1. Author: Please check and confirm that the authors and their respective affiliations have been correctly identified and amend if necessary.
   This is correct.

2. Author: Please confirm the section headings are correctly identified.
   This looks fine.

3. Author: The following references Boorse (1976, 1977) are not provided in the list. Please check and provide.
   Thank you for discovering this!


I tried to insert them in the list, but did not succeed. Could you help me inserting them properly?
The overdiagnosis of what? On the relationship between the concepts of overdiagnosis, disease, and diagnosis

Bjørn Hofmann,1✉2
Phone +4761135311 Email bjoern.hofmann@ntnu.no Email b.m.hofmann@medisin.uio.no
1 Institute for the Health Sciences, The Norwegian University Science and Technology, PO Box 1, 2802 Gjøvik, Norway
2 Centre for Medical Ethics, The University of Oslo, Oslo, Norway

Abstract

Overdiagnosis and disease are related concepts. Widened conceptions of disease increase overdiagnosis and vice versa. This is partly because there is a close and complex relationship between disease and overdiagnosis. In order to address the problems with overdiagnosis, we may benefit from a closer understanding this relationship. Accordingly, the objective of this article is to elucidate the relationship between disease and overdiagnosis. To do so, the article starts with scrutinizing how overdiagnosis can explain the expansion of the concept of disease. Then it investigates how definitions of disease address various challenges of overdiagnosis. The article specifically investigates recent attempts to clarify the relationship between the concepts of disease and overdiagnosis. Several shortcomings are identified and lead to a closer analysis of overdiagnosis in the diagnostic process. Contrary to recent contributions to the field, it is argued that cases of overdiagnosis are not cases of disease. They are non-verified labelling of disease. It is revealed how overdiagnosis establishes an unwarranted link between indicative phenomena, such as polyps or cell changes, and harm, and thereby generates a link to disease. One implication of this study is that we should stop attributing disease language to indicative phenomena. That is, we should stop calling it “cancer screening” when we are actually searching for polyps. Another implication is that we should strive for scientific progress in differentiating phenomena that are of negative value to us from those that are not. In overdiagnosis we diagnose something that is not disease; it is over-diagnosis.

Keywords

Overdiagnosis
Disease
Diagnosis
Definition

Introduction

The concept of overdiagnosis has gained increased attention in the professional literature, in the public, as well as in the philosophy of medicine (Blumenthal-Barby 2013; Carter et al. 2016a, 2015; Hofmann 2016; Morrison 2016; Rogers 2015; Rogers and Mintzker 2016a, b). Many definitions exist (Carter et al. 2016a) resulting in conceptual and practical challenges (Carter et al. 2015; Hofmann 2014). The crucial connection between overdiagnosis and disease has been acknowledged (Moynihan et al. 2013, 2014; Esserman et al. 2013, 2014; Black 1998) and recent work has tried to clarify the concept of overdiagnosis and its relationship to the concept of disease (Carter et al. 2015, 2016a; Rogers 2015). Despite these promising attempts, the relationship between overdiagnosis and disease is still unsettled. One of the reasons for this may be that current conceptual approaches do not provide sufficient means to analyse and limit overdiagnosis or the expansion of disease. This may be because we have not dug deep enough in the concept of diagnosis, and in particular its relationship with the crucial and value-laden concept of harm.

Accordingly, the objective of this article is to clarify the relationship between the concepts of overdiagnosis and disease, and in particular how they tend to expand. To do so the article will return to the a common roots of both the concepts of overdiagnosis and disease, i.e., the process of diagnosis. Understanding how diagnostics expands disease and generates overdiagnosis is crucial in order to hinder the unwarranted expansion of disease and to handle the challenges with overdiagnosis. Hence, the article will focus on what is diagnosed in overdiagnosis.

The article will investigate standard conceptions of disease and recent contributions in the field in order to address three core questions:

1. How do conceptions of overdiagnosis explain the expansion of the concept of disease?
2. How do definitions of disease address overdiagnosis?
3. How could better understanding of the process of diagnostics explain the extension of overdiagnosis and the expansion of disease?

In order to motivate this endeavour, allow me briefly to substantiate how overdiagnosis expands the conception of disease and clarify why it is relevant to scrutinize the concept of disease when addressing the issue of overdiagnosis.

How does overdiagnosis expand the conception of disease?

Overdiagnosis expands the disease concept in several ways. On the individual level it gives persons a disease label due to some diagnostic test; persons who would otherwise never have experienced any symptoms referable to the disease (Black 2000; Welch et al. 2011). On the population level, it increases the incidence of disease. On the conceptual level, the definition of the disease is expanded to include (ranges) of phenomena (e.g., indicators and preconditions) that were previously not encompassed. On the practical level, disease practices are expanded to include testing and diagnostic activities as well as extended follow up practices.

Why is it relevant to scrutinize the concept of disease when addressing overdiagnosis?

Overdiagnosis is said to expand the conception of disease beyond what is helpful. Accordingly, one strategy to reduce overdiagnosis would be to clarify the concept of disease in order to identify and eliminate cases of overdiagnosis. Another reason why the concept of disease is important for overdiagnosis semantically means too much diagnosis, and that diagnosis is closely related to the concept of disease. Disease is to a large extent what is overdiagnosed. Therefore, it is imperative to investigate the concepts of overdiagnosis and disease, as well as their relationship when addressing the challenges of overdiagnosis.

How do conceptions of overdiagnosis explain the expansion of the concept of disease?

In a recent article Rogers and Mintzker (2016a, b) propose a taxonomy of overdiagnosis to clarify the relationship between disease and overdiagnosis. They distinguish between “maldetection” and “misclassification” overdiagnosis, where maldetection overdiagnosis occurs where one is unable to discriminate between lesions or abnormalities that appear to be identical, but where only some of them will lead to harm. Screening-detected cancer would be a typical example of maldetection overdiagnoses. Misclassification overdiagnoses arise when one lowers the diagnostic thresholds for diseases that are defined in terms of physiological parameters that vary on a spectrum, so that many persons with a positive diagnosis do not experience symptomatic disease (Rogers and Mintzker 2016b). Lowering the threshold for high blood pressure would be one example of misclassification overdiagnosis. Table 1 gives an overview of Rogers and Mintzker’s conception of overdiagnosis.

Table 1

<table>
<thead>
<tr>
<th>Misclassification overdiagnosis</th>
<th>Maldetection overdiagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic problem</td>
<td>Setting the threshold</td>
</tr>
<tr>
<td>Type of problem</td>
<td>Ontological</td>
</tr>
<tr>
<td>Aim</td>
<td>Avoid setting over-inclusive thresholds</td>
</tr>
<tr>
<td>Definition</td>
<td>&quot;misclassification overdiagnosis occurs when diagnostic thresholds are set at levels such that many people have a positive diagnosis for the disease despite the absence of current or future symptoms and harm&quot;</td>
</tr>
<tr>
<td>Occurs in</td>
<td>Spectrum diseases</td>
</tr>
<tr>
<td>Explaining expansion of disease</td>
<td>Disease is expanded by lowering thresholds</td>
</tr>
</tbody>
</table>

Based on (Rogers and Mintzker 2016b) and (Hofmann 2016a)
How do definitions of disease address overdiagnosis?

10.1007/s11019-017-9776-z

How do definitions of disease address overdiagnosis?

Moreover, Rogers and Mintzker view the “pathognomonic and essential feature of overdiagnosis” to be “the identification of harmless disease” (Rogers and Mintzker 2016a). As I will elaborate on below, “harmless disease” is a contradiction in terms as harm is a key feature of disease. It is this harmful aspect of disease which justifies health care interventions. In another and more recent article Rogers and Walker define overdiagnosis as “the diagnosis of a condition that is not causing harm, and will not come to cause harm” (Walker and Rogers 2016, Epub ahead of print). Again, the problem with this account is that it is unclear what counts as disease and diagnosed conditions if such conditions are not related to harm, as ordinary conceptions of disease and diagnosis implicitly tend to be related to conceptions of harm. Excluding harm in this definition (“harmless”) either contradicts the value laden concepts of disease and diagnosis or it presupposes what is to be shown, i.e., to identify which detected conditions that are not harmful.¹

If something is diagnosed, it is more than a detected condition; it is a condition that is differentiated from something else because we have some knowledge of its characteristics (diagnosis). What tends to be common to all cases of diagnosis is that they are considered to be harmful in one way or another. For example, we can detect a range of changes in muscle fibres in a person without relating it to disease. Our attention to changes in fast-glycolytic fibres (and also to the gene for the protein dystrophin) is because we learned that people who have experienced harm appeared to have such fibres (and such mutations e.g. in Duchenne muscular dystrophy). Accordingly, a person who has a diagnosis is diseased (Svenaeus 2014). Hence, relating overdiagnosis to “harmless disease” and “harmless diagnosed condition” is a contradiction in terms as long as disease and diagnosis is related to harm in some way. I will return to this important issue below.

In another recent and important article Carter et al. aim at clarifying the concept of overdiagnosis in order better to address its ethical challenges (Carter et al. 2016a). In their definition overdiagnosis is a condition that is labelled with a diagnosis in a given population carrying an unfavourable balance between benefits and harms. They distinguish between Predatory overdiagnosis, Misdirected overdiagnosis, and Tragic overdiagnosis, where the first is where actors (doctors, industry) use diagnostics to gain themselves, the second “occurs when harm is avoidable, but the goal of the actor or actors is primarily to benefit others,” and the last is when harms are unavoidable, e.g., due to lack of knowledge.

In their theory and definition of overdiagnosis Carter et al. dismiss a “strong objectivist model of disease” and promote a “softening model of disease” (Carter et al. 2016a). They are quite vague on the concepts of diagnosis and disease. Instead they focus on the social process of (over) diagnosis, and found their conception of overdiagnosis is founded on professional community standards of practice and on a population-based conception of benefit (Carter et al. 2016a). This account apparently does a nice job in explaining the expansion of disease in terms of the widening of professional diagnostic standards at the basis of overdiagnosis. However, it comes at conceptual costs, reducing its explanatory power (Rogers and Mintzker 2016a; Morrison 2016; Hofmann 2016b). As pointed out by Rogers and Mintzker, the clause “the overall balance of benefits and harms of both diagnosis and any subsequent intervention” in the definition of overdiagnosis includes too much, as it may include any identified health-related problem where subsequent interventions fail to benefit instances of overdiagnosis (Rogers and Mintzker 2016a). In particular, it includes cases where “a woman is diagnosed with early breast cancer via mammography, receives treatment, but dies from a heart attack” (Carter et al. 2016b). This makes any person with manifest lethal disease overdiagnosed if the person dies in a car accident and not as a result of the disease. Although including competing mortality in the conception of overdiagnosis has some traction in epidemiology (Marcus et al. 2015) it still appears counter-intuitive. Nobody would blame health care for not taking into account that diagnosed persons may die in car accidents. Including competing mortality in the concept of overdiagnosis appears to confuse unnecessary diagnostics with unhelpful diagnostics. If the aim of clarifying and defining overdiagnosis is to improve health care, the target should be unnecessary diagnostics. No doubt we should avoid car accidents, but that may not be the task of health care. Hence, Carter et al. appear to have a vague (or “soft”) concept of disease based on professional standards of diagnostics, which tend to expand overdiagnosis. Moreover, overdiagnosis is expanded by including competing mortality.

Common to the approaches of Rogers and Mintzker and Carter et al. is that overdiagnosis is based on an expansive concept of disease (or diagnosis). Hence, expansion of disease to a large extent explains overdiagnosis (and its expansion). Thus, if the goal is to limit overdiagnosis (as defined as something undesirable), one potentially fruitful strategy would be to limit the expansion of the concept of disease. We therefore may investigate the latter. I will later argue that we need to return to the concept of diagnosis, but so far the most recent and promising search for an understanding and limitation of overdiagnosis (seen as something unfavourable or undesirable) has lead us to the concept of disease.

How do definitions of disease address overdiagnosis?
Let me start with some traditional and general conceptions of disease before I scrutinize some more recent approaches specific to overdiagnosis.

Traditional conceptions of disease and overdiagnosis

It has become commonplace to distinguish between naturalistic, normativistic, and hybrid/mixed/intermediate conceptions of disease (Ereshefsky 2009; Kingma 2014; Kovacs 1998). The first sees disease as given by nature, the second as given by culture, and the last as a combination. They offer quite different framings for the conception of overdiagnosis. Let me briefly investigate what they have to offer.

The most prominent naturalistic conception of disease is the biostatistical theory of disease (BST) according to which disease is defined as an internal state that reduces functioning to a level below what is statistically typical for the species and which does not contribute to biological goals, such as survival and reproduction (Boorse 1977, 555). Disease does not have to be harmful (Boorse 1976), but it must obstruct biological goals. If a certain internal state of the organism interferes with the performance of some natural function, it is disease. Anything beyond this may qualify as overdiagnosis. However, as BST’s definition of disease is based on a statistical conception of normality, it is not able to point out whether a single instance of diagnosis is an overdiagnosis. Moreover, BST has a “threshold problem” (Schwartz 2007) as it is not clear from nature where to set the limits to normal functioning. It may therefore be argued that BST does not provide sufficient resources to clarify overdiagnosis (Wakefield 2014; Walker and Rogers 2016, Epub ahead of print). As it is hard to read out of nature which forms of reduced functioning may be biologically disadvantageous, it becomes hard to define and delimit overdiagnosis.

Normativistic conceptions of disease see disease as established by social and cultural norms and values. These norms and values may vary among different groups (pluralism) and they may change over time. While they may be well justified within a given context, it can be difficult to find common ground for an overarching conception of disease. Accordingly, while such conceptions of disease can provide excellent guidance for defining and limiting overdiagnosis in specific contexts, it can be difficult to agree on overarching conceptions of disease that will do the same trick across such contexts. Moreover, it may be difficult to delimit contexts, and thus various conceptions of overdiagnosis.

Various hybrid conceptions may appear more promising. Expanding on the BST, Jerome Wakefield defines disease as harmful dysfunction (Wakefield 2014). According to the harmful dysfunction theory (HDT) of disease, a single instance of overdiagnosis is not disease, as it is not harmful. One may correctly have measured a(n indicative) condition or even a physiological or mental dysfunction, but when it is of no harm to the person it is not disease (Wakefield 2014). Conditions that are not harmful are not diseases. Asymptomatic carriers and neutral mutations are therefore not disease, and “harmless disease” is a contradiction in terms. The same goes for overdiagnosis on the populational level. Hence, the HDT appears to address overdiagnosis in providing firm limits to what counts as disease. Conditions that do not count as harmful dysfunctions are not disease, and hence are (potentially) overdiagnosis. As HDT rules out “harmless disease” it also rules out overdiagnosis. For example HDT dismisses both maldetection and misclassification overdiagnosis (Wakefield and Rogers 2016, Epub ahead of print).

Also expanding on BST, the risk theory of disease (RT) defines disease in terms of modifiable risk factors (Schwartz 2008). Hypertension and high cholesterol are considered to be diseases because they are modifiable risk factors for symptoms and death (of cardiovascular diseases). This conception of disease does not contribute to setting limits to overdiagnosis, as every condition that may represent a risk for symptoms and disease will be a disease, and not overdiagnosis. However, Schwartz wants to delimit “risk-based diseases” from diseases proper, because otherwise it would “warp the notion of disease” and because it has “the potential of undermining patient understanding” (Schwartz 2008) (p. 332). Hence, while RT does not contribute to clarifying overdiagnosis as it includes all risk factors (including indicative phenomena such as hypertension), the distinction between disease and modifiable risk-factors provided by Schwartz does, at least theoretically clarify the relationship between overdiagnosis and disease. The expansion of risk-based disease expands overdiagnosis. In practice however, the distinction may not be so big, and as Schwartz points out, risk-factors “should be seen as just as central to health care as treatments for disease” (Schwartz 2008). If you avoid or diminish certain risk factors, you can help prevent the disease.

The above conceptions of disease have been called “traditional” (Sadegh-Zadeh 2000, 2008) or “classical” (Walker and Rogers 2016, Epub ahead of print) as they are in some way defined in terms of set of conditions that are individually necessary and jointly sufficient for the application of the disease or in other ways have clear boundaries. There are several alternatives to such “classical” conceptions of disease. Let me present some such concepts which are especially elaborated to solve the challenges raised by overdiagnosis.

Alternative concepts of disease to explain overdiagnosis

As already introduced, Carter et al. promote a “softened” model of disease and base their conception of overdiagnosis on a professional community standard of diagnosis (Carter et al. 2016a). They disdain an objectivist approach to disease and define correct diagnoses as “those that are accepted as correct within a relevant professional community.” Accordingly, overdiagnosis is a population-based phenomenon which is defined in terms of conditions that are “labelled with diagnosis in the population” where “this identification and labelling would be accepted as correct
Although Carter et al. provide some means to explain overdiagnosis and explicitly discuss and allude to various conceptions of disease, they provide little guidance to determine diagnostic boundaries and to limit overdiagnosis (Rogers and Mintzker 2016a). Furthermore, the approach ignores that harm enters the definition of overdiagnosis both in the inherent concept of disease (or diagnosis), \(^1\) i.e., in the professional labelling with diagnosis, and in the assessment of benefits and harms (Hofmann 2016b). As for all normativistic, “softened,” or non-classical conceptions of disease, they are value-laden, and harm frequently is inherent (as definiens). \(^4\) I will return to this point below. For now it suffices to acknowledge that Carter et al. provide a definition of overdiagnosis which contains a non-classical concept of disease that to some extent may explain, but not prevent or limit, overdiagnosis. Moreover, harm (and benefit) enters the definition of overdiagnosis twice, both in terms of diagnosis (or disease) and in terms of net benefit. Although Carter et al. use the point that diagnoses result from social and political processes to defend that the fact that they enter harm twice (Carter et al. 2016b), the relationship between these conceptions of harm is unclear, and it is unclear what adding a second type or level of harm to the definition adds to the argument. The point here is that some notion of harm is at the basis of Carter et al. conception of overdiagnosis and that it is unclear how this relates to notions of harm inherent in their (vague) concept of disease.

In their recent and very interesting contribution Walker and Rogers explicitly suggest a conception of disease that intends to address the challenges with overdiagnosis (Walker and Rogers 2016, Epub ahead of print). In particular, they provide “a précising definition of disease designed to provide practical help in preventing and limiting overdiagnosis.” In doing so they “define disease as physiological dysfunction that has a significant risk of causing severe harm to the patient.” Accordingly, ‘significant risk’ and ‘severe harm’ requires specification in particular cases, depending on the best current information available on the condition, societal views, and individual patient preferences. In particular, “what will count as a significant risk, or a severe harm must be contextualised to the particular patient’s other health characteristics, and with reference to their values” (Walker and Rogers 2016, Epub ahead of print).

Although Walker and Rogers underline that the “précising definition of disease” (disease\(_{ODX}\)) intended to “provide practical help in preventing and limiting overdiagnosis” is a “classical concept with determinate boundaries” they fail to provide convincing arguments that it is different from “disease (simpliciter),” which they acknowledge to be a vague non-classical cluster concept (Walker and Rogers 2016, Epub ahead of print). Disease\(_{ODX}\) is “is for use only in borderline cases that are potential cases of overdiagnosis (such as diagnoses of early-stage, non-invasive cancers that are known to be commonly overdiagnosed, or diagnoses of spectral diseases that are close to the diagnostic thresholds and where patients have no other risk factors).” However, the “précising definition of disease” which is to “provide practical help in preventing and limiting overdiagnosis” contains the concept of overdiagnosis. This makes the “précising” definition circular. Providing a definition of disease that points out cases of overdiagnosis that are “known to be commonly overdiagnosed” does not add significantly to the conceptual clarity or to the practical guidance of how to handle or limit overdiagnosis. Moreover, it leaves us with the task of deciding what count as “borderline cases,” and thus which are potential cases of overdiagnosis. In other words, it brings us back to square one.

Additionally, when elaborating on key terms, such as “severe harm” and “significant risk,” the authors recur to vague terms. As they acknowledge, “the definition retains some vagueness” and seems “counterintuitive and with potential for significant relativism” (Walker and Rogers 2016, Epub ahead of print). Hence, it is far from obvious that it is a “classical concept with determinate boundaries.” Moreover, it is difficult to see how this adds much to the RT and to Schwartz’s distinction between disease proper and “risk-based diseases” and as described, disease\(_{ODX}\) appears as vague and non-classical as disease (simpliciter).

Hence, singling out a specific type of disease prone to overdiagnosis by a “précising definition” of disease (disease\(_{ODX}\)) does not “provide practical help in preventing and limiting overdiagnosis” as it depends on a diversity of individual and social values. It gives little guidance on how to assess and bar overdiagnosis\(^5\) as it lacks a clear concept of disease. Most importantly however, the concept of “dysfunction” in the definition of disease\(_{ODX}\) is quite unclear. As the authors explicitly claim: “the causal chains leading from a currently identified dysfunction to future harms are open.” At the same time the concept of “dysfunction” is declared to rule out conditions like pregnancy (Walker and Rogers 2016, Epub ahead of print). Moreover, the “dys” part of dysfunction appears to include some negative evaluation, and it is unclear how this inherent negative evaluation is related to the explicit harm of the definition, i.e., which has “a significant risk of causing severe harm to the patient.”

This problem comes out clearly when the authors claim that any instance of some detectable phenomenon that may result in disease, is diseased: “Someone with colonic polyp with malignant changes but no local or distant invasion, and no experienced harm, nonetheless does have a disease” (Walker and Rogers 2016, Epub ahead of print). However, here the authors presuppose what is to be shown, i.e., whether a given phenomenon is disease. Polyps are frequent in (the aging) population, and most polyps do not develop to colorectal cancer (disease).\(^6\) While I agree with Walker and Rogers that the definition of overdiagnosis in detection of polyps is hard, these cases appear to be non-false.

\(^1\) In their recent and very interesting contribution Walker and Rogers explicitly suggest a conception of disease that intends to address the challenges with overdiagnosis (Walker and Rogers 2016, Epub ahead of print). In particular, they provide “a précising definition of disease designed to provide practical help in preventing and limiting overdiagnosis.” In doing so they “define disease as physiological dysfunction that has a significant risk of causing severe harm to the patient.” Accordingly, ‘significant risk’ and ‘severe harm’ requires specification in particular cases, depending on the best current information available on the condition, societal views, and individual patient preferences. In particular, “what will count as a significant risk, or a severe harm must be contextualised to the particular patient’s other health characteristics, and with reference to their values” (Walker and Rogers 2016, Epub ahead of print).

\(^2\) Although Walker et al. provide some means to explain overdiagnosis and explicitly discuss and allude to various conceptions of disease, they provide little guidance to determine diagnostic boundaries and to limit overdiagnosis (Rogers and Mintzker 2016a). Furthermore, the approach ignores that harm enters the definition of overdiagnosis both in the inherent concept of disease (or diagnosis), i.e., in the professional labelling with diagnosis, and in the assessment of benefits and harms (Hofmann 2016b). As for all normativistic, “softened,” or non-classical conceptions of disease, they are value-laden, and harm frequently is inherent (as definiens). I will return to this point below. For now it suffices to acknowledge that Carter et al. provide a definition of overdiagnosis witch contains a non-classical concept of disease that to some extent may explain, but not prevent or limit, overdiagnosis. Moreover, harm (and benefit) enters the definition of overdiagnosis twice, both in terms of diagnosis (or disease) and in terms of net benefit. Although Carter et al. use the point that diagnoses result from social and political processes to defend that the fact that they enter harm twice (Carter et al. 2016b), the relationship between these conceptions of harm is unclear, and it is unclear what adding a second type or level of harm to the definition adds to the argument. The point here is that some notion of harm is at the basis of Carter et al. conception of overdiagnosis and that it is unclear how this relates to notions of harm inherent in their (vague) concept of disease.

In their recent and very interesting contribution Walker and Rogers explicitly suggest a conception of disease that intends to address the challenges with overdiagnosis (Walker and Rogers 2016, Epub ahead of print). In particular, they provide “a précising definition of disease designed to provide practical help in preventing and limiting overdiagnosis.” In doing so they “define disease as physiological dysfunction that has a significant risk of causing severe harm to the patient.” Accordingly, ‘significant risk’ and ‘severe harm’ requires specification in particular cases, depending on the best current information available on the condition, societal views, and individual patient preferences. In particular, “what will count as a significant risk, or a severe harm must be contextualised to the particular patient’s other health characteristics, and with reference to their values” (Walker and Rogers 2016, Epub ahead of print).

Although Walker and Rogers underline that the “précising definition of disease” (disease\(_{ODX}\)) intended to “provide practical help in preventing and limiting overdiagnosis” is a “classical concept with determinate boundaries” they fail to provide convincing arguments that it is different from “disease (simpliciter),” which they acknowledge to be a vague non-classical cluster concept (Walker and Rogers 2016, Epub ahead of print). Disease\(_{ODX}\) is “is for use only in borderline cases that are potential cases of overdiagnosis (such as diagnoses of early-stage, non-invasive cancers that are known to be commonly overdiagnosed, or diagnoses of spectral diseases that are close to the diagnostic thresholds and where patients have no other risk factors).” However, the “précising definition of disease” which is to “provide practical help in preventing and limiting overdiagnosis” contains the concept of overdiagnosis. This makes the “précising” definition circular. Providing a definition of disease that points out cases of overdiagnosis that are “known to be commonly overdiagnosed” does not add significantly to the conceptual clarity or to the practical guidance of how to handle or limit overdiagnosis. Moreover, it leaves us with the task of deciding what count as “borderline cases,” and thus which are potential cases of overdiagnosis. In other words, it brings us back to square one.

Additionally, when elaborating on key terms, such as “severe harm” and “significant risk,” the authors recur to vague terms. As they acknowledge, “the definition retains some vagueness” and seems “counterintuitive and with potential for significant relativism” (Walker and Rogers 2016, Epub ahead of print). Hence, it is far from obvious that it is a “classical concept with determinate boundaries.” Moreover, it is difficult to see how this adds much to the RT and to Schwartz’s distinction between disease proper and “risk-based diseases” and as described, disease\(_{ODX}\) appears as vague and non-classical as disease (simpliciter).

Hence, singling out a specific type of disease prone to overdiagnosis by a “précising definition” of disease (disease\(_{ODX}\)) does not “provide practical help in preventing and limiting overdiagnosis” as it depends on a diversity of individual and social values. It gives little guidance on how to assess and bar overdiagnosis as it lacks a clear concept of disease. Most importantly however, the concept of “dysfunction” in the definition of disease\(_{ODX}\) is quite unclear. As the authors explicitly claim: “the causal chains leading from a currently identified dysfunction to future harms are open.” At the same time the concept of “dysfunction” is declared to rule out conditions like pregnancy (Walker and Rogers 2016, Epub ahead of print). Moreover, the “dys” part of dysfunction appears to include some negative evaluation, and it is unclear how this inherent negative evaluation is related to the explicit harm of the definition, i.e., which has “a significant risk of causing severe harm to the patient.”

This problem comes out clearly when the authors claim that any instance of some detectable phenomenon that may result in disease, is diseased: “Someone with colonic polyp with malignant changes but no local or distant invasion, and no experienced harm, nonetheless does have a disease” (Walker and Rogers 2016, Epub ahead of print). However, here the authors presuppose what is to be shown, i.e., whether a given phenomenon is disease. Polyps are frequent in (the aging) population, and most polyps do not develop to colorectal cancer (disease). While I agree with Walker and Rogers that the definition of overdiagnosis in detection of polyps is hard, these cases appear to be non-false.
Hence, there are promising conceptions of disease that explicitly are elaborated or defined in order to address challenges with overdiagnosis. However, the analysis above indicates that they still have some unclear or unsettled issues making them less likely to prevent or limit overdiagnosis. There are two reasons for this. The first is that the attempts are more or less based on dynamic (expansive) or unclear concepts of disease. If overdiagnosis is based on an unclear or expanding concept of disease (Walker et al.) or diagnosis (Carter et al.), no wonder it is unclear or expansive. However, this also makes it more difficult to grasp or curb overdiagnosis. The second reason is that concepts that define overdiagnosis, such as disease, diagnosis, or dysfunction, have unclear relationships to the concept of harm. This can be seen in definitions of overdiagnosis in terms of “correct diagnosis” (Carter et al. 2016a), “non-harmful disease” (Rogers and Mintzker 2016b), or “inconsequential disease” (Hersch et al. 2015). Let me therefore take a closer look at such non-harmful conception of disease or diagnosis. Before doing so, allow me to clarify what I mean with the relationship between disease concepts and harm.

**Disease, dysfunction, diagnosis and harm**

By now it is widely acknowledged that disease is a value-laden concept (Agich 1983; Bolton 2008; DeVito 2000; Khushf 2007; Kingma 2014; Parsons 1958; Sade 1995; Stempsey 1999; Wakefield 2014). There is significant disagreement with respect to what values are involved, e.g., whether they are biological, social, or cultural. However, there seems to be a fairly broad agreement that disease is related to harm of some kind. The body (and the mind) may have millions of different conditions, process states, and mutations. To describe them all appears to be beyond reach, or—more importantly—beyond interest. When we describe, measure, or modify certain conditions, it appears to be because they are meaningful to us. In particular, we are preoccupied with conditions that are of negative value to us (Tranjøy 1967). Pathology appears to come before physiology (Canguilhem 1991), and pathology is related to pathos (suffering, feeling).

Hence, physical or mental conditions tend to come to our attention if they are (or can be) of negative value of some kind. As stated by Wakefield: “no one in medicine actually discusses the disorder status of, say, one harmless dead cell, and salient deficiencies discussed by medical professionals are almost always harmful.” (Wakefield 2014). There may of course be many types of negative value, be they physical, mental, emotional, social, or cultural harm. I have elaborated on the conception of harm and suffering elsewhere (Hofmann 2017), and this is not the place to enter the vast and vivid debate on disease axiology. The point here is that concepts, such as disease, dysfunction, and diagnosis, tend to be related to negative value of some kind. I will elaborate on this below, but what has been said so far suffices to point to some challenges with current conceptions of overdiagnosis that refer to “non-harmful disease” or “correct diagnosis.”

**Diagnosing non-harmful conditions**

One problem with conceptions such as “non-harmful disease,” “correct diagnosis,” and “inconsequential disease” is that they appear as contradictions in terms. If a disease is defined in terms of harm in some way, a “non-harmful disease” poses an inconsistency. Moreover, defining overdiagnosis in terms of “correct diagnosis” or “non-harmful disease” makes it prone to confusion with medicalization (Hofmann 2016). Yet another challenge is that the concept of “non-harmful disease” presupposes what is up for grabs, i.e., whether an instance of overdiagnosis is actually disease.

Such non-harmful conceptions of disease tend to ignore how the phenomena and conditions that define disease are selected. As argued, they are not selected capriciously, but based on expected disadvantages, either biologically (in terms of survival or reproduction), experiential (in terms of pain, reduced capabilities), social (in terms of behaviour), or cultural. Colorectal polyps are not selected by professionals as interesting because of their aesthetics, but because they can develop into harmful cancers which they believe can be prevented. However, colorectal polyps are not cancer, contrary to what Walker and Rogers claim (Walker and Rogers 2016, Epub ahead of print).

So far, I have (1) investigated how conceptions of overdiagnosis explain the expansion of disease and (2) how definitions of disease address the challenges with overdiagnosis. Although acknowledging the usefulness of traditional conceptions of disease and the valuable recent contributions from scholars, I have also identified what I think are weak spots. Let me therefore take the current status as a point of departure in order to see if it is possible to explain the extension of overdiagnosis and the expansion of disease in terms of the process of diagnostics.

**Understanding the process of diagnostics to explain the extension of overdiagnosis and expansion of disease**

In relation to the detection of non-harmful phenomena or conditions, concern about overdiagnosis refers specifically to the problem of attributing a disease status to phenomena or conditions that do not result in harm. Let me therefore investigate this in some detail.
Attribution of disease status to indicative phenomena that are not harmful

Let us call the phenomena that are considered to be related to disease (e.g., as a potential precursor) indicative phenomena (IP). Ductal carcinoma in situ, High- and Low-grade Squamous Intraepithelial Lesions (HSIL/LSIL), faecal occult blood, and adenomatous polyps are but some examples of IP. The crucial question of the overdiagnosis-debate is whether the IP are important for identifying disease (as ever defined) and thereby helping people (however that may be defined).

Accordingly, overdiagnosis can be defined as something that happens because we attribute disease status to indicative phenomena without justification, i.e., overdiagnosis is unwarranted attribution of disease status to IP.\(^8\)

Let us therefore scrutinize the process of how a disease is established, how something (such as an indicative phenomenon) obtains the diagnosis label, and how we can go astray and overdiagnose. Let diagnosis be the label (usually a specific name) given to a condition that is considered to be disease. “Through diagnosis the physician identifies and names disease’’ (Copeland 1977).

How diseases are established, diagnosis are set, and how something becomes overdiagnosis

Let us first start with the ultimate observable and undesirable phenomenon for (most) human beings and for health professionals, death,\(^5\) e.g., death from what we today call colorectal cancer. Death is usually considered to be of negative value.

In order to foresee death (prognosis) human beings in general, and health professionals in particular, have paid attention to symptoms as well as people’s communicated subjective experience, such as pain, feeling tired, reduced ability to perform ordinary tasks, i.e., illness. Attention to such symptoms is sparked by the negative experiences (pain etc.) and the negative value associated with death.

Correspondingly, members of society and social institutions have observed and addressed altered behaviour or reduced ability to fulfill social commitments, i.e., sickness. Again, these phenomena gain attention because they are considered to be of negative value.

Accordingly, medical professionals have paid attention to observable and measurable phenomena, such as blood in the stool, change in bowel movements, weight loss, tiredness, tumour, and ultimately death. Again, the reason why these phenomena gained attention is that they are of negative value to the person or are related to such a value. Traditionally, the relationship between what the professionals observed and measured and what the individual experienced and society assessed was fairly easy to verify by observation over time.\(^10\) If some phenomena were suitable to predict suffering (illness), altered behaviour (sickness), and ultimately death, then they qualified for being called disease. Moreover, examining skills and diagnostic tests were to be verified by autopsy. There was a comprehensible connection between disease and death, and between the identified phenomena, such as tumour, and events of negative value, such as suffering, altered behaviour, and death. See Fig. 1b.

Additionally, professionals and scientists have tried (and succeeded) in finding other phenomena that preclude disease and that indicate that a person will develop a disease. Peroxidase activity of the haeme moiety of the haemoglobin molecule may be an example of such an indicative phenomenon for colorectal cancer. It is neither a necessary nor sufficient condition for cancer and death (if left untreated), but only indicative. Hence, there is no direct link from the IP to negative value. Therefore, we need qualifying tests to verify the significance of the indicative phenomena. Colonoscopy may be one such test, identifying adenomatous polyps and adenoma. Hence, a “gold standard” can be used to verify the indicative phenomenon. See Fig. 1a. The “gold standard” ensured the link between the IP and disease, and thus to negative value (suffering or death).

However, there can be many gold standards, e.g., colonoscopy is the gold standard for colorectal cancer screening, while histology is the gold standard for diagnosing the disease (colon cancer). Therefore, it is important to notice that the “gold standard” does not provide a validated connection from the indicative phenomena to what really matters, i.e., what persons (and patients) experience and is valued as negative, e.g., whether they feel tired, are in pain, unable to work, and whether they die from what we characterize as colon cancer. See Fig. 1c. Hence, overdiagnosis breaks the link between the IP and negative value. Therefore, I think that Carter, Rogers and others are wrong when they claim that overdiagnosis includes “non-harmful disease,” “correct diagnosis,” or “inconsequential disease” because disease and diagnosis include something of negative value, e.g., harm. Correct diagnosis already takes account of harm. Adding harm (or lack of benefit) to this confuses more than it clarifies overdiagnosis.

And here lies the core problem with overdiagnosis. At present, we do not have any “diamond standard” to verify whether indicative phenomena, such as whether a positive faeces test will result in colorectal cancer or death (if not detected and treated) (Hofmann 2016a). Accordingly, overdiagnosis is a problem of lack of knowledge, and not an ontological problem (Rogers and Mintzker 2016b). Or, more precisely, overdiagnosis is a problem with acknowledging uncertainty.
Hence, at the core of the problem of overdiagnosis lies a problem of unwarranted disease labelling, i.e., stamping something as disease without a validated connection to negative experienced phenomena (such as suffering and death). Despite our lack of a “diamond standard” we label IP as disease and intervene against them believing it avoids disease and saves lives. This leap of faith is illustrated in Fig. 1c. Therefore, when Carter, Rogers, and others acknowledge that the IP are correct diagnoses or disease, they may contribute to the process of expanding disease and overdiagnosis instead of limiting it.

Fig. 1. The figure appears to be of poor quality. Can the quality be improved? If the figures need to be converted to grayscale, please make sure that the text is clear and the various shades of grey are discernable.

Conceptual overview of important aspects of diagnostics indicating the dynamic process of establishing diseases

The unwarranted link between indicative phenomena and harm

The unverified and unwarranted labelling of disease results from our strong belief in scientific and technological progress (Hofmann 2002, 2015), high hopes, and our desire to do good. Because the IP sometimes can lead to manifest disease, suffering, and result in death, we tend to label these conditions disease. Figure 2 illustrates this unwarranted labelling process at the core of overdiagnosis.

Fig. 2. The figure appears to be of poor quality. Can the quality be improved? If the figures need to be converted to grayscale, please make sure that the text is clear and the various shades of grey are discernable.

Conceptual overview of the disease labelling process

Discussion

Contrary to contemporary scholars, I have argued that cases of overdiagnosis are not cases of disease. They are non-validated and unwarranted labelling of disease or diagnosis. Disease has a negative value. So has a diagnosis. Both gain their meaning from the meaning of prognosis (Christakis 1999). Overdiagnosis presupposes a link between IP and harm (negative value), and thereby generates a link to disease. However, this link is not validated and unwarranted.

To reveal this unwarranted connection and to understand and bar overdiagnosis we need to acknowledge the process of diagnosis and how diseases are established. Accordingly, conceptions of “non-harmful diseases” and “correct diagnosis” does not pay sufficient attention to the process and making of disease and diagnosis. In particular they ignore the inherent (negative) value-ladenness of disease. Therefore, it is a dead end for defining and addressing the problem of overdiagnosis.
Accordingly, it is not primarily the concept of disease that needs clarification in order to capture and bar overdiagnosis, but the concept of diagnosis itself. In particular it is important to acknowledge the valuation defining the process from IP to diagnosis. Without proper validation of the relationship between IP and disease, we come to give unwarranted (negative) value to IP. As argued, the problem with overdiagnosis is that it is not proper diagnosis. It does not diagnose disease. This does not make overdiagnosis a contradiction in terms. 

However, as overdiagnosis is over-diagnosis, i.e., too much diagnosis.

This study has two important implications with respect to the question of how to bar overdiagnosis. First, we need to break the unwarranted link between IPs and disease and death. We must stop talking about “cancer screening” when we are screening for indicative phenomena of uncertain significance, IPUS. This does not mean that we must stop all screening for IPUS. Some such screening may certainly be valuable, but we must stop calling them “disease screening” (e.g., cancer screening) and end to link such phenomena to negative values and experiences, such as suffering and death. While these imperatives are not always followed at present, they are already a central focus of advocacy and the international research program on overdiagnosis. There is already some focus on using more appropriate names for the conditions that are screened for (Esserman et al. 2013). My hope is that the arguments presented here help bolster existing efforts.

Second, we need to look for validated and warranted links between IP and disease, i.e., we need to lend all our attention and efforts to finding “diamond standards” that can verify the relationship between the indicative phenomena and the ultimate experience (death). Although this may be difficult, and we may have to take small steps to reach this goal, we should always have this goal in mind. One way to approach this goal could be to perform more studies of what has been called “pathologic reserve,” i.e., of IP which people die with and not of. This would both provide us with useful and accurate accounts of the extension of overdiagnosis at the same time as it can generate fruitful hypotheses on what makes some IP develop to disease while others do not.

This study also underscores that it is important to differentiate between diagnosis and detection (of indicative phenomenon). Various instruments and technologies can measure a wide range of physiological, neurological, anatomical phenomena. These phenomena can be given specific characteristics and can be detected and recognized in various situations. However, they do not become diagnostic entities or diseases without some type of (negative) evaluation.

Although this study may appear as a naturalist analysis, it is not. It does not presuppose a Boorsean conception of biologically defined goals or values. I do certainly think that the values that underpin disease and diagnosis (as well as the thresholds that are implied) are defined by humans, i.e., by basic human experiences such as suffering and death. However, I do not subscribe to a hybrid conception of disease in the mode of Jerome Wakefield either. The reason is that I think that the basic mode of evaluation is related to “dysfunction” and that “harmful” adds little of importance, or even confuses the evaluation by including too much.

One problem with this approach can be that its insistence on that only harmful conditions are diseases could lead to underdiagnosis and scupper attempts at prevention. 

Here it is important to notice that the aim of health care is to help people and not harm them. If you make handling non-harmful diseases part of health care’s goal, then treating disease can no longer be part of health care’s goal (as it does not help people). I think there are good reasons to include “treating disease” in the goal of health care, aligned with the overall end to help and not to harm people. Accordingly, disease is something of negative value. However, this does not mean that you should not search for IP that can result in disease. On the contrary. The point made here is only that you should call them by their proper name (and not by disease names) and strive for validating the relationship between these IPs and disease. Hence, the way this approach tries to obstructing overdiagnosis is by barring the use of the term “diagnosis” when it is not warranted.

Then it does not result in underdiagnosis or scupper prevention.

It may also be argued that this conception of overdiagnosis is moving the debate too far away from the concern that accepted medical practices designed to help people was doing so much harm as to begin to undermine the value of any benefits it was delivering. However, moving the overdiagnosis debate from overall (population based) concerns to the diagnostic process itself may address important aspects of overdiagnosis that has not been attended to before. In a way the approach brings overdiagnoses back to its etymological origin, i.e., to diagnosis. Moreover, this may be a fruitful way to differentiate it from other concerns with “too much medicine” such as medicalization.

It may also be argued that most examples stem from screening and that overdiagnosis occurs in a much wider range of conditions. This is surely true. However, as overdiagnosis is most often discussed in screening I have chosen examples that would be familiar to the reader. However, this does not restrict the arguments to screening.

Conclusion

This article has tried to elucidate the relationship between disease and overdiagnosis. It started with investigating how current conceptions of overdiagnosis can explain the expansion of the concept of disease. Then it scrutinized how standard conceptions of disease can be used to address the issue of overdiagnosis. The article also investigated
recent attempts to clarify the relationship between the concepts of disease and overdiagnosis. Several shortcomings were identified leading to a closer analysis of the diagnostic process. Contrary to recent contributions to the field, I have argued that cases of overdiagnosis are not cases of disease. They are unwarranted labelling of disease. Overdiagnosis diagnoses something that is not disease—therefore it is over-diagnosis. However, it presupposes a link between indicative phenomena, such as polyps or cell changes, and harm, and thereby generates a non-verified and unwarranted link to disease. This implies that we must stop attributing disease language to indicative phenomena and strive for scientific progress in differentiating phenomena that cause harm from those that do not. If we pay close attention to what is over-diagnosed, we should be optimistic with respect to reducing overdiagnosis.

Acknowledgements
I am most thankful for wise comments and suggestions by two anonymous reviewers.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Compliance with ethical standards
Conflict of interest The author declare that he has no conflict of interest.

References


10.1007/s11019-017-9776-z


Please note that this is not a comment on the inconsistency of the authors' claim that overdiagnosis is both “harmless disease” and that it is harmful at the same time (Walker and Rogers 2016, Epub ahead of print). I think that Walker and Rogers will argue that the harm referred to in the definition of overdiagnosis is different from the harm resulting from (the events of) overdiagnosis. The first harm is pathogenic while the latter is iatrogenic and does not stem from the “diagnosed condition.” That is, there are several types of harm from diagnosis and disease, and only the iatrogenic harm is relevant for overdiagnosis.

1 Although Carter et al. indicate that professionals do not reject realist conceptions of disease (Carter et al. 2016b), they are unclear on this. Altogether, the reference to the professional community is quite vague.

2 Carter et al. are not clear on the distinction between disease and diagnosis. I will therefore address both and try to clarify the relationship below.

3 This does not mean that classically structured concepts cannot have definitions that are vague. Here the point is that normativistic conceptions of disease refer to norms and values of some kind, and frequently refer to harm.

4 There is also another circularity in the “precising definition,” as disease is defined in terms of a person already considered to be a patient, i.e., as “physiological dysfunction that has a significant risk of causing severe harm to the patient” (my emphasis). Although this may be a result of linguistic imprecision, it indicates the problem with this alternative conception of disease in order to clarify overdiagnosis: it lacks a clear concept of disease.

5 As will be argued later, the qualifier “malignant changes” presupposes what is to be shown, i.e., that it is a disease.

6 Walker and Roger also include non-harmful conditions in their conception of diagnosis. In fact they include it even in their definitions of disease, both in “diseaseODx” and in “disease (simplicer).” Addressing this issue is beyond the scope of this article.

7 This is what Walker and Rogers call “non-harmful disease,” but, as will be explained below, is not disease at all.

8 Just to be clear, this does not mean that death always is undesirable.

9 One of the reasons for this was that experiences, observable phenomena, and altered behaviors and social roles were quite immediate (i.e., late detection).

10 I owe this argument to one of the reviewers.

11 I owe this argument to one of the reviewers.

12 I owe this argument to one of the reviewers.