Advances in biomarker and monitoring diagnostics:

Great markets, not so great health effects

Bjørn Hofmann PhD 1,2
H. Gilbert Welch MD, MPH 2

1. Centre of Medical Ethics at the University of Oslo, and Department for Health, Technology and Society at the Norwegian University of Science and Technology in Gjøvik, Norway

2. The Dartmouth Institute for Health Policy and Clinical Practice, The Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA

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

Word count: 2616
References: 30
Figures: 2 (+ 1 supplementary figure)
Tables: 2

Key words: biomarker, test, overdiagnosis, overtreatment, stakeholder action, benefit, harm

Declaration: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that (1) BH has received funding by The Commonwealth Fund through the Harkness Fellowship at The Dartmouth Institute of Health Policy and Clinical Practice (TDI). The views presented here are those of the author and not necessarily those of The Commonwealth Fund their directors, officers, or staff or of TDI. (2) Neither BH or HGW have any relationships with any companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children have any financial relationships that may be relevant to the submitted work; and (4) BH and HGW have no non-financial interests that may be relevant to the submitted work. Both authors have contributed to the design of the study, data collection, data analysis, revision of the manuscript and both have approved the final manuscript. BH is the guarantor for this study. The study does not require ethics approval. BH affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide license to the Publishers.
Advances in biomarker and monitoring diagnostics: Great markets, not so great health effects

STANDFIRST:

Although intensively promoted new biomarkers and monitors may advance the time of diagnoses in selected patients, they will concomitantly increase the frequency of false alarms, overdiagnosis, and overtreatment in others. To better understand the market conditions and potential health effects of these burgeoning technologies, Bjorn Hofmann and H. Gilbert Welch review four case-studies and explore counterbalancing actions for clinicians to foster testing that provides more benefit than harm.
Tremendous technological advances and ample venture capital are combining to produce new medical diagnostics. New biomarkers are being identified to predict or detect a wide range of diseases; new devices are being developed to continuously monitor biologic parameters – often connecting with mobile devices to provide user-friendly updates of health status (mHealth). One vision is that these new diagnostics can transform medicine from treating disease to promoting health, from being reactive to being proactive, and from being general to being personal.\(^1\)

Another vision is less sanguine: namely, that it is equally possible that new diagnostics will warrant a black box warning.

<table>
<thead>
<tr>
<th>WARNING:</th>
</tr>
</thead>
<tbody>
<tr>
<td>The enclosed test results may induce worry – or, alternatively, a false sense of security. It is also possible that the results themselves are wrong. Any risk estimates provided may have a considerable margin of error and be subject to dramatic revision in the future.</td>
</tr>
<tr>
<td>The question of what to do with this information is unknown. The best course of action may be to do nothing, although you may experience considerable pressure for further evaluation and intervention that could pose additional health risks.</td>
</tr>
<tr>
<td>The resulting wild goose chase may lead to personal bankruptcy in some countries.</td>
</tr>
</tbody>
</table>

Efforts to detect disease early can always be accompanied by unintended harms. False alarms and indeterminate findings can worry patients, drive more testing, increase clinical workload, and distract clinicians from more important work. Overdiagnosis can lead to unnecessary treatments. Promotional campaigns will necessarily need to get people concerned about disease and suggest that the path to health is through testing – reinforcing health anxiety in some and distracting many from more important
health behaviours. And the process has real financial costs, some of which may directly fall on the patient.

In this article, we consider how clinicians could handle emerging diagnostics. We begin by investigating four case studies selected from the business database *Factiva*, based on both investor interest and their application in common diseases (Table 1). We go on to explore the market conditions, both the investment climate and the misleading feedback favouring further market growth. We conclude with specific actions for clinicians to minimize harm.

**FOUR CASE STUDIES**

**Immunosignature for cancer & infections**

*What is it?* Imunosignature, although not currently FDA approved, is an emerging technology that attempts to predict impending disease by analysing how an individual’s antibodies bind to proprietary arrays of random-peptides (numbering from the tens of thousands to several million). The technology has been investigated in diabetes, Alzheimer’s, infectious diseases, and cancer. A recent Nature Communications article reported on examining the sera of patients with 1 of 6 cancers or 1 of 6 infectious diseases (10 patients per disease). For each disease, researchers identified the top 50 most informative peptides – those peptides most able to discriminate among diseases.

*What are the claims?* HealthTell™, a company emerging from Arizona State University, argues that immunosignature represents “a new concept in healthcare: continuous monitoring of healthy people to detect and treat disease early … If widely utilized it has the potential to reduce medical costs greatly and can remodel our expectations of human health.”

*What are the potential benefits and what is the evidence?* Imunosignature can potentially detect and predict a wide range of diseases and become a useful tool in clinical practice. Initial studies report high sensitivity for multiple diseases (95%)\(^2\). Although no high quality evidence of improved patient outcome is available, researchers report that “to date, no diagnostic of which we are aware can simultaneously discriminate between six cancers and six infectious diseases using the same platform.”\(^2\). Reported sensitivities, however, may be overstated due to overfitting. While it may be possible identify variables that discriminate among conditions in a selected dataset, those variables may be less predictive in the general population. Sensitivities may also be overstated due spectrum bias, in which tests perform better at the extremes of the disease spectrum than they do in the grey area in-between, i.e. the clinical setting in which they are typically used.\(^5\) It is one thing to distinguish
overt cases of dengue fever from syphilis or myeloma from lung cancer; it is something quite different to distinguish who will and will not develop disease among a group of people at similar risk.

What are the concerns about harms and costs? Similarly, reported false positive rates (1-2%)\(^4\) may be understated as they were obtained from healthy volunteers (university students in these studies). Adding to the complexity, immunosignatures may vary greatly across healthy individuals and fluctuate over time within individuals. Adequately detecting impending disease may come at the cost of multiple false alarms and overdiagnosis. The technology will also raise a series difficult clinical, ethical, and health policy questions: What should be done for an individual who has concerning immunosignature, but who is otherwise well? Which consequences will it have for her employment and insurance? What are the downstream costs? Knowledge about such pertinent issues is missing.

**Breath test for lung cancer**

What is it? Breath testing, also characterized as a “cancer sniffing sensor,” is a technology that measures volatile organic compounds in an effort to detect disease.\(^6\) The most biologically plausible application for breath testing is lung cancer screening. Current efforts have involved measuring the concentrations of 4 to 33 distinct compounds which are then combined in a risk score and subsequently dichotomized into normal and abnormal. Breath tests are so far not FDA approved.

What are the claims? One developer of breath testing, Owlstone Nanotech, claimed that the device “could save 10,000 lives a year and save the NHS £245million.”\(^7\)

What are the potential benefits and what is the evidence? Breath testing could serve as an alternative to low-dose CT screening providing simple and widespread screening without ionizing radiation.\(^8\) Moreover, it could reduce screening costs, facilitate early detection of disease and reduce mortality. As clearly expressed by Dr. Norman Edelman of the American Lung Association: “It’s really the future of medical testing in general. We are just scratching the surface on the utility of breath testing in medical diagnosis. ... We could screen many, many more people for lung cancer.”\(^9\) Sensitivity is reported to be between 51% and 100%.\(^10\) This variation in reported sensitivity likely reflects the lack of standardization in breath collection and risk score methodology, as well as irregular validation in independent population samples.\(^10\)

What are the concerns about harms and costs? Reported false positive rates vary widely: between 0% and 83%.\(^10\) Many may get an incorrect and alarming notice about the risk of having a feared and deadly disease. False positive tests engender extensive follow-up testing, which in turn would produce considerable biopsy related risks and overdiagnosis, not to mention extra costs. Were the breath test priced low enough, it could easily be marketed well beyond the current target population for lung cancer screening (to those outside of the 55 to 80 year age range and, in particular, to non-smokers). If this occurs, the harms and costs would escalate – with little (or no) corresponding benefit.
Patch vital sign monitoring

What is it? Several FDA approved or CE marked patches are now commercially available for monitoring and transferring data wirelessly to another device (e.g. smartphone, tablet, central monitor). Patches can continuously measure a variety of parameters: ECG, heart rate, respiratory rate, activity, and posture (accelerometer). Accompanying software summarizes these data and provides customizable alarm thresholds, and has functions to assess heart rate variability, activity, energy expenditure, and balance.¹¹

What are the claims? Zephyr Technology, a producer of Biopatch, believes that “With Zephyr, you really can Measure Life...Anywhere!”¹²

What are the potential benefits and what is the evidence? The systems increase the portability of existing monitoring systems and may provide more accurate diagnosis and appropriate clinical management for a wide range of diseases.¹¹ Potential benefits include the detection of falls, cardiac events, and developing bed sores. Reported ECG parameters, heart rate, respiratory rate, and accelerometer measurements correlate closely with those obtained from conventional instrumentation. The performance data, however, frequently come from a small number of healthy males in controlled settings (exercise physiology labs), and likely overestimate the real-world performance for those who are less healthy and in less ideal measurement environments.

What are the concerns about harms and costs? System prices lie around $1,500 and expenses for disposable patches varies. Little is known about performance in everyday use – or the correctness and clinical importance of abnormalities detected by the systems’ functions and algorithms.¹¹ Statistical noise and artefacts (e.g. due to inadequate sensor contact) may result in frequent false alarms, and expectedly, unnecessary worries, visits, and costs. A randomized trial of real-time monitoring for a single physiologic variable – pulmonary impedance in heart failure patients – found that monitoring was associated with three times as many clinic visits and significantly more hospitalizations.¹³ Monitoring multiple variables might be expected to compound this problem, and widespread implementation would seem to be a recipe for multiple false alarms – resulting in more visits, more testing, more referral, and more fear.

Biomarkers for Alzheimer’s Dementia

What is it? Multiple blood based biomarkers have shown to be able to differentiate Alzheimer’s disease (AD) from healthy controls¹⁴ and to predict onset and progression AD many years before symptom debut.¹⁵

What are the claims? According to a press release from one research group, “This new blood test can accurately reflect development of AD up to 10 years prior to clinical onset.”¹⁶

What are the potential benefits and what is the evidence? Sensitivities over 95% for AD detection have been reported¹⁷ and over 90% for predicting AD.¹⁸ However, published results have been hard to replicate.¹⁴ Variability in assays and the
variability of biomarkers in the same individual over time are challenges yet to be accounted for. Many tests are developed and verified on the same population and thus lack real external validation. Test performance is frequently measured in two distinct populations: sensitivity among patients with overt Alzheimer’s disease, false positives among normal controls – again introducing spectrum bias. Adding to the complexity, different tests are tested against different gold standards for what constitutes Alzheimer’s disease.

*What are the concerns about harms and costs?* Reported false positive rates are high (10-30%) implying many may be given a false AD diagnosis. Even for persons correctly diagnosed or predicted, they will then have to face the challenge of what to do with a positive result – as the disease is not currently actionable. While early detection may help people plan and prepare, it can also result in emotional despair, stigma, and discrimination. In the want of clear definitions, clinicians may be tempted to use biomarkers as a quantitative and objective gold standard for the diagnosis of Alzheimer’s disease. This might not only increase the prevalence of AD, but also have serious implications for the right of persons to drive, make a will, and handle financial affairs. If biomarkers genuinely produce long lead times (e.g. 10 years before clinical onset), they will simultaneously produce ample potential for overdiagnosis – as many will die from other diseases before they develop overt Alzheimer’s disease.

**MARKET CONDITIONS**

**Investment Climate**

The producers of these four diagnostics are enthusiastic about the size of their potential market. HealthTell™ sees a substantial market for immunosignature, “With over 170 million people in the US being affected by neurological, autoimmune, oncologic, metabolic and infectious diseases, the importance of early detection and monitoring is paramount.” Investors apparently agree as the company has raised $40 million to development and commercialize immunosignature. While there are multiple potential producers of breath tests, Grand View Research expects the global Breath Analyzer Market is expected to reach USD 11.3 billion by 2024. Qualcomm Life also sees a big market for patch based vital sign monitoring systems, “There are 300 million people in Europe and North America and 860 million worldwide with at least one chronic disease.
It is estimated that 25% of patients would immediately benefit from wireless home monitoring solutions.”23 And producers of biomarkers for Alzheimer disease envision “sales to primary-care and neurology practices as a periodic, routine screening test recommended for individuals 65 years and older, representing a population of about 45 million people in the US, for a potential yearly market cap of roughly $3 billion.”24

The enthusiasm goes well beyond these four diagnostics, however, there are bullish expectations for the diagnostic industry in general (Figure 1). Revenues for the global biomarker market are expected to more than double between 2012 and 2018 (from $22.4 to $53.6 billion) and are expected to reach $100 billion by 2020.25 Global mHealth market revenues are expected to increase more than 10-fold between 2012 and 2018, albeit from a lower baseline (from $1.5 to $21.5 billion). Clearly, some see a vast market for new diagnostics – as we are all potential customers for early diagnosis.

Figure 1  Global mobile health (mHealth) and biomarker market revenues in billion USD in 2012 and estimated revenues for 2018.26 27
Misleading feedback favoring market growth

The early diagnostic market also has an idiosyncrasy: successful marketing can rapidly produce misleading positive feedback. Testing tends to promote the demand for more testing, regardless of the genuine utility of the test itself (Figure 2). At the population level, testing tends to increase the apparent prevalence of disease and abnormalities – fostering more concern and thus more testing. At the same time, testing tends to identify patients with milder forms of disease and abnormalities. These patients invariably do better than those diagnosed in the past – apparently reinforcing the utility of testing.

Figure 2  Misleading feedback favouring market growth both at the population and individual level.
Testing feedback is equally positive at the individual level, regardless of the test result. Because most tests are negative, most people will have the positive experience of being reassured by testing. Those whose results are shown to be falsely positive by subsequent testing, may nonetheless experience a sense of relief – perhaps explaining why false positives are not a significant deterrent to screening. The finding of consequential abnormalities provides the strongest positive feedback, as these patients are presumed to have benefited from the test and subsequent intervention. Ironically, those who have experienced the most substantial harm of testing, overdiagnosis, view themselves to be in the same group and are thus equally enthusiastic about testing.

Apparently favourable feedback nourishes the fresh wind of enthusiasm amongst investors, patients, and health policy makers. But they also reinforce the harms of testing: increasing health anxiety, false alarms and overdiagnosis. The resulting increased workload distracts clinicians from more important work and the focus on testing distracts patients from more important health behaviours. It is misleading feedback that requires a strong counterbalance.

**COUNTERBALANCING ACTIONS**

In countries where health care is market driven, payers may want to incentivize patients in testing decision-making by making them have “skin in the game.” For those employing cost-sharing strategies, we suggest they bundle the cost of expected downstream testing into one price. If a $100 test, for example, leads to a $2000 test 10% of the time – then the bundled test price would be $300 ($=|$100 +
Bundled pricing would provide the dual benefit of motivating careful consideration before testing, while covering patient downstream costs.

In countries where health care is better regulated, regulators will want to require rigorous assessments and force manufacturers to explicitly state how the new tests add clinical value before approval. Ideally, approval of new diagnostics would be contingent on a randomized trial demonstrating the beneficial effect on a patient-centered outcome. Practically, given the large sample size and long follow-up required, such trials will rarely occur.

Consequently, an important counterbalance will be rigorous assessments by cadre of researchers who represent the public’s interest, not that of industry. Three questions for researchers stand out. First, is the test reliably predictive of a health event that matters to patients? Second, is there an effective action that reliably lowers that risk? Although many tests fail on this account, those that produce benefit for some must be subjected to a final question: What happens to those who do not benefit? Answering this question requires routine surveillance looking for excessive false positive rates and excessive diagnostic yields (a warning sign of overdiagnosis).

Ultimately, however, we believe clinicians will serve as the most important counterbalance to these favorable market conditions. Four behaviors stand out (Table 2):

Educate patients: Most patients believe testing can only help them. Clinicians need to communicate that testing is a double-edge sword, with the potential to unleash clinical cascades that lead to distress, interventions, and physical harm. Self-testing should be discouraged. And when testing is warranted, clinicians should prepare patients for unexpected findings, e.g., a concerning immunosignature, and the possibility that ignoring them may be the best course of action.
Respect baseline risk: Diagnostics have been traditionally directed toward individuals with symptoms. The presence of symptoms is a powerful piece of information about an individual’s prognosis – one that suggests elevated risk. Here the harms of testing are typically small relative to the benefits. But moving testing and monitoring to people who have no symptoms changes this balance. There is less potential for benefit, simply because there baseline risk is low, but a similar potential for harm. If new diagnostics are to be used, they should be focused on those at the highest risk (e.g. breath test for heavy smokers) while avoiding testing in those at low-risk.

Think downstream: Before testing, consider the downstream implications. What will you do differently? If the answer is “nothing,” avoid testing. Consider not only whether a positive result is genuinely actionable (e.g. a patch alarm or a positive biomarker for Alzheimer’s), but also whether the result may lead to stigma and distress, unnecessary subsequent testing, overdiagnosis, and overtreatment.

Expect misleading feedback: Inoculate yourself, colleagues, and patients from apparently concerning reports following testing. Expect reports of rising disease prevalence following additional testing and recognize that epidemics may be more apparent than real. Inoculation is equally important for optimistic reports. Expect outcomes for the typical patient to improve and provide the alternative explanation for powerful stories of individuals who attribute their life to the test – they were overdiagnosed and needlessly treated.

Summary

Innovative technologies and ample venture capital are combining to produce new disease biomarkers and mobile monitoring devices. These new diagnostics represent tremendous technological advances, but do not automatically provide improvements in clinical care and population health. Diagnostic efforts to
determine what might be wrong – or what might go wrong in the future – can initiate a cascade of events that turn well people into sick patients. We must target new diagnostic tests to address real health problems, not to generate them.

KEY MESSAGES:
- Innovative technologies and ample venture capital are combining to produce new disease biomarkers and mobile monitoring devices.
- These new diagnostics represent technological advances, but do not automatically provide improvements in clinical care and population health.
- They have the potential to help some, but also to increase the frequency of false alarms, overdiagnosis, and overtreatment in others.
- Excessive testing and false alarms may increase health care workload and shift clinicians’ focus towards the healthy.
- Misleading feedback at both the population and individual level tends to favour further market growth.
- Clinicians must provide a strong counterbalance: educating patients, respecting baseline risk, thinking downstream, and expecting misleading feedback

TABLES
<table>
<thead>
<tr>
<th>Test (description)</th>
<th>Proposed Benefits</th>
<th>Potential Harms</th>
<th>Cost Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Immunosignature testing</strong> (255-350 antibodies measured in blood)</td>
<td>Early detection of cancer and infections.</td>
<td>Reportedly rare: 1-2% false positive rate</td>
<td>Unknown Overtreatment risk would seem high – despite absence of any data on what to do for a concerning immunosignature.</td>
</tr>
<tr>
<td><strong>Patch for vital sign monitoring</strong> FDA cleared (Skin patch measuring EKG, heart and respiratory rate, activity and posture.)</td>
<td>Detection of falls, cardiac events, and developing bed sores.</td>
<td>Likely common: due to statistical noise and artifacts (e.g. motion or faulty leads)</td>
<td>Unknown False alarms may result in additional testing and subsequent overdiagnosis and overtreatment.</td>
</tr>
<tr>
<td><strong>Biomarkers for Alzheimer's disease</strong> (9-163 biomarkers measured in blood)</td>
<td>Risk prediction and early detection of Alzheimer's disease.</td>
<td>Common: 10-30% false positive rate</td>
<td>High risk of overdiagnosis in those with limited life expectancy and long lead times of diagnosis. In this setting, proposed interventions would constitute overtreatment.</td>
</tr>
</tbody>
</table>
Table 2  Actions for clinicians to assure proper testing

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Educate patients</strong></td>
<td>Inform patients not only about the potential benefits but also of the dilemmas and harms that testing may entail. Prepare them for unexpected findings and inconclusive results.</td>
</tr>
<tr>
<td><strong>Respect baseline risk</strong></td>
<td>Avoid testing low-risk persons – particularly when false positives are common (or require invasive follow-up testing) and when the risk of overdiagnosis is high.</td>
</tr>
<tr>
<td><strong>Think downstream</strong></td>
<td>Consider all downstream implications before testing, in particular whether the test is actionable and whether it leads to distress, stigma or has implications for patients’ insurance. Avoid unnecessarily increasing the health care workload.</td>
</tr>
<tr>
<td><strong>Expect misleading feedback</strong></td>
<td>Expect incidence and prevalence to rise when trying to detect disease early or applying more sensitive tests. Expect outcomes to improve if you treat milder cases.</td>
</tr>
</tbody>
</table>

References


