

# UCLA

## UCLA Previously Published Works

### Title

Establishing a global quality of care benchmark report

### Permalink

<https://escholarship.org/uc/item/690500t9>

### Journal

Health Informatics Journal, 27(2)

### ISSN

1460-4582

### Authors

Sampurno, Fanny

Cally, Justin

Opie, Jacinta L

et al.

### Publication Date

2021-04-01

### DOI

10.1177/14604582211015704

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed



Published in final edited form as:

*Health Informatics J.* 2021 ; 27(2): 14604582211015704. doi:10.1177/14604582211015704.

## Establishing a global quality of care benchmark report

**Fanny Sampurno<sup>1</sup>, Justin Cally<sup>1</sup>, Jacinta Opie<sup>1</sup>, Ashwini Kannan<sup>1</sup>, Jeremy L Millar<sup>2</sup>, Antonio Finelli<sup>3</sup>, Andrew Vickers<sup>4</sup>, Caroline M Moore<sup>5</sup>, Christoph Kowalski<sup>6</sup>, Claire Foster<sup>7</sup>, Dan A. Barocas<sup>8</sup>, David Galvin<sup>9</sup>, Jean-Paul Van Basten<sup>10</sup>, John L. Gore<sup>11</sup>, Julia Ferencz<sup>12</sup>, Keith A. Lawson<sup>3</sup>, Khurshid R. Ghani<sup>13</sup>, Lorna Herbert<sup>14</sup>, Olli Saarela<sup>15</sup>, Sarah E Connor<sup>14</sup>, Sebastian Dieng<sup>12</sup>, Susan Linsell<sup>13</sup>, Timo Soeterik<sup>16</sup>, Paul Villanti<sup>17</sup>, Mark S Litwin<sup>14</sup>, Sue M Evans<sup>18</sup>**

<sup>1</sup> Department of Epidemiology and Preventive Medicine Monash University, Australia

<sup>2</sup> William Buckland Radiotherapy Centre, Alfred Hospital, Melbourne, Victoria, Australia

<sup>3</sup> Surgery, Princess Margaret Cancer Centre, Toronto, Ontario, Canada

<sup>4</sup> Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, USA

<sup>5</sup> Department of Urology, Division of Surgical and Interventional Science, University College London, London, UK

<sup>6</sup> Certification, German Cancer Society - Krebsgesellschaft, Berlin, Germany

<sup>7</sup> School of Health Sciences, University of Southampton, Southampton, UK

<sup>8</sup> Department of Urology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

<sup>9</sup> Department of Urology, Institution University College Dublin, Dublin, Ireland

<sup>10</sup> Department of Urology, Canisius Wilhelmina Hospital, Santeon Group, Nijmegen, The Netherlands

<sup>11</sup> Department of Urology, University of Washington, Seattle, Washington, USA

<sup>12</sup> Data Management, OnkoZert, Neu-Ulm, Germany

<sup>13</sup> Department of Urology, University of Michigan (on behalf of MUSIC), Ann Arbor, Michigan, USA

<sup>14</sup> Department of Urology, David Geffen School of Medicine, Department of Health Policy & Management, Fielding School of Public Health, University of California, Los Angeles, USA

<sup>15</sup> Dalla Lana School of Public Health, University of Toronto, Ontario, Canada

<sup>16</sup> Value-Based Healthcare, Santeon Group, Utrecht, The Netherlands

<sup>17</sup> Movember Foundation, East Melbourne, Victoria, Australia

<sup>18</sup> Cancer Council Victoria, Melbourne, Australia

---

**Corresponding details:** Fanny Sampurno, fanny.sampurno@monash.edu.

## Abstract

**Background**—The Movember funded TrueNTH Global Registry (TNGR) aims to improve care by collecting and analysing a consistent dataset to identify variation in disease management, benchmark care delivery in accordance with best practice guidelines, and provide this information to those in a position to enact change. We discuss considerations of designing and implementing a quality of care report for TNGR.

**Methods**—Eleven working group sessions were held prior to and as reports were being built with representation from clinicians, data managers and investigators contributing to TNGR. The aim of the meetings was to understand current data display approaches, share literature review findings and ideas for innovative approaches. Preferred displays were evaluated with two surveys (survey 1: 5 clinicians & 5 non-clinicians, 83% response rate; survey 2: 17 clinicians & 18 non-clinicians, 93% response rate).

**Results**—Consensus on dashboard design and three data-display preferences were achieved. The dashboard comprised two performance summary charts; one summarising site's relative quality indicator (QI) performance and another to summarise data quality.

Binary outcome QIs were presented as funnel plots. Patient-reported outcome measures of function score and the extent to which men were bothered by their symptoms were presented in bubble plots. Time series graphs were seen as providing important information to supplement funnel and bubble plots.

RMarkdown was selected as the software program principally because of its excellent analytic and graph display capacity, open source licensing model and the large global community sharing program code enhancements.

**Conclusions**—International collaboration in creating and maintaining clinical quality registries has allowed benchmarking of process and outcome measures on a large scale. A registry report system was developed with stakeholder engagement to produce dynamic reports that provide user-specific feedback to 132 participating sites across 13 countries.

## Keywords

International Registry; Benchmarking; Data Display; Data Science; Quality Indicators (Health Care); R Markdown

---

## BACKGROUND

A wide range of provider-specific quality of care indicators are reported by clinical registries to understand whether optimal care is being provided and to identify where gaps in practice and service provision occur (1). While clinical quality registries are increasingly viewed as important for quality assurance, the impact of performance feedback into practice hinges on the design and interpretability of the report as well as how the quality indicators are subsequently used in practice.

The TrueNTH Global Registry (TNGR) is a Movember funded, international registry established to understand variation and improve quality of care for men diagnosed with

localised prostate cancer (2). Movember is a global philanthropic organisation committed to improving men's health. Currently 132 participating sites across 13 countries contribute data to the registry. Data from local data centers (LDC) are transmitted each six months to the Data Coordination Centre located at Monash University, Australia. The governance structure of the TNGR, the role and responsibility of LDCs and participating sites have been previously reported (3). A set of 33 evidence and consensus-based quality indicators has been selected through a Delphi process (4).

Currently, no universal design tool or guidelines exist for how quality of care should be benchmarked and reported. It remains undetermined the extent to which statistical principles should be applied to the performance measures and how easy it is for recipients of the report who may have limited statistical literacy to decipher them. Working towards optimizing the balance between statistically sound data presentation and simplicity is the key. In an international registry, such as TNGR, diversity in languages is another important consideration. Information which is portrayed using figures is arguably more easily comprehended compared to a narrative description (5).

Here we discuss considerations of designing and implementing benchmark reports for an international clinical quality registry. During development, we sought to determine the most effective visual presentation of data quality, the quality indicators and a suitable reporting platform to distribute confidential reports to international contributors.

## METHODS

### Literature review

A scoping review was carried out to determine the various methods of visually displaying data in the field of health (Ovid Medline and EMBASE) and economics (EconLit). Search strategy is outlined in Supplementary file 1. In total, the review identified 320 articles on Medline, 323 articles on EMBASE and 1843 articles on EconLit. Three articles were particularly informative in outlining different ways of presenting data (6–8). Two overarching concepts for presenting data were identified, a) data at a single time point and b) data demonstrating change over time. We further categorised data displays into those which reported one or more than one outcomes on the same display.

### Working groups

**Primary working group**—A working group was convened to understand current data display approaches, share findings of the literature review and source ideas for new innovative data presentation approaches. Working group participants comprised members of the TNGR Executive Committee and representatives nominated by the Executive Committee (n = 12). The primary working group identified the purpose of the reports, viz:

- To provide participating LDCs/sites with feedback on their performance which is easy to interpret (read) and informative (clinically and statistically meaningful);
- To allow them to compare their results with other contributors;

- To produce visual presentations that facilitate a cycle of continuous improvement;
- To display early warning signal if outcomes begin to deteriorate;
- To maintain confidentiality of all participating LDCs/sites.

**Expanded working group**—As the project developed, the primary working group was expanded to gather a wider range of views on data presentation from TNGR participants. Local data centres were asked to nominate at least one working group participant to participate in the project. The skills of participants were diverse, with representation from urology, radiation oncology, hospital administration, epidemiology, statistics/biostatistics, psychology, data science, registry managers and the funder (Table 1).

## Surveys

Two surveys were conducted. The first survey was delivered to the primary working group members (Table 1) following a videoconference to obtain suggestions to supplement the literature review (Supplementary file 2a). Following the first survey, feedback from the working group was summarised into favourable and unfavourable factors of each design alternative. A subsequent survey was developed and distributed to the 38 members of the expanded working group following a workshop, described below (Supplementary file 2b).

## Workshop

A three-hour workshop was held with the expanded working group members. Participants were provided with (1) the literature review, (2) results of the initial survey and (3) the summary document of favourable and unfavourable attributes of each design developed by the primary working group. Workshop participants were asked to ameliorate the display or provide alternative methods of presenting data if they thought it could be improved.

## Survey design

The survey was designed to identify the best approach to displaying the following four types of data (Figure 1):

- Indicators with binary outcomes at a single (static) time point (requirement 1). The goal of the visualization was to provide a broad overview of how a site performed against an indicator associated with quality of care, and enable sites to compare their result with others in the specified reporting period. For this, participants were asked to preference simple bar charts, waterfall charts and funnel plots.
- Indicators with more than one outcome of interest at a static time point (requirement 2). The goal of the visualization was to provide a broad overview of how a site performed against an indicator with multi-dimensional outcomes, and enable sites to compare their result with others in the specified reporting period. For this, participants were asked to preference scatter-plot, eclipse chart, bubble chart and radar chart.

- Indicators showing change over time (requirements 3 and 4). The goal of visualization was to evaluate if the site's performance is improving/ deteriorating over particular reporting periods. For this, participants were asked to preference box-plots, line-plots time-series, sequential probability ratio test (SPRT) and dual y-axes time-series.
- Other questions in the survey included display options for a dashboard as a summary of performance (requirement 5). The goal of the dashboard was to provide an overview, 'one-page summary' of how the site performed across all indicators associated with quality clinical care. This included comparing the site's result in the current reporting period against previous reporting period and how the site performed in comparison with others. For this, participants were asked to preference traffic light and heat map.

Reports were adapted to cater to TNGR stakeholder requirements from reporting at a local level to a country wide level (LDC-level, a participating site-level or a country level).

### **Consultation sessions to develop reports**

Survey results were used to assist in developing a template quality of care report and inform the consultation process. In total eleven consultation sessions were undertaken with members of the expanded working group to further inform the final report. Participants of both workshops and subsequent consultation sessions are outlined in Supplementary file 3b. Consultation meetings were held after core sections of the report were developed. Each consultation meeting was held over one hour via videoconference. The developing work was provided in advance of each meeting and members were asked to either provide feedback in advance of, or during the meeting. Core questions addressed at these meetings related to the interpretability of the data displays, layout of the report and the text used in the report.

### **Statistical analysis**

Two methods were used to analyse the survey results. For survey questions which the participants had to select one preference among all the given options, percentages of respondents' selected individual graphical display were calculated. For the survey questions which the participants were asked to compare graphical displays to each other and rank them in order of preference, the average score was calculated by applying largest weight to the respondent's most preferred display and the least weight to their least preferred display.

### **Ethics approval**

Development of quality indicators is a component of the TrueNTH Global Prostate Cancer Outcomes Registry approved by the Alfred Health Human Research Ethics Committee (HREC/16/Alfred/98) Protocol version 2.0 (7 Sept 2017).

## **RESULTS**

Survey 1 achieved an 83% response rate (5 clinicians/ 5 non-clinicians) and survey 2 achieved a 93% response rate (17 clinicians/ 18 non-clinicians).

## Types and Frequency of Reporting

Clinician and non-clinician views differed regarding the health system level to which reporting should be delivered. While most non-clinicians preferred reporting at participating site, LDC and country level, clinicians did not feel country-level reporting was necessary, as some countries only comprised of one LDC and thus would misrepresent the data. There was consensus that to avoid reporting bias, the quality indicator report should provide information on the completeness of data used to generate the indicators.

Most working group participants indicated a preference for registry contributors to receive reports every six months to align with the TNGR data submission period and that time series reporting should include data aggregated over the 3-year reporting period.

## Report output

**Dashboard**—Consensus was reached on the creation of a dashboard to provide an overview of each of the 33 quality indicators and completeness of the reporting of data elements used in the indicators. A performance summary chart was used in the dashboard to provide a visual summary of the site's performance and how the performance in the current compared with that in the previous report.

Each quality indicator in the performance summary chart demonstrates the unit of measure (participating site/LDC) in relation to the population minimum, 1<sup>st</sup> quartile, median (2<sup>nd</sup> quartile), 3<sup>rd</sup> quartile, maximum and an aspirational target (determined using the *Aspirational Benchmarks of Care* approach described by Kiefe et al (9)). Figure 2 shows the “How to Interpret the Performance Summary Chart” section of the report where readers are given guiding information on how to read and interpret the plots that appear in the report.

### **Static reports displaying performance of quality indicator as binary outcomes**

—Funnel plots with control limits at 95% and 99.8% were preferred above other charts such as a bar chart, a line chart or box and whisker plot to display quality indicators with binary outcomes (Figure 3).

### **Static reports displaying performance of quality indicator for multiple outcomes**

—A bubble plot was preferred by the working group to display the multiple outcomes of interest in one chart. The bubble plot displayed the sexual, urinary and bowel domains individually. The function scores on one axis and the extent to which men were bothered by their symptoms on the other axis. The size of the bubble corresponded to the volume of survey responses received by the unit contributor, relative to the overall population. The plot was divided into four quadrants, where contributors who performed above their peers' average (better domain function score and lesser number of patients who reported moderate/big problem) were positioned in the right upper quadrant (Figure 4). Some of the respondents expressed concern that radar charts (10) would not be easily grasped by the registry audience. The use of confidence ellipses was also not highly preferred.

**Reports displaying change in performance of quality indicator over time—**

The purpose of a clinical quality registry program is to provide a platform for healthcare providers to monitor their performance over time whilst allowing them to compare their outcomes with others, or to external benchmarks. Viewing how performance measures change over time can reveal a story that cannot be seen in a static chart.

To accommodate this need, time-series figures were added to each indicator to show a site's performance (relative to the population's) in meeting the quality indicators over time (Figure 5). Some members of the working group expressed concern with presenting a sequential probability ratio test (SPRT) chart (11) due to the excessive complexity in its interpretation. Consensus was reached to aggregate three years of reporting period at six monthly intervals and to present 95% confidence interval to each time point to account for variability and precision in the data (Figure 5).

In terms of reporting multiple outcomes over time, the dual y-axes time series was initially considered. Function score was presented in the primary y-axis, the extent to which men were bothered by their symptoms was presented in the secondary y-axis and the period of follow up was presented on the x-axis. However, at the subsequent consultation session, the expanded working group noted that the graph suffered from "information overload" and advised to present each outcome as an individual time series.

**The reporting development platform—**Most respondents indicated a preference for an interactive report. However, given the size of the project (> 100 participating sites); consideration of time to set up a reporting platform which enables each LDC/site to visualise their own performance in comparison to other anonymised contributors; and expense in licencing subscription for commercial Business Intelligence solutions, a decision was made not to develop an interactive format in the early stages of the project.

R was selected as the software program in which to develop the reports, principally because of its excellent analytic and graph display capacity, open source licencing model and the large global community of developers and users sharing code to enhance the program. The use of RMarkdown provided a versatile document framework that can combine multiple coding languages (e.g., R, Python, STAN) and authoring languages (e.g., LaTeX, HTML). This allows for data to be processed, analysed, visualised and compiled in an accessible document format such as portable document format (pdf), HTML or word. In our case, it enabled generation of reports in a pdf, which allowed creation of paged and printable reports combining text and static graphs. Reports were typeset by embedding the LaTeX authoring language in the document, which enabled a versatile yet reproducible way of adhering with industry (Movember) style guideline.

**Report development—**The initial phase of report generation involved obtaining a database extract which was read, filtered and formulated into quality indicator scores using R scripts (Figure 6). Indicator-specific inclusion and exclusion criteria were used to identify the denominator of men eligible for each quality indicator during a defined period. The rate of performance for each unit (participating site/LDC) was presented as the proportion of eligible men attributed to that unit in the denominator who met criteria for the quality

indicator (e.g. men contributing to a LDC/site who had PSA level documented at diagnosis over a three-years period).

Aggregated data (numerator and denominator for each quality indicator) were used to generate tables and figures within R codes in the R Markdown document. These figures were formatted alongside dynamic descriptions, titles and statistics that were mapped from spreadsheets (*Calculation Details.xlsx* in Figure 6) accessible to collaborators external to the R Markdown interface.

Once data were formatted from the raw data to summarise data frames, they were exported to the RMarkdown template; here they dynamically filled tables, plots and text sections with results specific for the LDC/site. The report was then rendered into a pdf that was customised to a given LDC/site.

**Developing Design Features and Dissemination Strategy**—A number of design strategies were implemented to enhance the clarity of the report. An explanation of how to interpret each chart was incorporated at the front of the report. Color-coded side banners were added to pages to correlate with the subheadings banner in the dashboard (Figure 7). Funnel plots within individual indicator were constructed uniformly to reflect good performers at the top and the poor performers at the bottom. Similarly, the working group preferred having unidirectional performance summary charts in the dashboard which demonstrated better score as a contributor's dot moved to the right.

**Report Deployment**—A 'Report' page was added to the TrueNTH Global Registry data submission portal (3) which allowed authenticated users to download reports. Logics to manage access to files based upon LDC membership and email notification system to send emails to LDC users when a new report became available were created. LDC users also required to re-enter password to download their reports. TNGR administrators are able to log into the TNGR web portal to view activity in terms of report download by the LDC users.

Question and answer sessions were also held to provide opportunity for registry data managers and principal investigators to explore the reports and to ask questions. A "frequently asked question (FAQ)" page was created as a reference point for registry data managers. This page also contains information on indicator-specific inclusion and exclusion criteria and is accessed via the TrueNTH Global Registry data submission portal (3).

## DISCUSSION

Clinical quality registries provide regular feedback to individual providers with a view to developing best practices with respect to particular areas of care (12), reduce unwarranted variation and assist with implementation of practice changes (1). Despite their vast potential, reporting methodology and feedback mechanism vary considerably across registries and their impact has not been fully realised (13–16). Choosing the appropriate method of data presentation is essential to ensure that valid and accurate interpretations are made regarding the performance measure. Conversely reports including poorly-selected data presentation can convey misleading information.

Funnel plots with control limits at 95% and 99.8% were selected for use by TNGR to present quality indicators with binary outcomes. Although not as widely used as league charts (13, 15, 17), funnel plots accounted for varying provider's volume which suited the longitudinal nature of the registry. It should be noted that there were dissenting views on the use of control limits, with some participants feeling that inference was of little value in this setting and that use of control limits favours low volume providers, who are less likely to fall outside the control limits for a given level of suboptimal performance.

A study by Schmidtke et al (18) demonstrated that participants were likely to identify outliers when using funnel plots when compared to league charts. Being able to accurately identify the outlying data may motivate efforts by healthcare providers to reflect and instigate improvements in their local practice. Insufficient provider's volume and low event rates could compromise the precision of estimates of true underlying performance. A previous study examining surgical mortality as an indicator of hospital quality demonstrated that combining three years of data was considered sufficient for institution-level measures (19). However, it is acknowledged that the "volume" performed in an arbitrary three-year period by a provider is different for different medical conditions considered and for very high-volume interventions shorter periods are appropriate.

Scatter plots (20) and radar charts (10) have been used successfully by others to report multi-dimensional aspects of care. A previously published study plotted physician adherence to indicators in a bubble plot to motivate physicians' performance in adhering to guidelines for management of colorectal cancer (21). In this study, the standard error for individual adherence performances was expressed as a bubble. In our report, however, the bubble represented the population size of each LDC/site. Given the variability in how a bubble plot may be used across clinical practices, information and training are required to avoid misinterpretation of information.

The time-series graphs were seen as providing important information to supplement the static charts. Approaches used by existing registries varied widely, ranging from box plots of annual performance measures (13), bar charts broken down by quarterly performance measures (17), to more sophisticated statistical process control charts (22, 23). Sequential monitoring using control charts has been shown to be more sensitive than funnel plot as it displayed details of the history of outcomes at a particular LDC/site and thus facilitated early detection of outliers (23). It uses a cumulative log scale on the y-axis and time on the x-axis. The TNGR working group preferred simple time-series to SPRT charts because the y-axis showed an event rate (e.g. positive margin rates) which was easy for clinicians to interpret.

There is overwhelming evidence that performance feedback improves adherence to process guidelines and health outcomes (24, 25). The data specifically referencing dashboards as the mechanism of performance feedback, compared to reports or interactive systems, is less clear. This may be because what defines a "dashboard" compared to other forms of feedback is open to debate. The TNGR quality of care report incorporated the dashboard performance summary chart to provide a snapshot of contributor's performance in regard

to quality indicators. In the absence of an acceptable international benchmark (13, 14), the TNGR adopted the *Achievable Benchmarks of Care* (9).

Timely and routine feedback enables health providers to systematically monitor their quality of care. Diversity in the frequency of reporting is evident across clinical quality registries. Some registries provided hospital-level data to participating sites every three months (12, 15), others every six months (26) and others annually (13). In TNGR, quality of care reports are distributed to participating countries every six months to align with the TNGR data submission period. This approach enabled each performance measure to be reported based on the most up-to-date data.

### Strengths and limitations

There are several limitations when designing and implementing a quality of care report for the TNGR. There is inevitably a subjective aspect to the selection of the data presentation based on the level of expertise and personal experience of the working group participants. Variation in graphical attributes within the questionnaire such as color, line style and other design patterns may potentially confound the results of participants' preferences. It is acknowledged that with a different composition of working group, the selected data presentations could have been significantly altered. Comprehension of quality of care reports by diverse audiences is paramount to the success of a registry quality improvement program (27–29). As the TNGR quality of care report is published exclusively in English, this may limit the ability of non-native English-speaking audience to communicate the results to their local practitioners. A number of strategies were implemented to mitigate this issue. Feedback on the language and terms used in the reports was sought from our international contributors including those who came from non-English speaking background. We also added interpretation text, color coded report sections and provided user guides.

The strength of this project includes the heterogeneity of our working group participants bringing important local perspectives and experience to discussion. There were inevitably dissenting views among working group members. Finding the balance between simplicity and statistical fidelity was echoed throughout workshops and the subsequent consultation meetings. Data presentation must be easy to interpret with consideration that not all stakeholders are well versed in statistics, but data presentation cannot be oversimplified to prevent misinterpretation of LDCs/sites performance.

The working group members were also engaged throughout stages of report development to further modify layout, text, colour schemes and figures, particularly as reports progressed from theoretical models to those containing real data from contributing sites. As the TNGR is still in its infancy, we are unable to sufficiently quantify if this report will improve outcomes. However, to our knowledge, this is the first registry which strives to improve prostate cancer care globally by using a consistently defined international dataset.

### Future proposed activity

As the registry matures, evaluation of the existing designs will be undertaken to identify usability problems/ barriers, utilizing tools such as user experience evaluation questions by Lam et al (30). Translational research action plan will also be developed and implemented to

prioritize variances that need to be addressed. In addition, future implementation studies to assess the effectiveness of strategies to facilitate adherence and the impact of adhering to the evidence-based recommendations may also be carried out.

## CONCLUSION

Data visualization plays an integral role in communicating information derived from registry to healthcare providers. To enhance its effectiveness, performance measures should be reported in a comprehensive, concise and easy-to-understand manner to allow rapid identification of trends and outliers while adapting a scientifically sound data display technique. Using an R Markdown framework, we developed a registry report system that would produce multiple dynamic pdf reports that provide user-specific feedback to data managers and participating countries. We provide feedback through visualisation of a contributor's success rate in completing 33 quality indicators in comparison to other anonymised contributors. The TNGR quality of care and data completeness reports have undergone rigorous consultation during development. The TNGR performance metrics will be regularly re-evaluated to ensure continued validity and accuracy.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Abbreviations:

|             |                         |
|-------------|-------------------------|
| <b>LDC</b>  | Local data centre       |
| <b>TNGR</b> | TrueNTH Global Registry |

## REFERENCES

1. Evans SM, Scott IA, Johnson NP, Cameron PA, McNeil JJ. Development of clinical-quality registries in Australia: the way forward. *Med J Australia*. 2011;194(7):360–3. [PubMed: 21470087]
2. Evans SM, Millar JL, Moore CM, Lewis JD, Huland H, Sampurno F, et al. Cohort profile: the TrueNTH Global Registry - an international registry to monitor and improve localised prostate cancer health outcomes. *Bmj Open*. 2017;7(11).
3. Sampurno F, Kannan A, Lucas M, Liman J, Connor SE, Pearman E, et al. Development of Technologic Solutions to Address Complex Local Requirements of an International Prostate Cancer Clinical Quality Registry. *Jco Clin Cancer Info*. 2019;3.
4. Sampurno F, Zheng J, Di Stefano L, Millar JL, Foster C, Fuedea F, et al. Quality Indicators for Global Benchmarking of Localized Prostate Cancer Management. *J Urol*. 2018;200(2):319–26. [PubMed: 29477721]
5. In J, Lee S. Statistical data presentation. *Korean J Anesthesiol*. 2017;70(3):267–76. [PubMed: 28580077]
6. Jones Andrew Michael (2017). Data visualization and health econometrics. *Foundations and Trends in Econometrics*.
7. Schwabish Jonathan A. 2014. "An Economist's Guide to Visualizing Data." *Journal of Economic Perspectives*, 28 (1): 209–34.
8. Cleveland William S., and McGill Robert. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79, no. 387 (1984): 531–54.

9. Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback - A randomized controlled trial. *Jama-J Am Med Assoc.* 2001;285(22):2871–9.
10. Thaker NG, Ali TN, Porter ME, Feeley TW, Kaplan RS, Frank SJ. Communicating Value in Health Care Using Radar Charts: A Case Study of Prostate Cancer. *J Oncol Pract.* 2016;12(9):813–+. [PubMed: 27577622]
11. Spiegelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care.* 2003;15(1):7–13. [PubMed: 12630796]
12. <https://musicurology.com/wp-content/uploads/2018/10/Collaborative-Quality-Improvement.pdf> (Accessed October 2019).
13. Krebsgesellschaft Deutsche (ed) (2016) Annual Report of the Certified Prostate Cancer Centres (accessed October 2019).
14. Canadian Partnership Against Cancer. The 2014 cancer system performance report. Toronto: Canadian Partnership Against Cancer; 2014. Available from [systemperformance.ca](http://systemperformance.ca) (accessed October 2019).
15. Stattin P, Sandin F, Sandback T, Damber JE, Lissbrant IF, Robinson D, et al. Dashboard report on performance on select quality indicators to cancer care providers. *Scand J Urol.* 2016;50(1):21–8. [PubMed: 26162012]
16. Hoque DME, Ruseckaite R, Lorgelly P, McNeil JJ, Evans SM. Cross-sectional study of characteristics of clinical registries in Australia: a resource for clinicians and policy makers. *Int J Qual Health C.* 2018;30(3):192–9.
17. LaBresh KA, Reeves MJ, Frankel MR, Albright D, Schwamm LH. Hospital treatment of patients with ischemic stroke or transient ischemic attack using the “get with the guidelines” program. *Arch Intern Med.* 2008;168(4):411–7. [PubMed: 18299497]
18. Schmidtke KA, Watson DG, Vlaev I. The use of control charts by laypeople and hospital decision-makers for guiding decision making. *Q J Exp Psychol.* 2017;70(7):1114–28.
19. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality - The problem with small sample size. *Jama-J Am Med Assoc.* 2004;292(7):847–51.
20. Vickers AJ, Sjoberg D, Basch E, Sculli F, Shouery M, Laudone V, et al. How Do You Know If You Are Any Good? A Surgeon Performance Feedback System for the Outcomes of Radical Prostatectomy. *Eur Urol.* 2012;61(2):284–9. [PubMed: 22078336]
21. Chien TW, Lin YF, Chang CH, Tsai MT, Uen YH. Using a bubble chart to enhance adherence to quality-of-care guidelines for colorectal cancer patients. *Eur J Cancer Care.* 2012;21(6):712–21.
22. Noyez L Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interact Cardiovasc Thorac Surg.* 2009;9(3):494–9. [PubMed: 19509097]
23. Coory M, Duckett S, Sketcher-Baker K. Using control charts to monitor quality of hospital care with administrative data. *Int J Qual Health C.* 2008;20(1):31–9.
24. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review. *JMIR Hum Factors.* 2018;5(2):e22. [PubMed: 29853440]
25. Dowding D, Randell R, Gardner P, Fitzpatrick G, Dykes P, Favela J, et al. Dashboards for improving patient care: review of the literature. *Int J Med Inform.* 2015;84(2):87–100. [PubMed: 25453274]
26. Sampurno F, Earnest A, Kumari PB, Millar JL, Davis ID, Murphy DG, et al. Quality of care achievements of the Prostate Cancer Outcomes Registry-Victoria. *Med J Australia.* 2016;204(8).
27. Haddad FS, Manktelow AR, Skinner JA. Publication of surgeon level data from registers: who benefits? *Bone Joint J.* 2016;98-B(1):1–2.
28. Hibbard JH, Stockard J, Tusler M. Does publicizing hospital performance stimulate quality improvement efforts? *Health Affair.* 2003;22(2):84–94.
29. Jewett JJ, Hibbard JH. Comprehension of quality care indicators: Differences among privately insured, publicly insured, and uninsured. *Health Care Financ R.* 1996;18(1):75–94.

30. Lam H, Bertini E, Isenberg P, Plaisant C, Carpendale S. Seven guiding scenarios for information visualization evaluation. 2011

Author Manuscript

Author Manuscript

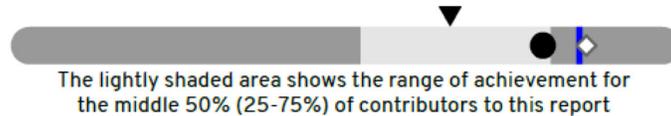
Author Manuscript

Author Manuscript

| CLINICAL   |          | PATIENT-REPORTED OUTCOMES (PROMS)  |          |
|--|----------|--|----------|
| CARE   | OUTCOMES | PROCESS OF CARE  | OUTCOMES |
| <b>Binary outcomes</b>   |          | <b>Multiple outcomes</b>   |          |
| <p><b>Requirement 1:</b> Provide a comparison with others of how a site performed against an indicator with binary outcomes in the specified reporting period</p> <p><b>Requirement 3:</b> Provide a temporal analysis of site performance against an indicator with binary outcomes</p> |          | <p><b>Requirement 2:</b> Provide a comparison with others of site performance performed against an indicator with multi-dimensional outcomes in the specified reporting period</p> <p><b>Requirement 4:</b> Provide a temporal analysis of site performance against an indicator with multi-dimensional outcomes</p> |          |
| <p><b>Requirement 5:</b> Provide a summary overview of site performance across indicators</p>  |          |  |          |

**Figure 1.**  
Requirements of the data presentation for the TrueNTH Global Registry Quality of Care report.

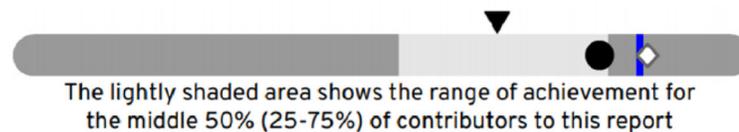
-  The black dot shows your LDC's result for men diagnosed between 01 Jul 2015 and 30 Jun 2018
-  The white diamond shows your LDC's result for men diagnosed between 01 Jan 2015 and 31 Dec 2017 as shown on report dated 27 May 2019
-  The triangle shows the median of all TrueNTH Global Registry LDCs for men diagnosed between 01 Jul 2015 and 30 Jun 2018
-  The vertical blue line shows the aspirational target for this indicator. The aspirational target is set using the *Achievable Benchmarks of Care* approach<sup>1,2</sup>, which estimates the aspirational target based on the top performing sites, which account for 10% of patients in the registry.



<sup>1</sup>Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. JAMA. 2001 Jun 13; 285(22):2871-9.

<sup>2</sup>Paddock SM. Statistical benchmarks for health care provider performance assessment: a comparison of standard approaches to a hierarchical Bayesian histogram-based method. HEALTH SERVICES RESEARCH. 2014 Jun; 49(3):1056-73.

-  The black dot shows your LDC's result for men diagnosed between 01 Jan 2017 and 31 Dec 2019
-  The white diamond shows your LDC's result for men diagnosed between 01 Jul 2016 and 30 Jun 2019 as shown on report dated 27 Nov 2020
-  The triangle shows the median of all TrueNTH Global Registry LDCs for men diagnosed between 01 Jan 2017 and 31 Dec 2019
-  The vertical blue line shows the aspirational target for this indicator. The aspirational target is set using the *Achievable Benchmarks of Care* approach<sup>1,2</sup>, which estimates the aspirational target based on the top performing sites, which account for 10% of patients in the registry.

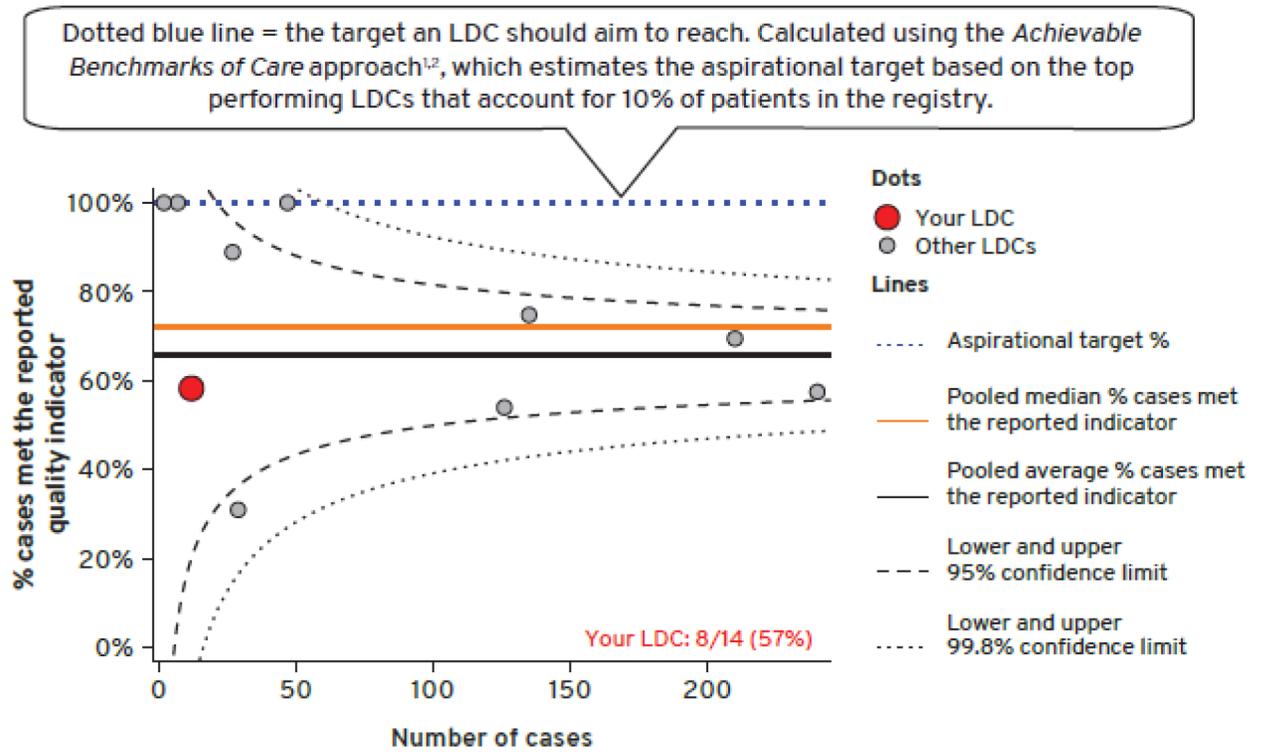


<sup>1</sup>Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. JAMA. 2001 Jun 13; 285(22):2871-9.

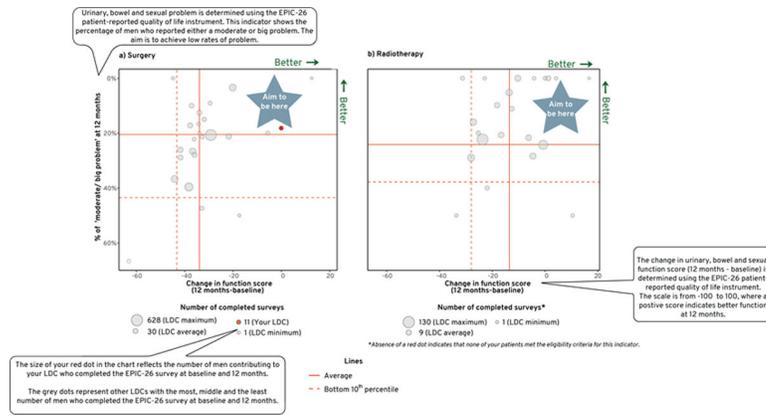
<sup>2</sup>Paddock SM. Statistical benchmarks for health care provider performance assessment: a comparison of standard approaches to a hierarchical Bayesian histogram-based method. HEALTH SERVICES RESEARCH. 2014 Jun; 49(3):1056-73.

### Figure 2.

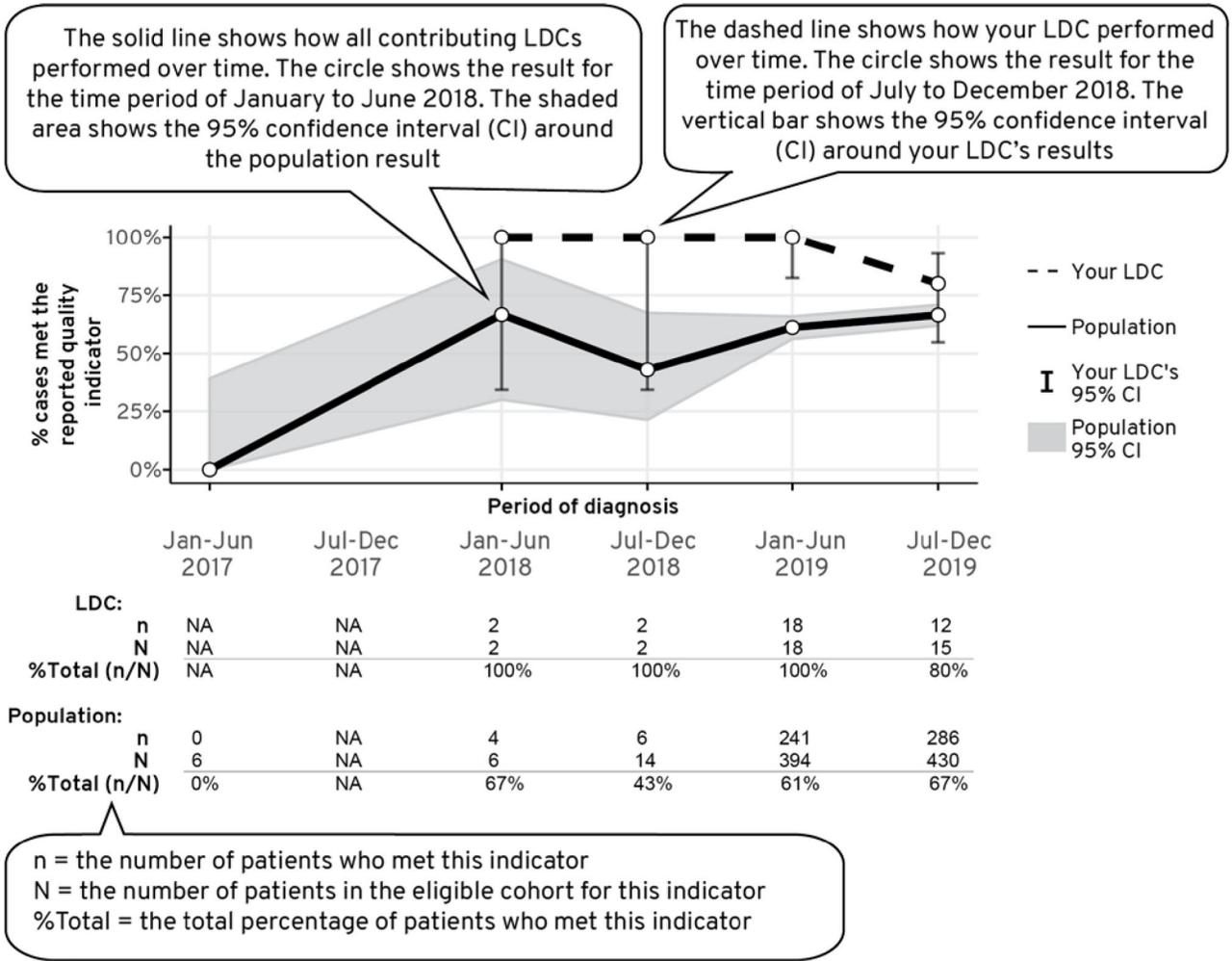
Performance summary charts were used in the TrueNTH Global Registry report to provide a basic overview of an LDCs' performance for each quality indicator in relation to the population minimum, 1<sup>st</sup> quartile, median (2<sup>nd</sup> quartile), 3<sup>rd</sup> quartile, maximum and an aspirational target.



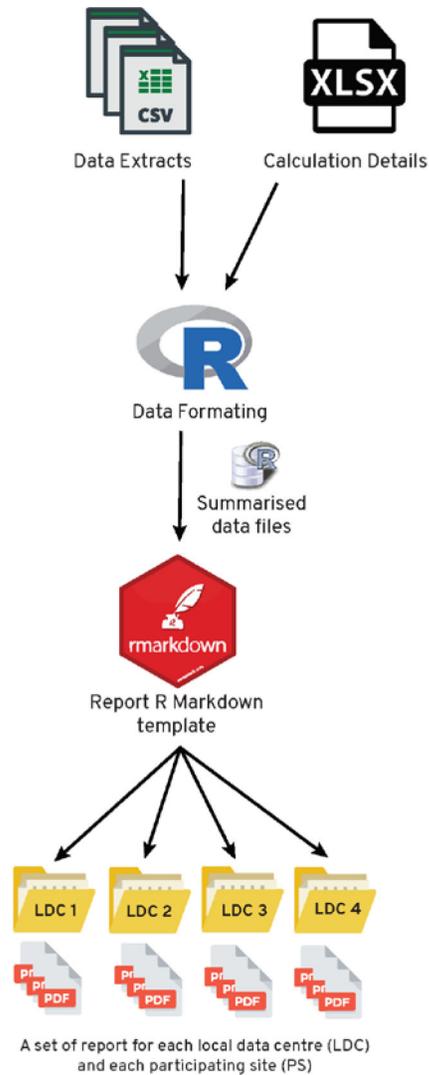
**Figure 3.** Interpreting the funnel plot was used to convey an LDCs’ performance for each quality indicator in comparison to all other LDCs’ over a three-years reporting period.



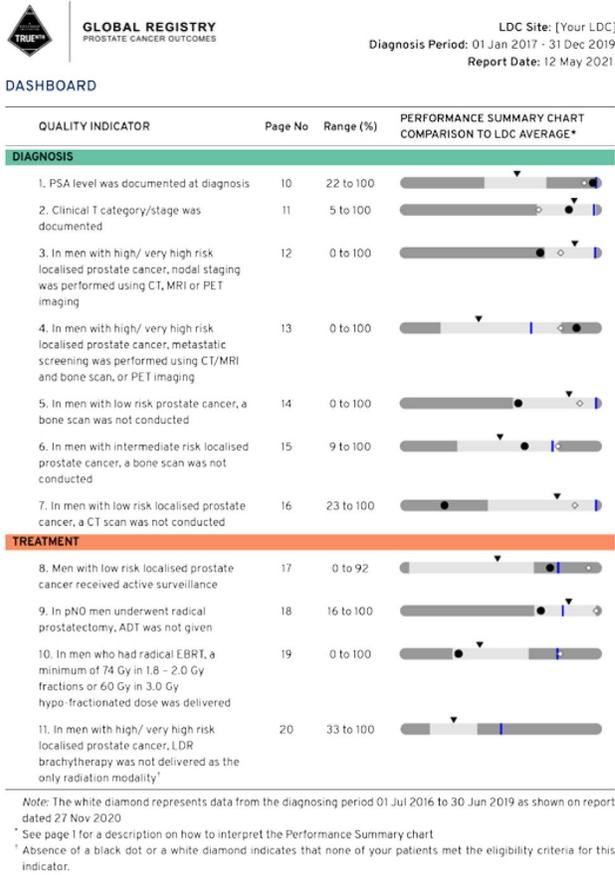
**Figure 4.** Bubble plot was used to convey an LDCs’ performance for sexual, urinary and bowel function scores and the extent to which men contributing to that LDC were bothered by their symptoms in comparison to all other LDCs’ over a three-year reporting period.



**Figure 5.** Time series chart was used to convey how a LDCs’ performance for each quality indicator varied across time in comparison to the population average.



**Figure 6.** The workflow for generating quality indicator reports in R. Data is mapped to calculation details for each quality indicator and formatted (i.e. summarized) for use within the plots of the report. Within an RMarkdown template data, text, tables and plots are dynamically generated for each LDC/site.



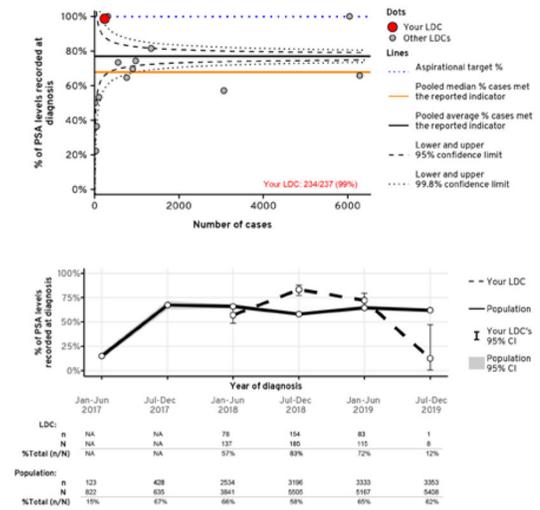
**1. PSA LEVEL WAS DOCUMENTED AT DIAGNOSIS**

This indicator reports on how well PSA was documented at diagnosis<sup>1</sup>. PSA levels are important to understand risk of disease progression. We only include men diagnosed via TRUS or transperineal biopsy as prostate cancer may be an incidental finding (e.g. TURP).

Figure 1 provides a summary of the completeness of PSA documentation for men at your LDC and trend in documentation over time including 95% confidence intervals (CIs).

|                        |   |
|------------------------|---|
| <b>Numerator (n)</b>   | Men diagnosed via TRUS or transperineal biopsy who had a PSA level documented within a six month period prior to the biopsy |
| <b>Denominator (N)</b> | Men diagnosed via TRUS or transperineal biopsy  |

Figure 1: Percentage (%) of PSA levels documented in contributing LDCs at diagnosis and trend



<sup>1</sup>NOTE: PSA level must be taken within a six month period prior to or up to date of diagnosis.

DIAGNOSIS

**Figure 7.**  
 An example of the TrueNTH Global Registry dashboard and the main report page

**Table 1.**

Specialties and countries of practice of the working group

|                       | Primary working group                              |   | Expanded working group                          |   |
|-----------------------|--|---|---|---|
|                       | Number of clinicians<br>(% responded to<br>survey) | Number of non-<br>clinicians (%<br>responded to survey) | Number of clinicians (%<br>responded to survey) | Number of non-clinicians<br>(% responded to survey) |
| <b>Austria</b>        |  |   | 1 (100%)  |   |
| <b>Australia</b>      | 1 (100%)   | 1 (100%)  | 3 (100%)  | 6 (50%)   |
| <b>Canada</b>         | 1 (0%)   |   | 1 (100%)  | 2 (100%)  |
| <b>Germany</b>        |  | 1 (100%)  |   | 1 (100%)  |
| <b>Hong Kong</b>      |  |   | 1 (100%)  |   |
| <b>Ireland</b>        | 1 (100%)   |   | 1 (100%)  |   |
| <b>Italy</b>          |  |   | 2 (100%)  |   |
| <b>Netherlands</b>    |  |   | 1 (100%)  | 1 (100%)  |
| <b>New Zealand</b>    |  |   | 1 (100%)  |   |
| <b>Norway</b>         |  |   | 1 (100%)  | 1 (100%)  |
| <b>Spain</b>          |  |   |   | 1 (100%)  |
| <b>United Kingdom</b> | 1 (100%)   |   |   | 2 (100%)  |
| <b>United States</b>  | 2 (100%)   | 4 (75%)   | 5 (100%)  | 7 (100%)  |