Knottnerus, J.A., Knottnerus, B.J. Decision-making given surrogate outcomes. Journal of Clinical Epidemiology: 2022, 145, p. 174-178

Postprint version	:	2.0
Journal website	:	https://www.jclinepi.com/article/S0895-4356(22)00009-9/fulltext
Pubmed link	:	https://pubmed.ncbi.nlm.nih.gov/35041971/
DOI	:	10.1016/j.jclinepi.2022.01.003

This is a Nivel certified Post Print, more info at nivel.nl

Decision-making given surrogate outcomes

J. André Knottnerus, Bart J. Knottnerus

[1] Professor emeritus of Primary Health Care, Department of Family Practice, Maastricht University Medical Centre, Maastricht, The Netherlands

[2] Senior researcher, primary care physician, Netherlands Institute for Health Services Research (Nivel), Utrecht, The Netherlands

Abstract

Background: Opinions differ about the extent to which intervention research should and can directly assess the main patient-important health outcomes, what role surrogate endpoints can play, and which requirements should then apply to the scientific underpinning of clinical and policy decisions.

Method: In a commentary we elaborate on this and provide guidance for dealing with related dilemmas.

Conclusions: Ethical, methodological and practical reasons for decision making based on surrogate endpoints can be that (1) reaching the intended patient-important health outcome would take too long to await direct RCT-based evidence, (2) experimental conditions have limited sustainability over time; and (3) the plausibility of an intervention's clinical efficacy, given the already available evidence regarding surrogate endpoints, goes beyond equipoise.

Given an expected increase of interventions with a long term patient-important health outcome perspective, dealing with surrogate endpoints will remain an important challenge.

Appropriately dealing with a surrogate endpoint includes (1) the assessment of its predictive value for the intended patient-important outcome, where GRADE guidelines for assessing 'indirectness' and 'causal chain analysis' can be helpful; (2) transparency of (absence of) evidence; (3) adequately updating the 'knowledge mosaic'; (4) weighing different perspectives and values, and (5)
monitoring whether adjustments need to be made.

The remaining level of uncertainty must be balanced against the urgency of clinical or societal decision making and the disadvantages of postponing this.

Criteria for using surrogate endpoints are suggested. Patients, citizens and policy makers can be involved in agreeing upon these criteria.

Introduction

Evidence-based medicine (EBM) generally requires that interventions, before being implemented, have been proven to benefit health. Does this mean that we should in principle discourage the use of surrogate outcome endpoints (biomarkers or physical signs that are used as a substitute for a patient-important outcome) and only use patient-important health outcomes such as mortality, clinically manifest morbidity, disability, and health-related quality of life? Or would this hinder the improvement and progress of healthcare? In addressing these questions, opinions differ about the extent to which intervention research should and can directly assess the main patient-important health outcomes, and about what role surrogate endpoints can play.[1,2,3,4] Linked to this is the question which requirements should apply in this regard to the scientific underpinning of clinical and policy decisions about healthcare intervention. In this commentary, we further elaborate on this and provide guidance for dealing with related dilemmas.

Achieving final health outcomes may take too long

For many morbidity patterns and intervention types, the follow-up time required for allowing the final health outcome of interest to occur is long. For example, vaccinating girls against human papillomavirus (HPV) reduces the occurrence of cervical intraepithelial neoplasia grade 2+ after some years,[5] while effects on cervical cancer incidence and survival can only be expected after decades. Population screening may detect risk groups and abnormalities at an early stage, but evaluating its impact regarding mortality and overdiagnosis needs much longer follow-up.[6,7] Primary care and occupational health interventions can favourably influence cardiometabolic disease risk profiles,[8] physical exertion[9] and smoking[10] after one year follow-up, but effects on morbidity and mortality can only be observed on the long term. And general practice-focused interventions can directly improve quality of care, such as antibiotic prescribing,[11] but how this will affect the health of the practice population, also in terms of antibiotic resistance, can only be measured after years.

In situations of demonstrated intermediate effects like those just mentioned, awaiting direct RCTbased evidence regarding effects on morbidity and mortality before making healthcare or policy decisions can be difficult and even controversial. On the one hand professionals, politicians and industry can exert pressure to introduce an intervention in practice, pointing to promising effects on intermediate outcomes without health improvement being directly observed in a trial. On the other hand, it is ethically debatable to postpone decisions about introducing an intervention for years or even decades after such effects are shown.

In such cases, responsible practitioners and policy makers cannot escape from weighing the risk of introducing a favourable intervention too late or not at all, thereby missing potential health gain via better prevention and care, against the risk of introducing that intervention ‘false-positively’, that is, ultimately without a positive health impact or having even a negative impact via adverse effects, overdiagnosis or resource waste. But this balancing act should be scientifically well informed, and linked to systematic surveillance, long term evaluation, and correction when necessary.[12]
Experimental conditions

Experimental conditions have limited sustainability over time

A practical-methodological problem is that sustaining an experimental setting – including the contrast between the intervention and the non-intervention or usual care condition - over a very long time is difficult or even impossible. Acceptance of an intervention given promising intermediate outcomes (as happened with HPV vaccination) will make the study contrast untenable. Emerging new technologies can even make it outdated, as policy makers, practitioners, and researchers will lose interest in the long-term outcome of a ‘former generation technology’, and patient organisations

request implementation of the more innovative interventions. This can lead to collapse of a longterm running study. Theoretically, a study can be continued while adding a parallel study evaluating a new technology, but a study on 'old technology' when new technology is available is clinically and ethically questionable. But even a completely new study, focused on the most innovative technology, will after some time itself be confronted with challenges related to even more advanced technology. So, in dynamic fields like these, the best achievable strategy may be continuous evaluation after introduction based on acceptable surrogate results.[8]

Plausibility beyond equipoise

Also in the evaluation of interventions with shorter term final health outcomes, already available evidence on intermediate outcomes can be a complicating factor. An example is influenza vaccination. While RCT results have shown that vaccination of the elderly can reduce the incidence of influenza with about 50%,[13] direct RCT-based evidence on the reduction of influenza-related complications (severe morbidity and mortality) is not available. But if it is considered plausible that reduction of influenza incidence will also lead to reduction of the incidence of its complications, would then a new placebo-controlled trial with severe complications as primary outcome be justified?[14] In our view, respecting the principle of equipoise, such a new, necessarily huge trial is ethically debatable if it would withhold many people an intervention halving the incidence of a disorder that can have severe complications.[15]

Society must deal with surrogate endpoints every day

Outside the health domain societal decision making based on surrogate outcomes is the rule rather than the exception. That does not mean that RCT-based policy making should not be pursued and applied much more in other domains,[16] but there are many cases in which awaiting (direct) evidence on final outcomes of interest before acting would last too long, be infeasible or even impossible. Examples are: evaluating school achievements as assumed predictor of professional career success; reduction of carbon dioxide emissions in the expectation that this will positively affect climate change and survival of mankind; and tightening up governmental supervision to improve banks' behaviour, in order to prevent new economic crises. Clearly, society accepts decision making based on surrogate outcomes if there is no reasonable alternative.

The challenge

In the medical and healthcare domain we expect many innovative interventions with a long term patient-important health outcome perspective, such as new vaccines against cancer and degenerative disorders,[17] molecular genetic screening methods,[18,19] and internet-based behavioural interventions.[20] This will be accompanied with an increase of potentially relevant surrogate endpoints along presumed or plausible causal pathways, in combination with the aforementioned obstacles to always conducting huge and long-term trials focused on patient-important health outcomes.

In view of these expectations, while fully RCT-based evidence focused on the intended patientimportant health outcome is preferable where possible, we should not reject the use of surrogate endpoints when these are the best we can get and if we want to avoid stagnation of healthcare, scientific and technological and progress. How to deal with surrogate outcomes in research and practice is an increasingly relevant challenge.

Guidance

For addressing this challenge, there is not one perfect, universally applicable approach. We need well-tailored, evidence-informed decision making, using the intended patient-important health

outcome where possible, but appropriately dealing with uncertainty where using surrogate outcomes is unavoidable. In support of the latter, we propose the following guidance.

Predictive value

The key question is how predictive a surrogate endpoint is for the intended health outcome of interest. The 'gold standard' procedure to clarify this is relating the surrogate to the intended patient-important outcome.[21,22] But if fully reaching the latter in one and the same study is impossible or infeasible, indirect estimations must be applied. In cases such as stopping smoking as a predictor of better health[9] this is associated with less uncertainty and outcome misclassification than in other cases such as early detection of an abnormality or disorder as a predictor of mortality reduction.[6] And using precancerous lesions as surrogate endpoints is likely to be better than just using blood test results.

In other words, surrogate endpoints are not homogeneous: some are (much) better than others, and often this is a matter of judgement as well as empirical evidence. In dealing with this type of 'indirectness' the GRADE guidelines can be helpful by rating down the quality of the evidence, generally by one or even two levels, when a surrogate endpoint has been used. Biomedical and clinical knowledge of the condition under study can be supportive in assessing how close the surrogate endpoints are to the putative causal pathway, and thereby in determining the degree of rating down.[23] Also, causal chain analysis to depict a causal chain of events, as is becoming increasingly common in systematic reviews of complex interventions, can help with explicitly identifying and evaluating the role of specific surrogate endpoins.[24]

The remaining level of uncertainty must then be balanced against the urgency of clinical or societal decision making and the disadvantages of postponing this.

Transparency of (absence of) evidence

To accomplish this balancing act in the best possible way, scientists (in planning research), practitioners (in making healthcare decisions), and policy makers (in taking their responsibilities) must be transparently provided with all relevant (indirect) evidence and uncertainties regarding the intervention's potential impact on final health outcomes. This should include what is known, not (yet) known and assumed about the association between the surrogate endpoint and the final patient-important health outcome, the predictive value of the first as to the latter, and what experts consider plausible expectations. In this context, systematic processes of expert judgment[25] must be applied, independent of commercial and non-commercial interests. In support of this, advanced statistical methods[26] and systems to rate the quality and validity of surrogate outcomes can be helpful.[23]

The knowledge mosaic

To properly assess the predictive value of the surrogate for the intended outcome in the absence of direct evidence, much insight is needed into the entire 'knowledge mosaic' of biology, pathophysiology, clinical phenomena and, where relevant, psychosocial mechanisms. Updating this predictive value, given new research, requires integrating the new knowledge pieces in the larger mosaic. It is not always necessary – if feasible at all - to carry out a new RCT every time a new mosaic piece comes up.[27] For example, if a more accurate breast screening test becomes available, its impact on the intended patient-important outcome can be assessed by substituting the outdated by the new accuracy data and connecting the latter to the current knowledge of effectiveness of early breast cancer treatment, using clinical decision analysis.[28]

Weighing different perspectives and values

In situations of uncertainty, in addition to expert judgment, the perspectives of patients, the public, policy makers and politicians should weigh heavily, especially when different values may lead to different conclusions.[29,30] In this context it makes a difference whether individual decisions (for 5 example, the decision whether or not to be vaccinated against HPV) or collective decisions (such as making vaccination mandatory in view of herd immunity) are at stake.[31] In the latter case a democratic majority decision must be taken, requiring that the public is optimally informed. When individual freedom of decision can be in conflict with collective interests, it is crucial to separate the roles of those who are responsible for producing evidence, those who advise individual subjects, and those who independently advise on general guidelines or policies.

Monitoring

Given inherent uncertainties, results after decisions based on surrogate endpoints must be carefully monitored to see whether one is on the right track or whether adjustments need to be made. This can be supported by interim analyses, and repeated review and scientific advice. This is, for example, the policy in the Netherlands regarding HPV vaccination.8 Instead of fearing such critical evaluations because these may contradict the primary decisions, policy makers should invest in them as cornerstones of good policy practice given uncertainties.

Agreeing upon criteria

Criteria for using surrogate endpoints, given the state of knowledge on their predictive values, can help as a basis for urgent decision making in the temporary or permanent absence of sufficient final patient-important endpoint data. Patients, citizens and policy makers can be involved in agreeing upon these criteria, which should cover:

- the urgency of taking a clinical or policy decision about the intervention to be evaluated;
- the impossibility or infeasibility of measuring the final patient-important health outcome, given the relevant time window, methodological limitations and ethical considerations;
- the biological and clinical plausibility of the proposed surrogate endpoint;
- a maximally evidence-based, credible assessment of its predictive value, an assessment of residual uncertainties left including a GRADE assessment of the quality of the evidence, with a motivation as to whether this is sufficient for making a decision;
- transparent weighting of the arguments for and against using the proposed surrogate endpoint from a scientific, clinical and societal perspective;
- a solid method for monitoring, evaluation and adaptation if needed;

If agreed criteria are not met, it is important to report that clearly too, in order to avoid misuse of surrogate endpoints such as unjustified limitation of endpoint measurements to cheap laboratory tests or using irrelevant short-term effects. Moreover, such reporting can stimulate research to collect better – and preferably direct - evidence about endpoints where necessary and possible.

Conclusion: cooperation of science, practice and society

The all-embracing question is which position science, healthcare practice and society must choose regarding decision making on (introducing) interventions when direct evidence on the intended patient-important health outcomes is lacking. Do we set the strict requirement of direct RCT-based evidence on such final outcomes in all cases, also when we know that such evidence is unlikely or even impossible to be completely available when clinical or policy decisions are required? Or do we make scientific, professional and societal agreements about what is considered the best possible and minimally required evidence, keeping a close eye on post-decision developments, and using all opportunities to learn and adapt? We believe that, in a context of ever-faster knowledge

development and societal decision making, the latter approach is best for continuously improving prevention and care, and for health care progress and innovation. For achieving this, appropriate guiding principles and good cooperation between science, practice and society are paramount.

Conflict of interest

The author reports no conflict of interest

References

1. Psaty BM, Weiss NS, Furberg CD, Koepsell TD et al. Surrogate end points, health outcomes, and the drug approval process for the treatment of risk factors for cardiovascular disease. JAMA 1999; 282 (8): 786–790. doi:10.1001/jama.282.8.786. PMID 10463718.
2. Dobler CC, Morgan RL, Falck-Ytter Y, Montori VM, Murad MH. Assessing the validity of surrogate endpoints in the context of a controversy about the measurement of effectiveness of hepatitis C virus treatment. BMJ Evid Based Med. 2018 Apr;23(2):50-53. doi: 10.1136/bmjebm-2017-110852.
3. Ronald L Koretz, Janus Christian Jakobsen, Christian Gluud. Letter to the editor re: Assessing the validity of surrogate endpoints in the context of a controversy about the measurement of effectivenes of hepatitis C virus treatment. BMJ Evid Based Med. 2018 Dec;23(6):240. doi: 10.1136/bmjebm-2018-111072. Epub 2018 Aug 31.
4. Bucher HC. Surrogate endpoint trials: benefit and pitfalls for clinical decision making]. [Article in German]. Internist (Berl). 2008 Jun;49(6):681-7. doi: 10.1007/s00108-008-2126-8.
5. Drolet M, Bénard É, Pérez N, Brisson M; HPV Vaccination Impact Study Group. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and metaanalysis. Lancet. 2019 Aug 10;394(10197):497-509. doi: 10.1016/S0140- 6736(19)30298-3. Epub 2019 Jun 26.
6. McKay AJ, Hogan H, Humphries SE, Marks D, Ray KK, Miners A. Universal screening at age 1-2 years as an adjunct to cascade testing for familial hypercholesterolaemia in the UK: A cost-utility analysis. Atherosclerosis. 2018 Aug;275:434-443. doi: 10.1016/j.atherosclerosis.2018.05.047. Epub 2018 Jun 4.
7. Johansson M, Brodersen J, Gøtzsche PC, Jørgensen KJ. Screening for reducing morbidity and mortality in malignant melanoma. Cochrane Database Syst Rev. 2019 Jun 3;6:CD012352. doi: 10.1002/14651858.CD012352.pub2.
8. Stol DM, Badenbroek IF, Hollander M, Nielen MMJ, Kraaijenhagen RA, Schellevis FG, de Wit NJ. Effectiveness of a stepwise cardiometabolic disease prevention program: Results of a randomized controlled trial in primary care. Prev Med. 2020 Jan 16:105984. doi: 0.1016/j.ypmed.2020.105984. [Epub ahead of print]
9. Edwardson CL, Yates T, Biddle SJH, Davies MJ, Dunstan DW, Esliger DW, Gray LJ, Jackson B, O'Connell SE, Waheed G, Munir F .Effectiveness of the Stand More AT (SMArT) Work intervention: cluster randomised controlled trial. BMJ. 2018 Oct 10;363:k3870. doi: 10.1136/bmj.k3870.
10. van den Brand FA, Nagelhout GE, Winkens B, Chavannes NH, van Schayck OCP, Evers SMAA. Cost-effectiveness and cost-utility analysis of a work-place smoking cessation intervention with and without financial incentives. Addiction. 2019 Dec 17. doi: 10.1111/add.14861.7
11. de Bont EGPM, Dinant GJ, Elshout G, van Well G, Francis NA, Winkens B, Cals JWL. Booklet for Childhood Fever in Out-of-Hours Primary Care: A Cluster-Randomized Controlled Trial. Ann Fam Med. 2018 Jul;16(4):314-321. doi: 10.1370/afm.2265.
12. Health Council of the Netherlands. Vaccination against HPV. The Hague: Health Council of the Netherlands, 2019; publication no. 2019/09. https://www.healthcouncil.nl/documents/advisory-reports/2019/06/19/vaccinationagainst-hpv

13. Govaert TM, Thijs CT, Masurel N, Sprenger MJ, Dinant GJ, Knottnerus JA. The efficacy of influenza vaccination in elderly individuals. A randomized double-blind placebocontrolled trial. JAMA. 1994 Dec 7;272(21):1661-5.
14. Knottnerus JA. Influenza vaccination in the elderly: current evidence and uncertainties. J Clin Epidemiol. 2009 Jul;62(7):675-6. doi: 10.1016/j.jclinepi.2008.10.006. Epub 2009 Feb 10.
15. Verhees RAF, Dondorp W, Thijs C, Dinant GJ, Knottnerus JA. Influenza vaccination in the elderly: Is a trial on mortality ethically acceptable? Vaccine. 2018 May 17;36(21):2991-2997. doi: 10.1016/j.vaccine.2018.04.027. PMID: 29691101 https://campbellcollaboration.org/
16. Malonis RJ, Lai JR, Vergnolle O. Peptide-Based Vaccines: Current Progress and Future Challenges. Chem Rev. 2019 Dec 5. doi: 10.1021/acs.chemrev.9b00472.
17. Korf BR, Rehm HL. New approaches to molecular diagnosis. JAMA. 2013 Apr 10;309(14):1511-21. doi: 10.1001/jama.2013.3239.
18. Jaffee EM, Dang CV, Agus DB, Alexander BM, Anderson KC, Ashworth A, Barker AD, Bastani R, Bhatia S, Bluestone JA, Brawley O, Butte AJ, Coit DG et al. Future cancer research priorities in the USA: a Lancet Oncology Commission. Lancet Oncol. 2017 Nov;18(11):e653-e706. doi: 10.1016/S1470-2045(17)30698-8.
19. David C Mohr, Stephen M Schueller, William T Riley, Hendricks Brown, Pim Cuijpers, Naihua Duan, Mary J Kwasny, Colleen Stiles-Shields, and Ken Cheung. Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies. J Med Internet Res. 2015 Jul; 17(7): e166.. doi: 10.2196/jmir.4391
20. Michiels S, Le Maître A, Buyse M, Burzykowski T, Maillard E, Bogaerts J, Vermorken JB, Budach W, Pajak TF, Ang KK, Bourhis J, Pignon JP; MARCH and MACH-NC Collaborative Groups. Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data. Lancet Oncol. 2009 Apr;10(4):341-50. doi: 10.1016/S1470-2045(09)70023-3.
21. Heerspink HJL, Greene T, Tighiouart H, Gansevoort RT, Coresh J, Simon AL, Chan TM, Hou FF, Lewis JB, Locatelli F, Praga M, Schena FP, Levey AS, Inker LA; Chronic Kidney Disease Epidemiology Collaboration. Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomized clinical trials. Lancet Diabetes Endocrinol. 2019 Feb;7(2):128-139. doi: 10.1016/S2213-8587(18)30314-0.
22. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011 Dec;64(12):1303-10. doi: 10.1016/j.jclinepi.2011.04.014.
23. Kneale D, Thomas J, Bangpan M, Waddington H, Gough D. Conceptualising causal pathways in systematic reviews of international development interventions through 8 adopting a causal chain analysis approach. Journal of Development Effectiveness 2018: 10: 422-437 https://doi.org/10.1080/19439342.2018.1530278
24. Tong A, Manns B, Wang AYM, Hemmelgarn B, Wheeler DC, Gill J, Tugwell P, Pecoits- Filho R, Crowe S, Harris T, Van Biesen W, Winkelmayer WC, Levin A, Thompson A, Perkovic V, Ju A, Gutman T, Bernier-Jean A, Viecelli AK, O'Lone E, Shen J, Josephson MA, Cho Y, Johnson DW, Sautenet B, Tonelli M2, Craig JC19; SONG Implementation Workshop Investigators. Implementing core outcomes in kidney disease: report of the Standardized Outcomes in Nephrology (SONG) implementation workshop. Kidney Int. 2018 Dec;94(6):1053-1068. doi: 10.1016/j.kint.2018.08.018.
25. Alonso A, Van der Elst W, Meyvisch P. Assessing a surrogate predictive value: a causal inference approach. Stat Med. 2017 Mar 30;36(7):1083-1098. doi: 10.1002/sim.7197.
26. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. BMJ. 2002 Feb 23;324(7335):477-80.
27. Lee JM, Kopans DB, McMahon PM, Halpern EF, Ryan PD, Weinstein MC, Gazelle GS.

28. Breast cancer screening in BRCA1 mutation carriers: effectiveness of MR imaging-- Markov Monte Carlo decision analysis. Radiology. 2008 Mar;246(3):763-71. doi: 10.1148/radiol.2463070224..
29. Soobiah C, Manley G, Marr S, Moore A, PausJenssen E, Teare S, Hamid J, Tricco AC, Straus S. A modified Delphi and cross-sectional survey to facilitate selection of optimal outcomes and measures for a systematic review on geriatrician-led care models. J Clin Epidemiol. 2019 May;109:117-124. doi: 10.1016/j.jclinepi.2019.02.005.
30. Knottnerus BJ. Cancer screening is not only about numbers. Eur J Cancer Care 2017;26(3):e12700.
31. Ellen Wolff, K Miriam Elfström, Hedda Haugen Cange, Sofie Larsson, Helene Englund, Pär Sparén, Adam Roth. Cost-effectiveness of Sex-Neutral HPV-vaccination in Sweden, Accounting for Herd-Immunity and Sexual Behaviour. Vaccine, 36 (34), 5160-5165 2018 Aug 16