MEDICAL AND SOCIAL
ENGLISH SPECIAL 2020

Science & Practice

Information from SBU – The Swedish Agency for Assessment of Health Technology and Social Services

Facing residual uncertainty

When evidence is scarce, decision-making in health and social care is a risky business. But some uncertainty is unavoidable – especially in areas where multiple factors interact and policy-makers disagree regarding values and goals.

Continued, page 2 ►

JOZEFMICIC / ISTOCK

SBU – ASSESSING HEALTH TECHNOLOGY AND SOCIAL SERVICES

The crystal ball is always hazy

FORECASTS ARE INTRIGUING – sparking both hopes and fears. Homeowners about to put their house on the market may recoil in response to the news flash 'Housing bubble about to burst!'. Investors about to choose an equity fund who read the headline 'Stock prices skyrocketing' continue reading with glee. And patients whose test results show 'a 42 per cent risk of dementia' will likely experience anxiety unless they have nerves of steel.

However, as we all know, prognoses are often wrong – and not just concerning how a new virus will spread around the world. Many Swedes will remember a particular summer when the weather prediction was glorious, only to find that the crispbread on the traditional midsummer table turned to rain-drenched mush. We wanted a detailed forecast with a precise prediction for our particular location, just for the tiny meadow where the picnic table would stand – not for the entire countryside – and for a given point in time; specifically, when we planned to be seated at the table, and not two hours later. Ideally, we would have had that prediction at least a few days in advance. In reality, the forecast that we actually received from meteorologists equipped with supercomputers, advanced mathematical models, extensive experience and dozens of measurements, was way off.

TWO HARVARD RESEARCHERS write* in the New England Journal of Medicine about questions that must be posed regarding the mathematical models used to predict the spread of infection during the COVID-19 pandemic. First: For what purpose was the model designed and for what temporal perspective – was it for the purpose of a short-term prediction, or to investigate how different assumptions may lead to potential future scenarios in the long term? A single model is seldom equally good (or equally bad) at everything. On what fundamental assumptions is the model based – for example, regarding the question of immunity and disease spread via asymptomatic individuals? How is contact tracing data used?

One important question pertains to how the uncertainty of data is calculated and reported, for example the confidence interval. In many cases, the more long-range a specific prediction, the greater the uncertainty. How reliable are the input data used in the calculation, and how different would the prediction be if the values changed somewhat, within seemingly reasonable intervals? Are the data based on confirmed or suspected cases of infection, or on documented deaths? If the model was developed from a database, was it national, regional or local? Is the model intended for general use or for a specific context – and if so, are the assumptions made when the model was constructed still valid in other contexts, in which population density and contact patterns may differ?

WHEN FORECASTS ARE VERY uncertain, one might question just how useful the underlying predictive model is. The forecast that Midsummer will be cloudy-if-not-gloriously-sunny-but-perhaps-stormy-with-torrential-rain – can that really be helpful? Of course it can, according to the NEJM authors – as long as we recognise, understand and take into account the uncertainty and the likelihood of local differences. This type of reasoning is reminiscent of the legendary Canadian physician Sir William Osler's description of medicine as the 'science of uncertainty and the art of probability'.

Ragnar Levi Editor

* Holmdahl I, et al. Wrong but useful – what covid-19 epidemiologic models can and cannot tell us. NEJM 2020; May 15. DOI: 10.1056/NEJMp2016822

UNCERTAINTY IS AN unwanted state. Nevertheless, it must be addressed – it is integral to daily life in health care and social services. Evidence from relevant, well-conducted studies, such as those reviewed by SBU, provides important support by clarifying anticipated consequences. But they fall far short of providing all the answers.

The rationale for basing health care and social service decisions on research evidence is that lives will improve and resources will be used more effectively. The aim is to reduce the risk of harm and wasting resources due to false expectations among policy-makers regarding consequences of interventions.

MEANWHILE, EXPECTATIONS relating to research may be inflated. The belief that simply more and better research could dispel all uncertainty is wishful thinking, according to researchers at the *Science Advice for Policy by European Academies* (SAPEA), a scientific advisory body to the EU Commission.

In its report *Making sense of science*[1], SAPEA emphasises that scientific knowledge should never be expected to provide perfect predictions, contribute absolute and universally applicable truths, or to adequately serve as the sole basis for decision-making. On the contrary, decision-makers are cautioned against such over-confidence – a warning that gained unanticipated relevance with the COVID-19 outbreak less than one year later. Decision-makers must take research findings into account, even when substantial uncertainty remains.

Researchers formulate hypotheses about reality and then subject them to systematic testing. Their assertions concern the nature of reality and how it functions, how various occurrences are related and – depending on subject – how the situation can, or in some cases should, be affected and changed. The endeavour to accurately describe reality is a common denominator.

However, this does not mean that all uncertainty is dispelled. According to the advisors at SAPEA, the scientific basis for decision-making will always be more or less uncertain – depending on the complexity of the issues, limitations in scientific knowledge and ambiguities concerning the ultimate goals of the decisions.

An issue becomes complex when different components in a system strongly interact so that whatever occurs in the

moment determines the likelihood of various subsequent events.[2] For example, the dynamics may depend upon how interactions between components can either boost or impede one another, the presence of control mechanisms that can either be turned on or off, and effects that may manifest at different rates and in different ways among different individuals.[3]

Greater complexity entails greater uncertainty concerning the benefits of interventions. When making decisions about complex issues, according to SAPEA, the system must be considered as a whole, and knowledge from multiple disciplines must often be applied – in order to be more confident about the outcome. To achieve the desired results, a whole range of simultaneous interventions must be combined, while carefully monitoring the effects so that decisions can be continuously adjusted as needed.

LIMITATIONS IN SCIENTIFIC knowledge pose another challenge for decision-makers, since researchers are unable to reliably assess the likelihood of various effects of an intervention. This situation may be caused by the absence of research, or the presence of findings that are ambiguous, inconsistent or contradictory due to random or systematic errors, bias. SBU's work clarifies for decision-makers both what the research shows, with varying degrees of scientific certainty, and what it does not show.

Scientific uncertainty may be rooted in methodological flaws, such as failure to take sources of error into account when designing trials; technical problems, for example related to poor or improperly used instruments for measurement or analysis; or epistemic uncertainty due to insufficient knowledge on underlying fundamentals or ignorance of alternative scenarios. Limited knowledge may also be related to the roles and incentives of scientists, as well as to who has the mandate to formulate research and to interpret and question the results.

UNCERTAINTY IN DECISION-MAKING

Complex issue – Involves multifaceted problem in which the many components either strongly facilitate or impede one another in a manner that is difficult to comprehend or predict.

Limited knowledge – Important information is missing, for example due to the absence of research, or research findings that are ambiguous, inconsistent or contradictory.

Contradictory points of view – Available knowledge is interpreted and assessed differently. The varying perspectives are difficult to reconcile, and differing goals are in conflict with one another.

CONTRADICTORY POINTS of view may ultimately be present, even when there is scientific certainty about the risks, costs and benefits of the interventions. Decision-makers and experts may differ in their interpretation of facts, as well as in their core values and outlooks on life. The same is true for the people who are affected by the decisions. Such different perspectives can be incompatible, though equally well-rooted in fact.

For example, although research findings may be unambiguous concerning the effects of measures against tobacco use and related costs, different decision-makers and different countries may have completely different opinions concerning the appropriate policy. Clearly, disagreements among experts regarding evidence-based policy are not necessarily due to scientific uncertainty.

Moreover, an individual may find it difficult to address his or her own conflicting goals and interests. A decision-maker to whom various conflicting goals are equally important will become uncertain when forced to make a choice, for example due to scarcity of resources. It will be difficult to determine which of all the critical goals take precedence and in each case, weigh what risks and costs become acceptable.

IT IS NOT UNCOMMON for the three types of uncertainty to occur simultaneously, in regard to one and the same issue, for which reason reducing uncertainty to zero is rarely possible. Nevertheless, presenting patients, practitioners and policy-makers with the most comprehensive and reliable evidence base possible should result in better informed and more transparent decisions. This practice makes it easier to distinguish what is reasonably certain from remaining uncertainties in specific areas, and of various types and degrees.

This approach makes it easier to discuss alternative actions and to cope with the inevitable residual uncertainty. Ultimately, many may find it easier to accept and comply with the decisions that are taken. ◆ **RL**

References

1. Science Advice for Policy by European Academies. Making sense of science for policy under conditions of complexity and uncertainty. Berlin: SAPEA, 2019. https://doi.org/10.26356/MASOS
2. Axelrod RM, et al. Harnessing complexity: organizational implications of a scientific frontier. New York: Basic Books, 2000.
3. Chu D, et al. Theories of complexity: Common denominators of complex systems. Complexity, 2003;8:19-30. doi: 10.1002/ cplx.10059

What predictions lead to better decisions?

Accurate predictions can help guide healthcare interventions. But some predictive modelling is misleading or otherwise inadequate. Risk and benefit must be considered before using models and algorithms for important decisions.

MANY DECISIONS IN health care and social services are based on assumptions about the future – how a disease or social problem will unfold. What happens if nothing is done – what is the risk for the individual? What assistance and interventions does the individual need?

A prognosis is an assessment of the likely course of a condition in an individual who has certain characteristics or who lives under certain circumstances. Many predictions are based on systematic observations of groups of individuals whose situation is similar, frequently obtained from large registry studies where many people were followed over time and in which a few developed the given condition.

THE PURPOSE OF a prediction can be purely informative – to gain knowledge concerning the risk that an individual will become afflicted by the given condition – or to provide a better foundation for decision-making in order to have an effect on the situation. In either case, the prediction should be as accurate as possible in order to provide a correct image of the future. The anticipated course of various conditions can be calculated using sets of mathematical instructions – algorithms – in which the various circumstances are combined and weighted in an effort to predict a certain condition in individuals, either short term or long term.

In the field of medicine, a number of prediction models have been developed; for example to assess the risk that an individual will develop cardiovascular disease. New models are constantly being published – but many are plagued by methodological problems and severe unreliability. Their accuracy may never have been compared with earlier models, and patient benefit may never have been demonstrated – let alone the associated risks.

THE MORE VERACIOUS and detailed the prediction, the more knowledge it provides and the better the decisions that should result. But when predictions are wrong, the consequences could be devastating. And this applies both to individuals and to groups. This is why it is so important to understand critical questions that must be answered before relying on predictions.

One aspect that may seem surprising is that prediction models can be accurate, even when the underlying causes of a condition to be predicted are unknown. In other words, although the aetiology of the condition remains unclear, it is still entirely possible to develop a model that provides accurate predictions – provided that the model is based on a sufficient number of correct observations and has been analysed using correct statistical methodology. ►

ALGORITHMS

Instructions to carry out calculations or to answer questions have been used in health care for purposes such as risk assessment for osteoporotic fractures, death in the ICU, and death due to coronary heart disease among individuals with hypertension or high cholesterol.

Vigorous research is currently underway to assess the benefit of algorithms for diagnostics and treatment. Researchers are exploring the role of machine learning with automatic feedback in order to 'train' the model to make more reliable predictions. When such models trigger automated actions within a closed system, they are referred to as 'intelligent' robots. Examples in medicine include insulin pumps and completely self-regulating ventilators which over time, become better adapted to the individual.

However, the efficacy, safety, costs and ethical consequences must undergo scientific scrutiny.

GETTY IMAGES / IMAGE SOURCE

▶ *Cont'd from page 4:*
However, a model constructed in this way cannot provide information as to what interventions are helpful. To do so would require efficacy studies that are able to distinguish causal factors from background and confounding factors.

Researchers who design a prediction model based on observational studies must avoid including irrelevant factors that simply occur by chance, along with the condition. Inclusion of the latter result in 'noise' that actually weakens the predictions as soon as the model is applied outside the framework of the studies.

ADVANCED MATHEMATICAL prediction models are often extremely sensitive. If the algorithm was developed and tested to make forecasts for a certain category of people, in a particular environment, it is far from certain that this model's predictions will be correct in another, similar context. Furthermore, similar algorithms may provide divergent predictions in the exact same setting. One common example is weather forecasts. Such predictions may be disparate and more or less reliable, depending on the models used by the different providers.

The main question is whether patients and users truly benefit from a certain algorithm when used as a basis for decision-making in health care and social services. The only way to approach this issue is to first test the model to ensure that it makes accurate predictions, and then study the model in practice to investigate the effects, both beneficial and harmful.

Various scientific requirements must be met. First, the mathematical model must be derived from accurate and complete data covering a large number of observations. Many diseases and conditions are multifactorial – the course for the individual is impacted by many factors and their context. In such cases, the algorithm must take many factors into account in order to yield correct and adequately detailed predictions.

IN ORDER TO avoid systematic errors, all factors that influence the prediction must be entered completely independent of the outcome. Data collection should be the same for all individuals regardless of their future prospects. Should data collection be influenced by the outcome, it could lead to erroneous predictions. For example, such a situation may arise when patients who appear to be in worse health are examined more thoroughly than others.

It is also important to test the predictive accuracy of the model in a context similar to the setting for which it is intended, such as within the same category of patients or service users. The model should also be tested in various sub-groups. In this way it can be calibrated and adjusted to not just accurately forecast the group average, but to also

WAS THE ALGORITHM DESIGNED WITH ATTENTION TO ...

- ... appropriate statistical analysis methodology?
- ... sufficient number of observed cases?
- ... correct handling of continuous and discrete variables?
- ... analysis of all individuals in the study/register?
- ... appropriate handling of attrition?
- ... consideration of complexity (e.g. competing risks)?
- ... testing predictions concerning who in the material is affected and who is not?
- ... testing such predictions for various sub-groups in the material?
- ... avoidance of over-adaptation of the model?
- ... consideration in relation to analysis of several variables simultaneously?

ANDREY-SHA74 / SHUTTERSTOCK

HOW WELL DOES THE PREDICTION MODEL WORK IN PRACTICE?

How clear and transparent is the prediction model?

- Is the model's approach to making predictions comprehensible?
- Does the model explain how reliable the predictions are?
- How is uncertainty, if any, communicated to users so as to avoid over-confidence in the results?

Is the prediction model used correctly?

- Was the model tested in different settings, using different data sources and in different patient or user groups?
- How do we know that the test environment for the model corresponds to the actual setting where the model is used and that it does not convey an erroneous impression?
  – Do the categories objectively reflect defined outcomes, or are they dependent on subjective assessments?
  – How often have the predictions of the model been accurate in the developmental environment?
  – Have the developers allowed leeway for the model to provide more or less correct predictions in various sub-groups of patients/users?
  – Will the model be used for the same type of assessments in the same type of contexts in which it was developed?
- How will the quality of the predictions be checked and how will the model be adapted based on these findings?

Does the prediction model entail risk to the individual?

- Does the model give sufficient consideration to the risk of serious consequences – is the precautionary principle applied in the predictions?
- Can the model recognise atypical, deviating data and handle them reliably in regard to the individual?

Does the prediction model contribute to better decisions?

- Does the model lead to better outcomes for patients and users? And to improved resource management? Or is there a risk that the model will lead to unnecessary or ineffective measures?
- What ethical consequences will the predictions have for the affected groups? Is use of the methodology consistent with generally accepted ethical principles?
- What type of decisions are reinforced by the model?
- Is there a risk that erroneous predictions affect decision-making so as to ensure verification of the prediction?

VECTORSMARKET / SHUTTERSTOCK

provide correct predictions for as many individuals as possible.

Before it can be concluded that the model meets at least basic requirements, it must be subjected to various statistical tests (such as cross-validation and bootstrap), a process known as internal validation. An assessment should also be made as to whether the predictions are equally accurate for other categories of patients and users, and in settings other than the one in which the algorithm was developed. This process is known as external validation. Many models published in scientific journals have not been sufficiently validated to be considered reliable.

THE STUDIES ON which the model is based, and that are used to test it, must apply uniform definitions and limitations of the conditions that the model is intended to predict. For example, should diagnostic criteria have varied over time or between countries, the results may be erroneous. The risk of non-uniform application of diagnostic criteria may be particularly high when these rely solely on a single practitioner's subjective assessment, without the support of objective measures.

Continual monitoring of the prediction model's accuracy is necessary. For example, every case of predictive failure should be assessed and used as input data to adjust the model accordingly. When such feedback and adjustment is automatic, it is referred to as machine learning or self-learning systems.

WHENEVER THERE IS a risk that an inaccurate prediction could lead to serious consequences such as severe injury or death, the algorithm must be designed to warn even at the slightest indication.

Yet another requirement is that the prediction model must describe the certainty of each prediction. An algorithm that provides precise though often erroneous predictions may be considerably less useful than a model that provides broader but more reliable predictions. The manner in which uncertainty is conveyed to users may hold great importance. Decision-makers must understand the degree of reliability of the calculation and take this into account.

Last but not least, the model must be useful in practice, provide more benefit than harm and be worth its price. For example, should the model require too much input of information that is not readily available, it may become useless in practice.

A PREDICTION MODEL must be neither overly optimistic (failing to predict the condition), nor too pessimistic (giving false alarms). Alarmist predictions risk generating anxiety and may lead to unnecessary measures. In contrast, overly optimistic predictions may create a false sense of security, which may also have serious consequences.

It is well to remember that even when a condition can be accurately predicted using an algorithm based on solid observational evidence, such an algorithm does not tell us what interventions are effective, safe and cost-effective for this condition. This would require a different type of study. ◆ **RL**

Further reading

1. Challen R, et al. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019;28:231-7.
2. Vollmer S, et al. Machine learning and artificial intelligence ... BMJ 2020;368:l6927.
3. Fall K, et al. Bra prognosstudier kan ge bättre kliniska beslut. Läkartidningen 2013;110:279-83.
4. Foroutan F, et al. Use of GRADE for the assessment of evidence about prognostic factors ... J Clin Epidemiol 2020;121:62-70.
5. Damen JAAG, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016;353:i2416.
6. Riley RD, et al. A guide to systematic review and meta-analysis of prognostic ... BMJ 2019;364:k4597.
7. Collins GS, et al. The TRIPOD Statement. BMJ 2014;350:g7594.

Reliable review requires thorough literature search

In order to extract the greatest possible knowledge from enormous research databases, knowledge of what keywords to use and how to combine them wisely is essential. The challenge of literature searches for systematic reviews is to miss nothing significant and to avoid drowning in irrelevant articles.

In order to obtain a correct idea of the benefit of an intervention, such as a treatment method, it is important to first correctly search for all of the research reports that have been published about the method in question. This challenge requires a literature search that is appropriately broad – neither so narrow that important articles are overlooked, nor so broad that they become impossible to sort out. At SBU, this balance is achieved jointly by information specialists, experts in the field and people trained in research methodology.

Within the framework of systematic reviews, SBU conducts literature searches in multiple steps. The point of departure for the search is to formulate structured questions that specify and define the subject. A typical review question includes what intervention should have been tested, in what population, compared with what interventions and how the outcome should have been measured. These elements are summarised by the acronym PICO – population, intervention, comparison (or control) and outcome.

Once the review authors have clarified their intention – and what criteria should be applied regarding study design, age of participants and language and year of publication of the articles – the next step is to select relevant databases and appropriate search terms. Interdisciplinary collaboration across professions and specialties is often required. Experienced researchers and practitioners know what search terms, synonyms and expressions they are likely to encounter in their respective fields, while information specialists are experts in various databases and what search strategy and vocabulary may be appropriate.

'One good method to begin looking for useful keywords is to start with existing, well-designed systematic reviews, since their search strategy should always be described,' says Klas Moberg, information specialist at SBU, with ten years of experience at the Karolinska Institutet University Library near Stockholm.

Another trick is to use known articles considered to be particularly relevant in the field as a point of departure.

'The databases can be searched for information on how the articles were categorised and labelled, in other words indexed, using controlled keywords – in-

STOCKBYTE / SHUTTERSTOCK

dex words. You can also find what other words and phrases, free text words, that researchers use in abstracts and titles.

The index words are obtained from the hierarchically arranged and controlled glossary – thesaurus – that is part of every large international database,' says Klas Moberg. The thesaurus in the Medline medical database (and therefore also in PubMed) is known as Medical Subject Headings (MeSH). The thesaurus in the PsycINFO psychology database is known as the Thesaurus of Psychological Index Terms, while Sociological Abstracts uses the Sociological Thesaurus.

'One advantage of using index words is that synonyms and conjugations need not be taken into account – a belabouring process which must be considered when using free text words. ▶

Can prognostic studies be trusted?

Certain critical basic questions must always be asked about research findings related to prognoses concerning people with a particular problem or medical condition.

▶ Have the investigators studied a representative sample, where all individuals were in the same phase, such as in the early phase of an illness? Ideally, all subjects were included at the same early stage, such as at the very onset of the problem.

▶ Have enough of those who were to be included in the study been followed for a sufficiently long period of time? Follow-up must be long enough to allow important outcomes to be identified. If there is attrition from the study, researchers must analyse the underlying reasons and check that these individuals did not differ in any essential way from those who remained.

▶ Were outcomes registered using an objective or 'blinded' approach? If measurement entails subjective assessments, ideally the assessors should have no information about study participants that may introduce bias, a process known as blinded assessment.

▶ If it is known that people with certain characteristics have a better or worse projection than others, have the researchers adjusted for such factors in their analysis? In the case of health, age or phase of illness may be of significance for outcome.

▶ Can a difference in the way results are reported yield a different picture? For example, five-year survival is often reported for cancer, i.e. the number of people in the group who are still alive at five-year follow-up. However, those who die may do so early or late during that interval. A survival curve with several measurement points can provide more information.

▶ How precise is the number reflecting prognosis, according to the 95-percent confidence interval of the assessment? The more observations, the narrower the interval and the more precise the result. At a late stage in the study, assessment figures tend to become less precise than in the beginning due to participant attrition.

▶ Is the reported prognosis valid and useful here and now? Or does the current situation deviate too much from the studied conditions? Is there any evidence that interventions improve the prognosis? ◆ RL

Paraphrased from: Centre for evidence-based medicine (CEBM), University of Oxford. Critical appraisal of prognostic studies. Worksheet. Downloaded July 2020 from https://www.cebm.net/wp-content/uploads/2018/11/Prognosis.pdf

ANAKEN2012 / SHUTTERSTOCK

▶ *Cont'd from page 9:*
However, index words also have their limitations.

'For instance, entry of new publications into the databases takes time. The very latest articles might not yet have been indexed,' says Klas Moberg.

DIFFERENT DATABASES MAY also use different index words for the same referent, and sometimes it is difficult to find a sufficiently precise word – such as in relation to a highly specific version of an intervention or therapy.

'That's why we usually combine index words with free text words using alternative spellings and conjugations. It is often possible to search using only the root of the word, the part of the word to which endings are added, followed by a truncation symbol, usually an asterisk.'

It is a good idea, according to Klas Moberg, to carry out sample test searches using index and free text words to see whether the articles in the hits are relevant to what you want to know.

'If the hits prove to be irrelevant we adjust the search strategy.'

Professional test searches also provide a rough idea of the scope of the literature in the field and how much time and effort might be needed for sorting and evaluating hits from the main search.

The next step of is to create search blocks, one for each element or concept to be searched. The search block includes the selected index and free text words with possible synonyms and phrases.

'Each block is first searched separately. Then the block searches are combined. Boolean operators such as OR, AND, and NOT are used to provide the database with specific commands.'

For example, the operator OR is used between synonyms and related terms within the same search block to indicate that it is sufficient for any one of the keywords to appear in a reference to generate a hit. The command OR makes the search broader and results in more hits.

The operator AND is used to combine different search blocks. This last command instructs that at least one word from each block must occur for a reference to be a hit. As a result the search becomes more specific and the hits fewer.

The Boolean operator NOT can also be used to narrow the search.

'However, the command NOT is often avoided in systematic reviews since it may exclude relevant references,' says Klas Moberg. 'In systematic review projects, having a few too many hits is preferable to missing something relevant.'

THE COMMANDS THAT can be used may differ from one database to another. In addition to OR, AND and NOT, many international databases also provide the option to use 'proximity operators', which control in what order and how close the search terms must be to provide hits.

'The exception is PubMed,' he says. 'It does not accept proximity operators.'

Literature searches pertaining to systematic reviews must be structured to use an exhaustive approach. Consequently, the search should identify as many research results as possible from everything that has been published about the PICO questions relating to the review. The challenge lies in finding as many relevant articles as possible, without obtaining too many irrelevant hits in the process.

'In the field, we say that a search has high recall when we can demonstrate that it successfully identified a large proportion of all conceivably relevant published articles. When a high proportion of the hits are relevant, the search is said to have high precision.'

AN IDEAL SEARCH should have both. But in practice you must often decide which is more important in meeting the purpose, Klas Moberg explains.

'In systematic reviews, high recall is considered to be especially important. This calls for a more inclusive, broader search instead of a narrower one, which is associated with greater risk of missing something essential. Often, we simply convert "population" and "intervention" in PICO into search blocks, to ensure that the search does not become too narrow.

'The price of choosing the broad search strategy is that it results in more hits that are irrelevant to the PICO framework. The result is more time spent sorting out irrelevant hits than would have been necessary with a narrower search.'

The job of the SBU information specialist is not just to know what databases are relevant to a particular field, but also how they work.

SBU is rarely satisfied with a single database and generally covers at least three. Since these may require different approaches, the search strategy may need to be adjusted. For healthcare-related questions, Ovid MEDLINE (or

PubMed), Embase and Cochrane are often used. For questions related to multidisciplinary issues and social work, PsycINFO, SocINDEX and Sociological Abstracts/Social Services Abstracts, plus Ovid MEDLINE (or PubMed) are used. The citation database Scopus may also be useful for the above searches.

'Sometimes the databases have default search filters or hedges that may be useful. These search tools can be combined with search blocks to find a certain category of studies, such as a particular study design.

'However, they don't always work perfectly, not even those that have been tested scientifically. And they may become obsolete when new index words are introduced.'

SINCE SUCH SEARCHES use an exhaustive approach, it is also important to remember that essential information may have been reported outside of scientific journals – for instance, in academic theses, guideline documents and research reports from authorities, organisations and businesses. Such sources, sometimes referred to as grey literature, may occasionally contain important and reliable information that may be difficult to access.

Transparency is required in systematic reviews, as well as in other scientific contexts. The authors of the overview must describe the details of their approach. The literature search must also be meticulously documented and presented so that it can be repeated with the same results, and to ensure that expert readers can evaluate its quality.

SOME KEY DATABASES

Systematic reviews, health care
- Cochrane
- Epistemonikos
- Evidence search (NICE)
- KSR Evidence, Kleijnen Systematic Reviews Ltd (KSR)

Systematic reviews, social work
- **Social Care Online**
- **Campbell Collaboration**

Assessments (using syst. overview)
- Agency for Healthcare Research and Quality (AHRQ), USA
- Canadian Agency for Drugs and Technologies in Health (CADTH), Canada
- Norwegian Institute of Public Health, Norway
- HTA database, The International Network of Agencies for Health Technology Assessment
- Swedish Agency on Health Technology Assessment and Assessment of Social Services (SBU)
- Social Care Online

Individual articles, multidisciplinary
- Scopus
- Web of Science
- Google Scholar

Individual articles, different disciplines
- CINAHL – nursing, physiotherapy, occupational therapy, etc.
- Cochrane Library – Cochrane Database of Systematic Reviews, Protocols, Cochrane Central Register of Controlled Trials
- Embase – medicine, incl. pharmacology
- PsycINFO – psychology, behavioural science and related disciplines
- PubMed – broad coverage of health and medicine
- SocINDEX – sociology, incl. anthropology, criminology, social psychology, social work, abuse and welfare
- Sociological Abstracts – sociology and related disciplines
- Social Services Abstracts – social work and welfare

'Literature searches for systematic reviews require both specialist knowledge and time,' says Klas Moberg, 'especially in areas with extensive literature from several research disciplines.'

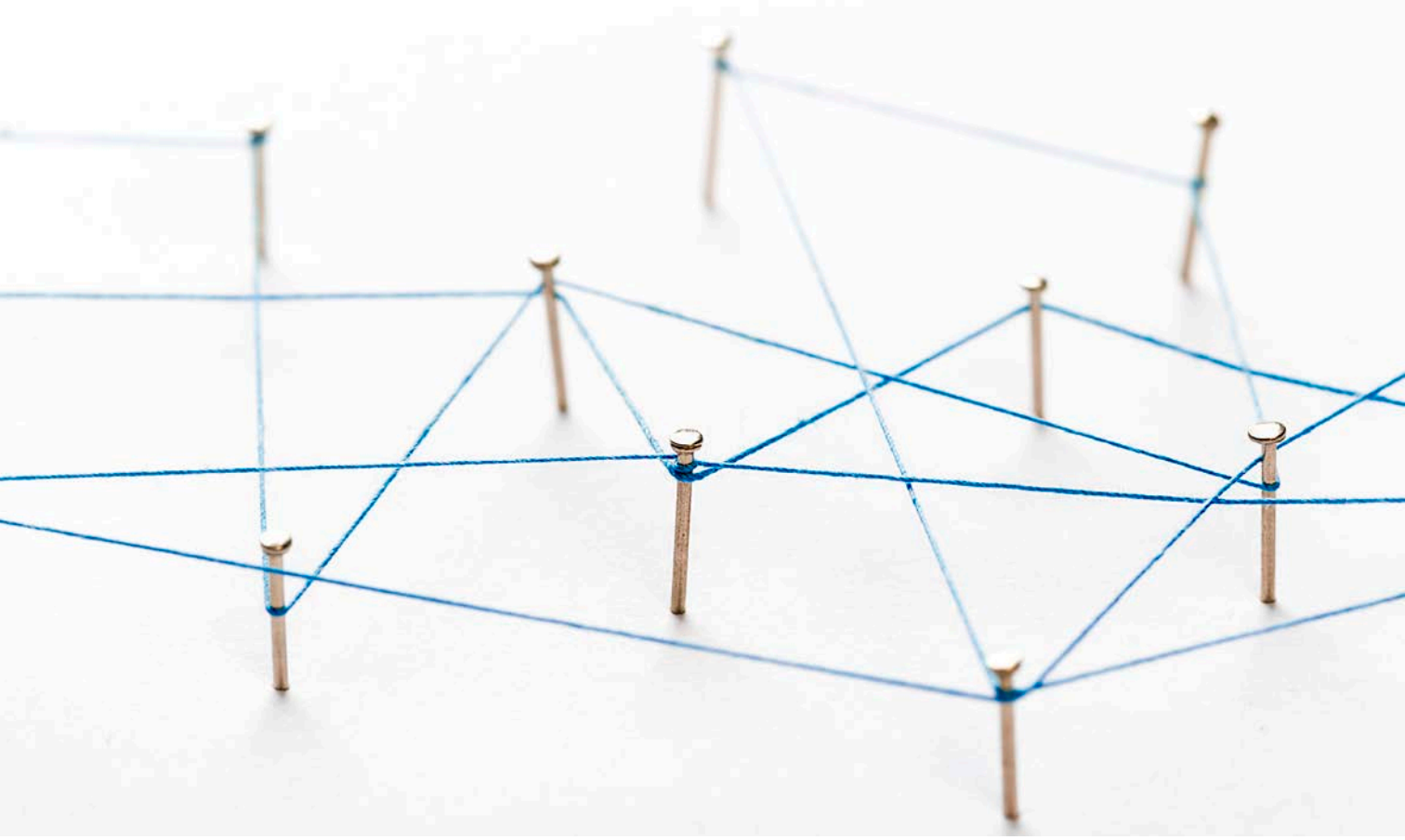
OWEVER, WELL-DESIGNED searches are absolutely crucial in order for the overall results of the reviews of existing studies to be reliable.

'If the search misses or eliminates essential findings, the composite picture may turn out completely wrong.' ◆ **RL**

Further reading

- SBU's Assessment of methods in health care and social services, see https://www.sbu.se/en/method/
- Cooper C, et al. Defining the process to literature searching in systematic reviews ... BMC Med Res Methodol 2018;8:85, https://doi.org/10.1186/s12874-018-0545-3
- Liberati A, et al. The PRISMA statement for reporting systematic reviews ... PLoS Medicine, 2009. https://doi.org/10.1371/journal.pmed.1000100
- Higgins JPT, et al. Cochrane handbook for systematic reviews of interventions. Version 6.0, 2019. See https://training.cochrane.org/handbook/current
- Karolinska Institutet library. https://kib.ki.se/sokavardera/systematiska-oversikter

CARLOS CASTILLA / SHUTTERSTOCK



NANA_STUDIO / SHUTTERSTOCK

'Correlation' does not equal 'causation'

The frequent occurrence of a certain factor together with a problem is not evidence that it is the cause of the problem – much less that elimination of the factor would cause the problem to disappear.

Not all correlations are causal. Factors that occur together with a health condition or problem, and that are statistically linked (associated or correlated) with the problem are often referred to as either risk factors or protective factors.

Sometimes research findings show that individuals who have a particular risk factor are also at higher risk of developing a certain condition or problem. Therefore, the presence of the risk factor in an individual predicts with a certain probability that they also have the health condition, or that they will develop it. An association is present.

Such associations, however, are often misinterpreted. Indeed, it may not be at all clear that the particular factor causes the condition. The demonstrated association could be causal (causal relationship), but must not necessarily be so.

To determine whether causality is involved, it is helpful to devise studies in which the believed cause can be manipulated, where its impact on the condition can be investigated. However, many ethical aspects must be taken into account in such studies. Sometimes human trials are clearly inappropriate. One example of an unethical study design would be to subject individuals who have never smoked to an intervention that is thought to induce a higher rate of smoking in the future; in other words, an intervention that is suspected of being harmful.

In such cases, researchers are instead relegated to conducting studies in which suspected harmful exposure occurs naturally in a group. Researchers can choose to study the health of the participants before and after exposure, or to explore whether the suspected negative health outcome arises more frequently among participants who have poor health. However, when considering whether an association demonstrated in such studies may be causal, several other circumstances must be taken into account.[1,2]

One key issue is the time perspective – whether the condition arose before exposure to the risk factor – in which case it is impossible for this factor to be the cause. The problem is that in many research studies it is difficult to determine which actually came first. For example, this could pertain to cross-sectional studies that investigate whether people with

a particular disease were also exposed to a suspected risk factor, compared with a healthy control group. In such studies, it is extremely difficult to determine whether exposure to the suspected factor actually preceded the disease. Instead, studies are needed that follow participants long enough for the condition to develop.

Moreover, it is necessary to rule out the presence of other causes which are common to both exposure and outcome; in other words, to rule out the presence of systematic error due to confounding factors, confounders or 'lurking variables', causing spurious correlations.

YET ANOTHER ISSUE that is often taken into account is the strength of the association (e.g. how often the risk factor and the problem occur together). The underlying reasoning is that the stronger the association, the more likely a causal relationship should be.

But this is far from certain. British epidemiologist Sir Austin Bradford Hill, widely recognised for his early ideas concerning causality (Hill 1965), pointed out that even weak associations may occur between cause and effect. He argued that it is unlikely that a strong association arises solely as a result of unknown underlying factors, measurement errors and selection errors. Were this to occur, the impact of the errors must be at least as strong as the association itself, and this is not usually the case, according to Bradford Hill.

However, others have pointed out that strong associations can also arise when statistical analyses are based on erroneous assumptions. ►

CORRELATIONS ARE SEEN WHEN ...

► ... two events are linked because one actually causes the other

► ... the two events are caused by a third underlying factor – known as a confounding factor (confounder)

► ... measurement or selection errors skew the results – information concerning the events is erroneous or mistakes were made when selecting study participants

► ... by chance, two events accidentally happen to covary.

Reliable findings are within reach

Many research findings regarding effects and associations are misleading or exaggerated. The responsibility weighs heavily on those who pay for and conduct the studies, according to Professor John Ioannidis at Stanford University in California. In a recent article, he urges his research colleagues and financial backers to adopt an array of measures.

Collaborate across the board with other researchers. Genetic epidemiology is one example of a field in which such collaboration between research groups has paid off.

Accept that findings must be replicated in new studies before the results can be considered correct, especially results from laboratory studies and small clinical trials.

Submit ongoing studies, protocols and data collections to registers in order to avoid unnecessary duplication of work and to ensure that all relevant information is accessible.

Share data, protocols and tools with other researchers. Then others can verify that the results are correct. This approach is already in use within biomedicine.

Fend off the influence of conflicted sponsors or authors, even when research findings are to be used in health economics, meta-analysis and guidelines.

Use more appropriate statistical methods and apply standardised definitions and analyses. False positive results must be minimised within fields such as epidemiology, psychology and economics. Be skeptical of alleged 'discoveries' or 'successes'.

Tighten requirements for study design and follow up using checklists for good research practices. Randomisation and blinding of investigators should be applied even in animal studies.

Improve peer review, reporting and dissemination of research. There are many suggestions for improvements, such as how to report various types of studies. (www.equator-network.org).

Provide researchers with better training in methodology and statistics. One example is the Clearinghouse for Training Modules to Enhance Data Reproducibility*, at the US National Institutes of Health, NIH. ◆ **RL**

Reference

Paraphrased from Ioannidis JPA. How to make more published research true. PLoS Med 2014;10:e1001747

* https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx

YETI STUDIO / SHUTTERSTOCK

▶ *Cont'd from page 13:*
A further aspect to be considered is whether variation in the intensity of exposure (dose) and the magnitude of the problem (response) seem to correspond. If greater exposure to a potential 'cause' is always followed by a greater 'effect', the causal inference is strengthened.

The same reasoning applies when a conceivable mechanism is found that could explain how the risk factor gives rise to the problem. Moreover, there may be experimental data to support causality – such as animal studies or studies that elucidate a mechanism of action.

In the case of human behaviour, making causal inferences is particularly difficult. This may be due in part to insufficient knowledge about the chain of events, emotions and thoughts that preceded a particular behaviour. For example, why do people start drinking alcohol and smoking cigarettes? Could it be that smoking triggers alcohol consumption – or vice versa? It is easy to envisage that there be many conceivable causes that are related through complex interactions.

IN THE CURRENT SBU assessment concerning associations involving e-cigarettes, 'snus' (moist tobacco) and smoking tobacco, the question becomes even more complicated since both exposure (use of e-cigarettes and snus) and results (tobacco smoking) are self-assessed and imprecise measurements.

In correlation studies, a particularly important challenge involves identifying and managing underlying factors that 'confound' the association that researchers actually want to investigate. While researchers must always take such confounding factors into account, it is not always obvious which ones are meaningful, nor is it certain that researchers have any information about them. The challenge is to avoid over- or under-estimating the impact of the confounding factors. Under-estimation may lead to spurious associations, while over-estimation could conceal associations that are actually present.

WHEN TWO OCCURRENCES often coincide, but in a different sequence on different occasions, it could be an indication that a confounding factor underlies the association. Should a study on a group of participants show that behaviour A precedes behaviour B among many participants, while many other participants demonstrate behaviour B prior to behaviour A, the association between A and B could be due to a common confounding factor.

For example, tobacco researchers have noted that snus users are more likely to eventually begin smoking cigarettes than are non-users. Similarly, cigarette smokers are more likely to begin using snus than are non-smokers.[3,4] In such cases, it may be important to consider the possibility of a third factor underlying both behaviours, such as the propensity to experiment with substances or to develop dependence.

Researchers usually have more confidence in a finding when it is replicated in different studies using one and the same design. However, it is important to remember that replication in itself is not proof of causality. The association may still be the result of similar studies repeatedly overlooking the same systematic error.

Meanwhile, causal inferences are strengthened when results from quite different types of studies overall point in the same direction – animal studies and mechanistic studies, as well as epidemiological and clinical studies of different designs. The investigation of a single question using several different approaches is referred to as triangulation.[5]

WHEN DISCUSSING CAUSALITY, it is ultimately important to avoid confusing necessary cause (that a certain factor is required to trigger a certain effect) with sufficient cause (that this factor by itself is sufficient cause). Even when one factor is necessary as a cause, other concurrent circumstances may be required to trigger an effect. As an example: consider two siblings who both inherit a trait for a hereditary disease, but only the one who is exposed to a particular environmental factor actually develops the disease. This hereditary factor was a necessary but not sufficient cause.

Such complexity is common. All factors that appear to be causes are not necessarily so, while many of the problems encountered in health care and social services have an array of interacting causes. Such problems are multifactorial.

A better understanding of causal relationships is crucial – especially when devising measures for scientific testing to ensure that it can be determined that the effects are truly those that were intended.

Lotta Ryk, Project Manager SBU
lotta.ryk@sbu.se
Ragnar Levi, Editor-in-Chief SBU

A CAUSAL LINK MAY BE PRESENT WHEN ...

▶ **... the suspected causal factor always precedes the effect,** which also occurs within a reasonable time interval. However, a delay between cause and effect may be difficult to establish through retrospective studies.

▶ **... the magnitude of dose and response correspond** – the stronger the 'cause', the greater the 'effect'. However, a common underlying confounding factor must be ruled out as an explanation for the link.

▶ **... the mechanism of action is theoretically feasible and consistent with known facts.** However, many important mechanisms of action are not yet known. What appears to be unreasonable today may be commonly accepted tomorrow.

▶ **... experimental data provide support for a causal relationship.** Studies in which the 'effect' increases or decreases when the suspected 'cause' is added or removed in laboratory studies may strengthen the likelihood.

▶ **... the association is strong.** However, some strong associations are due to incorrect statistical analyses, and a weak association does not rule out causality.

▶ **... the finding is consistent,** i.e., replicated by different researchers in different contexts at different points in time and in studies with different designs. However, replication may also be due to systematic errors in study design or conduct.

References

1. Hill AB. The environment and disease: Association or causation?. Proc R Soc Med 1965;58:295-300.
2. Howick J, et al. The evolution of evidence hierarchies... J R Soc Med 2009;102:186-94.
3. Haukkala A, et al. Progression of oral snuff use ...Addiction 2006;101:581-9.
4. Galanti MR, et al. Between harm and dangers. Oral snuff use ... Eur J Public Health 2001;11:340-5.
5. Munafò MR, et al. Robust research needs many lines of evidence. Nature 2018;553:399-401.

MARTIN BARRAUD / GETTYIMAGES

Critical issues for clinical guidelines

Health care and social services guidelines must be well-founded, clear and feasible to implement. The quality of the guidelines must be assessed using the AGREE tool, which raises the following questions.

Are both the scope and the purpose clear? Guidelines should clearly state the conditions or problems that they cover: to what patients or service users should the guideline apply and for which purpose – what effects are expected.

Who has been involved? It must be made clear what people and concerned professions did or did not participate when the recommendations were formulated, and in what way patients and users participated.

Have affiliations and conflict of interest been properly reported and handled? Special interests among initiators, authors, experts, reference groups and reviewers must be declared and addressed. The degree of independence of all contributors, special interest groups and sponsoring organisations must be described.

What is the factual basis? Recommendations concerning various interventions in health care and social services must be evidence-based. Guidelines must clearly describe the methods used to identify and aggregate the evidence. An exhaustive literature search should be included, as well as a systematic review and synthesis of results. If the guideline is based on calculations of cost-effectiveness and analyses of ethical and legal aspects, a description of the processes involved must be provided.

How have the relevance and reliability of the documentation been assessed? It must clearly be stated what scientific evidence or other basis was used to examine the benefits and risks of recommended interventions. Both the strengths and limitations of the evidence must be clearly presented. A description of how this material was assessed for relevance to the issue, scientific reliability (risk of random and systematic errors) and topicality must be provided. Recognised methodology must be used in this process.

What process has been used? The authors must describe the process used to formulate the recommendations and to resolve disagreements among participants. The factual basis on for each recommendation must be clearly described, and the recommendations must be specific and unambiguous. Various options for management of conditions or health problems must be clearly presented. Recommendations must be reviewed by external independent experts prior to publication, and review methodology described.

Can the recommendations be applied and followed up? It should be clear what circumstances and resources are required to implement the recommendations – and what factors can be expected to facilitate or obstruct their implementation. Suitable follow-up of expected and unexpected effects should be provided. It should also be clear how long the recommendations are expected to apply and how future needs for updating will be managed. ◆ **RL**

Further reading

More about the AGREE tool: https://www.agreetrust.org/agree-ii/Brouwers M, et al for the AGREE Next Steps Consortium. AGREE II: Advancing guideline development… Can Med Assoc J, 2010. doi:10.1503/cmaj.090449.

ONGOING

Some Current SBU Projects

CAESAREAN SECTION ON MATERNAL REQUEST
Sigurd.Vitols@sbu.se
Expected publ: autumn 2021

CHRONIC PAIN & COMPLEX INTERVENTIONS
Anna.Christensson@sbu.se
Expected publ: autumn 2021

CONTINUITY OF CARE
Per.Lytsy@sbu.se
Expected publ: summer 2021

COVID-19: LONG-TERM SYMPTOMS
Elizabeth.Ahsberg@sbu.se
Expected publ: winter 2020

DIABETES & DIETARY TREAMENT
Goran.Bertilsson@sbu.se
Expected publ: summer 2021

FEAR OF CHILDBIRTH, ANXIETY & DEPRESSION DURING PREGNANCY
Nathalie.Peira@sbu.se
Expected publ: spring 2021

FIBROMYALGIA
Anna.Christensson@sbu.se
Expected publ: autumn 2021

LIPEDEMA
Helena.Domeij@sbu.se
Expected publ: autumn 2021

MENTAL ILLNESS AFTER PREGNANCY
Goran.Bertilsson@sbu.se
Expected publ: spring 2021

MOBILITY AIDS FOR PHYSICAL DISABILITIES
Susanna.Larsson.Tholen@sbu.se Publ: autumn 2021

PERINEAL TEARS DURING CHILDBIRTH
Naama.Kenan.Moden@sbu.se Publ: spring 2021

SUICIDE PREVENTION PROGRAMS FOR YOUTH
Agneta.Pettersson@sbu.se
Expected publ: autumn 2021

VULVAR VESTIBULITIS
Karin.Wilbe.Ramsay@sbu.se
Expected publ: summer 2021

RETURN TO WORK AFTER LONG-TERM SICK LEAVE OR WELFARE
Elizabeth.Ahsberg@sbu.se
Expected publ: spring 2022

MEDICAL AND SOCIAL
Science & Practice

EXECUTIVE EDITOR Ragnar Levi, ragnar.levi@sbu.se • PUBLISHER Susanna Axelsson • TEXT Ragnar Levi [RL]
MAILING ADDRESS PO Box 6183, SE-102 33 Stockholm, Sweden • PHONE +46-8-412 32 00 • www.sbu.se
EMAIL registrator@sbu.se • TWITTER @SBU_en • DESIGN Alenäs Grafisk Form
TRANSLATION Susan & Charles Larsson • CIRCULATION (SWE) 145 000 • ISSN 1104-1250

VICTORIA SHAPIRO / SHUTTERSTOCK