Virtual Network Meeting - May 19, 2023

Phenotypic and Genetic Variation of Neurodevelopmental Conditions in Kenya and South Africa.

Patricia Kipkemoi (PhD student, Vrije Universiteit, Amsterdam)

Thank you, Bosiljka, for inviting me to this forum. As you've heard, my name is Patricia Kipkemoi. I'm a PhD student based in Kilifi, Kenya. Kilifi is this beautiful place that you are seeing on my screen. It's on the coast of Kenya, and it's a really lovely, quiet town. I'm based at the KEMRI-Wellcome Trust, which is in Kilifi, and I'm registered for my PhD at Vrije University Amsterdam. My background is in developmental psychology, and I'll be speaking with you today about neurodevelopmental disabilities in Africa, specifically focusing on a project that I'm working with called NeuroDev (which is a study that aims to factorize the genetic and phenotypic architecture of neurodevelopmental disabilities in South Africa and Kenyan populations). I'll share a bit of our findings from the Trio Pilot study. I'll be referring to neurodevelopmental disability as NDDs sometimes, so please pardon me if I switch between the abbreviations. I hope you can hear me well, and I'll just go ahead and start the discussion.

First, I'll give us a brief introduction to neurodevelopmental disabilities, although I'm sure many of us are versed in what they mean. I'll talk a bit about their context in Africa, I'll talk about the challenges in identifying NDDs in our setting, and the rationale for setting up the NeuroDev study. Then, I'll also talk about the methodology we employed in looking at our study and some of the findings, and wind up with some concluding remarks. We'll talk about our responsibility to the community and our way forward, and we can finish up with questions and a discussion section.

I can imagine we're all familiar with neurodevelopmental disabilities, but for our work in NeuroDev, we're mostly focusing on the APA definition of neurodevelopmental disability (that's the group of disorders with symptoms appearing during the developmental period). So, we're looking at Autism, ADHD, Intellectual Disability, Global Developmental Delay, Communication Disorders, and Specific Learning Disorders. We know that developmental differences are quite varied, and some children might have specific challenges in learning, concentrating, social skills, and cognitive ability. For one to meet the thresholds of Developmental Disability diagnosis, these differences usually have to produce challenges in the personal, academic, social, or occupational domains of functioning. We know that NDDs frequently co-occur with one another. So, for example, we might have a child who's autistic, and they might also have an Intellectual Disability. A child might have ADHD with a Specific Learning Disorder (and they also do sometimes frequently with other neurological conditions). In our cohort, we have a number of children with co-occurring epilepsy and other seizure disorders. We also know that the neurodevelopmental conditions sometimes also co-occur with other mental health conditions. So, a child with ADHD might have some anxiety challenges, they might, during adolescence or even in adults, develop mood disorders. So that's what we mean by the co-occurring nature of neurodevelopmental conditions.

Then we also know that NDDs have been termed as very heterogeneous (in their terms of the clinical characteristics, their causes, the treatment responses, and outcomes), meaning that one child might respond very well to a particular type of behavioral therapy, or not. They might respond well to a particular medication, let's say for ADHD, or not. And some of the clinical characteristics can vary in severity and how much they affect the child and their daily functioning.

So, with the factors associated with developmental disabilities, twin and family studies have demonstrated consistently that there's a shared genetic and environmental factor or contributions to these conditions. These risk factors are seen as contributory, not necessarily causal, but they could be reactive, independent, or contributory to neurodevelopmental disabilities. So, genes such as DDX3X, SHANK2, SCN2A, have been put forward and implicated with some neurodevelopmental conditions, some environmental factors, like advanced parental age, birth trauma (particularly hypoxia), infection during pregnancy… For example, in our setting, malaria is an infection of interest that seems to be associated with some developmental conditions. There are relatively higher rates of infectious diseases and prenatal complications in lower- and middle- income countries, such as where I'm based. So, it's important for us to also think about environmental factors that are contributory to these conditions.

Then I'll just talk a bit about the Global Burden of Disease project. So, the Global Burden of Disease project, by the WHO and the Institute for Health Metrics, has been mapping out the prevalence of a number of health conditions, including some neurodevelopmental conditions. There's a lot of criticism out there about the Global Burden of Disease study, one of them being that a lot of the estimates that they have about lower- and middle- income countries (LMIC) are made from very little data and many African countries fall into the LMIC bracket. So, in this map, for example, you can see we have the prevalence estimates of autism, which is one of the neurodevelopmental disabilities of interest, and it appears to have very low prevalence in many parts of Africa, many parts of Asia, and we know that from previous research, the global prevalence is around one to two percent. However, this map, because of the very little data that exists in our context, gives a very low estimate, which is more than likely an underestimate.

A majority of the world's children with developmental disabilities live in lower- and middle-income countries, and especially in Africa. However, we don't know much about the underlying etiology or causes or the profiles of autism and other developmental disabilities in many countries in Africa. There are very limited case control studies, and identifying environmental factors associated with autism, for example, pregnancy complications and adverse perinatal events (which are again, like I mentioned, a bit prevalent in our setting because of our resource constraints and limited access to health care, especially for pregnant women) and less than one percent of the world's autism research has taken place in Africa. This is a systematic review by Franz and colleagues that was done in 2017. I'm sure there's a bit of improvement since then, but I would imagine we're still under the five percent limit there. So, in terms of the prevalence of NDDs in Africa, as we've mentioned, there's limited data about the prevalence. Since 2015,

there have been two large-scale prevalence studies done in Africa. There was one done in Kampala, in a peri-urban setting, with a locally developed tool that was fashioned after the 10 questions questionnaire. They surveyed about 1169 children, and they found eight children with autism, so that's an adjusted prevalence of about 0.68 per 100. And, here in Kilifi, where I'm based, we did a survey of about 11260 children in a study called the NDD study. The data is still preliminary at this moment, but we identified about 92 children with autism. So, just to break it down a bit further, if we were to look at the crude prevalence rates, for autism we have about 7.1 per thousand (that's about 0.7), but if we were to adjust for attrition and even adjustable sensitivity, this goes up to about 1.3 percent. For ADHD, we see that it's about 7.39 percent, which map on really closely to what's out there in terms of global estimates. So, this work is still under preparation, but hopefully we'll be able to share the findings soon. Next, I just wanted to talk a bit about identification of NDDs in Africa, which can be a bit challenging for many reasons. First, is lack of awareness of developmental conditions by health and education staff and even by the public, especially in areas such as Kilifi, which is a rural community. There's also a lack of expertise, so we have very few psychologists, very few Child and Adolescent psychiatrists, with many of those who are actually practicing, practicing in urban areas as opposed to rural areas. We have a lack of validated screening and diagnostic tools, with many of these tools having been developed in the global North. So, before they're used in our setting, we need to adapt them for our setting. There's also a lack of affordable and sustainable and culturally homogeneous interventions that we can use in our setting.

I'll talk about the rationale for our study, NeuroDev, which is a genetic study of Developmental conditions. We know historically that genetic studies have relied heavily on samples of European ancestry. It's only about three percent genome-wide association studies including participants of African ancestry, so genetic studies in the NDD space follow a similar trend. So, for example, studies such as the Autism Sequencing Consortium, the Simon Simplex Cohort, the Spark study in the U.S, iPsych in Denmark, DDD in the UK are predominantly European. However, there is some diversity with the inclusion of people, let's say of Hispanic, Asian or African ancestry, but even with this, there's still a need to increase the diversity of participants in genetic studies. So that's where projects such as NeuroDev, and even DDD Africa, come into play, by including participants of African ancestry living on the African continent in genetic research for scientific and even ethical equity.

NeuroDev aims to collect and analyze extensive genetic phenotypic data from about 2 000 children over the next few years. Specifically, we aim to identify genes and genetic variants that are associated with neurodevelopmental differences in Kenya and South Africa. We want to characterize the phenotypic spectrum of our neurodevelopmental variation in our setting. We also hope to generate some exon data from controls (so these are children who are of matched ancestries as our cases), just to address the missing critical piece that is there in our current genetic reference databases, which are, again, leaning heavily towards people of European ancestry. There's also the aim to enhance neurodevelopmental research capacity in Kenya, and in South Africa, by training our clinicians and junior researchers (such as myself). We have two main recruitment sites. One is in Kilifi County, which is along the Kenyan Coast in collaboration with the KEMRI-Wellcome Trust research program led by Amina Abubakar and Charles Newton

as the PIs. Kilifi is in a health and demographic surveillance site and, as such, there's a bit of expertise in the measurement of cognitive and behavioral variation through a number of studies that have happened in this demographic site. Recruitment is done from existing registries from previous studies. We also engage with hospitals (especially the neurology clinics or the occupational therapy clinics), we engage with special education schools, and because Kilifi is along the Kenyan Coast, many of the participants are of Mijikenda or Swahili ancestry. The second site is in Cape Town, in South Africa, through the University of Cape Town and Red Cross Children's Hospital, led by Kirsty Donald. Our sample population in South Africa is quite diverse and includes people of mixed ancestry, Xhosa, and Afrikaans backgrounds. Then the team in Boston is headed by Elise Robinson at the Broad Institute who, in collaboration with all the other sites (KEMRI and University of Cape Town), we oversee aspects of data management, protocol alignment, and general study management. This is just a QR code of our website so feel free to scan it and I think Bosiljka also shared that with you. Feel free to have a look and you can learn more about the study and the tools that we use (which are actually available on the website).

So, more about the tools. NeuroDev deeply phenotypes all participants, with a range of assessment tools, and you can see them on our website. We have the demographic questionnaire, we have an asset index that is specific to Cape Town and also to Kilifi, we also have a Neuromedical Assessment, we use the Multeno Development Scales, and the Ravens Progressive Matrices to look at Child Development, cognition and non-verbal reasoning. We have tools such as the Social Communication Disorders Checklist (the SCDC), the M-Chat, and the 3Di, which look at autistic traits. We also have the SNAP which looks at ADHD traits, and we have the Child Behavioral Checklist (the CBCL), both preschool and school age, looking at internalizing or externalizing behavior.

So today, I'll just focus a bit more on the Trio Pilot, which is the genetic and phenotypic methods and findings from our pilot study which was which is data from our first year of collection of the project, which is around 2018 to 2019. We began data collection in South Africa in late 2018, around September 2018, and in Kenya we began in February 2019. The Trio Pilot concluded around July 2019. As you can see in this table here, the first year we enrolled about 219 cases, 195 mothers of the cases, 115 fathers and 92 unrelated child controls. We sequenced a hundred parent-child trios (so that's Mom, Dad, child) and included 99 trios in the pilot study. We had one family where they did not pass sample QC so we weren't able to use the data - at least the genetic data. We can see that the male to female ratio in cases is about 2.4 males to one female, which sort of tracks with what's out there in terms of neurodevelopmental conditions. The mean age of our cases is around seven and a half years, with our South African cohort having much younger children and our Kenyan cohort having older children.

The findings are discussed here, after the 99 parents-child trios. So, looking at the diagnostic overview, we have a majority of our children with autism and co-occurring Global Developmental Delay or Intellectual Disability. We have a number also with sole GDD or ID, a few with sole autism, and 5.5 percent with other NDDs (so these constitute Specific Learning

Disorders or Communication Disorders). So, looking at the children with autism in our cohorts specifically, so we looked at those children and counted the proportion of cases by the number of co-occurring adverse developmental outcomes (so that's whether the child presented with Intellectual Disability, whether they had seizures, and whether they had delayed walking). With the presence of one of each meant that they were added to the different groups here; so, if they had one of these co-occurring adverse events we added them here, if they had two, we added them there, to the two-count group, and if they had three of these adverse events, we added them to the three-count group. We conducted a similar analysis with the Simon Simplex cohort, which is our cohort based in the US. As you can see in the figure here, a number of our NeuroDev cases had at least one of the severe phenotypes included and far more NeuroDev cases had at least two or three of the counts in comparison to the Simon Simplex cohort. So, this shows us that there's a bit of an ascertainment bias in our population that results in having more severe cases being included into the study. This means that they have more co-occurring or comorbidities when they're enrolled into the study.

I'll just go ahead and talk about the genetic methods and findings. We collect samples from our participants, and we keep one sample in country. Following the H3 Africa guidelines, a part of the second sample is sent to the US, so our collaborators are at the Broad Institute. So, samples from both sides are shipped to the NIMH Biorepository at Rutgers University, where the DNA is extracted and finally makes the way to the Broad Institute, where the samples are genotyped and sequence of the genomics platform. Then, genetic analysis and variant calling is done by the Center for Mendelian Genomics team.

In terms of genetic findings, we currently have data from 75 trios in South Africa and 24 trios in Kenya. So, from the 75 trios in South Africa, we had a solved finding in 13 of the 75 trios. This means that we found a pathogenic, or likely pathogenic, variant in a known gene associated with neurodevelopmental disability in these individuals. We also observed five cases with a novel variant in a gene that has not otherwise been associated with neurodevelopmental disabilities. Of the 24 Kenyan trios analyzed, we had 10 of these cases with a solved finding and two cases with a novel variant in a gene not otherwise associated with a new developmental condition. These novel variants were submitted to a platform called Matchmaker Exchange, and we had about seven successful matches with other collaborators in different parts of the world. These are the gene variations that were implicated, so that's AG01, MYH10, CACNA1C, MAPK1, PPP2R5C, CACNA1E, SF1. I'm not a geneticist, so I'm not too sure about the biological pathways implicated here. MAPK1 and SF1 were noted in our two Kenyan cases, and I think the MAPK group of genes has been implicated in a condition called RASopathies. They sort of -the biological pathway is in protein transcription- so it kind of makes sense that if there's a challenge in how proteins are being made, that there might be a developmental challenge. Some of these have been published in certain case series, with a few more still in publication. I'm happy to share some of some of the papers that have already been published about these cases.

We also have a few single gene variant diagnoses in the cohort. These include variation in genes such as DDX3X, CREBBP, SYNGAP1, SCN2A, TLK2, MBD5, among others. It's important to note

that these variants are all de novo (they were not inherited), with something spontaneous that happened in the genome, and this is what seems to have been associated with the condition. I'll talk a bit more about the DDX3X diagnosis. It's associated with syndromic intellectual disability of the Snijders Blok type. From OMIM (which is an online catalog of human genes and genetic disorders), we can see that this variant is associated with a specific phenotype (that's differences in growth of the head, the neck, there's skeletal abnormalities, differences in skin, hair and nails, with some muscle, and soft tissue differences, some neurological differences). The features marked red are some of the features that we noted in our case: so low weight, they had dysmorphic facial features, especially in the midface, intellectual disability, and some behavioral challenges such as hyperactivity. At the moment, we're not able to share the photos of our case. We usually have to reconsent the participants before we can publish the photos, but this is just an example of the differences we also noted in our case. For example, in the midface.

We also found some structural variations, specifically copy number variants that are linked to neurodevelopmental conditions in our cohort. Some of these syndromes are very well known and have been discussed in literature. For example, 22q11.2 deletion has been associated with autism and other developmental disabilities. It's also important to know that most of these variants were de novo, except one which was inherited (and actually paternally inherited). Then I'll just highlight briefly one associated syndrome - so that's the 6q interstitial syndrome. Apologies for the many genetic terminologies, I'm also learning. It's pretty interesting to see that there's some convergence in what we're finding in our participants as well. So for the interstitial 6q micro deletion syndrome, that's been associated with neurodevelopmental conditions and has been thoroughly researched in other parts of the world. Individuals with this particular syndrome might have different phenotypic characteristics; they have cardiovascular defects, genitourinary challenges, head or neck challenges, nervous system issues, skeletal issues… The bits marked in orange are what we saw in our participants; so, they had a ventricular septal defect, also an abnormality of the midface, they have delayed speech, very delayed walking, delayed fine motor development, seizures, and microcephaly. So that's it for the findings that we've had so far.

So, what's next for NeuroDev? We're still collecting data in both the South Africa and Kenyan site. The South Africa site is set to wind up data collection in the next few months. For Kenya, we still have quite a bit of a way to go! We are hoping to wind up data collection at the end of 2025. We are still continuing with data analysis. We'll be sharing data in what we call waves. So right now, we've completed the Trio Pilot Wave and the manuscript has been accepted at Neuron, so once that's published, you'll be able to access some of this data through a resource called AnVIL (which is a controlled-access database). A number of trainees of the project also working on various manuscripts.

One of the things that I wanted us to maybe think about and talk about, from our experiences, is carrying out a genetic study in settings where genetic studies are not very common. We feel a very big responsibility to our community because we're dealing with families that have conditions that have previously been stigmatized in our setting / are still stigmatized in our

setting. We even have cases where children with severe developmental challenges are hidden, they are not able to access health services, education services... so we needed to think very strongly about the ethical and clinical responsibilities to our participants. How we think about it is our contribution to the community. First, we ensure that our participants are well informed about the study's aims before consenting; so, we do carry out a lot of community engagement. Thankfully, KEMRI has a very strong community engagement group so we're able to go into the community, engage with people in barazas (which are like community meetings), we go to hospitals, we go to schools, we meet people where they are and talk to them about generally all the research that happens in KEMRI and then also about the NeuroDev study. And because we're talking about genes, we, as a team, we sort of came up with an interesting way to talk about genes: we call it the 'Story of Maize' (I'll talk about it in just a bit). But another thing that was important for us, is to think about the assessment of capacity to consent, because we will be biobanking some of these samples for quite a while, so we needed to make sure that our participants truly understand the implications of their participation and what the study is truly about. So, we use the UBACC, which is the University of California Brief Assessment of Capacity to Consent. Our participant needs to get a particular score, so that's about 15 out of 20 - to get enrolled into the study, and we do it in sort of a conversational way, so if there's a question that maybe they're not answering correctly or they don't seem to have understood the parts of the protocol, we go over those in the information sheet again and again and again. We have up to four trials where we get to engage with the participants.

I'll talk about the Story of Maize briefly. We were thinking about a way to make genes more accessible to our participants because Kilifi is a very rural setting, there's very low literacy levels, and we still wanted to make sure that our participants understand what genes are and we try to make it as accessible as possible. So, we talked about it in terms of maize, so we call it the 'Story of Maize'. We usually ask “what does it take for a maize seed to become a maize plant?” And of course, there's air, nutrients, water... that's what our participants usually answer in our group, whenever we have our group discussions. We always say that a maize seed, with the right air, nutrients, water always becomes a maize plant. So there seems to be instructions in the maize that tell it to become a maize plant. So those are the building blocks that are in the base seed that give the maze instruction to become the plant. So that's how we sort of talk about genes there and then we talk about the different seeds and different varieties; we have specific local maize seeds, and they look quite different depending on where you are. We talk about the variation that even exists in plants; sometimes you might have a maize plant that's white with a bit of purple seeds in the cob. So that's an example of variation, and the degree of variation sometimes is quite high, or it can be quite low. So that's a way that we talk about traits and inheritance of traits. And then we also talk about how these instructions are coded; if you think about genes, there's the building blocks which is A, C, T, G. So, we talk about the instructions being in a coded way and how if – so the red is in Kiswahili, but you can just look at the English. We speak a lot of Kiswahili at the coast, or Kigiryama, which are the local languages. So, we just talk about how sometimes these mutations or changes in the code changes the instructions and therefore the instructions are understood differently. So, what's in the table here are examples of Kiswahili phrases. You can see that if you remove a particular part of the phrase, the meaning changes. If you add something, the meaning

changes. If you add a really big block, the meaning totally changes. So that's how we sort of discuss mutations or variation and how they lead to different characteristics and meanings. So, that's a brief of what we discussed there.

Another way we think about our responsibility to the community, is we sometimes have parents who know their child is developing differently, but they're not too sure what the condition is. We usually have debriefing sessions after the assessments, where we share some management tips that are sort of borrowed from the WHO caregiver skills training that we share with our participants. We also do our best to refer our participants appropriately for medical assistance. In Kilifi, we do have a neurology and developmental clinic that's free for the population: they come in and they are seen by a clinician, there's no consultation fee, and there's also no fee attached to the medications. That's our way of contributing to the community as well. We also train Community Health volunteers on the identification of neurodevelopmental conditions in the community health units. We also do this, so in South Africa they do return clinically relevant findings. However, in our setting, because we don't have similar resources, such as medical geneticists or genetics counselors, but we have found some a few participants who have clinically actionable findings. So, we're now having discussions on how to return these findings in a low-resource setting. So those are some of the ways that we're going about it, and of course we have a number of referral pathways that we try to utilize. We have specialists that might be available, we have education assessment resource centers in each sub-county in the country... So, whenever we have a child that comes in, that isn't going to school, we always try to encourage the parents to take them to school. We also partner with the National Council for Persons with Disabilities here in Kenya, where we try and group a number of families who don't have disability cards (which sort of give them a bit of reprieve in terms of like they can't get some medical aid, some education aid). So, we group our participants and partner with the National Council and they come and do mass registrations at our clinic. We also refer to advocacy groups, we encourage our parents to form support groups, and we also refer them to their teachers.

In conclusion, we've set up fairly complex genetic study in two sites on the continent. We actually have a third site in Nairobi, where I am currently out at Aga Khan University. The Trio Pilot highlights the success of the first year in that we're now sharing some of our findings. We have high quality data, there's a good number of hits in terms of genetic findings, there's clinical utility through the return of results, there's also scientific utility through data sharing, and of course we are continuing with data collection in the three sites and hopefully we'll be sharing more findings in the coming years.

NeuroDev is a tremendous, tremendous effort from so many people, so many teams. So many thanks to our teams, our PIs, the team in the US, in Kenya, and South Africa, funding bodies, and of course a special thanks to our participants.

So that's it for my presentation! I also wanted to do a quick plug of the Psychiatrics Genomics Consortium. It's a well-known Consortium that works on psychiatric genetics. However, as

we've seen through this presentation, and as we know, there isn't as much work (especially genetic work) being done in underrepresented populations / minority populations.
The PGC has an African working group, and the group is focusing on facilitating equity with African researchers into the world of psychiatric genetics. There's an Autism and ADHD working group. Myself and Celia van der Merwe (who's also on the NeuroDev study), we sort of coordinate these working groups. So, you are very welcome to join the group, and this is a QR code of the website. Please feel free to scan and have more info and learn more about the PGC Africa working group. And you can also reach out to me, this is my email, and I'd be able to share more information (and also add you to our working groups). Thank you so much for listening, and I welcome any questions.