Patient Diagnostics

International Edition: DOI: 10.1002/anie.201411741
German Edition: DOI: 10.1002/ange.201411741

From the Bench to the Field in Low-Cost Diagnostics: Two Case Studies

*Ashok A. Kumar, Jonathan W. Hennek, Barbara S. Smith, Shailendra Kumar, Patrick Beattie, Sidhartha Jain, Jason P. Rolland, Thomas P. Stossel, Catherine Chunda-Liyoka, and George M. Whitesides**

Keywords:
field trials · liver injury · point-of-care diagnostics · sickle cell disease · technology-translation

Despite the growth of research in universities on point-of-care (POC) diagnostics for global health, most devices never leave the laboratory. The processes that move diagnostic technology from the laboratory to the field—the processes intended to evaluate operation and performance under realistic conditions—are more complicated than they might seem. Two case studies illustrate this process: the development of a paper-based device to measure liver function, and the development of a device to identify sickle cell disease based on aqueous multiphase systems (AMPS) and differences in the densities of normal and sickled cells. Details of developing these devices provide strategies for forming partnerships, prototyping devices, designing studies, and evaluating POC diagnostics. Technical and procedural lessons drawn from these experiences may be useful to those designing diagnostic tests for developing countries, and more generally, technologies for use in resource-limited environments.

From the Contents

1. Introduction

One focus of lab-on-a-chip technologies has been the creation of point-of-care (POC) diagnostics.[1–3] Publications describing these efforts have appeared at an exponentially increasing rate (Figure 1). Despite this activity, the promise of devices that allow personalized, affordable healthcare remains, to a large extent, just that: a promise.[4,5] What steps are necessary to bridge the gap between publications on POC tests, and fully developed POC tests that are actually used to improve healthcare?

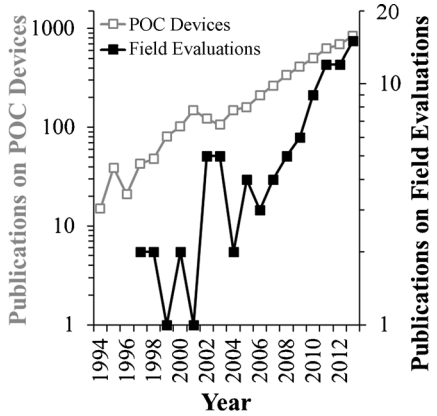


Figure 1. The number of publications with the topic of point-of-care tests or diagnostics (Publications on POC Devices) and publications on the field evaluation of such devices (Publications on Field Evaluations)—in clinical or low-resource settings—have both increased exponentially over the past decade. Publications on the devices themselves are about 60 times more frequent than publications about the field evaluation of the devices. Numbers of publications are based on results from the Web of Science (Thomson-Reuters) for the topic ("point-of-care" AND (diagnostic OR test)), for "Publications on POC Devices," and for the topic, (("point-of-care" AND (diagnostic OR test)) AND ("field trial" OR "field evaluation" OR "clinical trial" OR "clinical evaluation")) for "Publications on Field Evaluations".

Transitioning from a concept in a laboratory, to a product in the hands of users, is a difficult and expensive task in most fields, but it can become Herculean in medicine. Large clinical trials, and complex processes for regulatory approval, both require long times and substantial resources (financial and human). As the first of several steps, a POC test must go from the bench to the field.

There is an argument, sometimes heard in universities, that academic scientists should be concerned only with discovering new methods and enabling new technologies, and that actual reduction to practice—including testing in the field—should be left to companies. This procedural model works (although imperfectly) when potential profits are large and risks are low. In these circumstances, market forces will sometimes encourage companies to invest in the full development of early stage technologies. Technologies targeting the bottom of the pyramid,[6] however, do not fit this economic

[*] Dr. A. A. Kumar, Dr. J. W. Hennek, Dr. B. S. Smith, Prof. G. M. Whitesides
Department of Chemistry and Chemical Biology
Harvard University
12 Oxford St., Cambridge, MA 02138 (USA)
E-mail: gwhitesides@gmwgroup.harvard.edu
Dr. S. Kumar, P. Beattie, S. Jain, Dr. J. P. Rolland
Diagnostics for All
840 Memorial Drive, Cambridge, MA 02139 (USA)
Dr. T. P. Stossel
Hematology Division and Center for Biomedical Innovation
Brigham and Women's Hospital
One Blackfan Circle, Boston, MA, 02115 (USA)
Dr. C. Chunda-Liyoka
Department of Paediatrics, University Teaching Hospital
Nationalist Rd., Lusaka (Zambia)
Prof. G. M. Whitesides
Wyss Institute for Biologically Inspired Engineering
Harvard University
60 Oxford St., Cambridge, MA 02138 (USA)

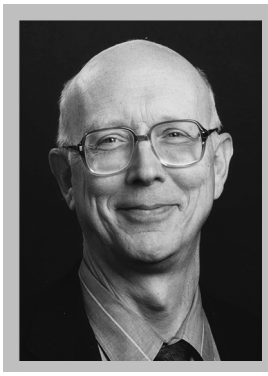
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201411741.

model because they often provide limited financial incentive and a high risk of failure in their intended application (often for reasons having little to do with the technologies).

By performing their own initial field evaluations, academic scientists and engineers can improve their technologies, lower the costs of further development, and encourage the additional investment needed to make a technology "real". This translational step is itself no small task, because it requires a set of skills and experiences that—while common in fields such as public health—are rare in the academic departments of chemistry, biology, and bioengineering, in which much of the advanced research in POC diagnostics has been done.

Publications on field trials and clinical evaluations of POC devices illustrate the difficulty of going from the bench to the field. Each year, only one paper is published about field testing or clinical evaluation of POC devices for every 60 papers published about laboratory tests of devices aiming toward POC diagnostics (in all countries). This ratio has been fairly consistent for the past two decades (Figure 1). Our objective in sharing our experiences in two representative technologies is to decrease this ratio by describing the less scientific, but remarkably interesting and difficult (technically and procedurally) challenges that must be overcome to convert publications into a field-tested prototype.

Here we describe a general framework for creating POC diagnostics, and illustrate its structure with two case studies. In the last seven years, we have taken two technologies from the bench to the field: 1) we used paper-based microfluidics[7] to create a liver function test[8,9] and evaluated the device in field trials in a hospital in Vietnam,[10] and 2) we developed self-forming step-gradients in density[11] to create a test for sickle cell disease,[12] and evaluated this test in a clinical setting in Zambia.[13] (In both cases, we have not yet obtained regulatory approval or created final products.) In one case, development and field testing were done through a strong partnership with a nonprofit company. In the other, our academic laboratory led the trials. By sharing these details, we hope that others interested in creating POC devices can benefit from the lessons we have learned, and can anticipate some of the challenges of developing and field-testing new technologies that are ultimately intended for reality.



George M. Whitesides received his AB degree from Harvard University in 1960, and his PhD from the California Institute of Technology in 1964 (with J.D. Roberts). He began his independent career at MIT, and is now the Woodford L. and Ann A. Flowers University Professor at Harvard University. His current research interests include physical and organic chemistry, materials science, biophysics, water, self-assembly, complexity and simplicity, origin of life, dissipative systems, affordable diagnostics, and soft robotics.

1.1. What is a Field Trial?

A scientist who has designed a new method to detect a disease in a simple, portable device might believe that the most appropriate way to test the device in the field would be to travel to rural clinics in low- and middle-income countries (LMICs) and begin using devices with patients. If the device is designed for the POC, should not the POC be the best place to test the device? Testing the performance of a device at the POC is essential, but not necessarily the first work that should be done in the field.

"Field trials" and, more generally, "field work", refers to a wide range of activities (Figure 2). Work in field settings can be broken down into four tiers:

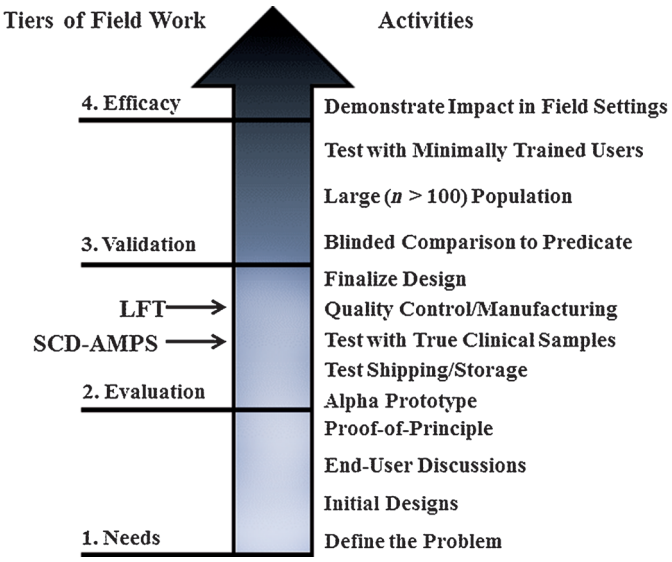


Figure 2. The spectrum of field work can be broken down into four tiers. Progressing upwards from an initial assessment of needs, the activities and milestones that a device must pass requires increased time and resources. Field trials of a POC diagnostic for low- and middle-income countries (LMICs) can refer to testing in an LMIC done anywhere above Tier 2. Field trials of the paper-based liver function test (LFT) and the test for sickle cell disease using aqueous multiphase systems (SCD-AMPS) were between Tier 2 and Tier 3.

1) needs: identifying a problem and understanding its context;
2) evaluation: testing a prototype of a device in the field to identify problems;
3) validation: demonstrating clinical performance in a field setting; and
4) efficacy: testing whether the use of the device has an impact on outcomes in health.

The two cases we describe were between Tier 2 and Tier 3 when they were evaluated in the field. In both cases, however, field work began much earlier, and included an assessment of needs, and an evaluation of designs with potential end users.

2. A Framework for Development

2.1. Defining a Problem

A successful technology must solve a real problem. Based on our experience and on reviews on POC diagnostics,[1,3,14,15] for a diagnostic for global health to make an impact, the problem addressed by the device should, generally, have four characteristics:

1) A substantial number of people should be impacted by the disease for a research program intended to ameliorate the problem to catch the attention of funding agencies, nongovernmental organizations (NGOs), and governments. That is, it should ultimately be sufficiently important for some agency concerned with its amelioration to be willing to implement (and help pay for) a successful solution.[16]
2) The diagnostic device should provide actionable information that can improve the well-being of the patient.
3) Simple interventions should already exist to treat the patient, or—for untreatable diseases (e.g. dengue, or, beyond symptomatic treatment, Ebola)—to point toward a public health intervention once a diagnosis has been provided.
4) Existing solutions to the problem should be inadequate or unaffordable.

The last criterion is especially important when the final goal is to create a useful product. Analyzing the market for current solutions provides an idea of the level of performance that must be surpassed to make a substantial impact.

To identify needs and outstanding problems in global health, reports from the World Health Organization (WHO) and the latest list of Grand Challenges from the Bill and Melinda Gates Foundation (BMGF), provide only a starting point. Both organizations do extensive research on the ground to develop these lists, but by their nature, these lists neglect the details. In POC applications, the detail is where the devil lives.

To understand better the problems in resource-limited environments that could be addressed by technology, experience at the point-of-care is necessary, whether it be in rural villages in Africa, urban slums in India, or forward-deployed military units in combat zones. When this experience is not directly available in a university laboratory, it is invaluable to find partners who understand these needs, and can explain them in the context of technical challenges.

2.2. Building a Team

Creating an interdisciplinary team with global scope requires little more than searching for potential partners, sharing ideas, and valuing what every member brings to the discussion. For scientists and engineers in academic departments, attending symposia aimed at medical audiences, or emailing doctors and researchers at local hospitals or nearby schools of public health are easy ways to find partners who can guide capabilities toward real problems.

International collaborations can often be initiated without boarding an airplane. In the US and Europe, many hospitals and medical schools have doctors who have worked in LMICs. Many of these doctors operate, or work with, NGOs that may be a resource for conducting trials. In developing countries, Ministries of Health and local NGOs can be valuable partners.

Trust and mutual understanding are important aspects to any partnership, but especially so in international collaborations. Online communities and forums, such as Global Health Delivery Online (http://www.ghdonline.org) and Engineering for Change (http://www.engineeringforchange.org), make it much easier to begin communicating with potential partners, but establishing a good working partnership is often best done in person. There are a number of international conferences and workshops on POC diagnostics each year. Universities also offer courses that take students and professors abroad to work with organizations in the field. Taking advantage of these opportunities can lead to partnerships and new research programs. Partnering with organizations that have experience doing field work is another option. Table 1 lists several organizations and resources to help establish partnerships.

2.3. Design Considerations for a Diagnostic Device

A well-defined problem requires that solutions fit specific constraints. The ASSURED criteria (affordable, sensitive, specific, user-friendly, robust, equipment-free, delivered), developed by the WHO,[17] provide a rough guide to, and check-list for, the design of POC devices, but should be understood in the specific context under which they were developed: criteria for rapid tests for sexually transmitted infections in a world before the now ubiquitous cell phones, and in which electricity was less available than it is (in many locations) now. These criteria are important, but they are guidelines and not a substitute for the considerations and requirements that are specific to a particular disease.[1] Some problems require quantitative measurements, while others only need a yes/no answer. In some cases, the ASSURED criteria may be too constraining. For example, the "equipment-free" in ASSURED may not apply for devices to be used in a district hospital that supports small-scale equipment. Similarly, a cell phone could count as a piece of equipment, but simple mobile phones are now common, even in many remote villages. (More advanced cell-phones that are now often proposed for use in resource-limited, POC applications, often are not available.)[18]

Those developing POC technology can visit sites where they intend their devices to function, or they can work with clinical and international collaborators to learn the detailed context of their diagnostic target.[19,20] Designing a device with a specific context and problem in mind avoids unnecessary restrictions, while recognizing constraints that are crucial to success in a field evaluation. For example, the allowable time for an assay and whether tests are run individually or in batches will depend on the daily workflow of the end user. Other considerations include understanding how end users

Table 1: List of potential partner organizations with global reach.

Organizations	Type	Headquarters	Countries of collaboration	Experience developing diagnostics
Bill and Melinda Gates Foundation (BMGF)	NGO	Seattle, WA, USA	>100	yes
Center for Emerging & Neglected Diseases (CEND)	university center	Berkeley, CA, USA	worldwide	yes
Center for Integration of Medical Innovation and Technology (CIMIT) & Center for Global Health (CGH)	consortium	Boston, MA, USA	>20	yes
Clinton Health Access Initiative (CHAI)	foundation	Boston, MA, USA	>25	yes
D-Lab (MIT)	university center	Cambridge, MA, USA	>20	yes
Engineers for a Sustainable World (ESW)	network	Pittsburgh, PA, USA	>9	
Engineers Without Borders (EWB)	network	Denver, CO, USA	>47	
Foundation for Innovative New Diagnostics (FIND)	NGO	Geneva, Switzerland	>60	yes
Global Scientific Solutions for Health, Inc. (GSSHealth)	consultants	MD, USA	>15 (Africa and Asia)	yes
John Snowe Inc.	consultants	Boston, MA, USA	>75	yes
Médecins Sans Frontières (MSF)	NGO	Geneva, Switzerland	worldwide	
National Center for the Advancement of Translational Science (NCATS-NIH)	government institute	Bethesda, MD, USA	worldwide	yes
National School of Tropical Medicine (Baylor)	university center	Waco, TX, USA	>7	
Partners in Health (PIH)	NGO	Boston, MA, USA	>12	
Program for Appropriate Technologies in Health (PATH)	NGO	Seattle, WA, USA	>70	yes
Sandra Rotman Centre (University of Toronto)	university center	Toronto, ON, Canada	worldwide	yes
Stanford Biodesign & Center for Innovation in Global Health (Stanford)	university center	Stanford, CA, USA	worldwide	yes
U.S. Agency for International Development (USAID)	government agency	Washington, DC, USA	>100	yes

typically receive medical supplies, where patients are currently sent to receive tests that are not available at the POC, how medical records are kept, what infrastructure exists (e.g. telecommunications, water, electricity), and how much time health workers spend with each patient.

In addition to the context in which a device is used, it is important also to consider the intellectual property landscape of a POC diagnostic at an early stage. In addition to searching the literature, research to develop practical technologies should include an evaluation of the "freedom to operate"—the freedom to develop and commercialize an idea without infringing upon the intellectual property of others. When working on technologies for LMICs, the freedom to operate may be less well-defined than in the US or Europe. Nevertheless, for both the technologies discussed, we applied for patents 1) to make it easier to attract commercial partners for development of products for noncompetitive developed markets,[21] 2) as a defensive strategy to preserve our freedom to operate,[22,23] and 3) to provide a tool to use in cross-licensing to ease our access to intellectual property belonging to others.[22] Since intellectual property is expensive to obtain and maintain, and since one good outcome of the type of technology we wish to develop would be to have it copied and developed by others in the LMICs, the best strategies for intellectual property are not always well defined.

2.4. Funding

Funding field trials is a challenge: these activities fall in the wasteland between traditional academic research and corporate development. A first step is to understand costs. Depending on the size and scope of the field evaluation, costs can range from $15000 (e.g. a week-long program in a rural setting to get end-user feedback) to over $100000 (e.g. a six-month clinical evaluation of performance on several hundred subjects). A frank and open relationship with the international partners on the team is required to estimate accurately costs for personnel, equipment, and local transportation. The team can then factor field testing into budgets for grants from familiar funding agencies, such as the NIH in the US and the European Commission in Europe, or foundations sympa-

thetic to global health causes, such as the BMGF or The Wellcome Trust.

Many institutions and nonprofit organizations offer "accelerator" or "translational" grants, designed to position a technology in such a fashion that it becomes easier to obtain funding from companies or venture capitalists. Design competitions or "hack-a-thons" can also provide money, insight from others in the field, and a mechanism to build a team.

Funding may also be available from LMICs interested in technologies that will benefit their citizens. Even when money is not available, in-kind services—such as access to space in a clinical laboratory or accommodations in dormitories—may reduce the funds needed from external sources. Donations can further reduce the cost of a field evaluation; companies that make laboratory and medical products are sometimes willing to donate ancillary supplies when there is a potential benefit to the public.

2.5. When Is a Device Ready for Field Work?

Ideally, one would test a device in field settings early and often throughout the entire spectrum of development and field testing (Figure 2).[24] The constraints of time and funding, however, require a more judicious use of resources. Much preliminary work can be done in a laboratory or with hospitals in developed countries. The state of development that a device must reach before testing in the field depends on the specific problem being addressed, and the objectives of the field work. How can one tell whether a device is in a phase of development that will benefit from evaluation in the field?

A field evaluation of a prototype device (between Tier 2 and 3) can quickly identify the most critical weaknesses of the device. The effect of environmental conditions (e.g. temperature and humidity), variations in biological specimens (i.e. testing on true samples versus surrogates), and problems with use and interpretation are all critical challenges that can be identified. These kind of studies can be short (<1 month) and require a more modest number ($n \approx 30$) of subjects than a higher-tier field trial because the objective is to identify critical issues with the device rather than subtle influences.

Despite the expectation of some degree of failure, a high level of confidence in the device is necessary to undertake a field evaluation. A proof-of-principle—demonstrations that the device works in a laboratory setting using clinically relevant samples or surrogates (e.g. serum spiked with an antigen)—provides a degree of confidence that, given proper settings and conditions, the device should work. Although initial validation on 30 or more samples is desirable to provide statistical power,[25] smaller sample sizes may still provide confidence in a result depending on the size of the effect being detected and the sensitivity required. As a minimum, however, one should not use less than seven independent samples.[25]

It is important to make sure that experiments with surrogate samples are independent, not just replicates;[26,27] for example, spiking different levels of an antigen into aliquots of plasma from the same sample would not provide the same amount of variation in the background as spiking antigen into plasma from different samples. If using surrogate samples, it is important to understand limitations of the surrogate in the test that the device performs. Using a sample representative of the sample in the field provides the best test of a device. A device designed and tested only on blood from venipuncture may behave differently with fresh blood from a fingerprick. (Obtaining a reproducible and high-quality sample from a fingerprick requires care and technique to avoid hemolysis or inclusion of large volumes of interstitial fluid.[28])

Testing the device with naïve users—people who were not involved in the development of the device—can provide important information about the design and operability of the device. Not only does a naïve user offer feedback about the usability of the test, the performance of the test being run and interpreted by such a user provides a more realistic estimate of performance in the field than use and interpretation by the developers. With both of our devices, most testing before field work was done with some involvement of the developers of each technology. If naïve users had been introduced during testing in the laboratory, perhaps we could have reduced the time spent on pilot trials, or identified areas to improve on the devices before they were evaluated in the field.

Devices designed to give a binary readout (i.e. "positive" or "negative") must be sensitive (able to detect positives) as well as specific (able to avoid classifying negative samples as positive). If a threshold is used to define whether a measurement is classified as positive or negative, a receiver operating characteristic curve provides a visual tool to understand the performance of the test.[29] If the device provides a quantitative measurement, comparison to measurements from a standard diagnostic test using a Bland–Altman plot can identify potential bias in the measurement.[30]

With confidence that the biological and technical side of the device can work, development efforts should focus on reducing the sources of confounding factors. In preparation for field evaluations, one needs to develop quality controls, identify suitable packaging, set storage requirements, and set shipping methods. Packaging devices, and then storing them in an oven or a high humidity environment, can provide quick tests for stability under extreme storage conditions. Packing some devices and sending them by a courier with a return service provides exposure to different shipping environments. The degree to which all these factors are understood and accounted for sets the level of confidence that a device will perform as expected in a clinical trial.

Field evaluations conducted during this phase of development, before the final design has been frozen, provide an important opportunity to test a device on clinically relevant samples, and to identify unexpected problems, before conducting a field trial on the level of the third tier of field work (Figure 2), such as a trial for regulatory approval. In both of the case studies that follow, the design was not finalized, but work to test for shipping, stability, and quality control had been done to different degrees. Table 2 provides a suggested timeline for development and field evaluation.

A field trial at or above Tier 3 requires high confidence that the device will work, and a fixed design (including packaging and storage conditions). This level of confidence

Table 2: Representative timeline for moving a device from the laboratory to a field evaluation.

Task	Timeline - 5 Year Plan: 1	2	3	4	5
1. Defining the Problem					
a) Team Building					
b) Needs Assessment					
2. Device Design & Testing					
a) Prototyping					
b) End User Feedback					
c) Validation on Clinical Samples					
d) Quality Control/Stability/Storage					
3. Supporting Activities					
a) Grant Writing					
b) IRB Approvals					
c) Institutional Agreements					
3. Field Evaluation					
a) Trial Design					
b) Shipment/Purchasing of Supplies					
c) Training					
d) Pilot Phase					
e) Full Study					
4. Next steps					
a) Analysis of Results					
b) Publication					
c) Discussion with Companies					

requires demonstration of diagnostic accuracy on larger numbers of clinically relevant samples ($n > 30$) on devices produced in different lots. Multiple users should both perform and interpret tests. The intention of a field trial at Tier 3 is not to troubleshoot, but to demonstrate the validity of the rapid test (often as a means towards regulatory approval). A field trial above Tier 4 requires a device that has passed a field trial at Tier 3 and is essentially in the form of a product. At Tier 4, field work aims to establish whether the use of the device at the POC provides a significant health benefit. For Tier 3 and above, the support of a company is often critical to bring manufacturing standards and scale to validation and efficacy testing.

2.6. Ethical Considerations

Research involving human subjects generally requires the approval of an Institutional Review Board (IRB). IRBs are committees common at universities and hospitals; they review all research proposals involving human subjects to ensure that studies are designed ethically and participants are properly informed and protected. (An important exception relevant to diagnostics is the use of existing samples that are either publicly available or obtained in such a manner that subjects are unidentifiable.) In any case, researchers involved in a field evaluation should complete training on human-subject research. The IRB review process, while sometimes cumbersome, is essential to protect the participants in a trial from physical or emotional harm. In fact, if approached properly, IRB committees can provide invaluable guidance to ensure the ethical and proper collection of data. Often, these committees are knowledgeable about regulatory requirements, and they can provide advice to ensure that the study is designed in a manner appropriate for regulatory approval. In general, field trials carried out abroad must be approved both by an IRB in the country of the trial—for which approval working with the local Ministry of Health may be necessary—and by a separate IRB in the country where the research project originates. Each IRB committee may have different requirements and standards; the process of reconciling these differences can take months.

2.7. Designing a Study

Once the prototype is ready to be tested in the field at or above Tier 2 (Evaluation), the design of the study becomes critical. One must lay out clear goals for the study. The entire team should know which tier of field work is expected because, as discussed earlier, the objectives and requirements for each tier differ significantly. During the design process, an institutional agreement between the research institution and the site of the field work should be established. This agreement can take the form of a subcontract or a memorandum of understanding. Clear, explicit expectations of work and commitments should be laid out, including precise language about financial commitments and oversight.

During a field evaluation, one may want to take a device to rural clinics to obtain feedback about the operation, interpretation, and design of the device. Such field work falls in Tier 1, and has different requirements for partners in LMICs and for the design of the study. A field evaluation, and an evaluation of usability at the POC, can be done concurrently, but each requires a specific set of goals and objectives. In some cases, they may each require separate IRB approvals.

Field evaluations past Tier 2, and before final efficacy testing (Tier 4), usually include checking the performance of the device on clinically relevant samples. Estimating performance requires comparison of measurements from the device to a "gold-standard test" (i.e. a widely accepted, standard clinical test). Often, these tests are not available at the POC. The requirement to compare results to a gold standard may mean that initial trials must be done in a regional hospital in-country, where the necessary equipment is available to perform a gold-standard test, rather than at rural clinics. If one aims to characterize device performance, it is essential to follow best practices[31,32] and sound statistics[33] (e.g. ensure adequate sample sizes for statistical power,[34] proper blinding of samples,[31] and well-defined inclusion and exclusion criteria[31]).

Recruiting and training staff to carry out the study can require weeks to months; the investment in time used to find qualified and committed staff members will allow a study to be more robust against unexpected external events. Training requires clear instructions on recruitment, workflow, sample collection and distribution, performing the rapid test, performing the standard test, and recording results. Organizations (Table 1) can sometimes help expedite this process at established sites. In addition to time for training, time should be set aside for a pilot study before the main study begins. A pilot phase allows logistical problems to be identified and remedied without compromising the quality of the data from the trial.

Instruments used for the collection of data (i.e. questionnaires and laboratory protocols) should include as much information as possible without becoming cumbersome. Moving from the laboratory to the field introduces many variables; unless these externalities have been tracked, the results of the study may be difficult to interpret. Events that could affect a device occur from the moment devices are prepared in the laboratory to the time that they are used, but some factors are not immediately apparent. For example, including a temperature logger when devices are shipped is often overlooked (see Sections 3.7, 4.7, and 4.10).

2.8. Context and Culture

Establishing a working relationship with partners overseas early in the development of a device reduces the risk of a misunderstanding later, during the implementation of a field trial or evaluation. Partners may not have protected research time, especially in clinical settings, and their work may prevent them from devoting as much time to run the study as needed. In such cases, it may be appropriate to hire a dedicated study coordinator for the project. Different countries have different hierarchical structures. Understanding the local culture can prevent a social faux-pas that can undermine a study or endanger a partnership.

2.9. Challenges

With all the complications of carrying out an international collaboration and field evaluation, unexpected challenges will inevitably arise. Strikes, natural disasters, and regional instability are just a few examples of unrelated events that can threaten a project. The project must have timelines, but also flexibility. Each obstacle taken in stride is an indication that the device is making its way toward something useful.

3. Case Study 1: Liver Function Test

3.1. The Problem

Efforts to combat HIV have enabled access to antiretroviral therapy (ART) in LMICs. As of 2013, over 10 million people were receiving ART.[35] Drug-induced liver injury (DILI) is a significant side effect associated with ART. Nevirapine-based ART—widely used in the developing world—is of particular concern, with rates of hepatotoxicity (a type of liver damage) exceeding 13 %.[36,37] Monitoring liver function provides an important tool to manage ART;[37,38] dosages and treatments can be adjusted if signs of liver damage or hepatitis appear. Tests to monitor liver function, however, are often unavailable in low-resource settings where many patients with HIV receive care. Levels of serum transaminases (aspartate aminotransferase, AST, and alanine aminotrasferase, ALT) provide a standard for monitoring DILI, but generally require centralized labs and venipuncture.[39]

3.2. The Test

AST and ALT are concentrated in the liver in hepatocytes, and usually are present only in low concentrations (5–40 units L^{-1} for AST and 5–35 units L^{-1} for ALT) in serum. If injury, toxicity, or inflammation damage the hepatocytes, they will secrete larger concentrations of these transaminases into the serum. The ratio of AST to ALT can be helpful in assessing the cause of liver injury; alcohol-related liver injury, for example, often results in a greater than 2:1 ratio of AST:ALT.[39] By choosing enzymes as markers for liver injury—in this case transaminases—we could create a colorimetric assay by providing a substrate specific to each enzyme that would result in a change in color. We describe the chemistry of the detection for AST and ALT in detail elsewhere.[8,9]

3.3. The Team

We built a team combining academia, industry, and medicine. A business plan competition at the Harvard Business School brought together business students and scientists from the Whitesides group. The Whitesides group had developed the initial idea of 3D paper microfluidic devices that formed the technical starting point for the development of the test for liver function.[7,9,40–42]

Diagnostics for All Inc. (DFA) was formed as a nonprofit (501-c-3) engineering organization to develop paper-based devices—or indeed, *any* type of device that suits a task—as diagnostics. Academic groups excel at basic research and innovation, but usually lack the proper resources, incentives, motivation, and experience to do the detail-oriented engineering required to execute trials, regulatory clearance, manufacturing, and quality-control for a product. DFA provided a vehicle to do high-quality engineering. The decision to create DFA as a nonprofit company stemmed from the idea that a for-profit entity driven by investors interested in financial return might be forced to focus on developed-world applications of paper-based diagnostics rather than first to apply the technology to address needs in the developing world. Whether a for-profit or not-for-profit model ultimately is more effective for POC diagnostics intended for resource-limited environments remains to be

seen. The development of a test for HIV by Sia (at Columbia) and Linder (at Claros, now OPKO) provides an alternative example.[43] Chin, Linder, and Sia also provide a helpful review of many of the companies working to commercialize POC diagnostics and their different funding structures.[4]

The team identified the need for a POC liver function test early on, and discussions with doctors in Boston-area hospitals and experts in public health confirmed the importance of the problem, especially in countries with large populations of patients being treated for HIV. After initially considering seven different assays, the team settled on ALT and AST.

DFA developed a robust assay and integrated sample acquisition, preparation, and evaluation into a single device. Dr. Nira Pollock at Beth Israel Deaconess Medical Center (BIDMC) provided medical guidance and led clinical validation in Boston. She identified a site in Vietnam to evaluate the test in a field setting. After initial validation, DFA called Bernhard Weigl, the director of NIH-funded Center for Point-of-Care Testing at the Program for Affordable Healthcare Technology (PATH), for assistance in running a field evaluation. (We mention by name some of the specific individuals who were particularly important in developing this test to emphasize the number and variety of people, skills, and connections necessary to accomplish field testing.)

3.4. Designing a Solution

To monitor DILI at the POC, we created a low-cost, rapid, liver function test (LFT) for serum transaminases using paper-based microfluidics.[8,9] Paper provides an inexpensive substrate for bioassays. Patterning paper enables multiplexed flow and small test zones; the latter reduce the need for large volumes of expensive reagents. Paper-based microfluidic devices provide an attractive system to develop POC tests for use where low cost is a key design criterion. Tests that are performed frequently, and potentially at high volume (e.g. tests to monitor liver function), are particularly sensitive to cost.

The specific problem of measuring liver function provided guidance in the design of the test. Although standard tests for serum transaminases provide a quantitative measurement, these measurements are generally interpreted in three bins. A rapid test needs to provide a semiquantitative readout that will allow results to be placed into the three bins, but further precision is unnecessary in practice. We, thus, made a colorimetric test with a read-guide (a standard color bar) that would allow users to provide a semiquantitative measurement of each test (Figure 3). Benchmark values for ALT and AST are based on measurements in blood serum. To separate serum (needed for the assay) from whole blood, we incorporated a plasma-separation membrane into the paper device.

With end users at the POC in mind, the device was designed to be as simple to use as possible, and interpretable by eye. The device was laminated in plastic to protect the paper test-zones from the outside environment, and a small hole in the lamination over the plasma-separation membrane provided an entry port for blood wicked directly from a fingerprick. Positive and negative controls built into the device provided an indicator of the validity of results (e.g. reagents are working, blood has not lysed). Several iterations of design incorporated improvements to the ease-of-use, sensitivity, and interpretation of the test (Figure 4).

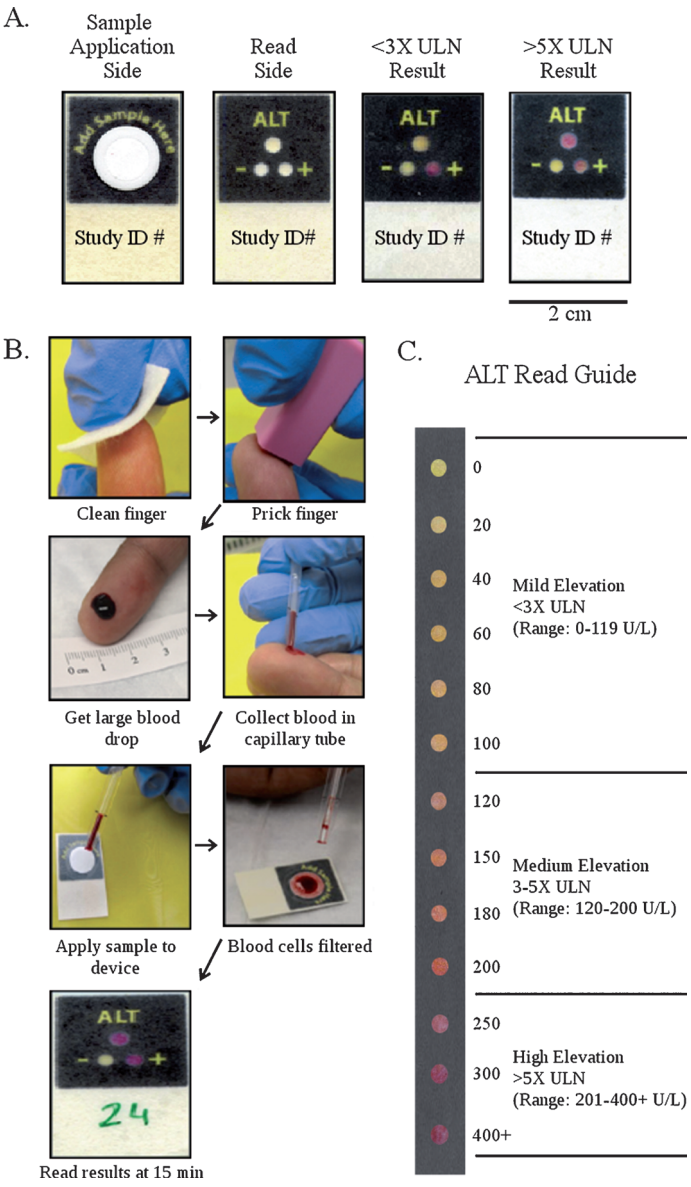


Figure 3. The design of a paper-based liver function test provides a semi-quantitative test for serum transaminases. A) The stamp-sized devices receive a sample of blood on the "application side" and provide a colorimetric readout on the "read side". B) The entire process of running the test requires minimal sample manipulation. C) Valid results are interpreted and binned into three levels based on elevation over the upper limit of normal (ULN) using a read guide.

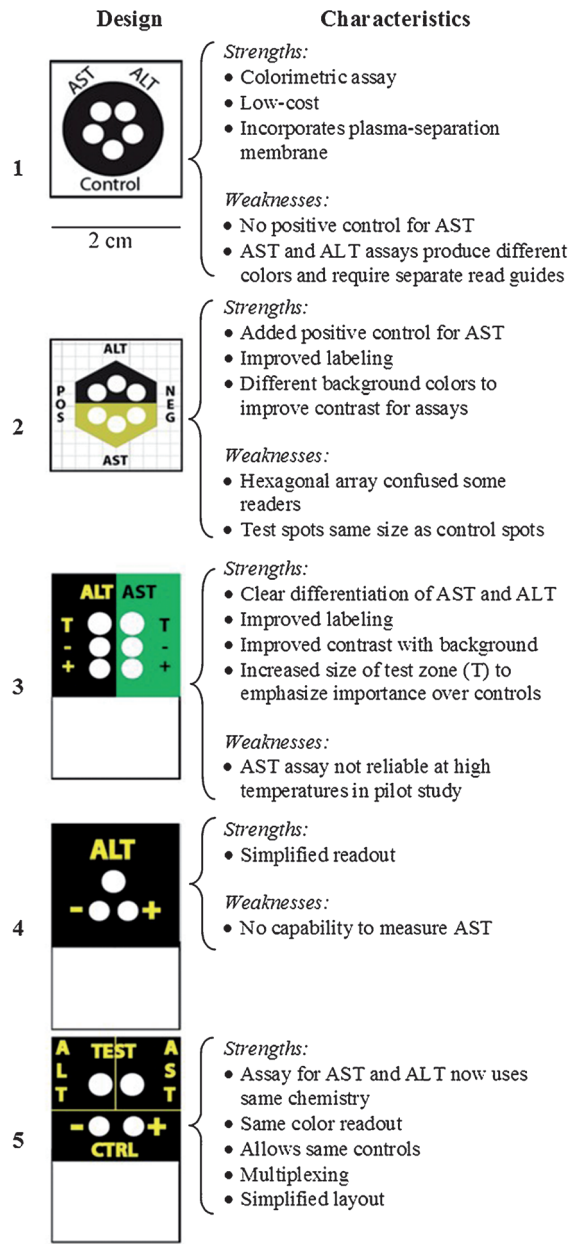


Figure 4. Iterations on the design of a rapid liver function test (1–5). Each design has the same scale. The characteristics of each iteration demonstrate improvements to the design to create a user-friendly device.

3.5. Validation and Preparation for Field Evaluation

We tested analytical, operational, and clinical performance of the LFT device.[8,9] Briefly, these studies included defining the limits of detection, assessing repeatability, checking for cross-reactivity and interference, optimizing the time for the assay, designing methods for metering the sample, comparing the performance of the device to gold-standard methods, and testing the environmental stability of the device. These studies were intended to validate the device and assess when to initiate field testing.

During this time, members of the team from DFA traveled internationally for conferences and meetings with potential end users in different healthcare settings. A trip to India in 2010 allowed a DFA member to show the design of the LFT at the time (Figure 4: Iteration 2) to clinical laboratory workers. Workers at the clinic felt that the test was too small, and demonstrated that ceiling fans in the hot environment had the potential to accelerate evaporation and change the dynamics of the test. Later iterations of the device increased the size and added a small white tab to make the test easier to hold (Figure 4**:** Iteration 3 onward). Protocols for running the test were modified to include placing the tests under a glass dish to minimize evaporation.

After three years of development, the team decided to perform a field evaluation to produce high-quality data for use in refining the device. Field evaluation would also assist in freezing aspects of the design that worked well; this step would then enable future trials for regulatory validation. Most importantly, we hoped the field study would provide a test to see how the device performed in a setting where it could have an impact on healthcare.

3.6. Funding the Trial

The initial work to develop the test on paper-based microfluidics was supported by a grant from the BMGF. This grant also enabled some early travel to countries where a large number of patients on ART could benefit from a test to monitor DILI. Dr. Pollock was supported by an NIH K23 grant to carry out the clinical validation in Boston. By working with PATH, we were able to fund the field evaluation through their NIH-funded Center to Advance Point-of-Care Diagnostics for Global Health.

3.7. Study Design

To obtain sufficient data, we wanted to enroll 600 patients in a controlled field setting. Working with BIDMC, we sought an international partner who worked with a large number of patients on ART and who would be able to benefit from a POC LFT. The partner also needed to have the infrastructure in place to measure serum transaminases as a gold-standard comparison to our test. BIDMC had a long-standing partnership with hospitals in Vietnam through the Harvard AIDS Initiative in Vietnam. Dr. Pollock connected with Dr. Donn Colby who had spent years living and working with hospitals in Vietnam as part of this program. Dr. Colby provided an essential link between the medical team developing the device and the local site in Vietnam. With Dr. Colby's assistance, the Hospital for Tropical Diseases (HTD) in Ho Chi Minh City was identified as a site for testing.

The HIV clinic at the HTD provided the right combination of appropriate patient population and supporting infrastructure. The clinic we chose saw 3000 HIV-positive patients per year who received ART. Of these, a significant proportion were on nevirapine-based ART (known to confer risk of DILI) or were at risk for co-infection with hepatitis B (HBV, 15% prevalence) and/or hepatitis C (HCV, 25% prevalence)—both of which can cause damage to the liver and

could benefit from monitoring liver function. The clinic also had an existing practice of routine transaminase monitoring (once every six months) for patients receiving HIV treatment. As a result, the clinic had capabilities to do standard tests, but also could benefit from a POC LFT for more frequent and less-expensive monitoring.

The study recruited and consented patients scheduled by their physicians for routine clinical ALT testing (specimen collected by venipuncture); after venipuncture, subjects proceeded to fingerstick collection for LFT testing. The team chose a target population of adults since this population would be one of the largest to benefit from the LFT and would be able to provide informed consent themselves (an important consideration to simplify IRB approval and the consenting process). The Vietnam field study was approved by the Institutional Review Boards of HTD as well as by Research Ethics Committees at PATH and BIDMC.

Working with HTD rather than a rural clinic enabled validation of results using standard tests. Resources at HTD (Roche Cobas 6000 analyzer) allowed automated testing of ALT. The study collected other clinical information, as available, that was relevant to the performance of the device: HBV status, HCV status, current HIV medications, current tuberculosis medications, and most recent CD4 count. Results of any laboratory tests ordered concurrently with ALT on the day of enrollment were also captured (e.g. AST, hemoglobin, hematocrit, platelet count, and creatinine).

Outside of clinical information, we also recorded other variables that could influence the performance of the LFT. Packaged LFTs were stored in ambient conditions, and, thus, a temperature and humidity logger, combined with historical weather data from Weather Underground (http://www.wunderground.com) provided a record of ambient environmental conditions. Initially, no data loggers were included during the shipment of LFTs to Vietnam, but the team added data loggers to the second batch.

Three activities occurred at the site of the study: 1) training, 2) pilot phase, and 3) study phase. The training was meant to familiarize all staff with the study and establish competency reading the LFT. The pilot phase was designed to recruit 50 subjects and ensure that study procedures worked as expected before beginning the study phase and recruiting 600 participants over six months.

3.8. Implementation

Approximately one month prior to the beginning of the study, representatives from PATH and DFA traveled to HTD to train nurses assigned to carry out the study. The training curriculum included the objective of the study, recruitment procedures, overview of the device function, steps for completing the fingerstick and transferring the sample to the device, and practice readings with mock devices. Nurses were specifically instructed to read and record device results privately, without interaction with any other individual. The nurses were required to pass a proficiency test using the mock devices (pass criteria: $>80\%$ bin placement accuracy and 100% determination of invalid tests) before patient enrollment could start. If they failed this test, they were retrained with the mock devices and given another test. Each nurse was allowed a maximum of two attempts to pass the test. During the pilot phase, the study nurses received immediate feedback on correct and incorrect use (including fingerstick, sample transfer procedure, and device reading) from an expert DFA representative. No additional training or feedback was given once enrollment for the study began.

The temperatures during the pilot phase were often higher than the range for which the LFT was designed (4–30 °C). The read-time for the test was adjusted for the higher temperature and the AST test was removed from the device because of the poor stability of reagents for that test at higher temperatures (current improvements in the AST test have addressed these issues and will be evaluated in future trials).

All tests were performed following a set of instructions provided with each product by DFA, the details of which are described elsewhere.[10] Notably, each test was read by two nurses. Neither the patients nor their doctors were informed of the results of their fingerstick testing. Although the field work took only six months, the total time to plan for the study, obtain IRB approval, and carry out the work at the field site was greater than a year.

3.9. Context and Culture

Working with multiple partners—academics (Harvard and BIDMC), a nonprofit company (DFA), an NGO (PATH), and a hospital (HTD)—provided the benefit of specialized knowledge and experience from each partner. Managing such a multi-institutional collaboration, however, also has challenges. Each partner has its own needs and its own timeframe. Academics want to publish results, hospitals need to maintain sufficient staff for their primary operations outside the study, companies want to learn as much as possible about their test for use in development, and NGOs need to ensure that the device is at a sufficiently mature stage to merit investing time and resources into a trial. In the case of the field trial in Vietnam, we believe everyone met their needs effectively, but this success required clear, open, and frequent communication at the outset, and throughout, the study. For example, DFA and PATH had wanted to have ten nurses reading tests independently to estimate concordance for the visual test. HTD was not willing to provide that many nurses to the study on the basis of the resources they had available. A compromise was reached in which HTD provided three nurses for the study. Rather than DFA or PATH hiring nurses, HTD received funds as a contractor and assigned nurses to the study.

Every healthcare system operates differently, and differences in norms have the potential to disrupt a study. At HTD, nurses rotate their positions every six months. The team from PATH and DFA arrived and began training nurses during the pilot phase in the middle of a rotation cycle; all the trained staff would have been rotated out of the clinic halfway through the study and could have caused major inconsistencies. The practice of rotating staff came to the attention of the PATH trainers, fortunately, during the pilot phase. Working

with HTD, they were able to negotiate for some of the key trained personnel to remain at the clinic for the duration of the trial.

Sometimes patients may be reluctant to participate in a study when there is no direct compensation. Working with an HIV clinic, where patients were already being monitored regularly for DILI, minimized the need to explain the potential future benefit of a rapid LFT.

3.10. Challenges

The study identified a number of parameters of the LFT that could be improved. Some involved ease-of-use (for example, the training required for accurate bin placement). Others related to quality control in components (a number of tests were invalid because a batch of plasma-separation membranes was faulty). Even though DFA tested the influence of environmental conditions (humidity and temperature) on shelf-life, the results from the study indicated that stability (of reagents, surprisingly, rather than enzymes) required improvement. Temperatures at the study site reached 36 °C during the course of the study. These temperatures were higher than those recommended for storage.

At the beginning of the study, nurses felt that devices were difficult to read; by the end of the study, with familiarity, they reported that tests were easy to read 90 % of the time, and that the instructions provided with the test were "very clear and easy to follow." The most frequently mentioned challenge to using the test was matching the color on the device to the color bar in the guide for the reader (Figure 3).

3.11. Lessons Learned

Field evaluation of tests identified several aspects that required improvement. For example, improving bin placement required expanding the dynamic range of the assay to give clearer visual differences in color between different levels of ALT. Minimizing invalid tests required increased quality testing of the commercial materials used in the assembly of the LFT, and tightening the requirements for the controlled manufacturing environment to reduce lot-to-lot variability. Stabilization of the reagents stored on the device was also required to improve shelf-life.

Like many rapid tests, the LFT requires that results be read within a specific window of time (after 12–14 min). Nurses with other duties and patients may have difficulty reading tests in this window. Several nurses in the study expressed the understandable desire that tests should be less sensitive to time; ideally, they wanted tests that could be read at any time over an interval of several hours after the assay had completed.

Although POC diagnostics using paper-based microfluidic devices have been explored extensively in academic laboratories, this study was the first large-scale field evaluation of a device of this type. As such, it provided evidence that paper-based devices could be used practically in LMICs. In particular, it established that healthcare workers performing tests on samples from a clinically relevant population agreed well in their evaluation of visual readings. Successful development of this device for clinical use will, nonetheless, require further, iterative, optimization. These initial field evaluations provided the guidance necessary for further development.

3.12. Next Steps

The newest prototypes have better accuracy, less sensitivity to environment conditions, and more importantly, a shelf life of greater than a year at 20–30 °C. Most of the issues identified in Vietnam have been addressed. We are currently preparing for field testing of the new prototype and beginning the process of validation of the product for regulatory approval (CE mark).

The intensive training of users allowed in the context of this first field study is unlikely to be feasible outside of study contexts. A thorough understanding of the minimal training requirements for novice users will ultimately be key to understanding the range of clinical environments in which this test can be used—whether that be in centralized clinics with trained staff, decentralized clinical settings with minimally trained health-care workers, or even at home. DFA is currently planning an "Untrained User Study" with 50–100 participants to obtain Clinical Laboratory Improvement Amendment-waived certification for the LFT device.

4. Case Study 2: Sickle Cell Diagnostic Test

4.1. Problem

Sickle cell disease (SCD) is an illness where an early diagnosis can have a major impact on health outcomes. Each year, over 300 000 children are born with SCD, most in sub-Saharan Africa and India.[44] In countries without early diagnosis and treatment, the mortality rates of children under 5 years old with the disease are 50–90 %.[45] Interventions as simple as prophylactic penicillin and parental education can have a significant impact on child survival rates.[46] The lack of a low-cost, rapid, POC test for SCD, however, means that the potential of these interventions goes largely unrealized.

4.2. The Test

The distribution of densities of red blood cells in SCD is heterogeneous compared to red blood cells from a healthy individual.[47,48] In SCD, sickling and dehydration leads to the formation of cells ($\rho > 1.12\ g\,cm^{-3}$) that are denser than the most dense red blood cells in a healthy individual ($\rho \approx 1.10\ g\,cm^{-3}$).[47–50] We developed aqueous multiphase systems (AMPS)—mixtures of polymers in water that spontaneously separate into immiscible liquid phases—to separate and visually identify the presence of dense cells characteristic of SCD.[12,13] The densities of the phases of AMPS provide a step-gradient in density. We designed a two-phase system

($\rho_{top} = 1.078\ g\,cm^{-3}$; $\rho_{bot} = 1.129\ g\,cm^{-3}$) and a three-phase system ($\rho_{top} = 1.077\ g\,cm^{-3}$; $\rho_{mid} = 1.108\ g\,cm^{-3}$; $\rho_{bot} = 1.120\ g\,cm^{-3}$) such that only dense cells characteristic of SCD would be able to sink through the bottom (most dense) phase and form a visible red layer at the interface between the bottom phase and the seal at the bottom of the container. Details of the test have been summarized elsewhere.[12,13]

4.3. The Team

The Whitesides group had developed AMPS.[12,51] A chance encounter with Dr. Thomas Stossel (Brigham and Women's Hospital, Boston) provided the incentive to think seriously about SCD. Dr. Stossel had spent the better part of a decade doing medical work in rural Zambia. As a hematologist, he knew the burden of undiagnosed patients with SCD in that country. He also understood the constraints that a test would have to meet to be useful in rural areas.

Once the team realized the potential impact of a density-based assay to identify sickle cell disease, we consulted Dr. Carlo Brugnara (Children's Hospital, Boston) as an expert on the density of red blood cells in SCD.[52,53] Dr. Stossel connected the team to Dr. Catherine Chunda-Liyoka, a physician at the University Teaching Hospital (UTH) in Lusaka, Zambia. Dr. Chunda-Liyoka complemented Dr. Stossel's knowledge of the needs for sickle cell diagnostics with her experience managing SCD in patients in Zambia.

The team moved from a conceptual idea to an evaluation in the field in Zambia in three years (Table 3). The speed of development was driven in part by the simplicity of the technology, and by the ability of team members to work together efficiently in the design and implementation of the field work.

Table 3: Timeline from the conceptualization of a density-based rapid test for sickle cell disease to a field evaluation in Zambia.

Task	Timeline: 2011	2012	2013	2014
1. Defining the Problem	■	■	■	■
a) Team Building	▒	▒	▒	
b) Needs Assessment	▒	▒	▒	
2. Device Design & Testing	■	■	■	■
a) Prototyping		▒	▒	
b) End User Feedback			▒	▒
c) Validation on Clinical Samples		▒	▒	
d) Quality Control/Stability/Storage		▒	▒	
3. Supporting Activities	■	■	■	■
a) Grant Writing		▒	▒	
b) IRB Approvals		▒	▒	
c) Institutional Agreements			▒	
3. Field Evaluation	■	■	■	■
a) Trial Design		▒	▒	
b) Shipment/Purchasing of Supplies		▒	▒	
c) Training			▒	
d) Pilot Phase			▒	
e) Full Study			▒	
4. Next steps	■	■	■	■
a) Analysis of Results				▒
b) Publication				▒
c) Discussion with Companies			▒	▒

4.4. Design considerations

Using AMPS, we generated thermodynamically stable, step-gradients in density to separate dense cells present in SCD, and provide a visual test.[12] The step-gradients that form in AMPS on settling in gravity, or—more rapidly—centrifugation, allowed us to make large batches of these mixtures of polymers and preload them into microhematocrit capillary tubes. The density of red blood cells can change in response to the characteristics of the medium (e.g. pH and osmolality). We, thus, designed our initial tests to match physiological pH (7.40 ± 0.02) and osmolality ($295 \pm 15\ mOsm\,kg^{-1}$).[54] (Other values of these parameters may also be useful for a rapid test. Changing either value may require adjusting the densities of the phases to reflect the shifts in density that can occur in red blood cells.)

Discussions with Dr. Stossel and Dr. Chunda-Liyoka indicated that it was important to minimize the requirements for power and time for the test, so that it could be run in rural clinics. Centrifugation is necessary for cells to move through an AMPS in the short time (minutes rather than hours) required. The faster the centrifuge, the more the power that is required, but the shorter the time that is needed for the test. We concluded that shortening time was more critical than reducing the use of electricity.

We could run the AMPS-based test in less than 15 min using a centrifuge capable of providing 13700 *g*; such a centrifuge requires power. For district and provincial hospitals in Zambia, power is accessed from the grid. Many Zambians, however, receive their care from rural clinics. In rural clinics and villages, one can often find solar panels charging car batteries. We used a DC-to-DC adapter to power the centrifuge with a car battery.

The entire process of the interaction of a patient with a diagnostic—from sample acquisition to reading the results—has to be carefully designed for ease of use and safety (Figure 5). We used polycarbonate capillary tubes, rather than glass tubes to avoided breakage, and the potential for puncture wounds. We designed the test so that tubes would be preloaded with solutions of polymers. Taking a sample of blood from a fingerstick and transferring it simply and practically, into a device is nontrivial. We iterated several designs (see Figure S1 in the Supporting Information) before finding a method—using capillary action—to add

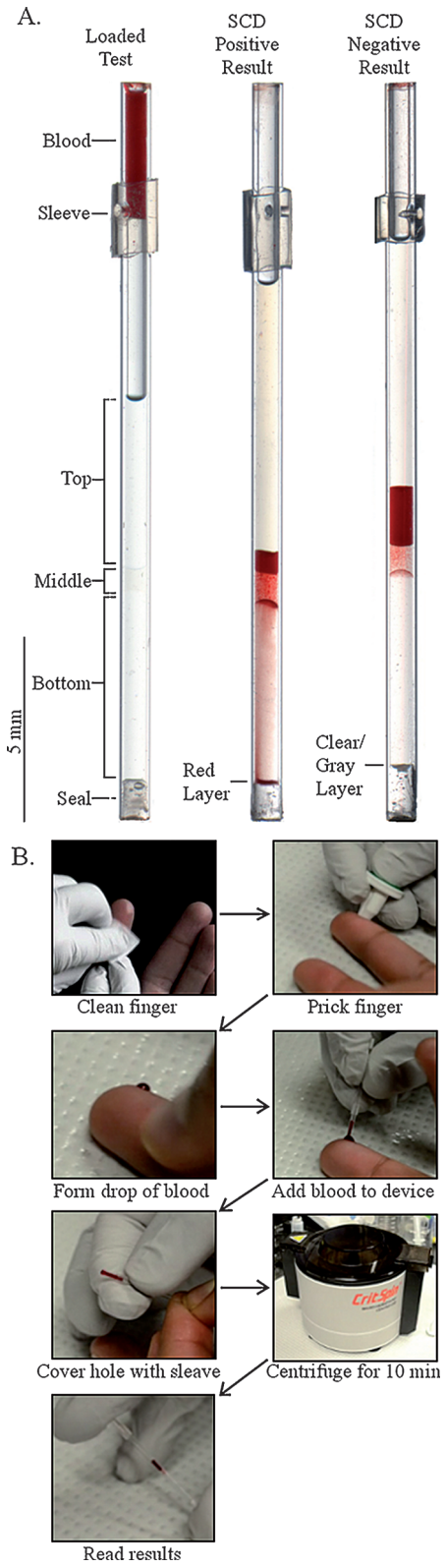


Figure 5. AMPS enable a simple, rapid test for SCD using density. A) A capillary preloaded with AMPS wicks a fixed volume of blood into the device. After covering the hole for filling the tube with a sleeve, centrifugation enables the rapid sedimentation of cells over the AMPS and provides a result that can be read by eye. A red layer above the seal at the bottom of the tube indicates the presence of SCD. B) The entire process of performing a test, including centrifugation, requires about 12 min.

a fixed volume of blood to a capillary that was already sealed on one end and partially filled with an AMPS. Visual readout removed the need for additional equipment or computing power. Including aspects of sample acquisition and readout into the design of the device ensured a smooth interaction with an end user: all that a user would have to do would be to load a drop of blood, and spin the small test in the centrifuge.

4.5. Validation Confirmed that the Technology Was Ready for Field Trials

Blood samples provided by our collaborators at Children's Hospital Boston allowed us to make rapid improvements on early iterations of the density-based test. Once the densities of the AMPS were identified that discriminated between normal blood, and blood from individuals with sickle cell disease, the design was frozen and testing was carried out on a larger number of samples ($n = 59$). To achieve this sample size, we collaborated with Prof. Sergey Shevkoplyas, then at Tulane University, and Dr. Julie Kanter, then at the Sickle Cell Center of Southern Louisiana.

For the initial proof-of-principle, we included the main two genotypes of SCD (HbSS and HbSC) and also tested sickle cell trait (HbAS) along with other non-sickle cell subjects (HbAA). Initial results showed a promising ability to discriminate between SCD and non-SCD with only 10 min of centrifugation and a drop (ca. 5 μL) of blood. To enable the test to leave the bench, we also began work on packaging and storage. We used various packaging materials and accelerated storage tests in an oven to assess the risk for evaporation before settling on a packaging system for the evaluation in Zambia.[13]

4.6. How Funding Was Obtained

Although the clinical need for a rapid test for SCD was clear, the way to fund such research was not. In general, SCD suffers from a low level of funding, as is often the case with neglected tropical and orphan diseases. We began applying to a number of innovation awards focused on global health. One difficulty with these grants was pressure to have data from the field to justify the award. Research to provide such data, however, required funding. This "chicken-and-egg" scenario is characteristic of projects in translational medicine.

To obtain our initial funding for SCD, we had to think of other applications for the AMPS technology that would be more attractive to granting agencies. By proposing to use AMPS both to diagnose SCD and to also explore circulating tumor cells, we were able to put together a successful proposal to the Blavatnik Biomedical Accelerator Fund. This award supported the development of the test and the field trial at UTH. A smaller award from the Harvard Global Health Institute enabled us to perform an evaluation of the usability of the test with end users in rural clinics in Zambia.

Data from our field trial aided us in securing funding for a second field trial from the Consortium for Affordable

Medical Technologies, a part of Massachusetts General Hospital's Center for Global Health.

4.7. Designing the Study

Dr. Stossel's connection to Zambia was one of the motivations for beginning work on a SCD diagnostic and, thus, work in Zambia was a natural choice for a field evaluation. The team from Harvard connected with Dr. Chunda-Liyoka at UTH and began to draft grant applications, study protocols, and IRB submissions.

The site for the field evaluation was the Department of Paediatrics and the Department of Haematology at UTH in Lusaka, Zambia. UTH could support equipment to do a gold-standard measurement of SCD (i.e. hemoglobin electrophoresis) and also had a large population of SCD patients. UTH regularly monitors patients with SCD through a program that includes regular evaluation and follow-up. As of 2014, UTH managed over 3000 patients with SCD, and provided follow-up to roughly 50 of these patients every Friday.

The primary relevance of a rapid test for SCD is for use in children, and hence, the study only enrolled subjects under 18 years of age. Working with children required consent of a guardian as well as assent for children above a certain age. Literacy and language differences added to the requirements of the consenting process to ensure adequate protection of subjects. Working in close consultation with both the local IRB in Zambia and the Committee on the Use of Human Subjects at Harvard was critical to ensure that all aspects of the ethics of the research were considered.

SCD screening at UTH relied primarily on solubility tests. Positive tests were followed up by gel electrophoresis. The electrophoresis equipment, however, was quite old and unreliable. Part of the study budget, therefore, was allocated to update the clinical laboratory at UTH with a semi-automated hemoglobin electrophoresis unit (SAS1/2, Helena). In particular, this system allowed quantitative measurements of different hemoglobin types, including fetal hemoglobin (HbF)—a parameter of interest for the density-based test. We also collected routine complete blood count (CBC) information on all subjects in the study, and used a questionnaire to gather basic demographic information, and information about factors that might constitute a confounding factor for our method (i.e. recent sickle crisis).

Outside of the medical data, we also designed methods to record tests that had packaging failures, temperature on the days testing was performed, and the time between the blood draw and the use of rapid and gold-standard tests. Guidelines for the use of samples for various tests (CBC, hemoglobin electrophoresis, and the rapid test) were decided based on guidelines from the manufacturers or, in the case of the rapid test, experience from initial validation. Samples that were processed with any method outside the recommended time-line would be invalid. One variable that was overlooked in the design phase of this trial was the inclusion of temperature and humidity loggers during shipment.

Through the Harvard Catalyst program, the team consulted with a biostatistician to design the size of the cohorts to be recruited in the study. The trial was designed to recruit about 600 subjects over six months. The process of designing the study, obtaining IRB approval, and installing the new hemoglobin electrophoresis equipment took over nine months.

In addition to the evaluation of the performance of the test, we also wanted to evaluate the test in rural clinics for ease-of-use. Working with UTH and the US Peace Corps in Zambia, we identified rural clinics to visit and designed a program to explain the test and receive feedback. The US-based IRB committee declared the survey to be exempt from full review as human subject research because of minimal risk and the nature of the information collected. The Zambian-based IRB committee, however, required a full review of the survey at the rural sites.

4.8. Implementation

A researcher from Harvard traveled to Lusaka at the beginning of the trial. Setting up the trial required a week. A full-day training of the study staff included the overall design of the study, workflow, recruitment, use of the rapid test, and management of data. Four readers were trained (two laboratory technicians and two nurses) using images of results of rapid tests and examples by an expert reader. A poster outlining each step of the use of the rapid test was placed prominently in the laboratory where rapid tests were run. A two-week pilot phase followed.

The pilot phase was critical for the success of the study. During this time, we evaluated initial concordance between the readers at UTH and the expert reader. We also identified and remedied potential problems with sample handling and workflow. For example, blood samples were collected in Vacutainers (Becton Dickinson) containing ethylenediaminetetraacetic acid (EDTA) as an anticoagulant. A fraction of each sample was transferred to a second tube. One tube of the sample went to the laboratory running hemoglobin electrophoresis while the other tube went to the laboratory running the rapid tests. Using separate laboratories aided with blinding the study. Initially, the second tube to which blood was added also contained EDTA. This additional EDTA could have caused dehydration of the cells and compromised the rapid test. This potential problem was identified during the pilot phase and all subsequent samples were aliquotted into untreated tubes after collection in anticoagulant-treated tubes.

Halfway through the study, the expert reader from Harvard returned to the study site and performed a blinded test for concordance with three of the local Zambian readers. Throughout the study, the team at Harvard made batches of AMPS solutions and assembled packages (by hand) with hundreds of rapid tests to ship to the study site.

4.9. Context and Culture

Recruitment of subjects with SCD was generally much easier than recruitment of subjects without SCD because

patients with the disease, and their parents, were more knowledgeable about the disease and the need for a rapid diagnostic test than the general population. Occasionally, parents chose not to participate in the study because there was concern about what could be done with the blood of their child. Some parents expressed a belief that a child has a finite amount of blood in their entire life and they were afraid that the child would not have blood left if they provided blood for the study in addition to a clinically indicated blood draw.

4.10. Challenges

Several unexpected obstacles threatened the completion of the study, but fortunately, all were overcome. At two points during the six-month trial, nurses at UTH went on strike. Although the nurses on the study did not strike, their workload for their nonstudy obligations increased. The continued recruitment of subjects during this time is a testament to the commitment of these nurses to the study.

Near the middle of the study, a major fire broke out in the Nairobi International Airport in Kenya. Although far from the study site, this airport was part of the delivery route for the batches of tests shipped to Zambia. The fire occurred just after one batch had been shipped overseas and, as a result, that batch took an extra week to arrive at UTH. This extra shipping time, and the much elevated temperatures encountered by the tests as they awaited shipment through the damaged airport, may have compromised the tests. The conditions during shipment, unfortunately, were not recorded. Indeed, the performance of the delayed batch was significantly worse than the other batches,[13] but without additional information, we could not justify exclusion from the data analysis. This experience demonstrates the importance of setting clear parameters for valid results, including shipping and storage conditions, along with the use of temperature and humidity loggers to capture the relevant data.

Although the nurses' strike did not halt the study, supply chain problems did cause a temporary pause. Initially, we purchased about 20% more electrophoresis gels than we believed would be required for the study. A fault in the heating unit in the electrophoresis instrument meant that only eight out of twelve lanes ran properly. Delays in getting a technician to fix the instrument meant that we ran out of gels before recruitment was complete. We suffered a one-month halt in the study while waiting for additional gels from the supplier. This delay prevented us from recruiting the original number of participants we had planned for, but we were still able to collect enough samples to retain statistical power before the budget for the study was spent. The dependence of the gold-standard device on maintenance and technical expertise highlighted the need for simplicity in the design of diagnostics for the POC.

4.11. Lessons Learned

The details of the performance of the device and user feedback are described in detail elsewhere.[13] Briefly, of the two systems tested, the best one had a diagnostic accuracy of 77%. This accuracy was an encouraging first step toward a useful test, but an increase in accuracy (to, perhaps, 90%) will still be required before the test becomes acceptable for clinical use. In general, false positives were more frequent than false negatives (more so than was found in the initial validation studies). Variation in performance between batches was fairly significant. We are actively working on improving quality controls and developing standards to use with the rapid tests to reduce the variability between batches. Concordance between readers was high, but could be improved by clearer guides for readers and more extensive training.

Apart from the technical knowledge gained from the field evaluation, we also gained significant contextual knowledge. The visit to the rural clinics was particularly informative. We were able to verify that appropriate interventions for SCD existed in rural clinics as well as off-grid access to power through car batteries charged by solar panels.

Perhaps one of the most interesting outcomes of the field evaluation came to light during exit interviews of the study staff in Zambia. A number of staff members commented—without any specific prompting—that they had a new perspective when thinking about developing technology for issues relevant to Zambia; if researchers abroad could come up with low-cost rapid tests, perhaps they—the Zambians themselves—could also come up with useful technologies. Although quantifying the impact of inspiration and modeling innovation is difficult, at best, the aspect of directly sharing knowledge and skills during such an international collaboration is a welcome side effect of co-creation and field evaluation.

4.12. Next Steps

The development of the next generation prototype of a rapid test for SCD is underway. Data from the field validation and initial work have enabled us to write new grants, and have sparked conversations with companies interested in developing the technology.

5. Conclusions and Recommendations

5.1. General Lessons Learned

In the two cases presented here, the technologies are not yet products, but field evaluation has significantly improved the POC devices, and has identified specific deficiencies to be addressed. Our experiences suggest five summary lessons:

1) *Begin planning and partnerships early.* Ideally, partnerships should be in place before applying for grants. Long-standing partnerships allow clear communications that enable a team to identify a problem, and also to work through challenges of performing a field trial.
2) *Get feedback early and often.* Even if a full field trial is not appropriate for a particular stage of development of a device, getting feedback about the design and use of

a device throughout development will improve its quality and the chances of success in a large field trial.

3) *Track every variable.* No matter how much validation is done, there will be variables that are not accounted for in the technical specifications of a device. Even if every variable is not controlled, most variables can be measured, and these data may prove to be the key to understanding the results from the field
4) *Aim for more subjects than the minimum needed.* Unexpected problems can pause or end a study prematurely. Understanding the statistical power of a study and the amount the sample size could be reduced before the study loses value is critical. Within reason, aiming for more subjects than needed provides flexibility with the timing of the study. In our experience, a buffer of 20–30% more subjects than required for the desired statistical power was appropriate.
5) *Have patience.* Taking the time to get IRB approval, defining every part of the protocol, working out the details of the supply chain, and doing a pilot study are all important steps to ensure that a field trial will be successful.

Above all, do not assume that the phase of invention and demonstration-of-principle in an academic laboratory is the hard, creative part, and the rest is technical detail. The field trials and optimization are certainly more expensive, organizationally difficult, and time consuming, and certainly not less demanding technically.

5.2. From the Bench to the Field, and the Field to the Shelf

Field evaluations are not only the purview of companies. Although a partnership between a not-for-profit company (DFA), with skill in bioengineering, and an academic group (the Whitesides group), more specialized in invention, developed the LFT, an NGO (PATH) led the field evaluation of the LFT device. An academic group (the Whitesides group), led the field evaluation of the AMPS-based rapid test for SCD with partners at UTH. The ability to demonstrate functionality (even imperfect functionality) in the field greatly reduces the risk in further development, and provides the groups involved the chance to obtain resources for further work (through grants, collaborations, or transitions of the technologies to larger organizations).

Many of the steps for the development of POC diagnostics for LMICs are also relevant in developed countries:

1) working with clinicians and end-users to identify a problem
2) obtaining IRB approval
3) validating the device on relevant samples with naïve users
4) comparing the performance to that of a gold standard
5) testing the stability of the device during storage and shipping

The constraints of cost and supporting technologies may be relaxed, but, as with POC diagnostics for LMICs, the requirements must be understood in the context of the specific use: consumers in the US may be able to afford a POC test priced ten times higher than would be affordable to consumers in India, but, in both settings, the test must be easy for consumers to use.

Someday, we hope that both the LFT and the SCD-AMPS will be available on the shelves of clinics as finished products, but successful field trials are only steps along that path. When the day arrives that these technologies do become products, the experience and lessons of early field evaluations will have played an important role.

The initial, critical funding for this work was provided by the BMGF in a program designed together with Dr. Boris Nikolic. We thank Dr. Nira Pollock (BIDMC), Sarah McGray (PATH), Dr. Bernhard Weigl (PATH), Dr. Sarah Vella (Xerox), Dr. Andres Martinez (California Polytechnic), and Sariah Khormaee (Harvard Medical School) for helpful conversations and discussions. We also thank the teams in the US, Zambia, and Vietnam who made field work possible. The initial work in paper-based diagnostics was carried out by Scott Phillips and Andres Martinez, and stimulated by work by Sam Sia and Vincent Linder. A.A.K. acknowledges support from a Graduate Research Fellowship from the National Science Foundation. J.W.H. and B.S.S were supported by BMGF (51308).

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 5836–5853
Angew. Chem. **2015**, *127*, 5932–5951

[1] P. Yager, G. J. Domingo, J. Gerdes, *Annu. Rev. Biomed. Eng.* **2008**, *10*, 107–144.
[2] S. K. Sia, L. J. Kricka, *Lab Chip* **2008**, *8*, 1982–1983.
[3] C. D. Chin, V. Linder, S. K. Sia, *Lab Chip* **2007**, *7*, 41–57.
[4] C. D. Chin, V. Linder, S. K. Sia, *Lab Chip* **2012**, *12*, 2118–2134.
[5] G. Whitesides, *Lab Chip* **2014**, *14*, 3125–3126.
[6] D. A. Pitta, R. Guesalaga, P. Marshall, *J. Consum. Mark.* **2008**, *25*, 393–401.
[7] A. W. Martinez, S. T. Phillips, M. J. Butte, G. M. Whitesides, *Angew. Chem. Int. Ed.* **2007**, *46*, 1318–1320; *Angew. Chem.* **2007**, *119*, 1340–1342.
[8] N. R. Pollock, J. P. Rolland, S. Kumar, P. D. Beattie, S. Jain, F. Noubary, V. L. Wong, R. A. Pohlmann, U. S. Ryan, G. M. Whitesides, *Sci. Transl. Med.* **2012**, *4*, 152ra129.
[9] S. J. Vella, P. Beattie, R. Cademartiri, A. Laromaine, A. W. Martinez, S. T. Phillips, K. A. Mirica, G. M. Whitesides, *Anal. Chem.* **2012**, *84*, 2883–2891.
[10] N. R. Pollock, S. McGray, D. J. Colby, F. Noubary, H. Nguyen, T. A. Nguyen, S. Khormaee, S. Jain, K. Hawkins, S. Kumar, J. P. Roland, P. D. Beattie, N. V. Chau, V. M. Quang, C. Barfield, K. Tietje, M. Steele, B. H. Weigl, *PLoS One* **2013**, *8*, e75616.
[11] C. R. Mace, O. Akbulut, A. A. Kumar, N. D. Shapiro, R. Derda, M. R. Patton, G. M. Whitesides, *J. Am. Chem. Soc.* **2012**, *134*, 9094–9097.
[12] A. A. Kumar, M. R. Patton, J. W. Hennek, S. Y. R. Lee, G. D′Alesio-Spina, X. Yang, J. Kanter, S. S. Shevkoplyas, C. Brugnara, G. M. Whitesides, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14864–14869.
[13] A. A. Kumar, C. Chunda-Liyoka, J. W. Hennek, H. Mantina, S. Y. R. Lee, M. R. Patton, P. Sambo, S. Sinyangwe, C. Kankasa, C. Chintu, C. Brugnara, T. P. Stossel, G. M. Whitesides, *PLoS One* **2014**, *9*, e114540.
[14] R. W. Peeling, D. Mabey, *Clin. Microbiol. Infect.* **2010**, *16*, 1062–1069.

[15] D. Mabey, R. W. Peeling, A. Ustianowski, M. D. Perkins, *Nat. Rev. Microbiol.* **2004**, *2*, 231 – 240.
[16] This criterion may be less important in cases where incentives are created, either by government initiatives or nonprofit organizations working to promote research on niche diseases. For example, the Food and Drug Administration (FDA) in the US has promoted research on orphan diseases through various incentives (http://www.fda.gov/forindustry/DevelopingProductsforrareDiseasesConditions/default.htm).
[17] H. Kettler, K. White, S. Hawkes, *Mapping the Landscape of Diagnostics for Sexually Transmitted Infections: Key Findings and Recommendations*, **2004**, Available at: http://www.who.int/tdr/publications/documents/mapping-landscape-sti.pdf.
[18] eMarketer Inc., "Smartphone Users Worldwide Will Total 1.75 Billion in 2014," can be found under http://www.emarketer.com/Article/Smartphone-UsersWorldwide-Will-Total-175-Billion-2014/1010536, **2014**.
[19] N. P. Pai, C. Vadnais, C. Denkinger, N. Engel, M. Pai, *PLoS Med.* **2012**, *9*, e1001306.
[20] B. H. Weigl, D. S. Boyle, T. de Los Santos, R. B. Peck, M. S. Steele, *Expert Rev. Med. Devices* **2009**, *6*, 461 – 464.
[21] J. Colyvas, M. Crow, A. Gelijns, R. Mazzoleni, R. R. Nelson, N. Rosenberg, B. N. Sampat, *Manage. Sci.* **2002**, *48*, 61 – 72.
[22] D. Somaya, *J. Manage.* **2012**, *38*, 1084 – 1114.
[23] D. Guellec, C. Martinez, P. Zuniga, *Econ. Innovation New Tech.* **2012**, *21*, 1 – 29.
[24] B. A. Caldwell, A. Young, J. Gomez-Marquez, K. R. Olson, *IEEE Pulse* **2011**, *2*, 63 – 67.
[25] C. VanVoorhis, B. Morgan, *Tutor. Quant. Methods Psychol.* **2007**, *3*, 43 – 50.
[26] G. Cumming, F. Fidler, D. L. Vaux, *J. Cell Biol.* **2007**, *177*, 7 – 11.
[27] D. L. Vaux, *Nature* **2012**, *492*, 180 – 181.
[28] National Committee for Clinical Laboratory Standards, Procedures and Devices for the Collection of Diagnostic Capillary Blood Specimens (Approved Standard HA-A5), NCCLS, Wayne, PA, **2004**.
[29] C. E. Metz, *Semin. Nucl. Med.* **1978**, *8*, 283 – 298.
[30] J. M. Bland, D. G. Altman, *Stat. Methods Med. Res.* **1999**, *8*, 135 – 160.
[31] S. J. Pocock, L. M. Friedman, C. D. Furberg, D. L. Demets, *Fundamentals of Clinical Trials*, Springer, New York, **2010**.
[32] R. Glasgow, D. Magid, A. Beck, *Med. Care* **2005**, *43*, 551 – 557.
[33] J. Lachin, *Control. Clin. Trials* **1981**, *2*, 93 – 113.
[34] K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, M. R. Munafò, *Nat. Rev. Neurosci.* **2013**, *14*, 365 – 376.
[35] United Nations, *Political Declaration on HIV and AIDS: Intensifying Our Efforts to Eliminate HIV and AIDS*, **2011**, Available at: http://www.unaids.org/sites/default/files/sub_landing/files/20110610_UN_A-RES-65 – 277_en.pdf.
[36] E. Martínez, J. L. Blanco, J. A. Arnaiz, J. B. Pérez-Cuevas, A. Mocroft, A. Cruceta, M. A. Marcos, A. Milinkovic, M. A. García-Viejo, J. Mallolas, X. Carne, A. Phillips, J. M. Gatell, *AIDS* **2001**, *15*, 1261 – 1268.
[37] Panel on Antiretroviral Guidelines for Adults and Adolescents., *Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents.* **2009**, Available at: http://aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf.
[38] G. Meintjes, G. Maartens, A. Boulle, F. Conradie, E. Goemaere, E. Hefer, D. Johnson, M. Mathe, Y. Moosa, R. Osih, T. Rossouw, G. van Cutsem, E. Variava, F. Venter, D. Spencer, *South. Afr. J. HIV Med.* **2012**, *13*, 114 – 133.
[39] *Current Medical Diagnostics and Treatment* (Eds.: S. J. McPhee, M. A. Papadakis), McGraw Hill, New York, **2010**.
[40] A. W. Martinez, S. T. Phillips, G. M. Whitesides, E. Carrilho, *Anal. Chem.* **2010**, *82*, 3 – 10.
[41] A. W. Martinez, S. T. Phillips, E. Carrilho, S. W. Thomas, H. Sindi, G. M. Whitesides, *Anal. Chem.* **2008**, *80*, 3699 – 3707.
[42] A. W. Martinez, S. T. Phillips, G. M. Whitesides, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 19606 – 19611.
[43] C. D. Chin, T. Laksanasopin, Y. K. Cheung, D. Steinmiller, V. Linder, H. Parsa, J. Wang, H. Moore, R. Rouse, G. Umviligihozo, E. Karita, L. Mwambarangwe, S. L. Braunstein, J. van de Wijgert, R. Sahabo, J. E. Justman, W. El-Sadr, S. K. Sia, *Nat. Med.* **2011**, *17*, 1015 – 1019.
[44] F. B. Piel, S. I. Hay, S. Gupta, D. J. Weatherall, T. N. Williams, *PLoS Med.* **2013**, *10*, e1001484.
[45] J. Hankins, R. E. Ware, *Lancet* **2009**, *374*, 1308 – 1310.
[46] M. Gaston, J. Verter, G. Woods, C. Pegelow, J. Kelleher, G. Presbury, H. Zarkowsky, E. Vichinsky, R. Iyer, J. Lobel, S. Diamond, C. T. Holbrook, F. M. Gill, K. Ritchey, J. M. Falletta, *N. Engl. J. Med.* **1986**, *314*, 1593 – 1599.
[47] M. E. Fabry, J. G. Mears, P. Patel, K. Schaefer-Rego, L. D. Carmichael, G. Martinez, R. L. Nagel, *Blood* **1984**, *64*, 1042 – 1046.
[48] S. Embury, M. Clark, *J. Clin. Invest.* **1984**, *73*, 116 – 123.
[49] W. H. Grover, A. K. Bryan, M. Diez-Silva, S. Suresh, J. M. Higgins, S. R. Manalis, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10992 – 10996.
[50] H. B. Weems, L. S. Lessin, *Acta Haematol.* **1984**, *71*, 361 – 370.
[51] A. A. Kumar, C. Lim, Y. Moreno, C. R. Mace, A. Syed, D. Van Tyne, D. F. Wirth, M. T. Duraisingh, G. M. Whitesides, *Am. J. Hematol.* **2015**, *90*, 31 – 36.
[52] C. Brugnara, T. Van Ha, D. C. Tosteson, *Blood* **1989**, *74*, 487 – 495.
[53] C. Brugnara, *J. Pediatr. Hematol. Oncol.* **2003**, *25*, 927 – 933.
[54] A. Kratz, M. Ferraro, P. M. Sluss, K. B. Lewandrowski, *N. Engl. J. Med.* **2004**, *351*, 1548 – 1564.

Received: December 5, 2014

