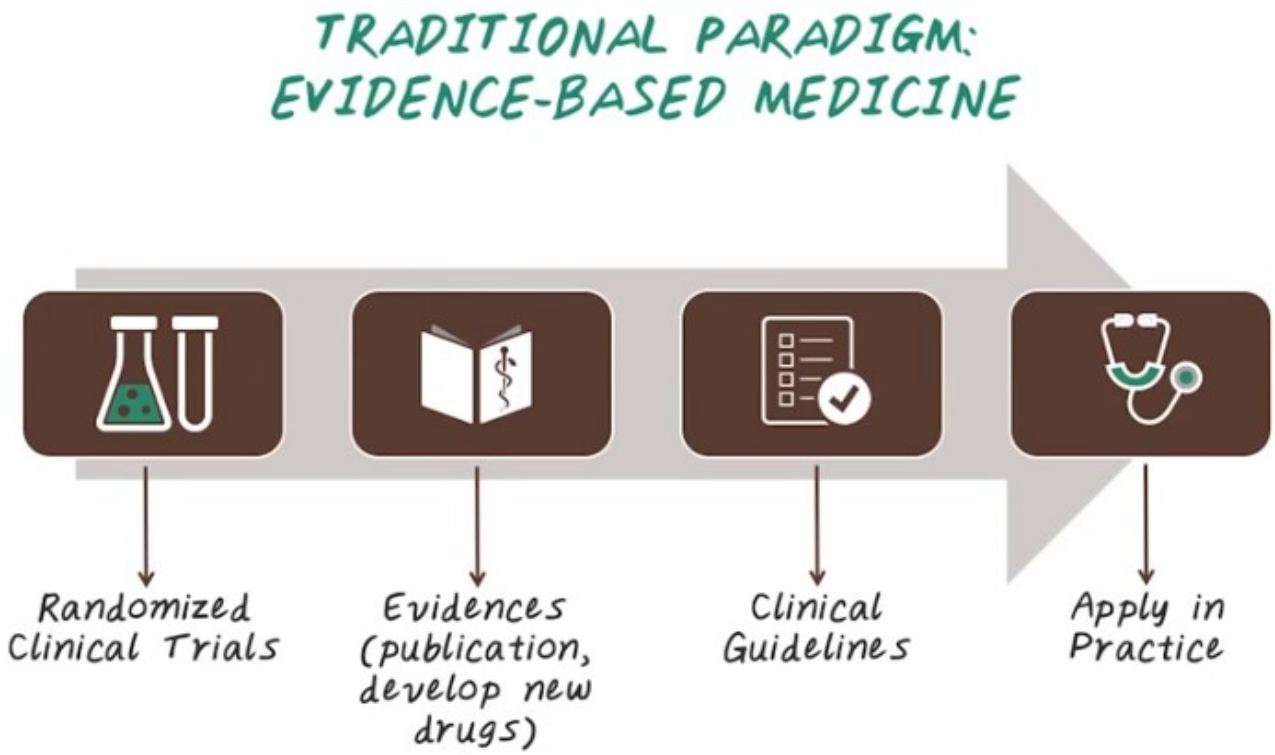
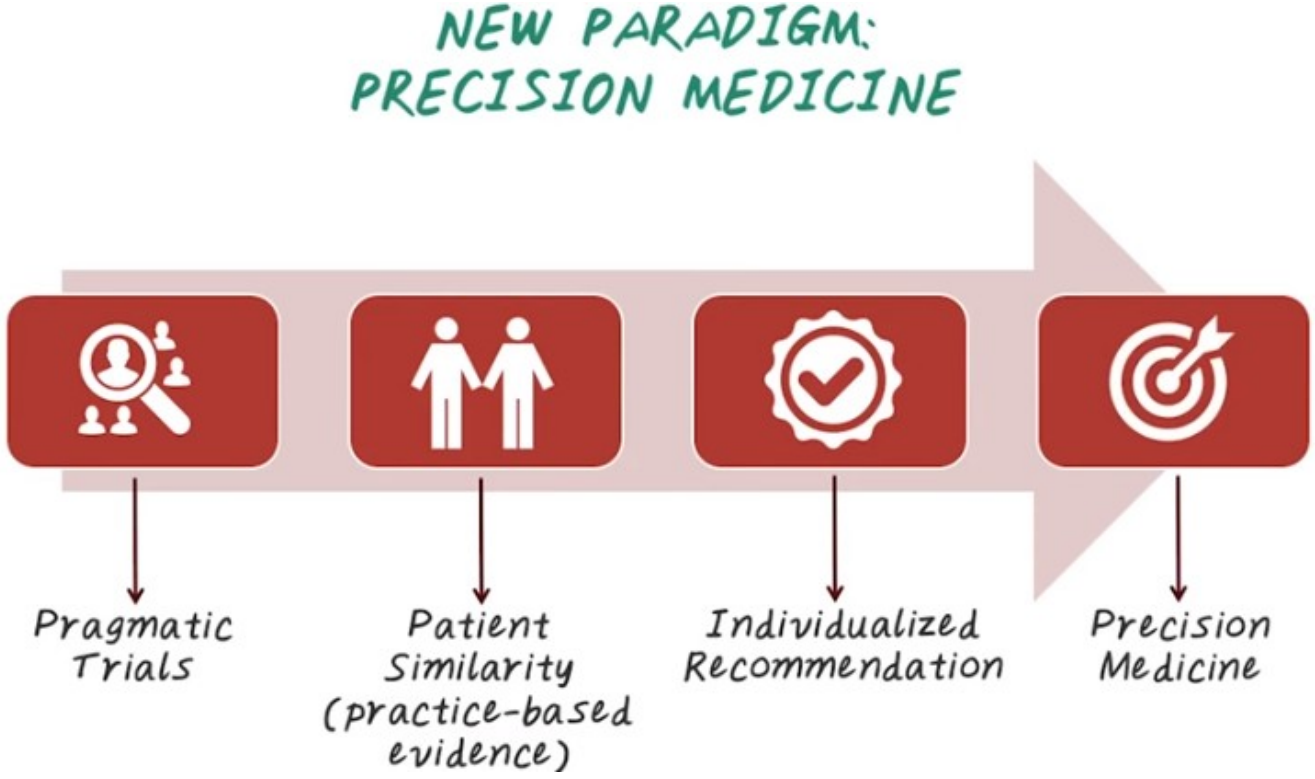
1. In this lesson we'll talk about how patient similarity can bring up a new paradigm of medical practice. We'll discuss how to find the most similar patient for a specific clinical context. We'll also talk about how patient similarity can support pragmatic trials and practice based medicine. Finally we introduce a supervised distance metric learning algorithm for patient similarity.

2. To motivate the importance of patent similarity search we need to understand the paradigm shift of medicine. Traditional paradigm considered randomized clinical trial as the gold standard for generating new evidence, and this paradigm is often called evidence-based medicine. The current recommendation for clinicians is to follow the evidence generated in the medical literature or clinical guidelines. And this evidence are largely created through RCT. There are several major challenges with this paradigm. For example, patients are heterogeneous, and can be very different from one another. And this one size fits all solution may not work in many clinical scenarios. Sometimes the guidelines are not up to date, sometimes they're not applicable to a specific patient. Thanks to the growth of electronic health records, we have a new paradigm that is emerging. We can conduct pragmatic trials based on EHR data. We can even consider doing practice-based medicine if the data driven evidence is strong. This new paradigm is called precision medicine, where personalized medical decision making is recommended. In this new paradigm, it is extremely important to be able to measure similarity among patients for a given clinical scenario. Next, let's elaborate both paradigms in more details.



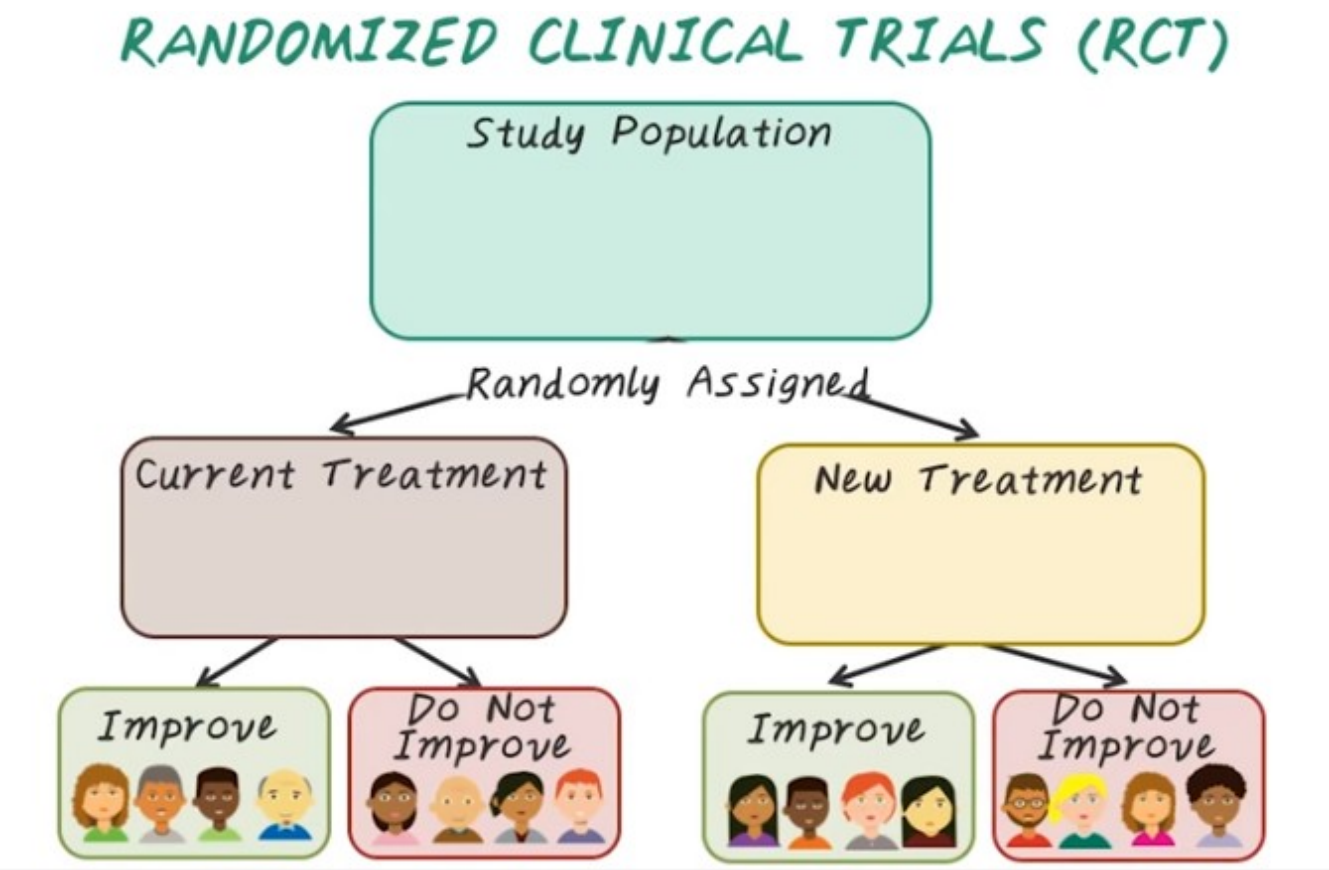
後面會詳細講 randomized clinical trials (RCT)

3. The traditional paradigm is sometime called evidence-based me dicine. The overall theme of evidence-based medicine is to make medical decisions based on the well designed and conducted research. And evidence-based medicine follows this four steps. It starts with perspective randomized clinical trials to test hypothesis. The successful hypothesis become evidence, which can be medical publications or new drugs that has been approved. Then medical experts work together to organize and prioritize all the related evidence into clinical guidelines. Finally, the clinicians apply this guideline in practice for treating patients.



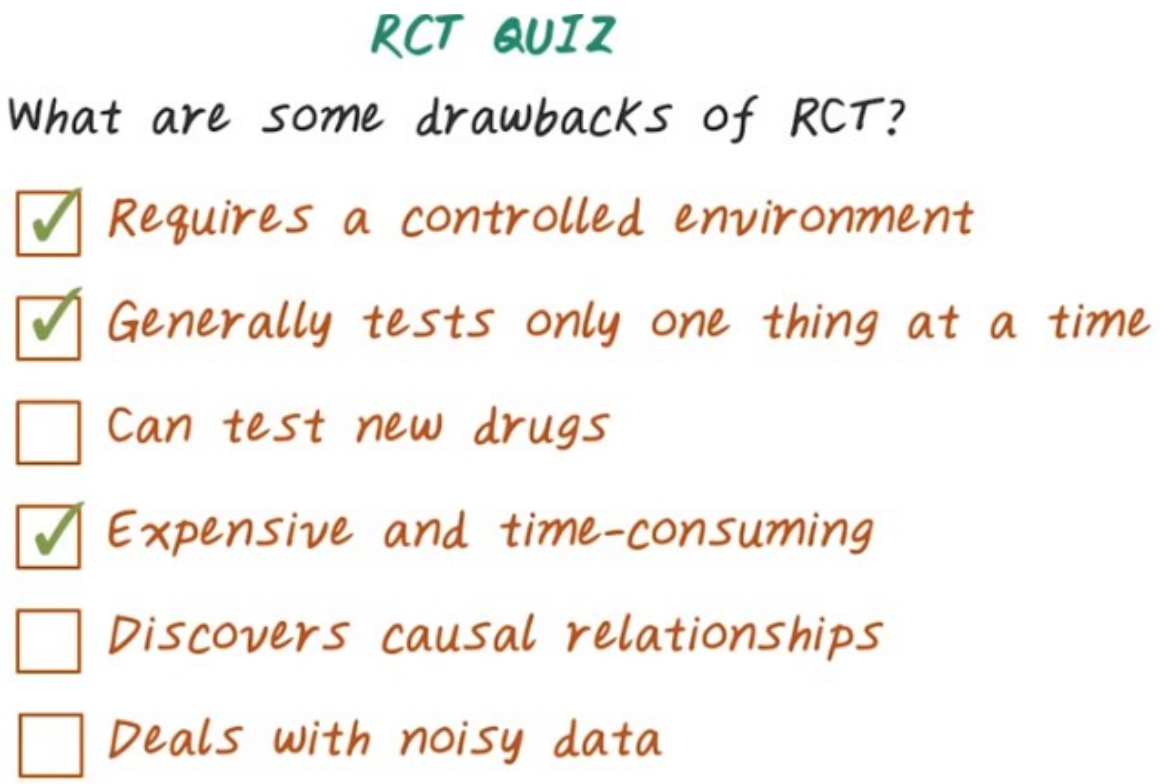
後面會詳細講 pragmatic trials
注意第二步弄出的 practice-based evidence, 後面會用到

4. Now let's talk about the new paradigm precision medicine. The goal of precision medicine is to create a new era of medicine in which researchers, health care providers, and patients all work together to develop personalized care. And precision medicine follows the following four steps. You start with pragmatic trials, which utilize large amount of historical data in the ehr systems to generate data driven evidence. Then we can apply patient similarity search for a given individual to find the similar patients. Then figure out what worked for those similar patients. Then recommend those treatment for the current patient. And this is often called practice-based evidence. If we follow practice-based evidence, we'll be able to create individualized recommendation or personalized care for a given individual. And this is how we achieve precision medicine.

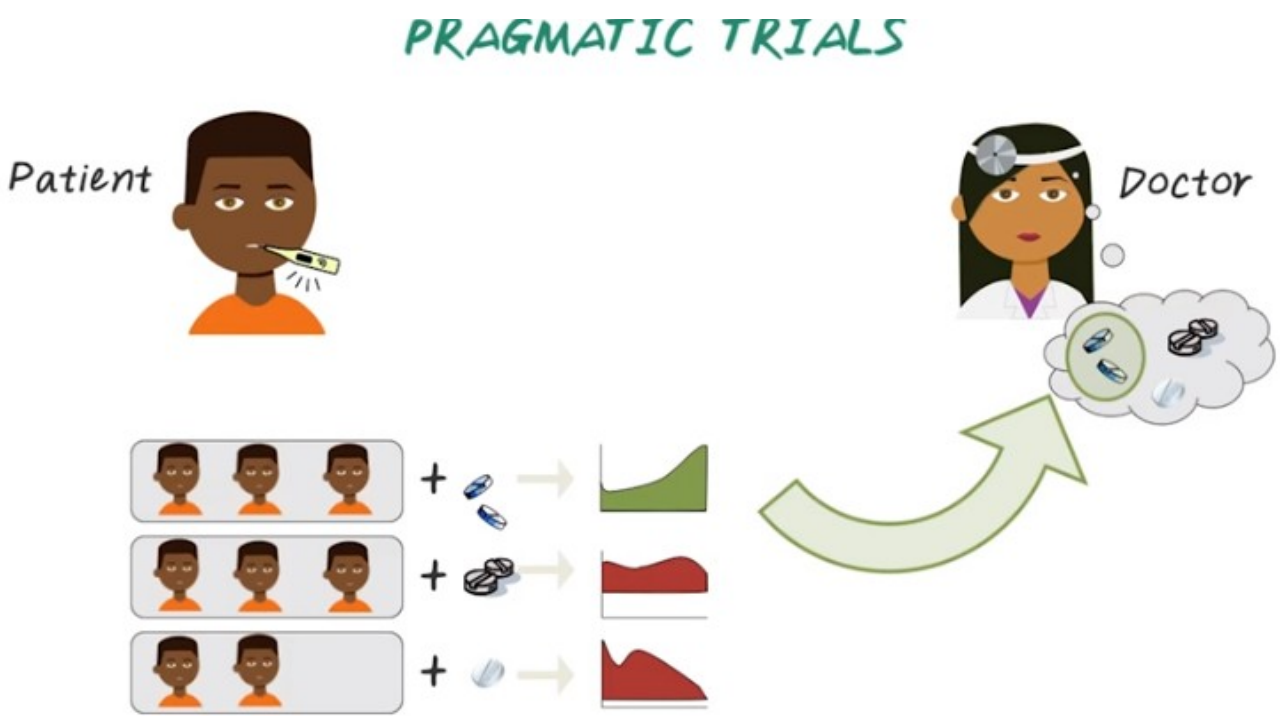


5. Next, let's talk about randomized clinical trials or RCT. To conduct RCT we start with the study population then we randomly assign everybody in this study population into two groups. In the control group, everybody is taking the current treatment or placebo(安慰劑) and in the treatment group, everybody is taking the new treatment we're testing. Then we look at the treatment outcome for both groups and there will be patients have improved outcome in the control group, and some do not have improved outcome. In the control group and similarly in the treatment group there will be people improve their outcome and some people do not improve their outcome. An RCT will compare the outcome from both group trying to figure out whether the treatment group on average have better outcome than the control group. And if the RCT confirmed the treatment group on average, have better outcome than the control group, would consider this trial as a success. Otherwise, this trial is a failure.

6. Now lets do a quiz on RCT. What are some drawbacks of RCT? Here are the options. It requires a controlled environment. It generally tests only one thing at a time. RCT can test new drugs. RCT is expensive and time-consuming. RCT discovers causal relationships. RCT deals with noisy data.

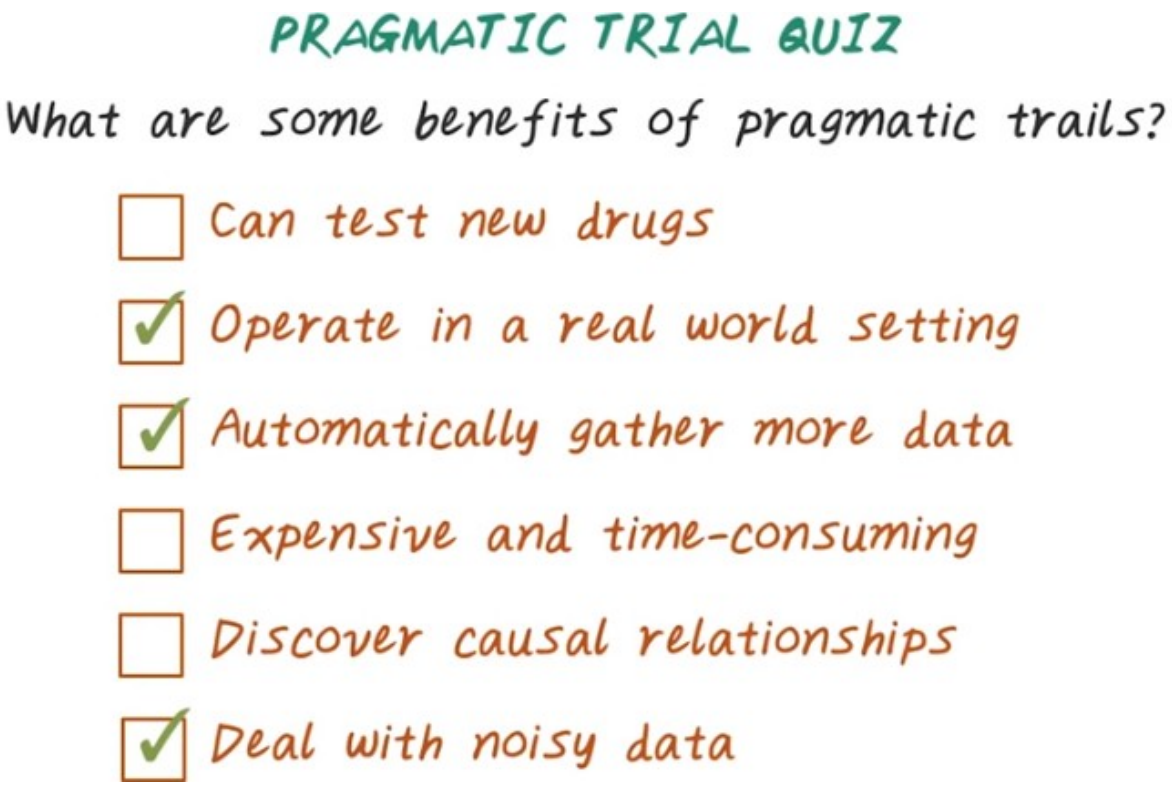


7. Here's the answers. RCT requires a controlled environment. The studied population in RCT are often times carefully selected with very strict inclusion and exclusion criteria. So the studied population often do not reflect the general population in the real world. Another drawback of RCT is it generally tests only one thing at a time. Oftentimes a RCT is specifically designed to test one drug. If the hypothesis of this RCT is rejected, the entire effort will be wasted. So in that sense, RCT can be very risky. And RCT does test new drugs, but this is not a drawback. Another drawback of RCT is, it is very expensive and time-consuming. Because we have to conduct a perspective study by recruiting patients and follow up with those patients, collecting data, then analyze that data to draw a conclusion so that can be very expensive and time consuming. RCT does discover causal relationship thanks for the randomization process, but this is not a drawback. Finally, because RCT is prospective study and the data are carefully designed and collected. Often time they are clean. So we don't have to deal with noisy data.

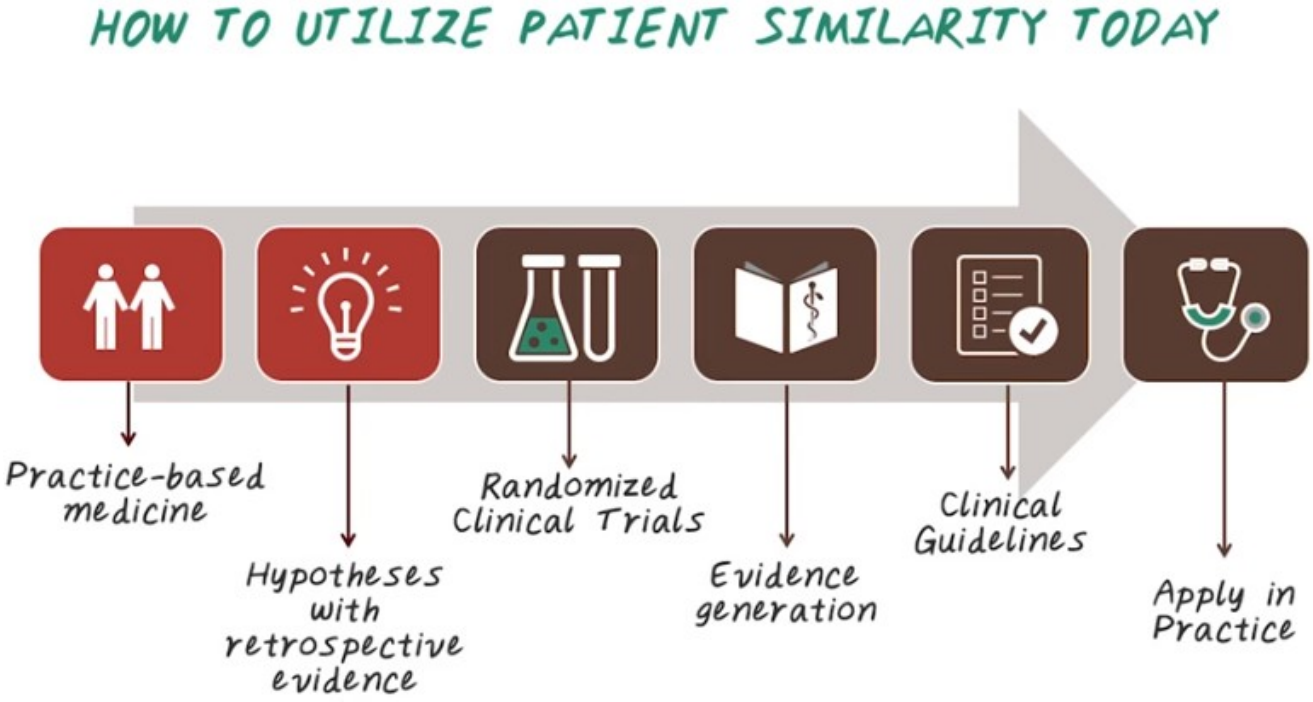


8. Next let's talk about pragmatic trials. So in traditional RCT, we generally measure the efficacy of a treatment that produces under ideal conditions. Often use carefully designed patient population in a research clinic. Pragmatic trials on the other hand, try to measure the effectiveness of a treatment that's produced in a routine clinical practice. For example, when a patient comes to a clinic, we can do a similarity search against a large patient database, try to find a similar patient to the current patient, then group those similar patients by treatment they have taken. Then look at the outcome they are getting, then recommend the treatment with the best outcome to the current patient. This is the overall idea for pragmatic trials. And the design of pragmatic trials reflects a variation between patients that occur in the real clinical practice and aims to inform choices between those treatments that works for a given individual.

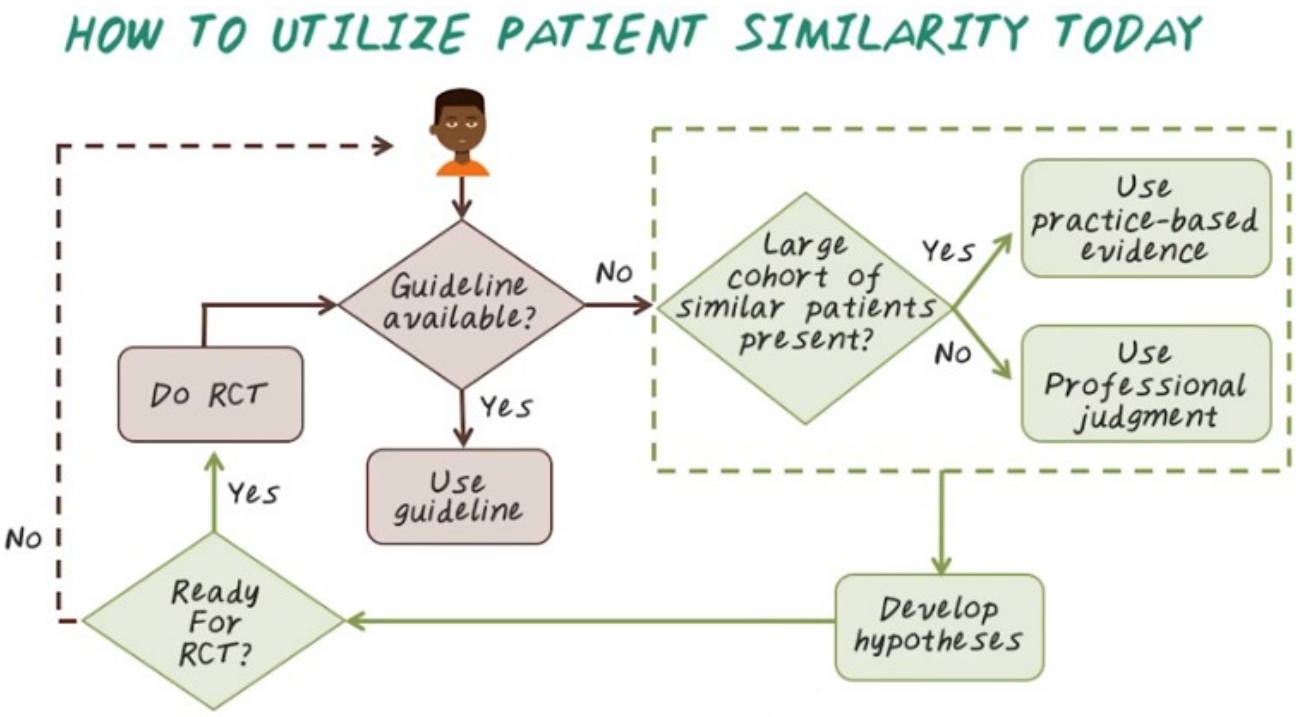
9. Now let's do a quiz on pragmatic trials. What are some benefits of pragmatic trial? Can it test new drugs? Can it operate in a real world setting? Can it automatically gather more data through this process? Is it expensive and time-consuming? Can it discover causal relationships? Doe it deal with noisy data?



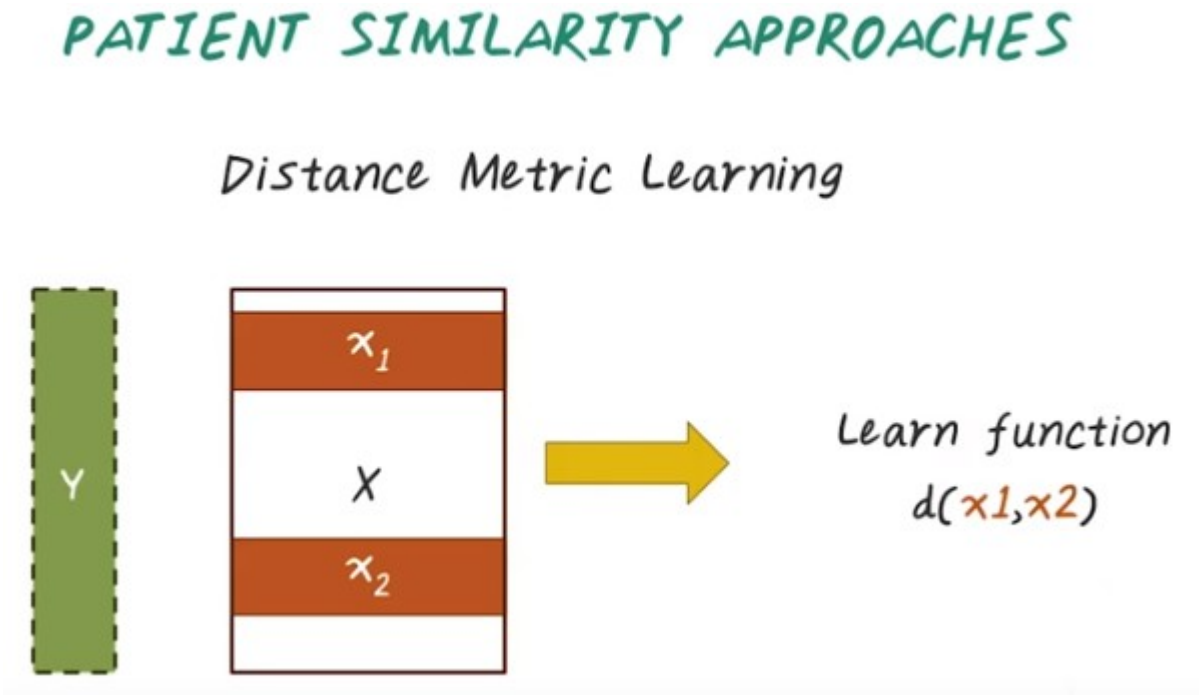
10. So let's go through this list, one by one. First, pragmatic trials cannot test new drugs. Because pragmatic trials depends on the historical data and all the treatments already happened in the past. In that case, we won't have any information about effectiveness of a new drug. And pragmatic trials do operate in real world setting, which is a benefit. Because it operate in the real world setting, pragmatic trials can automatically gather more and more data over time. Because every patient encounter will be able to generate new data that can be used for future pragmatic trials. Comparing to RCT, pragmatic trials is not expensive and time-consuming because we're dealing with historical data that already been collected. In general, pragmatic trials aren't able to discover causal relationships because randomization is not involved. Because we're operating in a real world setting, pragmatic trial has to deal with the noisy data generated in the electronic health record.



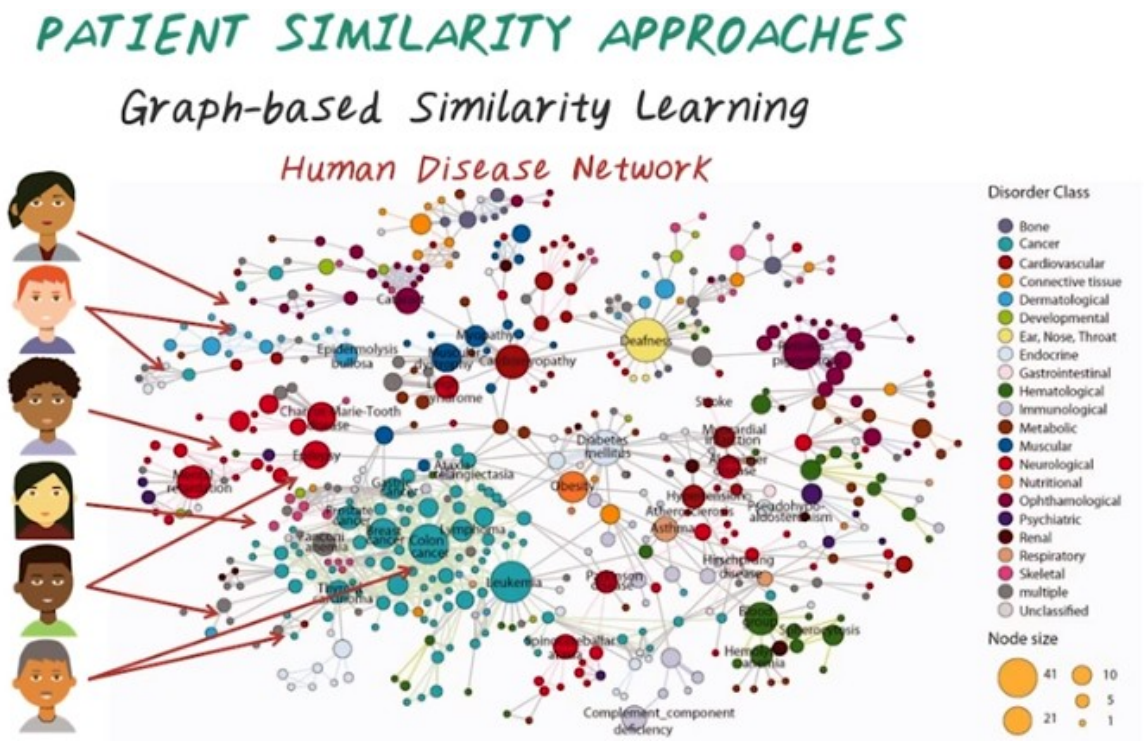
11. Now let's look at what's the process to utilize patient similarity today. We start with practice-based medicine. For a given patient, we'll look for similar patients. Then based on what happened to those similar patients, we can generate hypothesis, what could work best for the current patient. So those hypotheses, are generated based on retrospective evidence. In order to confirm those hypotheses, we oftentimes, have to go back to randomized clinical trials, to confirm those hypotheses through a prospective study. Once we generate those evidence, we can update the clinical guidelines, then apply those guideline in practice. Patient similarity, and practice-based medicine provide an intelligent way to generate hypotheses, in order to guide the randomized clinical trials, and evidence-based medicine.



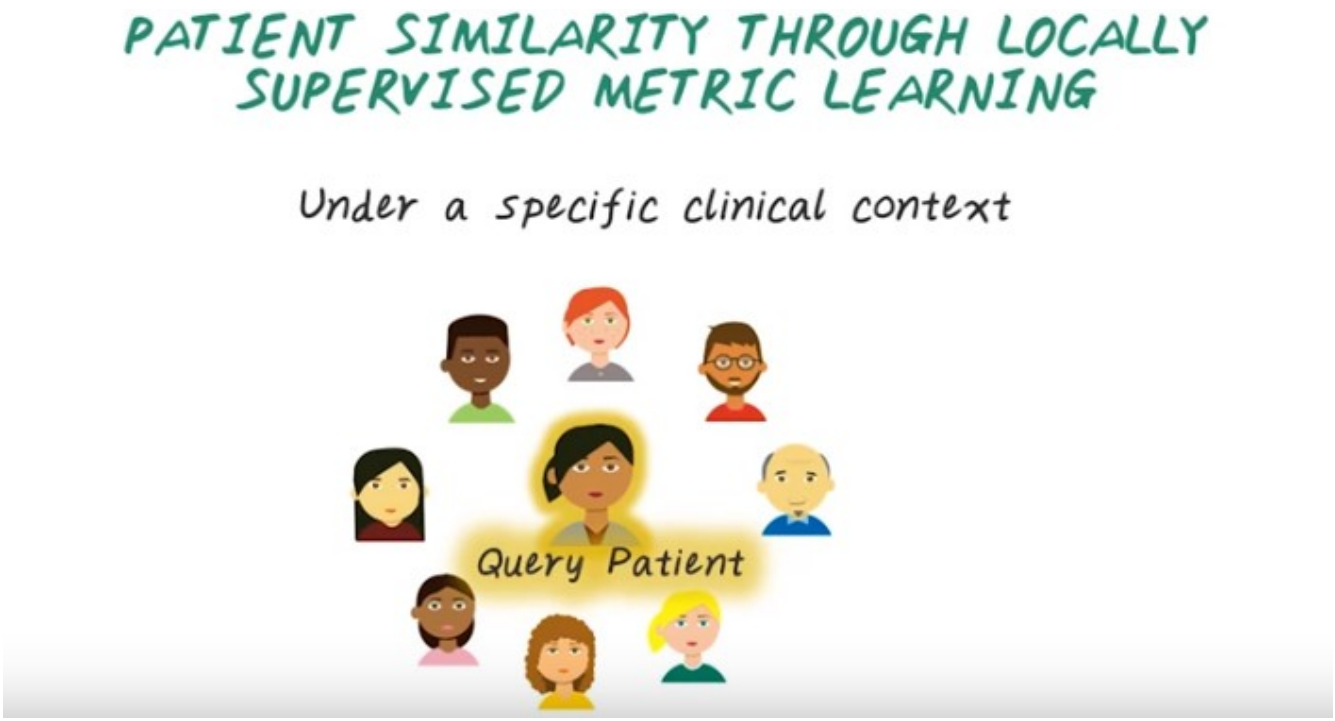
Here's another illustration of how we can use patient similarity in clinical practice. Imagine a patient comes into the clinic. The first thing we can do is, try to see whether appropriate guideline available to apply to the given individual. If there are a proper guideline, then we can directly use the clinical guideline to treat this patient. If we dont have a guideline, that is suitable for this individual, we could go ahead look for similar patient in the history. If we do have a large cohort of similar patient, we can use practice-based medicine, figuring out what treatment, were likely to work based on similar patients. If we don't have enough similar patients, available in the database, then we have to rely on professional judgment. And as we go through this practice-based medicines, many times, we can develop hypotheses, that worth conducting RCT. If we're ready for RCT, then we can conduct the RCT, to improve the guideline. If we're not ready for RCT, we can still use existing guidelines, along with similar patients' information to treat future patients. This region indicate evidence-based medicine, where clinical practice is largely depends on the guideline. And the green region, indicate a way to generate practice-based evidence from data. As you can see, both paradigm are quite complementary to each other. And they can work very nicely together, as indicated here. But the challenge is, how do we find similar patient from historical data? Next, we'll illustrate some of the algorithmic approach for patient similarity.



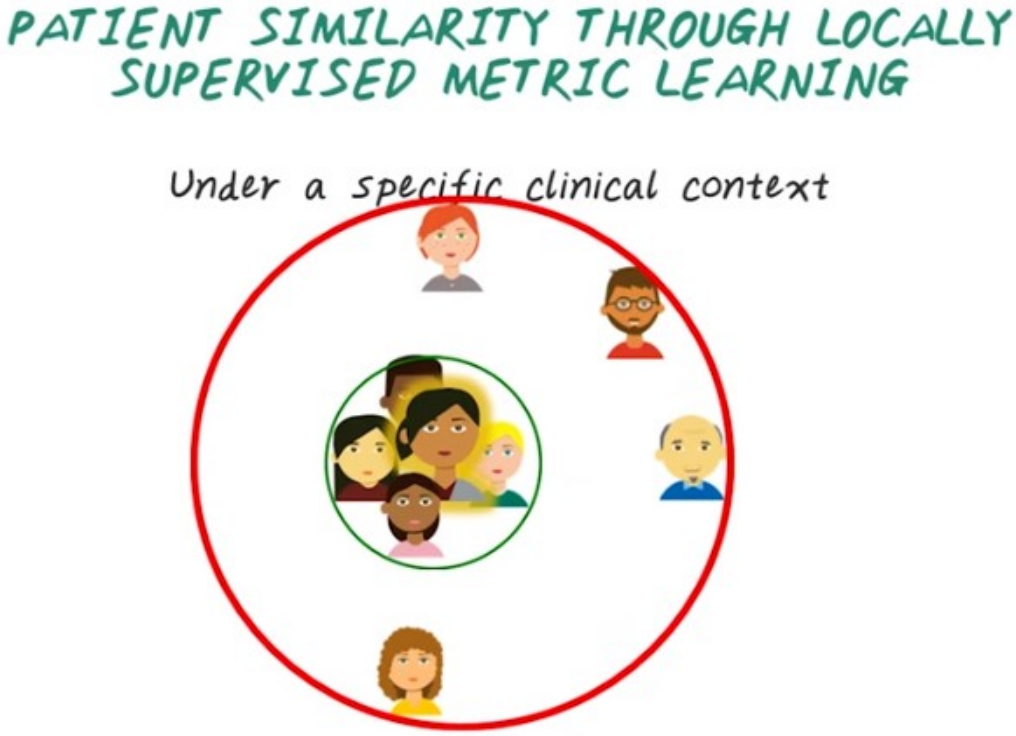
12. One way for solving patient similarity problem is to post this as distance metric learning problem. Assume we have a list of patient, and we know who is similar to whom. We also have a patient representation, and every patient is represented by a feature vector. For example, here are two patients, X1 and X2. And here, Y indicate the ground truth. For example, if two patient, X1 and X2, are similar, then they have the same label. If they are different, then they will have a different label. Then it's become a supervised distance metric learning problem. Given the ground truth label and feature vectors, we want to learn a distance metric, d(x1,x2). And this function will tell us the distance between those two patient. If they are similar, the distance will be small. If they are not similar, the distance will be large.



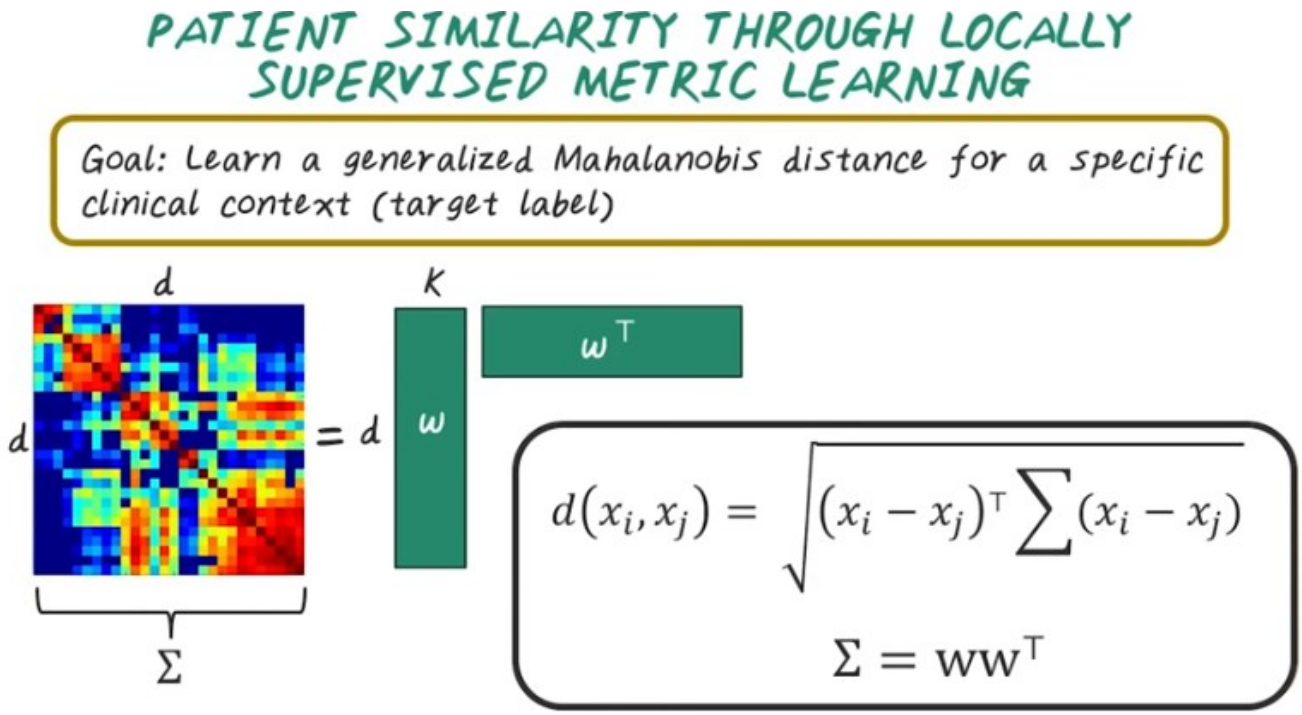
Besides distance metric learning, we can also use a graph-based similarity learning to figure out patient similarity. For example, given a set of patients, we want to figure out who is similar to whom. In medicine, we have a lot of medical knowledge that often represented as ontology, or a graph. Here is the human disease network. Every node indicate a disease, and every edge indicate a connection between two disease. And if you want to know more about medical ontology, we have a separate lecture specific on that. Now given the medical ontology or, in this case, a disease network, we can connect those patients to the diseases. For example, this patient have one disease, this patient have two. And the graph-based similarity learning is trying to figure out, given this heterogeneous graph that connecting patients to diseases. How can we figure out who is similar to whom?



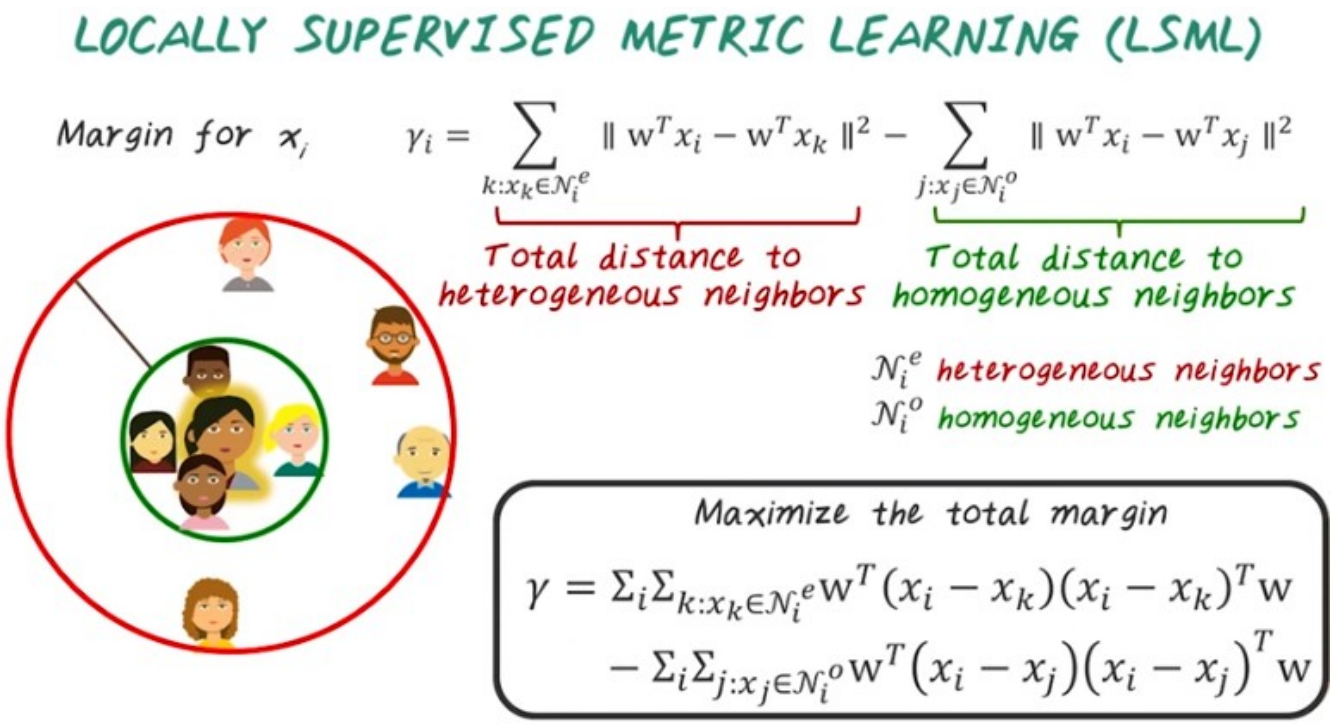
13. Now let's learn a specific distance metric learning algorithm. Called locally supervised metric learning. First, let's illustrate the intuition behind this algorithm. For example, we want to develop a distance metric under a specific clinical context. Such as, heart failure management. We have a query patient comes in and using some base line similarity measure. For example, Euclidean distance or cosine similarity we can retrieve a set of patients. That are potentially similar to this query patient. This algorithm is a supervised approach. So we have some ground truth label. For example we know, these four patients are indeed similar to this query patient. And we call them homogeneous neighbor. And we also know, these four patients are not similar to the query patient. We call them heterogeneous neighbors.



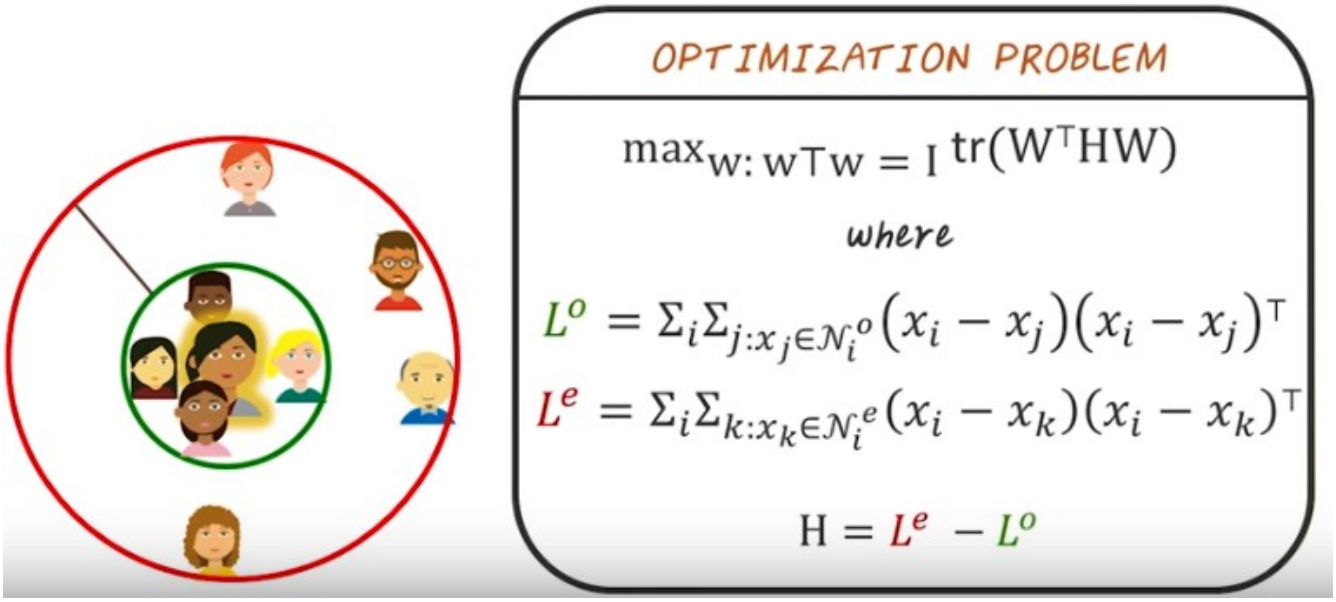
And given these two sets of neighbors, we want to change the underlying distance measure. So that the homogenous neighbor becomes closer and closer to the query patient. And the heterogeneous neighbor becomes further and further away from the query patient. And we want an algorithm that can do this automatically.



Now, understanding the intuition behind the algorithm. Now, let's formulate the problem mathematically. The goal of this problem, is to learn a generalized Mahalanobis distance for a specific clinical context. That is, we want to learn a sigma matrix. Which is d by d, where d is number of dimensions in the feature vectors. We assume the sigma matrix is symmetric and low rank. So that we can fracturize sigma, as w times w transpose. Where w is rectangular matrix of d by k, where k is much smaller than d. And in this case, the goal is to learn the w matrix. Mathematically, we want to learn this distance function, d(xi,xj). Which is very similar to equating distance. Except the sigma matrix in the middle. And the sigma matrix, is this symmetric low rank matrix, can be factorized as w times w transpose. And w, is what we need to learn from the data.



So the locally supervised metric learning, follows intuition we explained earlier. We want to define this margin for each patient. The margin is defined as, the total distance to the heterogeneous neighbors. Subtract the total distance to the homogeneous neighbors. Intuitively, it's indicated by this gap over here. And we want this margin to be large, so that the truly similar patient will be closer to the query patient. And we'd want to do this for all the patients. So we could define the total margin, which is the summation over all the margin for each patient. And the goal is to maximize the total margin by changing the W matrix.



Now we understand the objective function, is to maximize the total margin. We can rewrite the objective function in this matrix form, as trace of W transposed times H times W. And H is defined as, the difference between these two matrix. Le coming from all the heterogeneous neighbor, and L0 come from the homogeneous neighbor. Since H in this case, is a symmetric matrix. And the solution, is the eigenvectors of H with all the positive eigen values. And the complexity for solving this problem, is eigen validate composition of this matrix H.