

A Proposal to Speed Translation of Healthcare Research Into Practice

Dramatic Change Is Needed

Rodger Kessler, PhD, Russell E. Glasgow, PhD

Abstract: Efficacy trials have generated interventions to improve health behaviors and biomarkers. However, these efforts have had limited impact on practice and policy. It is suggested that key methodologic and contextual issues have contributed to this state of affairs. Current research paradigms generally have not provided the answers needed for more probable and more rapid translation. A major shift is proposed to produce research with more rapid clinical, public health, and policy impact.

(Am J Prev Med 2011;40(6):637–644) © 2011 American Journal of Preventive Medicine

Randomized controlled efficacy trials using precisely defined interventions and highly selected participants have been the preferred and often exclusive design of choice. Designed for narrowly focused pharmacology trials, when applied to the other major issues facing health care today, such trials are limited in their ability to address the complex populations and problems we face. A moratorium is proposed on such research for the next decade, and pragmatic, transparent, contextual, and multilevel designs that include replication, rapid learning systems and networks, mixed methods, and simulation and economic analyses to produce actionable, generalizable findings that can be implemented in real-world settings is suggested.

This shift would include greater focus on the needs of practitioners, patients, payers, and policymakers and generate more relevant evidence. Funding priorities would change to include greater focus on complex multimorbid patients in community settings. Changes would be made in grant review criteria and review sections would require reviewers with new methodologic skills and experience in pragmatic studies and contextual factors. The current situation demands study of complex interventions^{1,2} that produce complex outcomes. Relying on an efficacy-based RCT research paradigm established to answer questions under decontextualized, optimal conditions will not produce the solutions needed.

You cannot solve problems by continuing to use the same solutions that created the problem in the first place.

Albert Einstein

Healthcare researchers have generated an extraordinary literature concerning the relationship of biobehavioral factors to health and illness causation. There is an equally excellent literature demonstrating how nonpharmacologic interventions can improve both behaviors and biomarkers under optimal conditions.^{3,4} These efforts have left little time or funding to address the transfer of those results into healthcare or public health practice and policy.^{5,6} This has rendered that efficacy literature of very limited use to practitioners⁷ and ineffective at affecting real-world and large-scale practices and policies, or at achieving public health impact.

A considerable literature has documented the incomplete and often-lengthy gap between research and practice^{8–10} that has sometimes been characterized as “lost in translation.”^{11,12} Excellence in science generated research designs strong on internal validity but that have extremely limited or unknown external validity.^{6,13} Such research is not perceived by practitioners and policymakers as relevant to their practice settings and populations, nor of much use to decision makers searching for answers to specific complex questions to which they must respond.¹⁰ With healthcare reform focusing on comparative effectiveness research and evidence-based interventions, we are at a crossroads. The “excellence” in which we were trained enhances careers, but these efforts have little bearing on real-time needs for answers to complex implementation, dissemination, and policy questions. Here is an alternative approach that will substantially increase the probability that research findings will translate into

From the College of Medicine, University of Vermont (Kessler), Burlington, Vermont; and the Institute for Health Research, Kaiser Permanente Colorado (Glasgow), Denver, Colorado

Address correspondence to: Rodger Kessler, PhD, College of Medicine University of Vermont, Given Courtyard, 4th Floor, Burlington VT 05405. E-mail: rodger.kessler@uvm.edu.

0749-3797/\$17.00

doi: 10.1016/j.amepre.2011.02.023

action and policy, and be relevant to challenging, low-resource settings in the U.S. and abroad.

A Bold Proposal

Sometimes a problem reaches a point of acuity where there are just two choices left: bold action or permanent crisis.

David Rothkopf

We propose a 10-year moratorium on efficacy RCTs in health and health services research. This would provide the necessary time for researchers, practitioners, policy-makers, and citizens to collaboratively identify and evaluate innovations that have real potential for translation. With funding available for such efforts, investigators could design and conduct pragmatic trials¹⁴ and rapid learning studies¹⁵ with representative populations, staff, and settings that will more rapidly affect practice and policy.⁵ Such a shift is consistent, for example, with the FDA's recent draft guidance to focus on adaptive research designs based on analyses of interim data¹⁶ and recommendations of the CONSORT Working Group on Pragmatic Trials.^{14,17}

Why Such a Change?

Uncritically accepting designs that isolate, decontextualize, and simplify issues has dramatically decreased the applicability of the current results. The key problems of today are “wicked” problems¹⁸ that are multilevel, multiply determined, complex, and interacting. Physicists employing mechanistic and decontextualizing, isolation design approaches learned the limitations of such approaches at reductionism decades ago and have since moved to chaos and complexity theory, and more contextual approaches.^{19,20}

The RCT designs and hegemony around systematic reviews have worked well to create an initial body of research but have not worked for producing replicable results that matter or translate.²¹ The system that is built stifles creativity and thinking by holding that efficacy RCTs are always the highest or only type of evidence considered. A culture of funding, publication, and rewards has followed based on adherence to this model of deconstructive science.^{22,23} This has permeated all stages from training in research methods, to grant funding, to publication and review processes, to reward and tenure systems.

It takes years to conduct a large-scale RCT requiring grant approval, funding, protocol development, and IRB approval, recruitment, implementation, and follow-up. Frequently, only then can analyses begin, when investigators are blind to results and cannot begin analyses until

final follow-up data have been collected, making this an extremely inefficient paradigm.

No More Cookbook Randomized Controlled Trials

Efficacy-focused, RCT designs achieved prominence, at least in part, as a method for pharmaceuticals seeking FDA approval, where it can be argued that biological responses are standard (though even this assumption is inconsistent with personalized or genomic medicine). But this type of highly prescribed science has often been applied uncritically to *all intervention problems and questions*. These efforts have produced important outcomes²⁴ but even when a definitive result is produced, this provides little guidance as to how to translate findings into feasible and cost-effective programs for practice. Many political, economic, and historical reasons have contributed to the lack of rapid transfer.^{11,12} There is a clear dilemma that policy, organizational, and environmental issues and the interaction of these contextual factors with patient-centered healthcare, genomic, and biologic factors have not been elucidated by reductionistic cross-sectional correlational and efficacy RCT designs.

The difficulty emerges when these approaches are used for the “wicked”¹⁸ and complex, inter-related problems now driving current healthcare crises, both in the U.S. and internationally. Intervention solutions derived under ideal conditions are not the best solutions to complex, multiply determined problems in very different contexts.²⁵ Intensive, expensive interventions delivered in leading medical centers by world-class experts and requiring very skilled intervention delivery and high fidelity, administered to uncomplicated, highly motivated patients cannot be expected to work equally well in the messy, real-world, under-resourced public health settings around the world dealing with complex comorbid patients living in stressful, nonsupportive environments. The only thing surprising about this is why anyone would think they should.^{25,26} Beginning with designs that focus on rapid and relevant translation and learning will generate evidence more relevant for practice and policy uptake.

What Could Be Learned During a Moratorium

The types of research more likely to be applicable to real-world problems would be thoughtful alternative designs that fit the question—not projects that automatically use RCT designs for every question.²⁷ This emerging evidence would be practical, contextual, and transparent so others more easily could understand and build on its limitations.^{28,29}

Such a shift would help us to consider multiple principles of causation.^{30,31} In particular, greater attention would be paid to replication³² and robustness.³³ Often, the scientist who proposes to or performs an important replication is rejected by study sections and journals in favor of “new, innovative” surprising findings—even if cross-sectional and not replicated. This is especially true when one fails to replicate initial positive findings—which is usually interpreted to mean that the “evidence-based finding” has not been adequately implemented.³⁴ Instead, a great deal could be learned from failures to replicate—by understanding moderating factors and conditions responsible.^{33,35}

There are different types of replication: in his classic text Sidman³² distinguishes direct replication—conducted under conditions as close as possible to the original; from systematic replication—which explores boundary conditions by experimentally varying different factors. The more robust an intervention effect is across different conditions in systematic replications, the greater confidence one has that the effect is strong and causal.³⁶ Thoughtful replications should be considered important scientific contributions—not treated as inferior.

Decision makers often need relevant evidence in a matter of days or weeks, not years or decades.¹⁵ Etheredge and the IOM suggest “rapid learning healthcare evidence”¹⁵ as a different complementary type of research. Such studies use large data sets of real-world patients treated in real-world settings—such as HMOs, VA centers, or primary care-based networks.^{37,38} Substantial amounts of ARRA (the American Recovery and Reinvestment Act of 2009) money have been spent to create integrated data networks that combine data across multiple settings to provide information—often from electronic medical records—using principles of data harmonization to produce rapid, sometimes real-time, results on up to millions of real-world patients being treated under real-world conditions in different contexts.^{37,38}

A second way to foreshorten the amount of time to produce results for pressing problems is to utilize sophisticated models or simulations.^{39–42} Simulations have been used effectively to advance a number of sciences, including climate change, economics, and decision making. Recently, healthcare simulation models have been applied to model the results of various biomedical interventions and policies.^{43,44} Some of these models have been found to replicate the results of actual intervention trials quite closely.⁴³ It is beyond the scope of this article to discuss the strengths and limitations of simulation models or their variations, but they do offer another, seldom-used approach to produce relatively rapid answers. Such models including contextual factors to eval-

uate the potential impact of moderating conditions—could first be conducted to determine if there is the need for expensive multisite intervention trials. They can also be very useful in modeling costs and value, and in detecting potential unanticipated outcomes by varying assumptions and input factors such as patient characteristics and assumptions about adherence levels.

Implications of results from well-controlled, observational studies and simulations can in turn be relatively rapidly tested by designs such as multiple baseline across settings,²⁷ fractional factorial, preference, or other quasi-experimental designs in the rapidly advancing area of quality improvement evaluation.⁴⁵

Other Issues and Consequences

The types of important, practical questions that decision makers need answered are of the form: *What does it cost; how many and what types of people will participate and how do I know this will work in our setting?* If these questions cannot be answered to the satisfaction of the questioner, further consideration of the program or policy is extremely unlikely regardless of the amount, strength, or quality of data on its efficacy.

Cost questions are important and complex, and cost data must be a primary focus rather than an afterthought or “add-on.” Detailed discussion of economic analysis is beyond the scope of this paper, but useful references and guidelines are available.^{46,47} At a minimum, evaluations should include a basic analysis of the costs of delivering an intervention, using reasonably straightforward, transparent procedures.⁴⁷ Progressively more challenging questions concern cost effectiveness (and the related business perspective question of “return on investment” and cost-benefit analysis).⁴⁶ Particularly recommended are economic sensitivity analyses that can inform potential adopters of the impact of different recruitment and delivery options, delivery staff, intervention intensities, levels of scale, inflation assumptions, and other contextual factors. Such designs can test the limit of applicability to settings such as low-resource clinics and community health centers in the U.S. or low- and middle-income countries.

The second “field of dreams” question “if we offer it, will people come” is especially challenging to answer using traditional RCT designs. In RCTs, participants (and sometimes clinicians) must agree not only to randomization and research requirements but also to assignment to any of the experimental conditions. More real-world relevant designs to project uptake or reach participation rates⁴⁸ include preference designs⁴⁹ in which all or a subset of participants are allowed to choose

Table 1. Key pragmatic and translation issues in need of study by research design, intervention, and evaluation issues

Key complex T3 and T4 research translation issues in need of study				
Research issue	Practical and feasible interventions	Key contextual factors	Transparent reporting	Design fits question
Experimental design	Addresses issues relevant ^a to decision makers Representative settings and participants Includes complex patients ^a and realistic comparison treatment(s)	Heterogeneous or typical settings Study of moderating factors ^a Includes qualitative features	Reports modification and adaptation to recruitment and design across sites Local customization	Fits specific question Dynamic adaptive rapid and efficient Information for scale-up and robustness analyses Simulations
Intervention characteristics	Designed for broad adoption and implementation ^a Efficient MINC design approach Stepped care Scalable	Flexible Provides guidelines ^a for fidelity and customization Deliverable by variety of staff in typical settings	Reports on: Adoption ^a Implementation Modifications ^a Subgroup effects “CONSORT Plus ^a ” information ^b	Designed for healthcare settings of future ^a QI blends that get smarter over time Sustainable ^a with typical resources
Evaluation measures and analyses	Analyses of modifier and subgroup effects Effects of treatment intensity and staff expertise Cost, cost-effectiveness, and, sensitivity analysis ^a	Focus on policy, economic, and political context Assess impact on disparities ^a high-risk subgroups variation ^a across settings, staff, and time Generalization analyses ^a	Reach by condition Unintended results Quality-of-life impacts Implementation ^a by condition and over time Maintenance ^a at setting and individual levels	Evaluate systems impacts and unintended consequences Understand multilevel effects and mediators “Postmortem” interviews Long-term sustainability and program evolution

^aIssues especially in need of study

^bCONSORT Plus – see www.re-aim.org/tools/figures-and-tables.aspx. Extends reporting to include recruitment of settings and sustainability
CONSORT, Consolidated Standards of Reporting Trials; MINC, minimal intervention needed for change; QI, quality improvement

among alternative interventions, as is more typically done in real-world settings.⁴⁸

The third question of “will this work in our setting” raises critical issues for translation and dissemination that are seldom addressed in traditional RCT designs. As presented in Table 1, these issues can be summarized in the form of a matrix displaying the extent to which studies are practical, contextual, transparent, and thoughtful,³³ cross-indexed with how these issues apply to research design, intervention characteristics, and evaluation issues.

Here are a few highlighted issues from Table 1. Interventions should be developed from the outset with dissemination and scalability in mind.⁵⁰ The recent NIH program announcement #10-038 and establishment of a standing study section for implementation and dissemination research is a helpful step. It should assist systematic investigations of the specificity versus generalizability of intervention effects that are currently greatly understudied, as are the main and interactive effects of

health policies. If the resource demands for time, expertise, equipment, or commitment far exceed those in typical practice settings or for which it is possible to be reimbursed, it is unlikely that an innovation will be adopted, regardless of its effectiveness. Therefore, it is suggested that there be included an additional scored review element of feasibility of translation into typical practice settings.

With the advent of genomic medicine,⁹ this is even more true: mean results across genetically diverse samples are likely less informative than designs capable of answering questions about which intervention is most effective for which group of patients.³⁵ To do so leads us to consider the utility of practical or pragmatic trials.¹⁴

If we want more evidence-based practice, we need more practice-based evidence.

Larry W. Green, 2004²⁵

Thorpe et al. and the CONSORT Pragmatic–Explanatory Work Group¹⁷ have identified key elements of such trials.¹⁴ They suggest that applicability be an explicit element in planning of trials; trials be explicitly designed to produce results that have immediate and wide applicability; trials be reported in ways that make it easier to judge applicability; decision makers explicitly support and seek out trials with pragmatic attitudes that inform choices affecting clinical care, health services delivery, and policy; and funders request proposals for research on the major determinants of applicability.

Such trials^{14,28,35} fit well with the need to report findings more transparently (www.re-aim.org/tools/figures-and-tables.aspx). Information about recruitment procedures for settings and staff involved in a study, the number of settings excluded and reasons for exclusion, and the number and characteristics of settings and staff who are invited but decline to participate is also important for drawing conclusions.⁵¹ Many large-scale studies and multisite trials find substantial differences in implementation or outcomes across setting or staff types, and this should be reported more transparently.

Inclusion of such information in “CONSORT PLUS” summaries (www.re-aim.org) and diagrams as shown in Figure 1, including the long-term status of programs studied (e.g., if they are discontinued, modified, or adopted in entirety following the research), would greatly advance available knowledge of sustainability. Information on long-term program sustainability has consistently been found to be the characteristic least often reported across a variety of research areas and settings.⁵² Obviously, much greater attention to longer-term sustainability and its multilevel determinants is needed, and expanded reporting, such as that in Figure 1, would help to highlight this need.

The PRECIS criteria recommended by the Pragmatic–Explanatory Work Group on CONSORT¹⁴ represent an important effort to increase such transparency in reporting.¹⁴ The pragmatic–explanatory continuum indicator summary (PRECIS) has been developed to increase clarity about the extent to which a trial is applied and widely applicable (Pragmatic) or more basic and efficacy-focused (Explanatory) on ten dimensions: flexibility of the comparison and experimental interventions; practitioner expertise—in both experimental and comparison conditions; eligibility criteria; primary analysis; practitioner adherence; participant compliance; follow-up intensity; and outcomes. Routinely including information on the PRECIS dimensions would contribute toward transparency and broader understanding of results.

Pragmatic research applications may well “violate” rules many of us have been taught. For example, an intervention may need to be dynamic and adaptive⁵³ over time to address

emerging issues. Evaluations often need to employ both qualitative and quantitative methods and address multiple ecologic levels (e.g., patient; family; health care, community, policy). One implication of complexity theory illustrated by pharmacologic interventions is that there are often interactions and important unintended consequences.^{54,55} Several leading research groups are coming to view healthcare settings as complex, adaptive systems,^{19,54,55} and these systems often resist change or produce unexpected, nonlinear results when “perturbed.”^{54,55} Studies should be designed to detect such impacts—positive or negative. Because research follows the funding available for its conduct, development of private foundation, NIH, AHRQ, CDC, and other sustained announcements and support for pragmatic trials is encouraged.^{56,57}

Larger Impacts on Research

Here are likely consequences of a moratorium, using the NIH funding process as a jumping off point. Although only one-way research is funded, NIH accounts for a considerable amount of research funding, and the issues below generalize to other funding sectors. With an increased focus on research that rapidly responds to “messy and wicked, complex questions,” funding priorities, and announcements would change, especially those related to comparative effectiveness research.^{58,59} As complex patients having multiple morbidities in community settings become primary research foci⁶⁰ instead of excluded from research, there will be a need for significant rethinking of single-condition-focused funding initiatives.

Greater collaborative funding across DHHS agencies and public–private funding would enhance study of multimorbid complex patients and delivery systems in which they receive care. Grant review criteria would be reformulated to reward design criteria favoring thoughtful, pragmatic designs tailored to fit specific questions and conditions. Inclusion of broad, representative populations and diverse settings would become important criteria, and high levels of exclusions would be rated as limitations. Practice settings that reflected typical (or a range of) practice patterns and staffing models would be strengthened. Investigator and research team composition would need to include strong experience in applied settings. Community-based settings would be prioritized over academic expertise in specialized conditions. A new criterion might be added: *Will this research generate data likely to result in policy or practice improvement within 3–5 years?*

The moratorium would require substantial change in review sections. Many current study sections now consist predominantly of researchers trained in efficacy or drug RCTs, with a history of funding and a view of research that has efficacy RCTs as their hallmark. In contrast, new

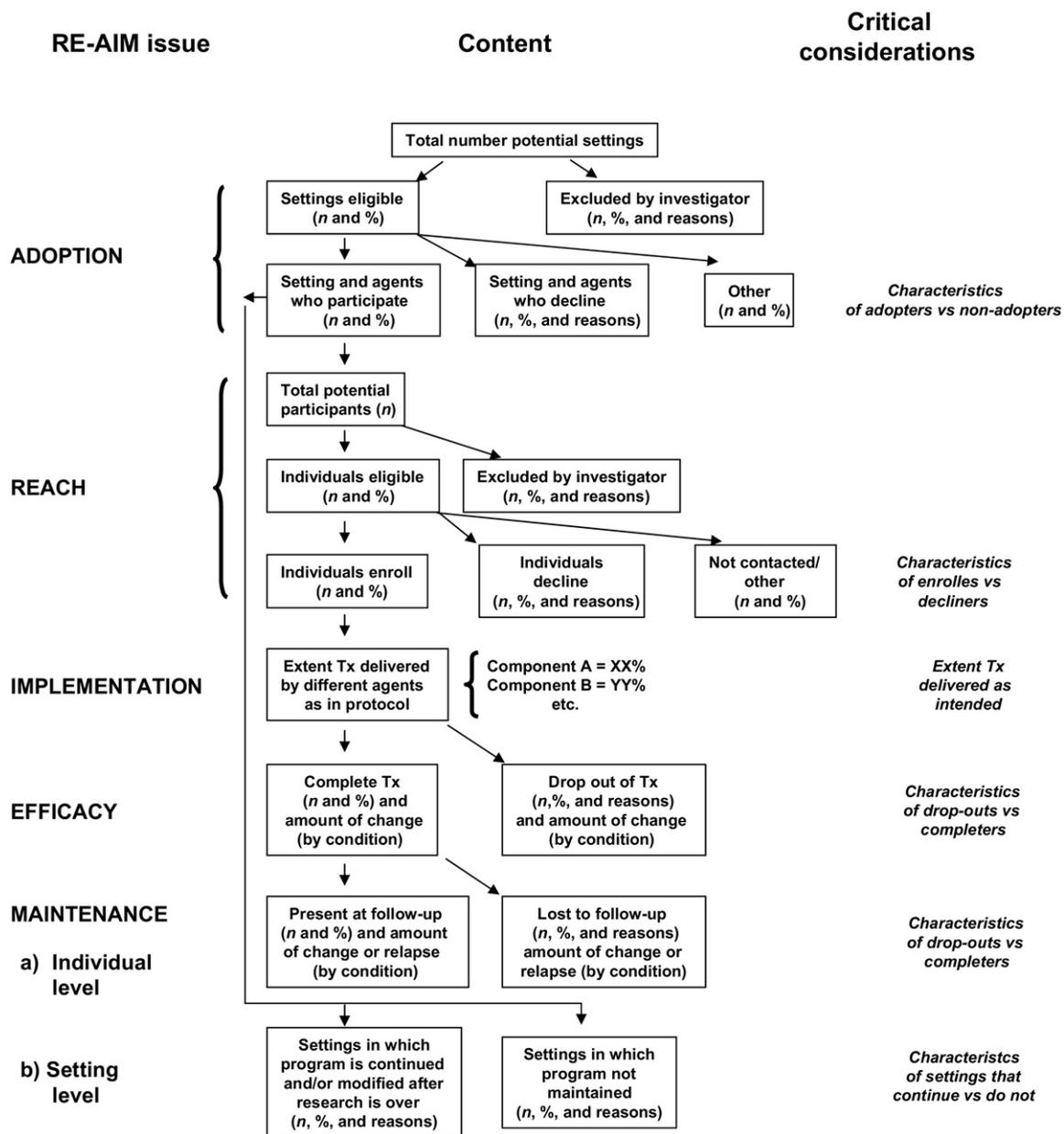


Figure 1. Extended CONSORT diagram

Note: At each step, record qualitative information and factors affecting each RE-AIM dimension and step in flowchart. Reprinted with permission from the RE-AIM website: www.re-aim.org/tools/figures-and-tables.aspx
 CONSORT, Consolidated Standards of Reporting Trials; RE-AIM, reach, effectiveness, adoption, implementation, maintenance; Tx, treatment

and more diverse methodologic skills and experiences in real-world settings would need to be qualifications of reviewers. This would likely include retraining current reviewers and recruitment of reviewers with those skill sets—and openness—to alternative designs.

Downstream consequences of a moratorium and corresponding realignment of research priorities would affect universities and medical schools, which would have greater incentives to develop community partnerships. Academic investigators would need to reinvent themselves. Training of investigators would

move from the lab to the community. Centers for Translational Science Institutes would shift their current emphasis from T1 (and some T2) research toward a much greater focus on what is coming to be called T3 and T4 research.⁹ T3 and T4 research extend translation from effectiveness to implementation and then to policy and broad public health implementation concerned with larger-scale public health impact. Such a refocus would surely respond to a historic conundrum of trials in resource-poor developing countries. Historically, in research trials, considerable segments of

the population are eliminated, and the design of the interventions preclude generalizability and availability of the intervention in most practices, even if the intervention is efficacious.^{61,62}

Anticipated Reactance

We have focused attention and identified the consequences of a moratorium on efficacy-style RCTs. This is good science and responds to a set of issues that has been severely limiting the impact, relevance, and timely application of the present science. Admittedly the ideas propose changes to the currently dominant philosophy of science held by many medical researchers—and proposing science that moves away from a mechanistic, reductionist view to a contextualist⁶³ or realist perspective.³⁵

This proposal embraces theories and suggests testing concepts of dynamic systems, organizational factors, complexity science, and unintended consequences: all necessary to do good science in the messy, real world. Such a shift would result in more qualitative and mixed-methods research,^{26,40} as well as modeling simulations, and force integration of that thinking into the current view of science. There would be a primary, rather than tertiary, focus on pragmatic trials, feasibility, cost effectiveness, generalizability, and external validity.^{6,9,13,14} In addition, if research no longer focused on isolated components but emphasized study of complex interventions⁶⁰ in multiple contexts (Table 1), then studies of variation⁶⁴ and impacts of variation in participants, settings, intervention staff, and delivery conditions would also be prioritized.

The definition of insanity is doing the same thing over and over again and expecting a different result.

Albert Einstein

Conclusion

The current model of mechanistic simplification and isolation of key factors in “efficacy-style RCT” intervention research clearly has not produced the results needed—and there is no indication that faster, more applicable results from this paradigm are on the horizon. The difficult issues to which answers are now needed—and quickly—can be summarized as complex problems of complex patients embedded in complex healthcare systems in complex and changing communities that require complex interventions^{61,62,65} embedded in changing socioeconomic–political conditions and health policies. Complex interventions^{1,2} that will produce complex outcomes must be studied, and it seems illogical to expect that a research paradigm established to answer quite different types of questions under simple, decontextualized optimal conditions will produce the answers needed. So perhaps a different—even radically different—approach is worth a try.

Dr. Glasgow is now Deputy Director for Dissemination and Implementation Science, Division of Cancer Control and Population Sciences, National Cancer Institute. The opinions in this article do not necessarily represent those of the National Cancer Institute or any other body.

No financial disclosures were reported by the authors of this paper.

References

- Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321(7262):694–6.
- Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *BMJ* 2004;328(7455):1561–3.
- Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med* 1986;15:451–74.
- Fisher JD, Fisher WA. Changing AIDS-risk behavior. *Psychol Bull* 1992;111:455–74.
- Woolf SH, Johnson RE. Inattention to the fidelity of health care delivery is costing lives. *Am J Public Health* 2007;97(10):1732–3.
- Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validity and translation methodology. *Eval Health Prof* 2006;29(1):126–53.
- Rothwell PM. External validity of randomised controlled trials: to whom do the results of this trial apply? *Lancet* 2005;365:82–93.
- Green LW, Glasgow RE, Atkins D, Stange KC. Getting evidence from science to practice: Slips “twixt cup and lip.” *Am J Prev Med* 2009;37(6S 1):S187–91.
- Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 2007; 9(10):665–74.
- Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;82(4):581–629.
- Green LW, Ottoson JM, Garcia C, Hiatt RA. Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annu Rev Public Health* 2009;30:151–74.
- Lenfant C. Clinical research to clinical practice—lost in translation? *N Engl J Med* 2003;349:868–74.
- Glasgow RE, Green LW, Klesges LM, et al. External validity: we need to do more. *Ann Behav Med* 2006;31(2):105–8.
- Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ* 2009;180(10):E47–57.
- Etheredge LM. A rapid-learning health system: what would a rapid-learning health system look like, and how might we get there? *Health Aff* 2007;Web Exclusive Collection:w107–18.
- Hsu HS, Gupta G. An overview of the FDA draft guidance on adaptive design clinical trials. 2010. www.fda.gov/.
- Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008; 337:a2390 (doi: 10.1136/bmj).
- Kreuter MW, DeRosa C, Howze EH, Baldwin GT. Understanding wicked problems: a key to advancing environmental health promotion. *Health Educ Behav* 2004;31(4):441–54.
- Axelrod R, Cohen MD. *Harnessing complexity: organizational implications of a scientific frontier*. New York: Simon & Schuster, 2000.

20. Thomas P. Integrating primary health care. Oxford, England: Radcliffe Publishing, 2006.
21. Ioannidis JP. Evolution and translation of research findings: from bench to where. *PLoS Clin Trials* 2006;1(7):e36.
22. Kuhn TS. The structure of scientific revolutions. 3rd ed. Chicago: University of Chicago Press, 1996.
23. Chalmers AG. What is this thing called science. 3rd ed. Lucia, Queensland, Australia: Queensland Press, 1999.
24. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002;346(6):393–403.
25. Green LW, Ottosen JM. From efficacy to effectiveness to community and back: evidence-based practice vs. practice-based evidence. Proceedings from conference: From Clinical Trials to Community: The Science of Translating Diabetes and Obesity Research; 2004. Jan 12–13, 2004; Bethesda MD. Bethesda MD: National Institutes of Diabetes, Digestive and Kidney Diseases, 2004.
26. Ottosen JM, Hawe P, eds. New directions for evaluation: knowledge utilization, diffusion, implantation, transfer, and translation: implications for evaluation. Hoboken NJ, 2009.
27. Hawkins G, Sanson-Fisher RW, Shakeshaft A, D'Este C, Green LW: The multiple-baseline design for evaluating population-based research. *Am J Prev Med* 2007;33(2):162–8.
28. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004;94:361–6.
29. Glasgow RE, Linnan LA. Evaluation of theory-based interventions. In: Glanz K, Rimer BK, Viswanath K, eds. Health behavior and health education: theory, research, and practice. 4th ed. San Francisco CA: Jossey-Bass, 2008:487–508.
30. Brownson RC, Baker EA, Leet TL, Gillespie KN, eds. Evidence-based public health. New York: Oxford University Press, 2003.
31. Bandura A. Self-efficacy: the exercise of control. New York: W.H. Freeman, 1997.
32. Sidman M. Tactics of scientific research: evaluating experimental data in psychology. 5th ed. New York: Basic Books, 1960.
33. Glasgow RE. What types of evidence are most needed to advance behavioral medicine? *Ann Behav Med* 2008;35(1):19–25.
34. Wade DT. Research into the black box of rehabilitation: the risks of Type III error. *Clin Rehabil* 2001;15(1):1–4.
35. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review: a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10(S1):S21–39.
36. Cronbach LH, Glesser GC, Nanda H, Rajaratnam N. The dependability of behavioral measurements: theory of generalizability for scores and profiles. New York: Wiley, 1972.
37. Agency for Health Care Research and Quality. Practice-based research networks. pbrn.ahrq.gov/portal/server.pt.
38. HMO Research Network. www.hmoresearchnetwork.org.
39. National Cancer Institute. Cancer Intervention and Surveillance Modeling Network (CISNET). cisnet.cancer.gov.
40. Schensul JJ. Community, culture and sustainability in multilevel dynamic systems intervention science. *Am J Community Psychol* 2009;43(3–4):241–56.
41. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. *Am J Prev Med* 2011;40(3):303–11.
42. Maglio PP, Mabry PL. Agent-based models and systems science approaches to public health. *Am J Prev Med* 2011;40(3):392–4.
43. Stern M, Williams K, Eddy D, Kahn R. Validation of prediction of diabetes by the Archimedes model and comparison with other predicting models. *Diabetes Care* 2008;31(8):1670–1.
44. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. *Am J Prev Med* 2011;40(3):353–61.
45. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff* 2005;24(1):138–50.
46. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost effectiveness in health and medicine. New York: Oxford University Press, 2003.
47. Ritzwoller DP, Sukhanova A, Gaglio B, Glasgow RE. Costing behavioral interventions: a practical guide to enhance translation. *Ann Behav Med* 2009;37(2):218–27.
48. Glasgow RE, Edwards LL, Whitesides H, Carroll N, Sanders TJ, McCray BL. Reach and effectiveness of DVD and in-person diabetes self-management education. *Chronic Illn* 2009;5:243–9.
49. Wermeling PR, van den Donk M, Gorter KJ, Ardine de Wit G, van der Graaf Y, Rutten GE. Towards a more efficient diabetes control in primary care: six-monthly monitoring compared with three-monthly monitoring in type 2 diabetes—The EFFIMODI trial. Design of a randomised controlled patient-preference equivalence trial in primary care. *BMC Fam Pract* 2010;11(1):35.
50. Klesges LM, Estabrooks PA, Glasgow RE, Dziewaltowski D. Beginning with the application in mind: designing and planning health behavior change interventions to enhance dissemination. *Ann Behav Med* 2005;29(Suppl):66–75.
51. Glasgow RE, Strycker LA, Kurz D, et al. Recruitment for an Internet-based diabetes self-management program: scientific and ethical implications. *Ann Behav Med* 2010;40(1):40–8.
52. Glasgow RE, Klesges LM, Dziewaltowski DA, Estabrooks PA, Vogt TM. Evaluating the overall impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues. *Health Educ Res* 2006;21(5):688–94.
53. Lei X, Yuan Y, Yin G. Bayesian phase II adaptive randomization by jointly modeling time-to-event efficacy and binary toxicity. *Lifetime Data Anal* 2011;17(1):156–74.
54. Stroebel CK, McDaniel RR Jr, Crabtree BF, Miller WL, Nutting PA, Stange KC. How complexity science can inform a reflective process for improvement in primary care practices. *Jt Comm J Qual Patient Saf* 2005;31(8):438–46.
55. Litaker D, Tomolo A, Liberatore V, Stange KC, Aron D. Using complexity theory to build interventions that improve health care delivery in primary care. *J Gen Intern Med* 2006;21(S2):S30–34.
56. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003;290:1624–32.
57. Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med Care* 2005;43(6):551–7.
58. Wu SY, Pearson M, Schaefer J. Assessing the implementation of the Chronic Care Model in Quality Improvement Collaboratives: does baseline system support for chronic care matter. Santa Monica CA: RAND, 2003.
59. Glasgow RE, Steiner JF. Comparative effectiveness research to accelerate translation: recommendations for “CERT-T.” Brownson RC, Colditz G, Proctor E, eds. New York: Oxford University Press, 2010.
60. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694–6.
61. Emanuel EJ, Wendler D, Killen J, Grady C. What makes clinical research in developing countries ethical? The benchmarks of ethical research. *J Infect Dis* 2004;189(5):930–7.
62. Benatar SR. Avoiding exploitation in clinical research. *Camb Q Health Ethics* 2000;9(4):562–5.
63. Biglan A. Changing cultural practices: a contextualist framework for intervention research. Reno NV: Context Press, 1995.
64. Selby JV, Schmittidiel JA, Lee J, et al. Meaningful variation in performance: what does variation in quality tell us about improving quality? *Med Care* 2010;48(2):133–9.
65. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:a1655.