Measuring the effectiveness of cataract surgery: the reliability and validity of a visual function outcomes instrument

D J Lawrence, C Brogan, L Benjamin, D Pickard, S Stewart-Brown

Abstract

Aims—To assess test-retest reliability and validity of the “TyPE” patient self assessed visual function questionnaire, as part of a study in two hospitals measuring the effectiveness of cataract surgery. The American TyPE questionnaire had minor adaptations made for use in Britain.

Methods—Test-retest reliability was assessed on 63 out of 378 adult cataract surgery patients in the study, using Spearman correlation coefficients and kappa coefficients of agreement. “Construct” validity was evaluated by comparing the association between changes in visual function questionnaire scores after surgery, with patients’ perception of change in visual function obtained by independent interview of 24 patients.

Results—The TyPE questionnaire items showed very good test-retest reliability. Average Spearman and kappa coefficients for 39 patients from hospital 1 were 0.93 and 0.84 respectively. Spearman and kappa coefficients of 0.9 and 0.81 were obtained for those nine patients in hospital 2 where both the test and retest questionnaires were filled in by the same people. However, for the 15 patients from hospital 2, where the questionnaire was filled in by different people in the retest, reliability was less good: the Spearman coefficients were still high, average 0.72, but the kappa coefficients were poor, 0.27. Good construct validity was exhibited, with a correlation of 0.79 between change in distance vision score from the questionnaires and the independent interview.

Conclusions—The adapted TyPE questionnaire is both very reliable and has good construct validity. The kappa coefficient should be used wherever possible to evaluate reliability. The test-retest reliability and validity and practicability of other visual function questionnaires have not been assessed adequately, and further development should be carried out of all such questionnaires, so that they may be introduced into routine clinical care.

(Br J Ophthalmol 1999;83:66-70)

unmet need for surgery.¹ There are still large variations between districts in surgery rates which have not been explained in terms of variations of need, and which indicate possible differences in severity thresholds for surgery.¹⁻⁵ Achieving equitable and optimal surgical rates depends on establishing criteria for which individuals should be operated on.⁶⁻⁸

Though there is much evidence from clinical observation that cataract surgery improves vision, only visual acuity has been systematically used as a quantitative assessment and outcome measure in routine clinical care.⁹ However, there is much evidence that Snellen visual acuity measurement does not validly measure impairment in everyday visual function caused by cataract.^{6-7, 10-12} For instance, it does not adequately measure the effects of glare in everyday situations.

Because cataract surgery produces changes in quality of life rather than reducing mortality, patients’ own assessments of visual function and health related quality of life are most appropriate for outcomes measurement.¹⁻¹³ However, until now, there has rarely been systematic measurement of patients’ visual function in cataract patients.¹²⁻¹⁴⁻¹⁵

There is no definitive ophthalmological test to determine severity thresholds for operating.⁶ American and British ophthalmological authorities’ guidelines do not consider it useful to define specific levels of visual acuity as sole cut off points for surgery in preoperative assessment.⁶⁻⁷ According to these guidelines, surgery should be undertaken when the benefits to the patient’s quality of life outweigh the risks, and this depends on the extent to which a cataract is interfering with the patient’s visual function.⁶⁻⁷

Some patients benefit less from cataract surgery than others, and the patient characteristics and other factors which are associated with this are only beginning to be researched.¹⁶ Also, although second eye surgery is known to benefit some patients¹⁷⁻¹⁸ still too little is known about what the relative visual function benefits are from second eye compared with first eye surgery.¹⁴⁻¹⁷

CATARACT SURGERY OUTCOMES USING VISUAL FUNCTION MEASUREMENT

To begin to resolve the above issues, it is essential to measure outcomes by asking patients with cataract about their visual function, using reliable and valid questionnaires.¹⁹

Reliability is the extent to which a vision test item, such as a Snellen visual acuity measurement, or a patient assessed visual function

Department of Public Health,
Buckinghamshire Health Authority,
Aylesbury, HP19 3ET
D J Lawrence
D Pickard

Directorate of Public Health, Anglia and Oxford Regional Office, NHS Executive, Milton Keynes MK14 6QP
C Brogan

Department of Ophthalmology, Stoke Mandeville Hospital, Aylesbury, HP21 8AL
L Benjamin

Health Services Research Unit, University of Oxford, Institute of Health Sciences, Oxford OX3 7LF
S Stewart-Brown

Correspondence to:
David Lawrence, Ealing, Hammersmith and Hounslow Health Authority, 1 Armstrong Way, Southall, London UB2 4SA.

Accepted for publication
31 July 1998

Though surgery for cataract is one of the most common elective operations in the NHS, there are still important unresolved issues about the procedure.

There are difficulties in estimating prevalence of cataract,¹⁻² and assessing population

questionnaire item, gives the same measurement or score when two or more measurements are made, while the patient's true visual function has not changed.

Validity is the extent to which a measurement instrument or test really measures what it is supposed to. It is difficult to evaluate the validity of visual function measurement instruments because visual function is a complex concept and there is no obvious "gold standard" (technically called "criterion" validity) for measuring it. However, it is possible to estimate "construct" validity by measuring the degree of agreement between two or more different ways of measuring visual function.

While research to develop patient assessed visual function and quality of life measurement in cataracts has been carried out in the USA,^{12 13 15} these instruments are still not widely used there.¹³ In Britain there has only been a little research on the development of such instruments,^{5 17 19} and they are not yet being used in routine care.

Our study was designed to test the reliability and validity of an established patient assessed visual function questionnaire, and to use it to measure the long term effectiveness of cataract surgery. In this article we report on the test-retest reliability, and validity of the questionnaire. The results of using this questionnaire to measure visual function and quality of life for up to 2 years after surgery are being reported in a separate paper, and further analyses of subsets of the data are in progress.

Methods

During the period March 1993 to August 1994, 378 patients being assessed for cataract surgery in two hospitals were entered into the study, and completed post operation questionnaires. Clinical data were collected using a protocol developed by a group of ophthalmologists in Buckinghamshire.¹⁹

Patients' visual function was measured using the American "TyPE" cataract visual function measurement questionnaire Version 1,²⁰ adapted by us for use in Britain. This questionnaire was developed in America after a comprehensive literature review, and used expert opinion of ophthalmologists, and a consensus process involving patient care committees.²⁰ There are questions on overall ability to see; how much vision hinders or limits usual daily activities, recognising people or objects across the street, reading, knitting, watching TV, and driving; and how much glare hinders day to day activities, reading, and driving under dazzling conditions. It uses five point scales of disability, from "not at all" to "totally disabled." We used especially large and clear fonts (a reduced example is shown in the appendix).

Our adaptations of the TyPE Version 1 changed "store" to "shop," removed "sewing" from the "reading" question, and created a separate question about "knitting/sewing." We also added a question on watching television. (Version 3 of the American TyPE, issued after this study was completed, has the same

changes to the "reading/sewing" questions which we have, and has also added a "watching TV" question.)

Following Mangioni *et al*'s methods,¹⁵ we empirically constructed four visual ability dimensions by combining related visual function questions. The distance vision function dimension combined the items "recognising people" and "watching TV"; the near vision dimension combined "reading price labels" and "reading magazines"; the glare vision dimension combined "usual activities glare," "glare when walking," and "glare when reading shiny paper"; and the driving vision dimension combined "day driving," "night driving," and "glare when driving."

RELIABILITY OF THE VISUAL FUNCTION QUESTIONNAIRE

In our cataract outcomes study, very nearly all patients in hospital 1 filled in the questionnaire twice preoperatively. The first time was at outpatients, as part of the routine preoperation assessment, and the second time was in hospital just before surgery. Test-retest reliability was assessed using as a sample the 39 patient visual function questionnaires where the interval between filling in the first and second preoperation questionnaires was less than 30 days. This 30 day limit between test and retest was chosen so that visual ability would not have changed significantly in the interval. Patients in hospital 2 filled in the questionnaire only once preoperatively. Therefore, for the reliability test-retest in this hospital, a random sample of 24 patients was asked to fill in the questionnaire again before operation, at home 4 days after filling in the first.

In both hospitals, the questionnaires were filled in either by the patients themselves, or with help from a friend or a health professional. We recorded which occurred. The data from both hospitals were analysed in two groups, according to whether the respondent (whether the patient responded alone or with help) remained the same from test to retest, or whether the respondent was different in the retest compared with the test.

Virtually all the test and retest questionnaires at hospital 1 were filled in at the hospital outpatient clinic. At hospital 2, all the test questionnaires were filled in at the outpatient clinic, whereas all the retest questionnaires were filled in at home.

The data were entered into a database (Microsoft Access), and statistical analyses were carried out using the Statistical Package for the Social Sciences (SPSS), Windows version 6.1. Spearman non-parametric correlation and Cohen's kappa coefficient of agreement for categorical data²¹ were used to measure the degree of agreement between test and retest questionnaire scores for all the visual ability variables.

VALIDITY OF THE VISUAL FUNCTION QUESTIONNAIRE

Construct validity of the questionnaire for measuring change in visual function as a result of surgery was evaluated by comparing the

Table 1 Retest reliability of visual function questionnaire items, hospital 1

Questionnaire visual function items	Hospital 1, groups 1 and 2 together		
	Number	Spearman correlation	Kappa coefficient
Rating of vision	36	0.96	0.80
Vision hinders usual activity	34	0.86	0.77
Vision hinders recognising people	36	0.93	0.93
Vision hinders reading price labels	36	0.89	0.85
Vision hinders reading magazine	36	0.96	0.86
Vision hinders knitting	37	0.90	0.89
Vision hinders watching TV	36	0.95	0.81
Vision hinders daytime driving	37	0.94	0.81
Vision hinders night-time driving	37	0.92	0.84
Glare hinders usual activities	36	0.93	0.86
Glare hinders reading shiny paper	36	0.93	0.78
Glare hinders driving towards the sun	37	0.98	0.90
Glare hinders walking outside on a sunny day	36	0.97	0.86
Overall average	36.2	0.93	0.84

change as measured by the visual function questionnaire with patients' perception of their change in visual function obtained by independent interview.

A random sample of 24 patients entering the study during the first 3 months in hospital 1 were interviewed by DP using "open ended" questions. Each was asked how much their visual function had changed as a result of surgery. The narrative responses for each interviewed patient were then assessed as falling into one of four categories of change in visual function—"worse," no change," "some what better," and "much better," by DL, without knowing what the changes in the visual function questionnaire scores were (that is, the assessment was "masked"). The extent of agreement between this categorised variable and the change in visual function scores from the questionnaires was tested using Spearman non-parametric correlation (kappa coefficients could not be calculated). The data were entered into Microsoft Access, and statistical analyses carried out using SPSS.

Results

RELIABILITY

The visual function questionnaire items overall showed very good test-retest reliability. For hospital 1, visual function test-retest Spearman coefficients averaged 0.93, with all above 0.85 and all but two above 0.90. Their kappa coefficients

averaged 0.84, and only two items—"vision hinders usual activities," 0.77, and "glare hinders reading shiny paper," 0.78—had coefficients below 0.80 (Table 1).

For hospital 2, where the retest questionnaire responses were filled in by the same person (or people) who filled in the test questionnaire (group 1 patients), the results were of similar reliability to those from hospital 1. Visual function test-retest Spearman coefficients averaged 0.90, with all above 0.70 and all but two above 0.80 (Table 2). Their kappa coefficients could only be calculated for three items, but for these they averaged 0.81, of similar order to those from hospital 1 (Table 2).

By contrast, where the retest questionnaires were filled in at home by a different person or people from the test questionnaire filled in at the hospital (group 2 patients), the reliability results were poorer. The average Spearman correlation coefficient was still quite high, 0.72, but the kappa coefficients, where they could be calculated, were low, average 0.27 (Table 2).

To calculate the kappa coefficient, the same categories have to be present in both test and retest. Because of small numbers, in hospital 2 for groups 1 and 2 separately, there were many variables in which some categories of response did not occur and the kappa coefficient could not be calculated.

VALIDITY

Patients' perception from the interviews of how their visual function had changed overall, correlated very well, 0.79, with the changes in the "distance" visual function dimension scores calculated from the questionnaires, and quite well, 0.62, with a composite "overall vision" score calculated from the original questionnaire data items (Table 3).

The correlations with "near," 0.48, and "glare," 0.43, visual function dimension scores, were only moderate. This might be expected if overall vision perception is more connected with distance vision than near vision or glare disability. There were insufficient car drivers in the interview sample to consider the "driving visual function" dimension.

Table 2 Retest reliability of visual function questionnaire items, hospital 2

Questionnaire visual function items	Hospital 2, group 1*			Hospital 2, group 2†		
	Number	Spearman correlation	Kappa coefficient	Number	Spearman correlation	Kappa coefficient
Rating of vision	8	0.89	0.80	15	0.77	NA
Vision hinders usual activity	8	0.93	0.80	14	0.66	NA
Vision hinders recognising people	8	0.95	NA	15	0.75	0.23
Vision hinders reading price labels	8	0.93	NA	15	0.66	0.27
Vision hinders reading magazine	8	0.93	NA	15	0.75	NA
Vision hinders knitting	9	0.74	NA	15	0.76	NA
Vision hinders watching TV	7	0.76	NA	15	0.56	NA
Vision hinders daytime driving	9	0.99	NA	15	0.7	NA
Vision hinders night-time driving	9	0.97	0.84	15	0.76	NA
Glare hinders usual activities	9	0.82	NA	15	0.68	0.23
Glare hinders reading shiny paper	9	0.90	NA	15	0.70	NA
Glare hinders driving towards the sun	9	0.97	NA	15	0.76	NA
Glare hinders walking outside on a sunny day	9	0.93	NA	15	0.83	0.33
Overall average	8.5	0.90	0.81	14.9	0.72	0.27

*Group 1 patients: those for whom the person(s) who responded in the retest was/were the same as in the test (see text).

†Group 2 patients: those for whom the person(s) who responded in the retest was/were not the same as in the test (see text).

NA = kappa coefficient could not be calculated because of unequal categories between the test and retest results for this data item owing to small numbers.

Table 3 Correlations of patients' perception of overall change in visual function (from the interviews), with changes in visual function dimension scores (from the questionnaires)

Visual function dimension	No of patients	Spearman correlation coefficient	Statistical significance (two tail)
Distance vision function	23	0.79	$p < 0.0001$
Near vision function	22	0.48	$p = 0.02$
Glare visual function	22	0.43	$p = 0.03$
Composite of visual function data item scores	21	0.62	$p = 0.03$

Discussion

RELIABILITY

These results show that the visual function questionnaire we used has very good test-retest reliability, with very high Spearman and kappa coefficients. The kappa coefficient of agreement is a much better measure of test-retest reliability than the Spearman correlation coefficient, because the kappa coefficient allows for agreement which would be expected by chance between the patients' test and retest scores.²¹ Perfect agreement is shown by kappa = 1, whereas if there is no agreement, other than that expected by chance, kappa = 0. Values of 0.75 and above are considered excellent agreement, with those between 0.4 and 0.74 moderate agreement.²²

The results for hospital 2 (Table 2) illustrate this. The Spearman coefficients are high for both groups 1 and 2, 0.81 and 0.72 respectively, but the average value of the kappa coefficients for group 2, 0.27, is much lower than for group 1, 0.81.

The test-retest reliability results reported here are as good as those reported from other work on patient assessed visual function questionnaires by Sanderson⁵ and Mangione *et al.*¹⁵ However, those two studies only reported Spearman correlation coefficients, and we have shown that these can give too optimistic results compared with use of the kappa coefficient. This is the first time that the kappa coefficient has been used to test reliability in these types of outcome instruments. The VF-14, another important visual function questionnaire, has been reported as being reliable by American¹³ and English¹⁷ studies, but no test-retest reliability results have so far been published by them. The American study . . . "elected not to assess the test-retest reliability of the VF-14 before surgery, but . . . to administer it at 4 and 12 months after surgery."²³ Because changes in vision are possible during this post operation period, and one purpose of these questionnaires is to detect such changes, we suggest that only preoperation test-retest reliability, with a short interval between tests, is valid for reliability testing. The English study did not assess test-retest reliability.

As far as we know this is the first study to use self administered questionnaires; other studies having used telephone or interviewer responses.^{12 15 17} In this study, patients who could not fill the questionnaire unaided were helped by an accompanying person or health professional, and we monitored this. We have shown that the conditions under which questionnaires are administered can affect reliability for self administered questionnaires. The very good results for group 1 patients from hospital 2 indicate that provided question-

naires are filled in by the same person or people, reliability is very good even when the milieu changes.

VALIDITY

The results provide evidence of good construct validity of the questionnaire items as a measure of patients' perception of change in visual function as a result of cataract surgery (Spearman correlation of 0.79 with distance vision function, 0.62 with a composite measure). These results also show that the questionnaire items are sensitive to change in visual function. That is, where patients report in interview that there is change in visual function due to operation, this is indicated by changes in questionnaire item scores.

Other studies have reported similar results for construct validity, though they have called it criterion validity (see below). These studies correlated questionnaire scores with other measures of patient assessed visual function, and not with change in scores before and after operation as we have done, but obtained similar levels of Spearman correlations (around 0.6).

No reported study, including this one, has evaluated the criterion validity of vision tests in measuring visual impairment from cataracts, using direct comparisons with patients' real ability to read, recognise people, or perform other everyday tasks which are vision dependent. Some studies have reported measuring criterion validity^{13 15 17} but did not in fact do so. These studies variously tested the visual function questionnaires against visual acuity,^{13 17} global self rating,^{13 15} patient satisfaction with vision,¹³ percentage binocular vision loss,¹⁵ and vision related quality of life.¹⁷ None of these is a direct measure of specific visual function dimensions.

In conclusion, this study provides important evidence that patient assessed visual function questionnaires have good test-retest reliability and construct validity for preoperative assessment and for measuring the impact of cataract surgery. Their potential for use in routine ophthalmic practice should be explored. Self assessed instruments are potentially more cost effective than questionnaires which require an interviewer.

Further research is required on the relative test-retest reliability of the various visual function questionnaires now available against each other and against visual acuity, using the kappa coefficient as a measure of agreement.

The criterion validity of visual acuity and visual function questionnaires needs to be established by testing them against patients' real ability to carry out visual function tasks at near and distance vision ranges. Further research is also necessary to examine how these visual function questionnaires compare with questionnaires which directly ask about *changes* in visual function rather than just ask about visual function status at different times.

We would like to thank all the patients and hospital staff who took part in the study, our research assistant Penny Drewett, the Bucks Association for the Blind volunteers who helped in outpatients, Angela Coulter, and Crispin Jenkinson for guidance, and Helen Doll for statistical advice. We would also

like to thank Dr Larry Chambers, Dr JA Muir Gray, and Dr Alison Hill for encouragement and advice. Buckinghamshire Health Authority gave much support, and the former Oxford Regional Health Authority provided funding.

- 1 Williams MH, Frankel SJ, Nanchahal K, et al. *Health care needs assessment: the epidemiology based needs assessment reviews. 11. Cataract surgery*. Winchester: Wessex Institute of Public Health Medicine, 1994.
- 2 Wormald RP, Wright LA, Courtney P, et al. Visual problems in the elderly population and implications for services. *BMJ* 1992;304:1226-9.
- 3 Mordue A, Parkin DW, Baxter C, et al. Thresholds for treatment in cataract surgery. *J Public Health Medicine* 1994;16:393-8.
- 4 Sanders D, Coulter A, McPherson K. *Variations in hospital admission rates: a review of the literature*. Kings Fund project paper 79. London: Kings Fund, 1989.
- 5 Sanderson HF. *Thresholds in cataract surgery: an investigation into the ease of admission for lens surgery in three English cities*. MSc Thesis. London: University of London, 1978.
- 6 Royal College of Ophthalmologists. *Guidelines for cataract surgery*. London: Royal College of Ophthalmologists, 1995.
- 7 Cataract management guideline panel. *Cataract in adults: management of functional impairment. Clinical practice guideline Number 4*. Rockville, MD: US Department of Health and Human Services Agency for Health Care Policy and Research, 1993.
- 8 Lee PP, Kamberg KJ, Hilbourne LH, et al. *Cataract surgery: a literature review and ratings of appropriateness and cruciality*. Santa Monica: Rand, 1993.
- 9 Courtney P. The national cataract surgery survey: (i) Method and descriptive features. *Eye* 1992;6:87-92.
- 10 Gibson RA, Sanderson HF. Observer variation in ophthalmology. *Br J Ophthalmol* 1980;64:457-60.
- 11 Elliott DB, Hurst MA, Weatherill J. Comparing clinical tests of visual function in cataract with the patient's perceived visual disability. *Eye* 1990;4:712-17.
- 12 Legro MW. Quality of life and cataracts: a review of patient centered studies of cataract surgery outcomes. *Ophthalmic Surg* 1991;22:431-43.
- 13 Steinberg EP, Tielsch JM, Schein OD, et al. The VF14: an index of functional impairment in patients with cataract. *Arch Ophthalmol* 1994;112:630-8.
- 14 Javitt JC, Brenner MH, Curbow B, et al. Outcomes of cataract surgery: improvement in visual acuity and subjective visual function after surgery in the first, second, and both eyes. *Arch Ophthalmol* 1993;111:686-91.
- 15 Mangione CM, Phillips RS, Seddon JM, et al. Development of the 'activities of daily vision scale': a measure of visual functional status. *Med Care* 1992;30:1111-26.
- 16 Schein OD, Steinberg EP, Cassard SD, et al. Predictors of outcome in patients who underwent cataract surgery. *Ophthalmology* 1995;102:817-23.
- 17 Desai P, Reidy D, Minassian DC, et al. Gains from cataract surgery: visual function and quality of life. *Br J Ophthalmol* 1996;80:868-73.
- 18 Laidlaw A, Harrad R. Can second eye cataract extraction be justified? *Eye* 1993;7:680-6.
- 19 Brogan C. *Patient assessed cataract outcome pilot*. Thesis accepted for Part II of the examination for membership of the Faculty of Public Health Medicine. London: Royal College of Physicians, 1994.
- 20 Javitt JC, Ware JE. *Users' manual: cataract TyPE specification*. Washington: Wothten Center for Eye Care Research, Georgetown University, 1989.
- 21 Siegel S, Castellan NJ Jr. *Non-parametric statistics for the behavioral sciences*. New York: McGraw-Hill. International edition, 2nd ed, 1988:284-91.
- 22 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- 23 Cassard SD, Patrick DL, Damiano AM, et al. Reproducibility and responsiveness of the VF-14. *Arch Ophthalmol* 1995;113:1508-13.

Appendix

Please answer these questions based on your best vision with both eyes open and wearing glasses or contact lenses if you usually do.

11 How would you rate your vision? (how well do you see?) Please circle one number

Poor	1
Fair	2
Good	3
Very good	4
Excellent	5

12 How much does your vision hinder, limit, or disable you in each of the following activities?

Please circle one number on each line

Activity	Not at all	A little bit	Some	Quite a lot	Totally disabled	Don't do for other reasons
Your usual daily activities	1	2	3	4	5	0
Recognising people or objects across the street	1	2	3	4	5	0
Reading price labels in shops and supermarkets	1	2	3	4	5	0
Reading a magazine, newspaper or book	1	2	3	4	5	0
Knitting or sewing	1	2	3	4	5	0
Watching television	1	2	3	4	5	0
Daytime driving	1	2	3	4	5	0
Night-time driving	1	2	3	4	5	0

13 How much are you hindered, limited or disabled by glare (dazzling light) in each of the following activities?

Please circle one number on each line

Activity	Not at all	A little bit	Some	Quite a lot	Totally disabled	Don't do for other reasons
Your usual daily activities	1	2	3	4	5	0
Reading shiny paper (such as a magazine)	1	2	3	4	5	0
Driving towards the sun or oncoming headlights	1	2	3	4	5	0
Walking outside on a sunny day	1	2	3	4	5	0

14 Who filled in this form? Please circle one number

I filled it out with no help	1
I filled it out with help from family and friends	2
I filled it out with help from a nurse or doctor or Bucks Association for Blind (BAB) volunteer	3
Family and friends filled it out	4
A nurse or doctor or BAB volunteer filled it out	5

15 Have you had a recent illness, injury, or emotional upset that has affected how you answer these questions?

Yes	1
No	2

Any comments