Real World Evidence for A.I.-Driven Coaching Chat-Bots

Jean-Luc Mommaerts

Bold = copied from article

PCT = pragmatic clinical trial
RWD: real-world data
RWE: real-world evidence
CCB: coaching chatbot

1 Intro

1.1 This article in bullet-points

- For several years, big healthcare payers have become more skeptical about the real-world value of RCTs.
- Delving into the lack of blindness in double-blind studies may enhance this substantially.
- Worldwide, more attention goes to pragmatic trials and RWE. There is a vast increase of articles about this in PubMed, at present +/- 1000 per year.
- The FDA is on the same track, working out guidelines for such.
- Such developments also have drawbacks. Mainly, they tend to relax rigorous science. Many are on guard for this.
- Psychotherapeutic studies are notoriously difficult to put in RCT format. Among other things, we see a *reproducibility crisis* in the highest-level studies of past decades.
- Meanwhile, mental health problems are surging worldwide, with tremendous cost (even before COVID).
- Chatbots are getting a lot of attention in many fields, including healthcare.
- A well-designed coaching chatbot amasses lots of valuable data, formalized and on-line.
- This can be enhanced thoroughly through Artificial Intelligence.
- ***We are developing an ambitious coaching chatbot, A.I.-enabled, on psychological and many psycho-somatic (functional) domains.***
- This brings the possibility of performing many pragmatic trials simultaneously, with a maximum score of pragmatism on the PRECIS-2 tool.
- At the same time, science on these real-world data enables formalization, possibly to the degree of an RCT.
- This may unearth new insights into the influence of the psyche on soma.
- These insights can be put into practice immediately and on a big scale.
- One can also envision using a coaching chatbot in pharmaceutical trials and non-mental medical practice to alleviate the burden on physicians.
- This can all lead to better health for many people and substantial cost savings for society.

1.2 Short-text

We envision an A.I.-driven coaching chatbot (CCB) that can 'talk' with many people simultaneously according to a specifically describable, proprietary technology of coaching (subconceptually humanistic). This CCB is A.I.-based: natural language understanding, speech synthesis, ontology formation, multiple rule bases, ANN technology, on-line reinforcement learning. We want to investigate to what degree the CCB is effective. For this, we are considering using pragmatic clinical trials (PCTs). Also, we foresee the availability of big data, thus a timely answer to payers' demand for RWE. We want to manage these scientifically and ethically from the start. We see specific hurdles and opportunities. The setting is very different from one in which a double-blind study is appropriate.

There will be a need to finetune a scientific methodology. Different CCBs will have different backgrounds. Thus, we will need a scientific investigation of CCBs' effectiveness in general and of each CCB individually. The effects of one are not the effects of another. Moreover, each CCB will keep

evolving, even due to its scientific data. Thus, the result of any scientific investigation are readily obsolete. This poses particular problems.

1.3 What is RWE?

In 2014, the National Library of Medicine (NLM) at the National Institutes of Health (NIH) created a new Medical Subject Heading 'pragmatic clinical trial.' However, there is a continuing debate about what constitutes a pragmatic trial, how it is distinguished from more traditional explanatory trials, and what are proper strategies or tools for designing and describing pragmatic trials. (1) Generally, a pragmatic study should fulfill at least two fundamental features (2):

- Its conduct should resemble usual clinical practice.
- The results should be applicable to multiple other settings (than the one lab of typical RCT) - done at several sites

A distinction may clarify (3):

- Explanatory trials are about ‘efficacy.’ They are primarily carried out in phases II-III of medical trials. They are about what can be, ideally, the effect of a medical product or procedure.
- Pragmatic trials are about ‘effectiveness.’ They are about what typically happens in the daily reality of medical care, leading to the following trial features:
    - ‘Management’ (experimental and comparison intervention) is flexible.
    - Participants = all people with the condition of interest
    - Cross-over permitted
    - No blinding
    - Unobtrusive measurement of subject compliance; no particular strategies to improve compliance
    - Unobtrusive measurement of physician adherence
    - All usual care providers are involved
    - Investigated outcome: broad overall health effects, including mortality, morbidity, functional status, well-being, and resource use [CCB9]
    - Physicians involved: also ones with low scientific expertise
    - Follow-up not with exceptional frequency
    - Primary analysis is of all participants who entered the trial (no ‘non-responders’ or ‘non-compliants’).

In theory at least, there is a continuum between both. Image from CCB1:

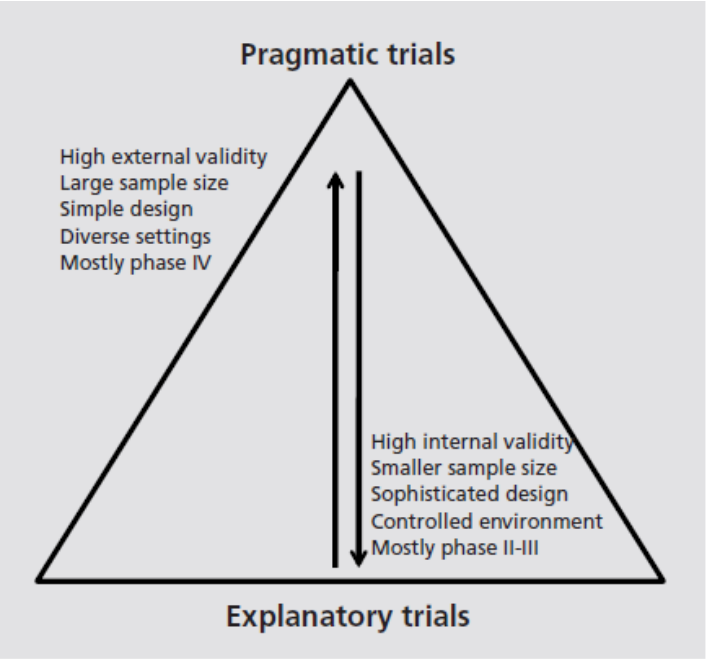

Figure 1. Schematic of the relationship between explanatory and pragmatic trials. The wide base of the pyramid depicts the relatively higher proportion of explanatory trials.

1.3.1 On big data

In a 2001 research report, Gartner analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional: increasing volume of data, velocity of data in and out, and variety of data types and sources. This requires new forms of processing to enable enhanced decision-making, insight discovery, and process optimization. (4) Much of the industry still uses this 3-V model for describing big data. Additionally, new Vs (for 'Veracity,' 'Value,' and 'Variability') have been added by some organizations.

Many RWD big-data sources are not designed for research purposes; therefore, the accuracy and reliability of data collected from these sources may be unknown. (5) Researchers must know how RWE can be used most effectively.

Big data RWE studies are intended to complement, not compete with, or replace RCTs.

CCBs may also provide big data. The quality of data may be very variable.

The ideal data source to generate RWE needs (6):

- **Patient population representative of those with the underlying condition (e.g., demographics, clinical comorbidities)**
- **Prospectively planned**
- **Continuously updated with minimal resources**
- **Longitudinal follow-up**
- **Rich clinical data: clinician-entered, patient-reported, and patient-generated**
- **Quality control measures in place**
- **Integrated within existing data systems**

1.3.2 Success but no success yet

The call for more pragmatism is not new. Differences between explanatory and pragmatic attitudes in trials were highlighted 50 years ago. [CCB27] The first article introducing the concept of pragmatism was published in 1967. [CCB1]

In recent years, the appreciation of the value of RWE has become more widespread. There are an ever-increasing number of publications presenting RWD/RWE: 930 publications in 2018 used RWE or RWD in either the title or abstract on PubMed, compared to 501 in 2017 and 326 in 2016. [CCB69] This interest is due, in part, to a 'voltage drop' in effectiveness as interventions tested in explanatory trials move into real-world situations. [CCB3]

Labeling an RCT as 'pragmatic' has become almost a badge of honor. [CCB2] As a result, trials with non-pragmatic features have also been tagged as pragmatic, even in major medical journals such as BMJ. Still, good pragmatic trials are seldom carried out. A serious impediment to using RWE is healthcare providers' lack of knowledge on understanding and interpreting RWE. [CCB3] Also, despite the various advantages of RWE, only a few guidelines currently use RWE to guide clinical practice recommendations. [CCB38] Given trials' social value, costs, and limited research resources, this situation is possibly unethical.

In the US, the value of RWE is recognized increasingly by the FDA. Same thing in China: RWE has gained wide attention in the past few years, and there are a few critical governmental initiatives. [CCB39] Yet, the use of RWE for healthcare practice and policy decisions is currently limited.

1.3.3 FDA and RWE

In December 2016, the US Congress passed the 21st Century Cures Act. Part of the aim was to establish a program to evaluate the potential use of RWE. It encouraged greater use by the FDA of RWE not obtained through RCTs. [CCB40] The 21st Century Cures Act says the FDA must draft guidance about acceptable standards and methods within five years in consultation with other stakeholders.

At this time, FDA regulatory scientists published a paper on RWE, clarifying that when the term RWE is used, the primary attribute that distinguishes it from other kinds of evidence is related to the context in which the evidence is gathered. Namely, in clinical care and home or community settings as opposed to research-intensive or academic environments. The distinction should not be based on the presence or absence of a planned intervention or the use of randomization." [5 in CCB60] According to the FDA's framework, RWE can also describe randomized trials in the clinical trial setting. These are typically very large RCTs and are referred to as PCTs. Likely, RWE can also be used by public health planners and payers to estimate the effect of a product or intervention on a large population. [CCB61]

Meanwhile, the FDA MyStudies App is designed to facilitate patients' RWD input directly. It can be linked to electronic health data supporting traditional clinical trials, pragmatic trials, observational studies, and registries. [CCB60]

2 Why RWE

> "Pragmatism in clinical trials arose from concerns that many trials did not adequately inform practice because they were optimized to determine efficacy. Because such trials were performed with relatively small samples at sites with experienced investigators and highly selected participants, they could be overestimating benefits and underestimating harm. This led to the belief that more pragmatic trials, designed to show the real-world effectiveness of the intervention in broad patient groups, were required." [CCB7]

2.1 Pragmatic reasons

An ideal trial includes a relevant population for the intervention, a control group treated with an acceptable standard of care, and meaningful outcomes. It must be conducted and analyzed at a high-quality standard. [CCB7] However, this ideal is not reached for most new treatments. On top of this, well-conducted RCTs are expensive. Additionally, there is much waste (85%). [CCB26] The RCT is still the design of choice to know which healthcare intervention is most effective. Randomization, however,

does not promote the applicability of the results to situations other than those of the trials. [CCB6] Indeed, RCTs on medicines before they are licensed (or assessing a new indication or dosage form) could hardly be pragmatic since they need to comply with clinical trial regulations bearing no resemblance to their subsequent use in routine care. [CCB2] Many regulatory requirements almost invariably prevent pre-approval trials from being pragmatic, disrupting any intention to mimic standard practice. [CCB2] The complexity of phase III trials has increased over time, thus moving away from pragmatism. This happened in the US mainly between 2001-2005 and 2010–2015.

Thus, the clinical evidence is insufficient to guide physicians and policymakers in choosing optimal medical treatment. The little systematic evidence has indicated the lack of external validity in RCTs. [CCB1] The conventional paradigm of EBM implies that the healthcare of individual diseases is based on the best available evidence. However, this model is not directly applicable for caring for complex patients and generally is no longer sustainable in the case of chronic diseases. [CCB54] Not surprisingly, guidelines based on RCTs are often disregarded in real life because of physicians' inappropriate prescribing and patients' marginal adherence. [CCB54] This all makes for increased doubt among healthcare policymakers as to whether the plethora of available evidence and ongoing data are translatable and usable in real life. Pragmatic trials might fill this gap, providing evidence on the relative effectiveness of a treatment in routine clinical practice rather than generating proof of the direct pharmacological or biological effect of the intervention. The idea is that this can be done in an early development phase while maintaining the strength of RCTs. [CCB8]

2.2 Internal/external validity

There are two types of validity in observational research:

- **Internal validity**: the study measured what it set out to measure. Thus, the ability to determine cause-effect relationships in a clearly defined population group. This is vulnerable to
    - **Selection bias:** absence of comparability between studied groups,
    - **Information bias:** incorrect determination of exposure or outcome — e.g., information being gathered differently between groups.
    - **Confounding:** relating an exposure to an outcome but measuring the effect of a third factor.
- **External validity (= applicability - whether an intervention works in real life)**: the ability to generalize from the study to many patients.

In short, we want the highest possible internal validity in combination with the highest possible external validity. But we need to make a choice:

	internal validity	external validity
RCT	+	-
RWE	-	+
CCB	+	+

Although RCTs, if correctly applied, lead to well-controlled trials with statistically credible results, the applicability of these results to real-life practice may be questionable. The external validity is low. Moreover, the same characteristics that contribute to the high internal validity of a trial (well-defined inclusion and exclusion criteria, blinding [CCB20], controlled environment) can hamper its external validity — the ability to generalize the results. The RCT trial can only prove the trial situation. [CCB17]

The main aim of pragmatic trials is to make results generalizable through higher external validity, as this is important for decision-making by policymakers, clinicians, and patients about a healthcare policy or practice. Therefore, key characteristics are broad eligibility criteria and patient-centered outcomes. [CCV27] In testing the effectiveness in routine clinical practice, in the full spectrum of everyday clinical settings, pragmatic trials should also try to retain internal validity. Pragmatic trials should measure a wide range of outcomes, mostly patient-centered ('patient-centered medicine'), with maximal heterogeneity in all aspects: patients, treatments, clinical settings, etc. They must be significant to attain this. They should mimic as closely as possible the users, settings, and circumstances in which the interventions under evaluation will probably be used. [CCB31]

2.3 Open vs. closed world

The difference between internal and external validity is relevant because of the difference between a *closed world* and an *open world*. An open world is one of complexity. Note the fundamental difference between complexity and complicatedness. Complicatedness (for instance, an airplane) is entirely conceptual, while complexity (any living organism, deep neural networks) is not. In complexity, emergences appear that are not tractably explainable by examining the separate elements.

The closed world assumption is perfect in a world of Newtonian physics, but the real human world is not *pragmatically* reducible to that. An open (= open-complex) world can, principally, not be well understood with conceptual means. An RCT strives to leave out complexity in a bid for conceptual control. That can only be done in lab circumstances. A clinical lab is an artificially closed world. The difference with the real, open world may be much more significant than many researchers imagine. Thus, change one element in the study design – even one that is difficult to discern – and the results may vary considerably. Humans are to be treated as open-complex systems — with no crisp borders and in continuous communication with the outside world. This demands pragmatic studies.

In theory, one can build a ladder to the moon — not in practice ('intractability'). Likewise, in theory, the behavior of an open-complex system, such as the human mind, is reducible to the behavior of its elements. In practice, one needs a computer more extensive than the universe to make all possible mental state changes explicit. Also, there is no one 'management path' to be chosen. *Every* user has a different path.

The classic clinical method of experimentation presupposes:

- Closed world assumption: clear borders of the experiment and elements within, with as little noise as possible. This means, in research: clear inclusion/exclusion criteria and crisp concepts all the way through.

  >< What if the real borders and concepts cannot be made clear and crisp? What if treating them as closed only brings more confusion?
- Limited number of experiments -> extrapolating the results.

  >< What if 'noise,' with a slight difference at the start, makes a significant difference on the road? Extrapolation cannot take this into account.

2.4 Other advantages of RWE studies

Some authors contend that "although real-world studies do not adhere to the same degree of controlled conditions and predefined patient management strategies as do conventional clinical trials, analyses resulting from these studies can be held to a high degree of validation and standardization, making them as meaningful as those from RCTs." [CCB33] A 2014 Cochrane review of the literature concluded that observational studies usually generate similar average treatment effects to those reported in clinical trials even when no attempt has been made to mirror the inclusion and exclusion criteria of the RCTs. [CCB56]

Additionally, "the unobtrusive collection of trial outcomes is attractive; it reduces the burden on the participants and investigators without introducing artificial aspects to follow-up." [CCB7] For instance, the absence of inclusion/exclusion criteria is a common feature of registries, enhancing their applicability to broader populations.

The idea that value can be derived from understanding how people feel and function in their everyday lives is finding its way to healthcare delivery, payment reform, product development, and regulatory decision-making. [CCB54] In traditional cost-utility analyses, this translates into the health value being based on population averages, calculating quality-adjusted life expectancy. [CCB43] Contrary to this, the pragmatic notion links in much better with the narrative of patient-centric care and the increasing involvement of patients in the drug development process, aiming to increase the relevance of new therapeutic interventions to the patient experience. [CCB32]

Some other advantages of PCTs versus RCTs:

- They are more time- and cost-efficient.
- They are relatively little labor-intensive.
- They provide more data on interactions with concomitant diseases or therapies.
- They provide more data on real-world adherence.
- They provide long-term safety data for unselected populations.
- They may cover various designs, from observational studies to interventional trials.

2.5 Drawbacks of RWE

The features of pragmatic trials that support the generalizability of their results to real-world practice may also limit the interpretation of results. These include heterogeneous patient populations in which some may not have the condition of interest. There is also a lack of blinding, absence of a placebo group, and suboptimal adherence to therapy. [CCB7] Thus, there is a risk of many low-quality studies nestling themselves in the RWE niche. An intense reproducibility crisis may lie ahead. Concerns that RWE is lowering the evidence for safety and efficacy should be addressed. [CCB41 Big pharma may influence consumers and physicians to heighten the placebo effect in real-world circumstances. RWE can be an additional tool to sell placebo-in-disguise, including CAM. Therefore, RWE should be accompanied by a clear ethical choice for non-placebo. Classic RCTs remain essential in this regard.

Unique personal distinctions are diminished if the RWE study combines different real-life groups to be more generally applicable. The results of the study may apply only to an exemplar subject. If, on the other hand, the RWE study distinguishes between many groups, then we get complex results that are not useful for clinical guidelines or individual clinical care. A CCB may solve both problems within its management (coaching) of users.

Major funders of clinical research (National Institutes of Health…) do not focus on supporting PCTs. The addition of such research within the health care delivery system demands a substantial increase in public and private funding. However, CCBs are at an advantage. Still, who should pay for these trials? Regarding drugs and devices, industry representatives may think they have already fulfilled their role in getting a drug to the registration stage.

Other problematic issues in RWE:

- Little standardization
- Bad data. Data should be reliable and complete. There is no escaping the ironclad rule of garbage in, garbage out. [CCB42] The complexity and robustness of data will vary for each source. This underscores why some healthcare providers and guidance committees have long-standing concerns. [CCB38]

- Big data may heighten cultural bias — for instance, a biased conviction of little mental influence on health. This may heighten medicalization.
- In healthcare, data is sticky. It tends to stay where it originated and does not get shared even when this would benefit patients, providers, and payers. [CCB41] There is also a lack of interoperability of databases (e.g., between electronic health records and analytical databases)
- What are the essential study elements to be reported?
- Watch out for 'weapons of math destruction.'
- Essentialism
- An extra burden to physicians. Pragmatic trials are prone to selective and low participation rates in a primary care setting as these sites may not be fully equipped or dedicated to supporting a clinical trial. [CCB8]
- Diversity in populations -> how to report about this?
- Social pressure on subjects to comply to the therapy. Patients may receive benefits from the health system obliging them to participate in research.
- 'Following everyday practice' = very diverse
- Huge diversity (in blinding, population, physicians, outcomes, protocols…) -> Any of this can skew the results.
- Is meta-analysis of such trials possible?
- Really 'pragmatic' = much larger studies necessary; large patient populations are needed.
- Uncertain availability, completeness, and coding variability of outcomes at follow-up
- Uncertain accuracy (confounding factors) and reliability (bias)
- Uncertain adherence to medications
- Uncertain methodological and analytical expertise, also of process optimization and quality of design
- Uncertain quality of results
- Uncertain adherence to privacy laws

Also, some ethical issues are raised about pragmatic trials [CCV27]:

- **Are activities in pragmatic RCTs research or practice?** How to satisfy current research regulations?
- **What level of oversight is required?** Some advocate that low-risk pragmatic RCTs do not require more stringent oversight than clinical practice.
- **Which study designs are appropriate for pragmatic RCTs?** They may favor novel designs, including cluster randomized, stepped wedge, registry, and cohort multiple designs
- **Who are the research participants in pragmatic RCTs?** This influences research ethics review, benefit–harm analysis, and informed consent procedures.
- **From whom, how, and when is informed consent required?** It's unclear what aspects of research must be disclosed.
- **Who are the gatekeepers in pragmatic RCTs, and what are their responsibilities?**
- **How should harm–benefit analyses be conducted in pragmatic RCTs?** For instance, in case of delayed implementation of an intervention. How should we identify and evaluate relevant risks and benefits to individuals or groups within a pragmatic RCT? What criteria should be

used to classify a pragmatic RCT as minimal risk? What evidence is required to substantiate a claim that an intervention or comparator is usual care and if this would constitute minimal risk? [CCB31]

- **How ought vulnerable groups to be protected?** Lower socioeconomic strata, children, pregnant women, prisoners.

There are risks to using big data, such as loss of patient confidentiality or misuse of data by insurers or other companies. In *Privacy in the Age of Big Data*, Payton and Claypoole stated that the biggest threat to individuals is from cyber-criminals, as most databases can be hacked, even secure government databases such as the Pentagon's. [24 in CCB43] Also, in *Obamacare* healthcare reforms, health insurance markets were not allowed to consider pre-existing conditions in pricing policies for individuals. The reason was the recognition that companies will soon be able to accurately predict healthcare costs by using big-data applications. [CCB43]

2.6 Combination RCT - RWE

RCT research, characterized as having the highest reliability in confirming a physiological or clinical hypothesis, and RWE research, which reflects the actual clinical aspects, can have a mutually supplementary relationship, forming the most potent evidence-based research method in medicine. [CCB72] There is a need to develop a hybrid trial methodology combining the best parts of both and that provides adequate scientific evidence for regulatory decision-making. [CCB32]

Bringing together the restrictions around RCT methodology and the breadth and strengths of RWE, it may not be until an RCT can be conducted within RWE trials that we see RWE being routinely used in decision-making beyond that of rare diseases or expansion of existing product labels. [CCB38]

3 PRECIS-2

3.1 What is PRECIS?

Some see 'pragmatism' as an attitude rather than a sum of specific characteristics. [CCB17]

The working group on pragmatic trials developed the Pragmatic-Explanatory Continuum Indicator Summary (PRECIS). This counted ten domains and was challenging to apply. It was further refined and validated as PRECIS-2 [CCB3], in which the requirements for pragmatism loosened substantially. [CCB7]

Currently, pragmatic trial investigators in clinical research, submissions of RCTs to funders, research ethics committees, and peer-reviewed journals should include a PRECIS-2 tool assessment done by the trial investigators. [CCB2]

Table 1 PRECIS-2 tool nine domains and scoring method [12]

Domain	*Comment*
Eligibility	Who is selected to participate in the trial?
Recruitment	How are participants recruited into the trial?
Setting	Where is the trial being done?
Organisation	What experience and resources are needed to deliver the intervention?
Flexibility: delivery	How should the intervention be delivered?
Flexibility: adherence	What measure are in place to make sure participants adhere to the intervention?
Follow-up	How closely participants are followed-up?
Primary outcome	How relevant is to participants?
Primary analysis	To what extent are all data included?

Score (each domain): from 1 to 5 using a 5-point Likert scale
1 = very explanatory, 3 = equally pragmatic and explanatory, 5 = very pragmatic

3.2 Hurdles

In one study, PRECIS-2 was used to assess five PCTs. Raters often struggled to use the tool, and significant differences in inter-rater reliability are observed despite having access to detailed study information and formal training. [CCB3] Its widespread application would require ongoing training in using the tool for both trialists and reviewers. Refinements would be helpful for raters, including creating additional rating criteria linked to numerical rating and exploring different formats for brief training of raters. [CCB4]

Moreover, as a trial design may change from initial concept to execution, the degree of pragmatism may also change. [CCB3]

PRECIS-2 has proved helpful for 'framing the conversation' about trial design among members of the Pragmatic Trials Collaborative Project. The authors note that design elements assessed by the PRECIS-2 tool may represent mostly stable decisions. Overall, there has been a positive response to using PRECIS-2 to guide conversations around trial design. [CCB5]

4 About CCBs

4.1 What and why?

We envision, in abstract terms, a psychological well-being-related CCB that will be a combination of coaching and meditative mental exercises. We do not see this CCB as ‘therapeutic’ but as ‘mental hygiene.’ People do not need to see themselves as ‘ill’ to use the CCB. This accords to a profound philosophy of patient-centeredness.

Our findings apply to CCB in general, for instance, with coaching towards dieting, patient compliance, joint (patient & caregiver) decision taking towards the right kind of care, etc. This can be performed in long-term studies.

Medicine is, at the core, focused on curing disease and relieving symptoms. There is no ‘inner growth’ as a goal. With CCB-coaching – as with coaching in general – one is not limited. Symptom relief may

rather be seen as a side effect of the primary goal. Still, it can also be a goal in the short and long term, which should interest physicians and psychologists.

4.2 Comparing CCB to other kinds of management

Medication: Very good standardization is possible in clinical trials. In practice, people use their meds very irregularly. Also, they cannot be monitored as to their mindset. In RCT, it is seldom done. Nevertheless, expectations can be pretty influential on the effects and side effects. Medications can readily go into cheap mass production. Nevertheless, they keep being expensive on a retail basis.
Psychologists: Standardization is difficult — probably a significant cause of the reproducibility crisis. The training of psychologists is expensive. They need to be trained on an individual basis.
Devices: Most are pretty expensive in production.

4.3 Advantages of science to CCBs

Physicians may refer to a CCB, but they will – rightfully – only do so when they trust it. So, there needs to be good communication. In studies, physicians and psychologists should also be involved whenever possible. Their experiences with patients/clients who also use a CCB are equally important. Their general impressions of the CCB may influence their patients' well-being.

4.4 Advantages of CCBs to science, including A.I.

"Big Data + Artificial Intelligence = Big Knowledge."

A CCB may ask users all kinds of things. Part of this may be oriented towards scientific investigation goals or towards the goal of understanding a specific kind of user, using this knowledge in the same session or next time. Standardization is possible.

CCBs can be open to users, providing them with their personal results in numbers and graphs. We may also provide continuous statistical information over many users, in due respect to privacy. In addition, a CCB can be an active searcher of information, forming and validating hypotheses as part of the coaching. This brings science and direct management closer together. A CCB can profit from big data crunching. Big data brings a way to manage open systems such as humans. Through this, the constraints on scientific experimentation can be relaxed:

- working within a more open world
- less limitation on data gathering -> less extrapolation needed.

During the coaching, a CCB can also be used to ask the patient about her other care/intervention/therapy. This way, we can compare [only CCB] to [CCB + other intervention(s)]. In case of positive results, for the sake of argument, in a longitudinal study:

> **IF** CCB gives equally good results with or without other intervention (no difference in the result of [only CCB] to [CCB + other intervention(s)] as we can investigate)
>
> **THEN** only CCB works.

A CCB can be used to heighten compliance in filling questionnaires for another therapy or for CCB itself.

CCB-RWE produces results that can be generalized. If they can be explained in human-understandable form, they can also be applied in routine practice settings.

A CCB can quickly go international in its related studies, finding cultural commonalities and differences. This is interesting to science while being also what payers want.

A frequent argument contra RCT is that it makes medicine too rational, disrespecting human factors in complex circumstances, such as chronic disease. With Lisa, we combine rationality and depth.

In being an active knowledge gatherer with many challenges to overcome in real-world, Lisa can get at the forefront of A.I., developing technology based on A.I insights and human cognitive psychology in proprietary ways.

Since a CCB will incorporate more and more A.I. technology, this can also be used for data-mining the results. Part of this is already done during the coaching. It can be augmented for the sake of pragmatic studies. This way, the coaching and the continuous PCT become one integrated happening. Also, whereas RCT data are used to test hypotheses, neural network models can be used to generate new hypotheses and look for links that may not immediately be apparent — 'unknown unknowns.' [CCB43]

5 Reasons for 'going RWE' with CCBs, more specifically Lisa

5.1 General

RWE is difficult for medication and devices: no blinding, placebo issues… In psychological interventions, it is even more difficult. Contrary to this, it is natural for CCBs to go for RWE while remaining standardized as much as possible.

Differentiation between spurious, indirect, and causal associations can be difficult. Criteria such as temporal sequence, strength and consistency of an association, and evidence of a dose-response effect lend support to a causal link. By managing these, Lisa can come closer to causality.

Moreover, RWE is about real-world people, including non-conscious processing, in mind-body unity. This is Lisa's habitat.

5.2 List of reasons for 'going RWE' with CCBs

In the case of Lisa:

- **A double-blind study is impossible**. People readily see what they are using. Lisa will be embedded in an environment in which people get much information at request. Openness in all this is part of Lisa methodology, which is based on principles of empathy. 'Blindness' is just intrinsically opposed to this.
- **Even an RCT is impossible** since the CCB will be widely available and cheap. This means that people in any no-treatment group have access quickly and without our knowledge, especially when using different log-ins. We cannot know the degree to which this may skew the results. Our openness makes this especially difficult.
- **We will change individual elements** in the CCB and see how this affects the effectiveness. This can be done on an ongoing basis. It enables us to investigate elements of effectiveness.
- **We will continuously improve** the CCB, also during the period of investigation. Improvements can be made in many details and more significant issues.
- **We will evaluate medication use, symptom reduction, and many more parameters.** This is feasible in an ongoing 'relationship' with the CCB, but not so much in a clinical trial.
- **We want to know the effectiveness within a general population** of patients (or 'users') in real life. Especially since the CCB has to be available stand-alone, we want to know whether people use it and are happy using it.
- **We will compare stand-alone use of the CCB vs. embedded use** (in combination with psychologists or physicians).
- **We will use the CCB on many domains**, health-related and more. To set up many traditional clinical studies would be expensive. PCTs within Lisa-setting lend themselves more readily.

- **We will put data from different domains together.** For instance: health-related data and contacts with the judicial system. This can also be done in the long term.
- **We will start prospective studies.** We think this will uncover a lot of influence of the mind on the body. For this, until now, there has been a lack of rigorous methodology. We believe the lack of exploring mind-body influences in the long term is highly costly to individuals and society.
- **We want to take full advantage of the CCB being computer-based** (also on smartphones and tablets). This asks for specific methodological developments. For instance, we can standardize the underlying principles of coaching much more than in human coaching.
- **We will perform continuous monitoring**, not at one time. This way, we will be able to see fluctuations in time.
- **We will follow up on our monitoring in the long term**, even for life. Of course, the 'users for life' are a specific subset. Nevertheless, interesting results may appear and indicate the need for more in-depth investigation.
- **We want to be able to investigate people from different cultures** and compare results.
- **We will perform our investigations simultaneously and on the same population as the 'active live users.'** For this, we need to integrate our pragmatic investigation into the flow of using the CCB as little intrusively as possible. Ideally, the questionnaires form part of the users' helpful experiences.
- **We envision many users, possibly millions.** This means we will need to manage big data with appropriate tools for the sake of the coaching as well as the investigation.
- **We will extract from our scientific data what is helpful to enhance the product.** We see this as an intrinsic part of development.
- **We want to make our clinical studies generalizable and individualizable (meaning: is it good for this** *specific person*?) The latter is profoundly 'patient-centered management.' A classical RCT suffers in both directions: extrapolation of the results and the applicability of the statistics to the individual case are notoriously hazardous.

In all this, one can see the relevance of specific methodological issues. We intend to pinpoint these as clearly as possible.

5.2.1 Opportunities

- Pragmatic trials are mainly done with one specific intervention as a focus, testing this in real life and comparing it with other interventions. Another goal may be to do a pragmatic study of 'treated people,' with no focus on any intervention or therapy. In this case, a CCB can be used to support the patient in choosing his preferred intervention and in follow-up.
- Since external validity is close to 100% (no extrapolation) and with continuous evidence flow, studies may be done prospectively in a 'retrospective' way within Lisa-knowledge. Specific subpopulations can be studied this way for specific kinds of decision-makers.
- A CCB can be linked to other data for in-depth cost comparisons.
- New patterns may be discovered and the relevance of existing ones investigated, especially in the mental domain.
- We can delineate more robust parts of knowledge within our knowledge repository.
- Integration is possible within Lisa-coaching of validated questionnaires on medication use, work, QoL, and providing information for decision-makers (payers, guideline developers, clinicians). This is also relevant for in-depth views in cases of possible benefit.
- Reproducibility +++.
- Lisa can do A/B testing on (blinded) groups. She can also 'act as-if blindly.' Thus, many small RCTs are continuously possible, more or less in double-blind. [Q: do users need to know this?]
- We can set up truly prospective studies (e.g., to test hypotheses).

- Some cons of RCTs are avoided, including their lack of double-blind (cf. STAT-article).
- No need for many trained researchers. Our research is centralized (>< many different CCBs would make this difficult)
- In case of a new indication, we can quickly build up evidence.
- We can show science results on the fly to users, building positive branding and motivation to use Lisa (public, physicians, coaches, reimbursement).
- We can ask questions in the long-term, also of people who have no subscription (?).
- We can combine our data with continuous physiological data trackers (IoT).
- We avoid the messiness of other RWE.
- Personalized care data can be analyzed and combined.
- Back-track studies: looking for patterns in people who fare better/worse
- We can educate users to take care of deep-mind health.
- We can focus on specific domains, such as cardiovascular risk -> Start with a questionnaire; offer specific sessions to a group; continuously look at results.
- We can release data for other researchers (not commercially!)
- We have lots of structured data and can do NLP on unstructured data.
- We can study the positive effects of Lisa on health and other effects on the change of behavior of users towards Compassion (?) through specific questions.
- We can see trends in QoL in specific subgroups and send alerts, e.g., when more burnout appears in one company.
- 'Precision-management' is possible.
- 'Organ-based medicine' can be surpassed while pain patterns can be seen more broadly — also clusters of pain with depression, anxiety… This may influence medical diagnosis and therapy.
- CCB may lead to fewer clinical investigations, therapy, and hospitalization. Thus: huge cost reduction.
- If there is a positive influence of working on one field towards effect on another field, this can be shown.
- Clarifying some disease etiologies: the importance of psychological factors above others.
- This may be the only valid source of evidence for mental influences in contrast to RCTs and other placebo-prone pragmatic trials.
- One can look for fuzzy patterns in effectiveness and risks.
- One can flexibly adapt to payers' needs for quality of data and evidence.
- The cost of science in CCB is low, even with long trials (>< medication). In case of failure, we can immediately enhance Lisa and restart.
- In principle, all elements of the 'ideal data source' for RWE are present.
- There may be much placebo in Lisa, but at least openly.
- One may work with a Total Cost of Care (TCoC) estimator to show economic benefit in cohorts, asking users specific questions in this sense.
- Many hypotheses can be generated about relevant patterns in the human condition, including patterns in the long term and with consequences also in the long term.
- One can perform studies about the impact of specific occurrences on mental health.
- There is less need for randomization since we bring the 'complete' reality into the picture.
- One can do hypothesis-generating case-control studies [= selection on the outcome, then looking at exposure] e.g., for rare outcomes.
- One can study possible confounders and automatically avoid them in future studies.

- PRECIS-2 (aiming to inform patients, clinicians, and policymakers) is meant to be as naturalistic as possible. Since CCB coaching IS the CCB study, Lisa is also as naturalistic as possible. Lisa scores on 'naturalism' with +/- 100%. Comparing Lisa to PRECIS-2, we see that Lisa scores positive on all dimensions.

5.3 Challenges

- A CCB keeps changing to serve users better, based on the science itself. Thus, science necessarily runs after the changes.
- Sloppy users might diminish the accuracy of the system, willingly or unwillingly.
- Sources of bias: not everyone is on Lisa; not everyone provides data to Lisa; not everyone provides good data.
- People use Lisa in many ways. The 'best way' is taken by those most motivated (= bias).
- Cyber-criminal theft of data and other privacy issues need to be considered.
- People may feel pressured to use the CCB or may feel continuously being monitored.
- The CCB might get stuck in spurious associations and irrelevant patterns.
- Questions about scientific rigor.
- People can feel manipulated.
- Users legally possess the data, but what with the knowledge?
- A CCB makes visible people's responsibility in health and illness. Are people ready for this? Also, it shows that present-day care may not be the best in many circumstances. Are practitioners ready for this?
- Which agency should control studies?
- Risk of data trawling.
- Ideally, the complete outcome should be studied. In an open world, this is not possible since it also involves outcomes in an indefinite future and in undefinable ways.
- Should a CCB-RWE, to be deemed scientific, include a clear *a priori* statement of a hypothesis to be tested or research question to be answered, and advanced analytics to be set out in advance of the study commencing, ideally in a published protocol?

6 Methodology

6.1 Double-blind studies?

Looking at the user-Lisa encounter, no rigorous double-blind studies are possible. The user knows when he uses the CCB or not. However, if we break the encounter into small elements, Lisa can do continuous A/B testing without the user being aware of this. With many users, Lisa can form A and B groups over any element, thus forming, on the fly, an RCT. This is 'blindness' from the user, even to whether or not he is being tested. The latter is a 'hidden administration' test. A specific A/B element can be whether Lisa uses a specific coaching technique or not. Of course, A/B testing can be performed on any element or combination of elements. Thus, continuously, a lot of 'experimental science' can be performed, with or without user consent, or with 'user prior consent' whereby users notify that A/B testing may be done without their specific consent for each specific occasion.

We thus have four groups related to user-blindness:

	Lisa coaches (A)	Lisa does not coach (B)
Lisa says she coaches (A)	open coaching	placebo

Lisa does not say she coaches (B)	hidden administration	no-coach group

Whether or not Lisa herself is 'blind' is also peculiar. She is 'blind' inasmuch as she does not reckon with the information she gets from the trial. Using neural network functionality, this may be difficult: The expectancy that the user will have a positive result may be present in the network representation as a bias.

Lisa can bring any technique at any dose. She can heighten or lessen the dose in time. This way, she can combine longitudinal studies with RCTs.

6.2 Longitudinal studies

Many longitudinal studies can be done either or not in combination with RCTs. Performing longitudinal studies is part of Lisa's coaching. Generally, this will be done in the sense of an open dialogue.

Lisa may use single questions, VAS, or questionnaires as part of her coaching. After several times, she can show the user the results in a graph. She can also show the average of other users' graphs, for instance, to show the user what he might expect with an optimal commitment to mental exercises. This way, the user becomes a researcher of his own case.

Longitudinal studies can be done very broadly or in specific settings with specific goals.

Validated questionnaires are possible on specific topics (such as chronic pain) and, more generally, quality of life. They can be taken in one go or in parts. Not being intrusive is a specific requirement during coaching.

There are several options for longitudinal study design in CCB setting:

- Cohort designs follow a group of people who possess a specific characteristic over time to see if they show a particular endpoint or outcome.
- In case-control studies, cases of a particular outcome (such as a particular adverse event) are gathered and controls without. Then Lisa looks backward to see what proportion had an exposure or characteristic of interest.
- A case-cohort design is a statistical variant of a case-control study. Controls are sampled from a list of people, with each person having an equal probability of being sampled.

6.2.1 User consent

From the start, we will ask all users whether they want to provide anonymized information toward science and improve the CCB. If they disagree, we ask them why. Users can withhold consent, but in some such cases, the coaching will necessarily be of lesser quality. This is also scientifically important.

Ideally, we would have data from before prospects become users.

6.2.2 Study data

We store many data, verbal and non-verbal, conceptual and non-conceptual.

Domains of data for the study

- the personal domain: patient demographics, medical history, health status, and patient identifiers
- the exposure domain: patient's experience with the technology or device
- the outcomes domain: primary endpoints (very broadly, including many rewards for Lisa), secondary endpoints, adverse events, and technology deficiencies

Data may also come from many other sources: electronic medical records, laboratory records, billing records, institutional or organizational databases, administrative databases, death and birth records, and existing registry databases.
We also collect information about potential confounders, such as a drug being taken to treat the same condition.

6.2.3 Analysis, interpretation, presentation

In what format do we present the results?

'Registry studies' are observational. Physicians can treat the condition however they want. Effectiveness is measured. This refers to how well a device performs as intended in the general population of patients and the general chaos of clinical practice. Within a CCB, registry studies can be done continuously.

Are the results applicable to the total population? Do the people who want to use the CCB have specific characteristics – such as belief in a CCB? How can we know?

Can we use fixed external comparison groups, and do we need them? Comparison groups may be internal (data collected simultaneously), external (data collected outside of the registry), or historical (data collected not simultaneously). As an alternative, we can put all subjects on a graph from high to low frequency users. Of course, this is biased, but that does not yet make the bias pragmatically important. If we see that all people who use the CCB (thus, who *want* to use it frequently) are better off, then we can give this message to the public and reimbursement instances. If we can make people *want* to use it, they are better off. Such branding may also be *effective*. Scientifically, we still want to know *what* works.

'Side effects' may also be positive.

What are the appropriate methods of analysis of big data?

We can set up conditions for more internal validity (more specifically controlled studies, with specific questionnaires, more RCT-like), and other conditions for more external validity (more RWE-like).

What we like to see:

Effect(CCB) + Effect(other therapy) = Effect(CCB)

➔ Whether Effect(other therapy) is equal to or greater than placebo, the policy maker favors cheap CCB, e.g., over antidepressants or CAM

= Lisa as 'alternative to the alternatives'

No comparison with a single other intervention is needed. We just put Lisa on top of care as usual and compare this with 'only Lisa' or look at the evolution of other therapy usage.

Another possibility is that the other therapy (medication…) slowly diminishes in needed dose through using Lisa.

6.3 Long term follow-up

Adverse event detection, collection, and reporting in the long term are the same as adverse event reporting for any other kind of trial.

For as long as users stay connected (lasting memberships), long-term follow-up is possible with Lisa. Thus, we can probe for specific outcomes and more general long-term effects, such as less burnout,

depression..., and even less dementia. The more general ones may be relevant also if the user does not explicitly aim for them. A reduction of burnout may be a positive, unintentional side effect.

This can be done at no additional cost for Lisa.

6.4 Privacy issues, ethics, data ownership

Under 'targeted' research consent, pragmatic trials that are similar to standard care need only include minimal additions to consent for standard care. [CCB13] Furthermore, the principles of ethics, data ownership, and privacy within trial conditions are the same for clinical studies. However, privacy issues are significant for CCBs. They may pose challenges because people readily provide personally sensitive information in a context of trust. Security must be at optimal level.

We don't 'sell' data, not even anonymized ones. Information about human depth may lead malevolent parties to abuse this as nothing else, not only to the person from whom the data come but to everyone.

CCB13

6.5 A new scientific method

Eventually, this is no longer hypothesis-based experimental science. This is not 'new' in a destructive way. It vastly augments. Experiments make way for a much more life-like sampling.

Crisp concepts are discrete patterns of necessary and sufficient characteristics. Those are not the only relevant ones. More distributed patterns may be even more critical for health and well-being. However, viewed from an exclusive focus on pure concepts, they may seem noisy. The aim should not be to make less clear what is clear but to make the less clear as clear as possible, not just throwing it on the noise-heap. Big Knowledge has this potential. It transcends the limitation within science on conceptual crispiness. In fact, with Lisa fully operational, there is no limitation on relevant data gathering except through external conditions: privacy issues, computer power… Through this, the immense amount of noise in most domains of medicine and healthcare will generally turn into many distinct shapes. A mist will clear up.

7 CCB for all medical studies?

A CCB can guide people through any medical study. Especially when people are already accustomed to CCBs, this does not interrupt their daily life. One can envision a future in which this is the 'new normal.' Thus, the CCB used in a medication trial can already support people, such as for chronic pain. This can be part of the treatment-as-usual.

Then a new analgesic procedure can be introduced, guided by the same CCB. In this setting, the CCB can support users in the medication arm more than is otherwise possible. The CCB can guide the subjects, motivate them to be compliant in their medication intake, ask them for reasons for non-compliance, measure safety, and QoL, ask for side effects and while doing so, also perform STAT (serial treatment assumption testing). This way, the STAT questions are part of the flow.

Could this use of CCB enhance the placebo effect more in the medication arm than in the treatment-as-usual arm? Or less? Probably both, one or the other depending on the way the CCB relates to subjects.

Since CCB data can be analyzed formally, this use of CCB in any medical study can combine characteristics of a closed and an open world, thus of classic RCTs and pragmatic trials. This can be done with both kinds of trials rather apart from each other or intrinsically combined. It enables researchers to bring closer together efficacy and effectiveness data. Analyzing these data can also provide more insight into why there may still be a difference between both. The goal is 'controlled

pragmatic trials' (CPT). The results are bound to be, at the same time, more individual (patient-centered) and general (reproducible). Both are flaws of many medical studies at present, and medicine in general, especially mind-body-related healthcare. In this, on the one hand, small differences in study design may yield large differences in outcome. On the other hand, there is a lack of pragmatic usefulness of many RCTs. A CCB can lead to more CPT results. Methods (rules, visualization) will be needed to make this more useful to humans in clinical guidelines and clinical practice. With such methods, given all assembled and semi-standardized data, a more pinpointed consultation of these data is possible — for instance, results only including women of a certain age, comparisons of results, and trends over longer periods. In combination with data from the EHR, even more knowledge can be generated. Care pathways can be optimally evaluated.

Pharmacological studies in this vein can be done explicitly and implicitly (without a formal setting). They can be done at any time, also when chronic medication has been taken for a long time. A CCB can be used in long-term prospective studies. Note the possibility to ping people on their smartphones when it is time for a few standardized questions. Some other barriers to the conduct of clinical trials are also alleviated: huge costs, inadequate knowledge of clinical research by the physician, costly monitoring of patients, and inadequate infrastructure. CPT can be mainly accomplished at home without burdening the physician. The data can be sent to the electronic health record. In case of need, the patient can be referred to his physician for further consultation.

Given the advantages, CPT seems to be the way to go for all future medical studies.

8 Importance to Lisa

8.1 + Individualization of AURELIS sessions

We strive for higher acceptability by users who get a 'life performance' feeling from the idea that 'this is made for me.' It also brings the sessions closer to coaching, thus more a part of the CCB functionality, lending a more prominent role to Lisa.

A/B testing is also possible in individualized sessions, searching for what works best for which people. This leads to more possibilities for RWE science.

8.2 Threats/opportunities to Lisa

Proving a substantial influence of the mind on the body lets more attention go to the mind in psycho-somatics. As a side effect, more attention will go to the placebo part of medication. Opening this makes people (and physicians) more open to the present-day lack of solutions for many problems. The result can be anxiety, to be alleviated by validated solutions. In due time, Lisa should be recognized by NHS, NIH, FDA, EMA, etc., for reimbursement.

A regulatory committee could prescribe how Lisa should react. This poses the question of who finally is to decide what general principles lie behind Lisa. There is a lot to find out about this.

A kind of medicine cannot get patented. CCB in health and wellness coaching may be seen as a kind of medicine. However, Lisa is <specifically this procedure>. Thus, she can be patented. In such a case, it is also essential to scientifically prove her worth.

9 Grants and funding

We will look for grants to perform these studies. We did not receive any grants until now for any part of this project.

10 Bibliography

1. Lipman PD, Loudon K, Dluzak L, Moloney R, Messner D, Stoney CM. Framing the conversation: use of PRECIS-2 ratings to advance understanding of pragmatic trial design domains. Trials. 2017 Nov 10;18(1):532.

2. Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic? BMC Med. 2018 Apr 3;16(1):49.

3. Sackett DL. Explanatory and pragmatic clinical trials: a primer and application to a recent asthma trial. Pol Arch Med Wewn. 2011 Aug;121(7–8):259–63.

4. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. Yearb Med Inform. 2014 Aug 15;9:14–20.

5. Webster J, Smith BD. The Case for Real-world Evidence in the Future of Clinical Research on Chronic Myeloid Leukemia. Clin Ther. 2019 Feb;41(2):336–49.

6. Corrao G, Cantarutti A. Building reliable evidence from real-world data: Needs, methods, cautiousness and recommendations. Pulm Pharmacol Ther. 2018 Dec;53:61–7.