Assessing and communicating heterogeneity of treatment effects for patient subpopulations: Keynote and panel discussion on communicating heterogeneous treatment effects across populations

James Heyward1 | Milena Lolic2 | Catherine Spong3 | David Atkins4 | Gene Pennello5 | Ravi Varadhan6

1Center for Drug Safety and Effectiveness, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
2Office of the Center Director, U.S. Food and Drug Administration, Silver Spring, Maryland, USA
3Department of Obstetrics and Gynecology, UT Southwestern Medical Center, Dallas, Texas, USA
4Health Services Research and Development Service, U.S. Department of Veterans Affairs, Washington, District of Columbia, USA
5Division of Biostatistics, U.S. Food and Drug Administration, Silver Spring, Maryland, USA
6Division of Biostatistics and Bioinformatics, Department of Oncology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

Abstract
How do we communicate nuanced regulatory information to different audiences, recognizing that the consumer audience is very different from the physician audience? In particular, how do we communicate the heterogeneity of treatment effects - the potential differences in treatment effects based on sex, race, and age? That is a fundamental question at the heart of this panel discussion. Each panelist addressed a specific “challenge question” during their 5-minute presentation, and the list of questions is provided. The presentations were followed by a question and answer session with members of the audience and the panelists.

KEYWORDS
clinical trials, drug trial snapshot, subgroup analysis, treatment effects

This article is adapted from a panel discussion given at the Collaborative Symposium, "Assessing and Communicating Heterogeneity of Treatment Effects for Patient Subpopulations: Challenges and Opportunities," hosted by Johns Hopkins University Center of Excellence in Regulatory Science and Innovation (JHU-CERSI) and the U.S. Food and Drug Administration on November 28, 2018. The panel discussion was part of Session 3: “Communicating Heterogeneity of Treatment Effects to Key Stakeholders.”

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. Pharmaceutical Statistics published by John Wiley & Sons Ltd.
1 | INTRODUCTION

How do we communicate complicated regulatory information to different audiences, recognizing that the consumer audience is very different from the physician audience? That is a fundamental question at the heart of this session. In the words of Sir William Osler:

Variability is the law of life, and as no two faces are the same, so no two bodies are alike, and no two individuals react alike and behave alike under the abnormal conditions which we know as disease.

This may seem self-evident, but the reality is that some people do not agree that there are differences based on sex, race, and age. So, how do we move everyone onto the same page to be able to have the discussions about the heterogeneous treatment effects (HTE)?

Each panelist addressed a specific “challenge question” during their 5-minute presentation. The list of the questions is provided in Table 1. The presentations were followed by a question and answer session with members of the audience and the panelists.

The remainder of the paper will have one section for each question or topic given in Table 1, followed by a section on the reactions to the panel presentations, a section on audience questions and panelists’ responses and a final section on the moderator’s thoughts.

2 | FDA’S DRUG TRIAL SNAPSHOT PROGRAM (MILENA LOLIC, MD, MS, LEAD MEDICAL OFFICER, PROFESSIONAL AFFAIRS AND STAKEHOLDER ENGAGEMENT, CDER OFFICE OF CENTER DIRECTOR)

In the response to stakeholder concerns about adequate and equal inclusion of women and minority groups in clinical trials, in FDASIA 2012 Congress directed FDA to take a closer look at and report on the inclusion and analysis of demographic subgroups in applications for drugs, biologics, and devices. Recognizing the lack of easily accessible information about participation in drug trials, in 2015 the US FDA Center for Drug Evaluation and Research (CDER) created an initiative in transparency called the drug trial snapshots (DTS). The purpose of the DTS program is to make demographic data more available and transparent by “providing consumers and healthcare professionals with concise information about who participated in clinical trials that supported the FDA approval of new drugs.” DTS also include information about trial design and “where the trials were conducted and whether there were any differences in the benefits and side effects among different demographic groups.”

| Panelist                  | Question/Topic                                                                                                                                                                                                 |
|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| John Whyte (Keynote)      | FDA’s drug trial snapshots (DTS) program                                                                                                                                                                         |
| Iris Masucci              | How is HTE information typically presented in prescription drug labeling, and are there opportunities available for enhancement?                                                                                  |
| Catherine Spong           | What are the main issues surrounding heterogeneity of treatments effects in pregnant and lactating women, and how do current communication avenues (ex. drug labelling) serve clinical decision making for both providers and patients? |
| Cindy Geoghegan           | From a patient perspective, who would be responsible for ensuring that patients have access to information on HTE at point of treatment decision making?                                                        |
| David Atkins              | How do we help clinicians communicate specific information for important patient groups without overwhelming them—how do we balance desire for personalized information vs. need for feasible approaches for practice?      |
| Gene Pennello             | How can individualized treatment effect estimates and their uncertainty be communicated?                                                                                                                                 |
| Daniel Caños              | How does CMS promote the assessment of heterogeneity of treatment effects while clinical evidence is being developed? What sub-groups are of interest to Medicare when generalizability is considered for National Coverage Determinations? |
| Paula Rausch              | How can social science research be used to help overcome some of the challenges of communicating HTE?                                                                                                                |
The goal is to publish a snapshot for each novel drug (new molecular entity) on the web within 30 days of approval. In addition, each year, a DTS summary report is published. Individual and yearly reports can be found on https://www.fda.gov/drugs/drug-approvals-and-databases/drug-trials-snapshots.

DTS have an intended audience of consumers, medical professionals and policymakers and differ from the drug label in a few major ways. Relative to the drug package insert or “label,” the DTS is not intended to inform prescribing decisions but rather to provide consumer-friendly information on the specific subgroups included in the drug trials that provided support for drug approval. In addition, the DTS may contain subgroup-specific safety and efficacy data that are not required by law to be included in the label. Statisticians, and anyone who is interested in the data and analyses underlying the information in the DTS, can find that using the “MORE INFO” icon in each snapshot. Drug labels, on the other hand, are intended for healthcare professionals and include a comprehensive resource for drug use including chemical, toxicology and pharmacology information.

There are a number of challenges and opportunities for use of the DTS as a tool to communicate HTE to the public. Namely:

1. Is there enough data to make conclusions about efficacy and safety for all subgroups?
2. When is generalizability of subgroup analyses ok?
3. When differences among subgroups are observed, are differences clinically meaningful?

Additional emergent questions are; how do regulators compete with the major sources of online information for consumers (“Dr Google”)? Consumers are not coming to FDA for information about drugs. They are using WebMD, or social media. Approximately the same number (72%) get their information online as from medical professionals. On social media, followers of Dr Oz far outweigh FDA, NIH, CDC, Mayo Clinic or even WebMD. But some are doing better than others: CDC has 5x followers than FDA—why is this and what can we learn?

Increasingly, consumers are empowered to inform themselves and to take actions, online, without consultation with a medical professional—for example, dermatologic lesion diagnosis and exam; and optometric exams for glasses. How do we ensure there are credible information sources at consumers’ disposal, and increase use of DTS in the intended manner? We can put credible content out there but it does not mean consumers will engage with it. How can we improve?

3 HOW IS HTE INFORMATION TYPICALLY PRESENTED IN PRESCRIPTION DRUG LABELING, AND ARE THERE OPPORTUNITIES AVAILABLE FOR ENHANCEMENT? (IRIS MASUCCI, PHARMD, SPECIAL ASSISTANT FOR LABELING, CDER OFFICE OF MEDICAL POLICY)

FDA-approved prescription drug labeling is the Agency’s primary communication tool for health care professionals. While labeling does not include everything that is known about the drug, it contains the information that is essential for safe and effective use of the drug. Information in labeling is used as the basis for other streams of information to the medical and patient communities, such as third party drug information providers, payors, and health care systems. Labeling also serves as the basis for what is permitted in prescription drug advertising and promotion.

HTE information is primarily conveyed in the “Clinical Studies” section of labeling. This section of labeling presents a summary of the clinical studies supporting the drug approval and includes findings from the required explorations of age, sex, and race. Here, the labeling will include the findings of any analysis that had the ability to detect a difference in such populations or will state when the sample size was too small permit any conclusions. FDA recommends that, when appropriate, a statement be included about the inherent risks of any unplanned subgroup analysis.

HTE-related information can be presented elsewhere in labeling, including in the “Use in Specific Populations” section—specifically, use of the drug in pregnant or lactating women, pediatric patients, geriatric patients, and other subpopulations as needed (eg, patients with renal or hepatic impairment, smokers, or patients with specific genetic variants). HTE information can also be found in other sections of the labeling if those findings result in important clinical consequences, such as specific dosing recommendations for a certain population of patients in the “Dosage and Administration” section.

HTE information appears in labeling in a variety of ways. The most common presentation in the “Clinical Studies” section is a short paragraph, typically at the end of the presentation of the study results, saying that, for example, an
examination of population subgroups did not reveal any clinically significant differences in responsiveness in the evaluated subgroups. An alternative presentation is the use of Forest plots. These figures typically include results in a variety of subpopulations, and the data are presented only descriptively. Forest plots in labeling are typically accompanied by a statement about the risks associated such unplanned subgroup analyses.

To date, Forest plots have appeared in labeling primarily when summarizing large outcome studies in the cardiovascular setting, which generally have thousands, if not tens of thousands, of patients. On the other hand, an oncology trial of 100 patients could pose a different challenge to consider the use of Forest plots in the labeling for those drugs. There is much discussion at the Agency about the utility of these plots in labeling, including the need to strike a balance between providing information and being transparent about what was done in a study, and the possibility of over-interpretation of the results or providing information that may be misinterpreted or misleading. To address these issues, it is important to remember the role of labeling—to provide the information that health care providers need for informed patient care decision-making.

**WHAT ARE THE MAIN ISSUES SURROUNDING HETEROGENEITY OF TREATMENTS EFFECTS IN PREGNANT AND LACTATING WOMEN, AND HOW DO CURRENT COMMUNICATION AVENUES (EX. DRUG LABELING) SERVE CLINICAL DECISION-MAKING FOR BOTH PROVIDERS AND PATIENTS? (CATHERINE SPONG, MD, PROFESSOR AND VICE CHAIR, DEPARTMENT OF OBSTETRICS AND GYNECOLOGY, UT SOUTHWESTERN MEDICAL CENTER)**

Pregnant and lactating women are a key group as their health impacts the entire population—each person was part of at least one pregnancy during their lifetime. Importantly, pregnant and lactating women are complex and extremely underrepresented in research and data collection. The numerous physiologic changes for both mother and fetus across gestation add to the complexity. These include changes in kidney and cardiac function, the slowing of the GI tract, and alterations in serum binding proteins in blood that affect drug activity, over the course of a pregnancy (Figure 1).

A well cited example is that of amoxicillin for treating anthrax. The recommended doses were not found to be effective when studied in pregnancy and suggest that amoxicillin concentrations adequate to prevent anthrax may be difficult to achieve during pregnancy and postpartum. Regarding lactation, not only does maternal physiology change across the postpartum period, the composition of human breast milk changes also changes, especially weeks 1 to 6, and within each lactation episode (foremilk vs hindmilk). Data in lactation is essential for both mother and baby. Limited data are available—for example, regarding asthma and pregnancy—Table 2 depicts the low availability of research on asthma drugs in that population despite it being a common condition. Almost all the pregnancy- and lactation-related research focused on pregnancy only, and not lactation. Of note, prevalence of asthma in pregnant women is ~8.5%, with 4% of women experiencing an asthma attack in the prior year.

**FIGURE 1** Physiologic changes across gestation
The HHS task force on research specific to pregnant women and lactating women (PRGLAC) had its genesis from the 21st Century Cures Act. The recommendations were submitted to Congress and the HHS Secretary and the task force now is working on implementation plans for the recommendations.

Regarding communication strategies—such as the drug label, PLLR modifications—these are vitally important for both providers and women. According to an FDA study, among 570 prescription label changes from 2015 to 2017, only 22% had data about pregnancy, 10% had data on lactation, and 70% had no human data for pregnancy or lactation. What information is available should also communicate what is not known. Opportunities for better accrual and communication of HTE exist: registries can be a tool to this end. The FDA maintains a list of active pregnancy exposure registries at http://www.fda.gov/pregnancyregistries, along with other resources including patient education resources. There are 67 active pregnancy exposure registries, and 11 active lactation exposure registries. To date, 8000 women have enrolled in pregnancy registries, and 27 in lactation registries. Ultimately, there is lots more to do in generating and communicating information about what we know and do not know about drugs in pregnant and lactating women.

5 | FROM A PATIENT PERSPECTIVE, WHO WOULD BE RESPONSIBLE FOR ENSURING THAT PATIENTS HAVE ACCESS TO INFORMATION ON HTE AT POINT OF TREATMENT DECISION-MAKING? (CINDY GEOGHEGAN, PRINCIPAL, PATIENT AND PARTNERS LLC)

Patients often seek HTE information when they are told “no”—such as by the insurer refusing coverage for a treatment, or when they are refused therapy at the point of care. The reason in cases such as these is that “we have no data on someone such as you.” Thus, from a patient perspective, it is important to have information about “patients like me.” When it comes to communicating HTE, it is important to consider that consumers have no understanding of the drug development process—health literacy is generally low—this is why Dr Oz is so popular. They may not recognize the importance of the FDA and of the clinical trials and may be more concerned with co-pay, and other aspects of coverage. When it comes to messaging about health, the consumers do not necessarily know what they are listening to and hearing about.

Another factor to consider when it comes to consumer decision-making is personal values—not everyone wants access to the latest drug or things others do not have access to. In addition, not all treatment effects matter to patients, so effects such as a biomarker or size of tumor, which a patient cannot feel, may not matter to them. However, it is ultimately important that patients have access to this information.

6 | HOW DO WE HELP CLINICIANS COMMUNICATE SPECIFIC INFORMATION FOR IMPORTANT PATIENT GROUPS WITHOUT OVERWHELMING THEM: HOW DO WE BALANCE DESIRE FOR PERSONALIZED INFORMATION VS NEED FOR FEASIBLE APPROACHES FOR PRACTICE? (DAVID ATKINS, MD, MPH, DIRECTOR, VA HEALTH SERVICES RESEARCH AND DEVELOPMENT [HSR&D])

When it comes to communicating HTE to stakeholders, some important questions to consider are:

1. When do we know that the average treatment effect does not apply to (ie, is clinically different than) the treatment effect in specific subgroups (true HTE)?
2. When do we know that specific factors will have important effects on the balance of benefits and harms (variation in net benefit)?

These questions come into relief when one considers the schematic above depicting treatment effects across clinical trial populations (Figure 2). Sample 1 and sample 2 represent different subgroups with distinct treatment effects, as in situation (1). But frequently even when treatment effects do not so clearly depend on individual factors (sample 3), clinicians may need to discuss how the net benefits may depend on a variety of considerations (eg, age, comorbidity).

The take home points are:

1. All clinical decisions should be individualized (Some of the pushback against this idea is not about whether evidence is right, but about how it is been applied in guidance).
2. What is beneficial or best treatment “on average” may not be beneficial to a given individual (and vice versa).
3. Benefits, harms, and balance of benefits and harms may vary.
4. Clinicians should consider and communicate those factors that are most important to balance of benefits and harms.
5. Heterogeneity is only one source (and possibly not most important source) of variation.

What are the sources of variation in treatment effect?

1. Genetics.
2. Genomics—bright future where our whole genome will be sequenced and we will know the right drug or condition.
3. *Heterogeneity of treatment effect*.
4. Applicability of trial results:
   - Package insert summarizes trial data—which has limited generalizability to the population at large—those excluded or not the cohort that tend to volunteer for trials.
5. Baseline risk of outcome being treated (age, other risk factors).
6. Risk of harms—overemphasis on harms may strike a bad balance with the benefits for a healthy patient.
7. Values may trump a lot of things the clinician thinks are important. Missing the forest for the trees.

One illustrative example of HTE communication in the field is the VeteransLikeMe study at the VA (Table 3). We mined VA data to make decisions for patients. This initiative began due to case in which a terrible prescribing decision made based on trial data—a 73 year old male with stent, on antithrombotics, who died from multiple bleeding complications, had an apparent 2% risk of dying with stent—but that was only one aspect of the patient's makeup. Having a stent implanted after a MI, on dialysis, over the age of 70, carries a much larger mortality risk, using VA data—2% risk goes up to 11%, and to 50% within 2 years (all cause). This particular initiative is a testament to the potential power of observational data combined with clinical trial data to get data more relevant to individual patients.
Patients are heterogeneous. Thus, in any given clinical study, treatment effects may be heterogeneous among subgroups. A priori, HTE may be anticipated or unanticipated.

The problem with subgroup analysis is that sample estimates of subgroup-specific treatment effects are subject to variation due to random sampling as well as variation in the true treatment effects. Thus, the sample estimates will tend to exhibit larger variation than the true treatment effects, which can lead random highs or lows that appear significant. For example, the true treatment effects may be homogenous, yet the sample estimates may exhibit considerable variation especially if based on small sample sizes, as illustrated in Figure 3.

Without correction, random sampling variation increases the chance that a subgroup analysis will falsely declare that one or more subgroup-specific treatment effects are statistically significant. To guard against this overinterpretation, several statistical approaches are possible. For brevity, one may consider communicating only the subgroups in which treatment effects vary significantly from the overall treatment effect after using a method to correct for random sampling variability.

Gelman et al use Bayesian hierarchical models to correct for random sampling variation. These models produce shrinkage estimates for the subgroups. Relative to the sample estimates, shrinkage estimates are pulled back to an estimate of average treatment effect across all subgroups, thus reducing the chance of random highs and lows that are falsely statistically significant. A shrinkage estimate for a subgroup borrows strength from the data on all the subgroups and thus is more precise than the sample estimate, which is only based on the data in that subgroup. For certain decision-theoretic models, Bayesian shrinkage estimates for subgroups control the directional false discovery rate.

### TABLE 3
Data from “VeteransLikeMe” research study (Reproduced with permission from Qing Zeng, PhD, unpublished)

| Query criteria                                      | 30 days (%) | 1 year (%) | 2 years (%) | Total patients |
|-----------------------------------------------------|-------------|------------|-------------|----------------|
| 1. Coronary stent                                    | 1235 (2)    | 4733 (8)   | 7334 (12)   | 62,146         |
| 2. Coronary stent + History of MI                    | 920 (3)     | 3081 (11)  | 4508 (16)   | 27,850         |
| 3. Coronary stent + History of MI + dialysis         | 87 (7)      | 301 (25)   | 435 (36)    | 1200           |
| 4. Coronary stent + History of MI + dialysis + 70-75 year old male Caucasian | 8 (11)      | 24 (34)    | 34 (48)     | 70             |

**FIGURE 3** The subgroup problem.  
*Source: Gelman et al* (Reproduced with permission from author and Taylor & Francis)
As described in Gelman et al and illustrated in Figure 4, a number of potential approaches exist to attenuate the risk of overinterpreting subgroup-specific treatment effects.

In the example, the subgroups are sites. In Figure 4, estimates and confidence intervals for the treatment effects within sites are illustrated for the three approaches. The Bonferroni correction retains the sample estimates but increases the width the confidence intervals relative to no correction (classical linear regression). In contrast, the multilevel hierarchical model provides shrinkage estimates with shorter confidence intervals than provided by no correction.

1. Classical linear regression.
2. Classical linear regression with Bonferroni correction—expands the CIs.
3. Multilevel model—Bayesian subgroup hierarchical model—shrinks estimates. Reduces type 1 and type 2 errors. Under the right conditions, shrinkage estimates might be the best one to use in a DTS.

### 7.2 Individualized treatment effects

Subgroup analysis is important for individualized medicine and clinical decision support (CDS). The most granular subgroup is the individual. Clinically, the individual is the most important subgroup. Communication of the individualized treatment effects can be difficult, but a prediction interval of an outcome in a future patient is designed to capture the full uncertainty in outcomes among individuals.

For example, CYP2C9 and VKORC1 genotypes are used to predict the stable therapeutic dose of warfarin in individuals. Genotype prediction is a useful guide for determining initial warfarin dose. Uncertainty in the actual stable dose in an individual is communicated with a prediction interval. To account for uncertainty in individual outcomes, a prediction interval is typically much wider than the corresponding confidence interval for the mean stable dose in a population of individuals with the same genotypes. While less uncertain, the confidence interval on the mean stable dose is not as clinically meaningful to communicate than the corresponding prediction interval of the stable dose for an individual. The prediction interval approach is illustrated for the warfarin dose by genotype example in slides on the symposium webpage on the JHU CERSI website.

Treatment effect estimates for every individual in a clinical trial can be communicated succinctly by ordering individuals by their estimates as in Figure 1 of Henderson et al. These authors used Bayesian additive regression trees to estimate individualized treatment effects. In general, prediction intervals should be used to account for uncertainty in outcomes for future individuals.

### 7.3 Confounding in subgroup analysis

Subgroup analyses can be confounded. For example, the subgroups based on one subgrouping variable can be confounded by the effects of another subgrouping variable. Multiway subgroup analysis strategies that correct for confounding are explained in Varadhan and Wang and Pennello and Rothmann.
Combination of categorical covariates can generate a large number of strata. For binary outcomes, the Cochran-Mantel-Haenszel (CMH) method or conditional logistic regression can be used to adjust odds ratios, risk differences, and risk ratios for stratum effects even when data are sparse within strata.\textsuperscript{10}

### 7.4 Causal effect estimation

If they can be estimated, the causal effects of treatments should be communicated. A causal effect is the difference in outcome if an individual received a new treatment than if she received the control treatment. Unfortunately, the outcome in an individual can be only be observed for one of the treatments. However, when treatment assignment is randomized in a clinical trial, the resulting treatment effect estimate is also the causal effect estimate.

### 8 HOW DOES CMS PROMOTE THE ASSESSMENT OF HETEROGENEITY OF TREATMENT EFFECTS WHILE CLINICAL EVIDENCE IS BEING DEVELOPED? WHAT SUBGROUPS ARE OF INTEREST TO MEDICARE WHEN GENERALIZABILITY IS CONSIDERED FOR NATIONAL COVERAGE DETERMINATIONS? (DANIEL CAÑOS PHD, MPH, DIRECTOR, CDRH OFFICE OF CLINICAL EVIDENCE AND ANALYSIS)

An illustration of how HTE information is considered in older populations is provided in CMS' medical device coverage determinations (risks/benefits within the Medicare population). Clinical trial sponsors are required by law to disclose on ClinicalTrials.gov how device would act in Medicare population (geriatric), and CMS regional offices (A/B MAC jurisdictions) make coverage decisions based on this data—they usually make a determination of universal coverage, no need for national coverage determination.

The national office is asked to weigh in and render a National Coverage Decision (NCD) when there is less data or when there are conflicting data from different subgroups (HTE). There are many potential issues with assessment of Medicare beneficiaries in clinical studies—many studies do not have many enrollees over the age of 65, because of exclusions due to:

1. heterogeneity—may have multiple comorbidities and/or be taking multiple medications.
2. nonadherence—may have difficulty following protocols and/or making study follow-up visits.
3. other considerations—measurement issues, cognitive function.

For this reason, sometimes coverage is granted while further evidence is awaited (Coverage with Evidence Development, [CED]). CED allows for positive coverage when evidence is insufficient for a more favorable decision. Evidence gaps may be due to a low number of beneficiaries in clinical studies, lack of meaningful health outcomes, limited generalizability, or inconsistency of study findings. Without a CED determination, some items or services would be non-covered. CED may involve ongoing randomized controlled trials, observational studies, and/or registries assessing a specific intervention, or specific aspects of the evidence base such as benefits and harms or health outcomes.

Shared decision-making is required in cases such as this—a discussion takes place before the therapy is administered, there is risk calculator, to determine what the benefits and risks are for that particular patient.

### 9 HOW CAN SOCIAL SCIENCE RESEARCH BE USED TO HELP OVERCOME SOME OF THE CHALLENGES OF COMMUNICATING HTE? (PANELIST: PAULA RAUSCH, PHD, RN, ASSOCIATE DIRECTOR OF RESEARCH AND RISK COMMUNICATIONS, CDER OFFICE OF COMMUNICATIONS)

Communicating broad messages to patients and health care professionals that enable them to have the targeted information they need is difficult. Dr Whyte discussed the DTS, and Dr Masucci discussed drug labeling, but CDER and FDA communicate about a lot of different topics in many different other ways also, which is where some of the issues related to HTE comes into play.
The question that I am going to look at today is how social science research can be used to help overcome some of these challenges related to communicating HTE. CDER/OCOMM conducts social science research on how to best communicate with various audiences about these topics with the objective of trying to understand behavioral or social processes that can predict or influence health outcomes or risk factors. The behavioral side involves actions people might take with respect to some health issue and the underlying psychological processes that might relate to those actions, and the interactions between them. On the social side we look at the roles played by culture, economics, and demographic status. We are also looking at bio-social interactions and various levels of social context on small groups to larger systems. We are using a variety of methodologies to do this. Much of what we do in the Office of Communications is applied research, but we also do some basic research. There are several cross-cutting themes that traverse social and behavioral research more generally. In the context of today’s symposium, what is most relevant is the emphasis on individuality, the variation therein, and the variation across socio-demographic categories such as gender, age, education, income, and health literacy.

When we are exploring these issues, we use both qualitative and quantitative research methods.

On the qualitative side, we start from the basics that would be include conducting reviews of the literature and environmental scans of various websites, and research of conversations occurring on social media platforms and other online sites. These methods are meant to be exploratory, drawing as much and as broad of information as possible about what is already known about the topic. We use that information to help narrow the focus to the most important areas of interest that we can ask people about through focus groups or interviews. These individual or group discussions give us an opportunity to get information directly from patients or health care professionals, including those of different races, income, education levels and gender. As part of these conversations, we can talk with them about things that are of interest or concern to them, and the kinds of information they need so that we can make better communication decisions. Workshops, guided listening sessions and direct observation are the other research methods that can be used to help us get the kind of in-depth understanding we are often seeking when we use qualitative methods.

On the quantitative side we try to get more representative data through surveys and other methods that allow us to look more specifically at differences across groups or individuals. For example, we explore different types of messaging and how different groups respond to those messages. Typically, we use both qualitative and quantitative methods to help us understand the diverse aspects of the topics being studied, as well as the diversity of knowledge, attitudes, beliefs, and behaviors among the heterogenous groups of individuals involved in these studies.

One important focus area is understanding of FDA messaging and the effects that messaging might have, for example, how clear they are, how relevant or important they are to the people reading or hearing them, and how they react emotionally to them. We also want to know how the information affects their perceptions, for example, how severe or susceptible they may feel toward a safety risk we are communicating about; how it may affect their decision-making; what their past experiences have been with something such as medications; their opinions and attitudes; and their actions and behavioral intentions. We can then make comparisons of all these things across various different socio-demographic groups in order to help us identify where and how best these different needs, desires, experiences, opinions, etc. may be able to be best addressed through communications and/or in other ways.

10 | AUDIENCE QUESTIONS AND PANELISTS’ RESPONSES

10.1 | Question 1

John Whyte:
I’ll start with a question for Dr Spong. You said we need to communicate what we know as well as what is unknown so how do we communicate what’s unknown if we don’t know it?

Catherine Spong:
I think a good example is simply from the labeling information or from the PLO data information [on the label] where you’re providing what information is available, and by adding “we have no studies in humans” that is conveying that we don’t have any information in this group. But, I think it’s important for people to recognize that and to communicate, understanding that communication [of not having data in the subgroup] is equally as important as the information that is known because without giving that then patients just assume it’s safe.

John Whyte:
—but the construct can also be that people don’t understand that the absence of evidence is not the evidence of absence. We have the real possibility that when we don’t have women in a trial that people feel the drug won’t work on women when the reality is we may not know. How do we address that?

Catherine Spong:
That’s a key point for our communications groups to figure out; how to communicate that—because not communicating that you don’t have information in that subgroup can lead to lots of different interpretations for patients. I think it’s really key as a provider, when I am talking with a patient who says “I am on this medication- is it safe for me?” to be able to say “this is the information that I know, this is what I don’t know and we’re going to have to work together to say how important is it for you to be on that medication compared to what information we currently have available,” and having that robust dialogue with that patient.

Cindy Geoghegan:
Yeah, I guess I kind of agree with both of you. You can’t say we don’t know anything, but when in specific cases you don’t have information, say, in lactating women I think it would be important to say that. I also think it would do a lot of service to clinical research if we were more apt to say “we don’t have answers in people like you” coming in the appropriate context so that people understand why they need to participate in research because again people don’t understand.

John Whyte:
—and how are you going to provide them that information is going to be in the label? Iris, where are they going to go to see that there were no people like that?

Iris Masucci:
There are circumstances in which the absence of information can be important to convey in labeling and needs to be clearly explained. For example, a statement that “safety and effectiveness have not been demonstrated” in a particular group may cause some readers to conclude that the drug has not been studied in that population, while other readers may conclude that studies have been performed but have failed to show a benefit.

Paula Rausch:
We just finished a research project where we developed our “Framework for Communicating Benefits, Risks and Uncertainties,” where we specifically tried to identify best practices on each of these, especially related to uncertainty, which we and others have found to be extreme difficult to communicate in ways that provide adequate understanding, particularly to lay audiences. When we started with the literature to try to identify best practices for doing that, there really weren’t many, and those we found were unable to address some of the unique constraints we faced or that addressed medication-related uncertainties. So, we really had to go through a very detailed series of studies that included focus groups followed by individual interviews followed by an experimental study to get an in-depth enough understanding of some of these things up that would provide us some practical guidance on how to explain uncertainties in ways that would better people understood them. Despite how difficult it was for people in our studies to understand uncertainties they were asked about, one thing that we heard almost universally was that people still wanted to be informed about them, essentially saying, “I don’t want you to overlook telling me that there are uncertainties. I want to know about them and have that information so that it can help me in my decision-making and it can help me in my discussions with my health care provider.” So figuring out how to explain and discuss uncertainties effectively is critical, no matter the audience.

10.2 | Question 2

Ravi Varadhan:
Something that Cindy said struck me; that the first encounter that the patient has with HTE is when they’re told “no” to a potential treatment that they would like to receive. So, I want to ask what makes the clinician say “no”—on what information was that decision based. In other words what sort of HTE assessment is done to justify that kind of a decision? So, I think it’s going back to whatever was in the labeling insert or was it some examination of the generalizability of the evidence to that particular type of patient, or what sort of things? Maybe there’s an example that could be shared.

Cindy Geoghegan:
I don’t know where the evidence came from- I’m just kind of sharing patient experiences. In many cases the only evidence that they hear about—they don’t need labels and they’re not going to the FDA website—but when you know you get a prescription you bring it to the pharmacy, or if you go to kind of pre-register for a procedure to have a device sensor implanted you’re told “this isn’t going to be covered.” So, the patient doesn’t have the information but then they
subsequently might learn the evidence does not support effectiveness in people that look like them. So, again I don’t know where that evidence is generated but you know I’ve been observing this. You used to go to CVS for the prescription and they would say “do you have any questions for your pharmacist?” Now, they say “are you aware of the copay on this” or “it’s not going to be covered” and people just walk away. So, I know that some of the purview of the FDA but when we’re talking about treatment effects we need to think about a longer term, you know, analytics and things that are going on to individualize decisions so do—.

John Whyte:
Where do you get your drug information?
Cindy Geoghegan:
I’ve been told I’m not a normal person—I do a lot of research, I look at labels, I look at all the research, I look at every trial and anybody that’s related to me, whether they want the information or not, will have it. But most people, and you’d really be surprised at the level of a lot of people that do not.

Daniel Caños:
Very good question. So, I mapped out earlier there are different layers to the onion that is CMS where there’s local coverage determinations for the Medicare Advantage program and there’s also for service on the national level. So, when you get into a noncoverage, that could be a local noncoverage where the other MAC jurisdictions would cover it, and you happen to be a jurisdiction that doesn’t cover it. They could have a noncoverage, so that’s one instance where you have noncoverage where one MAC finds the evidence not sufficient for a certain patient indication. From the national level you asked the question about where to get your out your evidence or where do you assess. We certainly look at the FDA label, look at professional society consensus statements, guidelines and peer reviewed publications. Again, go back to the implantable cardioverter defibrillator example. You know many of ours are based on evidence and if it’s consistent with the guidelines as far as the evidence audience goes you know we map that out to what those covered indications look like and that tracks there. So that would be an instance where we look at the actual indications I think for that actual device, ICDs at the FDA side is more functional indication—“if it defibrillates, then it defibrillates” and it’s not a patient-based indication, that’s a little different. But that’s where we would have some kind of differences as far as the indicated population goes. I’d also say, though, with many of these decisions that we say if you’re not part of the indication for the UM national coverage, the positive coverage on that, we also allow for those patients to be studied with an investigational device exemption study, a category B IDE study, premarket FDA study, so we still want to afford appropriate access to our beneficiaries for technologies even if there’s not an established evidence base there so we’ll allow for it to be covered under category B IDEs. That’s the case for ICDs where they fall out, they would be noncovered if they could be in a study that’s FDA approved and in CMS covering the study. Another option for that we allowed to provide appropriate access was the coverage with evidence development that I laid out before, that covers wear such as transcatheter or aortic valve replacement for some very interesting diagnoses where we have a nationally covered if you’re in a national registry and there’s covered with evidence developments that you can also be covered with the premarket clinical trial. So, you know from the national level to review these coverage determinations we do follow the guidelines consensus statement publications in the FDA label where applicable, and from there we do try to work to develop an evidence base where there is either of evidence for subgroups, or there’s inconclusive evidence as far as heterogeneity of treatment effect.

John Whyte:
Iris, you had a comment.
Iris Masucci:
FDA bases its approval of an indication on the data that has been submitted, and carefully considers how broad or narrow an indication should be. Most often, the indicated population mirrors the population in which the drug was studied. There are times, however, when FDA determines that an indication should include a broader population than was studied based on the data, and there are other times when an indication is narrower in scope than the population that was studied because perhaps the was benefit only in a certain subgroup, or there were safety concerns in another subgroup. Potential reimbursement coverage does not factor into these decisions, but the scope of the indication does, and gets a lot of review and discussion.

John Whyte:
I do have a question for Paula Rausch... what can we do to increase awareness? Do we need another campaign like the tobacco one to direct folks to FDA websites on drug information?
Paula Rausch:
So another thing that comes across, and this won’t be any surprise to anybody, is that people are only interested in and only have time for information that is directly related to them or their interests. Given the diversity of the people FDA provides information to, it’s really very difficult to provide targeted information that people can select to receive or not. We’re looking at ways to try and do that to the extent it might be possible. For example, with CDER’s Drug Safety Communications—which are the primary communication tool FDA uses to provide information about new or emerging safety issues with drugs after they’re on the market—we’re exploring different ways that allow individual patients and health care professionals to identify an receive just the information they want. Right now, we just don’t have those capabilities, but we’re hoping that technology will give us the ability to do that because we understand that targeting our information and having it relevant to people is key when we’re trying to make them aware of the information and having it be useful to them.

John Whyte:
Final question, microphone.

Steven Ruberg:

Most everything is supported these days out of clinical trials with an “intent to treat” population, which is kind of a just a combination of everybody randomized in the trial, and we figure out a way to include who don’t take the medications—dropout, have side effects, whatever and you give an overall treatment estimate. Akacha et al published something in Stats in Medicine in 2016 called the tripartite approach or estimation of treatment effects recognizing the treatments have different effects. They argue that out of clinical trials, we should publish or we should estimate the probability that you will not be able to take the drug because of the side effects, and then the probability that you won’t be able to take the drug because of lack of efficacy. And then for those people who can actually take the drug, what is the magnitude of the treatment effect if they have to take it? I’d be very interested in Cindy, from a patient perspective, hearing your thoughts about this. Those aren’t exactly the patient subpopulations we’ve been talking about but it is another way to say treatments have heterogeneous effects- not everybody can get efficacy, some discontinued because of safety problems. Iris, I’ve talked to folks who say “well that would be great,” that’s the kind of information that should be up front the label—when you take 10 milligrams of this drug there’s a 20% chance you’re going to have a side effect and a 20% chance it’s not going to work but for the 60% of you who can take it you’ll see 1 1/2% drop in your HBA1C or such and such. And then maybe from the CMS perspective, health economics version, those 3 estimations or groups of people who fall into those populations actually are very helpful for health economic modeling because they both they all have very different cost implications. People might take a drug for a short period of time, have a side effect and need lots of follow-up.

Cindy Geoghegan:

Patients would love to have access to that kind of information. I think the expectation is it’s going to work for them so I guess that’s a good idea.

Iris Masucci:

That would be challenging to do at the time of a drug approval, given how drugs are reviewed and how the data are summarized in labeling. The location of information in labeling is governed by regulation, so there is not a lot of flexibility available.

Daniel Caños:

So we generally don’t consider cost for coverage. We consider the clinically meaningful health outcomes, as you know those harder endpoints and don’t look at all these or QALYs or any such thing, generally speaking, except for services.

ACKNOWLEDGEMENTS
This work was supported in part by a Center of Excellence in Regulatory Science and Innovation (CERSI) grant to Johns Hopkins University from the U.S. Food & Drug Administration U01FD005942. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the HHS or FDA.

CONFLICT OF INTEREST
The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
James Heyward https://orcid.org/0000-0003-1375-9510
Ravi Varadhan https://orcid.org/0000-0002-8434-1034
REFERENCES

1. Food and Drug Administration Safety and Innovation Act of 2012. http://www.gpo.gov/fdsys/pkg/BILLS-112s3187enr/pdf/BILLS-112s3187enr.pdf
2. Andrew MA, Easterling TR, Carr DB, et al. Amoxicillin pharmacokinetics in pregnant women: modeling and simulations of dosage strategies. Clin Pharmacol Ther. 2007;81(4):547-556. https://doi.org/10.1038/sj.clpt.6100126.
3. Task Force on Research Specific to Pregnant Women and Lactating Women (PRGLAC). Report to Secretary, Health and Human Services and Congress. 2018. https://nichd.nih.gov/sites/default/files/2018-09/PRGLAC_Report.pdf
4. 21st Century Cures Act. https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf
5. Gelman A, Hill J, Yajima M. Why we (usually) don’t have to worry about multiple comparisons. J Res Educ Eff. 2012;5(2):189-211. https://doi.org/10.1080/19345747.2011.618213.
6. Lewis C, Thayer DT. A loss function related to the FDR for random effects multiple comparisons. J Stat Plan Inf. 2004;125:49-58.
7. Henderson NC, Louis TA, Rosner GL, Varadhan R. Individualized treatment effects with censored data via fully nonparametric Bayesian accelerated failure time models. Biostatistics. 2020;21(1):50-68. https://doi.org/10.1093/biostatistics/kxy028.
8. Varadhan R, Wang SJ. Standardization for subgroup analysis in randomized controlled trials. J Biopharm Stat. 2014;24(1):154-167.
9. Pennello G, Rothmann M. Bayesian subgroup analysis with hierarchical models. In: Peace KE, Chen D-G, Menon S, eds. Biopharmaceutical Applied Statistics Symposium Volume 2: Biostatistical Analysis of Clinical Trials. Singapore: Springer; 2018.
10. Agresti A, Hartzel J. Strategies for comparing treatments on a binary response with multi-centre data. Stat Med. 2000;19:1115-1139.

How to cite this article: Heyward J, Lolic M, Spong C, Atkins D, Pennello G, Varadhan R. Assessing and communicating heterogeneity of treatment effects for patient subpopulations: Keynote and panel discussion on communicating heterogeneous treatment effects across populations. Pharmaceutical Statistics. 2021;1–14. https://doi.org/10.1002/pst.2123