Appendix 2 (as supplied by the authors). Modified Delphi Consensus Process Methods and Results Summary

Appendix 2A. Methods Summary

The modified Delphi consensus process used two rounds to assess the level of agreement with, and facilitate refinements to, the consolidated screening principles[1-3]. Both rounds were conducted via online survey (Surveymonkey.com). We used four sources to identify international screening experts, including the lead authors of the included articles identified from our systematic review, members of the International Cancer Screening Network (ICSN), a targeted web search of screening councils/networks and screening-related conferences, and input from members of our research team and colleagues. We purposively selected experts that represented a range of fields/disciplines (medicine, public health, health care management/policy, health economics), areas of screening (cancer, prenatal, newborn, infectious disease), and jurisdictions. For the ICSN, we selected all members (n=19) as this was the preferred approach of the Network's leadership group. Standardized email invitations (and multiple email reminders) with consent information and a link to an online survey were sent to the selected experts directly (with the exception of ICSN members who were forwarded the email invitation and an email reminder via a network administrator).

Drawing on the findings of the systematic review, an online survey instrument was developed and pilot tested. In addition to seven questions on participant demographics and experience, participants were asked to assess each of the 12 consolidated principles, rating their level of agreement (on a seven-point Likert scale from strongly agree to strongly disagree) with two statements: "This principle is important and relevant for population-based programmatic screening decisions" (hereafter 'importance') and "This principle is clearly defined and understandable" (hereafter 'interpretability'). Participants were given the opportunity to provide qualitative feedback on each principle, and at the end of the survey were asked to note any apparent gaps and/or identify missing principles in the consolidated set, and to offer any other general comments.

Round 1 participants willing to participate in Round 2 were provided (i) a customized report that compared their Round 1 responses with anonymized/aggregated responses of all other Round 1 participants and (ii) a link to the Round 2 online survey. The Round 2 survey presented participants with revised principles (based on Round 1 feedback) and used the same rating questions for the importance and interpretability of each principle, again with the opportunity to provide qualitative feedback (see Table A2.1 for specific versions of the consolidated principles presented to participants in Round 1 and Round 2).

We assessed both consensus (level of agreement within rounds) and stability (consistency of results between rounds)[3] of the Delphi process using descriptive statistics (frequency distributions, measures of central tendency and dispersion) and assessment of the number and nature of participants' qualitative feedback[3,4].

Table A2.1 Comparison of Round 1 and Round 2 Versions of the Consolidated Principles

Round 1 Version of Consolidated Principles (Post-Systematic Review)	Round 2 Version of Consolidated Principles (Includes Post-Round 1 Refinements)
Principle 1. Epidemiology of the disease/condition: The epidemiology of the disease/condition should be adequately understood, and the disease/ condition should be an important health problem (e.g., high or increasing incidence/prevalence and causes substantial morbidity/mortality).	**Principle 1.** Epidemiology of the disease/condition: The epidemiology of the disease/condition should be adequately understood, and the disease/condition should be an important health problem (e.g., high or increasing incidence/prevalence and/or causes substantial morbidity/mortality).
Principle 2. Natural history of disease/condition: The natural history of the disease/condition should be adequately understood, the disease/condition is well-defined, and there should be a detectable preclinical phase.	**Principle 2.** Natural history of disease/condition: The natural history of the disease/condition should be adequately understood, the disease/condition is well-defined and, where relevant, there should be a detectable preclinical phase.
Principle 3. Target population for screening: The target population for screening should be clearly defined (e.g., with an appropriate target age-range), identifiable, accessible, and likely to participate.	**Principle 3.** Target population for screening: The target population for screening should be clearly defined (e.g., with an appropriate target age-range), identifiable, and contactable.
Principle 4. Screening test performance characteristics Screening test performance should be appropriate for the purpose, with all key components of the test being accurate (e.g., sensitive, specific, positive predictive value), reliable/reproducible, safe/ethical/acceptable, simple and cost-effective to perform/ administer to the target population.	**Principle 4.** Screening test performance characteristics: Screening test performance should be appropriate for the purpose, with all key components specific to the test (rather than the screening program) being accurate (e.g., in terms of sensitivity, specificity, positive predictive value) and reliable/reproducible. The test should be acceptable to the target population and it should be possible to perform/administer it safely, affordably and efficiently.
Principle 5. Target population for post-screening care: Screening test results should be clearly interpretable and determinate (e.g., with known distribution of test values and well-defined and agreed cut-off points) to allow identification of the screening participants who should (and should not) be offered diagnostic testing and other post-screening care.	**Principle 5.** Interpretation of screening test results: Screening test results should be clearly interpretable and, where appropiate, determinate (e.g., with known distribution of test values and well-defined and agreed cut-off points) to allow identification of the screening participants who should (and should not) be offered diagnostic testing and other post-screening care.
Principle 6. Post-screening care: There should be an agreed upon course of action for screening participants with positive screening results that involves diagnostic testing, treatment/intervention and follow-up care that will modify/alter the natural history and clinical pathway for the disease/condition, is available/accessible/acceptable to those affected and results in improved outcomes (e.g., survival, function, quality of life). The burden of post-screening care on all participants should be understood and the impact of false-positive tests should be minimized.	**Principle 6.** Post-screening test options: There should be an agreed upon course of action for screening participants with positive screening test results that involves diagnostic testing, treatment/intervention and follow-up care that will modify/alter the natural history and clinical pathway for the disease/condition, is accessible and acceptable to those affected and results in improved outcomes (e.g., increased functioning/quality of life, decreased cause-specific mortality). The burden of post-screening care on all participants should be understood/acceptable and the impact of false-positive and false-negative tests should be minimal.
Principle 7: Screening program infrastructure: There should be adequate infrastructure (e.g., financial resources, health human resources, information technology, facilities, equipment, test technology) for timely access to all components of the screening program (e.g., recruitment, testing, information access, diagnosis, referral, treatment, follow-up, patient education/support, staff training, program management/evaluation).	**Principle 7.** Screening program infrastructure: There should be adequate existing infrastructure (e.g., financial resources, health human resources, information technology, facilities, equipment, test technology) and/or a clear plan to develop adequate infrastructure that is appropriate to the setting to allow for timely access to all components of the screening program (e.g., recruitment, testing, information access, diagnosis, referral, treatment, follow-up, patient education/support, staff training, program management/evaluation).
Principle 8. Screening program coordination/integration: All components of the screening program (e.g., recruitment, testing, information access, diagnosis, referral, treatment, follow-up, patient education/support, staff training, program management/evaluation) should be coordinated and integrated with the broader health care system (including a formal system to inform, counsel, refer and manage screening participants) to optimize care continuity and ensure no screening participant is neglected	**Principle 8.** Screening program coordination/integration: All components of the screening program (e.g., recruitment, testing, information access, diagnosis, referral, treatment, follow-up, patient education/support, staff training, program management/evaluation) should be coordinated and, where possible, integrated with the broader health care system (including a formal system to inform, counsel, refer and manage screening participants) to optimize care continuity and ensure no screening participant is neglected.

Principle 9. Screening program acceptability: All components of the screening program (e.g., recruitment, testing, information access, diagnosis, referral, treatment, follow-up, patient education/support, staff training, program management/evaluation) should be clinically, socially, and ethically acceptable to screening participants, health professionals and society, and there should be effective methods for providing screening participants with informed choice, promoting their autonomy and protecting their rights.	**Principle 9.** Screening program acceptability/ethics: All components of the screening program (e.g., recruitment, testing, information access, diagnosis, referral, treatment, follow-up, patient education/support, staff training, program management/evaluation) should be clinically, socially, and ethically acceptable to screening participants, health professionals and society, and there should be effective methods for providing screening participants with informed choice, promoting their autonomy and protecting their rights.
Principle 10. Screening program benefits/harm: The expected range/magnitude of benefits (e.g., improved survival, function, quality of life) and harms (e.g., physical, psychological, social) for screening participants and society should be clearly defined, with high quality scientific evidence indicating that the overall benefit of the screening program outweighs its potential harms.	**Principle 10.** Screening program benefits/harms: The expected range/magnitude of benefits (e.g., increased functioning/quality of life, decreased cause-specific mortality) and harms (e.g., over-diagnosis, over-treatment) for screening participants and society should be clearly defined and acceptable, and supported by existing, high quality scientific evidence (or addressed by ongoing studies) that indicates that the overall benefit of the screening program outweighs its potential harms.
Principle 11. Economic evaluation of screening program: An economic evaluation (e.g., cost-effectiveness analysis, cost-benefit analysis, cost-utility analysis) of the screening program, conducted from a societal perspective, should be conducted to assess the full costs of implementing, operating and sustaining the screening program while clearly considering the opportunity costs and impact of allocating resources to other potential non-screening options (e.g., primary prevention, improved treatments, other clinical services) for managing the disease/condition.	**Principle 11.** Economic evaluation of screening program: An economic evaluation (e.g., cost-effectiveness analysis, cost-benefit analysis, cost-utility analysis) of the screening program, using a health system or societal perspective, should be conducted (or a clear plan to conduct an economic evaluation) to assess the full costs of implementing, operating and sustaining the screening program while clearly considering the opportunity costs and impact of allocating resources to other potential non-screening options (e.g., primary prevention, improved treatments, other clinical services) for managing the disease/condition.
Principle 12. Screening program quality and performance management: The screening program should have clear goals/objectives that are explicitly linked to program planning, monitoring, evaluating and reporting activities to ensure ongoing quality control and achievement of performance targets.	**Principle 12.** Screening program quality and performance management: The screening program should have clear goals/objectives that are explicitly linked to program planning, monitoring, evaluating and reporting activities, with dedicated information systems and funding to ensure ongoing quality control and achievement of performance targets.

Appendix 2B. Results Summary: Participant Characteristics

Fifty-six screening experts received an invitation to participate in the modified Delphi consensus process to assess the importance and interpretability of the consolidated principles developed through the systematic review. Of these, 18 (32%) responded to the Round 1 survey, with 16 indicating a willingness to participate in Round 2 and 12 ultimately responding to both the Round 1 and Round 2 surveys. The participants represented 11 countries, all fields/disciplines and diseases/conditions sought. Please see Table A2.2 for additional details on participant characteristics.

Table A2.2 Modified Delphi Consensus Process Participant Characteristics

Participant Characteristic	Round 1 Participants (n=18)		Round 2 Participants (n=12)	
	N	%	N	%
Age				
<45 years	3	17%	1	8%
45-64 years	11	61%	7	58%
65 years or more	4	22%	4	33%
Gender				
Female	8	44%	3	25%
Male	10	56%	9	75%
Country in which you have spent the majority of your professional career				
Canada	3	17%	1	8%
Denmark	1	6%		
Finland	1	6%	1	8%
Germany	1	6%	1	8%
Italy	1	6%	1	8%
Netherlands	1	6%	1	8%
New Zealand	1	6%		
Spain	1	6%	1	8%
Switzerland	1	6%	1	8%
United Kingdom	2	11%	2	17%
United States of America	5	28%	3	25%
Main areas of expertise related to screening				
Epidemiology	13	72%	9	75%
Health care management	2	11%	2	17%
Health economics	2	11%	2	17%
Health policy	9	50%	6	50%
Medicine	11	61%	6	50%
Public health	13	72%	8	67%
Main diseases/conditions that screening work focuses on				
Cancer	17	94%	11	92%
Infectious diseases	1	6%	1	8%
Newborn	5	28%	3	25%
Prenatal	3	17%	1	8%
Years of experience related to population-based screening research, evaluation or decision-making				
<15 years	3	17%	1	8%
15 years or more	15	83%	11	92%

Appendix 2C. Results Summary: Assessment of Importance of Each Consolidated Principle

For importance of the principles, the median level of agreement was high in both rounds, with agreement strengthening from Round 1 to Round 2. Only 1 participant in each Round indicated any level of disagreement with the importance of a principle (one participant noted disagreement with three principles in Round 1 and another participant noted disagreement with one principle in Round 2) (Tables A2.3 and A2.4). For most principles, measures of dispersion converged from Round 1 to Round 2 indicating stable consensus, with three exceptions (Principles 2, 4 and 8) (Table A2.5). Qualitative comments related to importance noted the need to consider health system arrangements (e.g., screening program integration/coordination principle may be less relevant to US health system context) and variable relevance of principles to specific types of screening test (e.g., genetic testing).

Table A2.3 Frequency Distribution for Level of Agreement of the Importance of Each Consolidated Principle (Round 1)

Round 1 (n=18) – Level of agreement with statement: *"This principle is important and relevant for population-based programmatic screening decisions"*

	strongly agree	agree	somewhat agree	neither agree nor disagree	somewhat disagree	disagree	strongly disagree
Principle 1	50.0%	38.9%	5.6%	0.0%	0.0%	5.6%	0.0%
Principle 2	52.9%	35.3%	11.8%	0.0%	0.0%	0.0%	0.0%
Principle 3	64.7%	23.5%	5.9%	0.0%	0.0%	5.9%	0.0%
Principle 4	58.8%	41.2%	0.0%	0.0%	0.0%	0.0%	0.0%
Principle 5	64.7%	29.4%	5.9%	0.0%	0.0%	0.0%	0.0%
Principle 6	58.8%	29.4%	11.8%	0.0%	0.0%	0.0%	0.0%
Principle 7	64.7%	23.5%	11.8%	0.0%	0.0%	0.0%	0.0%
Principle 8	50.0%	22.2%	22.2%	5.6%	0.0%	0.0%	0.0%
Principle 9	50.0%	44.4%	5.6%	0.0%	0.0%	0.0%	0.0%
Principle 10	58.8%	23.5%	5.9%	5.9%	0.0%	5.9%	0.0%
Principle 11	35.3%	41.2%	17.6%	5.9%	0.0%	0.0%	0.0%
Principle 12	76.5%	23.5%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A2.4 Frequency Distribution for Level of Agreement of the Importance of Each Consolidated Principle (Round 2)

Round 2 (n=12) – Level of agreement with statement: *"This principle is important and relevant for population-based programmatic screening decisions"*

	strongly agree	agree	somewhat agree	neither agree nor disagree	somewhat disagree	disagree	strongly disagree
Principle 1	83.3%	8.3%	8.3%	0.0%	0.0%	0.0%	0.0%
Principle 2	33.3%	50.0%	8.3%	8.3%	0.0%	0.0%	0.0%
Principle 3	72.7%	27.3%	0.0%	0.0%	0.0%	0.0%	0.0%

Principle 4	66.7%	25.0%	8.3%	0.0%	0.0%	0.0%	0.0%
Principle 5	54.5%	45.5%	0.0%	0.0%	0.0%	0.0%	0.0%
Principle 6	54.5%	45.5%	0.0%	0.0%	0.0%	0.0%	0.0%
Principle 7	25.0%	50.0%	25.0%	0.0%	0.0%	0.0%	0.0%
Principle 8	41.7%	33.3%	0.0%	16.7%	8.3%	0.0%	0.0%
Principle 9	75.0%	16.7%	8.3%	0.0%	0.0%	0.0%	0.0%
Principle 10	91.7%	8.3%	0.0%	0.0%	0.0%	0.0%	0.0%
Principle 11	50.0%	41.7%	8.3%	0.0%	0.0%	0.0%	0.0%
Principle 12	66.7%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A2.5 Descriptive Statistics for Level of Agreement and Stability of the Importance of Each Consolidated Principle[*¶]*

Round 1 (n=18) Importance				
Principle	**Mean**	**Median**	**Std. Dev.**	**Range**
Principle 1	6.2	6.5	1.22	5
Principle 2	6.4	7	0.71	2
Principle 3	6.4	7	1.27	5
Principle 4	6.6	7	0.51	1
Principle 5	6.6	7	0.62	2
Principle 6	6.5	7	0.72	2
Principle 7	6.5	7	0.72	2
Principle 8	6.2	6.5	0.99	3
Principle 9	6.4	6.5	0.62	2
Principle 10	6.2	7	1.38	5
Principle 11	6.1	6	0.90	3
Principle 12	6.8	7	0.44	1

Round 2 (n=12) Importance				
Principle	**Mean**	**Median**	**Std. Dev.**	**Range**
Principle 1	6.8	7	0.62	2
Principle 2	6.1	6	0.90	3
Principle 3	6.7	7	0.47	1
Principle 4	6.6	7	0.67	2
Principle 5	6.5	7	0.52	1
Principle 6	6.5	7	0.52	1
Principle 7	6.0	6	0.74	2
Principle 8	5.8	6	1.40	4
Principle 9	6.7	7	0.65	2
Principle 10	6.9	7	0.29	1
Principle 11	6.4	6.5	0.67	2
Principle 12	6.7	7	0.49	1

Round 2 - Round 1			
Change in Mean	**Change in Median**	**Change in Std. Dev.**	**Change in Range**
0.53	0.50	-0.59	-3.00
-0.33	-1.00	0.19	1.00
0.37	0.00	-0.80	-4.00
0.00	0.00	0.16	1.00
-0.04	0.00	-0.10	-1.00
0.07	0.00	-0.20	-1.00
-0.53	-1.00	0.02	0.00
-0.33	-0.50	0.42	1.00
0.22	0.50	0.04	0.00
0.74	0.00	-1.09	-4.00
0.36	0.50	-0.23	-1.00
-0.10	0.00	0.06	0.00

*Responses re-coded 'strongly agree' = 7, 'agree' = 6, 'somewhat agree' = 5, 'neither agree nor disagree' = 4, 'somewhat disagree' = 3, 'disagree' = 2, 'strongly disagree' = 1.

[¶]Shaded cell indicates change was unfavourable in terms of agreement/stability

Appendix 2D. Results Summary: Assessment of Interpretability of Each Consolidated Principle

For interpretability, the median level of agreement was also high, but generally lower than for importance. In Round 1, there was some disagreement regarding the interpretability of the principles with seven different participants indicating disagreement with at least one principle, and all but two principles receiving some level of disagreement (exceptions were Principles 8 and 12) (Tables A2.6 and A2.7). For Round 2, four participants indicated some level of disagreement for six of the 12 principles. The interpretability rankings improved from Round 1 to Round 2 for all but two principles and measures of dispersion narrowed for all but one principle (Principle 8) (Table A2.8). Most of the qualitative feedback related to interpretability suggested edits to the wording or language. A few comments recommended splitting up some principles while one participant raised concerns about the need for clarity in describing screening tests vs. screening processes, vs. screening programs, and another participant suggested that examples could help to illustrate the principles.

Table A2.6 Frequency Distribution for Level of Agreement of the Interpretability of Each Consolidated Principle (Round 1)

Round 1 (n=18) – Level of agreement with statement: *"This principle is clearly defined and understandable"*

	strongly agree	agree	somewhat agree	neither agree nor disagree	somewhat disagree	disagree	strongly disagree
Principle 1	33.3%	22.2%	5.6%	5.6%	0.0%	27.8%	5.6%
Principle 2	22.2%	27.8%	27.8%	0.0%	16.7%	0.0%	5.6%
Principle 3	58.8%	11.8%	11.8%	0.0%	11.8%	5.9%	0.0%
Principle 4	29.4%	41.2%	0.0%	11.8%	5.9%	11.8%	0.0%
Principle 5	41.2%	47.1%	5.9%	0.0%	0.0%	5.9%	0.0%
Principle 6	22.2%	38.9%	22.2%	0.0%	5.6%	5.6%	5.6%
Principle 7	44.4%	27.8%	22.2%	0.0%	5.6%	0.0%	0.0%
Principle 8	44.4%	38.9%	11.1%	5.6%	0.0%	0.0%	0.0%
Principle 9	38.9%	27.8%	16.7%	0.0%	5.6%	11.1%	0.0%
Principle 10	23.5%	35.3%	17.6%	11.8%	5.9%	5.9%	0.0%
Principle 11	27.8%	33.3%	16.7%	5.6%	11.1%	0.0%	5.6%
Principle 12	70.6%	17.6%	5.9%	5.9%	0.0%	0.0%	0.0%

Table A2.7 Frequency Distribution for Level of Agreement of the Importance of Each Consolidated Principle (Round 2)

Round 2 (n=12) – Level of agreement with statement: *"This principle is clearly defined and understandable"*

	strongly agree	agree	somewhat agree	neither agree nor disagree	somewhat disagree	disagree	strongly disagree
Principle 1	41.7%	25.0%	25.0%	0.0%	0.0%	8.3%	0.0%
Principle 2	8.3%	41.7%	25.0%	16.7%	8.3%	0.0%	0.0%
Principle 3	45.5%	36.4%	9.1%	0.0%	9.1%	0.0%	0.0%

Principle 4	33.3%	25.0%	41.7%	0.0%	0.0%	0.0%	0.0%
Principle 5	45.5%	45.5%	9.1%	0.0%	0.0%	0.0%	0.0%
Principle 6	18.2%	54.5%	9.1%	0.0%	18.2%	0.0%	0.0%
Principle 7	33.3%	41.7%	25.0%	0.0%	0.0%	0.0%	0.0%
Principle 8	33.3%	41.7%	16.7%	0.0%	8.3%	0.0%	0.0%
Principle 9	41.7%	33.3%	16.7%	8.3%	0.0%	0.0%	0.0%
Principle 10	41.7%	33.3%	25.0%	0.0%	0.0%	0.0%	0.0%
Principle 11	33.3%	41.7%	16.7%	0.0%	8.3%	0.0%	0.0%
Principle 12	41.7%	58.3%	0.0%	0.0%	0.0%	0.0%	0.0%

*Table A2.8 Descriptive Statistics for Level of Agreement and Stability of the Interpretability of Each Consolidated Principle**¶*

	Round 1 (n=18) Interpretability				Round 2 (n=12) Interpretability					Round 2 – Round 1			
Principle	**Mean**	**Median**	**Std. Dev.**	**Range**	**Principle**	**Mean**	**Median**	**Std. Dev.**	**Range**	**Change in Mean**	**Change in Median**	**Change in Std. Dev.**	**Change in Range**
Principle 1	4.8	6	2.29	6	**Principle 1**	5.8	6	1.47	5	1.06	0.00	-0.82	-1.00
Principle 2	5.2	5.5	1.69	6	**Principle 2**	5.3	5.5	1.14	4	0.08	0.00	-0.55	-2.00
Principle 3	5.9	7	1.69	5	**Principle 3**	6.1	6	1.22	4	0.21	-1.00	-0.47	-1.00
Principle 4	5.4	6	1.73	5	**Principle 4**	5.9	6	0.90	2	0.51	0.00	-0.83	-3.00
Principle 5	6.1	6	1.22	5	**Principle 5**	6.4	6	0.67	2	0.25	0.00	-0.54	-3.00
Principle 6	5.3	6	1.72	6	**Principle 6**	5.5	6	1.37	4	0.21	0.00	-0.35	-2.00
Principle 7	6.1	6	1.11	4	**Principle 7**	6.1	6	0.79	2	0.03	0.00	-0.32	-2.00
Principle 8	6.2	6	0.88	3	**Principle 8**	5.9	6	1.17	4	-0.31	0.00	0.29	1.00
Principle 9	5.6	6	1.69	5	**Principle 9**	6.1	6	1.00	3	0.47	0.00	-0.69	-2.00
Principle 10	5.4	6	1.46	5	**Principle 10**	6.2	6	0.83	2	0.76	0.00	-0.63	-3.00
Principle 11	5.4	6	1.69	6	**Principle 11**	5.9	6	1.17	4	0.53	0.00	-0.52	-2.00
Principle 12	6.5	7	0.87	3	**Principle 12**	6.4	6	0.51	1	-0.11	-1.00	-0.36	-2.00

*Responses re-coded 'strongly agree' = 7, 'agree' = 6, 'somewhat agree' = 5, 'neither agree nor disagree' = 4, 'somewhat disagree' = 3, 'disagree' = 2, 'strongly disagree' = 1.

¶Shaded cell indicates change was unfavourable in terms of agreement/stability

Appendix 2 References:

1. Okoli C and Pawlowski S. The Delphi method as a research tool: An example, design considerations and applications. *Information and Management* 2004;42:15-29.
2. Hsu C and Sandford B. The Delphi technique: making sense of consensus *Practical Assessment, Research and Evaluation* 2007;12:1-8.
3. von der Gracht HA. Consensus measurement in Delphi studies: review and implications for future quality assurance. *Technological Forecasting and Social Change* 2012;79:1525-36.
4. Holey EA, Feeley JL, Dixon J, et al. An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Medical Research Methodology* 2007;7:1-10.