

A Personal and Scientific 50-year Journey From Randomized Experiments to Artificial Intelligence:

A Comprehensive and Concise Introduction

Len Bickman



Current Problems

Are You Satisfied With the Current State of Mental Health Services?



Are they sufficiently available?



Are they well implemented and efficient?



Are they sufficiently effective?

Services Are Not Sufficiently Accessible

56.4% of adults with a mental illness received no treatment (Mental Health in America, 2018).

A fifth of adults with a mental illness reported that they were not able to receive the treatment they needed (Mental Health in America, 2018.

Less than half of adolescents with psychiatric disorders received any kind of treatment in the past year (Costello et al., 2013)

Over 60% of youth with major depression do not receive any mental health treatment (Mental Health in America, 2018).

Services Are not Well Implemented or Efficient

Large unmanageable caseloads – South Carolina 15% caseworkers handles more than 50 kids (Eggers, et al., 2017)

High provider turnover rate -25% in Texas (Eggers, et al., 2017)

Administrative burden of paperwork and notetaking can take 60% of caseworker time (Eggers, et al., 2017)

Language barriers for non-English speakers in California 220 different languages

Services are not Sufficiently Effective

32 RCTs found no compelling evidence to support any one psychosocial treatment over others (Hunt, et al. 2013)

Fongay & Allison (2017) state "The demand for a reboot of psychological therapies is unequivocal simply because of the disappointing lack of progress in the outcomes achieved by the best evidence-based interventions.

Meta-analysis of 19 conduct disorder intervention concluded that they have a small positive effect but there was no evidence of any differential effectiveness (Bakker, et al, 2017)

Weisz et al. (2107) in 50 year review of 447 RCTs found a randomly selected youth would do better 63% but did not show improvement over the 50 years or type of treatment.

Why did I think We Had These Problems?

- "Low use of measurement in routine care" (Bickman, 2012)
- Not using "advances in analytics and data mining" (Bickman, 2012)
- Focusing on systems of care instead of effectiveness of treatments (Bickman, et al.
- Lack of evaluation of treatment as usual (Bickman, 2008)
- Absence of measurement feedback systems and clinician training (Garland, et al., 2010)



The Large Investment in Systems of Care Has Distracted Us From the Effectiveness of Care

- Most popular and well funded federally (CMHS) supported intervention. Spent \$1 Billion since 1993.
- Approach focuses on services that are community-based, family-focused, youth-oriented, in the least restrictive environment, individualized, and culturally and linguistically proficient – all good things
- Ft. Bragg study \$6M cost effectiveness quasiexperimental evaluation – Army and NIMH funded and RCT Stark County with NIMH funding
- Both studies: System level variables affected – access improved and costs increased but no better clinical outcomes
- Still being funded (1\$billion) with large opportunity cost

The Five Problems That Contribute To Poor Services

The Five Problems That Contribute To Poor Services

-\/-	1. The diagnoses muddle
\checkmark	2. Poorly designed measures
	3. The primacy of RCTs
	The insufficiency of treatment precision
	5. The lack of learning or feedback

1. The Diagnoses Muddle

- External factors such as insurance influences diagnosis and it can be stigmatizing Perkins et al. (2018).
- Diagnosis of depression alone is not sufficient for treatment selection thus requiring additional information (Iniesta et al., 2016)
- Traditional diagnostic categories overlap and are not mutually exclusive (Bickman, et al., 2012)
- In practice medication is prescribed according to symptoms and not DSM diagnosis (Waszcuk, et al., 2017)
- "It is not clear how the new domains of the RDoC matrix map on to the current dimensions of psychopathology" (Heckers, 2015, p 1165)

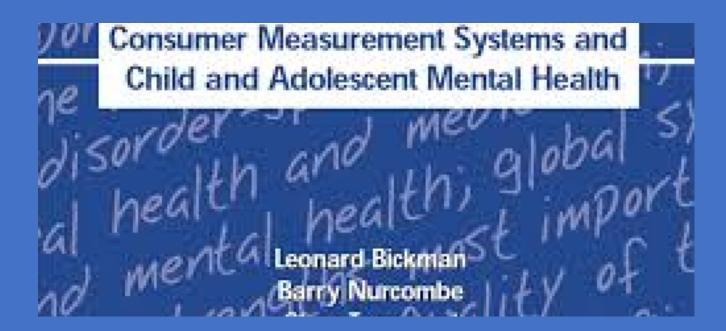
Journal of Methods and Measurement in the Social Sciences Vol.3, No. 1, 1-26, 2012

Problems in Using Diagnosis in Child and Adolescent Mental Health Services Research

Leonard Bickman	Lynne G. Wighton	E. Warren Lambert
	Vanderbilt University	
Marc S. Kar	ver I	indsey Steding
	University of South Florida	l

2. PoorlyDesignedMeasures

- Too few constructs
- Too long for practice
- A single respondent
- Not sensitive to change
- Too expensive
- No out of session measures
- Not contextual
- No feedback



3. The Primacy of RCTs.

Randomized Controlled Trials

A Gold Standard With Feet of Clay?

Leonard Bickman

Stephanie M. Reich

ndomized controlled or clinical trials or RCTs¹ have been taking on creasing importance, especially outside of the medical field. The num-RCTs is increasing as well as the number of areas in which they are ted (Bloom, in press; Boruch, Weisburd, Turner, Karpyn, & Littell, s). Moreover, these designs are being recommended and favored over designs by prestigious research organizations (e.g., Shavelson & ; 2002). In addition, several U.S. federal agencies deemed the RCT as d standard that should be used in considering the funding of research valuation but also in initiating and terminating programs (Brass, -Neto, & Williams, 2006). However, over the last several years, there is the the rest of the program in the program of the several years.



& SOCIAL EXPERIMENTATION



nphell's Legar

I FONARD BICKMAN

Social Research Methods

The SAGE Handbook of

For 50 years I Was a Committed Experimentalist

Pertti Alasuutan, Leonard Bickman and Julia Brannen

Our Experimental Methods Were Developed for Simpler Problems

- Mental health research more like weather forecasting with thousands of variables
- We need alternatives to the traditional way of doing research, service development, and service delivery
- Our research designs and analyses are based on over a century old model of evaluating simple agricultural experiments
- Oversimplification produces blunt tool



The Primary Clinical Problem with RCTs:

Prediction vs. Post- Diction or

Average Treatment Effects vs. Individual Prediction The fallacy of averages Subramanian et al., (2018)

- 1942 ideal woman measured 9 body dimensions then averaged each one. contest on who is average got 4000 responses, but no one matched on all 9.
- Air force in 1950 400 pilots on 140 body dimensions for cockpit – matching on just 10 were 0. Even when falling within 30%.

4. The Insufficiency of Treatment Precision

Only minor differential outcomes based on different therapies (Wampold, 2015)

Drug treatment choices in psychiatry are only successful in about 50% of the patients and as low as 11-30% for antidepressants (Bzdok & Meyer-Lindenberg, 2017; Dwyer & Falkai, 2018)

Antidepressants are more effective than placebos but with small ES and choice of specific medicine is trial and error (Perlis, 2016)

Relatively easy to distinguish one type of drug from another but not so for services, even dosage is hard to define

4. The insufficiency of treatment precision (cont.)

Little evidence that therapists influence psychotherapy outcomes of (King & Bickman, 2017)

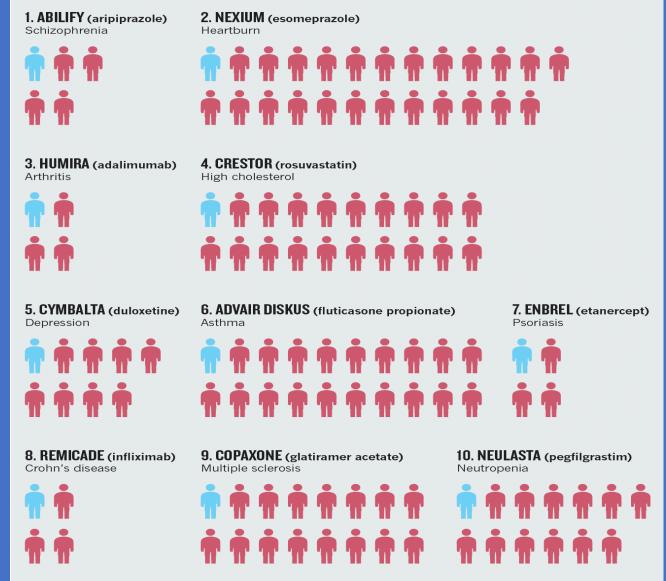
"Currently no objective, personalized methods to choose among multiple options when tailoring optimal psychotherapeutic and pharmacological treatment" (Dwyer et al., 2018 p.105)

After 46 years and 57 studies it is unknown which patients benefits from interpersonal psychotherapy (IPT) versus another treatment (Bernecker, et al., 2017)

Imprecise Drug Effects Not Limited to Mental Health Schork, 2015

IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).



Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at go.nature.com/4dr78f.

5. The Lack of Learning and Feedback

- Without accurate feedback there is little learning
- Principle holds true at the provider level
 - Clinicians do not get direct accurate feedback like carpenters
- Principle holds true at the field level
 - The field does not learn from treatment as usual, where most treatment occurs.
 - The lack of data collection and analysis of treatment as usual means we learn nothing. A wasted and costly opportunity to learn.



Precision Medicine and Artificial Intelligence (AI) as Solutions to These Problems

- Al produces effective solutions to problems similar to those faced by mental health services.
- The aim is what is the right treatment for the right patient at the right time.

read-only to prevent modification.

Adm Policy Ment Health (2016) 43:271–276 DOI 10.1007/s10488-016-0718-5



INTRODUCTION

Achieving Precision Mental Health through Effective Assessment, Monitoring, and Feedback Processes

Introduction to the Special Issue

Leonard Bickman¹ · Aaron R. Lyon² · Miranda Wolpert³

Published online: 18 February 2016 $\ensuremath{\mathbb{C}}$ The Author(s) 2016. This article is published with open access at Springerlink.com

Personalization and Precision: A New Paradigm

There is a sense of excitement and change occurring in mainstream medicine. President Obama, in his State of the Union address on January 30, 2015, announced a national Precision Medicine Initiative (The White House 2015). More recently, the United Kingdom's government innovation agency started a Precision Medicine Catapult designed to enhance the development of precision medicine in the UK (Precision Medicine Catapult 2015), Precision medicine is defined by the National Research Council as "the tailoring of medical treatment to the individual characteristics of each patient" (National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease 2011). This builds on an increasing interest in personalized medicine and, indeed, the terms "precision medicine" and "personalized medicine" are sometimes used interchangeably (Avitabile 2015). Common to both is an emphasis on

tailoring treatment to individual needs and, increasingly, on the role of technology to support that goal (Carney 2014; Sacchi et al. 2015).

Although much of the focus of medicine to date has been on biomarkers and genetics (McCarty et al. 2011), the concept is not limited to those factors. Just as critical, but less widely elaborated, are psychosocial variables that also fit under the umbrella of precision and personalized medicine. Increasing discussion has also focused on the relevance of precision medicine to mental health. Thomas Insel, a former Director of the National Institute of Mental Health (NIMH), has argued that the basic tenets of precision medicine are reflected in the NIMH Research Domain Criteria (RDoC) project, "which aims to develop more precise diagnostic categories based on biological, psychological, and socio-cultural variables" (Insel 2015). He adds that

"...precision medicine for mental disorders will not come from a single genomic glitch. Rather, like many other areas of medicine, many genes each contribute only a small amount of vulnerability as part of an overall risk profile that includes life experiences, neurodevelopment, and social and cultural factors. RDoC assumes that we will need many kinds of data to reach precision, more like triangulating to find your position on a map. These data will draw from many sources, including symptoms, genotype, physiology, cognitive assessment, family dynamics, environmental exposures, and cultural background ".

Authors contributed equally and are listed alphabetically

 Miranda Wolpert Miranda.Wolpert@annafreud.org

¹ Vanderbilt University, Nashville, TN, USA

Precision Medicine Description



The customization of healthcare, with medical decisions, treatments, practices, or products being tailored to the individual patient.



Usually based on the context of a patient's genetic content or other molecular or cellular analysis



Aspires to replace annual visits with coarse risk factors with personalized profiles and continuous longitudinal health monitoring (Gambhir et al., 2018).

From one-size fits all to Precision Medicine

- Not a paradigm shift but evolutionary
- Hippocrates advocated personalizing medicine
- Now use multiple sources of data to create individualized approaches

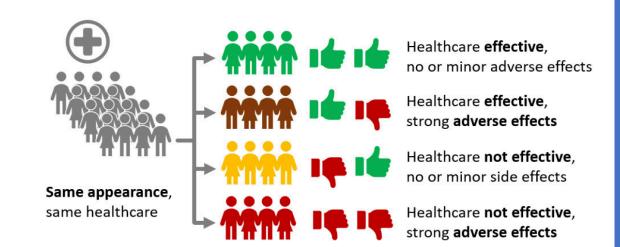
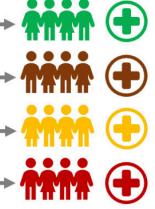


Figure 1(a). One-size-fits-all healthcare may not fit all.





What is Artificial Intelligence?

IA generally is thought to refer to machines that respond to stimulation consistent with traditional responses from humans, given the human capacity for contemplation, judgment, and intention.

Such systems have three qualities that constitute the essence of artificial intelligence: intentionality, intelligence, and adaptability.

Rather than being mechanistic or deterministic in how the machines operate, AI software learns as it goes along and incorporates real-world experience in its decision-making.

(adapted from West, 2018).

General Uses of Al

Al is currently used for predicting weather patterns, manufacturing, logistic planning (routing) banking, stock trading. Used in your smart phone, cars, planes, Siri, Alexa. In health care, decision support, testing and diagnosis, self-care aps.



Can sort through large data sets and uncover relationships that humans don't see. Also learns how to do it better with repeated use.

Artificial Intelligence and Precision Medicine

(Hinton, 2018)



use of algorithms to approximate human cognition in the analysis of complex data.



to analyze relationships between treatment techniques and patient outcomes.



AI have been applied diagnosis, treatment protocol development, drug development, personalized medicine, and patient monitoring and care.



Deep learning is best at modeling very complicated relationships between input and outputs and all their interactions

Applications of Precision Medicine

Ophthalmology – screen photos of diabetics with 94% specificity and 98% sensitivity . (Loh, 2018)

Skin cancer – scanned 129,450 clinical images with similar accuracy of board-certified dermatologists (Loh, 2018)

Cardiovascular risk prediction with ML is significantly improved over established risk prediction (Weng, et al., 2017)

Diagnostics – identify malignant tumors with 89% accuracy compared to 73% human pathologist (Loh, 2018)



IBM AI platform took 10 minutes to analyze a genome of a patient with brain cancer and suggest a treatment plan. Human experts took 160 hours (Loh, 2018)

Precision Psychiatry & Mental Health

XX

Unscramble traditional and new diagnostic categories based on biological/genetic and psychological data



Provide early indicators of success or failure of treatment

Ø

Provide feedback to clinicians at key decision points



Early warnings of relapse

The Solutions to the Specific Five Problems

Reprise: The Five Problems That Contribute To Poor Services

-\/-	1. The Diagnoses muddle
\checkmark	2. Poorly designed measures
	3. The Primacy of RCTs
	4. The Insufficiency of Treatment Precision
•••	5. The lack of learning and feedback

1. Solutions to the Problem with Diagnoses and Screening

NLP or text mining for identifying suicidal ideation in psychiatric research data base. Precision of 92% for suicide ideation and 83% attempts (Fernandes, et al., 2018).

A meta-analysis of 365 studies of predicting suicide only slightly better than chance and no change in 50 years (Loh, 2018)

AI can "bypass" the definitional problems by not using established categories but letting ML use a range of variables to describe the individual using classifier systems. (Tandon & Tandon, 2018)

Automated analysis of social media including tweets and Facebook can accurately detect depression, with AUCs ranging from .62 to .74 compared to clinical interviews with AUCs of .90. (Guntuku, et al, 2017)

1. Solutions to the Problem with Diagnoses and Screening (cont.)

Egger and her colleagues have used computer vision algorithms to code videos to elicit autism-related behaviors that can potentially be used to screen children for autism and other developmental disorders (Egger, et al.2018)

Prediction of 1479 adolescent suicides with ML showed high accuracy (AUC>.80) and outperformed traditional logistic regression analyses (.5-.6 AUCs) (Walsh et al., 2018).

Glen Saxe, has published a pioneering proof of concept paper that has demonstrated that ML methods can be used to predict child posttraumatic stress (Saxe, et al., 2017)

2. Solutions to the Problem with Measures

- Cover many domains
- Include various respondents
- Be very brief
- Longitudinal & sensitive to change
- Developed with modern psychometrics

Adm Policy Ment Health (2012) 39:3–12 DOI 10.1007/s10488-012-0404-1

ORIGINAL PAPER

The Peabody Treatment Progress Battery: History and Methods for Developing a Comprehensive Measurement Battery for Youth Mental Health

Manuel Riemer · M. Michele Athay · Leonard Bickman · Carolyn Breda · Susan Douglas Kelley · Ana R. Vides de Andrade

Published online: 11 March 2012

Personal Sensing

 The main analytical method used for personal sensing is machine learning (Peake, et al., 2018).



Ecological Momentary Assessments (EMA)

- EMA is the collection of real-time data in naturalistic environments
- Using a smart phone Identified gait features for estimating blood alcohol content (BAC) level (Gharani, et al. 2017)
- Distinguish sad vs. happy emotional states using movement sensor on smart watch (Quiroz, et al., 2018)
- mobility features from watch GPS trajectories successfully classified participants as high or low socially anxious (Boukhechba, et al., 2018)
- Described real-time fluctuations in suicidal ideation and Its risk factors using an average of 2.5 assessments per day (Kleiman et al., 2017)

3. Solutions to the Problem with the Primacy of RCTs

Saxe et al., (2016) has demonstrated the use of "Complex-Systems-Causal Network" to detect causal relationships among 11 variables and 167 bivariate relations in a psychiatry study using algorithms.

ML was more accurate that humans in predicting social and occupational disability with persons in high-risk states of psychosis (Koutsouleris, et al., 2018)

A comprehensive review and meta-analysis of machine learning algorithms that predict outcomes of depression showed excellent accuracy (.82) using multiple forms of data (Lee, et al., 2018)

4. Solutions to the Insufficiency of Treatment Precision

NLP used to code motivational interviewing fidelity content – modest relationship compared to human coders but very efficient (Tanana, et al. 2016)

Children's proclivity to interact with toys provides opportunity to use robots supported by AI to provide treatment (Kazdin, 2018)

Recent ML decision support aids using large-scale biological and other data have been useful in predicting responses to different drugs for depression (Dwyer et al., 2018)

Ecological Momentary Interventions (EMIs)

- EMIs are treatments provided to patients between sessions during their everyday lives (i.e., in real time) and in natural settings (Mohr, et al. 2017).
- Extends some aspects of psychotherapy into patients' everyday lives to encourage activities and skill building in a variety of conditions
- Uses momentary reminders for behaviors, e.g. medication adherence, management of symptoms
- More complex EMIs use algorithms to optimize and personalize systems
- "Bandit algorithm" that changes the likelihood of the presentation of a particular intervention overtime based past proximal outcomes
- Micro-randomized trials a sequential factorial design that randomly assigns an intervention component to each individual (Schueller , et al., 2017)

5. Solutions to the Problem of the Lack of Learning and Feedback

- Micro to the clinician
- Macro to the developer
- We have sufficient evidence that feedback improves outcomes but not clear what to feedback
- AI, especially neural networks, can be used to accumulate what has been learned and apply it to individualize new treatments or improve existing ones.

Effects of Routine Feedback to Clinicians on Mental Health Outcomes of Youths: Results of a Randomized Trial

Leonard Bickman, Ph.D. Susan Douglas Kelley, Ph.D. Carolyn Breda, Ph.D. Ana Regina de Andrade, Ph.D. Manuel Riemer, Ph.D.

Couple and Family Psychology: Research and Practice 2012, Vol. 1, No. 4, 274-284 © 2012 American Psychological Association 2160-4096/12/\$12.00 DOI: 10.1037/a0031022

The Technology of Measurement Feedback Systems

Leonard Bickman, Susan Douglas Kelley, and Michele Athay Vanderbilt University

Challenges and Advantages of these Solutions

Challenges with Al

nette

<u>.</u>	Legal liability – medical malpractice
~~ ~	Ethics - car manufacturer or driver if there is an accident, Trolley vigr
() ()	Accuracy and quality of original data set (Packin & Lev-Aretz, 2018)

- Large data set needed not common in mental health
- Need longitudinal data for prediction
- Data management of large data sets is complex

Challenges (cont.)

Lack of transparency of deep neural networks, not able to understand relationship because outcome does not describe process. Hidden layers, millions of parameters (De Choudhury & Kiciman, 2018).

However, new technologies are being developed to make more transparent AI approaches have higher accuracy (Valdes, et al., 2016)

Implementation issues using new technologies – the problem of engagement (Crutzen et al. 2014; Yeager & Benight, 2018)

Challenges (cont.)

Inherent bias in original data set that produced biased (racist) decisions (Williams, et al., 2018; Vale & Binns, 2017) but it can be counteracted (Fletcher-Watson et al., 2018)

Gary Marcus of NYU, an expert in ML, in a controversial paper, described 10 limitations of ML. He believes that ML "may well be approaching a wall" where progress will slow or cease.

Advantages of Al



Reduce costs – boring paperwork etc. through streamlining routine work

ğ

Integrative analysis of heterogeneous data from diverse sources

•••

Free up workers for more complex inquires



Can use remote sensing – body temp, movement, Paradigm shift



Massive investment in open-source and proprietary programs

Advantages of Al (cont.)

Handles many outcomes at once (Bzdok, 2018)

Produces individualized treatments not average

Learns how to do better with repeated applications

Identifies high priority cases

Works with human and machine in partnership collaboration AI augments person's tasks – keep humans in **Cen**ter

Al and the Future of Mental Health Services

The Critical Role of the Clinician

Changing clinician role raise big emotional, practical and economic issues

Other fields such as robotic surgery face similar issues

Not sufficient research on actual use of AI in therapy

But can examine research on therapies that have low or no clinician involvement for suggestions

Also look at Non-AI but technology-based treatments and small number of AI-based treatments Technology Based Intervention not using Al Internet delivered CBT or ICBT is major approach to using technology

9 meta-analyses of 166 studies found ICBT as effective as face-to-face (Andersson, et al., 2019)

Meta-analysis of 20 studies using behavioral intervention technologies (BIT) found equivalency (Carlbring et al., 2018)

Therapeutic alliance (TA) may be of modest importance, but early research shows TA can develop with online treatment (Fluckieger et al.,s 2018)

Bibliotherapy/ Self-Help

Does not involve a therapist or technology

A meta-analysis of 50 studies concluded that effects were similar to face-to-face treatment (Bennett et al, 2019)

A network meta-analysis of 155 studies of CBT using different formats (individual, group, telephone guided and unguided self-help) were considered equivalent alternatives to face-to face (Cuijpers et al, 2019)

Technological Interventions using Al

Use of conversational AI - real-time interchange between computer and person shows that this approach is low risk, high in consumer satisfaction and high in self-disclosure (Tuerk, 2019)

In a review of the literature form 1946-2018 on conversational agents used in treatments, Gaffney, et al,, (2019) found only 13 studies with four being RCTs. All studies showed reduced stress with five showing a significant reduction compared to control groups What Do Psychiatrists Think about the Future and the Role of Al?

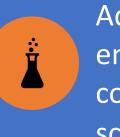
A survey of 791 psychiatrist from 22 countries were asked about how technology will affect their future practice (Doraiswamy, et al., 3030)

About 49% thought AI would have no or minimal effect on their work over the next 25 years. 47% thought it would have a moderate effect.

In his review of the research on medical specialist who may be replaced or augmented by AI, Ahuja (2019) said" Or, it might come to pass that physicians who use AI might replace physicians who are unable to do so"

My Current Work

Using Al Requires a Shift in Research and Development Culture



Adding engineers and computer scientists



From hypothesis testing to solution focused and systems oriented



From low-risk avoidance of type I errors to more risky avoidance of type II errors



Project 1: Using Al to improve treatment fidelity Clinical supervision may enhance fidelity

Under-resourced communities cannot afford clinical supervision

Al processing of audio recordings can provide fidelity feedback to clinicians

The technology is still in its infancy and evidence acceptance and usefulness is sparse

NIMH R-34 grant (M. Sibley) will test if an ADHD treatment's fidelity is enhanced with AI-powered feedback

Project 2: Reforming how advanced technology is disseminated and implemented

Over 20 years ago services found not sufficiently accessible or effective: What I call **Reform 1.0**

Evidence –based treatments

Systems of care and wraparound

Continuous quality improvement and decision support

Measurement Feedback Systems

Problems with Reform 1.0

Not a substantial improvement in access or effectiveness

High dropout rates

Low engagement of clients and clinicians

No way to personalize treatment

No way to improve treatment over time (selflearning)

Reform 2.0 Using Al and precision mental health to augment reform 1.0 services

Ameliorates problems through self-learning

Improves engagement and personalization that optimizes treatment for specific cases

Can better identify through predictions which data to collect and feedback

Can make better recommendations for supporting clinician decision making

The PRAC Alpowered platform and support services The Practice to research Acceleration Center (PRAC) is part of the Feedback Research Institute (feedbackresearch.org.)

PRAC focuses on collaborating with relevant stakeholders in implementing AI- enhanced services – providers, software vendors, researchers

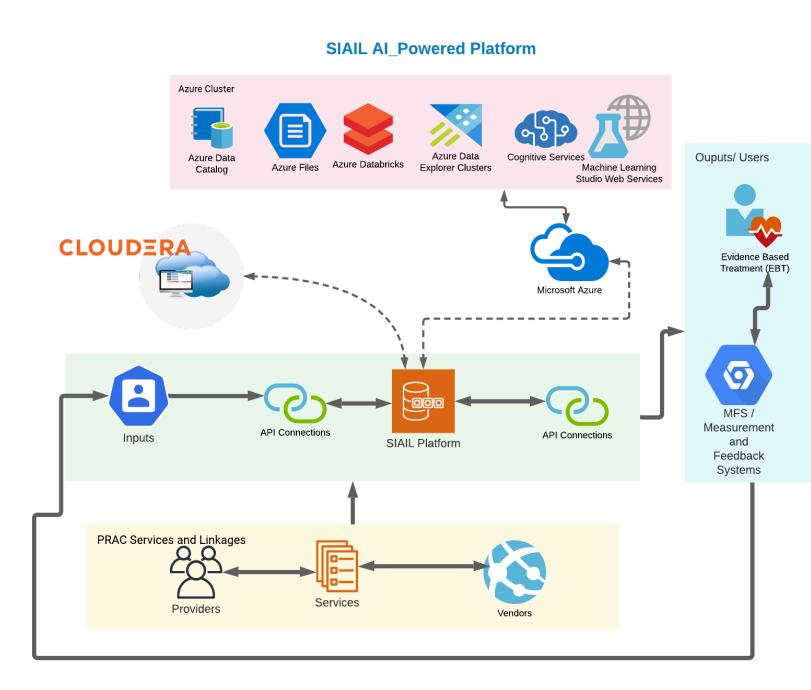
Providing services as a translator, conduit and implementation facilitator

Developing and maintaining a common data base with a focus on data collected contemporaneously with treatment

Provide an AI-enhanced platform that is capable of improving services by identifying the best data to collect and feedback as well as improvement over time.

Al Powered Platform

- Integrating and normalizing social/behavioral/MH data
- Common Data Model and Approach for Social/Behavioral Health
- Develop AI applications that provide real-time feedback to clinicians providing MH services
- Measurement Feedback System and Real-Time Decision Support
- AI applications leveraging both structured and unstructured data and models



Summary and Conclusions

Services are currently not individualized and are not effective with many clients. This is "hidden" in our statistical reporting.

Al can identify patterns and relationships that are not currently discoverable based on average responses.

Services can be made more effective and more engaging for clients and providers with the application of AI techniques.

We are at the embryonic stage of development in mental health, but we can build on the advances in other fields. We need new models of development, dissemination and implementation to assure that

technological advances are available to our poorest communities.