

This is the authors' final version, post peer-review, of an article published in *British Journal of Ophthalmology* 2014;98:464-8. The definitive version is available from bj.o.bmj.com

**Variation of Clinical Outcomes Used in Glaucoma Randomised Controlled Trials:
A Systematic Review**

Authors: Rehab Ismail, MSc, FRCS (Glasg), ¹ Augusto Azuara-Blanco, PhD, FRCS(Ed), FRCOphth ² and Craig R. Ramsay, PhD, PGDip ¹

Institution: ¹ RI, PhD Student, Health Services Research Unit, University of Aberdeen, Health Sciences Building, Foresterhill, Aberdeen, AB25 2ZD

² AAB, Clinical Professor of Ophthalmology & Honorary Consultant Ophthalmologist, Centre for Vision and Vascular Science, Queen's University Belfast, Institute of Clinical Science - Block A, Royal Victoria Hospital, Grosvenor Road, Belfast, BT12 6BA

¹ CR, Programme Director & Professor, Health Services Research Unit, University of Aberdeen, Health Sciences Building, Foresterhill, Aberdeen, AB25 2ZD

Corresponding author: Rehab Ismail

Tel: +44 (0)1224 438408

Fax: +44 (0)1224 438165

Email: r01rai11@abdn.ac.uk

Key words: outcomes, heterogeneity, trials, measures, glaucoma

Word count: 2496

Abstract

Purpose

In randomised clinical trials (RCTs) the selection of appropriate outcomes is crucial to the assessment of whether one intervention is better than another. The purpose of this review is to identify different clinical outcomes reported in glaucoma trials.

Methods

We conducted a systematic review of glaucoma RCTs. A sample or selection of glaucoma trials were included bounded by a time frame (between 2006 and March 2012). Only studies in English language were considered. All clinical measured and reported outcomes were included. The possible variations of clinical outcomes were defined prior to data analysis. Information on reported clinical outcomes was tabulated and analysed using descriptive statistics. Other data recorded included type of intervention and glaucoma, duration of the study, defined primary outcomes, and outcomes used for sample size calculation, if nominated.

Results

The search strategy identified 4323 potentially relevant abstracts. There were 315 publications retrieved, of which 233 RCTs were included. A total of 967 clinical measures were reported. There were large variations in the definitions used to describe different outcomes and their measures. Intraocular pressure (IOP) was the most commonly reported outcome (used in 201 RCTs, 86%) with a total of 422 measures (44%). Safety outcomes were commonly reported, in 145 RCTs (62%) whereas visual field outcomes were utilised in 38 RCTs (16%).

Conclusions

There is a large variation in the reporting of clinical outcomes in glaucoma RCTs. This lack of standardisation may impair the ability to evaluate the evidence of glaucoma interventions.

Introduction

Randomised controlled trials (RCTs) represent the gold standard for gathering evidence about treatment effectiveness. Clinical trials seek to evaluate whether an intervention is effective and safe by comparing the effects of different alternatives on outcomes that are chosen to identify its benefits and harms. There is recognition that insufficient attention has been paid to the selection of proper outcomes to be used in clinical trials. Outcomes need to be relevant to patients, clinicians and policy-makers if the research is to influence practice.(1) A variety of outcomes can be selected for clinical trials, and these can be measured and reported in different ways. Problems arise if the selection, measurement and reporting of outcomes are inconsistent across clinical trials. First, important outcomes can be overlooked. If trialists are not aware of the importance of outcomes, it is likely that factors relating to the conduct of the trial, such as sample size, will determine which outcomes are measured.(2) A lack of standardisation may also lead to inefficient trials that involve multiple outcomes, larger than required sample sizes and multiple interpretations of results.(3) Secondly, meta-analysis will be difficult if different outcomes have been used. Even if the same outcomes are selected, they may be measured and analysed in different ways, and this too can impair the ability to synthesise their results. Finally, the third problem is selective outcome reporting bias (SORB) defined as the results-based selection for publication of a subset of the original measured outcome variables (4); in the absence of a set of core outcomes, trialists may decide to omit certain results from the final report.(2) Missing outcome data can affect systematic review in two ways. Publication bias, where a study is not published on the basis of its results, can lead to bias in the analysis of a particular outcome. Whereas, in a published study that has been identified by the reviewer, SORB can arise if the outcome had been measured and analysed but not reported.(5) The selection of appropriate outcomes in a trial, therefore, is crucial to the assessment of whether one intervention is better than another. One way to address these issues is the development of agreed standardised sets of outcomes, known as 'core outcome sets' (COS).(6) Some work

has been undertaken to identify and validate the different patient reported outcome measures used in glaucoma trials. This previous effort provided a framework (and critique of existing tools) for selecting the core patient reported outcomes in glaucoma studies.(7) However, no attempt has been undertaken on the standardisation of the clinical outcome measures in glaucoma RCTs. Therefore, as a first step towards this goal, we aim to identify the breadth of clinical outcomes reported in glaucoma RCTs.

Methods

We conducted a systematic review of RCTs in glaucoma between January 2006 and March 2012. The reason of the time frame was that we were aiming to retrieve approximately 200 RCTs and we chose to start our search from the most recent ones. The n=200 is an arbitrary number that we proposed would be sufficient to identify the variation on outcomes reported in glaucoma RCTs. We searched for RCTs in glaucoma published in English only due to resources constraints; however there were no restrictions on the population studied, applied interventions, phases of the trials, or types of glaucoma.

The following electronic databases were searched: Ovid MEDLINE (R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE (R) from January 2006 to first week of March 2012. A sensitive search strategy with both controlled subject headings and text terms relating to glaucoma and randomised controlled trials was run. The abstracts were graded as “included”, “possible” or “excluded”. Abstracts were excluded if they were not evaluating glaucoma patients and/or the study design was not an RCT or not in English language. The investigators planned to identify further relevant RCTs (original studies) from references of the retrieved articles. All abstracts were reviewed independently by one reviewer (RI). Another senior investigator (AAB) reviewed approximately 10% of all abstracts (400 consecutive abstracts) to evaluate agreement. This was a pragmatic approach that would have resulted in a larger duplicate grading if the agreement had been suboptimal. Regarding the assessment of eligibility of RCTs full text copies were assessed by one reviewer (RI). A second masked senior reviewer (AAB) reviewed 10 consecutive papers.

All clinical outcomes and variations of reporting were included in the analysis. We collected information on population, duration of follow up (months), sample size, type of intervention (medical, surgical, laser, others) and outcomes. We recorded which outcomes were considered to be primary or secondary and documented methods used to measure them, by whom, and when they were measured and analysed. Other types of outcomes such as patient reported outcome measures (PROMs), pharmacokinetic evaluation as well as economic outcomes were excluded. We agreed to contact trial authors as needed to clarify any details necessary to make a complete assessment of the study.

The data were initially abstracted on "Microsoft Office Access 2007" database and then statistics was run on Excel database. The data were tabulated in an excel sheet (Microsoft Corporation, Redmond, WA). Comprehensive pilot testing with two investigators (a clinician (RI) and a glaucoma expert (AAB)) was undertaken. As an effect, taxonomy of glaucoma outcomes was developed. Outcomes were grouped into one of eight main outcomes: intraocular pressure (IOP), visual field (VF), safety, haemodynamics, anatomical, aqueous humor (AH) dynamics, surgical-related, and persistence/ adherence to therapy.

Outcomes were further classified into different measures (specific measurement, specific metric or methods of aggregation). For example, for IOP, the pre-defined measures were: (1) *IOP level*, which was further subclassified into three possible measurements (Table), (2) *IOP change from baseline*, which was also subclassified according to the methodology of measurement, (3) *composite definition of success/failure using IOP as part of the definition*, where three different measures were identified and (4) *other descriptions of IOP outcomes* as in correlations / associations using IOP (Table).

The pre-defined measures for the VF were: (1) *evidence of progression of the VF*, which was further subdivided into different metrics according to the methodology of measurement, (2) *rate of VF progression/slope* and (3) *time to VF progression*, which was also classified according to the methodology used. A similar classification of measures was used for the other clinical outcomes (see supplementary Table S1 (web only file) for classification).

Designated primary and secondary outcomes were also identified. Descriptive statistical analysis was undertaken.

Results

The search identified 4323 potentially relevant abstracts. Of these, 238 abstracts were retrieved as “include” and 56 retrieved as “possible”. From these abstracts two hundred and fifteen articles were included whereas 79 articles were excluded because either the study was not an RCT or it was an RCT but not for a glaucoma intervention. An additional 24 potential RCTs identified from the references with three of these were excluded for not being an RCT. These RCTs covered the period between 1988 and 2005 as regards recruiting participants, e.g. AGIS recruiting period ranged from 1988 to 1992.(8) Two hundred and thirty three full articles were finally included for data synthesis (Figure 1, modified PRISMA flow chart).(9) Regarding abstract selection there was one disagreement between the two reviewers (0.25%), where one abstract was identified by the second reviewer (AAB) as “possible” which had been excluded by the first reviewer (RI). Following full text review, there was a disagreement between the first two reviewers to include two articles which were excluded by the senior reviewer (AAB). This disagreement was resolved by discussion and the decision was to exclude them for not being RCTs.

A total of 967 clinical measures were reported. There were large differences in the definitions used to describe different outcomes and their measures. IOP was the most commonly reported outcome with a total of 422 times (44% of all measures) (Table). IOP was reported as an outcome in 201 of the 233 RCTs (86%). Among the IOP-related measures (Table), the most commonly used was *mean IOP* (n=143, 15% of all measures) while *composite definition of success/failure using IOP as part of the definition (e.g., achieving a percentage of IOP reduction or target IOP with or without medication)* was the second most commonly utilised (n=135, 14% of all measures) (Table). Aqueous humor dynamics was only reported in five of the 233 RCTs (2%). On the other hand, safety outcomes were reported in 145 RCTs (62 %). There was a total of 346 times (36%) when

safety measures were described (Table). Visual acuity (VA) was the commonest among the safety measures (n=75, 8% of all measures) (Table). On the other hand, visual field was utilised in 38 RCTs (16%), (reported 50 times, 5% of all measures) (table), where VF progression measured with global indices/score was the commonest (n=30, 3% of all measures) (Table). Ocular and systemic haemodynamics (reported in 32 RCTs (n=91, 9% of all measures), anatomical outcomes (reported in 25 RCTs) (n=34, 4% of all measures) as well as surgical outcomes (reported in 19 RCTs) (n=19, 2% of all measures) were measured and reported. Optic nerve head (ONH) morphology was the commonest among the anatomical-related measures (n=19, 2% of all measures) (Table).

Systemic haemodynamics (blood pressure (BP) & pulse rate (PR) was the most commonly reported measure among the haemodynamics (n= 44, 5% of all measures) (Table).

Primary outcomes were defined in 143 of the 233 RCTs (61%). Thirty one of the 143 studies (21.6%) determined more than one primary outcome; the total number of primary measures identified was 186, with IOP-related measures being the commonest (n=135, 73% of primary measures) (Table). Mean IOP/ mean diurnal IOP/ 24-hour IOP was the commonest to be used as primary measure (n=51, 27% of all primary measures) (Table).

Discussion

To the best of our knowledge, this is the first study to describe the selection of clinical outcomes in glaucoma RCTs. The study demonstrated a large variation in the way clinical outcomes were reported and a total lack of standardisation which may impair the ability to robustly evaluate the evidence of glaucoma interventions.

This variation was greatest for IOP and safety measures, the most commonly used clinical outcomes. Even though VF tests were uncommonly utilised as an outcome, a diversity of measurements was also apparent (with 12 different ways of evaluating VFs), whereas persistence/adherence to therapy has never been reported as an outcome.

Whilst the review was conducted in a rigorous, systematic manner, it is possible that some studies may have been missed by not searching the “grey” literature. The review had good

external validity - the studies we identified included different glaucoma populations as well as different interventions. We excluded observational studies, but we were confident that these reports would be unlikely to add to the breadth of outcomes observed in this review. Although the search for RCTs covered the period between January 2006 till March 2012, original reports of the included RCTs had covered the period between 1988 and 2005 as regards to study design and recruiting participants.(8) (10-13) However, we acknowledge the potential limitations of time frame and language restrictions. We acknowledge that different clinical trials may have different objectives that lead to variation in primary outcomes. Phase I and II trials might be most concerned with effects on IOP and later phase trials might be most concerned with effects on visual field. However, our aim was to determine the breadth of utilised and reported clinical outcomes as a first step in the development of core clinical outcome sets for glaucoma interventions. A Delphi survey is ongoing to get consensus on clinical outcomes and identify further clinical measures for implementation in glaucoma trials. The selection of appropriate outcomes is crucial when designing clinical trials in order to compare directly the effects of different interventions in a way that maximises evidence synthesis;. There is a growing recognition that insufficient attention has been paid to the selection of outcomes when conducting a clinical trial.(1) Clinicians often assess glaucoma treatment in terms of the effect on IOP. However, being a surrogate outcome, it may be inappropriate to use IOP as the sole consideration for the evaluation of a glaucoma intervention; (14) besides there are different ways of evaluating IOP changes. In the published literature over a 5-year period (between 2001-2005), Rotchford et al identified variation in reporting of IOP success rates in glaucoma surgical trials. The authors concluded that there were nearly as many different IOP-related definitions of success after glaucoma surgery as there were articles on the subject, and more than 70% of definitions appeared in only a single article.(15) The authors recommended that standardisation of published outcome parameters is essential to allow meaningful comparisons between different study reports.(15)

The outcomes reported by studies are the key for clinicians when making decisions about health care; however there is a general lack of consensus regarding the selection of outcomes in clinical settings, which affects trial design, analysis, as well as reporting.(16) Measuring outcomes that will not change healthcare decisions leads to a waste of resources and a failure to capitalise on the potential power of research to improve health care.(4)

The difficulties caused by differences in outcome measurement are well known to systematic reviewers. For instance, the five most accessed and top cited Cochrane Reviews in 2009 all reported problems related to outcomes in eligible trials due to inconsistencies in the outcomes reported in the primary reports.(4) There is still great uncertainty on how best to select the most appropriate outcomes for use in a clinical trial.(17) Some of the factors underlying this uncertainty may include ambiguity regarding which outcomes are of relevance to patients, unknown performance characteristics of potential outcomes(e.g. reliability), or improvements in general care of patients with emergence of new technologies, with the result that previously used outcomes are no longer relevant.(18)

The Core Outcome Measures in Effectiveness Trials (COMET) Initiative (19) was launched in January 2010 to address this lack of standardised outcomes in clinical trials and to develop a minimum set of measures named “core outcome sets” (COS), which include endpoints to be reported as a minimum. There is an expectation that the core outcomes will be reported to allow the results of studies to be combined as appropriate; and that researchers will continue to collect other outcomes as well.(6). If the outcomes in the COS are reported then reviewers or people looking at studies in isolation, will always have access to the same set of key data.(6)

To solve the issue of outcome heterogeneity, we aim to reach consensus on important clinical outcomes and how to measure the outcomes in glaucoma intervention trials by taking into account the views of clinicians and trialists through a Delphi survey of expert opinion and by implementing the COSMIN (**CO**nsensus-based **S**tandards for the selection of health **M**easurement **I**nstruments) approach (20) as a final step.

In summary, the selection of outcomes and their measures was inconsistent among glaucoma RCTs. In addition, there was a wide variety of description within individual clinical outcomes. There is an obvious need to define a core clinical outcome dataset for glaucoma trials.

Table legends

Table: Clinical outcome measures identified in randomised controlled trials between January 2006 and March 2012

Outcomes	Measures	N (%)	N primary (%)
IOP	Mean IOP/Mean diurnal IOP/24 hr IOP	143(15%)	51(27%)
	Level of IOP variation (Range/Fluctuation)	12 (1%)	3(2%)
	Other level of IOP/ unspecified criteria	7 (1%)	2(1%)
	Mean change of IOP/Mean diurnal IOP/Change of 24 h IOP (from baseline)	74 (8%)	35(19%)
	Change of IOP variation (Range/Fluctuation) from baseline	5(1%)	2(1%)
	Other change of IOP from baseline	1(0%)	0(0%)
	Composite definition of success/failure: IOP(with or without medication), target IOP, percentage reduction of IOP	135 (14%)	34(18%)
	Composite definition of success/failure: IOP after provocation test	6 (1%)	1(1%)
	Composite definition of success/failure: number of postoperative medications)	34(3%)	5(3%)
	Others (correlations/ associations using IOP)	5 (1%)	2(1%)
Total		422(44%)	135(73%)
VF	VF Progression: Global indices/score (MD, PSD, CIGTS score, AGIS score	30 (3%)	10(5%)
	VF Progression: Point/ location analysis	10 (1%)	4(2%)
	VF- Progression: Clinical judgement	3(0.3%)	0(0%)
	VF Progression: Others	2 (0.2%)	0(0%)
	Rate of VF progression: Point/Location analysis	2(0.2%)	2(1%)
	Time to progression: Global indices/score	1(0.1%)	0(0%)
	Time to progression: Clinical judgement	1 (0.1%)	1(1%)
	Time to progression: Others	1(0.1%)	0(0%)
Total		50(5%)	17(9%)
Safety	Functional outcomes (VA)	75(8%)	3(2%)

	Others	47(28%)	16(8%)
Total		346(36%)	19(10%)
Haemodynamics			1(1%)
	Systemic Haemodynamics (BP &PR)	44(5%)	
	Others	47(4%)	3(1%)
Total		91(9%)	4(2%)
Anatomical	ONH morphology/ RNFL	19(2%)	5(3%)
	Others	15(2%)	1(0%)
Total		34(4%)	6(3%)
Aqueous Humour dynamics			
	AH dynamics	5(0%)	1(1%)
Surgical outcomes		19(2%)	4(2%)
Persistence/adherence to Therapy		0(0%)	0(%)
Total Measures		967	186

Abbreviations used in table : IOP, intraocular pressure; VF, visual field;; VF, visual field; MD, mean deviation; PSD, pattern standard deviation, CIGTS, Collaborative Initial Glaucoma Treatment Study; AGIS, Advanced Glaucoma Intervention Study; TFCM, tear film confocal microscopy; OGSS, Oxford Grading System Score; VA, visual acuity; ERG, electroretinogram; blood pressure; PR, pulse rate; ONH, optic nerve head; RNFL, retinal nerve fibre layer

Acknowledgement

RI is funded by James Mearns Trust for PhD Studentship. The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Government Health Directorate. The views expressed are those of the authors and not necessarily those of the funding bodies. The authors declare no conflict or commercial interests

Copyright for authors

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above."

Grant funding for research but no other competing interest

"All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: RI had financial support from James Mearns Trust for PhD studentship for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work."

References

- (1) Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy* 2012 Jan;17(1):1-2.
- (2) Ian Sinha. **OUTCOMES IN CLINICAL TRIALS IN CHILDREN WITH ASTHMA** University of Liverpool: University of Liverpool; 2011.
- (3) Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum* 1997 Jul;40(7):1202-1209.
- (4) Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012 Aug 6;13(1):132.
- (5) Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010 Feb 15;340:c365.
- (6) Williamson P, Clarke M. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative: Its Role in Improving Cochrane Reviews. *Cochrane Database Syst Rev* 2012 Apr 13;5:ED000041.
- (7) Che Hamzah J, Burr JM, Ramsay CR, Azuara-Blanco A, Prior M. Choosing appropriate patient-reported outcomes instrument for glaucoma research: a systematic review of vision instruments. *Quality of Life Research* 2011 Sep;20(7):1141-1158.
- (8) The Advanced Glaucoma Intervention Study (AGIS): 1. Study design and methods and baseline characteristics of study patients. *Control Clin Trials* 1994 Aug;15(4):299-325.
- (9) Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J Craniomaxillofac Surg* 2011 Mar;39(2):91-92.

- (10) Krupin T, Liebmann JM, Greenfield DS, Rosenberg LF, Ritch R, Yang JW, et al. The Low-pressure Glaucoma Treatment Study (LoGTS) study design and baseline characteristics of enrolled patients. *Ophthalmology* 2005 Mar;112(3):376-385.
- (11) Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology* 1999 Nov;106(11):2144-2153.
- (12) Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol* 1999 May;117(5):573-583.
- (13) Kass MA, Heuer DK, Higginbotham EJ, Johnson CA, Keltner JL, Miller JP, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002 Jun;120(6):701-13; discussion 829-30.
- (14) Spaeth GL. Proper outcome measurements regarding glaucoma: the inadequacy of using intraocular pressure alone. *Eur J Ophthalmol* 1996 Apr-Jun;6(2):101-105.
- (15) Rotchford AP, King AJ. Moving the Goal Posts. Definitions of Success after Glaucoma Surgery and Their Effect on Reported Outcome. *Ophthalmology* 2010 January 2010;117(1):18-23.e3.
- (16) Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. *BMJ* 2011 Jan 6;342:c7153.
- (17) Devane D, Begley CM, Clarke M, Horey D, OBoyle C. Evaluating maternity care: a core set of outcome measures. *Birth* 2007 Jun;34(2):164-172.
- (18) Sinha I, Jones L, Smyth RL, Williamson PR. A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. *PLoS Med* 2008 Apr 29;5(4):e96.
- (19) Core Outcome Measures in Effectiveness Trials. Available at: <http://www.comet-initiative.org/>.

(20) COSMIN: **C**onsensus-based **S**tandards for the selection of health **M**easurement **I**nstruments. Available at: http://www.cosmin.nl/clinimetrics_1_19.html, 2013.