Looking for trouble?

Diagnostics expanding disease and producing patients

Running title: **Looking for trouble?**

Bjørn Hofmann PhD [1,2]

1. Institute for the Health Sciences at the Norwegian University of Science and Technology (NTNU) at Gjøvik, Norway

2. Centre of Medical Ethics at the University of Oslo, Norway

Corresponding author:

Bjørn Hofmann,

Centre for Medical Ethics

University of Oslo

PO Box 1130, Blindern

N-0318 Oslo, Norway

b.m.hofmann@medisin.uio.no

Financial support: There is no external financing of this research.

Word count of manuscript text: 2156

Key Words: overdiagnosis, diagnostic error, harm, concept of disease, indicator

Looking for trouble?

Diagnostics expanding disease and producing patients

Abstract:
Novel tests give great opportunities for earlier and more precise diagnostics. At the same time, new tests expand disease, produce patients, and cause unnecessary harm in terms of overdiagnosis and overtreatment. How can we evaluate diagnostics in order obtain the benefits and avoid the harms? One way is to pay close attention to the diagnostic process and its core concepts. Doing so reveals three errors that expand disease and increase overdiagnosis. The first error is to decouple diagnostics from harm, e.g., by diagnosing insignificant conditions. The second error is to bypass proper validation of the relationship between test indicator and disease, e.g., by introducing biomarkers for Alzheimer's disease before the tests are properly validated. The third error is to couple the name of disease to insignificant or indecisive indicators, e.g., by lending the cancer name to preconditions, such as *ductal carcinoma in situ*. We need to avoid these errors in order to promote beneficial testing, bar harmful diagnostics, and evade unwarranted expansion of disease. Accordingly, we must stop identifying and testing for conditions that are only remotely associated with harm. We need more stringent verification of tests and, we must avoid naming indicators and indicative conditions after diseases. If not, we will end like ancient tragic heroes, succumbing because of our very best abilities.

Summary points:

- Uncritical enthusiasm for early detection expands disease and results in overdiagnosis and overtreatment. There are three reasons for this:
    1. Test indicators are decoupled from harm: conditions that may not lead to disease are defined as disease and handled as harmful.
    2. The relationship between indicator and harm is poorly verified.
    3. Indicators are given names after the diseases they indicate, and thus appear more alarming than they are.
- To avoid this, we need to take three actions:
    1. We must stop identifying and testing for conditions that are only remotely associated with harm. In particular, we need to explain to patients that test indicators are indicators and not disease.
    2. We must apply stringent verifications of tests.
    3. We must avoid naming indicators and indicative conditions after diseases.

Three leaps of faith to avoid:

- From test indicator (e.g. polyps) to disease.
- From early (detection) to better health and no harm.
- From being able to do something to doing something good.

Introduction: what is the problem?

Novel tests provide fantastic opportunities for early diagnostics, enabling us to predict and prevent disease or to reduce its harm. Despite good intentions and promising results early diagnostics is increasingly criticized for creating overdiagnosis, i.e., the detection of (a condition that is classified as) disease without benefit for the person in terms of reduced morbidity or mortality, or increased quality of life [1-4]. No doubt, finding indicators, precursors, or signs of disease early can be life-saving to patients. However, in cases where persons would not have experienced symptoms, suffering, or reduced quality or length of life, if such entities were not identified, they may in fact be harmed by (subsequent) testing and unnecessary treatment. Additionally, it represents a misuse of resources.

Why has overdiagnosis increased? The first and trivial answer is that awareness has replaced previous neglect. Second, definitions of disease have expanded [5, 6]. By increasing sensitivity or lowering diagnostic thresholds we are identifying ever milder conditions, and by looking harder we find more[7, 8]. Third, we also find new signs, indicators, and markers to identify potential disease,[9] e.g., by using prostate-specific antigen (PSA) to identify prostate cancer risk or using multi-analyte blood tests to identify a wide range of cancers.[10] Fourth, new or different diagnostic methods are introduced to detect disease [5]. For example, CT angiography has replaced ventilation-perfusion (VQ) scanning, for identifying pulmonary embolism, doubling the number of persons being diagnosed [11]. Fifth, we look for and handle risk factors as diseases[12, 13]. Sixth, there is a strong and not always warranted belief in early detection and a cognitive bias that “early is better”[14].

The core of the problem is that we do not know whether what we detect will actually develop into disease [15]. More precisely, it is uncertain whether the *indicator* that we identify with our tests (also referred to as indicative phenomenon, indicating condition, or pre-disease)[8], such as fecal occult blood, polyps, ductal carcinoma in situ (DCIS), or biomarkers, will actually result in symptoms, disease, and death. **Table 1** shows some indicators with corresponding diseases and harms.

[Insert **Table 1** here]

Where does our very reasonable and highly plausible strategy of early detection go wrong? At what point do we expand disease beyond vindication? Where do we become tragic heroes failing to use our best assets for the best interests of patients? It seems we go astray at three crucial points:

1. when we *decouple diagnostics from harm*
2. when we *alter validation criteria*
3. when we *couple the name of disease to indicative phenomena of great uncertainty*.

1. Decoupling tests from harm

Diseases are traditionally established by connecting certain (biological) characteristics to harm and suffering. Blood in the stool, changes in bowel movements, weight loss, tiredness, and tumor correspond to pain and reduced functioning and eventually results in death of what is given the disease label *colorectal cancer* (CRC). Fairly robust methods for validation have warranted the connection between disease and suffering. Accordingly, overdiagnosis has not been a great issue.

However, as we find indicators, such as PSA, DCIS, and biomarkers, which are less firmly related to harm, and thus can result in overdiagnosis and overtreatment. Although we try to validate the relationship between indicators and disease by various "gold standard tests," such as biopsies, the connection to harm may be vague or unauthentic.

2. Altering validation criteria

Accordingly, we evade the traditional validation of how a test indicator relates to harm. New technologies may provide new standards for detecting diseases resulting in what has been called maldetection overdiagnosis[16] and overdetection overdiagnosis.[17] This has become more prevalent with screening, and will become even more ubiquitous with Big Data, artificial intelligence (AI), and predictive and personalized medicine.[18-22] By combining and analyzing a wide range of data, such approaches will be able to identify a vast variety of indicators. However, validation of how

the indicators relate to (and could avoid) people's sufferings are challenging.[23-25] Otherwise, indicators will come to define disease more than being validated by them.

3. Coupling disease name to uncertain phenomena

Decoupling testing from harm and lack of validation are only part of the problem. For sure, both introduce uncertainty. But uncertainty in diagnostics is not new. We have struggled with non-perfect diagnostic tests for ages, e.g., in terms of false (negative and positive) test results. However, when indolent lesions ("'ductal intraepithelial neoplasia" or "indolent lesions of epithelial origin") are labelled ductal *carcinoma* in situ or "pre-invasive breast cancer cells,"[26, 27] we produce connections that may not be warranted. In making the indicators *define* disease we decouple testing from harm and generate a new category: *the correctly identified person who would not be harmed*, i.e., the overdiagnosed. While high blood pressure can be an indicator and a risk factor of disease, it has become a disease of its own (I10 in ICD-10). Focusing on indicators may increase their importance for knowledge of disease beyond what is warranted. **Figure 1** explains how tests may be decoupled from harm.

[Insert **Figure 1** here]

It is important to notice that these three errors are not present in every case of overdiagnosis and are not three sequential steps in every case of overdiagnosis. They all merit careful considerations on their own right.

The slippery path from new tests to overdiagnosis

In the decoupling of test and harm we are subject to two leaps of faith. First we leap from test results (indicator) to disease. In situ neoplasms or "papillaryurothelial neoplasia of low malignant potential" come to designate cancers. While it can be perfectly warranted to pay attention to indicators and risks factors for disease, it is not warranted to handle risk factors as disease (per se).[12, 28] The second leap of faith is that we believe that early detection is better than later detection [14]. In this

we presuppose a linear or positive progression, which often is warranted, but not always.[29, 30] Early or small is not always the same as good [31].

Moreover, actionability overshadows uncertainty. When we are able to act or intervene, the indicators increase in importance. As we are able to remove polyps or DCIS, they become significant. No doubt, if we remove necessary conditions or risk factors for a disease, we may remove the disease. But at the same time we may come to confuse necessary conditions with the disease or with its causes. Although blood is a necessary condition for sepsis, we do not remove people's blood to avoid sepsis. Accordingly, not all indicators are important, and some indicators, while important, are not helpful in reducing harm.

What is the solution?

In the identified errors and leaps of faith lie the solutions. First, we must stop identifying and testing for conditions that are only remotely associated with harm. Second, we need more stringent verification of tests and, third, we must avoid naming indicators and indicative conditions after diseases.

Distinguish between uncertainties

The decoupling between tests and harm conflates two types of uncertainty. There exists one kind of uncertainty between diagnostic test results (for a given disease) and harm ("Verifying 1" in Figure 1b, *Disease test uncertainty*). This is the traditional uncertainty which comes with any diagnostic test due to lack of accuracy. It is expressed by sensitivity and specificity and measured in false (positive and negative) test results. The other type of uncertainty is about whether an indicator, such as DCIS will actually result in disease ("Verifying 2" in Figure 1b, *indicative test uncertainty*). It is this latter type of uncertainty that introduces overdiagnosis, i.e., identified indicators that will not result in harm.

Accordingly, there are *three types of uncertainty* related to tests of indicators and harm to persons. First, there is uncertainty with respect to what is found: Is the test result *true*, e.g., is the test result

of fecal occult blood correct? This is a question of *accuracy*. Second, there is uncertainty with respect to *whether it matters*? When we know that we correctly have identified the indicator, such as PSA, DCIS, or fecal occult blood (FOB), it is uncertain whether it matters to the person. This is a question of where to set the *threshold*. Third, there is uncertainty with respect to *how it matters*. A mammogram may be very good at detecting DCIS, but DCIS may not lead to any experienced harm from breast cancer. The various forms of uncertainty are illustrated in Figure 2.

[Insert **Figure 2** here]

Change the game of the name

As the test results may be decoupled from harm, the disease name may give them unwarranted importance. As illustrated in Figure 1, there is little or no validation of the relationship between the indicator test result and harm. Nonetheless, naming the indicator after the disease it indicates unwarrantedly couples it to harm. The term "carcinoma" in DCIS (*ductal carcinoma in situ*) indicates that there is a closer connection to (breast) cancer than what may be warranted [29]. It also alters conceptions and management of these conditions [27]. Hence, giving indicators disease-like names advances the impression that we know more than we actually do and contributes to the expansion of disease and overdiagnosis. The case of DCIS is particularly interesting because it was originally thought to be a variant of breast cancer, but (partly due to screening) showed more appropriately to be considered a precursor of the disease. Therefore, we should continue the encouraging improvements in the field, such as calling DCIS "indolent lesions of epithelial origin" (IDLE conditions),[32] using "urothelial neoplasia of low malignant potential" instead of grade 1 papilloma, and "papillaryurothelial neoplasia of low malignant potential" instead of papilloma and grade1 carcinoma.[29, 33]

Finding indicators and treating indicative conditions may clearly be warranted, but this can be done without calling them diseases. Instead of "colorectal cancer screening program" we could call it "polyp identification and removal program." Thus, although the decoupling of tests and harm has

some injurious implications, we can turn the trend and apply our diagnostic innovations to help people, and not to harm them.

Verify relationship between indicators and harm

Another way to bar the expansion of disease and overdiagnosis is to design studies to verify the relationship between indicators and harm. This would actually eliminate overdiagnosis, turning today's overdiagnosis into false positive test results. When the indicator test (and any following test) is positive but the person is not harmed, we have a false positive test. Hence, we could estimate the "true sensitivity" [34]. That is, the ultimate measure our tests, i.e., "the diamond standard,"[6] should be harm, and not some intermediate surrogate measure. No doubt, surrogate measures are useful on the way towards true diagnosis and treatment, and sometimes is the best we can get. However, they should not be confused with the ultimate measure of disease, which is harm.

When stopping to tag indicators with names of harm or disease we would undermine the unwarranted eager to find and eliminate indicators that do not develop and become harmful. That is, we would stop the uncontrolled expansion of disease. We therefore must stop talking about "unharmful disease"[35], "indolent disease, " "inconsequential disease"[11], or "low-risk disease."[30] Disease is harmful (in some sense). Conditions used as indicators may be indolent and become harmful, but they are not harmful *per se*. Clinicians need to be aware of this, and patients should be informed about it. Researchers must strive to verify the connection between indicators and harm, and health policy makers must avoid making "the war on cancer" to a "war on indicators." [9]

No doubt, identifying and targeting indicators and risk factors can be of outmost import for people's health. However, indicators and risk factors are not diseases.[12, 28] Both have uncertain connections to harm. While both risk factors and indicators can be principal to avoid or ameliorate harm, we do not whether they will do so – or whether they will result in subsequent unnecessary diagnostics and treatment, anxiety, or harm. This urges us to be cautious.

Conclusion

New tests may come to expand disease, produce patients, and cause unnecessary harm instead of early detection, avoided or reduced mortality, or mortality reduction. In order to avoid this we should stop identifying and testing for conditions that are only remotely associated with harm; we ought to verify tests more stringently; and we should avoid naming indicators and indicative conditions after diseases. In short, we should bring the naming of disease on par with the verification of harm. Instead of finding trouble in looking for dubious indicators, we should look for harm to avoid trouble. If not, we will end like the ancient tragic hero Oedipus, succumbing because of his very best abilities in the search for knowledge. Oedipus ended blinding himself because he realized that search for knowledge and his vision had hampered his wisdom.

Contributions

I am the sole author of this article, the content stems from me, and I have the sole responsibility for the content of the article. I owe the comparison to Oedipus and the Greek heroes to Jan Helge Solbakk.

Declaration of interests

I certify that there is no actual or potential conflict of interest in relation to this manuscript, and there are no financial arrangements or arrangements with respect to the content of this comment with any companies or organizations.

Acknowledgement

I am most thankful for wise comments and good suggestions from the anonymous reveiwers.

References

1. Siontis, K.C., et al., *Diagnostic tests often fail to lead to changes in patient outcomes.* J Clin Epidemiol, 2014. **67**(6): p. 612-21.
2. Welch, H.G., L. Schwartz, and S. Woloshin, *Overdiagnosed : making people sick in the pursuit of health*. 2011, Boston, Mass.: Beacon Press. xvii, 228 p.
3. Brownlee, S., et al., *Evidence for overuse of medical services around the world.* The Lancet, 2017.

4. Morgan, D.J., et al., *2017 update on medical overuse: A systematic review.* JAMA Internal Medicine, 2017.
5. Moynihan, R.N., et al., *Expanding disease definitions in guidelines and expert panel ties to industry: a cross-sectional study of common conditions in the United States.* PLoS Med, 2013. **10**(8): p. e1001500.
6. Hofmann, B., *The overdiagnosis of what? On the relationship between the concepts of overdiagnosis, disease, and diagnosis.* Med Health Care Philos, 2017. **20**(4): p. 453-464.
7. Welch, H.G., *Less Medicine, More Health: 7 Assumptions That Drive Too Much Medical Care*. 2015, Boston, MA: Beacon Press.
8. Hofmann, B., *The overdiagnosis of what? Barring the expansive conception of disease*, in *4th international conference on Preventing Overdiagnosis*. 2016: Barcelona.
9. Hofmann, B. and H.G. Welch, *New diagnostic tests: more harm than good.* BMJ, 2017. **358**.
10. Cohen, J.D., et al., *Detection and localization of surgically resectable cancers with a multi-analyte blood test.* Science, 2018: p. eaar3247.
11. Hutchinson, B.D., et al., *Overdiagnosis of Pulmonary Embolism by Pulmonary CT Angiography.* AJR Am J Roentgenol, 2015. **205**(2): p. 271-7.
12. Schwartz, P.H., *Risk and disease.* Perspectives in biology and medicine, 2008. **51**(3): p. 320-334.
13. Schwartz, P.H., *Progress in Defining Disease: Improved Approaches and Increased Impact.* The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 2017. **42**(4): p. 485-502.
14. Hofmann, B. and J.A. Skolbekken, *Surge in publications on early detection.* Bmj, 2017. **357**: p. j2102.
15. Zwaan, L. and H. Singh, *The challenges in defining and measuring diagnostic error.* Diagnosis (Berl), 2015. **2**(2): p. 97-103.
16. Rogers, W.A. and Y. Mintzker, *Getting clearer on overdiagnosis.* J Eval Clin Pract, 2016. **22**(4): p. 580-7.
17. Brodersen, J., et al., *Overdiagnosis: what it is and what it isn't.* Evid Based Med, 2018. **23**(1): p. 1-3.
18. Flores, M., et al., *P4 medicine: how systems medicine will transform the healthcare sector and society.* Per Med, 2013. **10**(6): p. 565-576.
19. Hood, L., J.C. Lovejoy, and N.D. Price, *Integrating big data and actionable health coaching to optimize wellness.* BMC Med, 2015. **13**(1): p. 4.
20. Topol, E., *The Creative Destruction of Medicine: How the Digital Revolution Will Create Better Health Care*. 2012, New York: Basic Books.
21. Torkamani, A., et al., *High-Definition Medicine.* Cell, 2017. **170**(5): p. 828-843.
22. Saracci, R., *Epidemiology in wonderland: Big Data and precision medicine.* European Journal of Epidemiology, 2018. **33**(3): p. 245-257.
23. Lipworth, W., et al., *Ethics and Epistemology in Big Data Research.* Journal of Bioethical Inquiry, 2017: p. 1-12.
24. Frické, M., *Big data and its epistemology.* Journal of the Association for Information Science and Technology, 2015. **66**(4): p. 651-661.
25. Ioannidis, J.P. and M.J. Khoury, *Improving validation practices in "omics" research.* Science, 2011. **334**(6060): p. 1230-1232.
26. Moynihan, R., et al., *What do you think overdiagnosis means? A qualitative analysis of responses from a national community survey of Australians.* BMJ Open, 2015. **5**(5): p. e007436.
27. Nickel, B., et al., *Words do matter: a systematic review on how different terminology for the same condition influences management preferences.* BMJ open, 2017. **7**(7): p. e014129.
28. Schwartz, P.H., *Small Tumors as Risk Factors not Disease.* Philosophy of Science, 2014. **81**(5): p. 986-998.

29. Esserman, L.J., I.M. Thompson, Jr., and B. Reid, *Overdiagnosis and overtreatment in cancer: an opportunity for improvement.* JAMA, 2013. **310**(8): p. 797-8.
30. Esserman, L. and I. Thompson, *Solving the Overdiagnosis Dilemma.* JNCI: Journal of the National Cancer Institute, 2010. **102**(9): p. 582-583.
31. Lannin, D.R. and S. Wang, *Are Small Breast Cancers Good because They Are Small or Small because They Are Good?* N Engl J Med, 2017. **376**(23): p. 2286-91.
32. Esserman, L., Y. Shieh, and I. Thompson, *Rethinking screening for breast cancer and prostate cancer.* JAMA, 2009. **302**(15): p. 1685-1692.
33. Epstein, J.I., et al., *The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder.* The American journal of surgical pathology, 1998. **22**(12): p. 1435-1448.
34. Prorok, P.C., B.S. Kramer, and A.B. Miller, *Study designs for determining and comparing sensitivities of disease screening tests.* J Med Screen, 2015. **22**(4): p. 213-20.
35. Rogers, W.A. and Y. Mintzker, *Getting clearer on overdiagnosis.* J Eval Clin Pract, 2016.

Tables

Table 1 Examples of indicators with corresponding diseases and harms. Harm of subsequent testing **and treatment is not included.**

Indicators	**Corresponding Disease**	**Harms (of the disease)**
Human Papilloma Virus (HPV) Biomarkers	Cervix cancer	Loss of appetite and weight, fatigue, pelvic pain, back pain, leg pain, swollen legs, heavy vaginal bleeding, bone fractures, leakage of urine or feces from the vagina, ultimately death.
Fecal occult blood, adenomatous polyps, adenoma	Colorectal cancer	Pain, feeling tired, reduced function, ultimately death
Microcalcification, Ductal Carcinoma in Situ (DCiS)	Breast cancer	Lump in breast, skin changes, burning, pain, swelling, spreading (metastases) to liver, lung, brain, ultimately death.
Plaque and Tangles Blood based biomarkers CSF based biomarkers	Alzheimer Type Dementia (ATD)	Loss of attentiveness, flexibility, abstract thinking, planning ability, and memory. Loss of speech, delusion, wandering, irritability, urinary incontinence, apathy, exhaustion, loss of muscle mass, (secondary) infections, and death.
Body Mass Index (BMI)	"Obesity"	Harms related to diseases, such as diabetes, cardiovascular disease, and renal failure. These diseases can ultimately lead to death.

Figure legends

Figure 1 Relationship between harm, disease, and test, where a) illustrates how symptoms and signs come to define disease due to validated connections with harm. b) indicates what happens when additional tests are introduced and validated by "gold standard" tests. c) explains how tests are named by the disease and its' expected harm without proper validation, which may ascribe unwarranted predictive power to the test.

Figure 2 Three types of uncertainties on the relationship between a test result of an indicator and harm to persons.