Hypothesis generating model-based wearable clinical trial

Wearable physiological sensors have the projected capability to detect unknown and unreported health conditions. Development requires rounds of discovery-oriented human subject research and confirmatory clinical trials. However, each study is a significant investment and difficult to justify in isolation. This impasse requires bootstrapping spiral device development through hypothesis-generating, model-based clinical trials. An unconventional clinical trial design addresses environmental health and infectious disease, through the day-to-day observation of diverse people who occupy a shared environment. The design utilizes a flexible suite of developmental diagnostic devices to detect the physiological impact of exposures. Through advanced data analysis, the devices provide information about deviations from normal parameters for each human subject. The correlation of these anomalies across the entire cohort generates hypotheses about exposures that impact health. These hypotheses can be investigated further in targeted studies and lead to simultaneous refinement of the devices.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2079v1 | CC-BY 4.0 Open Access | rec: 26 May 2016, publ: 26 May 2016

Title: Hypothesis Generating Model-Based Wearable Clinical Trial

Author: Benjamin C. Kirkup, Jr.

Affiliation: Center for Biomolecular Science and Engineering, Naval Research Laboratory

Corresponding Author: BC Kirkup, Jr.

Present Address: CODE 6910, Naval Research Laboratory, 4555 Overlook Ave SW, Washington DC 20375

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2079v1 | CC-BY 4.0 Open Access | rec: 26 May 2016, publ: 26 May 2016

Introduction

A world of complex environmental and infectious health effects is plagued with claims and counter claims (Parihar *et al* 2013, Magnavita 2015), shaping behaviors towards smog, pollen, indoor volatile organic compounds (VOCs), passive smoking, food products, drinking water, gluten, and even infectious agents like the flu. Establishing even a single causal relationship to enable diagnosis, treatment, and public health interventions requires decades of epidemiology because human and environmental diversity are compounded by the vagueries of self-reporting and the biases of controlled laboratory experiment. However, enabled by modern computing, sensors can characterize unreported health conditions by tracking novel physiological parameters. As an example, lung sound analysis extracts modes of respiratory disorder beyond coughs or wheezes (Fard *et al* 2015). Lung sound research is at the cusp of an unexpectedly fruitful stage (Bhattacharyya *et al* 2015, Joshi and Bartter), detecting conditions even beneath the notice of the patient (Jiang et al 2015, Yigla *et al* 2008). Other novel physiological measures, such as blood pressure waveforms (Townsend *et al* 2015) or bowel sounds (Goto *et al* 2015), also reflect cryptic underlying dysfunction. Characterizing this dysfunction while simultaneously developing the methods to detect it within the current structure of human subjects research is a study design challenge.

Novel physiological parameters are difficult to evaluate for a lack of historical data. A practical way to achieve informative correlation with exposures and outcomes is through monitoring of individuals during daily life (i.e. Holter monitor vs. stationary EKG). Continuous monitoring generates demographic-adjusted normal ranges, but also numerous repeated measurements during biological (e.g. diurnal, annual) cycles (Savage 2015). Longitudinal, particularly continuous, measurement addresses inter- and intra-personal variation which typically clouds the results of more traditional studies and requires repeated research initiatives to resolve. Equally importantly, continuous monitoring can capture clinical events, including the subclinical events that precede them and the sequelae. By capturing large numbers of events across diverse interpersonal backgrounds, wearable form factors enable even cheap, noisy sensors to yield medically relevant information (Bugane *et al* 2012, Louter *et al* 2014).

The value of broad sensor deployment in the absence of a clearly defined medical purpose is difficult to overstate. Seemingly silly sensors have advanced very rapidly by being widely deployed while still in the process of establishing utility. Opportunistic and iterative processes allow inexpensive sensors to find applications and subsequently be tuned for these new functions. Advanced sensors paradoxically suffer stagnation. Advanced sensors are deployed sparingly, under laboratory conditions, for short periods of time. Sophisticated sensors are only tested when the utility and prospect for FDA approval is evident in well in advance. Human testing is thus costly and rigid.

The hallmark of traditional diagnostic device development is the clinical trial, for validation. Clinical trials of therapeutics can be driven by outcomes, but diagnostics are validated against established standards (Hui and Zhou 1998, FDA 2007), if only a panel of medical experts (Rutjes *et al* 2007). The challenges of the imperfect gold standard are widely acknowledged if only inadequately addressed; but entering a trial without knowing what will be detected and having no standard transgresses the concept of a clinical trial. When deploying a novel wearable device, a traditional control cohort cannot be presumed to be healthy; certainly not for the duration of the study. Instead, health status, symptoms, and mechanism of illness are all latent variables, as is the performance of the instrument. Breaking

through this dialectical, reflexive ignorance requires simultaneous discovery (health status), development (refinement of the device), and validation (measurement of sensitivity and specificity).

Design Approach

Breaking this deadlock requires a novel study design: broadening the range of conditions under which the device is deployed, incorporating diverse suboptimal devices in a dynamic fashion, and accepting that results will be only gradually conclusive. The study itself is an entry into spiral development processes to address specific outcomes with increasing clarity. Ultimately, numerous clinically relevant disease states triggered by known exposures, monitored across demographically defined populations, will be diagnosed in a mechanistically relevant fashion.

The proposed study populations are cohorts with widely inclusive demographics who interact and traverse a shared physical space. The cohort is equipped with an *ad hoc* collection of prototype sensors to transduce various physiological parameters; medical history and journaling are adjunctive data streams. The sensors are expected to have substantial operational defects, including noise, malfunction, and bias. By detecting common underlying physiology, diverse sensors corroborate each other and overcome uncorrelated noise; by detecting distinct physiology (e.g. antagonistic, local to a body site), the sensors distinguish among modes of exposure and mechanisms of response.

During the study, sensors are baselined on each individual. From this baseline, excursions are reported as anomalies. Meaningful anomalies likely involve several sensors simultaneously (for example, excursions in the covariance of respiratory rate and heart rate would be distinct from excursions in the correlated rates of both). Salient anomalies are characterized and additional population data is polled to recruit less significant anomalies of a similar nature correlated in time and space, based on models of shared environmental exposure or infectious passage, including priming detection of similar anomalies from the same individual in the future. Temporal data also assists in disentangling reverse causality. During recruitment of anomalies, missing data may be imputed statistically. The strategy of recruitment leverages the breadth of the larger cohort to identify the extent of the exposure but the salient anomalies motivate biological significance and limit the potential for false discovery.

Sensors may be withdrawn, replaced, or inserted across all or part of the population during the study. Spiral sensor development enables increased focus on clinically relevant hypotheses. The study design can accept an arbitrary number of physiological sensors, to determine their utility in a dynamic fashion, and to substantiate or redirect contentious claims about causation in health and wellness. Faced with diverse environmental and infectious health challenges, a panoply of diverse sensors is not a problem, but an opportunity. Ultimately, the price of flexibility is that characterization of sensitive populations, exposures, and sensors must be confirmed in a (preferably pre-registered) clinical trial with a distinct subject population.

Addressing Statistical Significance

Determining statistical significance is difficult in an exploratory setting with repeated tests, multiple comparisons, and emergent hypotheses. Interactive determinations of statistical significance can limit false discovery. One example is the re-injection of putatively significant anomalies at random time points to attempt recruitment. If the subsidiary anomalies recruited during these injections is similar to those during organic events, then recruitment is not driven by shared exposures - but positive statistical

significance is only suggestive. Borrowing from Andrew Gelman (andrewgelman.com): Direction of research should flow not from statistical significance but instead effect size (clinical relevance), generalizability (public health relevance), and mechanistic plausibility.

Addressing Privacy/Confidentiality

The greatest safety concern is privacy/confidentiality. Three principles mitigate privacy risk. First, measurements should be transmitted parsimoniously. Second, measurements should be stored briefly. Third, data should be strategically degraded to frustrate data misuse.

Data from wearable sensors should be processed for anomalies on the subject with local computing to avoid transmission. Baselines should be stored in compressed form locally to enable streaming analysis; both refining the baseline and performing anomaly detection. Anomalies are transmitted for population-level analysis; but anomaly transmission could be opt-in, to mitigate subject privacy concerns. Second, baseline data should be compressed in a lossy fashion. Anomaly data can also be progressively reduced in precision as the study progresses. Third, precision location data is particularly sensitive; by defining neighborhoods to aggregate location data dynamically, numbers of subjects provides more than statistical power, but also a degree of privacy. Finally, given the recruitment method used to assemble suites of anomalies, fake anomalies should be injected into the transmitted data stream, generated from a plausible distribution. 'Blind injection' is employed in other disciplines (Cho 2016) to 'blind' investigators and test data analysis pipelines. In this case, injected anomalies would not effectively recruit correlated anomalies from the subject population, but would confound data misuse.

Device Safety and Behavioral Safety

Preventing physical harms from sensors is an anticipated requirement of any institutional review board. In addition, certain devices may pose unexpected behavioral risks (e.g. inconvenient exercise monitors may reduce exercise; cell phones are associated with pedestrian accidents). Behavioral safety may be more subtle but adverse behaviors are themselves events.

Social Interactions

Social interactions may be detected physiologically; equipping persons with a Holter monitor, a skin galvanic response device, and a respiratory monitor is perilously close to a walking polygraph. In a workplace, weekly meetings might be observed as environmental or contagious physiological stress. In a sense, this would be a success, and may even be relevant to a range of clinical outcomes (unless the meeting were a discussion of the sensors). This design concern is not trivially remedied during long term trials; but should be addressed during any confirmatory studies.

Conclusions

Wearable sensors are rapidly progressing in their application to health; but with a strong bias toward those with the least expensive preliminary design. These inexpensive sensors are broadly tested and rapidly refined as uses are discovered. More sophisticated sensing modes progress through a very different process, in the medical scientific literature, and despite tremendous promise, their development suffers by comparison.

We present a flexible study design which would facilitate the development of sophisticated sensors while preserving the advantages of formal studies. The joyful diversity of sensors, subjects, and exposures provides breadth and generalizability; and the concurrence both reduces the costs of subject recruitment and management, and also allows the sensors to inform on each other for the purpose of validation. Human subject protection concerns are also given due attention.

Because of the flexible nature of the study design, it is feasible to imagine a 'rolling' or 'continuous' study into which devices and subjects are continually recruited; time and geography being held relatively constant. This would optimize the use of the underlying infrastructure and limit administrative delays during initialization and termination of distinct studies. This model of study design could be conducted in any number of locally relevant environments (rural, urban, tropical, etc.) to localize and personalize medicine.

Acknowledgements

LCDR Yasuda is acknowledged for editing this manuscript. Greg Ellis is acknowledged for editing this manuscript. Greg Virgin and Aaron Jaggard are acknowledged for fruitful discussions of privacy and security. Jeff Byers is acknowledged for fruitful discussions of data analysis.

References

Bhattacharyya, P., Mondal, A., Dey, R., Saha, D. and Saha, G., 2015. Novel algorithm to identify and differentiate specific digital signature of breath sound in patients with diffuse parenchymal lung disease. *Respirology*, *20*(4), pp.633-639.

Bugané, F., Benedetti, M.G., Casadio, G., Attala, S., Biagi, F., Manca, M. and Leardini, A., 2012. Estimation of spatial-temporal gait parameters in level walking based on a single accelerometer: Validation on normal subjects by standard gait analysis. *Computer Methods and Programs in Biomedicine*,*108* (1), pp.129-137.

Cho, A., 2016. The scientist who spotted the fateful signal—and let the cat out of the bag. *Science*, *351*(6275), pp.797-797.

Fard, P.J.M., Moradi, M.H. and Saber, S., 2015. Chaos to randomness: distinguishing between healthy and non-healthy lung sound behaviour. *Australasian Physical & Engineering Sciences in Medicine*, *38*(1), pp.47-54.

Food and Drug Administration, 2007. Guidance for industry and FDA staff – statistical guidance on reporting results from studies evaluating diagnostic tests.

Goto, J., Matsuda, K., Harii, N., Moriguchi, T., Yanagisawa, M., Harada, D., Sugawara, H. and Sakata, O., 2015. New real-time bowel sound analysis may predict disease severity in septic patients. *Critical Care*, *19*(Suppl 1), p.P2.

Hui, S.L. and Zhou, X.H., 1998. Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research*, *7*(4), pp.354-370.

Jiang, H., Chen, J., Cao, J., Mu, L., Hu, Z. and He, J., 2015. Evaluation of Vibration Response Imaging (VRI) Technique and Difference in VRI Indices among Non-Smokers, Active Smokers, and Passive Smokers. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *21*, p.2170.

Joshi, M. and Bartter, T., 2015. Seeing what we hear: an eye to help the ear. *Respirology*, *20*(4), pp.524-525.

Louter, M., Maetzler, W., Prinzen, J., van Lummel, R.C., Hobert, M., Arends, J.B., Bloem, B.R., Streffer, J., Berg, D., Overeem, S. and Liepelt-Scarfone, I., 2014. Accelerometer-based quantitative analysis of axial nocturnal movements differentiates patients with Parkinson's disease, but not high-risk individuals, from controls. *Journal of Neurology, Neurosurgery & Psychiatry*, pp.jnnp-2013.

Magnavita, N., 2015. Work-related symptoms in indoor environments: a puzzling problem for the occupational physician. *International Archives of Occupational and Environmental Health*, *88*(2), pp.185-196.

Parihar, V.K., Hattiangady, B., Shuai, B. and Shetty, A.K., 2013. Mood and memory deficits in a model of Gulf War illness are linked with reduced neurogenesis, partial neuron loss, and mild inflammation in the hippocampus. *Neuropsychopharmacology*, *38*(12), pp.2348-2362.

Reitsma, J.B., Coomarasamy, A., Khan, K.S. and Bossuyt, P.M.M., 2007.Evaluation of diagnostic tests when there is no gold standard: a review of methods. *Health Technology Assessment*. 11(50).

Savage, N., 2015. Mobile data: Made to measure. *Nature*, *527*(7576), pp.S12-S13.

Townsend, R.R., Rosendorff, C., Nichols, W.W., Edwards, D.G., Chirinos, J.A., Fernhall, B. and Cushman, W.C., 2015. American Society of Hypertension position paper: central blood pressure waveforms in health and disease. *Journal of the American Society of Hypertension*.

Yigla, M., Gat, M., Meyer, J.J., Friedman, P.J., Maher, T.M. and Madison, J.M., 2008. Vibration response imaging technology in healthy subjects. *American Journal of Roentgenology*, *191*(3), pp.845-852.