The plight of patients with rare diseases is a critical unmet need of patients in healthcare. The statistics are frightening; there are 7000 rare diseases in the world that affect 350,000,000 people. One in eleven Americans has a rare disease. Three-quarters of patients with rare diseases are children and only half of patients receive an accurate diagnosis. The average delay for a patient to receive a diagnosis with a rare disease is 1 1/2 years. It is deeply concerning that one in four patients with a rare disease waits four years for an accurate diagnosis. There is an urgent need to communicate knowledge and expertise in the field of rare disease detection.

The journal Science, (American Association for the Advancement of Science) in collaboration with Fondation Ipsen delivers international science webinars for the general public. In 2021 these webinars focused on improving the detection of rare diseases. The Rare Disease Gazette is a magazine that broadcasts these discussions.
**The Conversation**

**Experts of the month**

**Tim Guilliams, Ph.D.** (Healx, Cambridge, UK)

**Marta Mosca, M.D., Ph.D.** (University of Pisa, Pisa, Italy)

**David A. Pearce, Ph.D.** (Sanford Health, Sioux Falls, SD)

**Sean Sanders, Ph.D.** (Science/AAAS, Washington, DC)

**Domenica Taruscio, M.D.** (National Centre for Rare Diseases, Rome, Italy)

**Sean Sanders (host):**

In this webinar we are going to turn to finding solutions, and I will be questioning our panel about current and future pathways to better detection, diagnosis and treatment of rare diseases.

**Domenica Taruscio:**

I am the Director of the National Center for Rare Diseases after the Italian Public Health Institute in Roma, in Italy, and I have always worked on rare diseases. My interests are from primary prevention to diagnosis, to surveillance, to education, and of course a translational research. I am interested in all aspects of rare diseases.

**Tim Guilliams:**

I am the co-founder and CEO of Healx. Healx is a mission-driven tech company focused on accelerating treatments for rare diseases, and it is an immense privilege to be here today. And our approach for finding rare disease treatments it is slightly different to trying do this with a new drug discovery paradigm, and at a different scale, so delighted to be here today.

**David Pearce:**

I am President for research innovation in the World Clinic Program for Sanford Health. I’ve been on an incredible journey in the last few years, having had a basic research lab, which was looking to understand a rare disorder called Batten disease, and go into the first ever clinical trial for that particular disease, which made me appreciate the need for registries and natural history studies. Without that, we would never have gone to a clinical trial. That journey has taken me to lead a large research institution with a focus on rare diseases and we have established a national, if not international, registry program for rare diseases. I also am the Vice Chair for the International Research Consortium.

**Marta Mosca:**

I am Professor of Rheumatology and Chief of the rheumatology unit at the University of Pisa in Italy. I am the coordinator of a European reference network. The name of the network is Reconnect and is the network dedicated to the rare and complex connective tissue and musculoskeletal diseases. So this is a wonderful project from the European Commission and dedicated to the care of patients with rare diseases. As a doctor, my mission is to diagnose and treat these patients and to research on how to better monitor and treat patients with rare rheumatic diseases.

**Sean Sanders (host):**

David, what can be done to get a more accurate diagnosis? This is a huge problem for anyone with a rare disease. What are we currently doing, and what do you think the future will hold for this?

**David Pearce:**

The diagnostic odyssey, as we often called it in rare diseases, is a problem that has been with us for many years. And it is a really difficult problem as a physician, if you encounter someone that has some of the symptoms with respect to the rare diseases because there can be multiple symptoms, which make it very difficult to get to the direct diagnosis because there could be a variety of different things that are presenting. So therefore, you might say, “Okay, well you have this symptom and this could lead to this particular disease.” But the problem is you may then you have multiple codes, and certainly in the United States, you have your ICD-10 codes for diagnosis. Therefore you may ultimately have many diagnoses, but not the correct one, because the vast majority of rare diseases do not have ICD-10 diagnostic codes. That makes it tremendously difficult with respect to putting the appropriate treatment in place. So through this international Rare Disease Research Consortium I’ve talked about, we have put workshops together. The idea is to better mine the electronic medical record and then probably use artificial intelligence and some sort of sophisticated approach to indicate, “Okay, these are the different symptoms that are associate with a given disease.” When it comes down to it, despite the fact we want a genetic diagnosis, most rare diseases still only have a clinical diagnosis based on the symptomatic presentation of the disease. I think we need to get providers, payers and researchers to use all that data and figure out a better way to expedite getting to a treatment and diagnosis.

**Domenica Taruscio:**

I would like also to add that there are many problems around this issue, not only coding, which is very important. There are also problems with nomenclature which we should use internationally. We need a set of agreed upon nomenclatures to understand each other and collect data which are also machinable. This is a big issue, and the diagnosis is still a problem for many patients that even with all these studies and efforts, they do not have a diagnosis. There is an international network on undiagnosed diseases, and this is dedicated at the global level to collect information for patients without diagnosis and share best practices, best protocols and experiences. So I think that there is a big, huge effort at the global level also to align the synergies.

**Marta Mosca:**

One important aspect is that we base our diagnosis on clinical symptoms, and sometimes, these symptoms are very mild in rare diseases. The very first step before coding, before genetics is the knowledge that there are rare diseases. We are here today discussing this. We need to make these conditions more clear. One important aspect is that we do not recognize what we do not know. All this information should be made available for all the doctors and the physician dealing with patients and then move onto codification, and nomenclature. Often, as said above, there are very mild clinical manifestations and collecting data and registries become very important. This entry step is very important for diagnosing a rare disease.
Tim Guilliams:
From a tech and AI perspective, what is going to be be possible soon, and is already possible but not being adopted yet, is to try and think about diagnosis in a less narrow way. A patient could potentially be 60% Bartter syndrome, 30% Pitt-Hopkins Syndrome and 10% Rett Syndrome, for example. We might think about diagnosis in a slightly different way and so form different ways of putting patients and diseases together. This may inform smart patient and disease clusters. We may use this approach to try and predict new treatments. But I think it could start with diagnosis where there are so many comorbidities, and so many links between rare and common diseases. The technology is available to make sense of that and maybe have a bit more holistic diagnosis so that it is not just one rare disease at a time.

Sean Sanders (host):
So just coming back to the training of physicians to think about rare diseases. How do we need to think about this? What needs to change in the training of physicians so that they are more aware of rare diseases and maybe do not just dismiss or misdiagnose, which is another, clearly another issue?

Marta Mosca:
We need to start to build up teaching and courses even at the university level on rare diseases. For example, I teach at university and there is no specific teaching for doctors, or medical students. I think that this is an important first step because they need to understand from the beginning of their careers, that there are rare diseases. Probably, there should be a dedicated course on rare diseases, on how we deal with this or the issues we have with patients with rare diseases, sometimes even in spite of the diagnosis or the type of disease. But then we need to move further. For example, we have educational projects for GPs, and healthcare professionals including nurses and physiotherapists. Diagnosis might come from any healthcare professional, not only physicians. This is very important because a patient might go to a physiotherapist or nutrition specialist for some symptoms, and these symptoms might be related to a rare disease. So this is something we do and different levels. We need to start from the basics and then move on, and continuously build on this.

Domenica Taruscio:
We need to start from the medical school and then go on, so that there is a continuum of education across all multi-stakeholder teams. We should also teach, I think, how to work together in a multi-stakeholder team because just one doctor or just one nurse working alone cannot do the best that they can do; they cannot make the right or correct diagnosis by themselves. Now, it is a must to work in a multi-dimensional team. Maybe we should use the European Reference net to deliver e-learning platforms, for example, to provide a second opinion consultation in one country or transversally in a multi-country level or international level. The picture is changing so fast, and we should really to take advantage of these.

Tim Guilliams:
There are 7,000 rare diseases. In rare diseases, it is really important to find the links with common diseases. That is how you can make breakthroughs faster. It would be great if some technology developers could come up with an app where you could almost use it per disease or per disease cluster. We need a design that is relevant for patients. We need technologies that are relevant. Then patients and families can use the app to talk to various stakeholders and educate the nurses, the doctors, everyone involved. My message to those who are developing apps, is could you please help us design something like this where you can go from the 7000, but then also something more specific, where we can examine one disease or one disease cluster at a time.

Domenica Taruscio:
There are many new activities, and new possibilities. We should synergize and make an effort to do our best. There are many societies, for example, ICORD, which is an International Conference Society For Rare Diseases and other initiatives, but there is not a unique hub where we can collect all this information. So maybe we need a unique hub at international level, not just European, or not just Australian, or USA, but globally. We must use these global instruments.

David Pearce:
The problem is that this is just a huge task, isn’t it? To get all of that information out there. When I think of the journey I’ve been on with the pediatric neurologist that I recruited to do this natural history study, when I first met him he said, “I’ve seen one child with the disease you’re talking about.”

Many physicians really in their lives will only see, maybe one of these children or adults with a rare disease. So in trying to map out that education for what we think it could be 7000 rare diseases, it is going to be be tremendously difficult. I think in my area how many diseases present with seizures. You have to say maybe take all the 300 or so disorders that present with seizures and follow that map along to each a diagnosis. We have to get there with the education, but how we actually get to the different sub-specialties is going to be critical. Primary care physicians in particular, have to manage this from the get-go. The first piece of the puzzle is to get patients to the right sub-speciality sooner rather than later.

Tim Guilliams:
I actually wanted to stay with you on a related topic, and that is just taking one step before diagnosis, looking at detection. A comment was made in one of our webinars that there are many people out there who do not know that they have a rare disease, and that this is a huge problem. So I am wondering how the tech side of things and about the work that you do to help identify patients, maybe through apps. Do you have any thoughts on that?

Sean Sanders (host):
You almost want to be preventative in a way. You want to let people know before the disease manifests.
people will understand that there is something going on. We often take eight to ten years to get a diagnosis, which is incredibly slow. I do not know what could be designed because it is a fine line. You almost want to go preventative, but also you do not want to alarm people when it is not needed.

**Domenica Taruscio:**

Well, of course, I think that we can also use a traditional tools such as genetic counseling, for example. So you can have great genetic counseling and see if there are families or members in a given family with high risk. We should push in this direction because it is not used in many countries. Genetic counseling is part of the followup, but it is not mandatory, for example. So I think that we can have this possibility for families at risk.

**Marta Mosca:**

I just wanted to highlight the fact that there are many different types of rare diseases. So for some diseases, this obviously can be done and there is something we can do on a family basis, but there are also rare diseases for which the diagnosis is based on the clinical manifestations. The idea is appealing to capture very early clinical manifestations. For example, this is my experience in rheumatological brain diseases where there is a delay in diagnosis and the disease onset, sometimes is very mild. The risk is to over-diagnose or have too many referrals for rare diseases, which cannot be done. Rare disease by definitions are rare and if you over refer patients it is going to be difficult to be supported. I think it is a tricky thing that is different for different rare diseases.

**Sean Sanders (host):**

Very briefly, what do you mean by high quality? Just so that the audience can understand what you mean.

**Domenica Taruscio:**

High quality means that there is internal quality control, external quality control and several other procedures that scientists and registry managers should follow. There are strict rules to be observed in order to obtain high quality data. Then we can analyze the data and obtain good results. Otherwise, if you do not observe these rules, you have a ‘garbage in’ and a ‘garbage out.’ If you want to have useful research registries to understand the pathogenesis and etiology of diseases and provide useful input for new therapies, we need high quality data.

**David Pearce:**

Domenica is absolutely right. The problem is that we are dealing with the heterogeneity of different rare diseases and the different types of data that you would want to collect, depending on how the disorder manifests. I have a slightly different approach; when families have a diagnosis of a rare disease, they reach out, they want answers and want information. First of all, it is okay to just have what I call a contact registry. If you do not have the individuals registered then you cannot go to the next step and maybe put a natural history studies together. And then depending on the disease and the group that you’re working with, what we do with our rare disease registries is allow them to put together their own survey questions, because they know the disease, they know themselves, the symptoms and the clinical profile better than anyone else.

That will evolve into that sophisticated database by scientists saying, “Hey, can we have access to your medical records?” or “Can we have a researcher or a physician who is very familiar with the disease, put a more sophisticated survey together?” A rare disease diagnosis or an undiagnosed disease - before you get to that diagnosis - is a terrible journey. So providing hope, I think is one of the underlying things of putting that first registry and natural history study together because it shows that there are people who are interested and want to help.
There is a lot of mistrust in the rare disease field. Sometimes, we have to walk before we can run, to actually get to that stage, and so build some trust. We cab develop simple data sets first and then get to that higher level.

“Patients, know the importance of registries, and they really want to be reassured, and they are really keen on participating and giving data because they really know that this is, I am not saying the only way, but a very, very important way to help them...”

Marta Mosca:
I would like to share with you the experience of ERNs, because within the ERNs, we need to develop registries. But I would like just to say that patients are willing to share their data. Our patients, know the importance of registries, and they really want to be reassured, and they are really keen on participating and giving data because they really know that this is, I am not saying the only way, but a very, very important way to help them. So within the ERNs, there are actions together with patients aimed at understanding, and Domenica knows this very well because she works with many ERNs and with us to find ways to share data. They want to be sure, they want their data to be used, but they want to give their data, and so we really need to find a way and to collect what we need. Sometimes it is epidemiological data; sometimes, and sometimes it is more detailed data. For some diseases, we really do not know much and so registries really are very basic, for us to gain information on.

Sean Sanders (host):
Tim, I wanted to ask you if you could speak to the quality of the data and how that has been used in the type of work that you do. Domenica mentioned ‘garbage in, garbage out’, which might be a term that some people are not familiar with so maybe, you can just explain what that means from a data perspective?

Tim Guilliams:
From a data perspective, this is incredibly important. So the amount of information that companies like Healx (https://healx.io/) could use to try and predict a treatment is not a snapshot, but something that keeps evolving because most rare diseases are progressive and so you want something longitudinal. You want information that goes with the patient. On the one hand, yes, there is this principle of garbage in, garbage out. However, there are also new algorithms, which allow you to basically work with much more noisy data and incomplete data. A big focus of Healx is that we have developed a rare disease knowledge graph. And a knowledge graph is basically a database of relationships. Like with Google when you search for something, usually in the top three, you have the right answer. But is our case, as part of this graph, you want different levels of certainty and relationship, and some of those relationships will be wrong, some of those will be right, but from imperfect data and existing relationships, you want to start predicting new relationships and gaps. This is a really big break-through, because it means that even with noisy, imperfect data, you may be able to use machine learning to try and make sense of imperfect information. If you could find a way to empower patients and give them the right tools, knowledge will advance. But how do we make it possible for patients and families to share their data: to empower them to do so, and also potentially help come up with solutions that are useful beyond the natural history and research?

I do not know the exact solution, but I think we should focus on the patient, their journey and making it as easy as possible for them to share information and to contribute. Patients and families are the experts. They need to help us develop the right treatments.

Marta Mosca:
I would like to share with you the experience of ERNs, because within the ERNs, we need to develop registries. But I would like just to say that patients are willing to share their data. Our patients, know the importance of registries, and they really want to be reassured, and they are really keen on participating and giving data because they really know that this is, I am not saying the only way, but a very, very important way to help them.

Domenica Taruscio:
There is a breakthrough to share with you from Italy. Since 2021, the National Registry for Rare Diseases, which is a national registry has collected information on more than 500 different rare diseases. This was for public health. We have a patients come to us, at Instituto (an Italian Public Health Institute) and say, “Listen, I would like as a patient, or an association to better understand our disease”, or, “Please can you help us to build a high quality registry?” From these requests, we have started ten specific rare diseases registries, for example, Prader Willi Syndrome and, cystic fibrosis. These registries have been started because patients want to better know their diseases, and improve the quality of their lives. This is another perspective you should be aware of – driven by patients. It is not the old perspective of collecting data like 20 years ago, by experts, epidemiologists, and statisticians. Now it has completely changed and the patient community comes first. The changes are being made first by patients and associations, and second by experts, stakeholders, expert teams and institutions.

Tim Guilliams:
This is a nice example of a true partnership where researchers, patient communities and focus groups work together to make a difference and align behind a common mission, which is to diagnose faster, and get treatments to rare disease patients faster. This is an area where patients’ families, patient advocacy groups can really help make a difference, and it is so important to get it right. There are many initiatives already happening, and I would like to encourage everyone to continue.

Tim Guilliams:
“This is an area where patients’ families, patient advocacy groups can really help make a difference.”

Domenica Taruscio:
There is a breakthrough to share with you from Italy. Since 2021, the National Registry for Rare Diseases, which is a national registry has collected information on more than 500 different rare diseases. This was for public health. We have a patients come to us, at Instituto (an Italian Public Health Institute) and say, “Listen, I would like as a patient, or an association to better understand our disease”, or, “Please can you help us to build a high quality registry?” From these requests, we have started ten specific rare diseases registries, for example, Prader Willi Syndrome and, cystic fibrosis. These registries have been started because patients want to better know their diseases, and improve the quality of their lives. This is another perspective you should be aware of – driven by patients. It is not the old perspective of collecting data like 20 years ago, by experts, epidemiologists, and statisticians. Now it has completely changed and the patient community comes first. The changes are being made first by patients and associations, and second by experts, stakeholders, expert teams and institutions.

Sean Sanders (host):
David, maybe you can just briefly explain for the audience what are natural history studies and how might they be used?

David Pearce:
I will use the example of a physician I met years ago, when he said, “I’ve only seen one child with Batten disease before.” So the natural history study is, first of all to examine and follow a patient over a period of time to become familiar with the symptomology and phenotype of the disease. I want to give a shout out to patient advocacy groups because when these groups meet, it is a fantastic opportunity for a physician to interact with group of individuals with the rare disease on a routine basis, follow them and see how the disease is progresses - that has what we have done that with Batten disease.
A natural history study can be as simple as, “Okay, so how are you feeling today? And how are you feeling tomorrow? And how are you feeling the next day?” in terms of what the phenotype is. This is regardless of whether it is a neurological disorder, kidney disorder, or liver disorder. A natural history study is tracking the progression of a disease and the effects it has on the individual. And then you can graph those symptoms and pinpoint exactly where somebody is during the course of this disease. This becomes incredibly important when you want to put a clinical trial together. A natural history study help is track the progression of a disease and the effects it has on an individual. When you have an intervention, you can see if you are actually slowing that progression or stopping the disease. A natural history study is incredibly important for understanding treatment options and their effectiveness.

Sean Sanders (host):
I would like to talk about treatments. In the research and development of treatments, what are some of the major barriers and the specific opportunities to eliminate these barriers?

Tim Guiliams:
There are a number of barriers and one of them is facilitating work with the patient communities. This is really important because patients know a great deal about what matters to them, what they have tried before, what did not work, and who is the expert. Patient groups can open all the doors. Healx was started because of a person called Nick Sireau who lives in Cambridge UK, and his son, Julian, was diagnosed with a retroviral disease called Black bone disease. We met Nick in 2014. He was on a journey to find a different treatment. He was trying to purpose raw chemical components to treat his children. This was successful and the treatment has now been approved by the EMA. This is an inspiration story of a super human, Nick, who has a child with a rare disease, but no scientific or AI background. They do not take no for an answer, they are going on a mission to find a treatment, and then they succeed, which is incredible. One of the key barriers for therapies is to partner with patient groups.

The second barrier is a more complex: the drug discovery paradigm. Traditional drug discovery is basically target-based drug discovery. It is based on one disease, one target, one drug. Human disease biology is more complicated than that, and so, particularly for rare disease patients, the industry massively under-delivers. You have 7,000 rare diseases, 95% of which do not have an approved treatment. Behind that is an awfully simple drug discovery paradigm. Because of this you end up with a very large failure rate, amazingly long development timelines and high costs, because you started with the wrong hypothesis in the beginning. It is important for companies to come up with new drug discovery paradigms that are more combinational, more complex, where you do not know why a particular drug is matched to a rare disease, but your algorithms identify that those two drugs could potentially help that rare disease. This type of novel approach is pioneering the next generation of drug discovery that is no longer based on ‘one disease, one target, one drug’.

Marta Mosca:
One of the major barriers are the numbers of the patients and the fact that these patients are scattered everywhere, so it is really difficult to make sense to find these patients and to treat them in a trial. Even if we had a treatment that we think might be very useful, we need numbers. There are solutions for this that include different techniques and statistical analysis which can be used. We can learn from the COVID pandemic. Maybe in the future, we might enroll patients into trials, even patients living far away using virtual care and telemedicine or on a virtual platform. We need to have networks of experts sand GPs. With these networks and virtual medicine platforms, we might be able to increase trial enrollment. This may enable patients to take an advantage of new treatments. I think this could be one method to improve the way we are working in this field.

David Pearce:
The human genome is around about 3.2 billion base pairs, and we maybe understand about 1% of it right now, and many rare diseases are of a genetic basis. One single change in that 3.2 billion letters can result in a rare disease, if you have two copies of that change. Figuring out how one missing piece, one genetic defect can result in a devastating neurodegenerative disease is a real challenge when we do
not understand how the brain works normally in its entirety, we do not understand the biology of something. To try and fix something, when you do not actually have a table of contents and a complete list of all the components, and what they do is very hard.

“It’s hard to try and fix something, when you do not actually have a table of contents and a complete list of all the components, and what they do is very hard.”

It is difficult to fix something when it is broken, and when you do not even have any of the tools and/or instructions to go with it; it is always going to be very difficult. The difficulty is exacerbated by the rarity of the diseases and the complexity of understanding what a single gene can do in a complicated human being. It is an exciting time on the basis of genetics right now. Specifically fixing that fundamental defect is certainly on the horizon and a reality in some cases right now. The bottom line is that there is a long way to go to understand the human, and the human condition.

Domenica Taruscio: I like the discussion about a new paradigm. Let us also speak about changing the EMA and the FDA [medicine regulatory bodies]. Sponsors have incentives, for example, at a European level, the FDA and in Japan. But are these incentives enough? Or are they the most appropriate? Maybe it is time to open a discussion about this because after 20 years many things have changed. I think it is time to have an open discussion on these new approaches.

Sean Sanders (host): This webinar is aimed at a general audience and particularly patients with rare diseases, the families of patients and rare disease organizations; what can they do? Is there any advice you can give them on how they can play a role, more of a role, particularly in clinical trials, getting clinical trials set up for new treatments?

Marta Mosca: Patients groups are very involved and they are essential. The paradigm has shifted thanks to patient organizations. As doctors and researchers, we need to understand this. We work for patients, and patients need to know this. We are there for them. So what should they do? They do many things already. Engage us, discuss with us and share you ordeals with us. For example, we work on patient pathways and we build patient pathways based on a patient’s point of view, not the physician’s point of view. I am a clinician, this is the very first step. Patient groups tell us, work with us and tell us about your needs, share with us your data, and guide us.

Sean Sanders (host): Thank you once again to our fantastic panel and to Fondation Ipsen for enabling this conversation through their kind sponsorship.

Journal Club

Article of the month

How to design a registry for undiagnosed patients in the framework of rare disease diagnosis: suggestions on software, data set and coding system


With approximately 30 million people in the EU and USA suffering from a rare disease, national strategies are being developed to improve medical care in rare diseases. There has been a call to develop registries of undiagnosed patients to accelerate diagnosis. This paper describes how such a registry for undiagnosed patients could be built and the information it should contain. The paper describes an open-Source Registry System for Rare Diseases (OSSE) to build the registry for undiagnosed patients based on the minimal data set for rare disease patient registries recommended by the European Rare Disease Registries Platform.

The next step, the authors suggest, is to implement the registry in medical centres for rare diseases. This could help rare disease diagnosis.
History Corner

Annals and Analytics: The Practice of History in the Age of Big Data

by Professors Ruis and Shaffer
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5426307/

"As much as historians would like to add a time machine to their toolkit, research in history is not problematic only because of insufficient data, nor because most extant data have passed through the filtration and interpretation of second-hand observation. On the contrary, for many historians the sheer quantity of available information – what William Turkel terms the ‘infinite archive’ of digital materials – cannot be processed using traditional methods alone. Far from solving this problem, a time machine would exacerbate it, adding more and richer data into the mix. In addition, there are important historical questions that cannot be answered solely through close readings of texts or through direct observation of the past; both cases overestimate the historian’s powers of observation, implying that critical analysis and ethnography can solve all historical puzzles.

Yet it is also dangerous to assume that more or more accurate data will necessarily lead to better understanding. The view that computers can take massive amounts of information and do most of our analytic thinking for us, a belief embraced by many data miners and glorified by tech evangelists, often yields statistically significant but conceptually meaningless results. We can and should outsource some of our thinking to smart machines, much as we have outsourced some of our memory to books and other media; but to do this well is to understand the limitations and leverage the affordances of different approaches to processing and analyzing information, both human and machine."

What’s up?

Highlights of the month

Movies to move you
by Cléa Stemitsiotis

Big Data is changing the landscape of our world and is the next frontier for rare disease diagnosis. The impact of Big Data on our lives has rightfully inspired many movie directors. Here is my selection of movies - exploring this new reality - that stand out:

1. Bennet Millie’s Moneyball (2011) demonstrates the amazing power of predictive analytics through the story of Oakland A’s general manager Billy Beane’s (played by Brad Pitt) successful attempt to put together a baseball team on a budget by employing computer-generated analysis to draft his players. The movie shows the importance of data in decision-making and the incredible potential of predictive modeling.

2. Far ahead of its time, Minority Report (2002), depicts how Big Data can be effectively used for predicting human behavior. Steven Spielberg’s film is based on famed science fiction writer Philip K. Dick’s story of a special police unit able to arrest murders before they commit their crimes. They base their decisions on the work of the Pre-Cogs, a small team of data scientists capable of predicting future crimes by analyzing large data sets.

3. In The Imitation Game (2014), Alan Turing (played by Benedict Cumberbatch), father of modern computer science, cracks the Enigma - a strategic code that the Nazis use to encode their messages - by creating a primitive computer to analyze all possible permutations.

4. Bonus: 21 (2008) is a fact-based story about six MIT students who were trained to become experts in card counting and head to Las Vegas to win millions. The movie shows how data-based decisions can be profitable for any business. Worth having a watch for all math and data fanatics out there!

5. Movies about big data: https://www.mygreatlearning.com/blog/top-data-science-movies/