

RESEARCH ARTICLE

Improving natural product research translation: From source to clinical trial

Barbara C. Sorkin¹ | Adam J. Kuszak¹ | Gregory Bloss² | Naomi K. Fukagawa³ |
 Freddie Ann Hoffman⁴ | Mahtab Jafari⁵ | Bruce Barrett⁶ | Paula N. Brown⁷ |
 Frederic D. Bushman⁸ | Steven J. Casper⁹ | Floyd H. Chilton¹⁰ | Christopher S. Coffey¹¹ |
 Mario G. Ferruzzi¹² | D. Craig Hopp¹³ | Mairead Kiely¹⁴ | Daniel Lakens¹⁵ |
 John B. MacMillan¹⁶ | David O. Meltzer¹⁷ | Marco Pahor¹⁸ | Jeffrey Paul¹⁹ |
 Kathleen Pritchett-Corning²⁰ | Sara K. Quinney²¹ | Barbara Rehermann²² |
 Kenneth D. R. Setchell²³ | Nisha S. Sipes²⁴ | Jacqueline M. Stephens²⁵ |
 D. Lansing Taylor²⁶ | Hervé Tiriac²⁷ | Michael A. Walters²⁸ | Dan Xi²⁹ |
 Giovanna Zappalá³⁰ | Guido F. Pauli³¹

¹Office of Dietary Supplements, National Institutes of Health (NIH), Bethesda, MD, USA

²National Institute on Alcohol Abuse and Alcoholism, NIH, Bethesda, MD, USA

³USDA ARS, Beltsville, MD, USA

⁴HeteroGeneity LLC, Washington, DC, USA

⁵University of California, Irvine, Irvine, CA, USA

⁶University of Wisconsin, Madison, WI, USA

⁷British Columbia Institute of Technology, Burnaby, BC, Canada

⁸University of Pennsylvania, Philadelphia, PA, USA

⁹Office of Dietary Supplement Programs, Center for Food Safety and Applied Nutrition, Food and Drug Administration (FDA), Hyattsville, MD, USA

¹⁰Department of Nutritional Sciences and the BIO5 Institute, University of Arizona, Tucson, AZ, USA

¹¹University of Iowa, Iowa City, IA, USA

¹²Plants for Human Health Institute, North Carolina State University, Kannapolis, NC, USA

¹³National Center for Complementary and Integrative Health, NIH, Bethesda, MD, USA

¹⁴Cork Centre for Vitamin D and Nutrition Research, School of Food and Nutritional Sciences, University College Cork, Cork, Ireland

¹⁵Eindhoven University of Technology, Eindhoven, Netherlands

¹⁶University of California, Santa Cruz, Santa Cruz, CA, USA

¹⁷University of Chicago Medicine, Chicago, IL, USA

¹⁸University of Florida, Gainesville, FL, USA

¹⁹Drexel Graduate College of Biomedical Sciences, College of Medicine, Evanston, IL, USA

²⁰Harvard University, Cambridge, MA, USA

²¹Indiana University, Indianapolis, IN, USA

²²National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD, USA

Abbreviations: AD, adaptive (clinical trial) design; ADME, absorption, distribution, metabolism, excretion; CT, clinical trial; FADS, fatty acid desaturase; FAIR, findable, accessible, interoperable, reusable (desirable characteristics for archived data); HTS, high-throughput screening; MPS, microphysiological system; NOD, non-obese, diabetic; NP, natural product; NPCT, natural product clinical trial; PBPK, physiologically based pharmacokinetic (a type of in silico model); PD, pharmacodynamics; PK, pharmacokinetics; PUFA, polyunsaturated fatty acid; QSP, quantitative systems pharmacology; RCT, randomized controlled trial; SCID, severe combined immune deficiency; VOI, value of information.

²³Cincinnati Children's Hospital, Cincinnati, OH, USA

²⁴National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA

²⁵Louisiana State University, Baton Rouge, LA, USA

²⁶University of Pittsburgh, Pittsburgh, PA, USA

²⁷University of California, San Diego, La Jolla, CA, USA

²⁸Institute for Therapeutics Discovery and Development, University of Minnesota, Minneapolis, MN, USA

²⁹Office of Cancer Complementary and Alternative Medicine, National Cancer Institute, NIH, Shady Grove, MD, USA

³⁰National Institute on Aging, NIH, Bethesda, MD, USA

³¹CENAPT and PCRPS, University of Illinois at Chicago College of Pharmacy, Chicago, IL, USA

Correspondence

Barbara C. Sorkin, Office of Dietary Supplements, National Institutes of Health (NIH), 6100 Executive Blvd., Ste. 3B01, MSC: 7517, North Bethesda, MD 20892, USA.

Email: barbara.sorkin@nih.gov

Funding information

HHS | National Institutes of Health (NIH), Grant/Award Number: NA

Abstract

While great interest in health effects of natural product (NP) including dietary supplements and foods persists, promising preclinical NP research is not consistently translating into actionable clinical trial (CT) outcomes. Generally considered the gold standard for assessing safety and efficacy, CTs, especially phase III CTs, are costly and require rigorous planning to optimize the value of the information obtained. More effective bridging from NP research to CT was the goal of a September, 2018 transdisciplinary workshop. Participants emphasized that replicability and likelihood of successful translation depend on rigor in experimental design, interpretation, and reporting across the continuum of NP research. Discussions spanned good practices for NP characterization and quality control; use and interpretation of models (computational through in vivo) with strong clinical predictive validity; controls for experimental artefacts, especially for in vitro interrogation of bioactivity and mechanisms of action; rigorous assessment and interpretation of prior research; transparency in all reporting; and prioritization of research questions. Natural product clinical trials prioritized based on rigorous, convergent supporting data and current public health needs are most likely to be informative and ultimately affect public health. Thoughtful, coordinated implementation of these practices should enhance the knowledge gained from future NP research.

KEYWORDS

clinical predictive validity, dietary supplements, model systems, rigor and replicability, value of information

1 | INTRODUCTION

1.1 | Background

Research on health effects of natural products (NPs) has often stemmed from traditional use, or from evidence that higher consumption of plant foods is associated with better health outcomes (see ¹ and refs. therein). This research has resulted in the successful development of many active ingredient-based pharmaceuticals from NPs. In contrast to a typical drug development pathway, studies based on traditional uses or epidemiological data rarely begin with a

precise understanding of the active molecules or their molecular mechanisms of action. A growing number of recent efforts to elucidate the mechanisms of action of traditionally used NPs have reported finding a complex mixture of bioactive constituents that act at multiple targets.²⁻⁵ Perhaps partly as a result of this still incompletely understood complexity, to date many NIH-supported clinical trials (CTs)^a of complex NPs (NPCTs), in particular the majority of large, randomized, controlled clinical trials (RCTs), have failed to reject the “null hypothesis,” that is, they detected no significant difference between the intervention and negative controls.⁶⁻⁸ While CTs that fail to reject their null hypothesis can

still provide invaluable guidance, the results of such trials are often met with concerns that different design choices—whether of product, dose, timing, participant eligibility criteria, outcomes, etc.—might have yielded a substantially different result. Each interventional trial represents a substantial investment, with well-powered Phase III CTs typically consuming more than \$10 million and many person-years.^{9,10} Additionally, as time and funds are limited, and participant pools may also be constrained, a decision to invest in a CT often precludes pursuing other research questions,¹¹ which may provide more useful public health-relevant knowledge than an RCT with an insufficient evidence base.

1.1.1 | Definitions

NPs were defined as including materials such as animals, marine organisms, plants/botanicals/herbals, macroscopic fungi, bacteria, probiotics, and minerals, and materials derived from them, including isolates, extracts, vitamins, and amino acids. NPs may vary in chemical complexity, ranging from crude to extensively purified materials, and were considered to include dietary supplements, certain foods, traditional medicines, and other products derived from multiple natural sources.

Complex NPs: NP preparations which retain a large number of chemical constituents whose relative concentrations and chemistry (metabolome composition) reflect the inherent variability of the sources, as opposed to products consisting of a very limited number of highly purified NP-derived constituents.

1.1.2 | Challenges to be addressed

For investigation of traditionally used NP, research design may be complicated by the chemical complexity and inherent variability of the NPs.¹²⁻¹⁴ The often incomplete and/or conflicting information on mechanisms of action, appropriate uses, and expected health effects poses challenges for formulating hypotheses and designing CT. The ready availability of a variety of preparations of similar NPs and their occurrence in complex matrices such as foods or dietary supplements further increases the challenges for the design, conduct, and interpretation of NPCTs.¹⁵⁻¹⁷ The variety of NPs in the marketplace poses challenges for product selection; in addition, participants must be monitored for inadvertent or otherwise unreported exposure to the intervention or components thereof (drop-ins). Here, consistent with the NIH definition, we use the term CT not only for efficacy trials but also for any study in humans which assesses NP safety, bioavailability, pharmacokinetics (PK), pharmacodynamics (PD), dose-response, mechanisms of action, and biological markers thereof.

1.2 | Focus and objectives

To discuss good practices for addressing these challenges, a two-day, public workshop on September 13 and 14, 2018 brought together speakers and attendees across disciplines and from academia, government, industry, and private practice at the NIH campus in Bethesda, Maryland. Past efforts to enhance rigor in NP research include the NCCIH Policy on Natural Product Integrity¹⁸ as well as numerous guidance documents addressing identification, reproducibility, study design, and execution, as well as conflict of interest issues.^{6,19-27} As was done for a recent workshop focused on fatty acids,²⁸ we assemble here relevant prior publications citing critical or state-of-the-art approaches drawn from a breadth of NP research and related disciplines, and address in greater depth NP-related topics that were specific foci of the workshop.

Reproducibility, replicability, and other critical characteristics of a solid evidential foundation have been variously defined in different sub-disciplines.^{29,30} Consistent with the 2019 NASEM Report on Reproducibility and Replicability in Science,³¹ this article uses the term reproducibility for “obtaining consistent results using the same input data, computational steps, methods, and code, and conditions of analysis,” and the term replicability for “obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.”

The workshop was predicated on the premise that adherence to a well-integrated set of good practices along the continuum from ethnobotany and epidemiology through basic biomedical research to CTs should enhance the effectiveness of translation from preclinical research and early phase CTs to information usable by consumers, manufacturers, practitioners, research funders, and other stakeholders. The good practices should also help reduce premature initiation of Phase III NPCTs, thus reducing instances producing strong grounds for post-trial concerns about the evidence base for a given trial. This article summarizes and updates the workshop discussions of such practices, which include those that apply to individual sub-disciplines, as well as productive combination of approaches across disciplines. The integration of all of these practices, along with the application of sound inductive reasoning, rigorous, appropriate, and transparent statistical analyses (Table 1),³²⁻³⁶ as well as targeting an effect size that is significant to clinicians and consumers³⁷ should increase the replicability and translational relevance of basic and preclinical NP research. Ultimately, this will serve to limit resources expended on flawed premises producing irreproducible data,³⁸⁻⁴⁰ and to ensure optimal knowledge gained from every NPCT, regardless of whether the results are consistent with benefit, harm, or absence of an effect on the assessed outcome(s).

This review summarizes areas of both consensus and contention that emerged from the workshop discussions.

TABLE 1 Highlights of broadly applicable good practices for enhanced replicability, transparency, and translational relevance of NP research

1. *Models*, by definition, differ from the actual subject of interest, however, when carefully chosen to address a clearly defined experimental question they can provide critical insights.²³⁵ Sound inductive reasoning, including careful consideration of differences between the model and the actual subject of interest, must be used in interpreting model system outcomes, for example, response in an experimental model may demonstrate a biologically relevant activity, but does not imply clinical efficacy.
2. *Quality control*. Ongoing, discipline-appropriate quality control is critical for all experimental assays and research components, from NP, through cell lines, microphysiological systems,⁵⁸ in silico, and whole-organism models, to CT procedures.
3. *Replication* is critical: within and across distinct model systems, with methods based on orthogonal principles, and in multiple labs. Results that replicate across models, methods, labs, are more likely to replicate in a CT.
4. *Rigor*. The same rigorous practices used in CT to avoid bias and enhance replicability should be applied in hypothesis-testing experiments that may be used in designing a CT. This includes randomization of wells/dishes/animals, masking/blinding of study personnel, and sample size calculations. Exploratory studies merit equally complete and transparent reporting but may accommodate less stringent procedures.
5. *Reporting*. Use approaches and venues that support more complete, transparent, and cumulative reporting.
 - a. *Registration*: use registered reports and pre-publication registries for preclinical in vivo studies as well as CT, for example, clinicaltrials.gov and preclinicaltrials.eu. Information needed to replicate research on complex NPs may go substantially beyond the intervention characterization required by these sites, see, for example, nccih.nih.gov/research/policies/naturalproduct.htm
 - b. *Outlet*: publish in journals that do not limit supplemental information.
 - c. *Publish*: exploratory studies (without forcing them into hypothesis-testing frameworks) and null results of hypothesis-testing studies.
 - d. *Openness*: curate key information for scientific research in publicly accessible databases, for example, curatescience.org.

While the workshop focused on approaches to improve the knowledge gained from future randomized, controlled, NPCTs, some attendees questioned the appropriateness of and need for such trials, especially for traditionally used botanical preparations, arguing that beneficial effects of such practices cannot be detected in the context of a conventional CT.

Moving from a narrower to a broader and more applied focus, this review begins with good practices for the application of assays used to elucidate NP bioactive constituents, and continues through the specification and characterization of NP interventions, endogenous and

exogenous factors that may influence the translational relevance of preclinical in vivo models, and the strengths and weaknesses of a broad range of state-of-the-art models. Subsequently, the review will address additional factors that influence the likelihood of successful translation to humans, including assessment of the supporting research, and concludes with approaches for evidence-based prioritization of this research. The review is organized into four main sections:

- Good practices for preclinical NP studies
- NP characteristics critical for replicability
- Translational relevance of preclinical NP research
- Good practices for moving to NPCTs

Figure 1 provides the graphical overview of the relationships among the covered topics, and Table 1 highlights broadly applicable good practices. The subsequent tables highlight good practices from each of the four main sections:

- Table 2 highlights the good practices for the use of in vitro and in silico models.
- Table 3 highlights the good practices for NP selection, formulation, and characterization.
- Table 4 highlights the good practices for the use of preclinical in vivo models.
- Table 5 highlights the critical considerations in NPCT design.
- Table 6 highlights the issues to consider in NPCT decision-making.

2 | GOOD PRACTICES FOR PRECLINICAL NP STUDIES

2.1 | Optimizing the use of in vitro, invertebrate, and in silico models and ethnobotanical approaches

Good practices in model systems (in vitro, invertebrate, in silico) and in translating from traditional medical systems^b to clinical research can enhance the replicability and translational relevance of preclinical NP studies. To date, the application of invertebrate model systems and some newer in vitro models has been very limited in NP research.

The measurement of activity in in vitro assays is a typical starting point for research on potential clinical application(s) of a NP. Good practices in proceeding from other starting points are discussed below. In vitro assays are also commonly used for monitoring the stability and potency of NPs, as well as for identifying fractions or components responsible for the

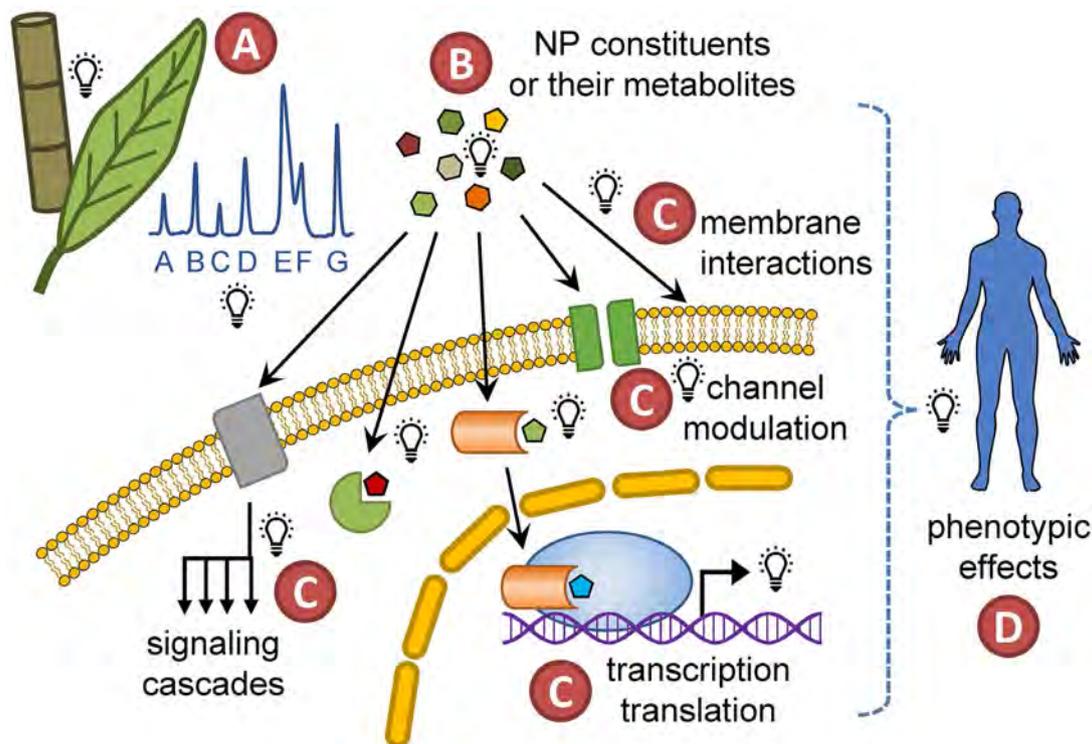


FIGURE 1 Critical information for the design, implementation, and interpretation of natural product clinical trials. Elucidating details (light bulbs) of molecular mechanisms of action prior to initiating a NPCT can help “de-risk” the planned trial by both providing evidence for the hypothesized outcome and its underlying causes, and supporting the development of methods to test that hypothesis. It is a good practice for researchers to establish standard procedures for product preparation (A), that are optimized for bioactivities and stability, and methods, preferably validated, for thorough characterization of product chemistry and stability. Elucidation of the bioactive NP constituents or metabolites (B) is critical for quality control, for assessing bioavailability, PK and PD, and determining whether minimal effective concentrations of bioactives have been achieved in plasma or at the targets (C). These data can support the development of methods to test mechanistic hypotheses, for example, by measuring in vivo engagement of the putative targets, proximate bioactivities, or PD, and/or assessing the relationships between these and (D) the primary and secondary NPCT outcomes, and any adverse events

assayed bioactivities. The replicability of these assays depends on many inputs, beginning with the test article, which often is a chemically complex NP. Good practices to enhance the replicability of vertebrate cell culture-based assays have been described elsewhere,^{38,39,41} as have recommendations for the use of antibodies.⁴² The following, therefore, focuses on issues more specific to the use of NPs in cell-based assays, and on approaches to enhance the translational relevance of those assays.

Data generated by high-throughput screens (HTS) and in vitro bioactivity assays pose many challenges for interpretation, importantly including their high potential for false discovery. Small molecule NP constituents and their spontaneous breakdown products and/or metabolites^{43,44} have significant potential to generate in vitro assay artefacts via aggregation,⁴⁵ fluorescence, light absorption, metal chelation, plasma or intracellular membrane disruption,^{46,47} protein denaturation, and/or specific and non-specific reactivity with biological nucleophiles (Table 2). For example, HTS and other in vitro studies of extracts and other preparations derived from *Curcuma longa* (turmeric; often improperly designated as “curcumin”) are susceptible to confounding due to the low

stability⁴⁴ and high reactivity of the curcuminoids, as well as the frequent lack of clarity as to the actual composition of a given preparation.^{48,49} Workshop speakers emphasized the need to design and conduct in vitro experiments in ways that mitigate and/or reveal such artefacts,^{41,50,51} for example, by time course and chemical analyses, as well as by comprehensive and transparent reporting of experimental parameters.

2.2 | In vitro models

In vitro approaches leverage simplified models of more complex biological systems in the interests of higher throughput and increased ability to decipher causal mechanisms (Table 1). In vitro models that better reflect the biology of an intact organism have been developed, taking advantage of advances including the use of primary, patient-derived cells and stem cells, direct differentiation of induced pluripotent stem cells, more precise methods of gene replacement, and identification and quantification of heterogeneity in vivo and in vitro. In vitro models are

TABLE 2 Highlights of good practices for the translational relevance of in vitro models assessing NP bioactivities and mechanisms of action

1. *Artefacts.* When using in vitro assays of NP activity, test for common types of NP assay interference, for example, fluorescence or absorption, membrane disruption, protein denaturation, and specific and non-specific reactivity with nucleophiles.⁴⁸
2. *Bioavailability.* If in vitro activity is observed only at concentrations substantially higher than those achievable in vivo in the tissue of interest, then the probability of seeing a robust clinical effect of consuming the NP is likely insufficient to form the basis for a NPCT.
3. *Non-specific molecular mechanism(s) of action.* Include controls for non-classical mechanisms of action such as metal chelation,⁸⁷ modulation of DNA modifying enzymes, or intercalation into plasma or subcellular membranes.^{46,87}
4. *Biomimetic in vitro models.* Consider pros and cons of more morphologically detailed “3D” models using human-derived cells. Organoids, “organs-on-a-chip” (OoC), and multi-organ biomimetic models (eg, human-derived intestine, liver, and kidney OoC coupled by appropriate microfluidics) can yield more translationally relevant high-throughput information than monolayer cell cultures, and are a powerful complement to preclinical in vivo models.

also increasingly leveraging advances in tissue engineering, microfluidics, and real-time biosensors of physiological processes. These in vitro models range in the extent to which they mimic in vivo conditions from relatively simple spheroids,⁵² through patient tissue-derived organoids, self-organizing stem-cell-based organoids,^{52,53} single “organs on chips” (OoC) under flow, and integrated, multi-OoC microphysiology systems (MPS).⁵⁴⁻⁶⁰ While still relatively high-throughput, these models from both healthy and diseased tissues have greater similarity to the chemistry, three-dimensional structure and component localization, environment, and possibly genetic stability of an intact organism than classical cell cultures grown on relatively flat substrates. Moreover, through the use of human-derived cells, these models may partially address the challenge of species differences in responses (Table 2).⁶¹⁻⁶⁴ MPS also allow targeted assessment of cell-type- or organ-specific dose responses (Table 4). Some major challenges in the use of MPS and other complex in vitro systems have recently been addressed, including the development and implementation of metrics for replicability and approaches to prevent the artefactual binding of hydrophobic molecules to the devices.^{59,65} Models incorporating elements of the human immune system are beginning to be developed.⁶⁶ However, understanding how to translate results from in vitro systems such as MPS to CT, for example, into calculated dose ranges remains challenging.

Only those effects observed in vitro at concentrations achievable in vivo following ingestion, or in systems that

incorporate upstream components modeling the functions of the gastrointestinal tract,⁶⁷ are likely to be clinically relevant to oral dosing. Activities observed only at higher concentrations than can be obtained following ingestion should not be over-interpreted in planning CTs of ingested products (Table 2). Notably, many NP constituents show very limited oral bioavailability. Ex vivo models may be used for extensive pharmacodynamic characterization of NP, but they do not fully recapitulate human physiology. For instance, a promising NP candidate in an in vitro cell screening assay may be metabolized by the liver into an ineffective or even toxic byproduct.⁶⁸ In vivo validation of in vitro results remains a key component of preclinical studies.

Speakers described several approaches to manage the sometimes limited or conflicting information on NP biological activities and their underlying mechanisms, including determining which NP components and cellular and molecular targets are required for activity in animals. The application to chemically complex NPs of agnostic (or non-targeted) in vitro assays⁶⁹ combining cutting-edge computational approaches, optimized mass spectrometric (MS) data collection and analysis to elucidate chemical composition, and cytological profiling^{70,71} now enables the generation of strong and specific hypotheses regarding bioactive NP components and the mechanisms through which they affect gene expression or cell phenotype.⁷²⁻⁷⁶ Such systems have the ability to interrogate multiple chemical components and their combinations simultaneously, without purification of individual constituents. This ability is critical, given accumulating evidence (consistent with tradition) that multiple chemical constituents in even a single botanical extract may contribute to bioactivity via distinct or similar molecular mechanisms.^{77,78} These may include effects by constituents on each others' solubility (or other aspects of bioavailability) or toxicity.^{5,79} These novel, non-targeted platforms also have the advantage of generating strong, testable mechanism of action hypotheses without requiring assumptions or prior knowledge regarding those mechanisms.

2.3 | Invertebrate in vivo models

Invertebrate models such as *Caenorhabditis elegans*^{80,81} and *Drosophila melanogaster* are popular in a variety of research areas for their ability to provide relatively high-throughput, but in vivo platforms. The conservation of many human genes and signaling networks, combined with the small size and relatively short lifespans of these models, facilitates the cost-effective evaluation of a variety of phenotypes (eg, reproduction, locomotion, cognition) in a large number of individuals (Table 4).⁸⁰ This provides the ability to study phenotypes that are challenging to assess in most vertebrate animal models, such as healthspan (the part of an organism's lifespan during which it is healthy).^{82,83} Invertebrate models can also be

used in more focused interventional and mechanistic studies that explore specific aspects of an established phenotype.⁸⁴⁻⁸⁶ For example, tea polyphenols were reported to modulate both stress responses and fertility in *D. melanogaster* strains⁸⁷ via a non-target-ligand mechanism of action (Table 2), by modulating iron homeostasis. Recent advances in omics technology and the use of outbred populations further expand opportunities for using *D. melanogaster* as a model system for NP preclinical studies.

2.4 | In silico models

When information on NP kinetics, dynamics, bioavailability, clearance, and pharmacologic activities is limited, one useful approach is to apply computational or in silico methods to leverage information from products with known chemistry, mechanisms of action, and activities. Data from more highly characterized products can be integrated into dynamic models, and in vitro to in vivo extrapolation (IVIVE) can be used to predict biological responses.⁸⁸⁻⁹² These approaches, as outlined by the Tox21 Federal collaboration and the U.S. Environmental Protection Agency's ToxCast program,⁹³⁻⁹⁶ can be used to focus high-throughput in vitro and in vivo testing on compounds and concentrations that are more likely to be biologically active and appropriate for each model.^{97,98} Such integrated approaches could be further developed to probabilistically model the pharmacological and toxicological kinetics and dynamics of complex NPs.⁹⁹

Speakers described several critical, often broadly applicable considerations for utilizing in silico models. First, the purpose of the model must be clearly defined, and the model appropriate and sufficient to fulfill the purpose. Second, as for any research, if experiments that provide the input data are not appropriately designed in the beginning, no amount of computational modeling will yield translatable information. Third, variability and uncertainty in the data that inform the model, as well as correlations between parameters must be considered. Uncertainty and variability in model outcomes are acceptable but must be quantified, and in optimized models they should be similar to real-world variability. Preliminary models may be continuously refined, using a "learn and confirm" strategy, or as understanding, for example, of NP pharmacology increases. Fourth, end users must understand the computational (or any other) model's limitations, the extent to which it is similar or dissimilar to the biological system of interest, and how the outputs compare with population-derived data, other real-world data, or with data from other models. Finally, computational models and the data used in their development must be made findable, freely accessible, interoperable and reusable (FAIR),¹⁰⁰ just as information for in vitro and preclinical in vivo models must be findable and sufficiently transparent to support replication.

2.5 | Ethnobotanical approaches

Another typical starting point for translational research is traditional NP use, particularly the use of plants in traditional medical systems. When basing NP research on traditional use, it is essential to consider and thoroughly document how the NP is sourced, identified, prepared, and used in the traditional contexts (Table 3). Intra-species variation and growing conditions strongly influence the metabolic profile of a plant and differences in harvest and preparation will also alter critical product characteristics.^{101,102} Many different uses may be described for a given species, and traditional healers have frequently been reported to use only specific subtypes of a given species. Awareness of such practices has the potential to be highly informative for the identification of bioactive components.¹⁰³ Analogous considerations apply to fungal and animal sources of NP. Good practices for reporting botanical NP sourcing, identification, characterization, vouchering and archiving have been previously described elsewhere.^{12,18,19,104-106} Comprehensiveness and transparency in reports of biomedical research on traditionally used plants have increased in recent years.¹⁰⁷ There

TABLE 3 Highlights of good practices for NP integrity including standardization, formulation, characterization, and quality control

1. *Quality control.* Identity and purity of all materials used, regardless of the supplier, should be verified (independent of supplier) prior to use. Appropriate storage and handling conditions should be maintained and monitored by stability testing at least throughout the duration of the study (consult, eg, the NCCIH Natural Product Integrity Policy¹⁸ and the NSF Stability Testing Guidelines for Dietary Supplements). To the extent feasible, comprehensive chemical analysis of product should be performed; minor product components or impurities can be responsible for or affect measured activities.
2. *Product specifications:*
 - a. *Chemistry.* Identification of product components required for the bioactivities studied provides critical information for studies of bioavailability, pharmacokinetics (PK), and pharmacodynamics (PD) and should enhance the replicability and clinical relevance.
 - b. *Formulation.* Consider the effects of NP formulation and matrix (eg, capsule, liquid, food) on bio-accessibility, bioavailability, PK, PD, and activity. These may differ significantly between an isolated NP constituent and the same compound in the context of a complex food matrix.
 - c. *Sensory characteristics.* Sensory acceptability must be considered for all in vivo research, and effectiveness of masking of placebos or active comparators must be considered in CT.
3. *Translation from ethnomedicine.* If basing an intervention on ethnobotanical observations, consider replicating traditional practice as closely as is feasible. For example, highly processed extracts will likely have different activities than a multi-botanical decoction (see 2a and 2b). Contextual differences (environment, behavior, genetics) may also affect the outcomes.

are cases, however, where the available records predate the availability of geolocation data on growing sites, or where the need to protect ecologically or culturally sensitive information may limit what can be published.^{108,109} Where a similar use or outcome is reported for the same or closely related species across cultures and geographical regions, those similarities may provide priority clues for translational research.

Translation from traditional to modern practice may be limited by important differences¹¹⁰ (Table 3). Traditional practices commonly combine multiple NP sources (eg, multi-botanical) into one intervention and are often associated with specific concurrent behaviors (eg, rituals, behavior changes, possibly including dietary changes). Moreover, if documented traditional practices and the NP outcomes to be studied differ substantially, it may be inappropriate to claim the traditional or typical use as supporting data; the differences between modern medical diagnoses and practices and traditional principles of diagnosis and practice may be too divergent to base claims for modern application on traditional use. Moreover, where a planned NPCT is expected to use a highly processed (single) botanical, information from traditional use may also lack relevance, as the bioactivity and/or the absorption, distribution, metabolism, and excretion (ADME) properties of the highly processed product may differ substantially from those of a (multi-component) traditional preparation. Where the NP and the study population are close to historical use and users, traditional practice may provide initial parameters for dose range to be used in research. In instances where no clear typical usage exists, for example with green tea extract capsules, speakers recommended that preclinical studies prioritize the determination of a therapeutic window and administer the lowest dose to achieve concentrations required for the biological and/or health effect of interest.⁹

3 | NP CHARACTERISTICS CRITICAL FOR REPLICABILITY

3.1 | NP specifications and integrity

The ultimate biological effect of an ingested, complex NP results from a dynamic interplay of factors: NP chemical composition, physical characteristics,^{111,112} bioaccessibility, bioavailability, metabolism, and target population characteristics including sex, age, genetics, background diet, and other environmental exposures.¹¹³ This subsection begins by summarizing issues inherent to the integrity of the NP intervention and subsequently addresses target-related sources of variability.

Replicability of outcomes depends on correct and clearly reported identification and archiving of source

materials^{12-14,17} and consistent preparation and activity of the NP.^{17,22,26,114} Some attendees felt that since the effects of complex NP preparations may not be due to only the identifiable components, transparent reporting of methods, including consistent product preparation and quality control via an in vitro assay relevant to bioactivity, is sufficient for replicable NP research. Others argued that in many research context it is also critical to understand which NP components are responsible for the bioactivities of interest, as this information can inform NP quality control, as well as being utilized in assessments of bioavailability, pharmacokinetics and to inform comparisons of different studies¹¹⁵ (Table 4).

Minor product components, impurities, instability, and chemical changes due to instability, resulting in the occurrence of degradation products can be responsible for some or even all measured activity. Instability has been demonstrated for a variety of prominent bioactive botanical marker compounds. These include ligustilide (a constituent of dong quai), which is very unstable and prone to redox reactions and photo-induced dimerization;^{43,116,117} desmethyloxanthohumol, a chalcone found in *Humulus lupulus*, or hops, that forms the phytoestrogen, 8-prenylnaringenin, via Michael addition;¹¹⁸ and curcumin, found in turmeric, which is known to form a myriad of degradation products and metabolites.⁴⁴ These chemical changes can modify bioactivity through a variety of mechanisms^{97,119,120} and represent factors that appear to be minor (Residual Complexity, see go.uic.edu/residualcomplexity), but may significantly alter biological activities and therefore require monitoring. A number of organizations have published general guidelines for stability testing, which are relevant for dietary supplements and other NPs, including the NCCIH Policy on Natural Product Integrity¹⁸ and guidelines published by the International Council for Harmonisation (ICH; ich.org).

Variations in growing conditions generate variations in relative levels of NP constituents; differences in constituent levels are also seen between closely related species interchangeably used in similar contexts NP.¹²¹ For example, the licorice source plants, *Glycyrrhiza glabra* L. and *G. uralensis* Fisch. ex DC differ significantly in the ratio of liquiritigenin to isoliquiritigenin, as well as in congeneric constituents, and are therefore expected to have different bioactivities.^{122,123} Characterization sufficient for the detection of such minor components, known or unknown, is therefore critical for quality control and for monitoring batch-to-batch variation and stability.^{15,124} Identity and purity of all materials used, regardless of the supplier, should be verified prior to use.^{14,124} Descriptions of approaches to NP selection and characterization have been published recently.^{17,22,114,125}

3.2 | Chemical and metabolic factors

ADME is a multistep, dynamic biological process that may be modified by product characteristics as well as by host characteristics^{115,126-129} including physiology and behavior. NP processing methods and final formulation can alter the chemical profile, stability, bioaccessibility, bioavailability, and metabolism of a NP.^{111,130} For example, soy germ isoflavone glucosides incorporated into pasta can be hydrolyzed to aglycones by wheat β -glucosidase.¹³¹

In contrast to most preclinical models, humans are generally diverse in genotypes, diets, other behaviors, and environmental exposures. Moreover, host genetics and cultural milieu profoundly affect flavor perception.¹³² If the NP will be delivered in or as part of a food, liquid, capsule, or tablet, acceptability of flavor, texture, and or tablet size are essential. For both CT and generally available commercial products, consumer burden is another critical factor; this may include the number of capsules required, ease of swallowing capsules or tablets, cost, and portability. Thus, sensory testing and burden, as well as stability of the NP preparation are critical for successful translation of an intervention (Table 3).

Human microbiota are generally much more varied than those of lab animal models, reflective of both genetic diversity¹³³ and differences in past exposures.¹³⁴ Run-in periods prior to pharmacokinetic and metabolic assessments can increase the relevance of preclinical animal models to long-term dietary patterns and provide important information for comparison with post-intervention metabolite profiles. A critical consideration is that, even where preclinical in vivo studies are designed to better model (CTs), the models usually differ from humans in their greater uniformity of background diet, control of prior exposures to the relevant NP, dietary requirements, and metabolic capacities (Table 4). For example, taurine is an essential nutrient for cats,¹³⁵ but synthesized by humans and rodents, while ascorbic acid is essential for humans, but synthesized by most other animals.¹³⁶

Human genetic differences and variations in gut microbiota cannot be considered to be independent, as both direct and indirect effects of host genetics on gastrointestinal ecology have been reported.^{133,134,137,138} Both background diet and bioactive NPs affect host physiology directly, as well as through modulation of the microbiota, and the microbiota in turn contribute to modulation of NP processing.^{139,140} Non-absorbed polymeric polyphenols may, for example, modulate absorption of polyphenol monomers in humans, such that history of exposure significantly influences pharmacokinetics.¹⁴¹ Effects on the host may, in turn, modulate microbial ecology and metabolism, while changes in microbial populations or gene expression may modulate host physiology, including ADME, collectively resulting in dynamic host-microbiota interactions that modulate or even generate the observed responses to ingested interventions.¹⁴¹⁻¹⁴⁶

Gastrointestinal microbiota have been shown in many cases to be required for the generation of bioactive metabolites of NPs. For example, gut microbiota can convert grape constituents to phenolic acids such as 3-hydroxy benzoic acid and dihydrocaffeic acid, which have been reported to inhibit β -amyloid and α -synuclein oligomerization, respectively, and epigenetically modulate gene expression.^{3,147} Speakers stressed the importance of a strong, though not necessarily complete, understanding of the pharmacokinetics and metabolism of a NP intervention for optimizing the design of large-scale CT, although opinions differed on whether such knowledge was necessary before pursuing early-phase trials.

Presenters highlighted cases where the biological activities of enantiomeric NPs differ significantly. For example, a cancer preventive effect of soy in rodents could be replicated using the *R*-(+) enantiomer of the isoflavone metabolite equol, but not the *S*-(-) isomer produced in vivo by the gut microbiota.¹⁴⁸⁻¹⁵² The microorganisms responsible for this conversion are widely present in rodents, but only found in only 20%-30% of adult humans, a difference that may contribute to discrepancies between observations in rodents and humans. This finding emphasizes the importance of considering, characterizing, and fully reporting the chirality of NPs and their metabolites, as well as of understanding the role of the gut microbiota. It was pointed out that translational relevance of in vitro HTS assays might be substantially improved through the integrated use of replicable model systems able to mimic the generation of translationally relevant metabolites by the target animal or its gut microbiota.

4 | TRANSLATIONAL RELEVANCE OF PRECLINICAL NP RESEARCH

4.1 | General considerations

Preclinical in vivo models are widely used in biomedical research, and are often considered a prerequisite for CT. These models provide critical information on the biological response to a product that in silico and in vitro models cannot currently provide in a reliable fashion, including on bioavailability, metabolic activation and deactivation, PD, and toxicology. In vivo testing also provides a basis for choosing dose ranges to be tested in CT.

In selecting any in vivo preclinical model system, investigators must keep in mind ways in which the differences between the model and the clinical application of interest are likely to affect inference from the study.²⁰ The model selected should be appropriate for the experimental question based on the organism characteristics considered most critical to the target biology. The likelihood of translational relevance of a preclinical model to a clinical or other experimental question

can be considered in terms of different types of model validity, including face validity (similar phenotype, lower probability of successful translation), construct validity (similar mechanism and phenotype, better probability of successful translation), and clinical validity (demonstrated clinical predictive success)¹⁵³⁻¹⁵⁶ (Table 4). For example, short-lived invertebrate models may be particularly advantageous to understand NP effects on healthspan, as these species conserve the target-of-rapamycin, insulin/IGF-1-like signaling, and sirtuin pathways, which have been shown to modulate healthspan in organisms ranging from yeast through *C. elegans* and *D. melanogaster* to mammals.⁸²

Many factors that differ between preclinical in vivo models and people may influence the biological effects of a NP. Background diet and dietary history can influence NP absorption and metabolism. For example, the absorption and PK of flavan-3-ols are altered by exposure history, as well as by host factors such as obesity and diabetes.^{112,141,157,158} This demonstrates the critical importance in research on humans of considering, assessing, and reporting background diet and history, as well as standard biometric parameters. Assessing correlations among human genomic, transcriptomic, proteomic, or metabolomic data, and between these and outcomes following a standardized exposure may elucidate mechanisms that underlie differences in the absorption or generation of, or exposure to, bioactive constituents or bioactive metabolites,¹²⁸ and allow prediction of responders versus non-responders. These approaches can enhance translation between epidemiological, preclinical, and clinical studies.

4.2 | Selection of a mouse model

Mice are frequently the model species of choice for in vivo preclinical research, due to their relatively small size and short lifespans, and the large number of strains and large body of literature available. Nevertheless, mouse and other in vivo preclinical models differ from humans in many known and unknown details, both intrinsic (eg, genetic, immune,¹⁵⁹⁻¹⁶⁴ surface area: volume ratio, lifespan¹⁶⁵) and extrinsic (eg, captivity, diet, history), which, depending on the research question, may substantially affect translational relevance.¹⁶⁶

It is critical to remember that each distinct inbred strain may respond differently to an intervention^{126,167} that the same strain supplied by a different vendor (or with different gut microbiota) may respond differently,¹⁶⁸ and that sex differences in response may appear only in specific species or strains.¹⁶⁹⁻¹⁷²

A wide variety of approaches are used to increase the predictive validity of mouse models. Transgenic mice are commonly employed to enhance construct validity; guidelines for rigor and transparency in research using mouse mutants have

TABLE 4 Highlights of good practices for preclinical in vivo research on NP

1. Product translational relevance and replicability:
 - a. *Equivalent product formulation.* NPs used in preclinical studies should use similar or analogous formulation(s) to the intended human studies.
 - b. *Biotransformation.* Track the actual bioactive compounds, which may be generated from precursors in the NP by the host or the host's gut microbiota. Either or both may transform NP constituents to active or inactive compounds. These transformations may differ between models and humans, and among humans, and affect replicability of results.
 - c. *Dose-ranging.* Doses used in models should be comparable to anticipated human dosage. Initial dose ranges may be based on traditional use or MPS dose-response studies, derived using PBPK modeling, or from preclinical in vivo studies using appropriate scaling approaches. A clinical study is still required to establish dose-response, assess safety, and determine PK and PD in the target population.
2. Model translational relevance:
 - a. *Model selection.* Consider the translational relevance for the specific research question of different species and strains, humanized, outbred, wildling, and/or genetically modified models. Heterogeneity in the model organisms may contribute to the ability to predict responders vs non-responders.
 - b. *Documentation.* Document as much as possible. Endogenous (eg, age, sex, physiological/health status, genetic), and external factors (eg, environment, including background diet, exposure history, enclosure size and enrichment, circadian timing, season, duration and timing of exposure relative to outcome assessment, handlers and handling methods) may affect outcomes. Make these records findable, accessible, interoperable, reusable (FAIR).¹⁰⁰
3. *Validated methods for de-risking the NPCT (see also Figure 1):* Develop and use validated assays to assess achievement of go/no-go criteria based on as many components of the mechanistic model as possible (see Figure 1), for example, achievement of effective concentrations of the bioactive component(s) at in vivo targets, target occupancy, and modulation of relevant biological networks.
4. *Specimen collection, analysis, and archiving.* Collect and appropriately document and store specimens needed for assays described in point 3, and as applicable, fecal, or other relevant samples to facilitate assessment of any association of the gut microbiota with differential responses. In analyses, include approaches to control for known artefacts, such as reagent contamination that can strongly affect outcomes in low biomass microbiome specimens.¹⁹⁵

been published.¹⁷³ Transgenic constructs allow experimentation on specific human pathways and targets of interest. Where the genetic mechanism of interest is unknown, for example in toxicity testing^{166,174} or research on neuropathic pain,¹⁷⁵ the genetic heterogeneity of outbred mice¹⁷⁴ may provide important information that is unavailable from inbred strains without otherwise affecting outcome variability.¹⁷⁶ Many different approaches to “humanizing” mice have been utilized to enhance predictive validity for specific research disciplines

or questions. Employing the severe combined immune deficiency mouse (SCID) is one such approach, whereby human peripheral blood cells, and cells of other human fetal immunological tissues such as bone marrow, thymus, or liver are introduced (usually an NOD-SCID- γ mouse).¹⁷⁷ Another approach involves developing transgenic and chimeric mice in which specific human genes are expressed. Various approaches have also been utilized to generate mice with “humanized” gut microbiota, which have been found to have significant clinical predictive validity in developing anti-infective, autoimmune disease, cancer, and transplant therapeutics.¹⁷⁸⁻¹⁸⁰ Inbred mice reconstituted with microbiota from wild mice have also been reported to have increased translational value.^{181,182} In their natural environment mice, like humans, are exposed to microbes, including pathogens, and thus should be regarded as metaorganisms, with their own specific microbiota (including bacteria, fungi, and viruses). Returning natural microbiota to standard laboratory mice was found to significantly improve the model’s ability to predict human immune responses in two preclinical studies.¹⁸³

4.3 | Model- and context-related considerations

Optimization of preclinical research for relevance to future NPCTs as well as general human use requires not only careful consideration of the choice of models but also of intervention dose,¹⁸⁴ composition, and delivery (Table 4). As discussed above, when research and intervention are founded on a traditional use, available ethnobotanical and ethnopharmacological information may be used for guidance (Table 3). Clinical and preclinical interventions should be as similar or, where more appropriate, as equivalent as is feasible, bearing in mind species differences,¹⁸⁵ as well as any age-, sex-, size-, or health status-related differences which may affect behavior, absorption, metabolism, and/or toxicity. Allometric or mechanistic modeling approaches can be used to scale to human dosing from preclinical *in vivo* data, and vice versa.^{151,184,186}

Background diet, as well as other environmental exposures, including medications, alcohol, smoking, and air quality may also have important effects on NP bioavailability, metabolism, and biological effects in both humans and animal models. In the context of NP-diet interactions, it is critical to consider species-specific dietary requirements and optimal diets (eg, differences in ascorbic acid requirements,¹⁸⁷ optimal amino acid or lipid compositions),¹⁸⁵ as well as species differences in biosynthetic abilities¹⁸⁸ and digestive anatomies.¹⁸⁹ When comparing results across studies, it should be noted that differences in local regulation of vertebrate animal research (eg, the European Union prohibition on feeding animal-derived ingredients to herbivores) result in substantial differences in commonly used basal diets

or chows. Researchers must carefully consider, for a given research question, whether background diets should be the same for different species, or at similar levels of optimization for each, or whether both possibilities need to be included as controls. As discussed below, physiologically based pharmacokinetic (PBPK) models may be helpful in making dose adjustments for different subpopulations.

Diet and circadian rhythm interact in their effects on health, gut microbiota, and metabolism.^{190,191} These effects are modulated by sex in organisms ranging from fish and mice through humans.¹⁹²⁻¹⁹⁴ Through their effects on gut microbiota, NPs may have important effects on their own metabolism and on the generation of active metabolites which, in turn, may significantly affect both PK and PD. In addition, changes in the gut microbiota may exert separate direct or indirect effects on animal biology and health.¹⁶⁸ Good practices for characterization of the gut microbiota have been discussed elsewhere.¹⁹⁵

Preclinical *in vivo* models are often used in environments that are optimized for the investigators and different from the species’ natural habitat. This context by itself is likely to affect experimental outcomes and may also modulate the effects of an experimental intervention. In either case, the experimental environment is likely to affect translational relevance. Investigators need to acknowledge these often unavoidable factors and consider them when inferring the translational relevance of studies.¹⁵⁶ In general, given the range of interrelated factors that may modulate the effects of a dietary intervention, it is critical that supporting data include experiments rigorously designed to falsify the hypotheses tested, for example, to test whether the proposed bioactive constituents or metabolite and/or proximate targets are indispensable and inducing for the effect studied (Figure 1).^{35,196}

The robustness of a result is optimally assessed by attempting to replicate the study not only in the same model, by those originally reporting the result and others, but also across models based on a range of species, genotypes, and phenotypes. Recording as much information as possible about the model and the environmental conditions under which it was used may help others to replicate the studies or facilitate exploring why a finding failed to replicate, or to translate to another model or species.

4.4 | Genetic diversity and gene-NP interactions

One critical difference between most preclinical *in vivo* models and humans is that the former are usually inbred, except as discussed above. Where the bioactive compounds are known, and population differences in health response have been reported, interrogating those differences may lead to the elucidation of underlying gene-environment interactions. For example, common genetic variants in two fatty acid

desaturase genes (FADS1 and FADS2) appear to modulate the response to dietary polyunsaturated fatty acids (PUFA), resulting in differences in resulting levels of pro- and anti-inflammatory lipid mediators and differences in risk for inflammation-associated health outcomes such as asthma, rheumatoid arthritis, and cancer.¹⁹⁷ Tracking the effects of prevalence of these variants might shed light on inter-trial differences in NPCT outcomes with PUFA interventions.^{197,198} Similarly, a variety of factors including biometrics and genetic polymorphisms appear to be responsible for observed differences between black and white American women in associations between serum 25-hydroxy vitamin D levels and bone fractures.¹⁹⁷⁻¹⁹⁹

4.5 | In silico modeling approaches

Quantitative or computational modeling and simulation are beginning to show their utility in optimizing NPCT design. Quantitative systems pharmacology (QSP) encompasses computational methods including statistics, pharmacometrics, PK and PD modeling, mechanistic or PBPK models, in vitro-in vivo extrapolation, and machine learning approaches. QSP is often employed in drug development to integrate data, optimize experimental dosing regimens, identify sources of inter- and intra-individual variability, predict efficacy or toxicity, or estimate appropriate doses for different subpopulations.²⁰⁰⁻²⁰² PBPK models incorporate physicochemical data (eg, molecular weight, pKa, lipophilicity), in vitro or preclinical data (eg, on metabolism and transport), and population-specific data (eg, organ volumes, tissue composition, blood flow) to model ADME. PBPK can be used to predict equivalent exposures for humans and other species, for both translational and reverse-translational studies (Table 4). Sensitivity analyses can allow comparison of the magnitudes of the effects of different parameters on the outcomes. These models may also provide insight into the interplay between metabolizing enzymes and transporters, inter-individual variability due to intrinsic (eg, pharmacogenomics, age, sex) and extrinsic differences (eg, drug, food, or other environmental exposures). Perhaps most importantly, these computational models can highlight the most significant gaps in the data,²⁰³⁻²⁰⁵ as well as support evidence-based selection of initial dose ranges for CT.

The application of quantitative pharmacology to complex NPs is in its infancy. In addition to the inherent chemical complexity of NPs, understanding of the pharmacology and metabolism of the individual chemical components is mostly very limited. Nevertheless, application to NPs of modeling and simulation approaches that incorporate uncertainty and variability can provide estimates of differences among populations, inform dose selection and go/no-go decisions, and provide early estimates of the probability of rejection of the null hypothesis.

5 | GOOD PRACTICES FOR MOVING TO A NPCT

As outlined above in the section “NP Characteristics Critical for Replicability,” speakers emphasized the importance of having rigorous and replicable data on the chemical components of the NPs required for the biological activities of interest, as well as evidence for the activity of interest itself from at least two distinct in vivo preclinical models. This information is critical for both NPCT planning and for post-trial interpretation of results (Figures 1 and 2).

5.1 | NPCT planning

The foundational information should rigorously support a clear and specific definition of the question to be addressed by the NPCT, framed in quantitative terms, and matched to the study population, the evaluation period, and the primary and secondary outcomes.^{206,207} When strong evidence supports the requirement for specific bioactive components and specific in vivo targets, this information can be used to develop approaches for de-risking a project, that is, assessing whether the accumulating data are consistent with the hypotheses and support continuation. This information enriches the ability to design a NPCT that will allow evidence-based interpretation and planning of next steps regardless of the completed outcome, as shown schematically in Figure 1, by providing the ability to:

- define minimal effective concentrations in plasma or at the targets;
- develop methods to assess in vivo engagement of putative targets, especially methods which may be translationally relevant, such as non-invasive methods;
- use those methods to assess the relationship between target engagement and outcomes;
- elucidate causal cellular or molecular mechanisms mediating the generation of the biological or behavioral responses of interest (ie, PD biomarkers).

This mechanistic information will be critical for clinical dose-ranging and dose optimization studies, as well as for detecting potential “drop-ins”: given the broad availability of many NPs, placebo group participants sometimes present with anomalously high levels of apparent off-protocol exposure to the intervention. Off-protocol exposures, whether accidental or intentional, may hamper the ability to interpret the results. Knowing which compounds are indispensable and inducing for the bioactivities of interest also allows investigators to ensure that sufficient

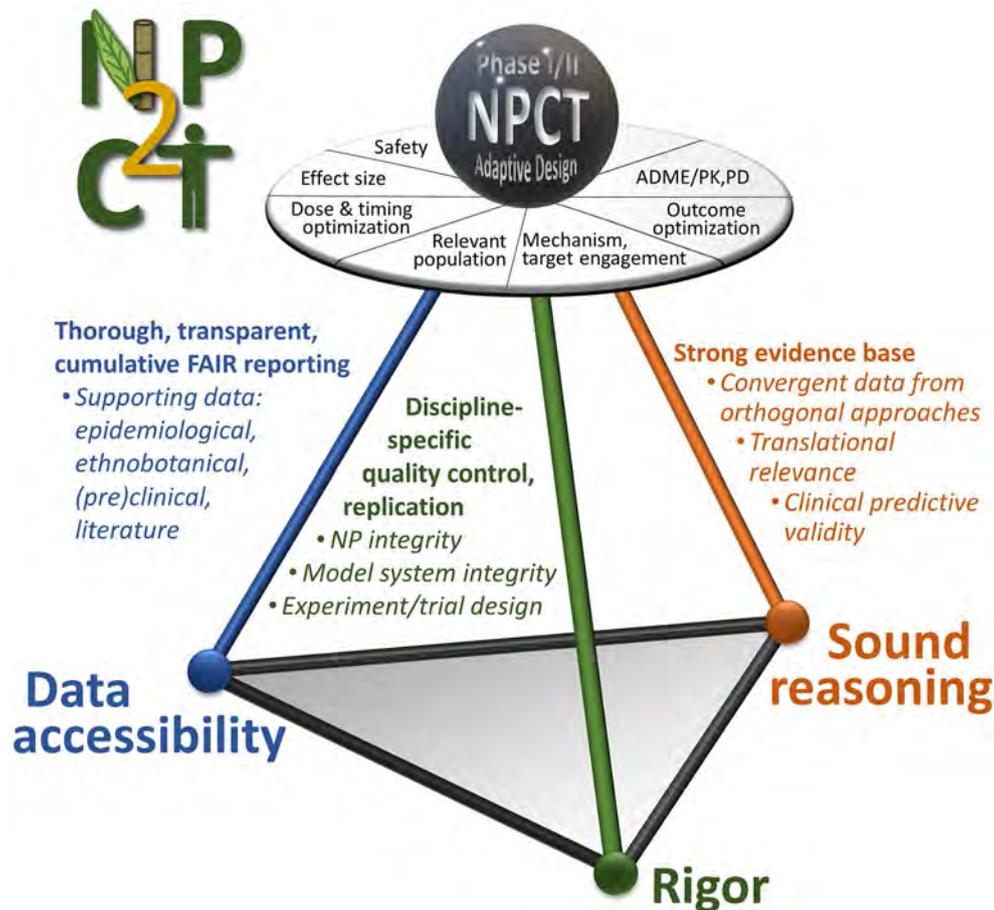


FIGURE 2 Good practices for actionable results in future translation of Natural Products research to Clinical Trials (NP2CT). Translation is most informative and cost-effective when CT Phase (I vs II) and design (eg, RCT, adaptive) decisions are built on a solid and rigorous evidence base (pyramid and tabletop). The choice to proceed to an NPCT and the determination of the most appropriate trial type (examples on sphere) depends on the extent and strength of supporting data available (table legs and segments). If multiple supporting data elements are missing or weak, an earlier phase CT may be more informative and cost-effective than a Phase III CT. If a substantial number of CT design choices or parameters lack supporting evidence, a more useful next step may be research addressing those gaps *via* either preclinical research or a mechanistic or early phase NPCT

time is allowed for washout between inventions in cross-over designs.²⁶ Equally important, knowledge, for example, of minimal effective concentrations and targets, or of metabolites indispensable for activity or genotypes required to generate those metabolites, can be used to generate minimal go/no-go criteria, which help identify CTs that are less likely to be informative, and ensure that only treatments with strong supporting evidence move forward to larger, confirmatory CT (Figure 2). Elucidation of such causal, molecular mechanisms of action may also be useful to predict which participants are likely to respond well or poorly to a specific intervention. Similarly, understanding the molecular mechanisms between an initial target and the final, assessed outcome should enable not only enhanced post-trial assessment of the validity of the pre-trial mechanistic hypotheses but also contribute to the ability to better predict individual responses to the intervention.

5.2 | Safety and dose selection

Safety issues that arise in preclinical testing must be carefully considered prior to CT, and safety outcomes must be included in all NPCTs.²⁰⁸ Safety in the context of a NPCT cannot be inferred from traditional nor contemporary use if the product, dose, route, schedule, and context of use differ (Table 3). Dosing in larger NPCT should be based on dose-response data from Phase I/III studies with safety data. Dose-ranging in Phase I NPCT may use allometric or mechanistic models to build from *in vivo* preclinical studies (Table 4), as described above, or, where appropriate, may be based on current clinical practice or traditional use.

As described above and summarized schematically in Figure 2, a variety of rigorous foundational information provides critical input for translation to later-phase efficacy CT, with much of this information requiring input from other types of NIH-defined CT. This foundational information includes the following:

- safety of the product to be used in the NPCT, in the population to be studied;
- product specifications (for interventions and any controls), including chemistry, stability, consumer acceptability (including sensory qualities), and packaging;
- bioavailability, PK, PD of the product to be used in the population to be studied;
- dose-response properties.

5.3 | Post-trial interpretation of results

In the absence of the foundational information described above, and without meeting the resulting go/no-go criteria, it will be more challenging to interpret a failure to reject the null hypothesis, as lack of effect could be due to many factors (Figure 1). These include, but are not limited to an incorrect hypothesis regarding the relationship between target engagement and outcome, or less interestingly, failure to achieve sufficient levels of bioactive compounds or target engagement. As the following variables could have an important influence on the results, obtaining information on these for each participant can also be critical to the interpretation of NPCT outcomes:

- Background diet—habitual and during the NPCT (the latter is potentially controllable).
- Age, sex, race/ethnicity, body size and composition, health status (especially with respect to outcomes of interest), co-medications, smoking, geography.
- Genetic and epigenetic data, including characterization of microbiota structure and gene expression, especially as they relate to ADME of the NP.^{134,168,195,209} When known, relevant genetic variants exist, for example, in metabolism, specimens for genetic analyses should at least be collected and saved.
- Sensory characteristics (texture, flavor, odor; pill, capsule, etc., number and characteristics) of products to be used in a NPCT must be acceptable to trial participants, who should not be able to distinguish amongst intervention(s) and controls. Information on participants' pre-trial beliefs about product efficacy, and post-trial conclusions with respect to the intervention they received should be collected as this may critically inform interpretation (Tables 3 and 5).

5.4 | Addressing expectancy effects

Human responses in NPCTs are also influenced by expectations, often resulting in either placebo or nocebo effects. These effects may vary substantially for different populations and interventions (eg, color, size, number of capsules, estimated cost).²¹⁰ Through conscious and subconscious processes,

people (researchers as well as participants) can be biased when assessing and reporting symptoms, functional abilities, and other patient-oriented outcomes. Whenever feasible, blinding/masking and other approaches should be employed to minimize such biases, including in preclinical studies (Tables 1 and 5).

Conscious and subconscious mind-body processes can also lead to real positive health outcomes, including in apparently objective, quantitative outcomes such as blood pressure.²¹¹ Because blinded “placebo” comparison groups are likely to achieve real health benefits through these processes, the magnitude of benefit estimated by comparing conventionally blinded placebo and treatment (intervention) groups is likely to underestimate the amount of benefit that would be observed when comparing non-blinded treatment to no treatment.²¹² Therefore, when feasible, no treatment and open-label treatment comparison groups should be included in RCT design, so as to arrive at best estimates of real-world effect sizes, as well as effect size from conventional blinded comparison.

5.5 | Sample and effect size

Sample size calculations must reflect simultaneous consideration of trial phase and design, estimated and targeted magnitude of treatment effects, and statistical power.²⁰⁷ The targeted effect size should be biologically plausible, and based on the degree of benefit that patients, consumers, or practitioners would consider sufficiently beneficial to be worthwhile, given other considerations such as cost, inconvenience, side effects, and potential for serious adverse effects²¹³ (Table 5). Effect sizes should not be based on estimates from the literature as a whole, which almost always leads to biased estimates,²¹⁴ inasmuch as the published literature contains a biased selection of all empirical research that is performed. Meta-analytic evaluations of the literature have consistently found that studies yielding null or mixed results are substantially less likely to appear in the published scientific record^{215,216} than those studies with strong statistical support for the tested hypothesis. This occurs even where the amount of collected data renders these “negative” studies highly informative. As a result, statistically significant published results often provide inflated effect size estimates.²¹⁷ Increased use of registered reports, for example, as provided by preclinicalresearch.eu, may begin to address this issue and decrease the bias in the scientific literature.

5.6 | Analysis of the literature

Beyond publication bias, two additional factors that challenge the translational relevance of the published literature have been widely agreed upon: analytic flexibility during the data analysis²¹⁸ and low statistical power.^{40,218,219}

TABLE 5 Highlights of major considerations for clinical trial design

1. *Statistical power.* Base sample sizes on the smallest effect size of interest, in the context of existing options and consumer burden. Use best current meta-analytic approaches to assess bias in the literature, and to correct for bias in reported effect sizes.²¹⁴
2. *Transparency and statistical analysis plans.* Clearly describe and document statistical analysis plans before beginning the trial.
3. *Comparison groups and expectancy/placebo.* Base controls and interventions on the supporting data and the resulting hypotheses. The inclusion of no-treatment, and open-label treatment groups, as well as conventional blinded comparison groups, in RCTs provides best estimates of real-world effect sizes. Development of inert but effectively masked placebos for NPs with strong color or odor (eg, with similar or masking odor or color) may be challenging. Assessing participant perceptions of product efficacy pre-trial, and of intervention allocation post-trial may provide critical information for interpretation of trial results.
4. *Optimal clinical trial paradigm.* Adaptive design (AD) NPCTs may be a more efficient approach to optimizing, and testing the efficacy of an intervention, if the evidence base for some CT parameters is lacking. AD CTs require statistical rigor to avoid bias in the interpretation of the results. ADs may be a good choice for Phase II trials.

All of these factors need to be taken into account when building on published research. Nevertheless, it should be kept in mind that a substantial portion of the scientific literature does consist of reliable and replicable findings. An important challenge for researchers navigating this intrinsically biased body of literature is to determine which research is reliable enough to build on, and which research is more likely to not be replicable.²²⁰ Because it is practically impossible to determine whether a finding will replicate based on a single study,²²¹ meta-analytic techniques have been developed to aggregate evidence across studies. However, analytic flexibility and selection for significant effects in small studies lead to upwardly biased effect size estimates. Novel bias detection techniques can identify different types of biases: P-curve analysis²²² can identify lines of research that consist of false positives, and under some circumstances PET-PEESE meta-regression^{222,223} can calculate bias-corrected effect size estimates.

5.7 | Clinical trial design

In the design of a later Phase NPCT, it is important to recognize that estimated treatment effects from pilot studies are necessarily imprecise. Pilot studies may under- or over-estimate the true effect size (but, as noted above, publication bias generally selects for overestimates). Effect size estimates and sample size calculations should, therefore, be conservative and based on prospective meta-analyses

(<https://methods.cochrane.org/pma/what-pma-0>) or other rigorous meta-analyses, that is, those which include all eligible studies designed and performed, or those which pass rigorous bias detection tests, such as the p-curve analysis described above, with PET-PEESE meta-regression used in estimating effect sizes. Results of pre-registered, rigorously performed, and analyzed pilot studies may also be used to estimate effect sizes when convergent, corroborating, or similar results are obtained in independent research from separate groups, using approaches based on different assumptions (triangulation or orthogonal approaches), such that any artefacts or sources of bias are likely to differ for the different approaches.^{224,225} Adaptive designs (AD) (see below) and sequential analyses (<https://methods.cochrane.org/pma/what-pma-0>) can be used to design informative studies when uncertainty about the effect size estimate remains. An argument can be made that later Phase NPCT intended to test whether an intervention is beneficial should be undertaken only where the estimated effect size from rigorous methods such as those described above is at least as large as the estimated smallest important difference or smallest worthwhile effect²¹³ for the intervention and outcome to be tested. This would not apply if the hypothesis to be tested is that the intervention has no effect, for example, on an adverse outcome.

Where the goal is to determine the maximum potential of an intervention, Phase II studies should enroll those subpopulations predicted to have strong responses, a practice called study “enrichment.” While enrichment, and the inclusion of multiple endpoints are appropriate for Phase II NPCT, it is important to recognize that these designs tend to bias the results of Phase II studies toward the conclusion that moving forward is worthwhile, although positive findings could be due to chance.

Because NPCT are costly, time-consuming,²²⁶ and dependent on voluntary human participation,¹¹ approaches that can support making trial design and prioritization decisions which optimize acquisition of replicable, actionable results (Tables 5 and 6) are in high demand. Multilevel modeling can be used in conjunction with existing data to identify parameters most strongly associated with the outcomes of interest.²²⁷ Simulation modeling can also be used to identify trials that are unlikely to reject the null hypothesis.²²⁸

AD can provide an efficient yet rigorous approach to honing the parameters of a NPCT while obtaining preliminary data on clinical effect of the intervention(s). An AD uses accumulating information to implement changes to the design during the course of the trial.^{22,229} Changes in the trial must be based on pre-specified decision rules. Changes made on an ad hoc basis may not preserve the type I error rate (false positives), and may lead to bias in parameter estimates or other problems with statistical properties. The many variations of AD include early-stage AD, which may focus on dosing, exposure, heterogeneous responses,

or biomarker responses. As it is assumed that AD are exploratory studies, there is less concern about the control of type I error. AD methods are more accepted in the learning (exploratory) stages of CT. Some key variations of AD include the following:

- *Adaptive enrichment designs*, which target therapies to participant subgroups most likely to benefit. In such designs, a trial initially enrolls a broad population. The initial study period (discovery phase) reveals participant groups most likely to benefit from treatment. In the second phase of the trial, only those subgroups most likely to benefit from treatment are randomized, providing increased power to detect an effect.
- *Sample size re-estimation designs* refer to AD that allow for adjustment of sample size based on an interim review of the data. Sample size re-estimation based on revised estimates of treatment effects has been controversial for the reasons cited above. However, internal pilot designs that re-estimate only “nuisance” parameters, such as the variance of control group event rate, are more generally accepted.
- *Platform trials* are useful for studying multiple targeted therapies in the context of a single outcome, with therapies allowed to enter or leave the platform on the basis of pre-specified stopping boundaries.²² This approach could be used to choose between several different products in a NPCT, for example, different preparations of NP material from the same source, such as whole dried plant, different extracts, and/or material enriched for one or more putative bioactives.²³⁰ Such studies could be done under a master protocol,²³¹ defined as a protocol encompassing multiple substudies, which may have different objectives. A master protocol supports coordinated efforts to evaluate one or more investigational interventions in one or more subpopulations within the overall trial structure.

Adaptive seamless designs combine objectives traditionally addressed in separate trials into a single CT. First-stage results inform the second stage, but final analyses use data from both stages. For example, the first stage may establish preferred dosage, whereas the second stage can assess the efficacy of that dose. To avoid bias in the interpretation of results, such designs require clear, comprehensive, and pre-specified criteria and raise tricky statistical issues, as data from both stages must be combined in a way that preserves key statistical operating characteristics.

In general, most AD require slightly larger sample sizes than a corresponding, conventional Phase II CT, but provide increased opportunity to detect an effect, and can also lead to timelier identification of ineffective (or clearly more efficacious) interventions, improving efficiency by allowing early discontinuation of trials for less effective or ineffective products.

TABLE 6 Highlights of major issues to consider in clinical trial decision-making

1. *Trial focus*. Is the proposed trial well-justified? Are the evidence for and magnitude of the effect, the size of the relevant population, and the need for the knowledge to be generated commensurate with the anticipated cost of a CT?
2. *Likelihood of actionable information*. What CT paradigm or other next step is most likely to provide actionable information?
 - a. *Next step*: Is the most informative next step likely to be a pilot trial, a larger CT, clinical research other than an efficacy study, or research other than a CT?
 - b. *NPCT protocol*: Have trial/protocol design decisions been sufficiently de-risked? Is there sufficient understanding of the path(s) from initial target(s) engagement to outcome, and availability of non-invasive assessments of these, to support evidence-based decisions regarding the next test of activity whether or not the null hypothesis is rejected upon trial completion?
 - c. *Bioavailability*: Are assays of bioavailability at the immediate in vivo target available and translationally relevant?
 - d. *Open ends*: Can any remaining questions be sufficiently addressed via use of a rigorous, adaptive trial design?

5.8 | Value of information

Whether and how to pursue a NPCT depends on the value of the information that the CT can be expected to generate. Given the high costs of conducting a CT, it is of critical importance to prioritize the questions to be addressed.¹¹ Value of Information (VOI) analysis provides a framework for generating quantitative estimates of the value of further research in a particular area. VOI estimates can serve as a guide for prioritizing research resources.²³²

The basic framework for estimating the VOI is the product of several factors: (a) the value of the information to be gained for individuals who would be affected by the findings, including both changes in health status and changes in costs, (b) the number of affected individuals in the population, (c) the extent to which the information generated would be adopted or implemented, and (d) the degree to which the value of the knowledge gained will depreciate over time as additional knowledge becomes available. This estimated quantity is typically summed over the relevant time horizon, adjusted by an appropriate discount factor reflecting the preference for current over future improvements in health.

Generating estimated values for the elements of the VOI calculation can itself be a costly and complex process. Depending on the NPs under consideration and the effects to be tested, different levels of VOI analysis may be undertaken. An initial level of conceptual VOI can assess whether any of the factors in the product are close enough to zero that the overall VOI would be very small,

indicating that substantial research investments would not be justified. Depending on other characteristics of the product under investigation, the health outcomes involved, and the specific data collection challenges to be addressed, progressively more detailed approaches to VOI exist, termed minimal modeling, full modeling, and maximal modeling, that provide more thorough estimates of the value of the information to be gained from pursuing additional research.^{232,233}

The use of VOI analysis to support decisions regarding which NPCT to pursue (Table 6) depends significantly on the available research foundation. Assessing the VOI to be gained to a representative affected individual requires definition of the outcome(s), and understanding of the effect sizes to be achieved. Defining and estimating the size of the relevant affected population requires understanding of the dietary, environmental, and genetic context of the effects. Use of VOI in priority setting is still relatively new and will benefit from further exploration and input from researchers and decision-makers. In particular, VOI should be assessed to understand the extent to which it changes decision-making relative to conventional approaches, and the health impact and monetary value that result from such changes.

6 | CONCLUSIONS

6.1 | The multi-disciplinary approach

Rigorous, evidence-based decision-making with respect to the design and prioritization NPCTs ultimately requires bringing together myriad research sub-disciplines. These disciplines span from ethnobotany through NP chemistry, food science, a range of modeling approaches, biomedical specialties including CT design and statistics, to value of information research. Notably, researchers who specialize in one sub-discipline often do not have ready access to collaborators from other fields, or are unaware of relevant state-of-the-art approaches from these areas. However, the likelihood that an NPCT will be optimally informative depends not only on the rigor of the information contributed by each sub-discipline but also the rigor with which the different threads are brought together, achieving collaborative synergy. Deploying and synergizing the good practices discussed by attendees at the September 2018 workshop, across the full range of this disciplinary continuum and from the earliest phases of developing data that may inform a future NPCT, should help focus translational NP research on objectives that are most likely to contribute to the design of, and increase the knowledge gained from future NPCTs (Figure 2).

6.2 | Controversies

Two major areas of controversy at the workshop related to translation from traditional health practices. The first of these areas of disagreement was the appropriateness of the RCT approach for the study of traditional practices, with several attendees arguing that RCT methodology is not well suited to studying them. Most attendees, however, supported the need for NPCTs to assess the safety and efficacy of tradition-based approaches in a 21st-century context, with the many associated differences, including in the interventions, and in individuals' histories, environments, and behaviors. A second area of disagreement was the extent to which it is necessary to elucidate causal, molecular mechanisms of action, including determination of bioactive substituents and their in-animal targets, to support the design of an optimally informative NPCT. A case was made that, in some contexts, adherence to current good manufacturing practices, including rigorous adherence to standardized growing and processing procedures, and in conjunction with an appropriate assay of systemic exposure, and, in some cases, bioactivity, is sufficient to ensure NP replicability. However, many attendees felt that product specifications, PK assessments, eligibility criteria, and outcome assessments should, to the extent feasible, be based on strong mechanism of action evidence, preferably derived from conceptually independent (orthogonal or triangulated) scientific approaches. Attendees considered such evidence to be critical both for optimal trial design and for the ability to interpret results after trial completion. Information on the mechanism of action, including compound(s) indispensable for inducing effects, as well as biological target(s) and response(s) indispensable for outcomes, may critically inform the design and/or interpretation of an NPCT, from product quality control through understanding the bases for individual differences in response, and may support the ability to develop biomarkers of proximate effects, or biomarkers useful to distinguish likely responders and non-responders.

6.3 | NPCT good practices

In preclinical research aimed at informing the understanding of purported, causal, molecular mechanisms of action, good practices include the use of biologically relevant concentrations of biologically relevant compounds and/or their in vivo metabolites, in one or more relevant models. Participants also emphasized the importance of including appropriate controls for common artefacts encountered in in vitro assays for the activities of phytochemicals. Further approaches that were highlighted as good practices for strengthening the foundational data for NPCTs included: (a) high-throughput, non-targeted approaches that can generate strong, specific

hypotheses regarding molecular mechanisms of action of chemically complex NP, (b) pre-registration of experiments at the hypothesis-testing stage, rigorous adherence to experimental protocols, randomization, and masking, which collectively minimize the risk that unconscious biases influence the results, (c) thoughtful consideration, when selecting model systems, of translational relevance and clinical validity for a specific research question, and (d) comprehensive and transparent reporting, combined with meaningful and FAIR data sharing. Participants discussed these issues for models ranging from *in silico* through the burgeoning variety of cell-based models, from organoids through MPS, through *in vivo* models from invertebrates to transgenic, outbred, wildling, and humanized mouse strains.

Good practices for NPCT decision-making based on the available foundational data were extensively discussed and gave rise to the third major area of controversy at the workshop. Some attendees felt that *n-of-1* trials^{23,4} based on traditional use and/or current practitioner experience provide sufficient evidence for safety and efficacy, and are better suited to tests of these approaches than RCTs. Others felt that RCTs are critical, but that NPCTs should be undertaken only when there is a strong likelihood that the outcome will influence practice. The foundation for such an assessment would come from public health needs and a rigorous, evidence-based effect size estimate, that predicts an effect magnitude considered worthwhile by practitioners and consumers, collectively supported by multiple, distinct lines of evidence.^{22,5} The evidence base for such assessments is strongest when it includes experiments designed to falsify the hypotheses tested, such as chemical, genetic, or other knockdown studies.^{11,35}

6.4 | Outlook

While many of the practices discussed at the workshop are already employed by numerous NP researchers, few, if any reported NPCTs have combined the full breadth of state-of-the-art approaches when investigating the effect of a NP on a specific health outcome. Integrating these practices in the development, implementation, and interpretation of NPCTs will require research teams to collaborate synergistically across disciplines. The anticipated benefit of combining the rigorous and trans-disciplinary approaches referenced herein increases the likelihood that, in the many instances where the outcome of a NPCT does not signal a change in practice, the trial will nevertheless have provided actionable new information, for example, about NP metabolism or the target biology.

ACKNOWLEDGMENTS

The authors thank Paul M. Coates, Joseph M. Betz, and F. Ellis McKenzie for inspiration and support, Wen Chen for

help with the initial concept, Wendy P. O'Meara for insightful editing, Sharon Ross and Harold Seifried of the NIH National Cancer Institute for invaluable input into the workshop, Miriam Sander for workshop transcription and notes, Cara Lynch, James Hubley, Katy Fuller, and Joyce Merkel for outstanding technical support. This project was supported by staff and funding from the NIH's Office of Dietary Supplements, National Cancer Institute, National Center for Complementary and Integrative Health, National Institute on Aging, National Institute on Alcohol Abuse and Alcoholism, and National Institute of Environmental Health Sciences. Support was also provided by staff from the Food and Drug Administration and the US Department of Agriculture.

CONFLICT OF INTEREST

MGF is a member of the Sensient Technologies Board of Directors and holds equity in Sensient Technologies. KDRS holds equity in the following companies: Asklepiion Pharmaceuticals, Baltimore MD, Ausio LLC, Cincinnati, Ohio, and Aliveris s.r.l, Italy; KDRS is a consultant to Retrophin, San Diego, and Sanitarium Health and Wellness Company, Berkleyvale, NSW, Australia and is an inventor on intellectual property related to equal. The remaining authors declare no conflicts of interest.

AUTHORS CONTRIBUTIONS

B.C. Sorkin performed literature searches, developed the plan for the workshop and review, obtained funding, organized co-author efforts, and wrote the paper. G.F. Pauli contributed to the development of the workshop and co-author team, contributed subject matter expertise to the manuscript as well as the development and conduct of the workshop, and wrote the paper. A.J. Kuszak, performed literature searches, contributed to the development of the workshop, the coordination of co-author efforts, and wrote the paper. G. Bloss, F.A. Hoffman, and N. Fukagawa contributed to the development of the workshop and co-author team, the conduct of the workshop, and the writing of the paper. D.C. Hopp, D. Xi, and G. Zappalá contributed subject matter expertise to the development and conduct of the workshop and to the writing of the paper. P. Brown contributed subject matter expertise and to the writing of the paper. B. Barrett, F.D. Bushman, S.J. Casper, F.H. Chilton, C.S. Coffey, M. Ferruzzi, M. Jafari, M. Kiely, D. Lakens, J.B. MacMillan, D.O. Meltzer, M. Pahor, J. Paul, K. Pritchett-Corning, S. K. Quinney, B. Rehermann, K.D.R. Setchell, N.S. Sipes, J.M. Stephens, D.L. Taylor, H. Tiriach, and M. Walters contributed subject matter expertise to the workshop and to the writing of the paper; all authors reviewed the final manuscript and agreed to submit it.

ENDNOTES

^a Per the 2014 NIH guidance, a clinical trial is defined here as a research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes (<https://grants.nih.gov/policy/clinical-trials/definition.htm>).

^b Traditional medical systems are defined by the WHO, in the WHO Global Report on Traditional and Complementary Medicine 2019 (<https://apps.who.int/iris/bitstream/handle/10665/312342/9789241515436-eng.pdf?ua=1>) as “the sum total of the knowledge, skill and practices based on the theories, beliefs and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health as well as in the prevention, diagnosis, improvement or treatment of physical and mental illness.”

REFERENCES

1. Baden MY, Liu G, Satija A, et al. Changes in plant-based diet quality and total and cause-specific mortality. *Circulation*. 2019;140:979-991.
2. Britton ER, Kellogg JJ, Kvalheim OM, Cech NB. Biochemometrics to identify synergists and additives from botanical medicines: a case study with *Hydrastis canadensis* (goldenseal). *J Nat Prod*. 2018;81:484-493.
3. Wang J, Hodes GE, Zhang H, et al. Epigenetic modulation of inflammation and synaptic plasticity promotes resilience against stress in mice. *Nat Commun*. 2018;9:477.
4. Boudreau A, Poulev A, Ribnicky DM, et al. Distinct fractions of an *Artemisia scoparia* extract contain compounds with novel adipogenic bioactivity. *Front Nutr*. 2019;6:18. <https://doi.org/10.3389/fnut.2019.00018>.
5. Munyangi J, Cornet-Vernet L, Idumbo M, et al. *Artemisia annua* and *Artemisia afra* tea infusions vs artesunate-amodiaquine (ASAQ) in treating *Plasmodium falciparum* malaria in a large scale, double blind, randomized clinical trial. *Phytomedicine*. 2019;57:49-56.
6. Hopp C. Past and future research at national center for complementary and integrative health with respect to botanicals. *HerbalGram*. 2015;107:44-51.
7. Manson JE, Cook NR, Lee I-M, et al. Vitamin D supplements and prevention of cancer and cardiovascular disease. *N Engl J Med*. 2019;380:33-44.
8. Manson JE, Cook NR, Lee I-M, et al. Marine n-3 fatty acids and prevention of cardiovascular disease and cancer. *N Engl J Med*. 2019;380:23-32.
9. Berman J. Clinical development of dietary supplements: the perils of starting at phase III. *Fitoterapia*. 2011;82:80-84.
10. Sertkaya A, Wong H-H, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin Trials*. 2016;13:117-126.
11. Zarin DA, Goodman SN, Kimmelman J. Harms from uninformative clinical trials. *JAMA*. 2019;322(9):813-814. <https://doi.org/10.1001/jama.2019.9892>
12. Applequist WL, Miller JS. Selection and authentication of botanical materials for the development of analytical methods. *Anal Bioanal Chem*. 2013;405:4419-4428.
13. Sorkin BC, Kuszak AJ, Williamson JS, Hopp DC, Betz JM. The challenge of reproducibility and accuracy in nutrition research: resources and pitfalls. *Adv Nutr*. 2016;7:383-389.
14. Simmler C, Graham JG, Chen S-N, Pauli GF. Integrated analytical assets aid botanical authenticity and adulteration management. *Fitoterapia*. 2018;129:401-414.
15. Fenclova M, Novakova A, Viktorova J, et al. Poor chemical and microbiological quality of the commercial milk thistle-based dietary supplements may account for their reported unsatisfactory and non-reproducible clinical outcomes. *Sci Rep*. 2019; 9:11118.
16. Rider CV, Walker NJ, Waidyanatha S. Getting to the root of the matter: challenges and recommendations for assessing the safety of botanical dietary supplements. *Clin Pharmacol Ther*. 2018;104:429-431.
17. Kellogg JJ, Paine MF, McCune JS, Oberlies NH, Cech NB. Selection and characterization of botanical natural products for research studies: a NaPDI center recommended approach. *Nat Prod Rep*. 2019;36:1196-1221.
18. National Center for Complementary and Integrative Health. *NCCIH Policy: Natural Product Integrity*. NCCIH/NIH. <https://nccih.nih.gov/research/policies/naturalproduct.htm>. Accessed August 14, 2019.
19. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Recommendations for reporting randomized controlled trials of herbal interventions: explanation and elaboration. *J Clin Epidemiol*. 2006;59:1134-1149.
20. Kilkeny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *J Pharmacol Pharmacother*. 2010;1:94-99.
21. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
22. Center for Drug Evaluation and Research. *Botanical Drug Development: Guidance for Industry*. U.S. Food and Drug Administration; 2016. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/botanical-drug-development-guidance-industry> Accessed August 14, 2019.
23. Kroeger CM, Garza C, Lynch CJ, et al. Scientific rigor and credibility in the nutrition research landscape. *Am J Clin Nutr*. 2018;107:484-494.
24. Smith AJ, Clutton RE, Lilley E, Hansen KEA, Brattelid T. Improving animal research: PREPARE before you ARRIVE. *BMJ*. 2018;360:k760.
25. Mebane CA, Sumpter JP, Fairbrother A, et al. Scientific integrity issues in environmental toxicology and chemistry: improving research reproducibility, credibility, and transparency. *Integr Environ Assess Manag*. 2019;15:320-344.
26. U.S. Food and Drug Administration. *Considerations in Demonstrating Interchangeability With a Reference Product (PDF)*. U.S. FDA. 2017. <https://www.fda.gov/media/102592/download>. Accessed August 14, 2019.
27. National Institutes of Health. *Enhancing Reproducibility through Rigor and Transparency*; NIH. <https://grants.nih.gov/policy/reproducibility/index.htm>.
28. Brenna JT, Plourde M, Stark KD, Jones PJ, Lin Y-H. Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids. *Am J Clin Nutr*. 2018;108:211-227.
29. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med*. 2016;8:341ps12.
30. Plessner HE. Reproducibility vs replicability: a brief history of a confused terminology. *Front Neuroinform*. 2017;11:76.

31. National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press; 2019. <https://doi.org/10.17226/25303>.
32. Morgan P, Van Der Graaf PH, Arrowsmith J, et al. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov Today*. 2012;17:419-424.
33. Karran E, De Strooper B. The amyloid cascade hypothesis: are we poised for success or failure? *J Neurochem*. 2016;139:237-252.
34. Karran E, Hardy J. A critique of the drug discovery and phase 3 clinical programs targeting the amyloid hypothesis for Alzheimer disease. *Ann Neurol*. 2014;76:185-205.
35. Wood AC, Wren JD, Allison DB. The need for greater rigor in childhood nutrition and obesity research. *JAMA Pediatr*. 2019;173:311-312.
36. Harrington D, D'Agostino RB, Gatsonis C, et al. New guidelines for statistical reporting in the journal. *N Engl J Med*. 2019;381:285-286.
37. Ferreira ML, Herbert RD, Ferreira PH, et al. The smallest worthwhile effect of nonsteroidal anti-inflammatory drugs and physiotherapy for chronic low back pain: a benefit-harm trade-off study. *J Clin Epidemiol*. 2013;66:1397-1404.
38. Freedman LP, Gibson MC, Wisman R, et al. The culture of cell culture practices and authentication—results from a 2015 survey. *Biotechniques*. 2015;59:189-190, 192.
39. Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: progress and priorities. *F1000Res*. 2017;6:604.
40. Button KS, Ioannidis JPA, Mokrysz C, et al. Empirical evidence for low reproducibility indicates low pre-study odds. *Nat Rev Neurosci*. 2013;14:877.
41. Sitta Sittampalam G, Grossman A, Brimacombe K, et al. *Assay Guidance Manual*; 2004. <https://www.ncbi.nlm.nih.gov/books/NBK53196/>. Accessed August 14, 2019.
42. Bradbury A, Plückthun A. Reproducibility: standardize antibodies used in research. *Nature*. 2015;518:27-29.
43. Schinkovitz A, Pro SM, Main M, et al. Dynamic nature of the ligustilide complex. *J Nat Prod*. 2008;71:1604-1611.
44. Schneider C, Gordon ON, Edwards RL, Luis PB. Degradation of curcumin: from mechanism to biological implications. *J Agric Food Chem*. 2015;63:7606-7614.
45. Duan D, Doak AK, Nedyalkova L, Shoichet BK. Colloidal aggregation and the in vitro activity of traditional Chinese medicines. *ACS Chem Biol*. 2015;10:978-988.
46. Ingólfsson HI, Thakur P, Herold KF, et al. Phytochemicals perturb membranes and promiscuously alter protein function. *ACS Chem Biol*. 2014;9:1788-1798.
47. Baell JB. Feeling nature's PAINS: natural products, natural product drugs, and pan assay interference compounds (PAINS). *J Nat Prod*. 2016;79:616-628.
48. Nelson KM, Dahlin JL, Bisson J, Graham J, Pauli GF, Walters MA. The essential medicinal chemistry of curcumin. *J Med Chem*. 2017;60:1620-1637.
49. Friesen JB, Liu Y, Chen S-N, McAlpine JB, Pauli GF. Selective depletion and enrichment of constituents in "curcumin" and other *Curcuma longa* preparations. *J Nat Prod*. 2019;82:621-630.
50. Dahlin JL, Walters MA. The essential roles of chemistry in high-throughput screening triage. *Future Med Chem*. 2014;6:1265-1290.
51. Dahlin JL, Baell J, Walters MA. Assay interference by chemical reactivity. In *Assay Guidance Manual* [Internet] Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2015. <https://www.ncbi.nlm.nih.gov/books/NBK326709/>. Accessed August 14, 2019.
52. Ramaiahgari SC, Waidyanatha S, Dixon D, DeVito MJ, Paules RS, Ferguson SS. Three-dimensional (3D) HepaRG spheroid model with physiologically relevant xenobiotic metabolism competence and hepatocyte functionality for liver toxicity screening. *Toxicol Sci*. 2017;160:189-190.
53. Tuveson D, Clevers H. Cancer modeling meets human organoid technology. *Science*. 2019;364:952-955.
54. Esch EW, Bahinski A, Huh D. Organs-on-chips at the frontiers of drug discovery. *Nat Rev Drug Discov*. 2015;14:248-260.
55. Dutta D, Heo I, Clevers H. Disease modeling in stem cell-derived 3D organoid systems. *Trends Mol Med*. 2017;23:393-410.
56. Skardal A, Shupe T, Atala A. Organoid-on-a-chip and body-on-a-chip systems for drug screening and disease modeling. *Drug Discov Today*. 2016;21:1399-1411.
57. Ewart L, Fabre K, Chakilam A, et al. Navigating tissue chips from development to dissemination: a pharmaceutical industry perspective. *Exp Biol Med*. 2017;242:1579-1585.
58. Low LA, Tagle DA. Tissue chips—innovative tools for drug development and disease modeling. *Lab Chip*. 2017;17:3026-3036.
59. Vernetti L, Gough A, Baetz N, et al. Functional coupling of human microphysiology systems: intestine, liver, kidney proximal tubule, blood-brain barrier and skeletal muscle. *Sci Rep*. 2017;7:42296.
60. Vernetti LA, Vogt A, Gough A, Taylor DL. Evolution of experimental models of the liver to predict human drug hepatotoxicity and efficacy. *Clin Liver Dis*. 2017;21:197-214.
61. Hartung T. Toxicology for the twenty-first century. *Nature*. 2009;460:208-212.
62. Seok J, Warren HS, Cuenca AG, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A*. 2013;110:3507-3512.
63. Baker LA, Tiriach H, Clevers H, Tuveson DA. Modeling pancreatic cancer with organoids. *Trends Cancer Res*. 2016;2:176-190.
64. Aberle MR, Burkhart RA, Tiriach H, et al. Patient-derived organoid models help define personalized management of gastrointestinal cancer. *Br J Surg*. 2018;105:e48-e60.
65. Li X, George SM, Vernetti L, Gough AH, Taylor DL. A glass-based, continuously zoned and vascularized human liver acinus microphysiological system (vLAMPs) designed for experimental modeling of diseases and ADME/TOX. *Lab Chip*. 2018;18:2614-2631.
66. Neal JT, Li X, Zhu J, et al. Organoid modeling of the tumor immune microenvironment. *Cell*. 2018;175:1972-1988.e16.
67. Minekus M, Marteau P, Havenaar R, Huis in't Veld, JHH. Multicompartmental dynamic computer-controlled model simulating the stomach and small intestine. *Altern La Anim*. 1995;23.2 <http://www.atla.org.uk/a-multicompartmental-dynamic-computer-controlled-model-simulating-the-stomach-and-small-intestine/>. Accessed November 18, 2019.
68. Peng Y-S, Liu B, Wang R-F, Zhao Q-T, Xu W, Yang X-W. Hepatic metabolism: a key component of herbal drugs research. *J Asian Nat Prod Res*. 2015;17:89-106.
69. Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1 000 000 profiles. *Cell*. 2017;171:1437-1452.e17.

70. Ochoa JL, Bray WM, Lokey RS, Lington RG. Phenotype-guided natural products discovery using cytological profiling. *J Nat Prod*. 2015;78:2242-2248.
71. Caicedo JC, Cooper S, Heigwer F, et al. Data-analysis strategies for image-based cell profiling. *Nat Methods*. 2017;14:849-863.
72. Potts MB, Kim HS, Fisher KW, et al. Using functional signature ontology (FUSION) to identify mechanisms of action for natural products. *Sci Signal*. 2013;6:ra90.
73. Vaden RM, Oswald NW, Potts MB, MacMillan JB, White MA. FUSION-guided hypothesis development leads to the identification of N⁶, N⁶-dimethyladenosine, a marine-derived AKT pathway inhibitor. *Mar Drugs* 2017;15:75.
74. Das B, Neilsen BK, Fisher KW, et al. A functional signature ontology (FUSION) screen detects an AMPK inhibitor with selective toxicity toward human colon tumor cells. *Sci Rep*. 2018;8:3770.
75. Litichevskiy L, Peckner R, Abelin JG, et al. A library of phosphoproteomic and chromatin signatures for characterizing cellular responses to drug perturbations. *Cell Syst*. 2018;6:424-443.e7.
76. Wang Z, Lachmann A, Keenan AB, Ma'ayan A. L1000FWD: fire-works visualization of drug-induced transcriptomic signatures. *Bioinformatics*. 2018;34:2150-2152.
77. Cech NB, Junio HA, Ackermann LW, Kavanaugh JS, Horswill AR. Quorum quenching and antimicrobial activity of goldenseal (*Hydrastis canadensis*) against methicillin-resistant *Staphylococcus aureus* (MRSA). *Planta Med*. 2012;78:1556-1561.
78. Elfawal MA, Towler MJ, Reich NG, Golenbock D, Weathers PJ, Rich SM. Dried whole plant *Artemisia annua* as an antimalarial therapy. *PLoS One*. 2012;7:e52746.
79. Caesar LK, Cech NB. Synergy and antagonism in natural product extracts: when 1 + 1 does not equal 2. *Nat Prod Rep*. 2019;36:869-888.
80. Lu T, Aron L, Zullo J, et al. REST and stress resistance in ageing and Alzheimer's disease. *Nature*. 2014;507:448-454.
81. Lithgow GJ, Driscoll M, Phillips P. A long journey to reproducible results. *Nature*. 2017;548:387-388.
82. Jafari M. *Drosophila melanogaster* as a model system for the evaluation of anti-aging compounds. *Fly*. 2010;4:253-257.
83. Jafari M. Healthspan pharmacology. *Rejuvenation Res*. 2015;18:573-580.
84. Schriener SE, Abrahamyan A, Avanesian A, et al. Decreased mitochondrial superoxide levels and enhanced protection against paraquat in *Drosophila melanogaster* supplemented with *Rhodiola rosea*. *Free Radic Res*. 2009;43:836-843.
85. Schriener SE, Lee K, Truong S, et al. Extension of *Drosophila* lifespan by *Rhodiola rosea* through a mechanism independent from dietary restriction. *PLoS One*. 2013;8:e63886.
86. Labachyan KE, Kiani D, Sevrioukov EA, Schriener SE, Jafari M. The impact of on the gut microbial community of *Drosophila melanogaster*. *Gut Pathog*. 2018;10:12.
87. Lopez T, Schriener SE, Okoro M, et al. Green tea polyphenols extend the lifespan of male *Drosophila melanogaster* while impairing reproductive fitness. *J Med Food*. 2014;17:1314-1321.
88. Judson RS, Magpantay FM, Chickarmane V, et al. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicol Sci*. 2015;148:137-154.
89. Sipes NS, Wambaugh JF, Pearce R, et al. An intuitive approach for predicting potential human health risk with the Tox21 10k library. *Environ Sci Technol*. 2017;51:10786-10796.
90. Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ Sci Technol*. 2015;49:8804-8814.
91. Bell SM, Chang X, Wambaugh JF, et al. In vitro to in vivo extrapolation for high throughput prioritization and decision making. *Toxicol In Vitro*. 2018;47:213-227.
92. Wambaugh JF, Hughes MF, Ring CL, et al. Evaluating in vitro-in vivo extrapolation of toxicokinetics. *Toxicol Sci*. 2018;163:152-169.
93. Kavlock RJ, Austin CP, Tice RR. Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal*. 2009;29:485-487; discussion 492-497.
94. Kavlock R, Chandler K, Houck K, et al. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. *Chem Res Toxicol*. 2012;25:1287-1302.
95. Thomas RS, Paules RS, Simeonov A, et al. The US federal Tox21 program: a strategic and operational plan for continued leadership. *Altex*. 2018;35:163-168.
96. Thomas RS, Bahadori T, Buckley TJ, et al. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol. Sci*. 2019;169:317-332.
97. Hubbard TD, Hsieh J-H, Rider CV, et al. Using Tox21 high-throughput screening assays for the evaluation of botanical and dietary supplements. *Appl In Vitro Toxicol*. 2019;5:10-25.
98. Ramaiahgari SC, Auerbach SS, Saddler TO, et al. The power of resolution: contextualized understanding of biological responses to liver injury chemicals using high-throughput transcriptomics and benchmark concentration modeling. *Toxicol Sci*. 2019;169:553-566.
99. Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK. Systems biology and combination therapy in the quest for clinical efficacy. *Nat Chem Biol*. 2006;2:458-466.
100. Wilkinson MD, Dumontier M, Jan Aalbersberg I, et al. Addendum: the FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2019;6:6.
101. Božin B, Kladar N, Grujić N, et al. Impact of origin and biological source on chemical composition, anticholinesterase and antioxidant properties of some St. John's wort species (*Hypericum* spp., Hypericaceae) from the Central Balkans. *Molecules*. 2013;18:11733-11750.
102. Ortiz D, Rocheford T, Ferruzzi MG. Influence of temperature and humidity on the stability of carotenoids in biofortified maize (*Zea mays* L.) genotypes during controlled postharvest storage. *J Agric Food Chem*. 2016;64:2727-2736.
103. Milanowski DJ, Winter REK, Elvin-Lewis MPF, Lewis WH. Geographic distribution of three alkaloid chemotypes of *Croton lechleri*. *J Nat Prod*. 2002;65:814-819.
104. Hildreth J, Hrabeta-Robinson E, Applequist W, Betz J, Miller J. Standard operating procedure for the collection and preparation of voucher plant specimens for use in the nutraceutical industry. *Anal Bioanal Chem*. 2007;389:13-17.
105. Klein MA, Nahin RL, Messina MJ, et al. Guidance from an NIH workshop on designing, implementing, and reporting clinical studies of soy interventions. *J Nutr*. 1192S;140:1192S-1204S.
106. Eisenman SW, Tucker AO, Struwe L. Voucher specimens are essential for documenting source material used in medicinal plant investigations. *J Med Act Plants*. 2012;1:30-43.

107. Gagnier JJ, Moher D, Boon H, Beyene J, Bombardier C. Randomized controlled trials of herbal interventions underreport important details of the intervention. *J Clin Epidemiol*. 2011;64:760-769.
108. McManis CR. The interface between international intellectual property and environmental protection: biodiversity and biotechnology. *Wash. ULQ*. 1998;76:255.
109. Battiste M, Henderson JY. Compulsory schooling and cognitive imperialism: a case for cognitive justice and reconciliation with Indigenous peoples. In: Trimmer K, Dixon R, Findlay YS, eds. *The Palgrave Handbook of Education Law for Schools*. Cham, UK: Palgrave Macmillan; 2018:567-583.
110. Cheng C-W, Wu T-X, Shang H-C, et al. CONSORT extension for Chinese herbal medicine formulas 2017: recommendations, explanation, and elaboration (traditional Chinese version). *Ann Intern Med*. 2017;167:W7-W20.
111. Gusev PA, Savarala S, Atkinson R, et al. Disintegration and rupture testing of prescription prenatal multivitamin/mineral dosage forms: a pilot study for the Dietary Supplement Ingredient Database (DSID). *Presented at the Current Developments in Nutrition*, Boston, MA.
112. Neilson AP, Ferruzzi MG. Influence of formulation and processing on absorption and metabolism of flavan-3-ols from tea and cocoa. *Annu Rev Food Sci Technol*. 2011;2:125-151.
113. Chen S-H, Forrester W, Lahav G. Schedule-dependent interaction between anticancer treatments. *Science*. 2016;351:1204-1208.
114. Catlin NR, Collins BJ, Auerbach SS, et al. How similar is similar enough? A sufficient similarity case study with *Ginkgo biloba* extract. *Food Chem Toxicol*. 2018;118:328-339.
115. Waidyanatha S, Ryan K, Roe AL, et al. Follow that botanical: challenges and recommendations for assessing absorption, distribution, metabolism and excretion of botanical dietary supplements. *Food Chem Toxicol*. 2018;121:194-202.
116. Sheng B, Vo Y, Lan P, Gardiner MG, Banwell MG, Sun P. Direct and metal-catalyzed photochemical dimerization of the phthalide (Z)-ligustilide leading to both [2 + 2] and [4 + 2] cycloadducts: application to total syntheses of tokinolides A-C and riligustilide. *Org Lett*. 2019;21:6295-6299.
117. Duric K, Liu Y, Chen S-N, et al. Studying mass balance and the stability of (Z)-ligustilide from *Angelica sinensis* helps to bridge a botanical instability-bioactivity chasm. *J Nat Prod*. 2019;82:2400-2408.
118. Chen S-N, Lankin DC, Chadwick LR, Jaki BU, Pauli GF. Dynamic residual complexity of natural products by qHNMR: solution stability of desmethylxanthohumol. *Planta Med*. 2009;75:757-762.
119. Jia M, Chen L, Xin H-L, et al. A friendly relationship between endophytic fungi and medicinal plants: a systematic review. *Front Microbiol*. 2016;7:906.
120. Li W, Chan C-K, Liu Y, et al. Aristolochic acids as persistent soil pollutants: determination of risk for human exposure and nephropathy from plant uptake. *J Agric Food Chem*. 2018;66:11468-11476.
121. Perry NB, Anderson RE, Brennan NJ, et al. Essential oils from dalmatian sage (*Salvia officinalis* L.): variations among individuals, plant parts, seasons, and sites. *J Agric Food Chem*. 1999;47:2048-2054.
122. Simmler C, Jones T, Anderson JR, et al. Species-specific standardisation of licorice by metabolomic profiling of flavanones and chalcones. *Phytochem Anal*. 2014;25:378-388.
123. Hajirahimkhan A, Mbachu O, Simmler C, et al. Estrogen receptor (ER) Subtype selectivity identifies 8-prenylapigenin as an ER β agonist from *Glycyrrhiza inflata* and highlights the importance of chemical and biological authentication. *J Nat Prod*. 2018;81:966-975.
124. Narayanapillai SC, Leitzman P, Gerard O'Sullivan M, Xing C. Flavokawains A and B in kava, not dihydromethysticin, potentiate acetaminophen-induced hepatotoxicity in C57BL/6 mice. *Chem Res Toxicol*. 2014;27:1871-1876.
125. Catlin NR, Collins BJ, Auerbach SS, et al. How similar is similar enough? A sufficient similarity case study with *Ginkgo biloba* extract. *Food Chem Toxicol*. 2018;118:328-339.
126. Barrington WT, Wulfridge P, Wells AE, et al. Improving metabolic health through precision dietetics in mice. *Genetics*. 2018;208:399-417.
127. Werz O, Gerstmeier J, Libreros S, et al. Human macrophages differentially produce specific resolvin or leukotriene signals that depend on bacterial pathogenicity. *Nat Commun*. 2018;9:59.
128. Zeisel SH. A conceptual framework for studying and investing in precision nutrition. *Front Genet*. 2019;10:200.
129. Fischbach MA. Microbiome: focus on causation and mechanism. *Cell*. 2018;174:785-790.
130. Gurley BJ, Yates CR, Markowitz JS. "...Not intended to diagnose, treat, cure or prevent any disease". 25 years of botanical dietary supplement research and the lessons learned. *Clin Pharmacol Ther*. 2018;104:470-483.
131. Clerici C, Setchell KDR, Battezzati PM, et al. Pasta naturally enriched with isoflavone aglycons from soy germ reduces serum lipids and improves markers of cardiovascular risk. *J Nutr*. 2007;137:2270-2278.
132. Dotson CD, Babich J, Steinle NI. Genetic predisposition and taste preference: impact on food intake and risk of chronic disease. *Curr Nutr Rep*. 2012;1:175-183.
133. Korach-Rechtman H, Freilich S, Gerassy-Vainberg S, et al. Murine genetic background has a stronger impact on the composition of the gut microbiota than maternal inoculation or exposure to unlike exogenous microbiota. *Appl Environ Microbiol*. 2019;85:e00826-19. <https://doi.org/10.1128/AEM.00826-19>
134. McBurney MI, Davis C, Fraser CM, et al. Establishing what constitutes a healthy human gut microbiome: state of the science, regulatory considerations, and future directions. *J Nutr*. 2019;149:1882-1895. <https://doi.org/10.1093/jn/nxz154>
135. Knopf K, Sturman JA, Armstrong M, Hayes KC. Taurine: an essential nutrient for the cat. *J Nutr*. 1978;108:773-778.
136. Shen L. More on vitamin C deficiency in mice and humans. *Hepatology*. 2011;53:2139; author reply 2139-2140.
137. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol*. 2011;9:279-290.
138. Goodrich JK, Waters JL, Poole AC, et al. Human genetics shape the gut microbiome. *Cell*. 2014;159:789-799.
139. Klurfeld DM, Davis CD, Karp RW, et al. Considerations for best practices in studies of fiber or other dietary components and the intestinal microbiome. *Am J Physiol Endocrinol Metab*. 2018;315:E1087-E1097.
140. Williams S. How the Microbiome Influences Drug Action. *The Scientist Magazine*. July 15, 2019. <https://www.the-scientist.com/features/how-the-microbiome-influences-drug-action-66081>. Accessed 18 November, 2019.

141. Novotny JA, Chen T-Y, Terekhov AI, et al. The effect of obesity and repeated exposure on pharmacokinetic response to grape polyphenols in humans. *Mol Nutr Food Res*. 2017;61:11. <https://doi.org/10.1002/mnfr.201700043>.
142. Ursell LK, Clemente JC, Rideout JR, Gevers D, Caporaso JG, Knight R. The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J Allergy Clin Immunol*. 2012;129:1204-1208.
143. Kleerebezem M. Microbial metabolic gatekeeping in the jejunum. *Nat Microbiol*. 2018;3:650-651.
144. Contijoch EJ, Britton GJ, Yang C, et al. Gut microbiota density influences host physiology and is shaped by host and microbial factors. *Elife*. 2019;8:e40553.
145. Frolinger T, Sims S, Smith C, et al. The gut microbiota composition affects dietary polyphenols-mediated cognitive resilience in mice by modulating the bioavailability of phenolic acids. *Sci Rep*. 2019;9:3546.
146. Zimmermann M, Zimmermann-Kogadeeva M, Wegmann R, Goodman AL. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*. 2019;570:462-467.
147. Ho L, Zhao D, Ono K, et al. Heterogeneity in gut microbiota drive polyphenol metabolism that influences α -synuclein misfolding and toxicity. *J Nutr Biochem*. 2019;64:170-181.
148. Setchell KDR, Brown NM, Lydeking-Olsen E. The clinical importance of the metabolite equol—a clue to the effectiveness of soy and its isoflavones. *J Nutr*. 2002;132:3577-3584.
149. Setchell KDR, Clerici C, Lephart ED, et al. S-equol, a potent ligand for estrogen receptor beta, is the exclusive enantiomeric form of the soy isoflavone metabolite produced by human intestinal bacterial flora. *Am J Clin Nutr*. 2005;81:1072-1079.
150. Setchell KDR, Cole SJ. Method of defining equol-producer status and its frequency among vegetarians. *J Nutr*. 2006;136:2188-2193.
151. Setchell KDR, Clerici C. Equol: history, chemistry, and formation. *J Nutr*. 2005;135:1355S-1362S.
152. Brown NM, Belles CA, Lindley SL, et al. Mammary gland differentiation by early life exposure to enantiomers of the soy isoflavone metabolite equol. *Food Chem Toxicol*. 2010;48:3042-3050.
153. Belzung C, Lemoine M. Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression. *Biol Mood Anxiety Disord*. 2011;1:9.
154. Garner JP. The significance of meaning: why do over 90% of behavioral neuroscience results fail to translate to humans, and what can we do to fix it? *ILAR J*. 2014;55:438-456.
155. McGonigle P, Ruggeri B. Animal models of human disease: challenges in enabling translation. *Biochem Pharmacol*. 2014;87:162-171.
156. Garner JP, Gaskill BN, Weber EM, Ahloy-Dallaire J, Pritchett-Corning KR. Introducing therioepistemology: the study of how knowledge is gained from animal research. *Lab Anim*. 2017;46:103-113.
157. Redan BW, Buhman KK, Novotny JA, Ferruzzi MG. Altered transport and metabolism of phenolic compounds in obesity and diabetes: implications for functional food development and assessment. *Adv Nutr*. 2016;7:1090-1104.
158. Chen T-Y, Ferruzzi MG, Wu Q-L, et al. Influence of diabetes on plasma pharmacokinetics and brain bioavailability of grape polyphenols and their phase II metabolites in the Zucker diabetic fatty rat. *Mol Nutr Food Res*. 2017;61:10. doi: 10.1002/mnfr.201700111
159. Beck JA, Lloyd S, Hafezparast M, et al. Genealogies of mouse inbred strains. *Nat Genet*. 2000;24:23-25.
160. Mestas J, Hughes CCW. Of mice and not men: differences between mouse and human immunology. *J Immunol*. 2004;172:2731-2738.
161. Zschaler J, Schlorke D, Arnhold J. Differences in innate immune response between man and mouse. *Crit Rev Immunol*. 2014;34:433-454.
162. Godec J, Tan Y, Liberzon A, et al. Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. *Immunity*. 2016;44:194-206.
163. Eruslanov EB, Singhal S, Albelda SM. Mouse versus human neutrophils in cancer: a major knowledge gap. *Trends Cancer*. 2017;3:149-160.
164. Herati RS, Wherry EJ. What is the predictive value of animal models for vaccine efficacy in humans? Consideration of strategies to improve the value of animal models. *Cold Spring Harb Perspect Biol*. 2018;11:a031583. <https://doi.org/10.1101/cshperspect.a031583>.
165. Talan M. Body temperature of C57BL/6J mice with age. *Exp Gerontol*. 1984;19:25-29.
166. Clark M, Steger-Hartmann T. A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. *Regul Toxicol Pharmacol*. 2018;96:94-105.
167. Becker KJ. Strain-related differences in the immune response: relevance to human stroke. *Transl Stroke Res*. 2016;7:303-312.
168. Velazquez EM, Nguyen H, Heasley KT, et al. Endogenous Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. *Nat Microbiol*. 2019;4:1057-1064.
169. Svenson KL, Von Smith R, Magnani PA, et al. Multiple trait measurements in 43 inbred mouse strains capture the phenotypic diversity characteristic of human populations. *J Appl Physiol*. 2007;102:2369-2378.
170. Morgan SJ, Elangbam CS, Berens S, et al. Use of animal models of human disease for nonclinical safety assessment of novel pharmaceuticals. *Toxicol Pathol*. 2013;41:508-518.
171. Justice MJ, Dhillon P. Using the mouse to model human disease: increasing validity and reproducibility. *Dis Model Mech*. 2016;9:101-103.
172. Salinero AE, Anderson BM, Zuloaga KL. Sex differences in the metabolic effects of diet-induced obesity vary by age of onset. *Int J Obes*. 2018;42:1088-1091.
173. Crusio WE, Goldowitz D, Holmes A, Wolfer D. Standards for the publication of mouse mutant studies. *Genes Brain Behav*. 2009;8:1-4.
174. Schmidt CW. Diversity outbred: a new generation of mouse model. *Environ Health Perspect*. 2015;123:A64-A67.
175. Costigan M, Scholz J, Woolf CJ. Neuropathic pain: a maladaptive response of the nervous system to damage. *Annu Rev Neurosci*. 2009;32:1-32.
176. Tuttle AH, Philip VM, Chesler EJ, Mogil JS. Comparing phenotypic variation between inbred and outbred mice. *Nat Methods*. 2018;15:994-996.
177. Shultz LD, Ishikawa F, Greiner DL. Humanized mice in translational biomedical research. *Nat Rev Immunol*. 2007;7:118-130.
178. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med*. 2009;1:6ra14.
179. Kashyap PC, Marcobal A, Ursell LK, et al. Genetically dictated change in host mucus carbohydrate landscape exerts a

- diet-dependent effect on the gut microbiota. *Proc Natl Acad Sci U S A*. 2013;110:17059-17064.
180. Hintze KJ, Cox JE, Rompato G, et al. Broad scope method for creating humanized animal models for animal health and disease research through antibiotic treatment and human fecal transfer. *Gut Microbes*. 2014;5:183-191.
 181. Masopust D, Sivula CP, Jameson SC. Of mice, dirty mice, and men: using mice to understand human immunology. *J Immunol*. 2017;199:383-388.
 182. Rosshart SP, Vassallo BG, Angeletti D, et al. Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell*. 2017;171:1015-1028.e13.
 183. Rosshart SP, Herz J, Vassallo BG, et al. Laboratory mice born to wild mice have natural microbiota and model human immune responses. *Science*. 2019;365:eaaw4361.
 184. Nair AB, Jacob S. A simple practice guide for dose conversion between animals and human. *J Basic Clin Physiol Pharmacol*. 2016;7:27-31.
 185. Piper MDW, Soultoukis GA, Blanc E, et al. matching dietary amino acid balance to the in silico-translated exome optimizes growth and reproduction without cost to lifespan. *Cell Metab*. 2017;25:1206.
 186. Sharma V, McNeill JH. To scale or not to scale: the principles of dose extrapolation. *Br J Pharmacol*. 2009;157:907-921.
 187. Tu H, Wang Y, Li H, Brinster LR, Levine M. Chemical transport knockout for oxidized vitamin C, dehydroascorbic acid, reveals its functions in vivo. *EBioMedicine*. 2017;23:125-135.
 188. Gabbay KH, Bohren KM, Morello R, Bertin T, Liu J, Vogel P. Ascorbate synthesis pathway: dual role of ascorbate in bone homeostasis. *J Biol Chem*. 2010;285:19510-19520.
 189. Selvaraj V, Zakroczymski MA, Naaz A, et al. Estrogenicity of the isoflavone metabolite equol on reproductive and non-reproductive organs in mice. *Biol Reprod*. 2004;71:966-972.
 190. Liang X, Bushman FD, FitzGerald GA. Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. *Proc Natl Acad Sci U S A*. 2015;112:10479-10484.
 191. Lévi F, Zidani R, Misset JL. Randomised multicentre trial of chronotherapy with oxaliplatin, fluorouracil, and folinic acid in metastatic colorectal cancer. *Lancet*. 1997;350:681-686.
 192. Ruben MD, Smith DF, FitzGerald GA, Hogenesch JB. Dosing time matters. *Science*. 2019;365:547-549.
 193. Liang X, Bushman FD, FitzGerald GA. Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. *Proc Natl Acad Sci U S A*. 2015;112:10479-10484.
 194. Davis DJ, Hecht PM, Jasarevic E, et al. Sex-specific effects of docosahexaenoic acid (DHA) on the microbiome and behavior of socially-isolated mice. *Brain Behav Immun*. 2017;59:38-48.
 195. Kim D, Hofstaedter CE, Zhao C, et al. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*. 2017;5:52. <https://doi.org/10.1186/s40168-017-0267-5>
 196. Yoshihara M, Yoshihara M. "Necessary and sufficient" in biology is not necessarily necessary—confusions and erroneous conclusions resulting from misapplied logic in the field of biology, especially neuroscience. *J Neurogenet*. 2018;32:53-64.
 197. Chilton FH, Murphy RC, Wilson BA, et al. Diet-gene interactions and PUFA metabolism: a potential contributor to health disparities and human diseases. *Nutrients*. 2014;6:1993-2022.
 198. Rizos EC, Ntzani EE, Bika E, Kostapanos MS, Elisaf MS. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. *JAMA*. 2012;308:1024-1033.
 199. Brown LL, Cohen B, Tabor D, Zappalà G, Maruvada P, Coates PM. The vitamin D paradox in Black Americans: a systems-based approach to investigating clinical practice, research, and public health—expert panel meeting report. *BMC Proc*. 2018;12(Suppl 6):6. <https://doi.org/10.1186/s12919-018-0102-4>.
 200. Sorger PK, Allerheiligen SRB, Abernethy DR, et al. *Quantitative and Systems Pharmacology in the Post-Genomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic Mechanisms*. Bethesda, MD: NIH; 2011.
 201. Hansen J, Iyengar R. Computation as the mechanistic bridge between precision medicine and systems therapeutics. *Clin Pharmacol Ther*. 2013;93:117-128.
 202. Stern AM, Schurdak ME, Bahar I, Berg JM, Taylor DL. A perspective on implementing a quantitative systems pharmacology platform for drug discovery and the advancement of personalized medicine. *J Biomol Screen*. 2016;21:521-534.
 203. Obach RS, Walsky RL, Venkatakrishnan K, Gaman EA, Houston JB, Tremaine LM. The utility of in vitro cytochrome P450 inhibition data in the prediction of drug-drug interactions. *J Pharmacol Exp Ther*. 2006;316:336-348.
 204. Di L, Feng B, Goosen TC, et al. A perspective on the prediction of drug pharmacokinetics and disposition in drug research and development. *Drug Metab Dispos*. 2013;41:1975-1993.
 205. Jones HM, Chen Y, Gibson C, et al. Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective. *Clin Pharmacol Ther*. 2015;97:247-262.
 206. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2-21.
 207. Gallin JI, Ognibene FP, Johnson LL. *Principles and Practice of Clinical Research*, 4th Ed. Cambridge, MA: Academic Press; 2017.
 208. National Center for Complementary and Integrative Health. *Data and Safety Monitoring of NCCIH-Funded Clinical Research*. NCCIH/NIH; 2012. <https://nccih.nih.gov/grants/policies/data-safety-monitoring>. Accessed August 14, 2019.
 209. Nehlig A. Interindividual differences in caffeine metabolism and factors driving caffeine consumption. *Pharmacol Rev*. 2018;70:384-411.
 210. WIRED Staff. *Placebos Are Getting More Effective. Drugmakers Are Desperate to Know Why*. WIRED; 24 Aug 2009. <https://www.wired.com/2009/08/ff-placebo-effect/>. Accessed 18 November 2019.
 211. Preston RA, Materson BJ, Reda DJ, Williams DW. Placebo-associated blood pressure response and adverse effects in the treatment of hypertension. *Arch Intern Med*. 2000;160:1449-1454.
 212. Barrett B, Brown R, Rakel D, et al. Placebo effects and the common cold: a randomized controlled trial. *Ann Fam Med*. 2011;9:312-322.
 213. Barrett B. Sufficiently important difference: concepts, caveats, and challenges. *Med Decis Making*. 2013;33:869-874.
 214. Albers C, Lakens D. When power analyses based on pilot data are biased: inaccurate effect size estimators and follow-up bias. *J Exp Soc Psychol*. 2018;74:187-195.
 215. Scargle JD. Publication Bias: The "File-Drawer" Problem. *J Sci Exploration*. 2000;14:91-106.
 216. Franco A, Malhotra N, Simonovits G. Social science. Publication bias in the social sciences: unlocking the file drawer. *Science*. 2014;345:1502-1505.
 217. Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology*. 2008;19:640-648.

218. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*. 2011;22:1359-1366.
219. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004;96:434-442.
220. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924-926.
221. Miller J, Schwarz W. Aggregate and individual replication probability within an explicit model of the research process. *Psychol Methods*. 2011;16:337-360.
222. Simonsohn U, Nelson LD, Simmons JP. p-Curve and effect size: correcting for publication bias using only significant results. *Perspect Psychol Sci*. 2014;9:666-681.
223. Stanley TD, Doucouliagos H. Meta-regression approximations to reduce publication selection bias. *Res Synth Methods*. 2014;5:60-78.
224. Kaelin WG Jr. Common pitfalls in preclinical cancer target validation. *Nat Rev Cancer*. 2017;17:425-440.
225. Munafò MR, Davey Smith G. Robust research needs many lines of evidence. *Nature*. 2018;553:399-401.
226. Shulman LP, Banuvar S, Fong HHS, Farnsworth NR. Discussion of a well-designed clinical trial which did not demonstrate effectiveness: UIC center for botanical dietary supplements research study of black cohosh and red clover. *Fitoterapia*. 2011;82:88-91.
227. Cui Y, Murphy B, Gentilcore A, et al. Multilevel modeling and value of information in clinical trial decision support. *BMC Syst Biol*. 2014;8:6. <https://doi.org/10.1186/s12918-014-0140-0>
228. Jayasekera J, Li Y, Schechter CB, et al. Simulation modeling of cancer clinical trials: application to omitting radiotherapy in low-risk breast cancer. *J Natl Cancer Inst*. 2018;110:1360-1369.
229. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J. Adaptive designs in clinical drug development—an executive summary of the PhRMA working group. *J Biopharm Stat*. 2006;16:275-283.
230. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med*. 2017;377:62-70.
231. Center for Drug Evaluation and Research. Expansion Cohorts: Use in First-In-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics. U.S. Food and Drug Administration; 2018. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/expansion-cohorts-use-first-human-clinical-trials-expedite-development-oncology-drugs-and-biologics>. Accessed 18 November 2019.
232. Meltzer DO, Hoomans T, Chung JW, Basu A. Minimal modeling approaches to value of information analysis for health research. *Med Decis Making*. 2011;31:E1-E22.
233. Hoomans T, Seidenfeld J, Meltzer D. Minimizing the costs of analyzing the value of health research. *Value Health*. 2012;15:A2.
234. Mirza RD, Guyatt GH. A randomized clinical trial of n-of-1 trials-tribulations of a trial. *JAMA Intern Med*. 2018;178:1378-1379.
235. Box GEP, Draper NR. *Empirical Model-Building and Response Surfaces*. New York, NY: John Wiley & Sons Inc; 1987.

How to cite this article: Sorkin BC, Kuszak AJ, Bloss G, et al. Improving natural product research translation: From source to clinical trial. *The FASEB Journal*. 2019;00:1–25. <https://doi.org/10.1096/fj.201902143R>