Green Park Collaborative

A partnership for innovation and effectiveness

RWE Decoder Tool

A Practical Tool for Assessing Relevance and Rigor of Real World Evidence

User's Guide

The Green Park Collaborative is a major initiative of the Center for Medical Technology Policy

February 7, 2017 | Version 1.0

BETTER EVIDENCE.
BETTER DECISIONS.
BETTER HEALTH.

Center for Medical Technology Policy
World Trade Center Baltimore
401 E Pratt Street, Suite 631
Baltimore, Maryland 21202

410.547.CMTP (P)
410.547.5088 (F)
info@cmtpnet.org (E)
@cmtp_baltimore (Tw)

www.cmtpnet.org

TABLE OF CONTENTS

OVERVIEW

The purpose of the GPC RWE Decoder Framework is to help decision makers more confidently and consistently consider available real world evidence (RWE) for the purposes of informing decisions. The objective is to foster greater confidence in, and transparency of, the use of RWE in decision-making. In turn, this may also lead to greater clarity for researchers and developers of RWE regarding the particular evidence needs of decision-makers and how the RWE generated by research will be accepted and utilized.

The GPC RWE Decoder Framework is shaped by the following assumption:
The two major dimensions for assessing a study are:

1. The methodological **RIGOR**, which impacts the trustworthiness of results, and
2. **RELEVANCE**, or direct relation of research findings to decision-makers' question.

How the two domains are prioritized is context-dependent, varying by question and by decision-making context, both within an organization as well as across stakeholder groups. Importantly, however, each dimension (Rigor and Relevance) can be assessed independently of the other.

What do we mean by Rigor?

We think about rigor as the application of methodology standards to reduce potential biases and inconsistencies in measurement so that a study is more likely to measure a true result. The more rigorous a study, the more confidence a decision maker can have in the study estimates (results).

What do we mean by Relevance?

We think about relevance as the degree to which a particular study relates to the decision to be made. In other words, the degree to which a study can potentially change or influences ideas, decisions, policies, or clinical practice.[1] Relevant research is more likely to yield "meaningful conclusions".[2]

The following sections provide guidance for completing each Module of the RWE Framework, with descriptions of each Module and examples of ratings.

[1] A Dictionary of Epidemiology (5th Ed). Porta M. Oxford University Press 2014.
[2] Kahn MG, Brown JS, Chun AT, Davidson BN. Transparent reporting of data quality in distributed data networks. eGEMs 2015; 3(1): Art 7.

MODULE 1. ARTICULATING YOUR QUESTION

MODULE 1. INSTRUCTIONS

Module 1 consists of two parts: Part 1 which is recommended and Part 2 which is required. To read more about what makes a good research question see Appendix 1.

Recommended (Part 1):
In your own words, respond to Questions 1-4.

1) What is the nature of your decision?
2) What do you want to know?
3) What do patients want to know (e.g. which outcomes are most important to patients?)
    1. Fill out PICOTS table (Part 2)
4) What is your research question, rephrased and following a PICOTS format?

Required (Part 2):

DOMAIN	DOMAIN DETAILS
Population	
Intervention	
Comparison	
Primary Outcome(s)	
Secondary Outcome(s)	
Timing	
Setting	

MODULE 1. EXAMPLE RESEARCH QUESTION (PICOTS)

Among adult patients (aged 18+) with chronic asthma (P), does inhaled treatment X (I) as compared to oral treatment Y (C) lead to reduction in asthma exacerbations and asthma-related ER visits (O), measured at 6 and 12 months (T), when provided by our covered network providers in usual primary care settings (S)?

It is helpful to fill out each PICOTS element of your research question in a table. Based on the example question for the long-term control of chronic asthma, an example table has been completed, for reference (Table 1). But how did we arrive at this formatted question? The next section on Module 2A explains the rationale behind the PICOTS elements and how to consider Relevance within each.

Table 1. Example Research Question (PICOTS)

Domains of Relevance	Example Entries
Population	Adult patients aged 18 & older currently being treated for chronic asthma
Intervention	Inhaled treatment X, as prescribed and used in routine asthma care
Comparison	Oral treatment Y, as prescribed and used in routine asthma care
Primary Outcomes	Asthma exacerbations
Secondary Outcomes	Emergency room visits; Patient reported outcomes x, y, z
Timing	12 months of follow-up, with measurements at 6 and 12 months
Setting	Adult primary care settings

MODULE 2A. ASSESSING THE RELEVANCE OF RWE STUDIES

Domains of Relevance: Population – **I**ntervention – **C**omparison – **O**utcome(s) – **T**iming – **S**etting

For detailed explanations of the domains, see Appendix 2.

MODULE 2A. INSTRUCTIONS

For each available study:

1) Consider the Relevance of each domain (on the spreadsheet, the user checks a box for each domain as it is considered). If a particular domain is not applicable or not determinable, leave that box unchecked.

2) Make an informed judgment of overall Relevance to you (this judgment is subjective, due to the different priorities of each domain to different contexts, e.g. Population may be essential for some users and low priority for other). The overall Relevance is informed by an understanding of each PICOTS domain, but does not have to be an average of all domains. On a scale of [1-4], **[minimally relevant (1), somewhat relevant (2), relevant (3), maximally relevant (4)]**, record your assessment of overall Relevance. Only enter "0" if the study completely lacks relevance in a domain that is essential to your decision, thus rendering the study unusable. (Studies assigned a "0" will be excluded from the final visual display).

Exceptions: sometimes, a good research question does not require every single domain of Relevance to be defined; it depends on the question. If you cannot identify a particular domain, leave it blank!

MODULE 2A. EXAMPLES

Table 2. Examples for Relevance

Domain	Less Relevant	More Relevant
		
Population	Excludes many of the types of patients who would receive usual care in your covered population, e.g. the study population is similar to your covered population on two or more demographic characteristics OR comorbidities.	Study population is nearly identical to patients receiving usual care in your covered patient population, e.g. by age, race, sex, comorbidities and other demographic characteristics.
Intervention	The intervention is administered, as it is *intended* to be used, but is perhaps: a) strictly administered in a way does not allow for the variation of standard practice, b) administered by researchers or specially trained providers, or c) combined in some way with other elements.	The intervention is administered as it would be in usual care delivery settings, e.g. same dose and schedule as standard care within your network (or as labeled), administered by usual care provider(s), and allowed flexibility, reflecting real world conditions.
Comparison	The comparator or control is not directly relevant to your evidence needs, for example, a) does not reflect the flexibility of real world administration, b) is administered e.g. by researchers, c) is combined in some way with other elements or has small modifications from how it is delivered in usual care.	The comparator or control is another drug or intervention directly relevant to your research question, and (if applicable) is administered in a pragmatic way, identical to usual care.

Outcomes **-Primary** **-Secondary**	Primary (and key secondary) outcomes are relevant to your needs, but perhaps not directly linkable or comparable to the outcomes measured within your own population. Or, outcomes are surrogate.	Primary (and key secondary) outcomes are of obvious importance to your decision-making needs (e.g. clinical and patient-important outcomes, as measured in your own system).
Timing	The duration of follow-up is not what is preferred, but does not affect ability to answer the research question reasonably well.	The duration of follow-up is completely relevant to the decision-making context and research question at hand.
Setting	Settings in which the intervention is administered do not resemble provider settings within your own network or usual care settings for your patients, however, do not dramatically alter the ability of the study to answer your research question.	The study setting(s) in which the intervention is administered are very similar to settings within your provider network, or to settings where eligible patients would receive intervention as part of usual care.

Table 3. Potential reasons for "0" Relevance, by domain

Domain	Critically Not Relevant (0)
Population	The study population is not similar to your covered population, and comprises a narrow sample of young and otherwise healthy adults (zero comorbidities).
Intervention	Intervention does not reflect how it would be used in your population.
Comparison	The comparator is not at all relevant to your question. For example, perhaps it is a placebo, or different drug or intervention you are not interested in, or currently do not provide. Perhaps it is a highly specialized and tightly monitored care protocol. This is more problematic in comparative effectiveness research questions in particular.
Outcomes	Primary (and key secondary) outcomes are surrogate outcomes of little clinical meaning or importance to decision makers, including patients and providers.
Timing	The duration of follow-up is too short to aid your decision-making needs or otherwise not of appropriate length to answer your question.
Setting	Study settings are drastically different than any settings where the intervention would be delivered to your patients, results won't be useful to your decision-making needs.

MODULE 2B. ASSESSING THE RIGOR OF RWE STUDIES

Domains of Rigor	Subdomains
Quality of Research Question	PICOTS Stated
	Appropriateness of Study Design
	Scientific Argument
Risk of Bias	Confounding Bias
	Selection Bias
	Bias from Classification of Intervention
	Bias from Deviation from Intended Intervention
	Bias from Missing Data
	Bias from Measurement of Outcomes
	Bias in Results Reporting
Precision	Confidence Interval
Data Integrity	Data Source & Intention
	Completeness
	Fidelity
	Plausibility
	Cohort Construction (& Linkage)

For detailed explanations of the domains and subdomains, see Appendix 3.

MODULE 2B. INSTRUCTIONS

For each available study:

1) Consider the Rigor of each subdomain (on the spreadsheet, the user checks a box for each subdomain as it is considered). If a particular subdomain is not applicable or not determinable, leave that box unchecked.

2) **OPTIONAL:** For each domain (there are four), make an informed judgment of the overall Rigor for that domain. On a scale of [1-4], **[minimally rigorous (1), somewhat rigorous (2), rigorous (3), maximally rigorous (4)]**, record your assessment of overall Rigor for each domain. If a particular domain is not applicable or not determinable, you may leave that field blank. Only enter "0" if, within that domain, the study completely lacks a feature of rigor that is essential to you, thus rendering the study unusable.

3) Make an informed judgment of the overall Rigor of the study. The overall Rigor is informed by an understanding of the strength of each of the four domains, but does not have to be an average. On a scale of [1-4], **[minimally relevant (1), somewhat relevant (2), relevant (3), maximally relevant (4)]**, record your assessment of overall study Rigor. Only enter "0" if you have rated one of the four primary domains as "0".

MODULE 2B. EXAMPLES

Table 4. Examples for Rigor >> Quality of the Research Question

Subdomain	Less Rigorous	More Rigorous
		
PICOTS Stated	Incomplete or poor description of (applicable) PICOTS elements in study objectives or aims, or lacking the level of detail expected of research questions.	A clearly articulated research question defining each (applicable) element of PICOTS is stated in the objectives session, along with clear study aims.
Appropriate Study Design	Study design was selected in part for being a feasible test of the hypothesis, but there are limitations to its findings.	Study design and statistical analyses are clearly and directly appropriate for testing the hypothesis as stated.

Table 5. Examples for Rigor >> Risk of Bias[3]

Sources of Bias Specific to Non-Randomized Studies			
Subdomain	**Less Rigorous (Higher risk of bias)**		**More Rigorous (Less risk of bias)**
			
Confounding bias	At least one known important domain not appropriately measured or controlled for; OR Reliability or validity of measurement of an important domain was low enough that we expect serious residual confounding; OR Confounding inherently not controllable OR the use of negative controls strongly suggests unmeasured confounding.[4]	Confounding expected, all known important confounding domains appropriately measured and controlled for; and Reliability and validity of measurement of important domains sufficient, serious residual confounding not expected.	The study is nearly comparable to a well-performed randomized trial with regard to this domain. Investigators have conducted and reported findings from sensitivity analyses, and demonstrated convincingly that bias has a negligible effect on the results.
Selection bias	Selection into the study was related to intervention and outcome; or Start of follow up and start of intervention do not coincide, and a potentially important amount of follow-up time is missing from analyses, and the rate ratio is not constant over time; Selection into the study was strongly related to intervention and outcome; or A substantial amount f follow-up time is likely to be missing from analyses, and the rate ratio is not constant over time. If randomized: selection into trial based on a nonrandom rule or sequence, or based on clinician judgment, preference of the patient, etc.	Selection into the study may have been related to intervention and outcome, but the authors used appropriate methods to adjust for the selection bias; or start of follow up and start of intervention do not coincide for all participants, but (a) the proportion of participants for which this was the case was too low to induce important bias; (b) the authors used appropriate methods to adjust for the selection bias; or (c) the review authors are confident that the rate (hazard) ratio for the effect of intervention remains constant over time.	All participants who would have been eligible for the target trial were included in the study and start of follow up and start of intervention coincide for all subjects. If randomized: investigators document and clearly describe a random component in the generation of the allocation sequence to the study arms, for example, via a random number table, computer-based random number generator, coin toss, shuffling of cards, etc.

[3] Table 5 describes greater and lesser risks of bias found in studies (greater risk of bias = less rigor, lesser risk of bias = more rigorous). In some cases, there may be a critical known source of bias that should be addressed in responsible research. If a study fails to address a potential source of bias that is well known for a particular topic or type of study design, thus rendering it unusable for decision-making, you may wish to assign a "0" for Rigor.

[4] Descriptions of subdomains under Risk of Bias pull from the work of the Cochrane Bias Methods Group and Cochrane Non-Randomised Studies Methods Group; Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ROBINS-I: a tool for assessing Risk Of Bias In Non-randomized Studies of Interventions, Version 7 March 2016. Available from http://www.riskofbias.info [accessed {date}]; Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0, 24 September 2014. Available from http://www.riskofbias.info [accessed June 24, 2016].

Bias from classification of intervention	Intervention status is not well defined, or major aspects of the assignments of intervention status were determined in a way that could have been affected by knowledge of the outcome; (Unusual) An extremely high amount of misclassification of intervention status, e.g. because of unusually strong recall biases.	Intervention status is well defined but some aspects of the assignments of intervention status were determined retrospectively.	Intervention status is well defined and based solely on information available at the time of intervention.
Bias due to deviations from intended intervention	Switches in treatment, cointervention or problems with implementation fidelity are apparent and are not adjusted for in the analyses; substantial deviations from the intended intervention are present and are not accounted for in the analysis.	Bias due to deviation from the intended intervention is expected, and switches, co-interventions, and some problems with intervention fidelity are appropriately measured and adjusted for in the analyses. Alternatively, most (but not all) deviations from intended intervention reflect the natural course of events after initiation of intervention.	No deviation from the intended intervention is expected, e.g. if both intervention & comparator are implemented over a short time period, and subsequent interventions are part of routine medical care, or if the specified comparison relates to initiation of intervention regardless of whether it is continued.
Bias due to missing data	Proportions of missing participants differ substantially across interventions; or Reasons for missingness differ substantially across interventions; and Missing data were addressed inappropriately in the analysis; or The nature of the missing data means that the risk of bias cannot be removed through appropriate analysis; (Unusual) There were critical differences between interventions in participants with missing data that were not, or could not, be addressed through appropriate analysis.	Proportions of missing participants differ across interventions; or Reasons for missingness differ minimally across interventions; and Missing data were not addressed in the analysis.	Data were reasonably complete; or Proportions of and reasons for missing participants were similar across intervention groups; or Analyses that addressed missing data are likely to have removed any risk of bias.
Bias in measurement of outcomes	The methods of outcome assessment were not comparable across intervention groups; or The outcome measure was subjective (i.e. likely to be influenced by knowledge of the intervention received by study participants) and was assessed by outcome assessors aware of the intervention received by study participants; or Error in measuring the outcome was related to intervention status; The methods of outcome assessment were so different that they cannot reasonably be compared across intervention groups.	The methods of outcome assessment were comparable across intervention groups; and The outcome measure is only minimally influenced by knowledge of the intervention received by study participants; and Any error in measuring the outcome is only minimally related to intervention status.	The methods of outcome assessment were comparable across intervention groups; and The outcome measure was unlikely to be influenced by knowledge of the intervention received by study participants (i.e. is objective) or the outcome assessors were unaware of the intervention received by study participants; and Any error in measuring the outcome is unrelated to intervention status.

Bias in selection of reported result	Outcome measurements or analyses are internally or externally inconsistent; or there is a high risk of selective reporting from among multiple analyses; or the cohort or subgroup is selected from a larger study for analysis. Critical, there is evidence or strong suspicion of selective reporting of results, and the unreported results are likely to be substantially different from the reported results.	The outcome measurements and analyses are consistent with an a priori plan; or are clearly defined and both internally and externally consistent; and there is no indication of selection of the reported analysis from among multiple analyses; and there is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results.	There is clear evidence (usually through examination of a preregistered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and subcohorts.

Sources of Bias Specific to Randomized Studies[5]

Subdomain	Less Rigorous (Higher risk of bias)	More Rigorous (Less risk of bias)
		
Selection bias	See above	
Reporting bias	See above	
Performance bias	• No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.	• No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding; • Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.

[5] Descriptions of Risk of Bias for randomized studies from: Assessing Risk of Bias in Included Studies. Cochrane Handbook, Chapter 8. Available online at: [http://methods.cochrane.org/bias/assessing-risk-bias-included-studies]

Detection bias	• Not all of the study's pre-specified primary outcomes have been reported; • One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; • One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); • One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; • The study report fails to include results for a key outcome that would be expected to have been reported for such a study.	• No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding; • Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.
Attrition bias	• Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size; • 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization; • Potentially inappropriate application of simple imputation.	• No missing outcome data; • Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); • Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size; • Missing data have been imputed using appropriate methods.

Table 6. Examples for Rigor >> Precision

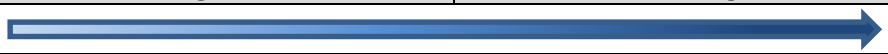
Subdomain	Less Rigorous	More Rigorous
		
Confidence Interval	Wide interval or limited meaningfulness of estimate	Narrow and meaningful

Table 7. Examples for Rigor >> Data Integrity

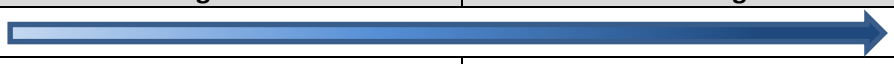
Subdomain	Less Rigorous	More Rigorous
		
Data Source & Intention	High potential for miscoded data or data entry mistakes, intention of data collection is clearly misaligned with goals of study, questionable choice of database	Well documented, defined and cleaned (e.g. clear codebook and data phenotypes, data entry and cleaning processes described), intention of data collection closely aligned with goals of study
Completeness	High rate of missingness in variables that are important for analysis	Low or no missingness in dataset or in the variables included in analysis
Fidelity	High risk of data errors and/or false values	Data reliability and validity tested and well demonstrated
Plausibility	The data paint a highly unbelievable story based on existing evidence of disease rates, treatment use, existing trends or other known information	The data paint a believable story based on existing knowledge of the sample and target populations, and which aligns with a sound scientific argument.
Cohort Construction	Poor reporting of linkage, unaccounted loss of records, etc.	Records linked using unique identifier or well-validated algorithm

Table 8. Examples of "0" for Rigor, By Domain

Domain	Critically Not Relevant (0)
Quality of Research Question	The study population is not similar to your covered population, and comprises a narrow sample of young and otherwise healthy adults (zero comorbidities).
Risk of Bias	Critical risk of bias observed from one or more known sources and failure to address through study design or analytic plan.
Precision	Very wide confidence interval such that result lacks meaning for decision making
Data Integrity	High potential for miscoded data or data entry mistakes; High risk of data errors and/or false values; Poor reporting of linkage, unaccounted loss of records, etc.

MODULE 2C. MAGNITUDE AND DIRECTION OF EFFECT

Effect sizes can be quite large or quite small, however, whether or not a small difference carries any meaning depends on your situation and the environmental context in which you will decide to make any changes.

MODULE 2C. INSTRUCTIONS

For each available study:

1) Consider the findings for the study's primary outcome, or whichever specific outcome for which you are primarily interested in the study.

2) One a scale of [1-3], **[not very meaningful (1), meaningful (2), very meaningful (3)]** record your assessment of the meaningfulness of the effect size for each study.[6]

3) Indicate the direction of the estimate **[negative (1), positive (2)].** Select positive if there is a greater occurrence of the outcome in the intervention group compared to the control, select negative if there is a lesser occurrence, or reduction of the outcome in the intervention group. If study results indicate *no difference* in effect, select negative (1) for direction.[7]

4) To highlight study titles for which results suggest no difference in effect, select "Yes" in the far right column next to each study that observed no difference.

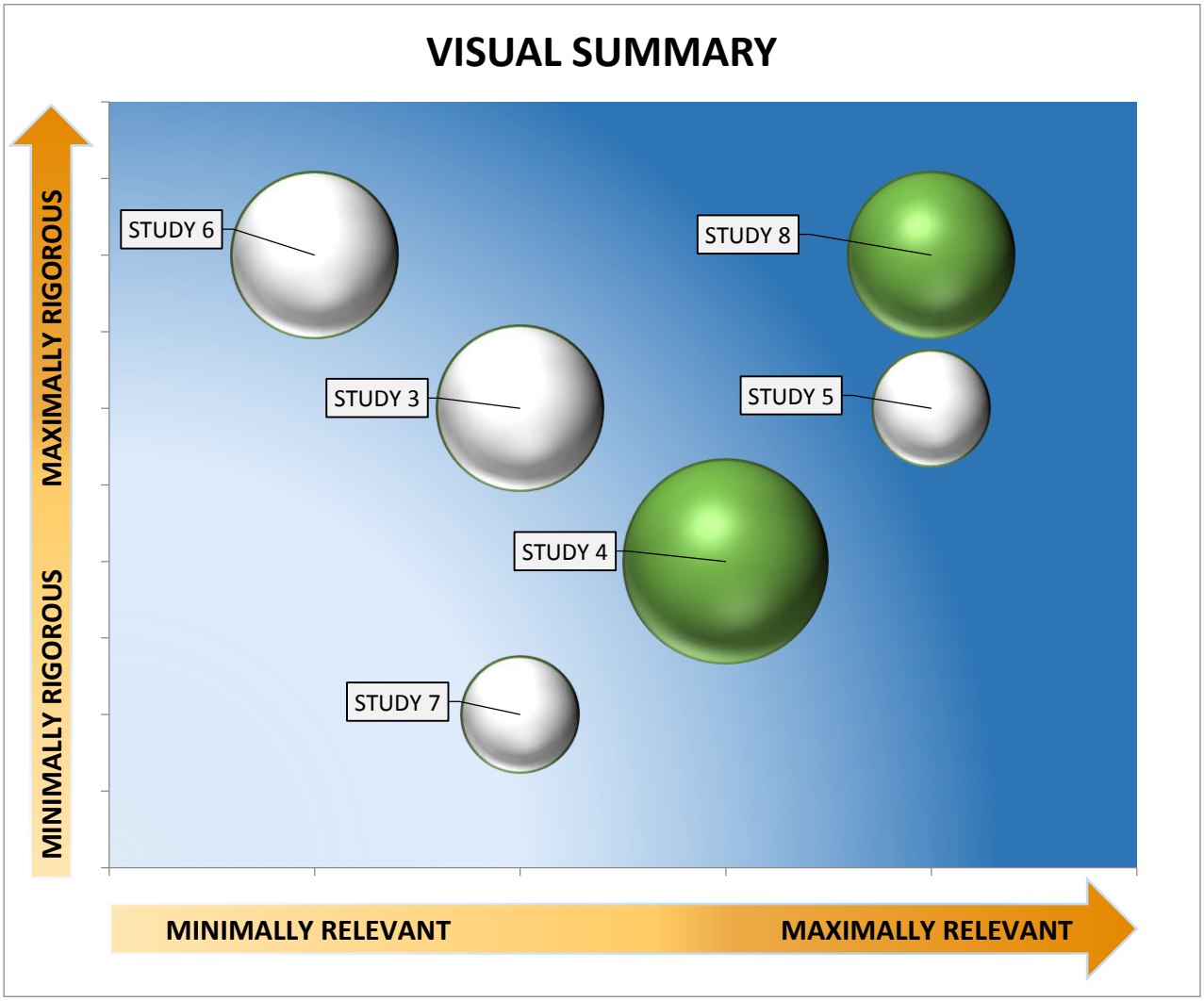
[6] The meaningfulness of an effect size may depend on the size of the study population, or on the size of the target population for which you are making a decision. For example, a small but statistically significant effect size observed in a very large study population might be quite meaningful to your decision-making considerations. And small differences applied to a large patient population might generate measurable change at the population level. Therefore, we recommend thinking about the study population size, the total population to which your decision may apply, and what may constitute a meaningful range of difference in effect to you, before completing Module 2C.

[7] Please Note: This version of the GPC RWE Decoder does not differentiate between negative direction and "no difference" of effect. Future iterations with improved functionality will have the ability to make this distinction.

MODULE 3. VISUAL SUMMARY OF RESULTS

The figure below is a hypothetical output of Module 3. Having completed Modules 2A, 2B, and 2C, the RWE Decoder Tool (Version 1.0) will automatically generate a visual summary of your results. The size (area) of the bubble corresponds to the meaningfulness of the Effect Size. The color of the bubble corresponds to the direction of the estimate (green = positive, white = negative or no difference). Values for **RELEVANCE** fall along the x-axis, and values for **RIGOR** fall along the y-axis. Any study assigned a "0" for one or more domains is automatically excluded from the graph.

Studies plotted in the uppermost right-hand corner are highly rigorous and highly relevant. Most studies will likely fall in the middle. However, depending on the decision making needs of you and your colleagues, how you utilize studies of low rigor or relevance are dependent on your priorities and the impact of the decision you face.



APPENDIX 1: DEFINING A GOOD RESEARCH QUESTION

To help you complete Module 1 we first discuss how you might articulate a question of importance to your decision-making needs and translate that into a structured research question that is recognizable to researchers and other health care stakeholders alike.

WHAT IS A GOOD RESEARCH QUESTION?

A research question is an "answerable inquiry"[8] into a particular issue or topic area. In health care research, including real world research, a good research question needs to be as clear and precise as possible. The appropriate level of detail is not always intuitive to those of us who do not spend their time designing research studies. What can occur is a gap in communication between the stakeholders who have important questions or decisions to make, and the researchers who need to design a study to be able to answer a particular question.

As a patient, you might ask, "which blood pressure drug is the best option for me?" As a researcher this question would become a recipe based on your health care setting, your personal and medical characteristics, your provider's characteristics. For example, if you are a Hispanic woman who has just been diagnosed with high blood pressure, non-smoking, and 55 years old with a family history of [x, y, or z], your question may need to be translated into a population-level question that can be answered for you and other patients who are similar to you. You (the patient) may also be more interested in particular outcomes or side effects than your doctor, or health insurance provider, who are likely more interested in specific clinical outcomes of disease progression or outcomes of resource utilization. Your priorities and the outcomes that matter may also shape a question. Your question of "which blood pressure drug is the best option for me" in a conversation between you and your doctor, might, in a different conversation between you and a researcher, translate into something more structured and specific, such as*:*

> *"Among Hispanic women between the ages of 45-60 with a family history of [x, y, or z], who are recently diagnosed and have never previously been treated for high blood pressure, is [Drug X] more effective than [Drug Y] at reducing [Outcomes]?*"

This "translated" version of the question is structured around a framework called PICOTS – Population, Intervention, Comparator, Outcomes, Timing, and Setting.

[8] Lai WW, Vetter VL, Richmond M, et a. Clinical research careers: reports from a NLHBI Pediatric Heart Network clinical research skills development conference. AHJ 2011; 161(1): 13-67.

APPENDIX 2. DOMAINS OF RELEVANCE EXPLAINED

1. Population

There are many different ways to define a population. The **SAMPLE POPULATION** is the set of individuals enrolled in a particular research study or included in a particular analysis and can be described by their collective characteristics. This set of individuals might exclude particular subgroups of people, depending on the criteria which was applied to allow or restrict people to participate in a particular study. For example, a study of asthma treatment may be limited to non-smokers, or to patients under a certain age, or to those who are not also on medicine for another major condition such as hypertension.

Importantly, the **TARGET POPULATION** is the broader population of patients or individuals for which the question is being asked, or who will be influenced by decisions based on this research. It is important to articulate your target population in Module 1 based on the scope and reach of your decision and who will be impacted. For example, your target population might be adult asthma patients within your regional health system. What does that population look like? What is their age range, sex, race? What proportion of them are high adherers or consistent users of the health services you provide? How many of them also suffer from other chronic diseases? The answers to all these questions make up the characteristics of your target population. Your target population may differ in characteristics from a particular study's sample population. These characteristics may also differ from the national average. And likely, they also differ from another major health system in another region of the country. Once you understand your own target population, it becomes possible to assess how closely a sample population resembles that target population. Most likely you won't find an exact match unless you simply enroll everyone from your system into your own research study. More likely, you will understand what are particular characteristics that are important to you and that would cause you to identify a study as more or less relevant to your needs.

2. Intervention(s)

In research, the **INTERVENTION** is the treatment or element of care that is being provided to patients, and which you want to study. The intervention could be anything from a drug (e.g. a blood pressure medicine) or medical device, to a particular surgery, to a lifestyle change (e.g. enrolling in a program to quit smoking or following a work out plan), to a disease management strategy; or it could be something else. In some research studies, the intervention is intentionally changed or manipulated and then its effect studied over time. In other studies, patients who received or were exposed to the intervention for other reasons, for example, in their usual course of health care, are identified and their data observed in order to study the effect of the intervention. However, in order to understand the relevance of the intervention in a study to the intervention for which you need to make a decision (e.g. prescribing to your patients, or covering on a drug formulary), the details can become quite important. Let's take the asthma example. As a provider or other decision maker, you may also want to study the particular way in which the intervention and comparison drugs are taken, such as whether they're taken at morning or at night, or when taken in combination with other treatments. The more of these details you understand about the question you need to ask on the intervention, the better able you will be to assess whether existing studies resemble your situation. Another way to think about intervention and its relevance to you, is to ask yourself: would I take the intervention *as is* and provide it to our patients, or would I make modifications first?

3. Comparator

The **COMPARATOR** is the benchmark against which the intervention is compared. It is any alternate treatment that might be taken instead of the intervention treatment; it might be the existing standard of care against which you would like to test a *new* treatment to see whether one out performs the other. Some comparators are easy to define, such as placebo. However, in the case of real world evidence, comparators can often be

"usual care" – a term that encompasses any number of standard treatments or practices. Depending on your question, usual care may be as basic as a competitor drug. It could be more complicated, like a complicated surgical procedure that is often full of unexpected curveballs. It could also be a care pathway involving multiple clinicians – nurses, specialty physicians, medical assistants – who are all supposed to follow particular steps, but in reality, juggling serious time restraints and conflicting priorities, might deviate from the protocol in seemingly small but inconsistent ways. Sometimes more than one comparator might be scientifically appropriate (there are more than two ways to treat high cholesterol, hypertension, cancer, diabetes, etc., there are multiple antibiotics to treat infections, there are multiple drugs and non-traditional practices to treat pain). Sometimes, the comparator is a treatment or approach used in your health system that you might ultimately *replace with* the intervention, if enough evidence demonstrates that the intervention is more effective (or *as* effective, but perhaps with less side effects or lower costs). When using this framework, consider what is standard practice or considered the "norm" in your own setting, among your own patients or colleagues.

4. Outcomes

Where do you want to see a difference or a change? The **OUTCOME** is that measurable thing upon which the intervention is supposed to have an effect. For example, the outcome of studying the effect of chronic asthma treatment is asthma exacerbations (or "flare-ups"), the outcome of studying the effect of blood pressure medicine is blood pressure. There can be multiple outcomes, of varying importance depending on who is asking the question. As another example, say you are a patient starting a diet and exercise regimen; there are several different ways to define your outcome: weight (measured in pounds), size (measured in inches), energy (self-reported), strength (self-reported or measured), or perhaps even improvement in a clinical measure such as blood pressure, cholesterol, etc. The relative importance of outcomes can also vary between decision-making groups. For example, in the case of asthma, a provider may want to see less emergency room visits, a patient may simply want to be able to exercise regularly without difficulty breathing. The outcome is one aspect of a study that reminds us that relevance is in the eye of the beholder. As a decision maker, you need to ask yourself, what outcome(s) are most important to you and your organization, and also, what outcomes are important to patients and their doctors involved in care decisions? Once you are able to answer that question, you know what to look for when assessing relevance in real world studies.

The **PRIMARY OUTCOME** is the main outcome of the study, and for which the study and statistical analysis is designed to best capture. **SECONDARY OUTCOMES** are additional outcomes that are captured in the study. They are included because they are also important, perhaps to patients or other stakeholders, but are not intended to be the primary finding. Whether or not findings from the secondary outcomes are as statistically robust as findings from the primary outcomes depend on the study design and analytic plan.

5. Timing

Sometimes an intervention can create an effect almost immediately; other times it can take weeks, months or years. For example, antibiotics usually work in a matter of days, a new diet and exercise regimen may take several weeks before you start to see a measurable difference, and it may take several months on a new asthma treatment regimen to finally get symptoms under control. When thinking about your question, and what types of research will be relevant, think about **TIMING** as the necessary duration of follow-up in a study (how long after an intervention the study continues to collect data from patients) in order to capture *meaningful* differences in your outcome.

Think about the relevance of timing another way: a study assessing the effectiveness of a 5-day course of antibiotics shouldn't require months of follow up data collection from patients – it is unnecessary, because the effect of the drug will already have taken place. Alternately, a study assessing the effectiveness of a cancer

treatment needs more than just a few weeks of follow-up data collection, because otherwise the timing won't be long enough since it takes several months to notice a change in tumor growth.

6. Setting

The **SETTING** is where a research study takes place. For instance, a study may be conducted at an academic hospital in a very urban area, or a community hospital, or perhaps at primary care doctor's offices that are part of a particular health care system or network in a defined geographic area. Setting is an important domain, particularly when assessing real world evidence, because our health care system is so diverse and complicated. The way care is provided or funded in one hospital or clinic can be very different than another. And if you're a health care provider wondering whether or not you should adopt a new treatment or program for patients in your system, you ideally want to see real world evidence from systems that resemble your own.

APPENDIX 3. DOMAINS OF RIGOR EXPLAINED

1. Quality of the Research Question

a. Picots Stated in Research Question or Study Aims

As mentioned previously, a **RESEARCH QUESTION** is an "answerable inquiry". From another point of view, it is the question that a study sets out to answer. (It is also possible for a study to be designed to answer more than one research question.) When assessing the rigor of a study, it is helpful to look for a clear account of the research question by the study designers, before the study has been implemented. A clearly stated research question with elements of PICOTS well-articulated and preceding discussions of the study design are encouraging. Decision makers may also have an added boost in confidence of the quality of the research question if relevant stakeholder communities played a role in shaping and refining the research question, and there is a clear account of how their feedback was incorporated.

When assessing the analytic design, consider the study hypothesis. A hypothesis is a statement arising from a research question, translated in such a way that research can either support or contradict. A hypothesis can take the form of: 1) a supposition, arrived at from observation or reflection that leads to refutable predictions, or 2) any conjecture cast in a form that will allow it to be tested and refuted.[9] The key idea here is that a hypothesis is written in such a way that it can be tested, so that, through a well-designed study, evidence can be generated to either support or refute a hypothesis. Therefore, the analytic plan should be developed so that it is appropriately able to test the conditions articulated by the hypothesis.

b. Appropriateness of Study and Analytic Design

An appropriate **STUDY DESIGN** is one developed to, within reason and feasibility, address the unique needs laid out by the research question. The analytic design as the architecture of the study: the structure, specific details of the population, time frame, methods and procedures involved, and ethical considerations.[10] Look for descriptions of how study investigate believe that the study design they have chosen is appropriate to generate the evidence that will answer the research question they have articulated. Again, any additional stakeholder engagement to shape or develop the study protocol adds additional confidence that the analytic design was developed with the intent of generating the most meaningful results. Another benefit of a clear and well-thought out analytic design is that it can be replicated by other researchers. On the path of knowledge generation, replicating results by different teams and in different settings leads to greater strength and confidence that the evidence reflects a true effect.

2. Risk of Bias

BIAS is a deviation from the truth.[11] In research, bias is a systematic deviation of the results from the truth. Systematic is the opposite of random. Something creating a bias in a study can influence the value of the results. Bias in research isn't necessarily intentional. In most cases, it is usually unintentional, and in some cases unanticipated. Before assessing the risk of bias, ask yourself: Are there any systematic features in the study implementation or known factors that could be influencing the results so that they are inaccurate?

Ideally, a study result, or estimate of an effect of an intervention or treatment, should isolate and represent changes in the outcome due only to that specific intervention. Assessing risk of bias means assessing the risk

[9] A Dictionary of Epidemiology (5th Ed). Porta M. Oxford University Press 2014.
[10] A Dictionary of Epidemiology (5th Ed). Porta M. Oxford University Press 2014.
[11] Bias in the RWE Decoder does NOT refer to the researcher's conclusion being influenced by political or ideological ideas or personal economic incentives.

that the study findings are unduly influenced by factors other than the intervention and have not been appropriately accounted for in the final estimates. Demonstrating knowledge of potential sources of bias and providing a thorough discussion of design or analytic decisions to address these potential sources of bias is how the rigor of a study can be improved. Bias typically arises as some consequence of a design flaw or analytic aspect of a study, and can occur from several different sources. The subdomains, or factors to consider within this domain are common potential sources of bias. They are listed and described in Table 9. These sources of bias have been well documented and articulated by the Cochrane Collaboration.[12] While users of the GPC RWE Decoder may have varied or limited familiarity with different sources of bias, lessons can be pulled from other initiatives as well. Recent validation work from the GRACE initiative, for example, has found that the single most predictive hallmark of methodological quality, or rigor, in observational comparative effectiveness research, is whether the investigators have applied sensitivity analyses to their findings.[13] (Sensitivity analyses examine the extent to which results are affected by changes in assumptions, in methods, models, or values in unmeasured confounders. Results of sensitivity analyses can help describe the extent to which bias has an impact on study results.[14])

An important caveat is that bias is extremely common. Even in the most well-designed studies with highly sophisticated analytic plans, there can still be potential risk of bias. As you assess consider the risk of bias within your overall assessment of methodological rigor, look for indications that potential sources of bias have been considered or discussed by investigators, and contingencies have been built into their study designs or analytical plans to attempt to account for those biases.

Table 9. Sources of Bias and their Meaning

Sources of Bias Specific to Randomized Trials	
Source of Bias	**Meaning**
Confounding	Confounding, or confounding bias, is bias of the estimated effect of an exposure (or intervention) on an outcome, due to the presence of a common cause of the exposure (or intervention) and outcome.[15] At least one known important domain is not appropriately accounted for; OR confounding inherently not controllable OR the use of a negative control group (e.g. a group in which no response to the intervention is expected) strongly suggests unmeasured confounding.
Selection Bias	Selection into the study was related to intervention and outcome; or Start of follow up and start of intervention do not coincide, and a potentially important amount of follow-up time is missing from analyses, and the rate ratio is not constant over time; Selection into the study was strongly related to intervention and outcome; or A substantial amount of follow-up time is likely to be missing from analyses, and the rate ratio is not constant over time. [16]

[12] Sterne JAC, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016; 355:i4919.

[13] Dryer NA, Bryant A, Velentgas P. The GRACE Checklist: a validated assessment tool for high quality observational studies of comparative effectiveness. J Manag Care Spec Pharm. 2016;22(10):1107-13.

[14] A Dictionary of Epidemiology (5th Ed). Porta M. Oxford University Press 2014.

[15] A Dictionary of Epidemiology (5th Ed). Porta M. Oxford University Press 2014.

[16] Descriptions of subdomains under Risk of Bias pull from the work of the Cochrane Bias Methods Group and Cochrane Non-Randomised Studies Methods Group; Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ROBINS-I: a tool for assessing Risk Of Bias In Non-randomized Studies of Interventions, Version 7 March 2016. Available from http://www.riskofbias.info [accessed {date}]; Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0, 24 September 2014. Available from http://www.riskofbias.info [accessed June 24, 2016].

Classification of Intervention	Intervention status is not well defined, or major aspects of the assignments of intervention status were determined in a way that could have been affected by knowledge of the outcome; (Unusual) An extremely high amount of misclassification of intervention status, e.g. because of unusually strong recall biases.
Deviations from Intended Intervention	Bias due to substantial deviations from the intended intervention, such as: switches in treatment, co-intervention or implementation of the intervention. In real-world studies, sometimes deviations from intended intervention are a consequence of the intervention. For example, in a comparison of a once-daily oral medication vs. a medication that is injected multiple times per day, researchers might observe higher discontinuation, lower adherence, and more errors with the injected medication. This source of bias should be considered within the context of the research objective, the characteristics of the intervention, whether the primary analysis is as treated or intent-to-treat, etc.
Missing Data	Proportions of missing participants differ substantially across interventions; or Reasons for missingness differ substantially across interventions; and Missing data were addressed in appropriately in the analysis; or The nature of the missing data means that the risk of bias cannot be removed through appropriate analysis; (Unusual) There were critical differences between interventions in participants with missing data that were not, or could not, be addressed through appropriate analysis.
Measurement of Outcomes	The methods of outcome assessment were not comparable across intervention groups; or The outcome measure was subjective (i.e. likely to be influenced by knowledge of the intervention received by study participants) and was assessed by outcome assessors aware of the intervention received by study participants; or Error in measuring the outcome was related to intervention status; The methods of outcome assessment were so different that they cannot reasonably be compared across intervention groups.
Selection of Reported Result	Outcome measurements or analyses are internally or externally inconsistent; or there is a high risk of selective reporting from among multiple analyses; or the cohort or subgroup is selected from a larger study for analysis. Critical, there is evidence or strong suspicion of selective reporting of results, and the unreported results are likely to be substantially different from the reported results.
Sources of Bias Specific to Randomized Trials[17]	
Selection Bias	(See above) In addition, the concern in a randomized trial is that allocation of participants to the intervention arms of the study are not randomly determined, but instead determined through some other nonrandom means, or that randomization was inadequate and that there is some systematic difference between the groups in each intervention arm (or between control and intervention). Alternately, if the allocation of a participant to a study arm (e.g. intervention or control) is not adequately concealed, such that patients or doctors can see which group a participant will be assigned to, that may also bias the selection of participants into the trial.
Selection of Reported Result	(See above)

[17] Descriptions of Risk of Bias for randomized studies from: Assessing Risk of Bias in Included Studies. Cochrane Handbook, Chapter 8. Available online at: [http://methods.cochrane.org/bias/assessing-risk-bias-included-studies]

Performance Bias	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest. After enrolment into the study, blinding (or masking) of study participants and personnel may reduce the risk that knowledge of which intervention was received, rather than the intervention itself, affects outcomes. Effective blinding can also ensure that the compared groups receive a similar amount of attention, ancillary treatment and diagnostic investigations. Blinding is not always possible, however. For example, it is usually impossible to blind people to whether or not major surgery has been undertaken.
Detection Bias	Detection bias refers to systematic differences between groups in how outcomes are determined. Blinding (or masking) of outcome assessors may reduce the risk that knowledge of which intervention was received, rather than the intervention itself, affects outcome measurement. Blinding of outcome assessors can be especially important for assessment of subjective outcomes, such as degree of postoperative pain.
Attrition Bias (similar to bias from missing data, above)	Attrition bias refers to systematic differences between groups in withdrawals from a study. Withdrawals from the study lead to incomplete outcome data. There are two reasons for withdrawals or incomplete outcome data in clinical trials. Exclusions refer to situations in which some participants are omitted from reports of analyses, despite outcome data being available to the trialists. Attrition refers to situations in which outcome data are not available.
Other biases	Depending on the design of the randomized trial, (e.g. in cluster-randomized trials potential risk of recruitment bias, in cross-over trials potential risk of carry-over). There may also be special circumstances depending on the nature of the intervention or setting, for example, where contamination (mixing of intervention and control) presents a risk of bias.

3. Precision

PRECISION is the minimization of random error, or additional "noise" surrounding an estimate. A precise answer, or research finding, is for which the estimation error bounds are relatively small. It should be noted, however, that precision does not assume or indicate accuracy. You may have a highly biased (inaccurate) but highly precise finding. As a metaphor, think about two players throwing darts at a dartboard, and neither player is accurate. One player is somewhat inconsistent, but manages to at least land all their darts within the first couple rings of the board. The darts are somewhat scattered and not very precise. Another player may throw very consistently every dart at a single spot in the right corner of the board. While his darts were very precise, his game is still not accurate, since he did not hit the bullseye.

Depending on the research question or the needs of decision makers, sometimes a highly precise number that is somewhat biased is still acceptable. That is a point worth discussing amongst colleagues when assessing RWE, which can often, given its nature to observe the "messy" real world, produce noisier, less precise findings. When assessing the precision of a research finding, ask yourself: How clear-cut is the observed estimate? What is the observed variability (do the researchers report standard deviation)? How narrow is the window in which outcomes are observed? Is the confidence interval narrow or wide?

4. Data Integrity

DATA INTEGRITY can be understood to include the extent to which data sources are understood and can be trusted for use in research, the extent to which constructed datasets can be reproduced, and the processes or safeguards used to clean data and reduce errors.

For example, if the data source is an administrative claims database, and part of your research question involved patient adherence to a drug treatment regimen, the intention of the data that was collected is still somewhat aligned with your research question, though research was not the original intention of the data collectors.

Table 10. Subdomains of Data Integrity and Their Meanings

Subdomain	Meaning
Data Source & Intention	Origination of Data, Intention of Original Dataset (its purpose and intended uses), also any information on how data were entered or coded.
Completeness	The absence of missing data. This by itself is distinct from bias due to missing data… missing data may be random or systematic. This domain relates only to the degree to which the data are complete.
Fidelity	Correctness, or whether the values accurately describe the truth. For example, is a subject coded as female actually female, does the calculated age make sense for the corresponding year of birth?
Plausibility	Believability of the data. For example, are rates within the sample population so extremely different from existing knowledge of the target population as to warrant some healthy skepticism?
Cohort Construction	The processes applied to original data in order to generate the final dataset used for the study. For example, do data go through certain extraction procedures to reach outside users? This subdomain is perhaps more applicable when multiple sources of data are linked to form a more comprehensive dataset for research. How fields are linked, or different definitions of an outcome are reconciled can mask some of the original detail or original meaning of some variables, which important to consider when assessing RWE.