



## Prioritisation of elective surgery in New Zealand: The Reliability Study

Carolyn Doughty, Andrew MacCormick, Justin Roake, John Fraser, Phil Hider, Ray Kirk, Bryan Parry, Andre van Rij, Jean-Claude Theis

### Abstract

**Aims.** This paper describes the rationale and methodology of a study assessing the reliability of tools for clinical prioritisation (Clinical Priority Assessment Criteria [CPAC]) of patients for elective surgery in New Zealand.

**Methods.** Surgeons from three specialties (general, vascular, and orthopaedic surgery) completed a computerised evaluation rating clinical vignettes across a range of diagnoses using several priority tools. The study design is described and an outline of the individual tool development and definitions is given.

**Results.** Of the 124 surgeons that participated in this study, 48% (60) were general surgeons, 21% (26) were vascular surgeons and 31% (38) were orthopaedic surgeons. The response rates in the first phase of data collection were 67%, 79%, and 63% for general, vascular, and orthopaedic surgery respectively. Completion rates were high with 100%, 93%, and 98% of the same groupings of surgeons completing the first round evaluations. A further 77% to 89% of the participants from the first round also completed the re-test evaluation.

**Conclusion.** This study indicates that assessment of the reliability of CPAC tools currently in use in New Zealand is feasible using a vignette-based approach. In the future, study designs that allow for some face-to-face contact may be preferable for achieving optimal response and completion rates among surgeons. Further work from this study will focus on the individual results for each specialty and examining whether altering ethnicity status in vignettes had any effect on scoring behaviour.

The New Zealand Priority Access Project was developed in the 1990s by the funders of health care to provide criteria to prioritise patients for elective surgery.<sup>1</sup> In conjunction with a system of booking specific times for surgery, the project was designed to provide more consistency and transparency to the access for elective surgery across the country.<sup>2</sup>

Clinical Priority Assessment Criteria (CPAC) for several procedures were developed that included clinical and social parameters aimed at estimating a patient's ability to benefit from a procedure.<sup>1</sup>

Procedure-specific CPAC tools were initially established in three clinical disciplines amidst considerable international interest and some local controversy.<sup>1,3</sup> Although the tools were developed to improve the judgement of clinicians, research findings indicated they had some limitations. Coronary artery bypass grafting CPAC scores for example did not predict cardiac events among patients on a waiting list for bypass grafting better than clinical judgement alone.<sup>4</sup>

Widespread introduction of CPAC throughout New Zealand was problematic. In the public sector, universal use of CPAC tools was made compulsory in order to receive funding for surgery. To clear the backlog of people waiting for care and to facilitate the introduction of the system, additional funding was provided over several years.

Funding was intended to be available only where clinical priority assessment criteria and booking systems were in place, including audits of waiting lists being completed and financially sustainable thresholds established to determine access.<sup>5</sup> As there were few nationally agreed tools, local versions were developed and (in the process) the procedure-specific focus was generalised to encompass entire specialties. This was the origin of the Generic Surgical Priority Criteria (GSPC) tool used in general and vascular surgery.

The tools were often applied as protocols rather than decision support tools,<sup>6</sup> and many were introduced without any formal assessment of their reliability. Those few that were compared, were associated with relatively low levels of inter-rater reliability.<sup>7-9</sup>

In recognition of the deficiencies of many of the generic point scoring prioritisation tools introduced in 1998, the National Waiting Times Project (NWTP) initiated development of further CPAC tools with a specialty-wide (rather than condition- or procedure-specific) focus. This was done by specialty groupings of interested surgeons and general practitioners.

The approaches to the task were varied, but tools collectively known as 'Integrated Scoring Systems' (ISS) have emerged as leading contenders for specialty-wide prioritisation.<sup>10</sup> The principle behind ISS is that a panel of specialists develop a table indicating the range of scores-specific conditions treated within that specialty may attract. This, in effect, gives some ranking to those conditions but with wide overlaps and is undertaken as a transparent process. As the score ranges are not concerned with individuals the specialists are not confronted with a conflict of interest.

When assessing individuals specialists are then required to grade priority on a numerical (Likert) or visual analogue scale. Priority assigned is condition-specific; for example, maximum priority would represent the severest presentation of that condition together with high capacity to benefit. The grade is assigned utilising the specialist's clinical judgement without the need to assign scores to specific domains. By combining the specialist assigned grade and the tabulated range of possible scores a specialty-wide priority scoring system is developed.

There has also been interest in the reliability of global clinical judgement, without the constraints of tabulated ranking of conditions, as a prioritisation tool.<sup>11</sup> Under this system, priority would simply be assigned by using a visual analogue scale (VAS), but further research on how surgeons intuitively prioritise is still needed.

The CPAC Evaluation Consortium was commissioned to broadly evaluate the introduction of CPAC tools into New Zealand. The study rationale and methodology described here was designed to evaluate the reliability of CPAC tools in general use as well as tools proposed by the specialty groups of the NWTP. The study aimed to determine the reliability of the CPAC tools for inter-surgeon and intra-surgeon scoring of patients. The method employed clinical vignettes, as a surrogate for real patients, which were scored by surgeons from three specialties (general surgery,

vascular surgery, and orthopaedic surgery) using three separate priority tools within each specialty.

By using vignettes, the responses of surgeons to a series of uniform and clinically relevant scenarios were gathered. Under similar circumstances, vignettes have been demonstrated to reliably represent the decisions made by clinicians with simulated patients presenting with the same problem.<sup>12</sup>

## Methods

**Selection of specialties and surgeons**—The three specialties selected for study presented spectra of conditions and circumstances such that the CPAC tools could be evaluated under a wide variety of circumstances. Three parallel (but not identical) studies involving each specialty were conducted across the country.

Vascular surgery was included because it encompassed a small number of distinctive clinical strands (e.g. chronic venous insufficiency) with problems ranging from life or limb threatening to relatively minor. Furthermore, the number of vascular surgeons nationally was sufficiently small that the study could be designed to include the whole population who could be engaged through a strong national society to which all practising vascular surgeons belong.

Orthopaedic surgery was included because of its focus on a large number of high-volume and high-cost procedures that mainly sought to enhance quality of life. Surgeon numbers were such that only a random sample could be included (see surgeon selection below).

General surgery was included as possibly the most difficult case for prioritisation having a large number of heterogeneous conditions and procedures; however, a history of research into priority systems in general surgery<sup>7,8</sup> made it a potentially informative specialty for inclusion. As with orthopaedics, surgeon numbers were such that only a random sample could be included.

Eligible subjects for the three studies were all surgeons in each specialty that were practising at least part time in the public health system. General surgeons were those listed on the Medical Council of New Zealand's General Surgery Vocational Register, vascular surgeons were those listed on the register of the New Zealand Society of Vascular Surgeons, and orthopaedic surgeons were those listed on the New Zealand Orthopaedic Association Register.

Surgeons who practised full time in the private system (some general and orthopaedic surgeons but no vascular surgeons) were excluded, as they were unlikely to be familiar with the need for prioritisation of patients for surgery. Surgeons from each specialty who developed the patient vignettes were also excluded. Eligible surgeons were contacted by letter and telephone to determine their willingness to participate in the study. General surgeon and orthopaedic surgeon participants for the study were then selected from the eligible surgeons who had indicated their willingness to participate. In vascular surgery, all eligible surgeons willing to participate were included.

**Development of vignettes and selection of clinical diagnoses**—The three studies were undertaken using electronically administered patient vignettes. Vignette presentation was standardised across specialties and included the patient's diagnosis, age, gender, ethnicity, comorbidity, clinical history, and clinical examination as well as relevant investigation results.

For general surgery, 32 vignettes that covered four common generic surgical diagnoses; hernia (8), cholelithiasis (8), haemorrhoids (8), and right-sided colon cancer (8) were developed. The vignettes were constructed on a fractional factorial design<sup>13</sup> and included cues on each of seven criteria that surgeons had previously indicated they used in prioritising patients for elective general surgery. The vignettes were piloted with six surgeons purposely selected because they represented a range of ages and locations.<sup>11</sup>

For vascular surgery, 60 vignettes covering three diagnoses (varicose veins [20], carotid artery stenosis [20], and intermittent claudication [20]) were constructed to cover the possible spectrum of severity for those conditions.

For orthopaedic surgery, 50 vignettes were developed. They included two specific patient complaints, carpal tunnel syndrome, and painful knee, each with 15 vignettes and a group of 20 vignettes covering a range of diagnoses including painful hip, disc prolapse, painful shoulder, trigger finger, torn meniscus, tennis elbow, and Dupuytren's contracture.

**Selection of CPAC tools for evaluation**—In each specialty, three priority tools (each utilising a different approach to prioritisation) were used to assign priority to each vignette.

The three approaches are summarised as:

- Global assignment of priority based upon clinical opinion (i.e. VAS tools in each specialty) included on the basis of prior research,<sup>8</sup>
- Point-scoring systems (i.e. GSPC in general and vascular surgery and the Generic Orthopaedic Access Criteria tool) included as examples of the tools in use at the time, and
- Three specialty-specific integrated scoring systems (ISS) as products of the NWTP. The details of the tools used in each specialty are described below and the complete range of tools is summarised in Table 1.

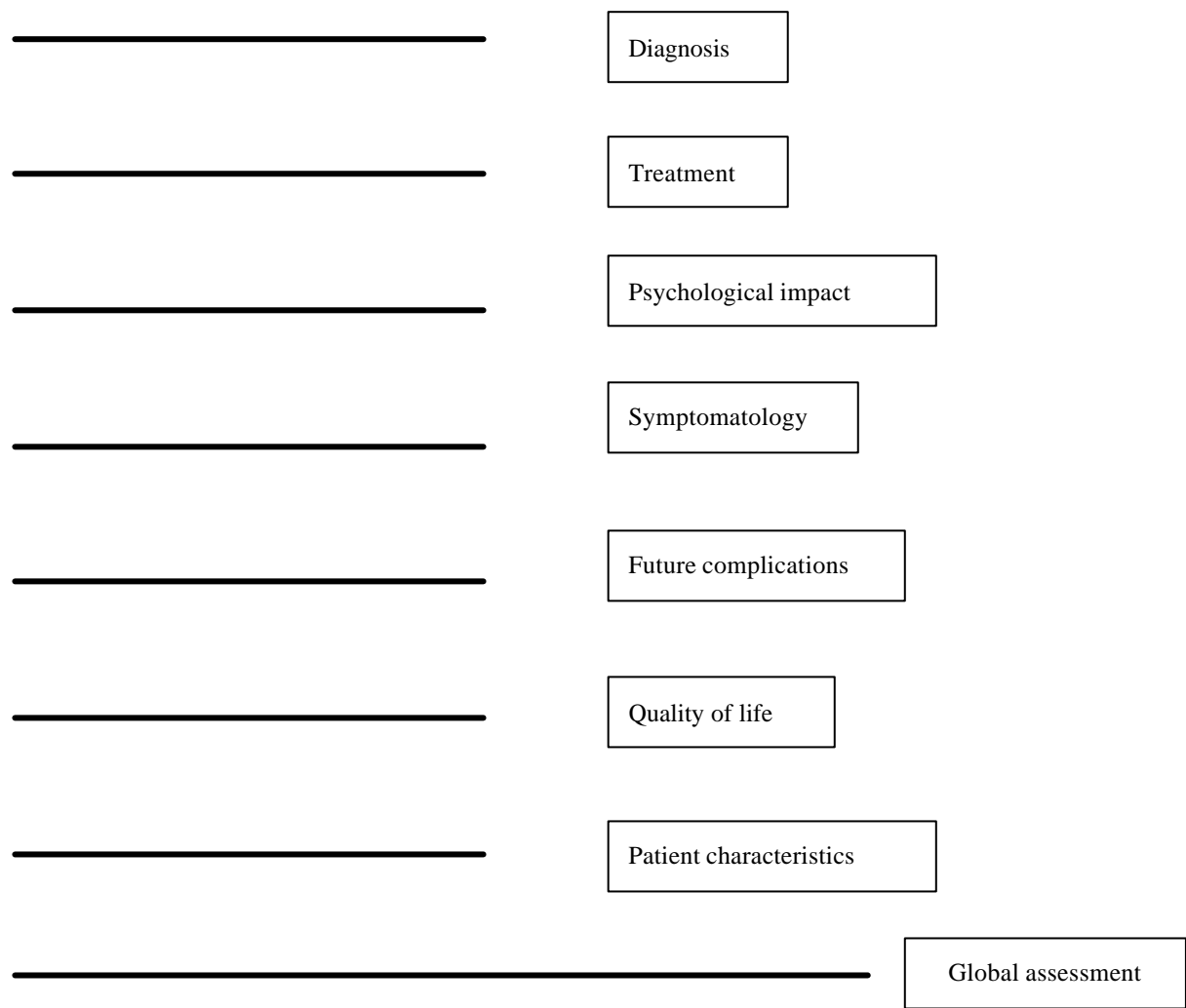
**Table 1. Summary of the specific dimensions of the CPAC tools evaluated for each specialty**

Scoring of Priority <sup>1</sup>			
Tool	General Surgery	Vascular Surgery	Orthopaedic Surgery
GSPC <sup>2</sup> / GOAC <sup>3</sup>	Severity	Severity	Natural History
	Suffering	Suffering	Pain
	Disability	Disability	Personal Functional Limitation
	Clinical Cost of Delay	Clinical Cost of Delay	Social Limitation
	Ability to Benefit	Ability to Benefit	Potential to Benefit
	Degree of Improvement <sup>4</sup>	Degree of Improvement <sup>4</sup>	
	Likelihood of Improvement	Likelihood of Improvement	
VAS	Diagnosis	Suffering	Suffering
	Psychological Impact	Disability	Disability
	Treatment	Clinical Cost of Delay	Clinical Cost of Delay
	Symptoms		
	Patient Characteristics		
	Future Complications		
	Impact on Life		
	Global Assessment (0 to 100)	Global Priority (1 to 100)	Global Priority (1 to 100)
ISS	Lowest to Highest (1 to 9)	Lowest to Highest (1 to 5)	Lowest to Highest (1 to 5)

<sup>1</sup> All formulae produce a score between 1 and 100; <sup>2</sup> For the GSPC the score is the product of factors severity and ability to benefit. These are defined as the sum of their components; <sup>3</sup> For the Generic Orthopaedic Access Tool the score is the sum of the factor; <sup>4</sup> Anticipated improvement; CPAC=Clinical Priority Assessment Criteria; GOAC=Generic Orthopaedic Access Tool; GSPC=Generic Surgical Priority Criteria; ISS=Integrated Scoring Systems; VAS= Visual Analogue Scale.

**General surgery**—The VAS tool was a set of eight visual analogue scales (Figure 1). Seven of these were 50 millimetre subordinate scales. These seven scales were for each of seven criteria identified from surgeons as relevant to prioritisation;<sup>11</sup> this enabled the measurement of the impact of a criterion on the patient’s priority. Subsequently, a 100 millimetre visual analogue scale measured the subjects’ global judgement of priority. This measurement was converted into a score out of 100.

**Figure 1. Visual Analogue Scale containing seven criteria as sub-scales**



The GSPC tool consisted of five criteria. The algorithm consists of the sequential addition and multiplication of the criteria as in the box below.

Severity = suffering (maximum 7) + disability (maximum 7) + clinical cost of delay (maximum 6)

Ability to Benefit = degree of improvement (maximum 2) + likelihood of improvement (maximum 3)

GSPC Score (100 points) = severity (20 points) x ability to benefit (5 points)

The definitions of the criteria included in the GSPC are as follows:

- **Suffering:** Suffering is most commonly pain but would also include any other disvalued symptom or experience for example, nausea, vomiting, vertigo and tinnitus. The other component, which could be included unless evaluated separately, is anxiety about the illness.
- **Disability:** Disability caused or exacerbated by this condition.
- **Clinical cost of delay:** Clinical cost of delay refers to the possibility that delay will lead to future suffering or premature death and/or result in a poorer operative result.
- **Degree of improvement anticipated:** 'Gold standard' of benefit, where this procedure worked well for a patient.
- **Likelihood of improvement:** Likelihood of achieving that level of benefit in this case.

The ISS in general surgery consisted of a nine point diagnosis-specific Likert scale combined with tabulated score ranges determined for each diagnosis from that observed in previous work<sup>14</sup>. The Likert scale points (from 1 to 9) were anchored to the 0, 5, 10, 25, 50, 75, 90, 95, and 100th percentiles respectively (Table 2). This relationship implies that, when scoring a patient, there is an assumed normal distribution of a given diagnosis over the Likert scale of 1–9.

**Table 2. Relationship of Likert score to percentile and priority score for given diagnoses in general surgery**

Likert score	1	2	3	4	5	6	7	8	9
<b>Percentile</b>	<b>0</b>	<b>5</b>	<b>10</b>	<b>25</b>	<b>50</b>	<b>75</b>	<b>90</b>	<b>95</b>	<b>100</b>
Anorectal	10	28	36	50	65	78	90	100	100
Cholelithiasis	20	33	41	55	70	86	99	100	100
Hernia	4	19	28	45	61	77	89	97	100
Colon cancer	45	80	86	95	100	100	100	100	100

**Vascular surgery**—The VAS in vascular surgery was similar to general surgery except that it utilised only three subordinate scales namely suffering, disability, and clinical cost of delay. These criteria were chosen to provide transparency to the overall priority assessment and cross correlate with elements of the GSPC.

The vascular (and orthopaedic) VAS tool(s) differed in this respect from the general surgery tool for two reasons. First, the subscales of the general surgery tool were derived from interviews of general surgeons and no analogous process had been undertaken in vascular (or orthopaedic) surgery. Second, to achieve sufficient statistical power, the study required each vascular (and orthopaedic) surgeon to score more individual vignettes than in general surgery. The number of vignettes used had to be balanced against the pragmatic need to consider the workload of participating surgeons and ensure high levels of completion. The GSPC in vascular surgery was exactly the same as used in general surgery.

The ISS for vascular surgery differed from general surgery. As a method of measuring diagnosis-specific clinical judgement, a single 100-millimetre visual analogue scale was used (instead of a Likert scale as used in general surgery). Each point of the scale was then converted to a score using tabulated diagnosis-specific score ranges. These tables had been constructed by consensus between members of the vascular surgical NWTP group. Within each diagnostic group, it was assumed that the scores would be normally distributed about a mean. Scores were calculated using an inverse normal distribution function where the mean and standard deviation were determined by the NWTP group and the extremes of the VAS represented the 2nd and 98th percentiles.

**Table 3. Relationship of Visual Analogue Scale (VAS) to priority score for given diagnoses in vascular surgery<sup>1</sup>**

Point on VAS (mm) <sup>2,3</sup>	£2	5	10	20	30	40	50	60	70	80	90	95	≥98
Varicose veins	0	6	12	18	23	27	31	35	39	44	50	56	62
Intermittent claudication	17	21	24	28	31	34	36	38	41	44	48	51	55
Carotid stenosis	35	41	47	53	58	62	66	70	74	79	85	91	97

<sup>1</sup> Scores were calculated using an inverse normal distribution function where the mean and standard deviation (SD) for each diagnostic group was determined by the NWTP group; <sup>2</sup> Selected examples shown (intermediate values to nearest millimetre were included); <sup>3</sup> Mean and standard deviation for varicose veins ( $31 \pm 15.2$ ), intermittent claudication ( $36 \pm 9.1$ ) and carotid stenosis ( $66 \pm 15.2$ ); NWTP=National Waiting Times Project.

**Table 4. Relationship of Likert score to quintile and priority score for given diagnoses in orthopaedic surgery**

Likert score	1	2	3	4	5
<b>Knee</b>					
Painful knee	40	52	65	77	90
<b>Wrist</b>					
Carpal tunnel syndrome	25	39	52	66	80
<b>Mixed category</b>					
Malunion supracondylar fracture humerus	20	32	45	57	70
Tennis elbow	20	30	40	50	60
Osteoarthritis ankle	40	51	62	74	85
Bunion / Hallux valgus	20	31	42	54	65
Ingrowing toenail	5	16	27	39	50
Ganglion wrist	5	17	30	42	55
Leg length discrepancy	30	42	55	67	80
Healed fracture femur	5	19	32	46	60
Trigger finger	35	42	50	57	65
Osteoarthritis hip	40	52	65	77	90
Loosening total hip replacement	60	69	77	86	95
Dupuytren's contracture finger	20	34	47	61	75
Rupture extensor tendon rheumatoid arthritis	40	50	60	70	80
Torn meniscus	40	50	60	70	80
Disc prolapse	50	60	70	80	90
Rupture rotator cuff shoulder	30	40	50	60	70
Painful shoulder	40	49	57	66	75

**Orthopaedic surgery**—The VAS in the orthopaedic study was identical to vascular surgery.

The Generic Orthopaedic Access Criteria tool was a self-explanatory tool with specified anchor points for point allocation within each of the five criteria below.

Factors for this were additive, and the definitions are as follows:

- **Natural history of the problem:** Condition is life/limb threatening or likely to deteriorate without treatment.
- **Pain:** Severity of pain in four categories from none to severe with explicit descriptors in each category.
- **Functional limitation:** Severity of physical disability in four categories from none to severe with explicit descriptors in each category.

- **Social limitation:** Severity of social disability in four categories from none to severe with explicit descriptors in each category.
- **Potential to benefit from surgery:** Five categories from no benefit to return to near normal likely.

The ISS consisted of a five-point Likert scale with tabulated score ranges determined by clinical consensus.

**Data collection**—The vignettes were presented to each surgeon using a custom designed computer programme on a Delphi platform integrated with two Microsoft Access databases. One database contained the patient vignettes whereas the second was for data storage.

A research assistant demonstrated operation of the computer software to each surgeon. Vignette and associated prioritisation tool pairs were then presented to the surgeon as separate screens scrolling one at a time on the computer. The initial segment of the evaluation comprised of introductory screens explaining the purpose of the study and its format.

The presentation order of the vignettes and tools was randomised, and clinicians were not able to go back and check their previous answers. The research assistant was present initially to help with use of the computer software, but they did not assist with the meanings of the questions. At the completion of the vignettes, each clinician was briefly interviewed regarding the use of the prioritisation tools and any comments they had on CPAC tools generally.

Test-retest reliability was assessed on a second round of data. The data were collected via a web-based platform that mirrored the interface of the original computer programme. The shift to a web-based programme was made after feedback from participant surgeons indicated the need for a more flexible mode of access over time.

**Sample size**—Preliminary calculations were made to estimate the sample size and the number of vignettes required, to achieve sufficient power for each specialty. For example, vascular surgery (as the smallest specialty of the three) had 35 surgeons nationally; with a design involving 30 surgeons and 60 vignettes, the 95% confidence intervals for the intra-class correlation was approximately plus or minus 0.15 for true values of the intra-class correlation within the range 0.4 to 0.7.

For the second round, power calculations confirmed that fewer data points were required. Accordingly, the number of general surgeons participating was reduced to 30 while maintaining 32 vignettes. Conversely, the number of surgeons remained constant for vascular and orthopaedic surgery but the number of vignettes was reduced to 39 and 40 respectively.

**Statistical analyses**—Analysis of data from each specialty was initially undertaken using a standard frequentist approach to calculating the intraclass correlation coefficients (ICC). This method uses two-way analysis of variance, where the effects of subjects are important in the analysis and the unit of reliability of interest is the individual score.<sup>15</sup> The form of the equation chosen for general and orthopaedic surgery was that for a random sample of surgeons where the individual surgeon's reliability was of interest.

The design of the vascular surgery study was that of each vignette being rated by the same subjects, who are the only judges of interest. This was because the intention was to sample the whole population of vascular surgeons in New Zealand. Therefore the form chosen for vascular surgery was that for a fixed sample of surgeons where the individual surgeon's reliability was of interest.

The initial analyses were performed using SAS version 8.2 (SAS Institute Inc, Cary, NC, USA) and a macro that calculates reliabilities for intraclass correlations (INTRACC.SAS).

Test-retest analysis was conducted using a mixed model. Surgeon and vignette were incorporated as random effects to represent that we were not interested in the particular surgeons and vignettes per se but wanted to treat them as a random sample of all surgeons and patients. Time was also included as a random effect as we were interested in the difference between any occasion the tools were applied for prioritisation.

Detailed results from these and further model-based statistical analyses for each specialty will be reported in a separate article.

## Results

For general surgery, the total population for publicly operating surgeons was 99 surgeons, however six surgeons who were used to develop the vignettes were



excluded from this sample. Of those eligible for inclusion, 31 refused to participate and a sample of 60 was selected from the remainder using a random number generator. The final response rate for general surgery was therefore 67% (62/93). For vascular surgery, 33 surgeons were eligible for inclusion, 85% (28/33) of these agreed to participate. In orthopaedic surgery, the total population for publicly operating surgeons was 153; of these, 62 were randomly selected to participate in the study and 65% (40/62) of these agreed to participate.

Within the group of those who had (from the outset) agreed to participate, completion rates were high with 60/60 (100%) of the general surgeons, 26/28 (93%) of the vascular surgeons, and 38/40 (95%) of the orthopaedic surgeons completing the evaluation. There did not appear to be any differences in non-response across the regions, and the primary reason cited for initially refusing to participate or subsequent failure to complete was 'lack of time'.

The characteristics of participants versus non-participants were assessed using information gleaned from the vocational register for each specialty. No statistical differences between the two groups were found on a variety of measures including years in practice and geographical location. The majority of participants were computer literate and quickly learned how to navigate the programme. A small proportion had difficulties restoring their session to the correct starting point when completing their evaluation, but all technical problems encountered related to use of specific hardware and were able to be resolved over the phone or by email.

A looping problem in the software prototype meant that in the early phase of the study some participants initially scored a few additional scenarios. Almost without exception, those surgeons who started the evaluation completed it. Of the small number of surgeons that did not complete (or proceeded by zero-scoring selected vignettes), three individuals reported difficulties scoring specific conditions (primarily carotid stenosis) because they did not routinely encounter this condition in their current clinical practice. Missing or erroneous data was either omitted or taken into account in subsequent analyses.

Conclusion of the second round of data collection, which relied predominantly on telephone and email reminders to the participant surgeons proved to be more difficult. This was reflected in the lower response rates across specialties (general surgery 76.7%, vascular surgery 88.5%, and orthopaedic surgery 86.8%) and the longer time period to completion encountered during the retest.

## Discussion

Although priority tools are in use in several countries, the extent to which they have been formally validated or tested for inter-rater or intra-rater reliability varies. The Western Canada Waiting List Project has conducted reliability testing on priority scoring tools for general surgery, cataract surgery, as well as hip and knee replacement<sup>16</sup> however this work was on the basis of different tools and used on average only 11 surgeons and 6 videotaped vignettes.

While priority scoring tools are used in the United Kingdom, we are unaware of any reliability testing that may have been performed. Within the New Zealand context, Dennett et al<sup>7</sup> have explicitly drawn attention to the problem of inadequate prior testing of reliability; stating that (ideally) inter-rater reliability and tool validity should

have been established by those who developed the methods before they were introduced.

Therefore (to our knowledge) this study is the first within New Zealand to specifically address the issue of reliability among scoring tools (already in use) for prioritisation of surgery. As this was the first study of its type among surgeons in New Zealand, appropriate methodology had to be developed and applied by the principal investigators. Several strengths and limitations inherent in the methods used were identified throughout the course of the project.

This study of reliability was able to be conducted nationally with research assistants based in all three of the main centres (Auckland, Wellington, and Christchurch), whereas previous studies looking at different aspects of the clinical priority assessment criteria have been conducted regionally.<sup>8</sup>

The moderate-to-good response rates obtained in the first round of data collection was consistent with other studies that have examined medical practice using clinical vignettes.<sup>17-19</sup> However research based on the responses of surgeons is unusual; previous studies incorporating vignettes have usually involved general practitioners or psychiatrists.<sup>19</sup>

Studies based on computer-based vignettes are especially rare, and have been confined to primary care practitioners.<sup>17</sup> In this study, excellent first-round completion rates were achieved among surgeons from all three specialties across the country. Most surgeons were very positive about participation, with less than 1% of surgeons who initially agreed to participate subsequently expressing an objection in principle to any evaluation of CPAC tools. Importantly, randomisation of the order in which the tools were presented was also incorporated in the study design, thereby minimising the influence the first score could have on any subsequent scores.

The shift to a web-based platform for the second round, although more convenient for the participating surgeons resulted in an increase in the time it took to collect data. It also made follow-up of the surgeons who were slow to complete more challenging, possibly because of the lack of opportunity for face-to-face contact with a research assistant. It is difficult to know to what extent the lack of any specific incentive to complete or that this was the second time each surgeon had been asked to complete the evaluation, contributed to the lower completion rates in the second round.

In a recent randomised trial Gattelari and Ward<sup>20</sup> reported that surgeons offered an indirect incentive to participate in an Australian survey were actually significantly less likely to respond, and (as a group) were more tardy in their response. Contrary to expectation, offering donations to the surgeons (college) was counterproductive in enhancing response rate. In the present study, the only incentive offered to surgeons was a bottle of wine on completion (and prior to completion in the second round), however it was also clear from feedback from participating surgeons that the single overriding factor in delayed response or non-completion was (not surprisingly) 'time constraints'.

The present study is limited by certain factors. Firstly, it only uses vignettes, and despite their widespread use, there remains uncertainty and controversy about the degree to which vignettes reflect actual clinical practice.<sup>21</sup> Although Peabody et al<sup>12</sup> have suggested that vignettes are a useful and valid way of measuring physician

practice, for practical purposes it was necessary to restrict the size of each vignette (and the extent of detail that could be provided about each paper patient to the surgeon) so that the evaluation would not be too long. It is feasible that additional radiographs, clinical photographs, and videos of real patients would have made the cases more realistic, and may have added to the overall consistency of the scoring.

The two other main sources of bias in studies such as this arise from selection of the sample and non-response. Poor response rates have the potential to compromise the validity of postal surveys and other types of studies<sup>22</sup> and several techniques may need to be used to maximise response rates.<sup>23</sup>

Although this study had reasonable first-round completion rates, surgeons who chose not to participate did not do so randomly, therefore the potential effect of selection bias must be acknowledged. Selection bias is more likely to be problematic in the general and orthopaedic surgery arms of the study because of the greater proportion of surgeons who refused to participate or who were ineligible to participate.

While in practice all surgeons must use the scoring tools, it is possible that they may work better for those surgeons that support their use. In the worst case scenario, results obtained might suggest that inter-rater reliability was much higher than what would otherwise be found if all surgeons within a given specialty were included in the sample. As relatively little information was available on the characteristics of those surgeons who responded compared to those who did not, it is difficult to rule out the possibility that the sample may be a source of bias in the present study. Regardless of this limitation, some selection bias in a study of this nature is probably inevitable.

In general surgery, the second round response rate was slightly lower than for other specialties (76.7%), although this figure remains very comparable to what has been reported elsewhere in male physicians.<sup>24</sup> Ward et al demonstrated significant differences in response rates between male and female general practitioners in a study that looked at response-aiding strategies, with females (74% versus 63%) being more likely to respond to a survey. We note that our sample of surgeons, across all specialties, was predominantly male (99%).

This collaborative study indicates that evaluation of the reliability of CPAC tools currently in use in New Zealand is possible using a vignette-based approach. Study designs that allow for some ongoing face-to-face contact with surgeons, although potentially more expensive may ensure better rates of response and completion especially where re-test reliability is to be assessed.

**Author information:** Carolyn Doughty, Research Fellow;<sup>1</sup> Andrew D MacCormick, Surgical Registrar and Lecturer;<sup>2</sup> Justin Roake, Professor and Head of Department;<sup>3</sup> John Fraser, CPAC Evaluation Project Manager;<sup>4</sup> Phil Hider, Public Health Physician;<sup>1</sup> Ray Kirk, Director, New Zealand Health Technology Assessment Unit;<sup>1</sup> Bryan Parry, Professor and Head of Division;<sup>2</sup> Andre van Rij, Professor and Head of Section;<sup>5</sup> Jean-Claude Theis, Associate Professor and Head of Section;<sup>6</sup>

<sup>1,3</sup>Departments of Public Health and General Practice<sup>1</sup> and Surgery,<sup>3</sup> Christchurch School of Medicine and Health Sciences, University of Otago, Christchurch

<sup>2</sup>Department of Surgery, Faculty of Medicine and Health Sciences, University of Auckland, Auckland

<sup>4</sup>Health Services Research Centre, School of Government, Victoria University of Wellington, Wellington

<sup>5,6</sup>Departments of Surgery<sup>5</sup> and Orthopaedic Surgery,<sup>6</sup> Dunedin School of Medicine, University of Otago, Dunedin

**Acknowledgements:** This study was carried out as part of the CPAC Consortium Evaluation Project—funded by the Ministry of Health. We acknowledge the additional contributions of Patrick Graham, Sally Hamilton, Monir Hossain, Alexandra Macmillan, Roshan Perrera, Elizabeth Robinson, and Nicola Trolove.

**Correspondence:** Dr Carolyn Doughty, Department of Public Health and General Practice, Christchurch School of Medicine and Health Sciences, PO Box 4345, Christchurch. Fax: (03) 364 3697; email: [carolyn.doughty@chmeds.ac.nz](mailto:carolyn.doughty@chmeds.ac.nz)

## References:

1. Hadorn DC, Holmes AC. The New Zealand priority criteria project. Part 1: overview. *BMJ*. 1997; 314:131–4.
2. Fraser G, Alley P, Morris R. Waiting lists and waiting times: their nature and management. Wellington: National Advisory Committee in Core Health and Disability Support Services; 1993.
3. Hadorn DC, Holmes AC. The New Zealand priority criteria project. Part 2: coronary artery bypass graft surgery. *BMJ*. 1997;314:135–8.
4. Seddon ME, French JK, Amos DJ, et al. Waiting times and prioritization for coronary artery bypass surgery in New Zealand. *Heart*. 1999;81:586–92.
5. New Zealand Ministry of Health. Purchasing for your health: 1995/96. Wellington: Ministry of Health; 1997.
6. Adams J. The national booking list: sliced bread or shuffled deckchairs? New Zealand Medical Association Newsletter. 1998;1:2.
7. Dennett ER, Kipping RR, Parry BR, Windsor J. Priority access criteria for elective cholecystectomy: a comparison of three scoring methods. *N Z Med J*. 1998;111:231–3.
8. Dennett ER, Parry BR. Generic surgical priority criteria scoring system: the clinical reality. *N Z Med J*. 1998; 111:163–6.
9. Halliwell T. How fair is cataract prioritisation? *N Z Med J*. 1998;111:405–7.
10. Derrett S. Surgical prioritisation and rationing: some recent changes. *New Zealand Bioethics Journal*. 2001;2:3–6.
11. McCormick A. Identification of criteria for the prioritisation of patients for elective general surgery. *J Health Serv Res Policy*. 2004;9:28–33.
12. Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA*. 2000;283:1715–22.
13. Cochran W, Cox G. Experimental designs. 2nd ed. New York: John Wiley and Sons; 1957.
14. McCormick A, Plank L, Robinson E, Parry B. Prioritising patients for elective surgery: clinical judgement summarised by a linear analogue scale. *Aust N Z J Surgery*. 2002; 72:613.
15. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–8.
16. Western Canada Waiting List Project. From chaos to order: making sense of waiting lists in Canada. Final report, March 2001. Edmonton: WCWL; 2001.

17. Walton RT, Gierl C, Yudkin P, et al. Evaluation of computer support for prescribing (CAPSULE) using simulated cases. *BMJ*. 1997;315:791–5.
18. Quine MA, Bell GD, McCloy RF, et al. Appropriate use of upper gastrointestinal endoscopy—a prospective audit. Steering Group of the Upper Gastrointestinal Endoscopy Audit Committee. *Gut* 1994;35:1209–14.
19. Jorm AF, Korten AE, Jacomb PA, et al. Beliefs about the helpfulness of interventions for mental disorders: a comparison of general practitioners, psychiatrists and clinical psychologists. *Aust N Z J Psychiatry*. 1997;31:844–51.
20. Gattellari M, Ward JE. Will donations to their learned college increase surgeons' participation in surveys? A randomized trial. *J Clin Epidemiol*. 2001;54:645–9.
21. Jones TV, Gerrity MS, Earp J. Written case simulations: do they predict physicians' behavior? *J Clin Epidemiol*. 1990;43:805–15.
22. Evans SJ. Good surveys guide. *BMJ*. 1991;302:302–3.
23. Etter JF, Perneger TV. Analysis of non-response bias in a mailed health survey. *J Clin Epidemiol*. 1997;50:1123–8.
24. Ward J, Bruce T, Holt P, et al. Labour-saving strategies to maintain survey response rates: a randomised trial. *Aust N Z J Public Health*. 1998; 22(3 Suppl):394–6.