Population based glaucoma screening
Stoutenbeek, Remco

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Stoutenbeek, R. (2010). Population based glaucoma screening. [s.n.].

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 8

General discussion
Chapter 8

The main issue that this thesis tries to address, is whether or not population based glaucoma screening should be introduced in the Netherlands. Chapter two (Literature review) gives an overview of current literature regarding different aspects of screening. Although most of the Wilson and Jungner criteria for population based screening are satisfied, two criteria in particular remain problematic. First and foremost, cost-effectiveness remains dubious (see paragraph 2.10). Second, although diagnostic tests for glaucoma have improved substantially in recent years, screening will still generate a lot of false positive subjects. Lack of a screening test with near-perfect specificity contributes to an unfavourable cost-effectiveness ratio of the screening programme (see paragraph 2.5). The five studies described in this thesis (chapter three to seven) were aimed at these two key matters.

8.1 Cost-effectiveness

Cost-effectiveness is usually expressed as costs per Quality Adjusted Life Year (QALY). It is a measure that indicates whether a particular intervention is good value for money. An intervention is said to be cost-effective when the costs per QALY gained is within the limits of society's Willingness to Pay (WTP), the exact amount varies per country. In the Netherlands, the cost-effectiveness threshold is set at €20,000/QALY for preventive interventions.\textsuperscript{1,2} In the United Kingdom, WTP is determined by the National Institute for Clinical Excellence to be about £30,000 (or €35,000).\textsuperscript{3} In the United States, WTP is $50,000 (also equalling €35,000).\textsuperscript{4,5} Economic implications of the introduction of population based screening are difficult to foresee in sufficient detail. As described in paragraph 2.10, the two major studies on cost-effectiveness of glaucoma screening\textsuperscript{6,7} disagree on the economic feasibility thereof, because their respective models yielded considerably different final costs per QALY: €9,000 as found by the Finnish model\textsuperscript{6} is well within the WTP limits, while around €70,000 (rough approximation) as estimated by the UK model\textsuperscript{7} clearly is not. So which one of them is right, or at least closest to the actual costs? Several points of criticism were already made in paragraph 2.10 regarding the Finnish study. The various elements that make up the overall cost of a screening programme are summarized in table 2.10.1 (Costs and benefits of a glaucoma screening programme). Several modifiers exert influence on the total costs of glaucoma screening. The most important parameters in this respect are: test specificity, glaucoma prevalence, age at which screening starts, and screening interval. The latter two are sufficiently covered in chapter two (Literature review). However, additional comments are necessary regarding test specificity and glaucoma prevalence, related to the cost-effectiveness studies mentioned above.
Test specificity

In paragraph 2.8, the number of false positive test subjects produced by population based screening in the Netherlands was estimated, assuming a screening test specificity of 95% based on data presented in paragraph 2.5. However, most of the studies cited in paragraph 2.5, concerning both perimetry and imaging tests, determined specificity on a by-eye basis (i.e. one eye is selected as the study eye; the fellow eye is excluded). In real life, most individuals have two eyes, hence two chances to be classified as abnormal. Data from the FDT perimetry measurements collected for the third chapter of this thesis was reanalysed to calculate the by-eye specificity and compare it to the original by-patient specificity. Out of 108 healthy elderly subjects, 11 had VF loss in one of their eyes (five right eyes and six left eyes), and only two had VF loss in both eyes. This corresponds to a by-patient specificity of 88%, and a mean by-eye specificity of 93%. The same subjects also underwent GDx imaging as part of the Groningen Longitudinal Glaucoma Study. These data were also re-analysed. By-eye specificity was set at 94% by choosing the appropriate cut-off point for the GDx test. The corresponding by-patient specificity was 90%. If we adjust for the discrepancies of by-eye versus by-patient specificity in regard to the specificity rate of 95% as applied in table 2.8.1, then the actual maximum feasible specificity of glaucoma screening for either a single perimetry test or a single imaging test would be closer to 91%. In that case, the clinical workload would nearly double, approaching 500,000 false positives instead of 277,000 every five years! The impact on cost-effectiveness of so many false positives is unacceptable, and needs to be negated. In chapter four, several strategies for increasing the test specificity were explored. The Finnish cost-effectiveness study assumed a test specificity of 98%, based on combining different diagnostic tests, and repeating abnormal test results. The highest specificity we were able to obtain was 94% by combining FDT and GDx results (classification on by-patient basis), see chapter four. Repeating abnormal tests in addition to combining different types of tests will probably result in a modest additional increase in specificity. However, a specificity of 98% remains unlikely to be achieved without a significant drop in sensitivity, including loss of sensitivity to moderate and severe glaucoma, which is especially undesirable.

Glaucoma prevalence

The prevalence of glaucoma in the Netherlands is about 2% in the general population aged >40 years (see paragraph 2.1). At least half of all glaucoma patients are undetected, so the prevalence of undetected glaucoma is approximately 1%. Glaucoma screening would naturally become much more cost-effective if prevalence were higher. This gain is clearly shown in the various sensitivity analyses carried out in the UK cost-effectiveness study discussed earlier. Therefore, the authors of that study suggest screening targeted at certain subpopulations as a potential alternative to screening of the general population. The aim is to attain a subpopulation in which the prevalence of glaucoma is
increased, by limiting the screening programme to subjects with one or more risk factors for developing glaucoma. Important risk factors for glaucoma are summarized in the last section of paragraph 2.2. Positive family history of glaucoma, black ethnicity, and myopia are relevant in this regard. Screening in risk factor defined subpopulations can be effective if two conditions are met. Obviously, the Relative Risk (RR; also termed Risk Ratio) of a specific risk factor must be significantly increased. But if the risk factor itself is rare, then the screening programme targets only a fraction of the population, with little impact on the overall burden of a particular disease. The Population Attributable Risk (PAR) takes this into account. Since the proportion of the Dutch 50+ population with black ethnicity is less than 1.3%, the PAR of black ethnicity is insignificant even though it carries a considerably increased RR of 3.8. Positive family history of glaucoma might be a better candidate for targeted screening. Out of 6773 subjects that participated in the Rotterdam Study, 8.5% had a positive family history of glaucoma (see chapter six, discussion section). The RR may be as high as 9.2 according to Wolfs et al., partially based on data from the Rotterdam Study. They reported a PAR of 16.4%, which could be considered worthwhile. However, as a result of current opportunistic case finding, the proportion of undetected glaucoma patients that have a positive family history of glaucoma is very small, whereas the proportion of known glaucoma patients with positive family history is far greater. Specifically 2% among undetected glaucoma patients and 39% among detected patients have a positive family history of glaucoma, P<0.001, see chapter six (percentages presented in table 6.1). Since the majority of patients who have a positive family history of glaucoma are apparently already detected quite efficiently through current practice of opportunistic case finding, screening targeted at this risk factor is redundant. That leaves myopia as the remaining option for targeted screening. Myopia is a common condition with an estimated prevalence of 26% in the Netherlands (see chapter five, table 5.2). However, the RR of myopia is only 1.7 (based on pooled data from three population based glaucoma studies). This modest RR is probably not worth the additional effort of inviting only myopic subjects for screening, improvement in cost-effectiveness will be minimal. Also, it is hard to justify from an ethical point of view that the non-myopic population is not eligible to glaucoma screening, based on such a minor difference in risk of developing glaucoma. A more practical approach would be to intensify opportunistic case finding taking place at opticians. Myopics need to visit an optician anyway because their spectacles will need adjustment or replacement periodically, so most of them will be accessible for evaluation. This alternative will be discussed further below.

In summary, strategies to improve the cost-effectiveness of glaucoma screening by targeting subpopulations with higher glaucoma prevalence as a result of risk factors are not viable in the Netherlands.
8.2 Screening bias

Both observational studies that investigate ongoing screening programmes as well as modelling studies that try to predict the yield of a screening programme are susceptible to bias. There are at least three types of screening bias: length bias, lead time bias, and class bias. See paragraph 2.3 for background information about terminology used in this context.

Length bias
The progression rate of diseases in general and of glaucoma in particular can vary considerably. For some individuals glaucoma progresses rapidly, whereas others have a slowly progressing variant. Those with a rapidly progressive variant remain for only a relatively short period of time in the detectable preclinical phase (D-PCP), a stadium in which the disease is detectable by a screening test but does not yet cause any symptoms. In contrast, individuals with a slowly progressive variant will remain in the D-PCP for many years. Among cases identified by a periodic screening programme, the slowly progressive ones will be over-represented. After all, rapidly progressive cases will more often become symptomatic (manifest disease; clinical phase) during the screening interval (the time period in between two consecutive screening rounds), before their next screening round was due. Consequently, the benefits of screening fall short: the decrease in morbidity and/or mortality will turn out to be smaller than expected. This phenomenon is called length bias, and is essentially a kind of selection bias towards mild disease cases. Its effect in glaucoma screening is probably significant, because glaucoma is characterised by a long D-PCP on average, but at the same time there is considerable individual variation in the rate of progression. The negative impact of length bias on screening yields is difficult to quantify. The study described in chapter six was originally designed and intended to explore the presence of length bias in glaucoma, and give a rough estimate of the effect magnitude. Unfortunately, baseline differences came to light between opportunistically detected and undetected incident glaucoma patients, which precluded a final statement on the presence or absence of length bias in glaucoma screening. Still, glaucoma patients that were discovered through opportunistic case finding outside the Rotterdam Study had worse glaucoma than those that had remained undetected (see discussion of chapter six for more details).

Lead time bias
Screening allows for early detection of diseases compared to normal diagnosis which is initiated in a later stadium, when complaints or symptoms start to occur. Time gained by early detection is called lead time, and is equivalent to half the D-PCP (see paragraph 2.3). Survival is defined as the time span between biological onset of disease (in this case glaucoma) and end stage thereof (blindness). If a disease is detected earlier due to screening, then survival will increase.
automatically by an amount equal to the lead time. This is not a beneficial effect caused by screening, but purely the arithmetical result of starting to count from an earlier disease phase. Lead time bias occurs when this deviation is not taken into account. Evaluation of ongoing screening programmes is prone to lead time bias, whereas studies conducted prior to introduction of screening as well as modelling studies are not. Therefore, the impact of lead time bias is irrelevant at present.

**Class bias**

Individuals from a higher socio-economic class have a healthier lifestyle than individuals from a lower socio-economic class. They are also more likely to participate in health promoting projects such as a screening programme. If screening is attended predominantly by healthy people, then less disease will be detected. This problem is called class bias. Whether a healthy lifestyle has any influence on glaucoma onset or progression is questionable. Black ethnicity is a risk factor for glaucoma, and is also associated with a lower socio-economic class, which may cause some class bias. However, as mentioned earlier in this chapter, the proportion of Dutch 50+ inhabitants with black ethnicity is less than 1.3%. Any effects of class bias must therefore be minimal.

In summary, among the three existing types of screening bias, only length bias is relevant for glaucoma screening in the context of this thesis. The modelling studies (discussed in paragraph 2.10 and also in this chapter) do not account for length bias, hence their cost-effectiveness estimates regarding glaucoma screening are too optimistic.

### 8.3 Verdict on glaucoma screening

Most aspects relevant to glaucoma screening were reviewed in chapter two. An important argument against screening is the problem that screening is in all probability not cost-effective. This complicated issue is discussed in paragraph 2.10, as well as in paragraph 8.1 and 8.2, where the negative impact on cost-effectiveness of suboptimal by-patient specificity and length bias are assessed. Several additional aspects against or in favour of the introduction of a screening programme in the Netherlands will be discussed in this section.

Part of the incident glaucoma cases identified by the Rotterdam Study were found to have also been detected by regular ophthalmic care outside the study, whereas other cases had remained undetected during the entire follow-up interval (see chapter six). Twenty-three cases (29%) had already been detected, 55 cases (71%) remained undetected. The severity of glaucoma was worse in detected cases compared to undetected cases ($P=0.009$). The additional yield of screening is therefore lower than would be expected from prevalence data: we
estimated that only about one in 1000 screened individuals could be saved from bilateral end-stage glaucoma. This estimate might be somewhat conservative, since it is based on the assumption that only patients with >10 missed points on STP in their better eye will reach bilateral end-stage glaucoma. In chapter seven we found that >10 missed points on STP corresponds to a HFA MD < -7.5 dB (moderate glaucoma). A lower cut-off point is arguably more sensible. For >5 missed points on STP (equivalent to HFA MD < -3.8 dB, early glaucoma) estimated yields will be better: about one in 550 screened individuals might be saved from blindness. Nevertheless, yields are limited, and for every 1000 screened individuals, 50 cases will still require further investigation as a result of false positive test results (see paragraph 2.5 and 2.8).

In the last section of paragraph 2.4, studies by Hattenhauer\textsuperscript{19,24} and Chen\textsuperscript{25} were discussed that reported on the cumulative probability of bilateral blindness from glaucoma, defined as either VF constriction to ≤20° or visual acuity ≤2/20 in the better eye. Most eyes were classified as blind due to VF constriction; only a small fraction had a visual acuity of ≤2/20. Although severe constriction of the VF is unquestionably debilitating, the impact on quality of life is limited as long as central visual acuity is preserved. Much et al. investigated the long-term survival of central visual field in end-stage glaucoma.\textsuperscript{26} Eighty-four eyes of 64 patients were included. Only fourteen eyes (17%) lost more than three lines of visual acuity during a mean follow-up period of 8.3 years. Thus, despite being considered legally blind based on VF criteria (see table 2.1.1 for definitions), central visual acuity of end-stage glaucoma patients may be preserved for many years.

Definitions of blindness, low vision, and visual impairment are based on the remaining amount of visual function in the better seeing eye (see table 2.1.1). Glaucoma reduces health-related quality of life (QoL) mainly in advanced stages of the disease, when severe visual field damage has also occurred in both eyes.\textsuperscript{27,28} Therefore, glaucoma severity of the better eye is emphasized throughout this thesis, while the worse eye is essentially being ignored. An alternative point of view is that screening should be aimed at retaining useful vision in both eyes, and thus at preventing end-stage glaucoma in the worse eye. QoL in individuals with good vision in both eyes is higher than QoL in functionally monocular individuals.\textsuperscript{29} But the difference is only marginal, and consequently cost per QALY gained by a glaucoma screening programme designed to save the worse eye would be astronomical. Still, the introduction of a screening programme set to prevent bilateral blindness would as an added benefit also lower incidence of unilateral blindness with a concurrent modest increase in QoL.

Verdict

The cost-effectiveness studies\textsuperscript{6,7} discussed in paragraph 2.10 and this chapter are a good starting point for determining whether or not a periodic population based glaucoma screening programme should be introduced in the Netherlands. The
UK study\textsuperscript{7} concluded that screening was not cost-effective by a wide margin. The Finnish study\textsuperscript{6} concluded that screening was indeed cost-effective, and in certain age cohorts even dominant. However, their results seem over-optimistic due to several apparent weak points in study design and chosen parameter values; see criticism expressed in relation to the discussion of this study in paragraph 2.10. In this thesis, several factors are identified that have a further negative impact on the already dubious cost-effectiveness of glaucoma screening. These factors are:

- length bias (chapter 6 and paragraph 8.2)
- suboptimal by-patient test specificity (paragraph 2.5 and 8.1; chapter 3 and 4)
- large clinical workload from false positives (paragraph 2.8 and 8.5)

This leads to the conclusion that at present periodic population based glaucoma screening is not feasible in the Netherlands.

### 8.4 Alternatives

There are at least three commonsense alternatives to periodic population based screening. The most straightforward option is to maintain the current practice of opportunistic case finding. Another approach is once in a lifetime screening instead of periodic screening. This second option seems counter-intuitive since glaucoma prevalence slowly increases with age, and there are no specific opportunities during the course of glaucoma that warrant screening at any specific moment. This strategy is not explored in this thesis. The third and most interesting option is to intensify opportunistic case finding taking place at opticians, and is discussed in the following section.

In chapter five it is reported that 80\% of Dutch inhabitants aged $>$40 years visit an optician at least once every five years.\textsuperscript{20} As mentioned in paragraph 2.8, participation rates for breast cancer screening vary considerably, and 70\% is considered to be achievable.\textsuperscript{30} This means that optician based case finding will potentially reach at least as many individuals as a nation wide screening programme. Optician based screening might therefore be a more valid designation than optician based case finding. Opticians are comfortable with providing ophthalmic care, and most of them already perform non-contact tonometry as a manner of glaucoma case finding voluntarily. 97\% of optician shops is equipped with a non-contact tonometer (unpublished data, obtained from the questionnaires described in chapter five). There are approximately 3300 optician shops in the Netherlands.\textsuperscript{31} In a small pilot study among 50 opticians, 37 of whom responded to the questionnaire, 91\% expressed willingness to participate in an extended glaucoma screening programme (i.e. more extensive than tonometry).\textsuperscript{20} Financial compensation was not ascertained, and may not be necessary. After all, opticians have acquired non-contact tonometers for competitive reasons of their own accord. The same trend is now discernible with
respect to the FDT perimeter and the GDx nerve fibre analyser. In paragraph 2.8, it is established that a lower age limit of 50 years is suitable for a glaucoma screening programme in the Netherlands. For optician based screening, it may be more appropriate to start at age 45, so as not to miss individuals that visit an optician because of presbyopia.

Optician based screening would reduce costs considerably as compared to a normal screening programme. See table 2.10.1 for an overview of the different types of screening related costs: all direct screening costs as well as the societal costs related to productivity loss and transport are practically eliminated in optician based screening. Still, substantial clinical costs remain because of the large clinical workload that results from screening, principally due to false positive test results (discussed in paragraph 2.8). Optician based screening is far more likely to be cost-effective than a normal glaucoma screening programme, but a definite statement requires additional research (see paragraph 8.5).

If optician based screening were to be instituted, ophthalmologists should be made aware that around 90% of individuals that are referred by opticians because of an abnormal test result, are expected to turn out to be false positive cases. Left unaware, ophthalmologists might at least initially assume that the referring opticians are incompetent. A positive attitude and readiness to cooperate are important. Encouragement of opticians to participate in screening efforts may be a suitable task for ophthalmologists’ departments. Referral should not require an appointment with the general practitioner.

The screening test strategy must be kept simple so that any employee in the optician shop can carry out the test, and understands what to do with the results. Cut-off points for screening tests should be chosen aiming for high specificity. An example of a simple decision tree for screening with the FDT perimeter is shown below in figure 8.4.1. The FDT can be substituted by another glaucoma screening device such as the GDx or HRT without the need to alter the rest of the decision tree. Repeat testing is probably necessary for all of current diagnostic devices in order to attain a high specificity. Regardless of the format of glaucoma screening, both opticians and ophthalmologists still need to perform tonometry on a routine basis. The reason is that individuals with normal screening tests but a (very) high IOP can progress so rapidly that a five year screening interval is inadequate.
Figure 8.4.1 Decision tree optician based glaucoma screening by FDT perimetry
8.5 Suggestions for further research

During the process of writing this thesis, several issues were identified that would benefit from further research.

False positives
The majority of glaucoma screening participants with a positive (abnormal) test result will be classified as normal upon further evaluation. These individuals are called ‘false positives’ (see paragraph 2.8). There are several options regarding how to deal with false positives in the course of the screening programme. Viable approaches are: return false positives to periodic screening; retain false positives under clinical follow-up; or, exclude false positives from screening indefinitely and rely on opportunistic case finding only. If false positives are allowed to return to periodic screening, accumulation of normal subjects that repeatedly fail their screening test every screening round will likely ensue, which will negatively affect the cost-effectiveness of the screening programme. It would be interesting to know what proportion of false positives falls into this category of ‘reoffenders’. If the proportion is significant, then it becomes important to determine the reason why false positives occur. This knowledge can help to choose the best course of action. For example, a false positive subject who has cataract may return to periodic screening after phacoemulsification is performed. Some of the false positives will prove to be unsuitable for either imaging or perimetry (see first section of paragraph 2.5 for overview of causes thereof). They can return to periodic screening if an alternate screening test modality is used (i.e. switch from perimetry to imaging or vice versa); if the screening programme by default relies on a combination of two screening test modalities, then the inappropriate test is left out and cut-off points for the remaining test may be adjusted. False positives who are unsuitable for both perimetry and imaging should be excluded from further screening rounds, case finding only seems the most cost-effective for this group. Borderline cases should be followed clinically.

Increasing the specificity
High specificity is important in a screening setting, as stressed in paragraphs 2.5 and 8.1. The Finnish cost-effectiveness study discussed in paragraphs 2.10 and 8.1 assumed a very high by-patient test specificity of 98% for their Markov model, attained theoretically by combining different diagnostic tests and repeating abnormal test results. A study is needed to explore what effects such an extensive screening test strategy will have on sensitivity in general and sensitivity to moderate and severe glaucoma in particular (with cut-off points set to achieve a 98% specificity).

Blindness from glaucoma
Why do glaucoma patients in the Netherlands go blind? This question is of fundamental importance with respect to the yield of glaucoma screening. If the
predominant cause of glaucoma related blindness is late detection, then screening will have the greatest impact on preventing blindness (see paragraph 2.4 for advantages of early detection). In contrast, if suboptimal treatment of glaucoma progression during follow-up also plays a significant role, then the beneficiary effects of screening will be limited. To clarify this issue, a study is required that determines the proportion of blind glaucoma patients in the Netherlands that was treated suboptimally.

Continuation of the current line of research
Further research in the field of glaucoma screening in our department is aimed at optician based glaucoma screening. Preparations for a pilot project investigating this matter were recently started in collaboration with opticians in the North of the Netherlands who are willing to participate in glaucoma screening. Severity of glaucoma at diagnosis will be determined prior to and after introduction of the pilot project. This will provide more definite data on the yield of glaucoma screening. Cost-effectiveness of optician based glaucoma screening should be explored before it is introduced nationwide.
References


5. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What is the price of life and why doesn’t it increase at the rate of inflation? Arch Intern Med 2003;163:1637-41.


10. Jansonius NM, Heeg GP. The Groningen Longitudinal Glaucoma Study. II. A prospective comparison of frequency doubling perimetry, the GDx


General discussion


