

# Mistakes in epidemiological study design in gastroenterology and how to avoid them

Tomer Adar<sup>†</sup> and Kate M Fleming

Conducting primary research and creating new knowledge is how we push the boundaries of science, how we assess new and/or alternative treatments, and ultimately what underpins decisions on who to treat, when, and how to treat them. It is fundamental to the concept of evidence-based medicine (EBM). Many people working within the field of gastroenterology will be directly involved in conducting epidemiological research. In this article, we discuss some of the common pitfalls that may render such research less useful than hoped. This is, of course, a brief overview and if we were able to stress one thing from our experience of conducting research it would be to ensure that you have the right mix of people and skills in place from the outset of your project. Collaborative research allows people with different backgrounds and knowledge bases to address a single problem from different angles, and ultimately makes for better, and more enjoyable, research.



## Mistake 1 Asking an irrelevant or unoriginal research question

What makes us want to conduct a piece of research? Have we seen something unusual, or all too common, in the clinic? Is there something that is a major issue of clinical or public health importance? Has there been particular media interest in a specific disease and/or treatment? Is there a call for research funding into a specific topic?

When something piques our interest, it is crucial to find what other research on the topic has already been done and what is underway. This knowledge helps to determine whether or not we should pursue our interest. We may wish to apply the 'FINER' criteria to our work—is our study feasible (see mistake 8), interesting, novel, ethical (see mistake 9) and relevant?<sup>1</sup>

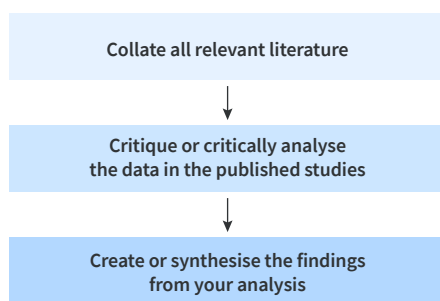


Figure 1 | Researching your research question.

If we begin with the assumption that our initial thought is highly unlikely to be novel, then the logical first step is to find out what has already been done on the topic, what the results were, and in what setting or populations was it performed. Doing such due diligence will generate ideas regarding the appropriate methods of enquiry and potentially identify limitations in previous research that could be improved on. Essentially, a thorough search of the existing literature and knowledge of ongoing research in the field of interest allows a broad area of interest to be refined into a specific and well-formulated question (see mistake 2).<sup>2</sup>

Frequently, people assume there are only two steps to a literature review—gathering information and then synthesising it—but there is more to it than that (figure 1). Crucial to a good scientific literature review (systematic or otherwise) is to critically analyse the literature before synthesising it.<sup>3</sup> Only by critiquing the literature can you assess its validity and context in terms of your specific research question. This is increasingly important in the world of science and medicine, characterised as it is by information overload and the proliferation of journals, some of which have questionable peer-review processes. By identifying gaps and weaknesses in the research that has gone before, you are more likely to be able to refine your own study question into a relevant and answerable hypothesis.

It is not always necessary to perform a systematic review, although doing so, particularly if you are embarking on a larger programme of primary research, can be extremely beneficial and potentially justify publication in its own right.<sup>4</sup> Remember that not all research is published in paper form, so knowing who is doing what (perhaps from conference attendance) means that you can find out more information and also see what the current trends are in your research field.<sup>5</sup>

## Mistake 2 Failing to frame a clear research question

Spending a great deal of work, time and resources on a research project before the question of data analysis is raised (as may commonly occur during abstract submission deadline season) is extremely detrimental to its chances of success. By not prioritising the formulation of the precise research question, it can become obvious that the data collection was incomplete, or that the question cannot be answered without the need for additional work that may or may not be feasible (e.g. the patients filling a questionnaire, etc.) (figure 2).

Framing a clear, well-defined research question is key to a successful research project. Such a question should be generated in the early stages of the research project and provide the rationale for most of the decisions subsequently made during the study design process.

© UEG 2020 Adar and Fleming.

Cite this article as: Adar T and Fleming KM. Mistakes in epidemiological study design in gastroenterology and how to avoid them. *UEG Education* 2020; 20: 20–24.

<sup>†</sup>Tomer Adar (deceased July 30, 2018) was an Advanced Fellow in IBD at Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA.

Kate Fleming is a Senior Lecturer in Social Epidemiology at the Institute of Population Health, University of Liverpool, Liverpool, UK.

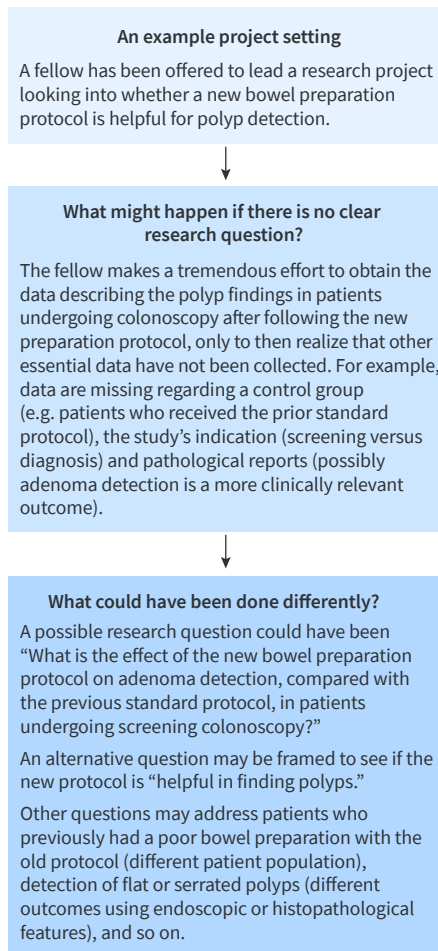
**Acknowledgements:** This article was written as a collaboration between a clinical gastroenterologist and a nonclinical epidemiologist, with grateful thanks for the opportunity to learn, to debate and to write together. It is dedicated to the memory of Tomer Adar.

Illustrations: J.Shadwell

Correspondence to: kate.fleming@liverpool.ac.uk

**Conflicts of interest:** The authors have no conflicts of interest to declare.

Published online: October 2, 2020.



**Figure 2 |** The importance of having a clear research question.

Clear research questions in epidemiological study design are often framed using a PICO format, with a clearly defined patient population, intervention (or exposure), comparison and outcome.<sup>6</sup> Only once a research question has been framed, can a clear plan for the study design, data collection and analysis be completed (further information can be found in an article we wrote examining this 'mistake' from a Young GI angle<sup>6</sup>).

### Mistake 3 Presuming only a randomized controlled trial will provide the answer to your question

The prevailing school of thought in the teaching and practice of medicine across Europe is that medicine should be evidence based.<sup>7,8</sup> Indeed, this seems rather intuitive—who would not want to be treating their patients, or treated as a patient, on the basis of the best available knowledge at the time?

Ubiquitous within the teaching and scholarship of evidence based-medicine (EBM) is the evidence hierarchy.<sup>9</sup> Most people will be familiar with this as a pyramidal representation of different study designs, with the 'lowest' form of evidence provided by case reports and case

series at the base, moving up through various epidemiological study designs, case-control studies, cohort studies, and randomized controlled trials (RCTs), with the top of the pyramid culminating in systematic reviews and meta-analyses. Following our literature review and research question refinement, we determine we need to conduct primary research (thus removing the option of a systematic review or meta-analysis) and are left with RCTs as the 'gold standard' in research.

Trying to push our research into the highest tier of the EBM pyramid on no other basis than believing it will therefore provide 'better quality evidence' is a foolish way to embark on research.<sup>10</sup> Indeed, we recommend that you examine the question and choose the appropriate study design. Each epidemiological study design has its merits and its weaknesses, and each is better suited to answering certain types of question.<sup>11</sup> In particular, recent advances in causal inference methodology have removed some of the previous postulated weaknesses of observational studies not being able to answer causal questions.<sup>12</sup> Arguably, a well-conducted causal observational study will provide better evidence of a real world application than an RCT, conducted as they usually are on a highly sampled population (see mistake 7).

In addition, there are several pertinent research questions that can only be answered by eliciting patient or physician preferences and viewpoints. Although this article is considering quantitative epidemiological study designs only, we cannot stress strongly enough that the findings of a well-conducted qualitative study are just as important in adding to our evidence base if the question being asked is best answered by a study of this nature. As EBM pioneer David Sackett so eloquently said more than 20 years ago, "Each method should flourish because each has features that overcome the limitations of others when confronted with questions they cannot reliably answer."<sup>13</sup> Precisely!

### Mistake 4 Not working with clearly defined parameters

Defining key data items such as primary exposure or intervention, primary outcome and covariates is clearly important for the data collection and analysis process.<sup>14</sup> If the parameters are not clearly defined nonuniform and/or unsystematic data collection results, and the data generated will be uninformative, or worse, may lead to incorrect inference (bluntly speaking: garbage in = garbage out). This point is especially important if multiple individuals are to be involved in the data collection process.

While this may be less of an issue if the data item is an objective measurement (e.g. CRP levels), even if a time frame is defined (e.g. CRP levels after 3 months of treatment), the situation becomes

more challenging when more objective findings need to be subjectively evaluated (e.g. severity of mucosal inflammation demonstrated by colonoscopy), and especially when the entire evaluation is based on subjective information (e.g. wellbeing or quality of life [QoL]). The utility of subjective measurements should not be discounted however. There is a clear shift to including patient reported outcome measures (PROMs) and patient reported experience measures (PREMs) alongside 'hard' endpoints to allow full assessment of healthcare quality.<sup>15</sup>

Using previously validated scores and questionnaires (e.g. from the International Household Survey Network (IHSN) Question Bank [<https://qbank.ihsn.org/>]) is a way to ensure the scientific value of any data collected and goes some way to meet reviewers' expectations for publication. Reviewing the relevant literature and consulting an expert in the field as part of the protocol preparation process may help to identify appropriate instruments and scoring systems. Of note, some may require permission for use. Using standardized instrumentation is also beneficial in other ways: a clear methods section becomes easier to write, and the possibility that the work can be reproduced and replicated is increased.<sup>16,17</sup>

### Mistake 5 Failing to account for multiple comparisons

In the age of electronic records and large datasets, it is important to keep the focus on the predefined research question and to be careful with any additional analysis that may be performed. One of the pitfalls to avoid is blindly running statistical tests across all parameters in a large dataset, hoping to stumble across a significant finding (again, this is more likely to occur when time pressured and is sometimes referred to as 'fishing' or 'p-hacking').<sup>17</sup>

The reason why running numerous statistical tests can be a mistake is both logical and mathematical.<sup>18</sup> The basis for many statistical hypothesis tests is evaluating the chance of obtaining the observed results compared with what would have been expected (e.g. based on the general population or the study's control group), with a commonly used threshold of 5% for stating that the observed data are significantly different to the expected (this is viewed by many as a mistake in itself,<sup>19</sup> but further discussion is beyond the scope of this article). In other words, there is a 5% chance of getting results that appear significantly different from the expected ones, even though the reality is that there is no significant difference (this is a type-1 error). The more statistical tests we run, the more these 5% start to add up, and, therefore, when multiple comparisons are performed the threshold for determining significance is made more conservative. The Bonferroni correction is often

used in this regard, though other methods also exist.<sup>20</sup>

Of note, certain projects have more than one outcome. In these cases, the primary outcome and any secondary outcomes should be clearly predefined, and the analysis should be performed accordingly. Finally, important discoveries can still be made outside those predicted in the analysis plan or even by the original question. In these cases, it is important to realize whether the findings are indeed a true signal or an artefact from multiple comparisons. Additional confirmatory studies, focusing on the new question, may be needed.

**Mistake 6 Conducting the study on the wrong number of participants**

Fundamental to conducting a good study is the awareness of sample size (or study size).<sup>21</sup> If an effect truly exists then the study should be performed on a sufficiently large sample to detect it. Too small a sample size can also increase the likelihood of generating a false-positive effect because smaller sample sizes can be more prone to certain biases (see mistake 7).

What the sample size should be for any particular study depends on a number of parameters. There are several pieces of knowledge that are fundamental to all sample size calculations (Figure 3).

The first consideration is the prevalence of the disease in the unexposed population, which may be known from previously conducted research or from other similar populations. There is a level of guesstimation involved in determining this parameter, but basing it on prior knowledge is, of course, beneficial. The second factor is the minimum effect size, that is, the clinically relevant difference that we wish to detect (e.g. a reduction of bilirubin levels of 1mg/dL or 0.1mg/dL). The smaller the difference to be detected, the larger the sample size required. The third parameter is the significance level (or alpha value) of the study, which is the probability of rejecting a null hypothesis given that it is true (type-I error) and is frequently set at 5%. The final component when calculating sample size is the power of a study, the probability that it will detect an effect when there is one to be detected—power is frequently set at 80% or 90%.

With these parameters known to us or chosen, the required sample size can be calculated. Other factors that may need to be taken into consideration include the ratio of controls to cases and the likelihood of loss to follow-up in the patient population. The more complicated the study design, the more complicated the calculation of sample size becomes. Whilst there are several statistical packages available that allow the calculation of sample sizes (some opensource and others proprietary),<sup>22</sup> as with so many aspects of study design it is best to ensure

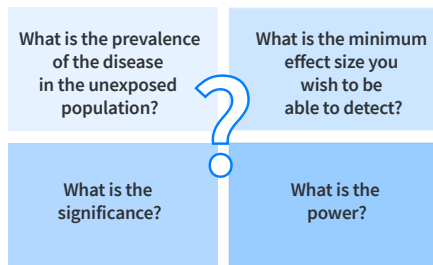


Figure 3 | Sample size considerations.

you have the correct skills mix and support for your project from the outset. Consulting or collaborating with a statistician at this stage of project development can be extremely beneficial.

In practice, sample size considerations are often based as much, if not more, on resource implications as they are on the purely mathematical components mentioned above. If we know there is a finite number of cases, an adjustment to other parameters will have to be justified to allow the study to go ahead, such as increasing the numbers of controls to cases, or increasing the effect size to be detected. Conducting a study on too many cases (perhaps by wishing to detect a very small effect size) may be considered a waste of resources and, therefore, the study runs a higher risk of not being funded (see mistake 8).

**Mistake 7 Not considering potential sources of bias from the outset**

The avoidance of bias is one of the cornerstones of good epidemiologic study design. Bias has been described as "... the deviation of results or inferences from the truth or processes which lead to such deviation."<sup>23</sup> There are a huge number of defined biases and it is not possible to discuss every single one of them here. Broadly, bias in epidemiological studies can be categorized into selection bias and misclassification (information bias; see table 1).

Selection bias arises when there is a systematic difference in the characteristics of people within a study compared with those who are not. This can include differences or errors in those enrolled in the study in the first place compared with the target population (sampling bias), through the inclusion of an inappropriate control population (e.g. choosing hospital controls who have a particular condition), and through following up participants differently dependent on their exposure status (attrition bias or follow-up bias).

Misclassification is a deviation in the measurement of one or more factors of interest in a study. Misclassification may affect all groups equally (nondifferential misclassification) or, perhaps more frequently, it may be more likely to occur in one group than another (differential misclassification, or information bias).

Information bias includes observer bias (where a study team member differentially records observations based on known exposure or outcome status), recall bias (where a participant is more or less likely to remember accurately an exposure based on their outcome status) and reporting bias (where a participant is more or less likely to reveal or suppress certain information, particularly when there may be stigma attached to it [e.g. alcohol use]).

We also include confounding in this section. There is debate as to whether confounding is strictly speaking a type of bias because it does not arise from a systematic error in the design of a piece of research, but rather from the nonrandom distribution of risk factors in the population being studied. A confounder is a third variable that is associated both with the disease and exposure of interest in any given study. If a confounding variable has been measured then it is usually possible to adjust for it in the analytical stages of the study through stratification or modelling. The study design can also be adjusted to minimize or remove confounding by restricting the study population or matching the population based on certain known (or suspected) confounding factors. If this is not possible or appropriate, careful thought as to the likely confounding variables in any study will ensure that appropriate data have been collected to allow adjustment at the analysis stage.

**Mistake 8 Undertaking unfeasible projects**

Resources for research are often limited, especially for those in the early stages of their career. Research that requires recruitment of patients is often orders of magnitude more expensive than research that can be performed on data collected via other mechanisms. Of course, although funding is a crucial point, it is not the only one, and once a research question has been framed, and a protocol has been finalized, it is important to evaluate whether the proposed project is feasible or not.

At times, a protocol that on paper seems perfectly tailored to answer a research question is not practicable and will be impossible to execute in the real world. In such cases, the protocol and research question should be revisited, and a more practical approach considered. Although this may be frustrating because plenty of time and work will have already been invested in the project, it is vital to recognize unfeasible projects, and identify them as 'nonstarters'. This course of action may ultimately be more helpful to the researcher's professional development than the alternative of starting a never-ending project.

Many factors may render a project unfeasible. Insufficient protected time for carrying out the research, and an insufficiently framed research question are pitfalls that early-stage researchers should avoid. Finding

Type of bias, misclassification or confounding	Description
Sampling bias	Differences or errors introduced during the enrolment of study participants compared with the target population
Control selection bias	A type of selection bias when controls are not truly representative of the population from which the cases were drawn
Attrition/follow-up bias	Unequal loss of participants during a study; can be due to differential follow-up of participants dependent on their exposure status
Nondifferential misclassification	Deviation in the measurement of one or more factors of interest in a study that impacts all groups equally
Differential misclassification/information bias	Deviation in the measurement of one or more factors of interest in a study that impacts one group more than another
Observational bias	Differential recording of observations based on known exposure or outcome status
Recall bias	Increased or decreased likelihood that a participant will remember accurately an exposure based on their outcome status
Reporting bias	Increased or decreased likelihood that a participant chooses to reveal or suppress certain information; more likely to occur if there is a stigma attached (e.g. alcohol use)
Confounding	Arises from the nonrandom distribution of risk factors in the population being studied. A confounding variable is associated both with the disease and exposure of interest—if measured it can be adjusted for during the data analysis. Confounding can be minimized or removed during the study design process by restricting the study population or matching the population based on certain known (or suspected) confounding factors.

**Table 1 |** Bias, misclassification and confounding. This is not an exhaustive list of terms or considerations in bias, misclassification and confounding.

good mentors and experienced collaborators (or those with special attributes relevant to the project, such as access to a unique subset of patients) can provide tremendous support—with their input it may be possible to turn an unfeasible project into a project that can be successfully completed.

### Mistake 9 Overlooking ethical and regulatory aspects

Overlooking ethical and regulatory aspects is no less important a mistake than any of the others already considered, but such considerations are not unique to epidemiological studies. Studies recruiting participants or using data obtained from humans (most epidemiological studies) have specific ethical considerations that were enshrined in the Declaration of Helsinki in 1964 and are updated periodically at the World Medical Association General Assembly.<sup>24</sup> Different considerations are required for studies involving animals (including ensuring the welfare of animals and the principles of replacement, reduction and refinement). Ethical conduct, in clinical practice and in research, should always be front and centre in a clinician's mind.

A study protocol may be well designed to answer the research question but fall short of meeting ethical and regulatory requirements. For example, collecting 50 ml of blood may be sufficient to run all the tests needed for a new study on necrotizing enterocolitis in preterm babies, but such a blood draw may be considered unethical in

this population. This point also corresponds with issues of feasibility (see mistake 8), but it is to be hoped that such a request would not be approved by the local review board for ethical reasons before empirical data indicated the feasibility challenge.

Ethical and regulatory aspects should, therefore, be integrated into the study design thought process, so that the end result is a protocol that would meet ethical expectations, receive the necessary approvals and meet publishing standards. For studies involving humans, the informed consent process, risk/benefit ratio, conflicts of interest, adherence to good clinical practice (GCP), and a clear distinction between routine clinical practice and study-related procedures are a partial list of the elements that need to be addressed when a new application for study approval is submitted. Specific considerations regarding the use of data are covered in the EU's General Data Protection Regulation (GDPR) and most ethics applications now also require data management and processing considerations.<sup>25</sup>

Of course, different regions may have different regulatory requirements, and these can also vary based on the role of the researcher in the study staff. For early-stage investigators, it is important to be aware of the requirements that have to be met both on the personal and protocol-associated level. In addition to the often-required formal training, having a good mentor and consulting colleagues may prove invaluable when trying to overcome any hurdles being faced for the first time.

### Mistake 10 Going it alone

For many of the mistakes discussed above, our opinion is that early collaboration and consultation with other colleagues and experts is the best course of action. This includes, but is not limited to, working with other clinicians, epidemiologists, statisticians, triallists, ethicists, methodologists, librarians, clinical coders, patients and public representatives. Indeed, trying to go it alone is invariably a mistake because collaboration and consultation generate the best, and most enjoyable, research, allowing us to question each other's positionality, to refine our thinking, and design the best study available to answer our question. For certain questions, the magnitude of what we hope to achieve can only be attained through multisite, perhaps multicountry, collaboration. Fostering a good network from an early stage of your career, building relationships, and not being afraid to drop a potential collaborator an email or DM, can all lead to fruitful research opportunities and ultimately better designed and conducted studies.

### References

- Hulley S, et al. *Designing clinical research*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, 2007.
- Haynes RB, et al. How to do clinical practice research: a new book and a new series in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol* 2006; 59: 873–875.
- Booth A, Papaioannou D and Sutton A. *Systematic approaches to a successful literature review*. London: Sage Publications Ltd, 2012.
- Poklepić Peričić T and Tanveer S. 'Why systematic reviews matter. A brief history, overview and practical guide for authors', Elsevier Authors' update [https://www.elsevier.com/connect/authors-update/why-systematic-reviews-matter] (2019, accessed 28 August, 2020).
- Mahood Q, Van Eerd D and Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. *Research synthesis methods* 2014; 5: 221–234.
- Richardson WS, et al. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12–A13.
- Guyatt G, et al. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; 268: 2420–2425.
- Straus SE, et al. *Evidence-based medicine: How to practice and teach it*. 4th ed. Edinburgh: Churchill Livingstone, Elsevier, 2010.
- Guyatt GH, et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995; 274: 1800–1804.
- Deaton A and Cartwright N. Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* 2018; 210: 2–21.
- Bonita, R, et al. *Basic epidemiology*. 2nd ed. World Health Organization, 2006.
- Maathuis MH and Nandy P. A review of some recent advances in causal inference. In: Bühlmann P, et al. (eds) *Handbook of Big Data*. New York, Chapman and Hall/CRC, TaylorFrancis, 2016.
- Sackett DL and Wennberg JE. Choosing the best research design for each question. *BMJ* 1997; 315: 1636.
- Allen M. Defining variables, In: *The SAGE Encyclopedia of Communication Research Methods*. USA, SAGE Publications, Inc., 2017.
- Nelson EC, et al. Patient reported outcome measures in practice *BMJ* 2015; 350: g7818.

16. The Lancet. Research: increasing value, reducing waste [https://www.thelancet.com/series/research] (2014, accessed 28 August, 2020).
17. Academy of Medical Sciences. Reproducibility and reliability of biomedical research: improving research practice [https://acmedsci.ac.uk/file-download/38189-56531416e2949.pdf] (2015, accessed 28 August, 2020).
18. Ioannidis JP. Why most published research findings are false. *PLOS Medicine* 2005; 2: e124.
19. Concato J and Hartigan JA. P values: from suggestion to superstition *J Investig Med* 2016; 64: 1166–1171.
20. Vickerstaff V, Omar, R & Ambler G. Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. *BMC Med Res Methodol* 2019; 19: 129.
21. Greenland S. Power, sample size and smallest detectable effect determination for multivariate studies. *Stat Med* 1985; 4: 117–127.
22. Porta MS, Greenland S and Last JM. *Dictionary of Epidemiology*. 5th ed. USA, Oxford University Press, 2008.
23. World Gastroenterology Organisation. Sample Size Calculation [https://www.worldgastroenterology.org/education-and-training/clinical-research-tools/sample-size-calculation] (accessed 28 August, 2020).
24. World Medical Association. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. [https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/] (1964–2008, accessed 28 August, 2020).
25. European Union. General Data Protection Regulation (GDPR) [https://gdpr-info.eu/] (2018, accessed 28 August, 2020).

## Your epidemiology study design briefing

### Meetings

- Young Investigators' Meeting [https://ueg.eu/education/face-to-face-education/young-investigators-meeting]
- Evidence Based Medicine Course [https://ueg.eu/education/face-to-face-education/evidence-based-medicine-course].

### Tools

- Division of Cancer Epidemiology and Genetics at the National Cancer Institute. Discovering the causes of cancer and the means of prevention: Tools and resources [https://dceg.cancer.gov/tools/design].
- World Gastroenterology Organisation. Sample Size Calculation [https://www.worldgastroenterology.org/education-and-training/clinical-research-tools/sample-size-calculation].
- International Household Survey Network (IHSN) Question Bank [https://qbank.ihsn.org/].

### Reference material

- Bonita, R, et al. Basic epidemiology. 2nd ed. World Health Organization, 2006. [https://apps.who.int/iris/handle/10665/43541]

- Centers for Disease Control and Prevention (CDC). Principles of epidemiology in public health practice (3rd ed.): An introduction to applied epidemiology and biostatistics [https://www.cdc.gov/csels/dsepd/ss1978/index.html].
- Munnangi S. Epidemiology of study design. *Stat Pearls* [https://www.statpearls.com/kb/viewarticle/21207/].
- Public Health Action Support Team (PHAST). Research methods appropriate to public health practice, including epidemiology, statistical methods and other methods of enquiry including qualitative research methods. In: *Public Health Textbook* [https://www.healthknowledge.org.uk/public-health-textbook/research-methods].
- Talley NJ, et al. (eds). *GI Epidemiology: Diseases and Clinical Methodology* (2nd Ed.), 2014 John Wiley & Sons, Ltd [https://onlinelibrary.wiley.com/doi/book/10.1002/9781118727072].
- University Library, Georgia State University. Literature reviews: Types of clinical study designs [https://research.library.gsu.edu/c.php?g=115595&p=755213].