as the average of the sites’ scores for GCP compliance and completeness of study data submitted in each study. These measures assess the node’s effectiveness in assuring regulatory compliance and scientific integrity at its respective sites.

5.4.4. Project completion measures

Nodal project completion was measured by aggregate enrollment success across all sites and manuscript completion by lead investigators for each study. These measures assess nodal leadership during study implementation and publication.

5.5. Timing of evaluations

The report cards are produced annually. The timing was originally planned to coincide with the end of the network funding cycle so report card results and feedback to sites and the network could be incorporated to improve the next funding cycle application.

5.6. Implementation of the network performance measure tool

The Steering Committee gave the QA Subcommittee responsibility for the oversight of the report cards. There
were separate report cards developed: one for each Hospital ED, and one for each of the research nodes. A detailed manual of operations (MOO) was developed to explicitly define each measure, to describe how each item was measured and scored and the person responsible for collecting the information. Scores for each performance measure were contributed by site, nodal, or data coordinating center investigators and research staff, depending on the specific data requirements. The data coordinating center was primarily responsible for collating the scores and distributing the report card. Report cards were prepared annually for presentation and distribution at the Steering Committee meeting. A blinded summary report was distributed to the entire committee. Site investigators received their individual detailed report card. The report also provided a blinded ranking so that investigators could compare themselves to other sites based on their report card score. Annually,

<table>
<thead>
<tr>
<th>Key metric</th>
<th>Research node performance measure corresponding to key metric</th>
<th>% score</th>
<th>Quantitative or qualitative measure</th>
<th>Description of scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>General contributions</td>
<td>Participation in PECARN meetings</td>
<td>7.5</td>
<td>Quantitative</td>
<td>A roster is kept of each meeting and each node should have full representation from sites and nodal leadership at each meeting. Electronic voting occurs and nodes are scored based on full representation from their nodes on each ballot. Each node should have 2 members present at each subcommittee meeting. PI of each study grades leadership from each node and nodal participation in each project.</td>
</tr>
<tr>
<td>General contributions</td>
<td>Voting participation</td>
<td>7.5</td>
<td>Quantitative</td>
<td></td>
</tr>
<tr>
<td>General contributions</td>
<td>Subcommittee contribution</td>
<td>7.5</td>
<td>Quantitative</td>
<td></td>
</tr>
<tr>
<td>General contributions</td>
<td>Leadership and participation by node and its nodal leader for each project (to be graded by the study PI)</td>
<td>10</td>
<td>Qualitative</td>
<td>PI of each study grades leadership from each node and nodal participation in each project</td>
</tr>
<tr>
<td>Intellectual development</td>
<td>Grant submissions</td>
<td>20</td>
<td>Quantitative</td>
<td></td>
</tr>
<tr>
<td>Intellectual development</td>
<td>Original concepts submitted to Steering Committee and successful protocol development</td>
<td>15</td>
<td>Quantitative</td>
<td></td>
</tr>
<tr>
<td>Intellectual development</td>
<td>Author participation and timely manuscript completion</td>
<td>10</td>
<td>Qualitative</td>
<td>For each manuscript timeliness of manuscript completion and author participation from each node is averaged. Count of number of concepts submitted to Steering Committee and approved which are developed into protocols.</td>
</tr>
<tr>
<td>Project implementation</td>
<td>GCP compliance, document completeness (aggregate score for all sites)</td>
<td>7.5</td>
<td>Quantitative</td>
<td>Study monitor evaluates GCP during study site monitoring visits using data center template. Score is averaged for all sites within a node. Varies per study but usually a count of % eligible per study averaged annually over all studies and all sites in a node.</td>
</tr>
<tr>
<td>Project completion</td>
<td>Enrollment success</td>
<td>15</td>
<td>Quantitative and qualitative</td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Research node performance measure report card.

Fig. 2. Site scores 2005.
the QA Subcommittee reviewed the criteria in the report card for the Steering Committee, and as the network matured it became important to “tweak” or customize the report card from year to year. This customization was detailed in the manual of operations so sites were aware of the changes in measures/criteria and how they were scored each year. Although this customization was important to meet the needs of the network, it made it difficult to compare report card results for each customized measure between years.

Twenty-five different centers were evaluated using the report cards over the three year period. After year 1, some sites were eliminated from the network based on performance and other issues and new sites were added. The overall mean site score was 75.8%, 83.8% and 83.0% in years one through three, respectively. Figs. 2–4 show the distribution of scores throughout the network for each year. These figures show that the difference between the best and worst performing sites decreased in subsequent years, reflecting general improvement of performance across the network.

Timely submission of the protocol to the local IRB is a key factor in getting multiple sites prepared to begin a clinical study. The data center tracked sites’ IRB submission times before and after the report card, offering an objective measure of the impact of the report card. IRB submission time was significantly improved after implementing the report card. The time interval was measured from the date the protocol was submitted to the local IRB to the date the protocol was approved.

![2006 Site Scores](image)

**2006 Site Scores**

Average 83.8%

Distance from Average

*Fig. 3. Site scores 2006.*

![2007 Site Scores](image)

**2007 Site Scores**

Average 83%

Distance from Average

*Fig. 4. Site scores 2007.*
made available to sites to the date the site submitted the final protocol to their local IRB. In theory, this time interval could be as short as one day. In 2004, sites were asked to submit a new network protocol to their IRBs for review, but no deadline was given. Only one site submitted in less than 3 weeks. In 2007, 3 new projects required protocol submission to the sites’ IRBs. Sites that were participating were given a 3 week deadline to submit their IRB application and were informed by the data coordinating center that this would be part of the annual report card score. Ninety-five percent of the sites submitted each of these protocols to their IRBs within the required 21 days. The average time interval for IRB submission decreased from 5.7 weeks (early 2005) to 3.2 weeks (2006).

6. Discussion

The development and implementation of report cards may have improved performance in the PECARN. The process helped establish and maintain high research standards and motivated sites to improve performance. The investment of time and effort was valuable, and the result was a set of common goals that all sites could work toward. It is impossible to assess whether the results from the report card could have been predicted without its use however the report card helped provide a more objective and impersonal evaluation of a site than could otherwise have occurred.

A recent publication “Evaluating Emergency Care Research Networks: What are the Right Metrics?” made consensus based recommendations that metrics for EC research networks must be measurable, explicitly defined and customizable [10]. These qualities were important to our network in the initial development of the performance measures. As we implemented the report card it became very clear that each metric needed to be explicitly defined in greater detail than initially anticipated. Therefore the manual of operations was revised several times to reflect these details. As the network matured and the types of research studies changed it became important to customize or “tweak” the report card from year to year. It was important to describe this customization in the manual of operations so sites were aware of the changes in measures and how they were defined and scored from year to year.

When we set out to define measures we purposefully included some subjective qualitative measures but our experience was that despite varying performances, most sites and nodes scored full marks in these areas. This reinforced the need for objective quantitative measures within our network. The qualitative or quantitative scoring of each metric is detailed in Tables 1 and 2.

Enrollment success is likely one of the more important site measures in a research network. Initial guidance for the network membership encouraged regional representation with diverse emergency departments; consequently PECARN is comprised of sites with a wide range in the number of pediatric ED admissions. For the report card, enrollment success thus had to be evaluated based on the sites potential for enrollment rather than actual patients enrolled. Estimating these numbers tended to be difficult and thus enrollment success did not play as large a role in the report card in the initial years of the network. Therefore, for reasons related to varying size of institutions, difficulty establishing initial enrollment targets, seasonality and variability of disease presentation, and customization of this measure, we did not have adequate data to show an impact of the report card on enrollment. In addition, retrospective studies that we were doing as foundation for later work did not contribute much to an enrollment measure. Therefore, in the early years of the network, we scored items that established the quality of the network function such as “IRB submission time” and placed less emphasis on enrollment numbers. Enrollment success is much more of a focus for us now that the network has been established and we have several randomized controlled trials starting or in progress.

The development of the report cards served two important purposes for the PECARN. Defining the report card measures allowed the network subcommittee members to define the site and node performance measures on which their sites and nodes would be judged. This process was important in gaining consensus across the network on what defined successful performance. Second, formal distribution of the report card in an annual network meeting focused attention on the importance of the report card and set the bar for what defined a successful site. We believe this method of distribution communicated the importance of the report card and may have motivated sites to improve their performance in the next year.

The implementation of the PECARN report cards showed an early impact on site performance in terms of IRB submission. Site investigators and research coordinators rapidly became more responsive to study deadlines after becoming aware that the delays would be reflected in their report card scores. Sites that had previously had significant delays in submitting to their IRB were able to submit within the deadlines. The report card also identified low performing sites so that early intervention with additional training could be provided. Site scores were provided annually to nodal leadership to help them identify and address poorly performing areas for their sites. For example, if Good Clinical Practices (GCP) was an area of poor performance for a site, nodal site monitors concentrated on educating the site about GCP. If “timely data entry” was consistently low for a site then site and nodal leadership worked together to figure out ways to improve this metric at that site.

There was no report card overall score or bar set below which a site would no longer participate in the network. The decision to discontinue a site’s membership within the network was made by nodal leadership and was based on many factors which included the report card score. The report cards were developed in the spirit of a performance improvement tool and not as a punitive tool.

There were several challenges that were encountered during the report card development process. The report card development process required many meetings and discussions and took two years. Gaining consensus on metrics and measures was challenging due to the different perspectives of the committee members. Determining the details of exactly how elements would be scored was complicated by the size and volume differences among sites as well as by the diverse types of studies conducted by the network. For example, retrospective chart abstraction studies and registry submissions might be evaluated by data submission times, whereas randomized controlled trials and observational studies could evaluate enrollment and regulatory compliance.
The report card components are intended to reflect the combined performance of investigators and research coordinators at each site; unfortunately, the report cards have not allowed separate assessment of investigator and coordinator performance. Although we did not specifically investigate the perceptions of the Site Investigators and Research Coordinators towards the report cards, we believe that the report cards led them to prioritize their workload and focus on activities that would improve the scientific integrity of each research project.

6.1. Limitations of the report card

Certain performance measures were customized from year to year and this may have affected comparisons of overall scores between the three years, this also made it difficult to compare customized measures between years. However, the report card is a living document which reflects an evolving network and revisions are necessary from year to year to reflect different study stages and types and the changing composition of the network. While the specific performance criteria were not publicized in the first year, sites became aware of the criteria as soon as the first report card was published. This makes it possible for sites to improve performance in specific areas that would result in optimal report card scores. This could result in an improved score without an improvement in overall site performance. Finally, since the report card was developed in the early stages of the network, it is possible that some of the improvements noted were due to the learning curve of participation in the network rather than the report card.

7. Conclusions

Collaborative development and implementation of report card performance measures helped PECARN focus on performance improvement and may have contributed to improved operations and efficiencies within the network. Lessons learned from the development and implementation of the PECARN performance measures may be useful in other research settings.

Acknowledgments

The Pediatric Emergency Care Applied Research Network (PECARN) is supported by cooperative agreements U03MC00001, U03MC00003, U03MC00006, U03MC00007, and U03MC00008 from the Emergency Medical Services for Children (EMSC) program of the Maternal and Child Health Bureau, Health Resources and Services Administration, US Department of Health and Human Services.

Participating centers and site investigators are listed below in alphabetical order: Atlantic Health System/Morristown Memorial Hospital ( ); Bellevue Hospital Center ( ); Calvert Memorial Hospital ( ); Children’s Hospital of Buffalo ( ); Children’s Hospital of Michigan ( ); Children’s Hospital of New York — Presbyterian ( ); Children’s Hospital of Philadelphia ( ); Children’s National Medical Center ( ); Cincinnati Children’s Hospital Medical Center ( ); DeVos Children’s Hospital ( ); Harlem Hospital Center ( ); Holy Cross Hospital ( ); Howard County Medical Center ( ); Hurley Medical Center ( ); Johns Hopkins Medical Center ( ); Marquette General Hospital ( ); Medical College of Wisconsin/Children’s Hospital of Wisconsin: Newark/Beth Israel Medical Center; University of California Davis Medical Center; University of Michigan; University of Rochester; University of Utah/Primary Children’s Medical Center; Upstate Medical Center; Washington University/St. Louis Children’s Hospital; University of Maryland.


References


