Why and how should we integrate biomarkers into complex trials? A discussion on paradigms and clinical research strategies.

Frauke Musial[1], Scott Mist[2], Sara Warber[3,] Mary Jo Kreitzer[4], Cheryl Ritenbaugh[5], Christian Kessler[6]

Authors' affiliations:

1. The National Research Center in Complementary and Alternative Medicine (NAFKAM), Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Tromsø, Tromsø, Norway. (frauke.musial@uit.no)
2. Oregon Health & Science University, School of Nursing, Portland, Oregon, USA. (mists@ohsu.edu)
3. Department of Family Medicine, University of Michigan, Ann Arbor, Michigan USA. (swarber@med.umich.edu)
4.
5. Earl E. Bakken Center for Spirituality and Healing, University of Minnesota, Minneapolis MN, USA. (kreit003@umn.edu) Department of Family & Community Medicine, The University of Arizona, Tucson Arizona, USA. (ritenbau@email.arizona.edu)

6. Department of Internal and Complementary Medicine, Immanuel Hospital and Institute of Social Medicine, Epidemiology & Health Economics, Charité-University Medical Center, Berlin, Germany. (kessler.christian@gmail.com)

Running head: Biomarkers in complementary medicine

Corresponding author

Frauke Musial, PhD

The National Research Center in Complementary and Alternative Medicine (NAFKAM),

Department of Community Medicine

Faculty of Health Sciences

UiT The Arctic University of Norway

9037 Tromsø

Norway

Tel. (+47) 77 64 92 82 (Frauke Musial)

Tel. (+47) 77 64 66 50 (NAFKAM)

Fax. (+47) 77 64 68 66

e-mail: frauke.musial@uit.no

Abstract

Background: Research in complementary and alternative medicine (CAM) encounters a variety of challenges, such as potentially synergistic, multimodal, and complex interventions which are often dependent on the relationship between practitioner and patient, on specific settings, and on patients' individual preferences, expectations, beliefs and motivations. Moreover, patients seeking CAM care often suffer from chronic disease conditions, and multiple symptoms and/or pathologies. On the other hand, CAM interventions are often challenged as being solely dependent on subjective and non-specific factors without biologically-based mechanisms of action. If we agree that biomarkers as outcomes are important for the understanding of CAM interventions, a hypothesis- and strategy-driven process for the selection of the most appropriate biomarkers is needed. **Methods:** This paper presents the results of an expert panel on how to integrate biomarkers in whole system research of an interdisciplinary workshop on research methodology in complementary and alternative medicine held in November 2012. **Results:** The following main CAM research challenges were identified as a) finding appropriate biomarkers, which are able to picture the complex pathophysiological pathways and likewise complex interventions under study; b) integrating these biomarkers into clinical trials in CAM; and c) identifying the biomarkers specific to the particular CAM intervention being applied. **Conclusion:** The paper provides a disease/condition/symptom and intervention driven strategy regarding how to identify the outcomes of interest and possible related biomarkers. The research approach presented here allows the selected biomarkers to be grounded in conventional physiology/pathophysiology as well as complementary and alternative concepts, including traditional systems of

medicine. The goal is to provide researchers in the field with a framework on how to integrate biomarkers into complex trials.

Key words: Complementary and Alternative Medicine (CAM), biomarkers, trial methodology, complex interventions, Traditional Medicine, Whole Medical Systems, integrative medicine

Hintergrund: Die Forschung im Bereich der Komplementär-und Alternativmedizin (CAM) muss sich mit einer Vielzahl von Herausforderungen auseinander setzen. Dazu gehören potenziell synergistische, multimodale und komplexe Behandlungsmodelle, die häufig sehr von der Therapeut – Patient Beziehung geprägt werden ebenso wie die Tasache, dass eine Behandlungssituation und Wahl sehr von den individuellen Präferenzen des Patienten, seinen Erwartungen, Überzeugungen, und Motivationen abhängig sein kann. Darüber hinaus leiden Patienten, die eine Komplementär- oder alternativmedizinische (CAM) Behandlung wünschen, häufig an chronischen Krankheitszuständen sowie komplizieren Symptombildrn, und / oder mehreren Pathologien zugleich. Eine weitere Herausforderung ist, dass CAM-Interventionen oftmals auf überlieferten Traditionen und Erfahrungswerten beruhen, ohne dass mögliche biologische Wirkmechanismen bekannt sind. Gerade aus diesem Grund kann es zielführend und interessant sein, Biomarker in eine wissenschaftliche Untersuchung zu integrieren, um so mehr über die potenziellen Wirkmechanismen von teilweise historisch sehr alten und traditionellen Therapien zu erfahren. Unter der Voraussetzung, dass Biomarker für das Verständnis von CAM-Interventionen nützlich sein können, ist ein hypothesen- und strategiegesteuerter Prozess für die Auswahl der am besten geeigneten Biomarker erforderlich. **Methode:** Es werden die Ergebnisse eines Expertenpanels zur Integration von Biomarkern vorgestellt. Das Panel war Teil eines interdisziplinären Workshops zu Forschungsmethoden in der Komplementär- und Alternativmedizin der im November 2012 stattfand. **Ergebnisse:** Als die wichtigsten Herausforderungen wurden a) Die Identifikation geeigneter Biomarker, die in der Lage sind, die komplexen pathophysiologischen Bedingungen chronischer Erkrankungen und ebenso komplexer Interventionen abzubilden; sowie b) Die praktische Integration dieser Biomarker in klinische Studien in CAM Bereich; und die c) Selektion derjenigen Biomarker, die für die

jeweilige CAM-Intervention spezifisch sind, identifiziert. **Schlussfolgerung:** Der Artikel schlägt eine an Krankheit / Zustand / Symptombild orientierte Strategie vor, mit der die wichtigsten Hauptzielgrössen für eine Studie und damit potenziell assoziierte und valide Biomarker identifiziert werden können. Das Ziel dieser Stratgie ist es, klinischen Forschern ein System zur Einbindung von Biomarkern in komplexe klinische Studien anzubieten. Mit dem hier vorgestellten Forschungsansatz werden Biomarker auf basis konventioneller Physiologie / Pathophysiologie ausgewählt, behalten aber gleichzeitig ihre Relevanz und Validität für die zumeist komplexen, komplementären und alternativen Therapien.

Schlüsselwörter: Komplementärmedizin, Alternativmedizin, Biomarker, klinische Studien, komplexe Interventionen, traditionelle Medizin, holistische medizinische Systeme, integrative Medizin

INTRODUCTION

This paper is a consensus paper as one of the results of the workshop "Designing research aligned to the whole systems model of health, disease and healing" as part of NAFKAM's 2012 Northern Lights Workshop series held in Sommarøy, Tromsø, Norway. The overall goal of the 2012 workshop was to define issues that need to be included in whole systems research and point to relevant methodological issues. The workshop started with the whole group of participants applying a World Cafe style brainstorming on topics and issues for whole systems model research designs. The potential role of biomarkers and whether and how they should/could be integrated in complex trials was identified as one of several relevant topics and the group of co-authors was assigned as "expert" panel for this topic.

Within the expert panel, the participant began with a short brainstorming session on why biomarkers can be a useful outcome in complex trials. Taken from there, the expert panel developed a principal design template and discussed its operationalization with several, specific examples of increasing complexity. Since the group was small and rather homogenic in their opinions, the discussion was free and not formalized. The paper represents the consensus statement of the expert panel.

The problem: evaluating the impacts of complex interventions

Research in complementary and alternative medicine (CAM) encounters a variety of challenges, such as potentially synergistic, multimodal, and complex interventions, aiming at symptom amelioration, and psychosocial and behavioral changes in the patient with long-

term effects. Furthermore, these interventions are often dependent on the relationship between practitioner and patient, on specific medical settings, and on patients' preferences, expectations, beliefs and motivations. Moreover, patients seeking CAM care often suffer from chronic diseases and multiple pathologies [1]. The challenges these complex and multifactorial conditions impose on clinical research design are particularly relevant for whole medical systems. Typical examples for whole medical systems are traditional medical systems such as Chinese Medicine, Ayurveda, Kampo, Tibetan Medicine, Unani-Tibb, Native American Medicine, as well as homeopathy, anthroposophy, and the rapidly emerging field of integrative medicine. These whole systems not only include multiple treatment modalities, but also alternative diagnoses and patient-practitioner interactions, as well as techniques for changing the patients´ (or in case of primary prevention, clients´) behavior, all of which are frequently implemented in highly individualized fashions and in system-specific settings [2] [3-5]. Nonetheless, rigorous investigation of clinical effects in controlled clinical trials is possible [6] [7] [8] and even randomized clinical trials have been suggested [9]

Moreover, research in CAM frequently faces the challenges that treatments are not derived from conventional biological hypotheses and the therapies are implemented regardless of whether the "conventional" biological mechanisms, comparative effectiveness, component efficacy, or even safety aspects are documented in the research literature [10]. On the other hand, many of these interventions, in particular in whole medical systems [2], have been practiced for centuries or even millennia, so that at least an implicit empirical knowledge on clinical effectiveness and safety within certain cultural settings is available [11, 12].

All these preconditions, the complexity of the interventions and the treatment setting, the fact that many of the patients suffer from several chronic conditions or diseases, and the fact that there is generally no clear hypothesis on the biological mechanism of action, make the applicability of the gold standard in clinical research, the randomized controlled trial (RCT), particularly challenging [1, 2, 10, 13] although, as we see it, not impossible. In the field of CAM, there has been a growing awareness that the development of appropriate methodological and statistical frameworks for the investigation of complex interventions is one of the key answers to the quest [1, 2, 8, 10], including a partial re-evaluation of the "outcomes" concept [13].

Why biomarkers as outcomes for complex interventions?

Considering all the previously discussed challenges, why should we even attempt to utilize biomarkers as outcomes in complex trials? Treatment effects, measured by biomarkers, are usually considered a sign of specific efficacy. However, could and should we not be content if we are able to show the clinical effectiveness of CAM interventions to the benefit of the patient?

We believe that there are at least three good arguments for research strategies to also include biomarkers:

1) **Pragmatism**: CAM interventions in conventional medical care settings are often delivered in addition to conventional care or as being part of integrative medical care. Enhancing the understanding of complex CAM interventions with regard to the conventional biomedical model is likely to facilitate the integration of the various

approaches and prevent possible undesirable interactions among different treatment modalities. The synergistic effect of a common basis of understanding would improve care, increase patient safety, and improve practitioner- and patient-satisfaction.

2) **Epistemology/hermeneutics:** Many systems of traditional healing are based on theoretical models about health and disease, or even broader, the organization of life in general. Broadening the spectrum of outcomes in complex CAM trials from emotional/motivational, functional/behavioral to biological/physiological mechanisms may allow for a "cross talk" between the traditional and "western" natural science models of health and disease with the overarching goal of a common understanding and a better approximation of "medical realities".

3) **Strategic:** Even though we recognize that the currently acknowledged "bio-psycho-social model" reaches substantially beyond the boundaries of the "biomedical model" of health and disease, we observe that much of the current medical research, also within scientific CAM contexts, is still being conducted along the tenets of the biomedical model. If the aim is to establish complex CAM interventions in a model of good patient care, demonstrated effects on biomarkers will increase the potential for integration of CAM interventions in the predominant models of contemporary medicine.

What frameworks can guide biomarker-selection for complex interventions?

If we agree that biomarkers as outcomes are important for the understanding of CAM interventions, a hypothesis- and strategy-driven process for the selection of the most

appropriate biomarkers is needed. Most important, the chosen biomarkers should be valid to the indication/condition/syndrome studied. This is best achieved, if they are oriented towards ***the core subjective health complaint or symptom*** of the patientMoreover, as Paterson et al. [13] consequently emphasize, outcomes which are appropriate for complex health interventions ideally are able to reflect ***changes and dynamics.*** In complex health interventions we often see that a treatment, even under strictly controlled conditions, shows its effects in the way a subject adapts to a challenge. As Paterson et al. [13]] point out, a process is something that enables the individual to adapt to varying experiences. Therefore, while trying to design a trial and searching for the right outcome, we may be trying to quantify something like a ***"Flexibility-Stability" continuum***. With regard to including biomarkers among the outcomes, this means that we have to be aware of whether we expect a treatment to show its effect in the **magnitude** of an outcome, or by altering its **variability.**

Two biomarkers exemplifying this principle are body temperature and heart rate variability. A stable body temperature (thermoregulation) is a core requirement in endothermic animals (like mammals) and a mechanism that provides a stable environment for the chemical reactions in the body cells. Body temperature undergoes slight variations due to circadian rhythms and activity; it is a classic example for homeostatic regulation [14]. Consequently, deviances from the normothermic optimum, such as fever, are generally a sign of disease or, at least, a prodromal symptom. Heart rate, on the other hand, is a classic example for a physiological variable made for flexibility and adjustment. Heart rate varies with regard to physical (e.g. sport) or psychological (e.g. stress) challenges and supports behavioral actions needed in response to the external (e.g. climbing stairs) or internal (e.g. thoughts and emotions) environment. The loss of heart rate variability is generally a pathological sign of

cardiovascular disease. Consequently, heart rate variability would be a classic example for a dynamic variable, where a desirable treatment outcome could be an increase of heart rate variability [15].

These two examples illustrate that changes in processes (e.g. heart rate variability) as well as levels (e.g. homeostatic body temperature) may be indicators of improved health. It is highly relevant to consider the nature of the underlying mechanism when choosing a biomarker. First, the biomarker should relate directly to the question (e.g. whether the mean or median of a measure or an index of variability is appropriate) and will to some degree impact the study design. However, the question of how sensitive the biomarker is to change will to a large extent drive the study sample size needed to detect an effect.

These conceptual issues may be further challenged when trying to integrate biomarkers into complex interventional trials, since biomarkers are usually identified and selected on the basis of a clear, biological hypothesis investigated in a design which allows for the application of Wilhelm Wundt's (1832-1920) fundamental "principle of isolated variation" in experimentation (e.g. cited in [16]). The principle of isolated variation requires that the treatment groups are varied according to only one particular variable, which can be expected to exert an effect on the dependent or outcome variable. All other sources of variation must remain constant. The classical pharmacological RCT is a typical example in that it only varies the content of the active drug against a placebo. The application of the "principle of isolated variation" is mandatory when biological mechanisms or the efficacy of a specific component of the therapy (component efficacy) of a treatment are to be studied (see also [10]).

If the aim is to investigate the conventional biological underpinnings of CAM interventions, biomarkers are mandatory. With regard to biomarkers and the "principle of isolated variation," complex trials in the CAM field struggle with two major challenges:

a) The mechanisms of action of the intervention are commonly unclear and the explanatory models are rarely grounded in conventional physiology/pathophysiology. Therefore, it is often difficult to decide which variation of the experimental or treatment conditions is being isolated, and often several variables unavoidably change simultaneously. Consequently, biomarkers may also vary in uncontrolled ways in several of the treatment conditions. This can make the study results difficult to interpret.

b) The CAM interventions are complex and often address several aspects of the patient and his or her symptoms, which from an understanding of the (e.g. organ-) specific conventional pathophysiology may be seen as unrelated to the disease or at least only relevant at the second or third level. However, these (cumulated) second or third level effects may actually have an impact on the specific pathophysiology, but the impact may show itself on a different, likely a longer, time line.

In the light of the complexity of the problem, can we integrate biomarkers into these multidimensional trials and is there a need to do so? The answer to both questions is clearly "yes" [2, 13] and there has been an explicit call for "objective change indicators" which should preferably be biomarkers [13].

METHODS

A critical reflection on the use of biomarkers and a "caveat"

Even though this group of authors favors or is at least supportive to the concept of the use of biomarkers, we are well aware that a wrongly selected biomarker will not add to our understanding of mechanisms of effect of an intervention, quite the opposite. Since biomarkers are often seen as making study outcomes more credible and they are therefore "fashionable", it is also fundamental, that the biomarkers used are in fact valid for i) the symptom pattern, and ii) the intervention.

It is tempting to choose biomarkers which are convenient or readily available. A good example is the biomarker "salivary cortisol" which has often been utilized when stress reduction was suspected to play a role. However, salivary cortisol as a hormonal measure is a very dynamic stress marker, vulnerable to many different factors that play specific roles in patients, such as medications, co-morbidities, and every day stressors [17-19]. Therefore salivary cortisol is not particularly suitable to detect long-term effects. It can, however, be used validly in standardized sub-experiments, where a dynamic response is systematically evoked and where salivary cortisol is a well-established marker for the stress response [20, 21] (see example 4).

It is exactly for these reasons that the biomarker expert panel was established at the Northern Lights workshop "Designing research aligned to the whole systems model of health, disease and healing". The research strategy developed and suggested here intends to guide researchers in their selection of biomarkers, in order to avoid designs, where the chosen biomarker is not valid for the study. Such a situation is at best, a waste of resources, but can also, depending on the nature of the biomarker, set the study participants at risk, without adding useful information to the understanding of the mechanisms of effect and should be avoided.

Strategies for the selection of specific biomarkers in complex clinical trials

The central question is: how can biomarkers be selected strategically, so that the likelihood that they are valid for the intervention and the design is maximized? The challenge can be addressed through the implementation of two general principles which refer a) to the process of ***selecting*** the ***appropriate biomarkers*** and b) the application of small "***sub-experiments***" embedded within complex trials.

a) The selection of biomarkers should be driven by the physiological system where the treatment-induced change can be expected, such as the central nervous system, the autonomic nervous system, the immune system etc. Moreover, the level where the change is to be expected needs to be taken into account (e.g. for the CNS: at the level of the brain or below?) [22, 23]. The most rational guidance is provided by the ***relevant symptoms*** of the patient. If this is, for example, pain, then a pain-related biomarker would be appropriate.
b) "Sub-experiments" or challenges to the system are needed when complex interventions cannot be expected to alter biomarkers at a resting level/state, but may alter the physiological response to psychophysiological challenge (e.g. mental stress, physical activity). In these sub-experiments, specific challenges to the physiological system in focus, e.g. a stress test, are applied and the ***change in response to the challenge,*** meaning an **evoked response** (e.g. change in heart rate variability), is measured as outcome (see also Fig 4).

In the following section, we suggest a generalized trial design which provides a rational guideline for the selection of biomarkers, taking into account the challenges and frameworks

identified above. Furthermore, we provide three different virtual trial examples to serve as examples for the application of the template, with different indications (low back pain, esophageal reflux disease, breast cancer survivors) and different complex interventions (Traditional Chinese medicine, Ayurveda and Integrative Care).

RESULTS

A trial design template for a strategic biomarker selection

Usually in a trial, the study participants are selected according to a health condition or disease by a diagnosis that is operationalized by inclusion and exclusion criteria. From here, either all or a selected subgroup of patients will receive a CAM treatment. The CAM treatment usually also has i) its own definition of health and disease and ii) its own theory of pathodynamics. Consequently, a second set of selection (inclusion and exclusion) criteria according to the premises of the whole medical system or other complex intervention may or may not be implemented. Figure 1 describes this process.

In the first step, the key symptom/variable which is expected to be addressed by the intervention needs to be identified, and an appropriate outcome measure as operationalized for this symptom/variable needs to be selected. This may often be a self-report outcome. Self-reports are of immense importance even though this paper focuses on biomarkers, because self-reports provide a measure for the most relevant aspect of any trial: they represent a measure of suffering, the outcome with the highest relevance for the patients and thus for the whole trial! A common example would be subjectively perceived pain as the key symptom and the visual analogue scale as its operationalization for measurement (see example TCM,

figure 2). It is assumed that any type of successful intervention should be able to induce symptom relief.

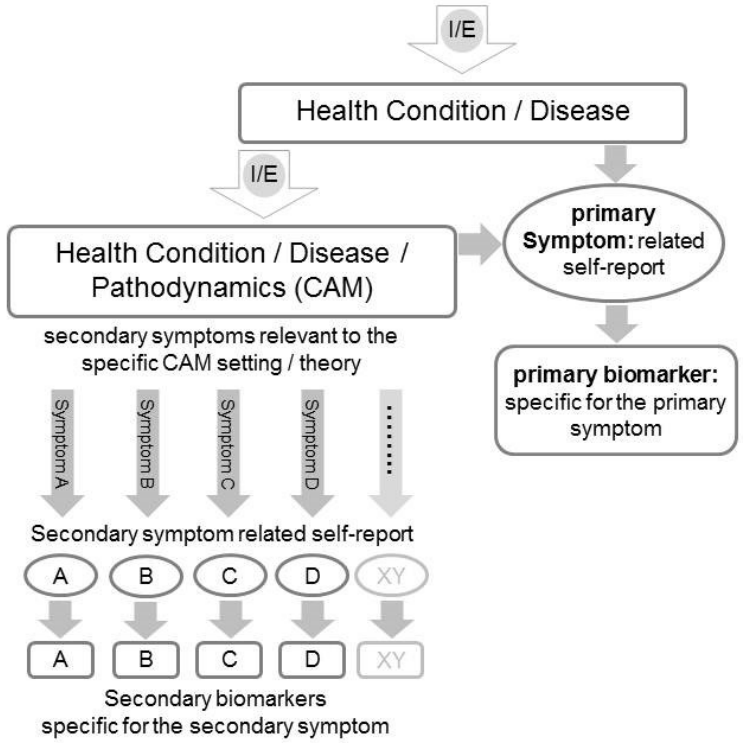

Figure 1:

Trial design template for the selection of suitable biomarkers. The biomarker selection is symptom driven. The primary biomarker is closely related to the primary symptom of the health condition or disease, while the secondary biomarker is closely related to a secondary symptom (A, B, C…XY), which is derived from the CAM-defined pathodynamics. I/E = Inclusion/Exclusion criteria.

The next step is to identify a biomarker, which is as closely related to the selected main symptom and/or self-report as possible. This biomarker will be the main or primary biomarker for the trial. For pain, this could for example be a neurophysiologic measure taken from the quantitative sensory testing system (QST). QST includes a series of tests applying challenges in the form of controlled pain stimuli to the pain processing system [22, 24-26] (see example TCM, figure2). If the trial would result in a change of the subjective measure ***and*** the primary biomarker, then this would document a health condition- / disease-related specific effect of the treatment.

However, *every* intervention that somehow addresses the main symptom is likely to also influence the related biomarker. For many of the purposes that biomarkers are integrated into complex trials, this would be sufficient. Nonetheless, in order to identify *specific intervention-related* mechanisms of action, it may make sense to select other, additional biomarkers, which are less related to the main or core symptom, but more closely related to other aspects of the intervention. In order to be as specific to the intervention as possible, other symptoms, which are of relevance for the particular intervention, need to be identified. This is of major relevance in all whole medical systems trials, as the whole system interventions as a general rule address symptom complexes. One or more of these additional symptoms need to be identified. In our TCM example (figure 2), this is sleep disturbance, and consequently, the most appropriate biomarker for this symptom needs to be identified, which in this case could be a sleep EEG. We advocate that this additional symptom and biomarker should also be associated with a validated self-report measure.

As mentioned above, a theoretical example from a TCM trial on low back pain illustrates this strategy well (figure 2). Patients with low back pain are identified and a selected

subgroup is subjected to a complex TCM intervention. The visual analogue scale is chosen as subjective measure for the most relevant symptom, pain. Since the quantitative sensory testing system (QST) provides a biomarker directly targeting the pain processing pathways and networks, it is a commonly utilized biomarker for pain related interventions [22, 24-26]. In TCM theory, a TCM diagnosis for the pain syndrome could be associated with sleep disturbances therefore sleep disturbances are a second, TCM-related symptom. In addition to a validated self-report, such as the Pittsburgh Sleep Quality Index [27], sleep quality could be assessed by EEG in a sleep laboratory. Consequently, if the TCM intervention reduces pain and shows a related pattern in the QST measures, then this outcome supports a specific effect on the pain processing networks through the TCM intervention. If, in addition, the sleep EEG and PSQI are also indicative of a better sleep quality, then this result shows a pattern, which can be directly related to TCM theory. If the design was such that the TCM intervention was compared with other active conditions, which likewise reduced pain and related QST measures, but only the TCM interventions showed an effect on sleep disturbance, or demonstrated the largest effect independent of pain score change, then this could be interpreted as a specific effect of the TCM intervention.

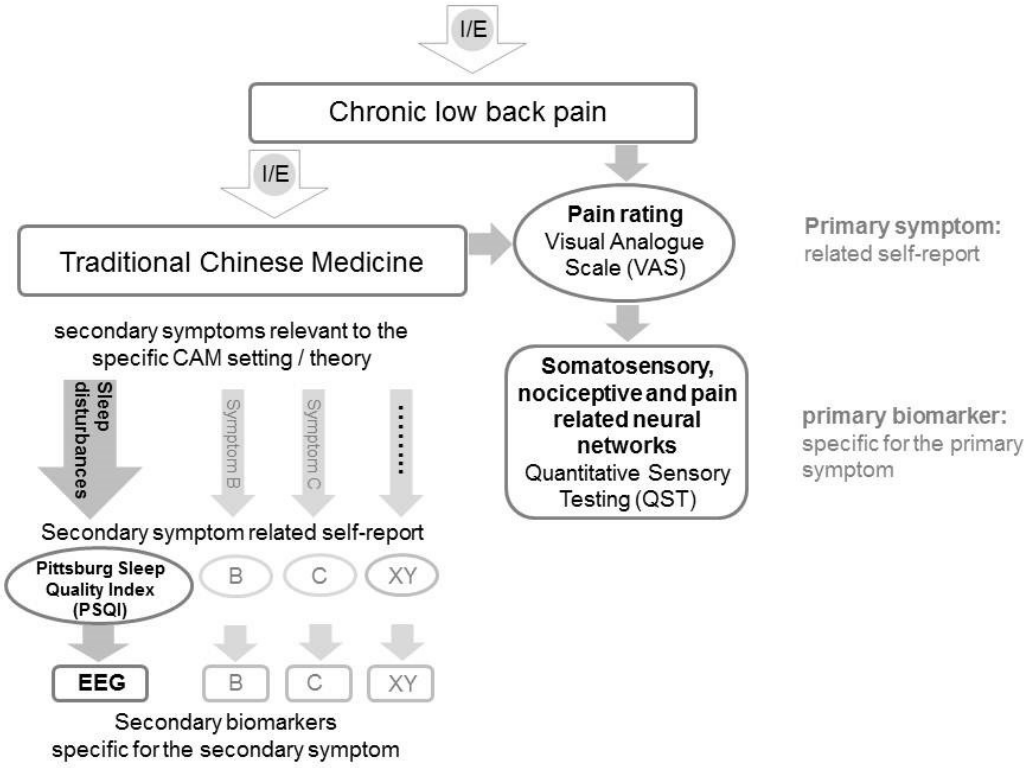

Figure 2:

Example from a TCM trial on chronic low back pain. The primary biomarker is QST which is a neurophysiologic pain marker and thus closely related to the primary symptom back pain, measured by the pain rating. According to TCM theory, sleep quality is a relevant symptom related to chronic low back pain. Therefore, sleep quality is the secondary symptom relevant for TCM theory and measured with the PSQI. The appropriate secondary biomarker for sleep quality is the sleep EEG.

Our next hypothetical example (Figure 3) illustrates the principle of a trial on gastro-esophageal reflux disease (GERD), treated with a complex Ayurvedic intervention. After selecting subjects according to conventional criteria for GERD, a subgroup of patients suffering from *Amla-Pitta* according to Ayurvedic diagnosis could be further selected [26]. The main symptom in both the conventional and the Ayurvedic disease entity is heartburn (due to acid reflux), which can be measured by esophageal pH monitoring. Therefore, the related biomarker could be the determination of esophageal pH with an esophageal sensor monitoring pH using a pH catheter. The esophageal pH monitoring is also a good example for the earlier discussed principle, that disease is often characterized by the loss of flexibility and that a healing process may be shown in an increased flexibility after treatment. The subjective impact on the patients´ life and their suffering can be measured with a variety of instruments e.g. [28-30]. The example here includes the reduction in quality of life specifically related to gastrointestinal reflux disease, measured with the QOLRAD [31]. In addition the Ayurvedic Amla-Pitta syndrome may also include a disposition for Irritable Bowel Syndrome with diarrhea (IBS-D), and a tendency for aggressive behavior. Consequently, gastrointestinal transit and salivary testosterone were selected as additional biomarkers [32, 33], specifically chosen for the Ayurvedic theory and intervention. A bowel habit patient diary can be used as self-report measure for GI-transit and the aggression questionnaire (AQ) as subjective measure for aggression [34, 35].

As in our TCM example, a relief of heartburn together with an increase in esophageal pH would indicate a treatment effect on GERD. If this effect were associated with a reduced tendency for IBS-D and normalized gastrointestinal transit, as well as a reduced tendency for aggressive behavior and normalized salivary testosterone, and this pattern of effects would

not show in other active control conditions, then this result would clearly relate back to the particularities of the Ayurvedic treatment and theory.

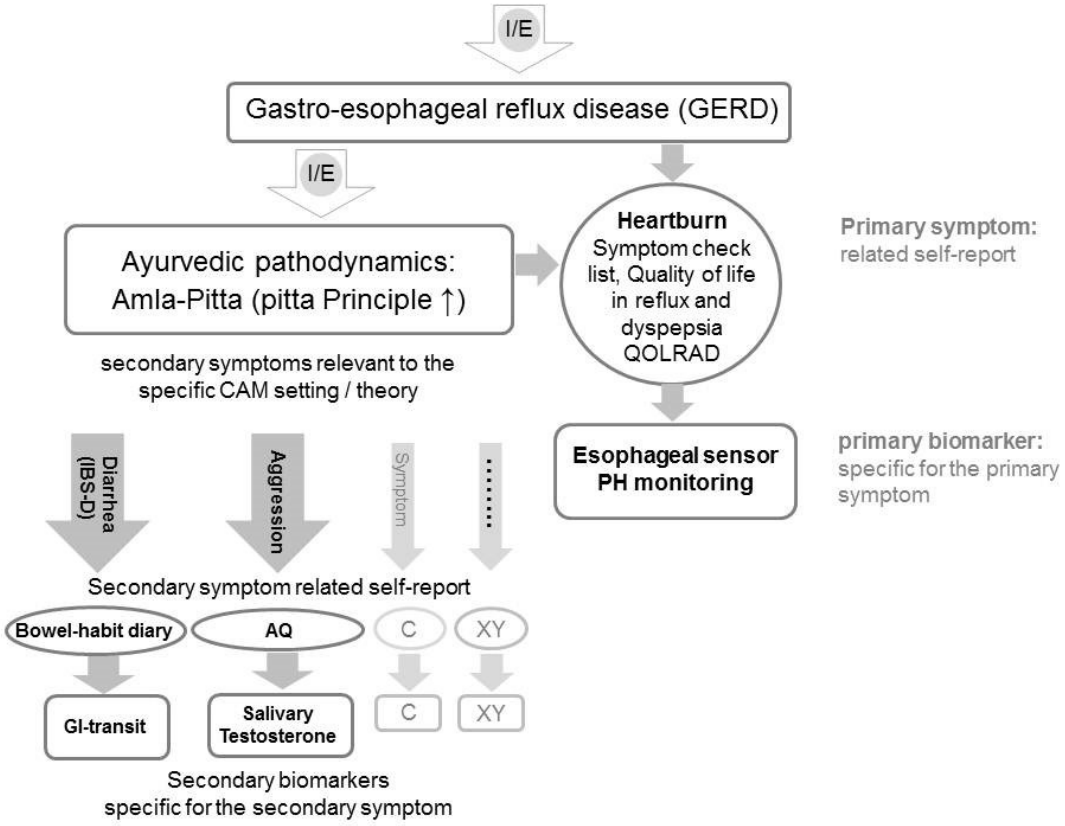

Figure 3:

Gastro-esophageal Reflux Disease and Ayurveda. The main symptom is heartburn which is measured by a symptom check list and its impact on the quality of life, measured by QOLRAD. Closely related to the main symptom is the esophageal PH as a biomarker. In addition the Ayurvedic Amla-Pitta syndrome may also include a disposition for Irritable Bowel Syndrome with diarrhea (IBS-D), and a tendency for aggressive behavior. Therefore gastrointestinal transit and salivary testosterone were selected as secondary biomarkers.

Our last example constructs a trial of a multi-modal complex treatment program for breast cancer survivors (figure 4). The main symptom, which should be addressed by this treatment program, is fatigue. Therefore, the spontaneous EEG could be selected as the appropriate biomarker in order to indicate the CNS correlates of fatigue, namely increased slow wave activity [36-40]. The FACIT-fatigue scale provides a reliable, functional measure of fatigue [41, 42]. In addition, a major component of the multi-modal treatment program focuses on stress and stress management under the assumption that the diagnosis of a potentially life threatening disease and the subsequent therapy, often including surgery and chemotherapy, are major stressors. Of course, to measure stress, many self-report instruments are available and should be carefully selected and applied. In our example, the Perceived Stress Scale (PSS), one of the many well-established and valid instruments, is chosen [43, 44]. Moreover, here stress is seen as a variable that reflects a certain level (stress level) and would thus be a *static* condition, but which can also be seen as a *dynamic* event, the stress response to a challenge. Natural killer cells could be selected as a suitable biomarker for the measurement of stress level [45, 46]. As for the measurement of the stress response, a stress interview, the Trier Social Stress Test (TSST) [47], could be administered and skin conductance, salivary cortisol response, and other cardiovascular measures (e.g. heart rate response) measured [15, 20, 21]. Thus, the stress interview is a good example of a "sub-experiment" embedded in a trial [47].

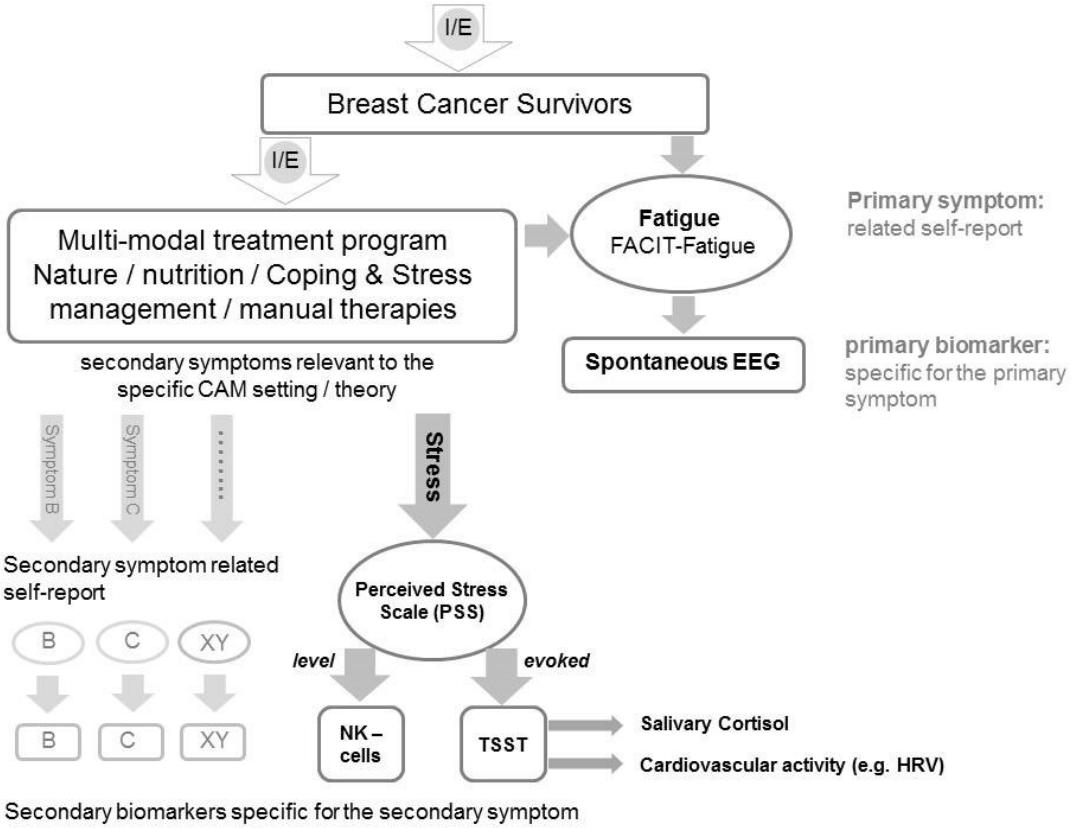

Figure 4:

A multi-modal complex treatment program for breast cancer survivors. The stress interview (TSST) is a good example of a "sub-experiment" or a challenge to the system in question embedded in a larger trial.

DISCUSSION / CONCLUSION

We hope that the suggested strategies will assist other researchers in selecting biomarkers that are meaningful and valid for the interventions studied. As complex as the challenge is, these authors are convinced that a hypothesis- and symptom-driven framework is the most promising with regard to the validity of the selected outcomes. Nonetheless, integrating biomarkers in complex CAM trials is logistically challenging, tedious, time consuming, and costly.

Why should we do it? Research on CAM struggles with the fact that complex interventions have often been applied over centuries or even millennia, and the clinical effectiveness of some is known and is sometimes even striking. However, there is rarely a good understanding (at least from the perspectives of conventional hermeneutics) as to why they work. Why is that so important?

It is widely accepted in the CAM field that a profound knowledge of the clinical effectiveness and the relevant context factors of these complex interventions is required in order to design patient-centered care models that are accepted by the global medical community and health care stakeholders. However, if we want to improve CAM interventions, and tailor them around the complexity of the individual patient's needs, then knowledge of the physiology involved in these interventions is mandatory. Moreover, this approach may be a valuable research tool in the attempt to analyze complex and whole medical systems *as they currently exist*.

We are well aware that others have previously made useful and explicit statements in relation to the integration of meaningful biomarkers into clinical trials [2, 10, 13]. We wished to continue and extend this discussion in order to provide strategies on how this can be

done most appropriately, guided both by the symptom pattern and the particularities of the CAM intervention chosen.

List of abbreviations:

CAM	complementary and alternative medicine
RCT	randomized controlled trial
TCM	Traditional Chinese Medicine
I/E	Inclusion/ exclusion criteria
QST	Quantitative sensory testing
VAS	Visual analogue scale
PSQI	Pittsburg Sleep Inventory
EEG	Electroencephalogram
GERD	Gastrointestinal reflux disease
IBS-D	diarrhea predominant irritable bowel syndrome
QOLRAD	Quality of Life in Reflux And Dyspepsia
AQ	Aggression Questionnaire
GI-transit	Gastrointestinal transit
FACIT	Functional Assessment of Chronic Illness Therapy - Fatigue scale
PSS	Perceived Stress Scale
NK-cells	Natural killer cells
TSST	Trier Social Stress Test

HRV Heart rate variability

Competing interests:

The authors declare that there are no competing interests.

Author's contributions:

This paper is the result of a workshop which was part of NAFKAM's 2012 Northern Lights Workshop in Sommarøy, Tromsø, Norway. Therefore, all authors were involved in the development of the methodology and the concepts presented, and in the drafting of the overall paper outline. FM and SM developed the outline of the workshop CK provided the conceptualization of the graphs. SM, CK, and SW provided specifics for the examples. All authors revised the manuscript carefully and approved the final version of the manuscript.

Acknowledgements:

The Northern Lights Workshop in Sommarøy, 2012 was supported by a grant from the faculty of health sciences, University of Tromsø, Norway**.** We are thankful to Åsa Sohlen for indispensable technical support.

References:

1. MacPherson H, Peters D, and Zollman C, *Closing the evidence gap in integrative medicine.* BMJ, 2009. **339**.
2. Ritenbaugh C, et al., *Whole systems research becomes real: new results and next steps.* J Altern Complement Med, 2010. **16**(1): p. 131-7.
3. Kessler C and Michalsen A, *The role of whole medical systems in global medicine.* Forsch Komplementmed, 2012. **19**(2): p. 65-6.
4. Kessler, C.S., et al., *Effectiveness of an Ayurveda treatment approach in knee osteoarthritis - a randomized controlled trial.* Osteoarthritis Cartilage, 2018. **26**(5): p. 620-630.
5. Kessler, C.S., et al., *Reliability of Ayurvedic Diagnosis for Knee Osteoarthritis Patients: A Nested Diagnostic Study Within a Randomized Controlled Trial.* J Altern Complement Med, 2019.
6. Furst DE, et al., *Double-blind, randomized, controlled, pilot study comparing classic ayurvedic medicine, methotrexate, and their combination in rheumatoid arthritis.* J Clin Rheumatol, 2011. **17**(4): p. 185-92.
7. Furst, D.E., et al., *Well controlled, double-blind, placebo-controlled trials of classical Ayurvedic treatment are possible in rheumatoid arthritis.* Ann Rheum Dis, 2011. **70**(2): p. 392-3.
8. Witt, C.M., et al., *Comparative effectiveness of a complex Ayurvedic treatment and conventional standard care in osteoarthritis of the knee--study protocol for a randomized controlled trial.* Trials, 2013. **14**: p. 149.
9. Ernst E and Furst DE, *A blueprint for placebo-controlled double-blind studies of complex, individualized interventions.* Focus on Alternative and Complementary Therapies, 2011. **16**: p. 49-50.
10. Fonnebo, V., et al., *Researching complementary and alternative treatments--the gatekeepers are not at home.* BMC Med Res Methodol, 2007. **7**: p. 7.
11. World Health Organization (WHO), *WHO traditional medicine strategy*. 2002, World Health Organization: Geneve.
12. World Health Organization (WHO), *Traditional Medicine. Report by the Secretariat. A 56/18*. 2003, World Health Organization Geneve.
13. Paterson, C., et al., *Evaluating complex health interventions: a critical analysis of the 'outcomes' concept.* BMC Complement Altern Med, 2009. **9**: p. 18.
14. Breedlove SM, Rosenzweig MR, and Watson NV, *Chapter 13. Homeostasis: Active regulation of internal states*, in *Biological Psychology. An introduction to behavioral, cognitive, and clinical neuroscience. 5th Edition*. 2007, Sinauer Associates Inc: Sunderland, MA, USA. p. 388-393.
15. Thayer, J.F., et al., *A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health.* Neurosci Biobehav Rev, 2012. **36**(2): p. 747-56.
16. Krauth, J., *Experimental Design. A handbook and dictionary for medical and behavioral research.* . Techniques in the behavioral and neural sciences., ed. J.P. Huston. Vol. 15. 2000, Amsterdam: Elsevier.
17. Granger, D.A., et al., *Medication effects on salivary cortisol: tactics and strategy to minimize impact in behavioral and developmental science.* Psychoneuroendocrinology, 2009. **34**(10): p. 1437-48.
18. Hellhammer, D.H., S. Wust, and B.M. Kudielka, *Salivary cortisol as a biomarker in stress research.* Psychoneuroendocrinology, 2009. **34**(2): p. 163-71.
19. Kudielka, B.M., D.H. Hellhammer, and S. Wust, *Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge.* Psychoneuroendocrinology, 2009. **34**(1): p. 2-18.

20. Campbell, J. and U. Ehlert, *Acute psychosocial stress: does the emotional stress response correspond with physiological responses?* Psychoneuroendocrinology, 2012. **37**(8): p. 1111-34.
21. Kudielka, B.M. and S. Wust, *Human models in acute and chronic stress: assessing determinants of individual hypothalamus-pituitary-adrenal axis activity and reactivity.* Stress, 2010. **13**(1): p. 1-14.
22. Musial, F., D. Spohn, and R. Rolke, *Naturopathic reflex therapies for the treatment of chronic back and neck pain - Part 1: Neurobiological foundations.* Forsch Komplementmed, 2013. **20**(3): p. 219-24.
23. Musial, F., A. Michalsen, and G. Dobos, *Functional chronic pain syndromes and naturopathic treatments: neurobiological foundations.* Forsch Komplementmed, 2008. **15**(2): p. 97-103.
24. Rolke, R., et al., *Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values.* Pain, 2006. **123**(3): p. 231-43.
25. Rolke, R., et al., *Quantitative sensory testing: a comprehensive protocol for clinical trials.* Eur J Pain, 2006. **10**(1): p. 77-88.
26. Spohn, D., F. Musial, and R. Rolke, *Naturopathic reflex therapies for the treatment of chronic pain - Part 2: Quantitative sensory testing as a translational tool.* Forsch Komplementmed, 2013. **20**(3): p. 225-30.
27. Buysse, D.J., et al., *The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.* Psychiatry Res, 1989. **28**(2): p. 193-213.
28. Murthy, S., *Madhava Nidanam (Roga Viniscaya) of Mahavakara. A Treatise On Ayurveda.* 51. Reprint. ed. 2001, Varanasi: Chaukambha Orientalia.
29. Shaw, M.J., et al., *Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease.* Am J Gastroenterol, 2001. **96**(1): p. 52-7.
30. Wahlqvist, P., et al., *The Work Productivity and Activity Impairment Questionnaire for Patients with Gastroesophageal Reflux Disease (WPAI-GERD): responsiveness to change and English language validation.* Pharmacoeconomics, 2007. **25**(5): p. 385-96.
31. Wiklund, I.K., et al., *Quality of Life in Reflux and Dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD).* Eur J Surg Suppl, 1998(583): p. 41-9.
32. Carre, J.M., et al., *Changes in testosterone mediate the effect of winning on subsequent aggressive behaviour.* Psychoneuroendocrinology, 2013. **38**(10): p. 2034-41.
33. Carre, J.M. and C.M. McCormick, *Aggressive behavior and change in salivary testosterone concentrations predict willingness to engage in a competitive task.* Horm Behav, 2008. **54**(3): p. 403-9.
34. Buss, A.H. and M. Perry, *The aggression questionnaire.* J Pers Soc Psychol, 1992. **63**(3): p. 452-9.
35. O'Connor, D.B., J. Archer, and F.W.C. Wu, *Measuring aggression: Self-reports, partner reports, and responses to provoking scenarios. .* Aggressive Behavior, 2001. **27**: p. 29-101.
36. Tanaka, M., et al., *Fatigue-associated alterations of cognitive function and electroencephalographic power densities.* PLoS One, 2012. **7**(4): p. e34774.
37. Tanaka, M., et al., *Effect of mental fatigue on the central nervous system: an electroencephalography study.* Behav Brain Funct, 2012. **8**: p. 48.
38. Craig, A., et al., *Regional brain wave activity changes associated with fatigue.* Psychophysiology, 2012. **49**(4): p. 574-82.
39. Alvarez, J., et al., *The effect of EEG biofeedback on reducing postcancer cognitive impairment.* Integr Cancer Ther, 2013. **12**(6): p. 475-87.
40. Le Bon, O., et al., *Ultra-slow delta power in chronic fatigue syndrome.* Psychiatry Res, 2012. **200**(2-3): p. 742-7.

41. Butt, Z., et al., *Measurement of fatigue in cancer, stroke, and HIV using the Functional Assessment of Chronic Illness Therapy - Fatigue (FACIT-F) scale.* J Psychosom Res, 2013. **74**(1): p. 64-8.
42. Kosinski, M., et al., *Qualitative validation of the FACIT-fatigue scale in systemic lupus erythematosus.* Lupus, 2013. **22**(5): p. 422-30.
43. Cohen, S., T. Kamarck, and R. Mermelstein, *A global measure of perceived stress.* J Health Soc Behav, 1983. **24**(4): p. 385-96.
44. Golden-Kreutz, D.M., et al., *Assessing stress in cancer patients: a second-order factor analysis model for the Perceived Stress Scale.* Assessment, 2004. **11**(3): p. 216-23.
45. Kennedy, S., J.K. Kiecolt-Glaser, and R. Glaser, *Immunological consequences of acute and chronic stressors: mediating role of interpersonal relationships.* Br J Med Psychol, 1988. **61 (Pt 1)**: p. 77-85.
46. Kiecolt-Glaser, J.K. and R. Glaser, *Psychoneuroimmunology and cancer: fact or fiction?* Eur J Cancer, 1999. **35**(11): p. 1603-7.
47. Kirschbaum, C., K.M. Pirke, and D.H. Hellhammer, *The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting.* Neuropsychobiology, 1993. **28**(1-2): p. 76-81.

1

2

3

4