

OPEN

EXPERT CONSENSUS DOCUMENT

Semantics in active surveillance for men with localized prostate cancer — results of a modified Delphi consensus procedure

Sophie M. Bruinsma¹, Monique J. Roobol¹, Peter R. Carroll², Laurence Klotz³, Tom Pickles⁴, Caroline M. Moore⁵, Vincent J. Gnanapragasam⁶, Arnauld Villers⁷, Antti Rannikko⁸, Riccardo Valdagni⁹, Mark Frydenberg¹⁰, Yoshiyuki Kakehi¹¹, Christopher P. Filson¹² and Chris H. Bangma¹ on behalf of the Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) consortium

Abstract | Active surveillance (AS) is broadly described as a management option for men with low-risk prostate cancer, but semantic heterogeneity exists in both the literature and in guidelines. To address this issue, a panel of leading prostate cancer specialists in the field of AS participated in a consensus-forming project using a modified Delphi method to reach international consensus on definitions of terms related to this management option. An iterative three-round sequence of online questionnaires designed to address 61 individual items was completed by each panel member. Consensus was considered to be reached if $\geq 70\%$ of the experts agreed on a definition. To facilitate a common understanding among all experts involved and resolve potential ambiguities, a face-to-face consensus meeting was held between Delphi survey rounds two and three. Convenience sampling was used to construct the panel of experts. In total, 12 experts from Australia, France, Finland, Italy, the Netherlands, Japan, the UK, Canada and the USA participated. By the end of the Delphi process, formal consensus was achieved for 100% ($n = 61$) of the terms and a glossary was then developed. Agreement between international experts has been reached on relevant terms and subsequent definitions regarding AS for patients with localized prostate cancer. This standard terminology could support multidisciplinary communication, reduce the extent of variations in clinical practice and optimize clinical decision making.

Active surveillance (AS) is an accepted option for the initial management of carefully selected men with localized, well-differentiated prostate cancer who are thought to have a low risk of progression^{1–4}. AS is broadly described as a management option for patients with low-risk prostate cancer, which involves the postponement or avoidance of invasive treatment, with a switch to curative treatment if evidence is obtained that the patient has an increased risk of disease progression or if the patient expresses preference for it. However, semantic heterogeneity exists in the literature and guidelines. For instance, the specific definitions of the terms AS and watchful waiting (WW) are inconsistent in the published

literature and can elicit considerable confusion. The terms AS and WW are frequently used interchangeably, but they refer to very different observational approaches. AS involves the avoidance or postponement of immediate therapy combined with careful surveillance; definitive treatment is then offered if there is evidence that the patient is at increased risk of disease progression⁴. AS differs from WW, which is based upon the premise that men will not benefit from definitive treatment of clinically localized prostate cancer owing to limited life expectancy, comorbidity, and the prolonged natural history of the prostate cancer⁴. Patients managed using a WW protocol undergo observation consisting of a lesser

Correspondence to S.M.B.
s.m.bruinsma@erasmusmc.nl
Department of Urology,
Erasmus Medical Center,
Wytemaweg 80, 3015 CN
Rotterdam, Netherlands.

doi:10.1038/nrurol.2017.26
Published online 14 Mar 2017;
corrected 30 Mar 2017

Key points

- Active surveillance (AS) is broadly described as a management option for men with low-risk prostate cancer, but semantic heterogeneity exists in the literature and guidelines
- An urgent need for uniform terminology exists to establish active communication and collaboration between research groups around the world
- Agreement between international experts has been reached on 61 relevant terms and subsequent definitions regarding AS for patients with localized prostate cancer
- This standard terminology could support multidisciplinary communication, reduce the extent of variation in clinical practice and optimize clinical decision making

degree of monitoring than those on AS might receive, and in which palliative treatment is generally instituted if metastases or local symptoms develop. Interpreting and comparing research results is difficult owing to the nonstandardized use of the terms AS and WW and their intended and often mixed treatment objectives (curative or palliative)⁵.

This semantic heterogeneity is also reflected in AS guidelines⁶. In these guidelines, AS is primarily recommended for patients with low-risk tumours; however, these guidelines contain various definitions of low-risk prostate cancer, as specified by different combinations of clinical criteria including clinical and pathological characteristics (such as tumour stage, serum PSA levels, biopsy Gleason score, tumour volume and serum PSA density). Furthermore, definitions of disease reclassification and progression differ among published guidelines and multiple criteria for initiation of curative treatment are proposed⁶.

Problems resulting from the use of ambiguous language include hindered clinical decision making, particularly in multidisciplinary collaborations, and limited opportunities for research⁷. Moreover, such ambiguity has raised a barrier that hampers exchange of knowledge within and between fundamental domains of research

and research groups⁸. An urgent need exists for uniform terminology to help clarify scientific discourse: that is, we need to speak the same language. The primary purpose of this study was to reach international consensus on definitions of terms that are often used in AS undertaken for carefully selected men with localized, well-differentiated prostate cancer.

Methods used to develop definitions

Expert panel. Convenience sampling was used to construct the panel of experts. The Movember Foundation's Global Action Plan Active Surveillance (GAP3) project is an integrated project lasting 30 months that is being implemented across 14 countries in the five Movember regions (Australasia, Europe, the UK, Canada, the USA)⁹. Milestones of the project include the development of a global AS database for clinical, biospecimen, imaging and biomarker data, worldwide, tailor-made guidelines on AS and a web-based AS platform for patients and providers. The experts for the panel were selected within the Movember Foundation's GAP3 consortium (BOX 1), consisting of urologists, acclaimed scientists, a pathologist and radiation oncologists with expertise in the field of AS. We aimed at including at least one expert per participating institute (currently, $n = 15$) in the GAP3 consortium. Eligible experts were invited by e-mail to participate in the study. After participating, the experts were asked to provide personal information, such as their specialty⁸.

List of terms. As part of the GAP3 initiative, a narrative review of available AS guidelines provided a comprehensive overview of recommendations regarding patient selection, frequency and type of monitoring and the criteria for initiation of definitive treatment⁶. This review has been used as a starting point to produce a list of potentially important terms (led by Sophie Bruinsma, a researcher in the field of prostate cancer). Subsequently,

Author addresses

¹Department of Urology, Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, Netherlands.

²University of California, San Francisco, 550 16th Street, Department of Urology, 6th Floor, Mailbox Code 1695, San Francisco, California 94143, USA.

³University of Toronto, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, #MG 408, Toronto, Ontario M4N 3M5, Canada.

⁴University of British Columbia, Department of Radiotherapy & Developmental Radiotherapeutics, Radiation Oncology, BC Cancer Agency, 600 West 10th Avenue Vancouver, British Columbia, V6R 2T9, Canada.

⁵University College London and University College London Hospitals Trust, 4th Floor, Rockefeller Building, 74 Huntley Street, London, WC1E 6AU, UK.

⁶Academic Urology Group, Department of Surgery and Oncology, University of Cambridge, Box 279 (S4), Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK.

⁷Department of Urology, Hôpital Huriez, Place de Verdun, Centre Hospitalier Régional Universitaire, 59037 Lille, France.

⁸Helsinki University and Helsinki University Hospital, Department of Urology, PL340, 00029 HUS, Helsinki, Finland.

⁹Department of Oncology and Hemato-oncology, Università degli Studi di Milano, Radiation Oncology 1, Fondazione IRCCS Istituto Nazionale dei Tumori, Via G. Venezian 1 - 20133 Milano, Italy.

¹⁰Department of Urology, Monash Health; Department of Surgery, Faculty of Medicine, Monash University, 322 Glenferrie Road, Malvern, Melbourne 3144, Victoria, Australia.

¹¹Department of Urology, Kagawa University Faculty of Medicine, 1750-1 Ikenobe, Miki-cho, Kita-gun, Kagawa 761-0793, Japan.

¹²Department of Urology, Winship Cancer Center, Emory University School of Medicine, 1365 Clifton Road NE, Suite B1400, Atlanta, Georgia, USA.

Box 1 | Members of the Movember Foundation's GAP3 consortium

Principle Investigators

Bruce Trock (Johns Hopkins University, The James Buchanan Brady Urological Institute, Baltimore, USA), Behfar Ehdai (Memorial Sloan Kettering Cancer Center, New York, USA), Peter Carroll (University of California San Francisco, San Francisco, USA), Christopher Filson (Emory University School of Medicine, Winship Cancer Institute, Atlanta, USA), Jeri Kim (MD Anderson Cancer Center, Houston, USA), Todd Morgan (University of Michigan and Michigan Urological Surgery Improvement Collaborative, Michigan, USA), Laurence Klotz (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada), Tom Pickles (University of British Columbia, BC Cancer Agency, Vancouver, Canada), Eric Hyndman (University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada), Caroline M. Moore (University College London and University College London Hospital Trust, London, UK), Vincent Gnanapragasam (University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK), Mieke Van Hemelrijck (King's College London, London, UK & Guy's and St Thomas' NHS Foundation Trust, London, UK), Prokar Dasgupta (Guy's and St Thomas' NHS Foundation Trust, London, UK), Chris Bangma (Erasmus Medical Center, Rotterdam, Netherlands), Monique Roobol (Erasmus Medical Center, Rotterdam, Netherlands), Arnould Villers (Lille University Hospital Center, Lille, France), Antti Rannikko (Helsinki University and Helsinki University Hospital, Helsinki, Finland), Riccardo Valdagni (Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy), Antoinette Perry (University College Dublin, Dublin, Ireland), Jonas Hugosson (Sahlgrenska University Hospital, Göteborg, Sweden), Jose Rubio-Briones (Instituto Valenciano de Oncología, Valencia, Spain), Anders Bjartell (Skåne University Hospital, Malmö, Sweden), Lukas Hefermehl (Kantonsspital Baden, Baden, Switzerland), Lee Lui Shiong (Singapore General Hospital, Singapore, Singapore), Mark Frydenberg (Monash Health; Monash University, Melbourne, Australia), Yoshiyuki Kakehi (Kagawa University Faculty of Medicine, Kagawa, Japan), Byung Ha Chung (Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea), Kwang Suk Lee (Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea).

Pathologist

Theo van der Kwast (Princess Margaret Cancer Centre, Toronto, Canada).

Technology Research Partners

Henk Obbink (Royal Philips, Eindhoven, Netherlands), Wim van der Linden (Royal Philips, Eindhoven, Netherlands), Tim Hulsen (Royal Philips, Eindhoven, Netherlands), Cees de Jonge (Royal Philips, Eindhoven, Netherlands).

Regional statisticians

Mike Kattan (Cleveland Clinic, Cleveland, Ohio, USA), Ji Xinge (Cleveland Clinic, Cleveland, Ohio, USA), Kenneth Muir (University of Manchester, Manchester, UK), Artitaya Lophatananon (University of Manchester, Manchester, UK), Michael Fahey (Epworth HealthCare, Melbourne, Australia), Ewout Steyerberg (Erasmus Medical Center, Rotterdam, Netherlands), Daan Nieboer (Erasmus Medical Center, Rotterdam, Netherlands); Liying Zhang (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada)

Clinical Research Partners' IT Experts

Wei Guo (Johns Hopkins University, The James Buchanan Brady Urological Institute, Baltimore, USA), Tanya Milan (Memorial Sloan Kettering Cancer Center, New York, USA), Nicole Benfante (Memorial Sloan Kettering Cancer Center, New York, USA), Janet Cowan (University of California San Francisco, San Francisco, USA), Dattatraya Patil (Emory University School of Medicine, Winship Cancer Institute, Atlanta, USA), Rachel Sanford (MD Anderson Cancer Center, Houston, USA), Tae-Kyung Kim (University of Michigan and Michigan Urological Surgery Improvement Collaborative, Ann Arbor, Michigan, USA), Alexandre Mamedov (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Canada), Vincent LaPointe (University of British Columbia, BC Cancer Agency, Vancouver, Canada), Trafford Crump (University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada), Rifat Hamoudi (University College London and University College London Hospital Trust, London, UK), Jenna Kimberly-Duffell (University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK), Aida Santaolalla (King's College London, London, UK and Guy's and St Thomas' NHS Foundation Trust, London, UK), Daan Nieboer (Erasmus Medical Center, Rotterdam, Netherlands), Jonathan Olivier (Lille University Hospital Center, Lille, France), Emanuele Bianchi Janetti (Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy), Tiziana Rancati (Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy), Helén Ahlgren (Sahlgrenska University Hospital, Göteborg, Sweden), Juanma Mascarós (Instituto Valenciano de Oncología, Valencia, Spain), Annica Löfgren (Skåne University Hospital, Malmö, Sweden), Kurt Lehmann (Kantonsspital Baden, Baden, Switzerland), Catherine Han Lin (Monash University and Epworth HealthCare, Melbourne, Australia), Hiromi Hiram (Kagawa University, Kagawa, Japan).

Research Advisory Committee

Guido Jenster (Erasmus MC, Rotterdam, Netherlands), Anssi Auvinen (University of Tampere, Tampere, Finland), Anders Bjartell (Skåne University Hospital, Malmö, Sweden), Masoom Haider (University of Toronto, Toronto, Canada), Kees van Bochove (The Hyge B.V. Utrecht, Utrecht, Netherlands), Ballentine Carter (Johns Hopkins University, Baltimore, USA).

Management team

Rachelle Kirk-Burnnand (Movember Foundation, Melbourne, Australia), Sam Gledhill (Movember Foundation, Melbourne, Australia), Mark Buzza (Movember Foundation, Melbourne, Australia), Sophie Bruinsma (Erasmus Medical Center, Rotterdam, Netherlands).

GAP3; Global Action Plan Prostate Cancer Active Surveillance.

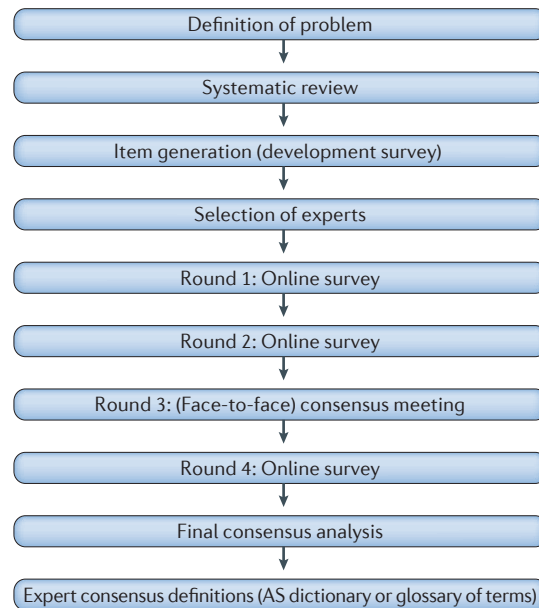


Figure 1 | Key elements of the proposed modified Delphi study. The Delphi method is a widely accepted technique of structured and systematic information gathering from a group of experts (termed the Delphi Panel) on a specific topic using a series of questionnaires. The Delphi method enables a panel of experts to provide insights and opinions, even when they are not located in the same geographic area. The current study uses a modified Delphi method, in which an iterative sequence of online surveys were combined with a physical meeting of the panel of experts

this list has been critically reviewed by Monique J. Roobol (an epidemiologist) and Chris H. Bangma (a urologist), from Erasmus MC, the Netherlands, who are experts in the field of prostate cancer, and items were added if considered needed. In total, the initial list addressed 53 individual items.

Procedure for determining definitions

Consensus definitions were derived using a Delphi method. The Delphi method is a widely accepted technique of structured and systematic information gathering from a group of experts (termed the Delphi Panel) on a specific topic using a series of questionnaires¹⁰. The Delphi method enables a panel of experts to provide insights and opinions, even when they are not located in the same geographic area. According to this formal consensus-building method, participants were asked to fill out an iterative sequence of surveys, in this case in the form of online questionnaires. The current study uses a modified Delphi method, in which online surveys were combined with a physical meeting of the panel of experts (FIG. 1).

In the first round, the experts were asked to provide definitions of 53 terms related to AS for low-risk prostate cancer according to their personal opinion. These experts were informed that this list might not be exhaustive and were asked to add potentially missing items and their corresponding definitions to the list at the end of

the survey. The open comments made by the experts were carefully considered by the referee group⁸, consisting of Bruinsma and Roobol, and, based on the input of the experts, temporary definitions were formulated and clustered into themes. These temporary definitions were presented to the experts in a second-round survey and they were asked whether they agreed or disagreed with the proposed definitions; if they did not agree with a certain definition, the experts were asked to clearly state why. Consistent with other studies^{11–13}, consensus was considered to be reached if $\geq 70\%$ of the experts agreed on a definition⁸. If consensus was achieved on definitions, these were added to the AS dictionary or glossary of terms (hereafter, referred to as the glossary) (TABLE 1). Subsequently, to facilitate a common understanding among all experts involved and to resolve potential ambiguities, a consensus, in-person meeting was held, which was attended by representatives from the majority of the countries participating in the GAP3 consortium (the third round). This meeting was organized in conjunction with the annual conference of the European Association of Urology (EAU) in March 2016, hosted in Munich (Germany). During this meeting, the majority of the terms and subsequent definitions on which no consensus had been reached in the previous surveys were further discussed by the experts. Based on the second survey and the input of the experts during the face-to-face consensus meeting, a third and final survey was designed (the fourth round). This final survey consisted of: terms and subsequent definitions on which no consensus had been reached in the second survey, and were adapted based on the input of the experts from this survey; terms and subsequent definitions on which no consensus was reached in the second survey, that were discussed during the consensus meeting, and subsequently adapted based on the experts' input at the consensus meeting; and terms and subsequent definitions on which consensus was reached in the second survey, but were adapted based on suggestions for improvement from the experts in the second survey. We asked the experts whether they agreed or disagreed with the formulated consensus definitions. If they did not agree with a certain definition, the experts were asked to clearly state why they did not agree with the proposed definition. Consensus was considered to be reached if $\geq 70\%$ of the experts agreed on a definition. Consensus definitions were added to the glossary (TABLE 1).

Survey administration. Three rounds of surveys were conducted between January 2016 and April 2016. The experts were given ~2 weeks to complete each survey round, and several reminder e-mails were sent. In the first round, the preliminary survey was sent to all experts. The second-round survey was only sent to the responders from the initial round. All experts who responded to the first and second survey were permitted to participate in the consensus meeting (third round). The third and final survey (round four) was disseminated to all participants from round two and to those present at the consensus meeting who did not participate in round two.

Table 1 | Glossary of terms related to active surveillance

Term	Definition	% of agreement
Background		
Overdetection	The detection of cancers that would not have been discovered in the absence of screening (such as PSA), and would not have caused any (clinical) problems	100
Overdiagnosis	The diagnosis of a clinically insignificant cancer that would not have been discovered in the absence of screening (such as PSA), and that would not harm the patient	100
Overtreatment	The treatment of clinically insignificant cancer that would otherwise not have harmed the patient during his lifetime	100
Risk groups and surveillance		
Risk	The likelihood of a defined event	100
Risk group	A system classifying patients into categories sharing approximately the same likelihood of a defined event (such as risk), based on clinical and pathological findings (including tumour stage, serum PSA levels and biopsy Gleason score)	100
(Clinical) risk stratification	The process of assigning patients with prostate cancer to risk groups based on clinical features (such as PSA level, DRE and biopsy findings)	100
Very-low-risk prostate cancer	Prostate cancer with a minimal risk of progression on repeat biopsy (such as an increase in Gleason score or an increased number of cores positive for cancer) and with very good prognosis	75
Low-risk prostate cancer	Prostate cancer with a low risk of progression on repeat biopsy (such as an increase in Gleason score or an increased number of cores positive for cancer) and with a good prognosis	75
Intermediate-risk prostate cancer	Prostate cancer with a moderate risk of progression on repeat biopsy (such as an increase in Gleason score or an increased number of cores positive for cancer) and with a reasonable prognosis	83
High-risk prostate cancer	Prostate cancer with a high risk of metastasis or cancer-specific death if left untreated	92
Cancer definitions		
Localized prostate cancer	Cancer that is confined within the prostate (for example, not spread outside the prostate), classified by clinical stage <T3	100
Indolent tumour	A tumour that is slow growing, or not growing at all.	75
Early prostate cancer	Small-volume, localized prostate cancer	100
Clinically insignificant prostate cancer	Prostate cancer that is, despite the absence of treatment, unlikely to cause symptoms, or metastasize (even in the absence of symptoms) or to cause mortality during a man's lifetime	100
Favourable-risk disease	A low-risk prostate cancer, characterized by T1(c) or T2, PSA<10 ng/mL and biopsy Gleason score of ≤6	90
Biopsy		
Protocol-based biopsy	A biopsy scheduled as part of a predefined AS protocol	100
Non-protocol-based biopsy	A biopsy outside the predefined AS protocol	90
Diagnostic biopsy	The first biopsy is positive for prostate cancer	100
Confirmatory biopsy	The prostate biopsy following a positive diagnostic biopsy (such as the first biopsy positive for prostate cancer) that is intended to confirm clinical insignificance of the previously diagnosed prostate cancer, is typically performed within 12 months after diagnosis and might target previously undersampled areas, according to a protocol that specifies minimum standards	92
Targeted biopsy	Sampling of specific areas of the prostate that are suspicious for cancer, usually based on DRE or imaging findings	100
MR-targeted biopsy	Any biopsy technique (such as cognitive, fusion, in-bore) in which an MRI scan is used to determine the location of a suspicious target prior to biopsy	100
Treatment choice		
Active surveillance	A monitoring strategy for patients with prostate cancer with the aim of avoiding or deferring curative treatment	100
Watchful waiting	Management of patients with a limited life expectancy, in whom palliative treatment (without curative intent) is initiated if symptoms develop	100
Definitive treatment	Any treatment with curative intent	80
Active treatment	Treatment (such as surgery, radiotherapy or focal therapy) for prostate cancer with the primary aim of curing cancer	83
Immediate treatment	Treatment with curative intention, typically performed within 6 months of diagnosis, without any time period of intended surveillance	90
Treatment shift	Generic expression indicating a change in the therapeutic plan of a patient	90

Table 1 (cont.) | Glossary of terms related to active surveillance

Term	Definition	% of agreement
Reclassification/progression		
Reclassification	A change in risk group as a result of re-evaluation of clinical or pathological parameters, unlikely to be caused by actual changes in cancer biology	80
Upgrading	Change in risk group owing to an increase in Gleason grading on repeat biopsy	100
Upstaging	Change in risk group owing to an increase in the extent of the disease (such as stage) based on DRE or imaging findings	100
Progression	A broad term indicating worsening of the disease, based on an increase in grade or extent of disease after a follow-up period, unrelated to resampling	92
Tumour progression	Clinical evidence of an increased tumour risk, based on clinical, imaging or pathological findings	75
Clinical progression	New cancer related signs (such as an increase in tumour grade and/or tumour volume and/or clinical stage and/or PSA level) or the development or worsening of symptoms, assessed by physical examination (such as DRE and biopsy) and/or imaging, unrelated to resampling	92
Biochemical failure/ recurrence	An increase of PSA after curative treatment above a defined threshold	92
Biochemical progression	An increase of a biochemical marker over time while on AS, above a defined threshold or kinetic parameter	92
PSA progression (failure/ recurrence)	An increase in PSA over time more than a predefined level, calculated PSADT or PSAV, as an indicator of disease progression	80
Pathologic progression	Change in pathological characteristics as shown by an increase of tumour grade (based on rebiopsy Gleason score) and/or volume (based on the number of positive cores or the maximum extent of cancer per core) after a follow-up period, unrelated to resampling	92
Symptomatic progression	Progression of cancer such that it causes symptoms	80
MRI progression	Stage progression or increase in volume or risk score of an MRI defined prostate cancer lesion, or the appearance of a new lesion since last MRI	100
Markers		
PSA doubling time (PSADT)	The time needed for PSA to double (an increase of 100%)	100
PSA velocity (PSAV)	An increase in serum PSA level in a unit of time (usually per year)	100
PSA density (PSAD)	Total serum PSA level divided by total prostate volume (ng/ml/cc), either measured by TRUS or MRI	100
Free PSA	Serum PSA that is unbound to other proteins in the blood	100
Prostate Health Index (PHI)	A PSA-based test that combines the results of three PSA measurements (standard PSA, free PSA, and [-2]pro-PSA) to improve accuracy for cancer	100
Prostate Cancer Antigen 3 (PCA3)	A prostate specific gene, used as a urine-based biomarker, to determine risk of progression	83
Other terms		
Baseline	The time period during which the patient was referred to AS	80
Gleason score	A grading system for prostate cancer as a method for predicting the behaviour of this disease (in terms of aggressiveness)	80
Grade, grading	A system for classifying cancer cells based on how abnormal they appear when examined under a microscope, to provide information about the aggressiveness of the tumour and its tendency to spread in the body, according to well established criteria such as Gleason grading (ISUP)	90
Downgrading	Decrease in Gleason score when comparing prostate biopsy with radical prostatectomy Gleason score	100
Stage, staging	Cancer extension in the body as per the TNM staging system	90

AS, active surveillance; DRE, digital rectal exam; ISUP, International Society of Urological Pathology; PCA3, Prostate Cancer Antigen 3; PHI, Prostate Health Index; PSAD, PSA density; PSADT, PSA doubling time; PSAV, PSA velocity; TRUS, transrectal ultrasonography.

Results of the consensus initiative

Participation of experts. In total, 17 experts from the 15 participating institutions were invited to take part in the first-round survey, of which 14 completed the survey and three did not respond to the invitation. These 14 experts were invited to complete the second-round survey, of which 10 responded and four did not respond to the invitation. In total, seven of the 10 experts who responded to the second survey were present at the

consensus meeting (round three). In addition, two experts who participated in the first-round survey but not in the second-round survey were present at the consensus meeting and participated in the semantics discussion. The 10 experts who participated in the second survey and were also present at the consensus meeting, and the experts who participated only in the first round survey and were present at the consensus meeting were invited for the third and final survey

(round four). In total, 12 experts were invited, all of whom participated (FIG. 2, TABLE 2 and [Supplementary information S1](#) (table)).

Survey design and reaching a consensus. Initially, 53 terms relating to AS were included in the first-round survey. Subsequently, eight new terms were added to this list by participants in the first round. Thus, the second survey consisted of 61 terms. Terms were classified according to the following themes: background ($n=3$), risk groups and surveillance ($n=7$), cancer definitions ($n=6$), biopsy ($n=13$), treatment choice ($n=8$), reclassification/progression ($n=13$), markers ($n=6$) and other AS terms ($n=5$). At the end of round two, 64% ($n=39$) of the survey items achieved consensus. The majority of the items and their subsequent definitions on which no consensus was reached in the second survey (19 of 22) were discussed in more detail during the consensus meeting. Some items were not discussed owing to time constraints; these terms included localized prostate cancer, indolent tumour and tumour progression. Based on the discussions, some of the terms were excluded from the glossary (TABLE 3). Reasons for exclusion included unfamiliarity with the concept ($n=2$), insufficient evidence to determine the definition of a term ($n=1$), or the experts considered them unclear and not useful in the field of AS ($n=9$). Based on the results of the second survey and the

consensus meeting, a third survey was designed. This final survey consisted of 23 items: definitions on which a consensus was not reached in the second survey and were adapted based on the input of the experts in this survey ($n=5$); definitions on which a consensus was not reached in the second survey, which were discussed during the consensus meeting and adapted based on the experts' input ($n=8$); and of definitions upon which a consensus was reached in the second survey, but were adapted slightly based on suggestions for improvement from the participants in this second survey ($n=10$). Consensus was reached with respect to definitions of all these terms.

Results of the Delphi process. By the end of the Delphi process, formal consensus ($\geq 70\%$ agreement) was achieved on 100% of terms ($n=61$). In total, consensus definitions were formulated for 51 terms (TABLE 1). The additional 10 terms were excluded from the AS glossary, as all experts agreed these terms are unclear and unnecessary in the field of AS (TABLE 3). Of the 51 terms, 25 definitions reached full consensus (100% agreement). Complete agreement was reached on definitions of key terms such as AS, WW, upgrading and upstaging. For 26 items, consensus ranged from 75% to 92%. Small ambiguities were encountered with definitions related to the various risk groups that are used to stratify patients with prostate cancer ($n=4$), cancer definitions ($n=2$), biopsy terms ($n=2$), treatment choices ($n=4$), reclassification/progression ($n=9$) and other AS terms ($n=5$). Furthermore, a semantic model has been developed, representing an AS timeline (from prostate cancer diagnosis to long-term evaluation of AS), including associated terms and definitions from the glossary (FIG. 3). This overview includes some key terms from the glossary, but is not exhaustive. The first event to occur in this AS timeline is the diagnosis of prostate cancer. Once the diagnosis of prostate cancer is established, further evaluation that incorporates known risk factors is required to determine appropriate treatment options. During AS, the prostate cancer is closely monitored over time. If repeated risk evaluation shows changes in the condition of the patient, treatment plans can be adapted accordingly.

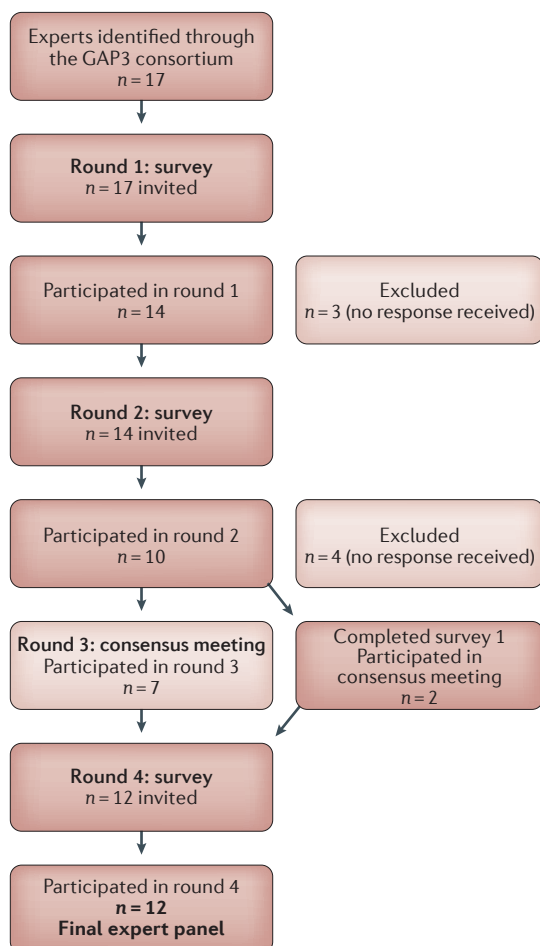


Figure 2 | Recruitment of experts to the Delphi study.

In total, 17 experts from the 15 participating institutions were invited to take part in the first-round survey, of which 14 completed the survey and three did not respond to the invitation. These 14 experts were invited to complete the second-round survey, of which 10 responded and four did not respond to the invitation. In total, seven of the 10 experts who responded to the second survey were present at the consensus meeting (round 3). In addition, two experts who participated in the first-round survey but not in the second-round survey were present at the consensus meeting and participated in the semantics discussion. The 10 experts who participated in the second survey and were also present at the consensus meeting, and the experts who participated only in the first round survey and were present at the consensus meeting were invited for the third and final survey (round 4). In total, 12 experts were invited, all of whom participated making up the final expert panel. GAP3; Global Action Plan Prostate Cancer Active Surveillance.

Table 2 | **Characteristics of the Delphi expert panel**

Attribute	Characteristic	Participants (n)
Gender	Male	11
	Female	1
Background or specialty	Urologist	10
	Radiation oncologist	2
Country	USA	2
	Canada	2
	UK	2
	Australia	1
	France	1
	Finland	1
	Italy	1
	Netherlands	1
	Japan	1

Challenges in achieving a consensus

An urgent need exists for uniform terminology regarding AS in order to aid communication and collaboration among research groups around the world. The purpose of this study was to reach international consensus on definitions of terms often used in AS for carefully selected men with localized, well-differentiated prostate cancer. Using a modified Delphi method in which 12 known leaders in the field participated, agreement has been reached on 61 relevant terms and subsequent definitions relating to AS for prostate cancer.

Several findings deserve particular attention. The experts encountered difficulties regarding the definitions of the various risk groups used to stratify patients with prostate cancer, namely very-low-risk, low-risk, intermediate-risk and high-risk prostate cancer. The explanation for why these difficulties were encountered seems to be multifactorial. Firstly, too many doubts existed on the combinations of clinical criteria — including clinical and pathological characteristics — that differentiate the various risk groups. Many questions were raised by the experts, including whether only clinical stage, serum PSA level and Gleason score should be included; whether other

criteria, such as PSA density and maximum percentage of cancer per core biopsy are also pertinent and at what level the cut-off values should be set. Secondly, difficulties were encountered with regard to the cancer-specific survival rates that are associated with these risk groups. Thirdly, the experts had different perspectives on the definition of risk. For example, whether this term refers to the risk of metastasis or cancer-specific mortality. The experts concluded that, at present, the risk groups should not be defined, apart from general concepts, owing to the fact that robust data from men with clinically insignificant prostate cancer who are undergoing AS, especially from studies with >10 years mean follow-up duration, remains limited. The Movember Foundation's GAP3 project, which was launched in August 2014, can make a substantial contribution to the collection of robust data^{6,14}. By combining data from existing AS databases (including clinical, biopsy sample, imaging and biomarker data) from all over the world, the largest centralized prostate cancer AS database to date has been created, which will be updated annually. By subsequently analysing data from the majority of patients who are currently undergoing AS worldwide, appropriate definitions of the various risk groups will likely be delineated.

In addition to concerns regarding definition of the risk groups, intensive and complex discussion occurred on the distinction between the concepts of AS and WW. A formal consensus has been reached on the definitions of both management strategies, but the experts involved indicated that it was difficult to dichotomize surveillance. According to the experts, the intensity of surveillance gradually declines over time. An agreement seems to exist regarding both ends of the spectrum: protocol-based surveillance (that gives rise to curative treatment) at one end and no surveillance at the other end. However, the existence of a grey zone in between these strategies has been acknowledged (FIG. 4). This grey zone was described by the experts as a phase of active (regular) annual monitoring of serum PSA levels (no biopsies undertaken) with the aim of palliation when deemed necessary. Many clinical questions arose, including what this strategy should be called in practice — the terms 'slow surveillance' and 'AS light' were proposed but not agreed upon — or whether it should simply be referred to as 'non-protocol-based

Table 3 | **Terms excluded from the active surveillance glossary**

Term	Reason for exclusion	% of agreement
First biopsy	Unclear and not useful in the field of active surveillance	100
Second biopsy	Unclear and not useful in the field of active surveillance	100
1st and 2nd biopsy	Unclear and not useful in the field of active surveillance	100
Initial biopsy	Unclear and not useful in the field of active surveillance	100
Repeat biopsy	Unclear and not useful in the field of active surveillance	100
Serial biopsy	Unclear and not useful in the field of active surveillance	100
Systematic biopsy	Unclear and not useful in the field of active surveillance	100
Slow surveillance	Unfamiliarity with the concept	100
Active surveillance light	Unfamiliarity with the concept	100
Radiological progression	Insufficient evidence as yet	100

CONSENSUS STATEMENT

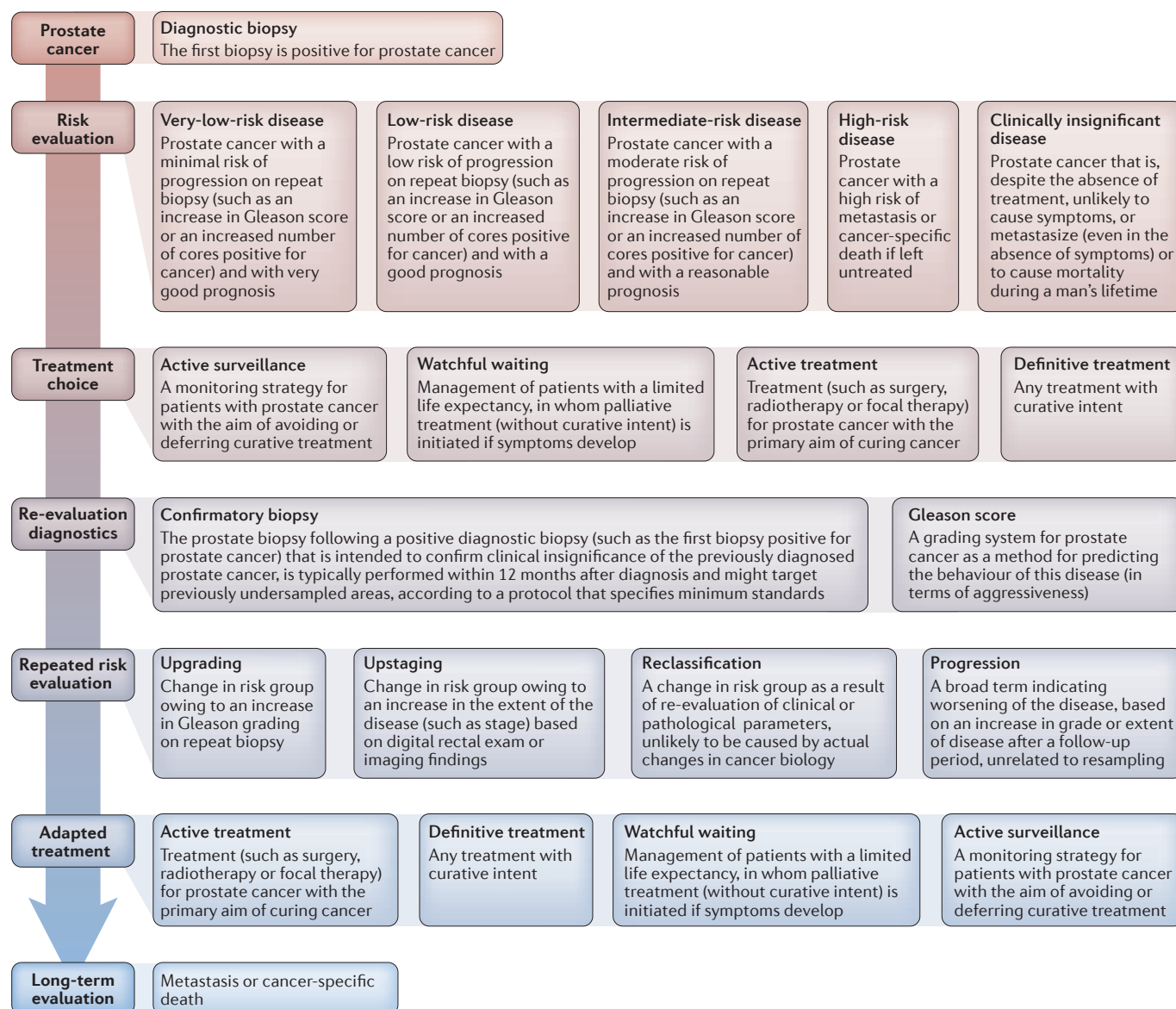


Figure 3 | **Semantic model of the active surveillance (AS) timeline (from diagnosis to long-term evaluation), including associated terms and definitions per stage.** The first event is diagnosis of prostate cancer by biopsy sampling. Patients are then evaluated and stratified by the risk category of their disease: very-low-risk, low-risk, intermediate-risk, high-risk or clinically insignificant. A treatment choice appropriate to their risk category is then made, choices of therapy include AS, watchful waiting, active treatment and definitive treatment. Patients can then undergo re-evaluation diagnostics including a confirmatory biopsy and assessing the Gleason score of the cancer, after which a repeat risk evaluation can be undertaken. Based on this new risk evaluation, treatment can be adapted accordingly and patients enter into a long-term evaluation protocol.

surveillance'. Also, the questions of what patients call this type of surveillance and how they perceive it, arose during the consensus procedure. Furthermore, how AS should be practiced was discussed, including aspects such as whether a serum PSA test should be performed every year or, for instance, every 5 years, and what the best strategy for deciding to stop the AS protocol is. Clearly, despite the availability of considerable consensus on the key terminology, continued research is necessary to gain a deeper understanding of these clinical aspects of undertaking and practising AS.

Biopsy sampling and analysis has a role in the risk assessment of patients with prostate cancer who are eligible for AS. After initiation of an AS programme, most guidelines recommend use of surveillance biopsies to check for and identify pathological indications of tumour progression. Many biopsy-related terms were found in the literature and several more were raised by the experts in the survey rounds of our study, including initial biopsy, first biopsy, second biopsy, repeat biopsy, serial biopsy and systematic biopsy. All experts agreed that the majority of these terms are unclear and

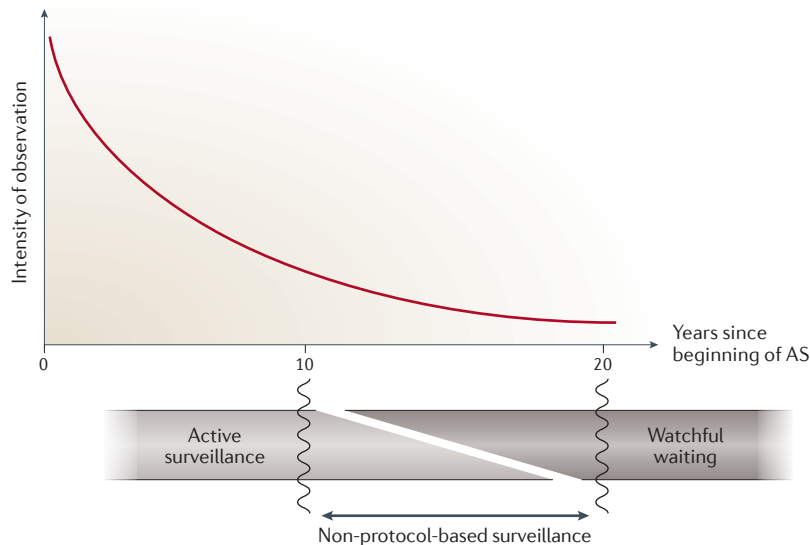


Figure 4 | The grey zone between active surveillance (AS) and watchful waiting based on expert consensus. This grey zone was described by the experts as a phase of active (regular) annual monitoring of serum PSA levels (no biopsies undertaken) with the aim of palliation when deemed necessary.

unnecessary in the field of AS, and, therefore, should not be included in the glossary (TABLE 3). The experts strongly advise clinicians, researchers and patients to limit the terms to diagnostic biopsy, confirmatory biopsy, protocol-based (surveillance) biopsy and non-protocol-based (surveillance) biopsy to avoid confusion in clinical practice.

Many definitions on which consensus was reached by the panel of experts incorporate references to the Gleason grading and scoring system. A group from Johns Hopkins Hospital led by Dr Epstein first proposed grouping scores into five prognostic categories, termed Grade Groups 1–5 (REF. 15). A subsequent multi-institutional study of >20,000 men validated these Grade Groups, which resulted in its acceptance by the International Society of Urological Pathology, the WHO and the College of American Pathologists¹⁶. Importantly, these new grades are likely to enter mainstream practice in the near future, which will, in turn, potentially influence AS terminology.

Many projects that aim to introduce standard terminology in clinical practice are unsuccessful, perhaps because standard terminology is rarely used in clinical practice⁷. A number of opportunities exist to consider how to most effectively implement standardized terminology for AS into clinical practice. The aim of the Movember Foundation's GAP3 initiative^{6,14} is to offer standardized, evidence-based guidelines on AS. The glossary of terms can be added to these guidelines and form the basis for a full understanding of the presented recommendations. Additionally, the homogeneous semantics should be used in presentations at major meetings of national and international associations, and included in papers that will be published in national and international scientific journals and specialists journals.

Strengths and weaknesses

The modified Delphi method seems to have been successful for deriving consensus definitions. Furthermore, the face-to-face consensus meeting of the referee group enabled the in-depth exploration of the reasons for disagreements on definitions. These discussions accelerated the consensus process and revealed new areas of interest (such as the grey area between AS and WW). Nevertheless, this study has some limitations. As purposive sampling was used (and participants were, therefore, not randomly selected), representativeness cannot be assured. However, the whole premise behind the Delphi theory is that the panel members are in fact experts in their field, therefore, yielding results of increased accuracy, instead of selecting a representative sample of the population. Furthermore, the number of experts that participated in this Delphi study (sample size) was relatively small. The number of participants could have affected the potential for ideas as well as the amount of data analysed. However, no agreement on the panel size for Delphi studies exists, and neither do recommendations or unequivocal definitions of small or large samples¹⁷. Many published Delphi studies use panels consisting of 10–100 or more panellists¹⁷. Official consensus was obtained regarding all 61 definitions, but not all experts fully agreed with all final definitions. Consensus was considered to be reached if $\geq 70\%$ of the experts agreed on a definition. In the current study, consensus varied between 75% and 100% per item. Also, one term (radiological progression) has been excluded from the AS dictionary because insufficient evidence exists to include it as yet. In a systematic review on the use of MRI in men with low-risk or intermediate-risk prostate cancer who were considered suitable for AS, MRI was demonstrated to be useful for the detection of clinically significant disease at initial clinical assessment of men considering AS¹⁸. In addition, MRI might be useful for confirming the absence of any large anterior lesions that could have been missed during routine diagnosis¹⁹. However, robust, formally published data are needed before widespread implementation of MRI for this purpose^{18,20}. Most likely, the semantics of AS will need to be augmented with MRI-related definitions as new evidence becomes available. During the course of the GAP3 project, many current patient series have been found to lack sufficient volume to be analysed appropriately. Additional funding has been committed by the Movember Foundation to assess the value of MRI within AS. In conclusion, the findings of this Delphi consensus procedure represent expert opinion, rather than indisputable fact, which should be kept in mind²¹.

Conclusions

Agreement between international experts has been reached on relevant terms and subsequent definitions regarding AS for patients with localized prostate cancer. This standard terminology could support multidisciplinary communication, reduce the extent of variations in clinical practice and optimize clinical decision making. International debate on all aspects of AS might be strengthened by an improved understanding of the concept of AS.

- Dall'Era, M. A. *et al.* Active surveillance for prostate cancer: a systematic review of the literature. *Eur. Urol.* **62**, 976–983 (2012).
- National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology. NCCN http://www.nccn.org/professionals/physician_gls/pdf_guidelines.asp (2017).
- Morash, C. *et al.* Active surveillance for the management of localized prostate cancer: guideline recommendations. *Can. Urol. Assoc. J.* **9**, 171–178 (2015).
- Klotz, L. Active surveillance for men with low-risk, clinically localized prostate cancer. *UpToDate* <http://www.uptodate.com/contents/active-surveillance-for-men-with-low-risk-clinically-localized-prostate-cancer> (2017).
- Ip, S. *et al.* An evidence review of active surveillance in men with localized prostate cancer. *Evid. Rep. Technol. Assess. (Full Rep.)* **204**, 1–341 (2011).
- Bruinsma, S. M. *et al.* Active surveillance for prostate cancer: a narrative review of clinical guidelines. *Nat. Rev. Urol.* **13**, 151–167 (2016).
- Stallinga, H. A. *et al.* Does language ambiguity in clinical practice justify the introduction of standard terminology? An integrative review. *J. Clin. Nurs.* **24**, 344–352 (2015).
- Kleynen, M. *et al.* Using a Delphi technique to seek consensus regarding definitions, descriptions and classification of terms related to implicit and explicit forms of motor learning. *PLoS ONE* **9**, e100227 (2014).
- Bruinsma, S. M., Bangma, C. H., Obbink, H. & Roobol, M. J. Active surveillance for low risk prostate cancer: the study protocol of the Movember Global Action Plan 3 (GAP3) project [abstract 1036]. *Eur. Urol. Suppl.* **14**, e1036–e1036a (2015).
- Yeh, J. S., Van Hoof, T. J. & Fischer, M. A. Key features of academic detailing: development of an expert consensus using the Delphi method. *Am. Health Drug Benefits* **9**, 42–50 (2016).
- Mokkink, L. B. *et al.* Protocol of the COSMIN study: Consensus-based Standards for the selection of health Measurement INstruments. *BMC Med. Res. Methodol.* **6**, 2 (2006).
- Zafar, S. Y. *et al.* Consensus-based standards for best supportive care in clinical trials in advanced cancer. *Lancet Oncol.* **13**, e77–e82 (2012).
- Hasson, F., Keeney, S. & McKenna, H. Research guidelines for the Delphi survey technique. *J. Adv. Nurs.* **32**, 1008–1015 (2000).
- Bruinsma, S. M. Global action plan on active surveillance for low risk PC: Movember Foundation launches integrated project on active surveillance. *Eur. Urol. Today* **26**, 26 (2014).
- Pierorazio, P. M., Walsh, P. C., Partin, A. W. & Epstein, J. I. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU Int.* **111**, 753–760 (2013).
- Epstein, J. I. *et al.* A contemporary prostate cancer grading system: a validated alternative to the Gleason score. *Eur. Urol.* **69**, 428–435 (2016).
- Akins, R. B., Tolson, H. & Cole, B. R. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Med. Res. Methodol.* **5**, 37 (2005).
- Schoots, I. G. *et al.* Magnetic resonance imaging in active surveillance of prostate cancer: a systematic review. *Eur. Urol.* **67**, 627–636 (2014).
- Sankineni, S., Osman, M. & Choyke, P. L. Functional MRI in prostate cancer detection. *Biomed. Res. Int.* **2014**, 590638 (2014).
- Moore, C. M. *et al.* Reporting MRI in men on active surveillance for prostate cancer — the PRECISE (Prostate Cancer Radiological Estimation of Change in Sequential Evaluation) recommendations: a report of a European School of Oncology task force. *Eur. Urol.* <http://dx.doi.org/10.1016/j.eururo.2016.06.011> (2016).
- Powell, C. The Delphi technique: myths and realities. *J. Adv. Nurs.* **41**, 376–382 (2003).

Acknowledgements

This study is linked to a larger project, The Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) initiative, which is a collaboration between institutions, hospitals and research centres from the USA, Canada, Australia, Singapore, Japan, Korea, UK, Ireland, the Netherlands, France, Sweden, Finland, Switzerland, Italy and Spain. The Movember Foundation has invested €1,664,950 in GAP3 to create the largest centralized prostate cancer active surveillance database to date. This funder did not have any role in the study design, collection, analysis or interpretation of data, or in the drafting of this paper. For information, contact M.J.R. m.roobol@erasmusmc.nl.

Author contributions

All authors researched data for and reviewed and/or edited the manuscript before submission. S.M.B. wrote the article and S.M.B., M.J.R. and C.H.B. provided a substantial contribution to discussions of content.

Competing interests statement

The authors declare no competing interests.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this

article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

SUPPLEMENTARY INFORMATION

See online article: [S1 \(table\)](#)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF