Project Notes:

Project Title:
Name: Agarwal, Naaisha

Note Well: There are NO SHORT-cuts to reading journal articles and taking notes from them. Comprehension is paramount. You will most likely need to read it several times, so set aside enough time in your schedule.

Contents:

Knowledge Gaps:

This list provides a brief overview of the major knowledge gaps for this project, how they were resolved and where to find the information.

Knowledge Gap	Resolved By	Information is located	Date resolved
AUC-ROC Curves	Searching up an explanation	https://www.geeksforgeeks.org/machine-learning/auc-roc-curve/	10/5/25
Linear regression	Searching up an explanation	https://www.ibm.com/think/topics/linear-regression	10/5/25
Cox proportional hazards model	Searching up an explanation	https://www.geeksforgeeks.org/data-science/cox-proportional-hazards-model/	10/5/25

Literature Search Parameters:

These searches were performed between (Start Date of reading) and XX/XX/2019.
List of keywords and databases used during this project.

Database/search engine	Keywords	Summary of search
WPI library	Gut-brain axis	Found various papers identifying different factors of the gut brain axis and how it affected mental health
WPI library	Digital twins	Found papers about personalized surgery, data, and models
WPI library	Gut microbiome	Found papers related this to the gut brain axis, and mental health
WPI Library	Migrants Environment Health	Found papers that related to how health effects changed based on different environments migrants lived in
WPI Library	Effect of change on environment on health	Found many papers on the effects of climate change on people and also effects of moving on people's health

Tags:

Tag Name	
#machine-learning	#environment-for-health
#disease-detection	#model-organism-study

#methodology	
#neural-decoding	

Article #0 Notes: Template

Article notes should be on separate sheets
**All citations are included at the top in order to have proper formatting to copy and paste and reference. However, the citations are also included in their respective boxes in order to remember the context for each one, especially the follow up citations. The format in the table is not the correct formatting for APA citations, which is why the proper citations are included above.

KEEP THIS BLANK AND USE AS A TEMPLATE

Source Title	
Source citation (APA Format)	
Original URL	
Source type	
Keywords	
#Tags	
Summary of key points + notes (include methodology)	
Research Question/Problem/ Need	
Important Figures	
VOCAB: (w/definition)	
Cited references to follow up on	
Follow up Questions	

Article #1 Notes: Detection of Disease Features on Retinal OCT Scans Using RETFound

Article notes should be on separate sheets

Citation:

Du, K., Nair, A. R., Shah, S., Gadari, A., Vupparaboina, S. C., Bollepalli, S. C., Sutharahan, S., Sahel, J.-A., Jana, S., Chhablani, J., & Vupparaboina, K. K. (2024). Detection of Disease Features on Retinal OCT Scans Using RETFound. *Bioengineering (Basel, Switzerland)*, *11*(12), 1186. https://doi.org/10.3390/bioengineering11121186

Related Works Citations:

Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., & Zeng, X.-J. (2018). A Bibliometric Analysis and Visualization of Medical Big Data Research. *Sustainability*, *10*(1), 166. https://doi.org/10.3390/su10010166

Andreu-Perez, J., Poon, C. C. Y., Merrifield, R. D., Wong, S. T. C., & Yang, G.-Z. (2015). Big Data for Health. *IEEE Journal of Biomedical and Health Informatics*, *19*(4), 1193–1208. https://doi.org/10.1109/JBHI.2015.2450362

Source Title	Detection of Disease Features on Retinal OCT Scans Using RETFound
Source citation (APA Format)	Du, K., Nair, A. R., Shah, S., Gadari, A., Vupparaboina, S. C., Bollepalli, S. C., Sutharahan, S., Sahel, J.-A., Jana, S., Chhablani, J., & Vupparaboina, K. K. (2024). Detection of disease features on retinal OCT scans using RETFound. Bioengineering, 11(12), 1186. https://doi.org/10.3390/bioengineering11121186
Original URL	https://www.mdpi.com/2306-5354/11/12/1186

Source type	Journal Article
Keywords	retinal imaging; optical coherence tomography; machine learning; age-related macular degeneration; foundational model; automated report generation
#Tags	#machine-learning, #disease-prediction, #methodology
Summary of key points + notes (include methodology)	Current existing methods to diagnose OCT scans include manually doing it which is labor intensive or using deep learning or CNN models to classify them which requires an immense amount of data and processing power. This study proposes a foundation model RETFound that requires unlabelled data and a smaller amount to be able to improve on the accuracy of previous methods. They found that their model was able to perform significantly better than the state of the art CNN models on both single task and multi task situations, and it required less data and the data was all completely unlabelled. Overall this will help make models to analyze OCT scans more available so doctors need to put in less manual labor in order to accurately diagnose scans and patients can get their diagnosis fasters in order for more time for preventative action.
Research Question/Problem/ Need	Current existing methods to diagnose OCT scans include manually doing it which is labor intensive or using deep learning or CNN models to classify them which requires an immense amount of data and processing power, both of which are not effective and efficient methods.
Important Figures	1770 Scans in Dataset – This was the number of labelled scans in the dataset used to train RETFound. This is important to understand the scale of the data that was needed and to consider the effects of this amount.
VOCAB: (w/definition)	Foundational model: large pretrained model designed to understand general patterns in data Encoder: part of neural network that processes input data and converts it into a numerical represenation Drusen: yellow or white deposits that accumulate under the retina, often assoicated with age-related macular degeneration PED: pigment epithelial detachment – condition where retinal pigment epithelium (RPE) layer separates from underlying tissue Hyperreflective dots: tiny bright spots on OCT images caused by depositis, cells, or debris in retina Subretinal fluid: fluid between retina and RPE layer Intraretinal fluid: fluid within retinal layers
Cited references to follow up on	Andreu-Perez, J.; Poon, C.C.; Merrifield, R.D.; Wong, S.T.; Yang, G.Z. Big data for health. IEEE J. Biomed. Health Inform. 2015, 19, 1193–1208. https://10.1109/JBHI.2015.2450362

	What are some sources to get data for health and how are they being utilized? Liao, H.; Tang, M.; Luo, L.; Li, C.; Chiclana, F.; Zeng, X.J. A bibliometric analysis and visualization of medical big data research. Sustainability 2018, 10, 166. https://doi.org/10.3390/su10010166 What are the current advancements in using data in health and what are some gaps?
Follow up Questions	- Can this foundation model be used to diagnose other diseases than AMD? - How early can this model accurately predict the likelihood of a person getting the disease? - Can this foundation model use pictures of eyes (such as ones taken with smartphones) to classify the disease or only OCT scans?

Bulleted Notes:

- Reading OCT scans manually takes a lot of effort for doctors
- Current methods for deep learning or neural networks take up too much data that is unrealistic
- RETFound is a foundation model (more generalized for more tasks)
- It needs less data and all the data can be unlabeled
- Automate detection of pathological features
- Early detection of these features is crucial for identifying and diagnosing diseases sooner
- Wants to see if improved compared to normal CNN approach (ResNet-5)
- data source: Cirrus 5000 OCT device
    - still uses OCT – can the model take other modalities of data and still perform as high
    - if not, can other models perform as high?
    - What causes the lower performance?
- 110 unique eyes in the dataset
    - Is this a large enough dataset
    - Does this contain lots of different people and eye types in order to reduce risk of bias
- Annotated features like healthy, pigment, drusen
- Trained RETFound using a masked autoencoder approach
    - Self supervised pretraining method
- Different task setups
    - Single task models
    - Multitask models
        - What are the benefits of the different types of setups
        - Why would we need these types of setups
- Layers were fine tuned using adom optimizer
- Evaluated on metrics:
    - Accuracy
    - AUC-ROC
        - What is this?
        - How is it helpful over accuracy?

Article #2 Notes: This brain implant lets a man who lost his speech to ALS produce natural-sounding sentences instantaneously

Article notes should be on separate sheets

Citation:

This brain implant lets a man who lost his speech to ALS produce natural-sounding sentences instantaneously. (n.d.). Retrieved September 16, 2025, from https://www.science.org/content/article/brain-implant-lets-man-who-lost-his-speech-als-produce-natural-sounding-sentences

Follow Up Citation:

Wairagkar, M., Card, N. S., Singer-Clark, T., Hou, X., Iacobacci, C., Miller, L. M., Hochberg, L. R., Brandman, D. M., & Stavisky, S. D. (2025). An instantaneous voice-synthesis neuroprosthesis. *Nature*, *644*(8075), 145–152. https://doi.org/10.1038/s41586-025-09127-3

KEEP THIS BLANK AND USE AS A TEMPLATE

Source Title	This brain implant lets a man who lost his speech to ALS produce natural-sounding sentences instantaneously
Source citation (APA Format)	Reardon, S. (2025, June 11). This brain implant lets a man who lost his speech to als produce natural-sounding sentences instantaneously | science | AAAS. Science. https://www.science.org/content/article/brain-implant-lets-man-who-lost-his-speech-als-produce-natural-sounding-sentences
Original URL	https://www.science.org/content/article/brain-implant-lets-man-who-lost-his-speech-als-produce-natural-sounding-sentences
Source type	News Article
Keywords	Neural Decoding, ALS, speech, assistive technology
#Tags	#neural-decoding
Summary of key points + notes (include methodology)	In this news article, a brain implant allowed an ALS patient to speak in real time. This patient can still make sounds, but the speech isn't intelligible right now. The implant can read the brain waves and listen to the sounds of

	what the patient is trying to say. Using this, the model can be trained to analyze and predict what the patient is trying to say. Then it says the speech out loud. The big improvement with this implant is that it can produce words that the model has never been trained on, can emphasize certain words, create nonsense speech like hmm and eww, and can even sing for some time. This allows locked in patients to communicate in more real time conversations. Although, it is not perfect. This new implant has been shown to have around 60% understanding from others, which is an improvement from the 4% without any help, but not as good as the 98% understanding with the previous BCI that produced text. Another future direction that this model provides is the ability to learn new languages without being trained on them. Some things to consider also include how the implant will proress as the patient's motor cortext degenerates and how its ability to work will compare on other individuals with different speech impairments, or using more electrodes for more intelligible speech. This relates to my prospective project because I want to explore BCI and how it could possibly help with mental health.
Research Question/Problem/ Need	Need: There is a need for a way for ALS patients to be able to communicate even though they have lost a lot of their speech and mobility abilities.
Important Figures	60% of volunteers understood speech generated by the machine using this method vs the 4% of speech understood without→ this figure stood out to me because it was a big improvement from the amount of speech understood without and shows how it would improve communication for ALS patients
VOCAB: (w/definition)	Motor cortex: region of brain responsible for planning, controlling, and executing voluntary actions Waveform: shape of electrical signal recorded from brain over time
Cited references to follow up on	Wairagkar, M., Card, N. S., Singer-Clark, T., Hou, X., Iacobacci, C., Miller, L. M., Hochberg, L. R., Brandman, D. M., & Stavisky, S. D. (2025, June 12). *An instantaneous voice-synthesis neuroprosthesis*. Nature News. https://www.nature.com/articles/s41586-025-09127-3 --> more information about creating a neuroprosthesis for voice syntheis
Follow up Questions	What is currently causing the lack of complete understanding from the volunteers? How well does this work if the person changes his thoughts of what he wants to say in real time? Can this work for bilingual patients? Does the accuracy/performance change based on different brain diseases or a lack of brain disease?

Bulleted Notes:

- Researchers developed brain computer interface (BCI)
    - Also considered neuropreosthesis
- Focused on helping ALS patient who couldn't speak
    - Wanted to help them communicate directly from what they were thinking and outputting that into audio that everyone else could hear
- Something really cool is that it can generate words it has never seen before
    - I think this is really interesting because we wouldn't be able to think of new words as humans, and neural networks are modeled off of our brains
    - Maybe it does it by sound?
- Real time speech synthesis allows for live conversations → better for patient because it brings them closer to talking with others
    - Also allows for emphasis and intonation – key parts of natural speech
    - Also sings?
        - I wonder what parts of the brain are so different that it is able to pick up on all these subtle nuances
- A big drawback is that it has less accuracy than the brain to text BCIs, however still much more accuracy than the patient without the BCI
    - I think the benefit of being able to have real time conversations and all of these subtle nuances is better

Article #3 Notes: Brain Neuroplasticity Leveraging Virtual Reality and Brain–Computer Interface Technologies

Article notes should be on separate sheets

Drigas, A., & Sideraki, A. (2024). Brain Neuroplasticity Leveraging Virtual Reality and Brain–Computer Interface Technologies. *Sensors*, *24*(17), 5725. https://doi.org/10.3390/s24175725

Follow Up Citations:

Frontiers | Recovery after brain injury: Mechanisms and principles. (n.d.). Retrieved September 16, 2025, from https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2013.00887/full

Mechanisms Underlying Recovery of Motor Function After Stroke | Cerebrovascular Disease | JAMA Neurology | JAMA Network. (n.d.). Retrieved September 16, 2025, from https://jamanetwork.com/journals/jamaneurology/fullarticle/787250

Mirelman, A., Maidan, I., Herman, T., Deutsch, J. E., Giladi, N., & Hausdorff, J. M. (2011). Virtual Reality for Gait Training: Can It Induce Motor Learning to Enhance Complex Walking and Reduce Fall Risk in Patients With Parkinson's Disease? *The Journals of Gerontology: Series A*, *66A*(2), 234–240. https://doi.org/10.1093/gerona/glq201

KEEP THIS BLANK AND USE AS A TEMPLATE

Source Title	Brain Neuroplasticity Leveraging Virtual Reality and Brain–Computer Interface Technologies
Source citation (APA Format)	Drigas, A., & Sideraki, A. (2024). Brain Neuroplasticity Leveraging Virtual Reality and Brain-Computer Interface Technologies. Sensors (Basel, Switzerland), 24(17), 5725. https://doi.org/10.3390/s24175725
Original URL	https://www.mdpi.com/1424-8220/24/17/5725
Source type	Article

Keywords	Brain-computer interface, neural decoding, virtual reality, neuroplasticity
#Tags	#neural-decoding
Summary of key points + notes (include methodology)	This article starts off by outlining how VR and AR have been used increasingly in therapy treatments to create immersive scenarios to help patients face their fears among other things. It also explains how BCIs (brain computer interfaces) have been used to read brain waves and predict what the user is thinking, or anyalze the brain wave charts for other other patterns. This article set out to test how the use of BCI and VR could be combined to increase neuroplasticity in users. It created a set up of a VR simulation similar to ones done in the past. However, the user also had a BCI implanted. This BCI analyzed the patterns of the brain to see how the user is reacting to the VR simulation and adjust it in real time to better target certain areas of the brain. The results showed significant improvement in both therapy style simulations and increased neuroplasticity. This article relates to my idea of how BCI could be used to improve mental health by showing one technique and analyzing the results.
Research Question/Problem/ Need	Is it possible to utilize a combination of neural decoding and VR/AR for adaptive simulations to improve neuroplasticity?
Important Figures	
VOCAB: (w/definition)	Neuroplasticity: brain's ability to adapt and change in different situations Neurofeedback training: individuals learn to regulate their own brain activity
Cited references to follow up on	1. Nudo, R.J. Recovery after brain injury: Mechanisms and principles. *Front. Hum. Neurosci.* **2013**, *7*, 887. [Google Scholar] [CrossRef] [PubMed] → more about understanding how the brain recovers 2. Ward, N.S.; Cohen, L.G. Mechanisms underlying recovery of motor function after stroke. *Arch. Neurol.* **2014**, *61*, 1844–1848. [Google Scholar] → stroke recovery 3. Mirelman, A.; Maidan, I.; Herman, T.; Deutsch, J.E.; Giladi, N.; Hausdorff, J.M. Virtual reality for gait training: Can it induce motor learning to enhance complex walking and reduce fall risk in patients with Parkinson's disease? *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2013**, *68*, 764–776. [Google Scholar] [CrossRef] [PubMed] → use of neural decoding for non – speech purposes to help patients with Parkinson's disease
Follow up Questions	Could neural decoding be used like this for other real time therapy adaptations? What other biomarkers could be used to help in adaptive simulations?

Bulleted Notes:

- Interesting how vr and ar are used in combination
    - I wonder why not one over another
- Also really interesting how this was done specifically to improve neuroplasticity
    - I wonder if this could be used in schools
    - Is it used for therapy?
- AR and VR have already been used in therapy to create immersive scenarios for people
- I wonder what benefits this has over for example a therapist adjusting based on what signals they see
- Maybe they could add other signals other than just the brian waves
    - Heart beat
    - Sweating?
    - Facial expressions
    - Body signals (foot tapping, nervous hand gestures)
- Does this need long term (repeated) use or is it showing improvement after one use

Article #4 Notes: Stabilizing brain-computer interfaces through alignment of latent dynamics

Article notes should be on separate sheets

Karpowicz, B. M., Ali, Y. H., Wimalasena, L. N., Sedler, A. R., Keshtkaran, M. R., Bodkin, K., Ma, X., Rubin, D. B., Williams, Z. M., Cash, S. S., Hochberg, L. R., Miller, L. E., & Pandarinath, C. (2025). Stabilizing brain-computer interfaces through alignment of latent dynamics. *Nature Communications*, *16*(1), 4662. https://doi.org/10.1038/s41467-025-59652-y

Follow up references:

Ma, X., Rizzoglio, F., Bodkin, K. L., Perreault, E., Miller, L. E., & Kennedy, A. (2023). Using adversarial networks to extend brain computer interface decoding accuracy over time. *eLife*, *12*, e84296. https://doi.org/10.7554/eLife.84296

Sani, O. G., Abbaspourazad, H., Wong, Y. T., Pesaran, B., & Shanechi, M. M. (2021). Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nature Neuroscience*, *24*(1), 140–149. https://doi.org/10.1038/s41593-020-00733-0

Deo, D. R., Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2024). Brain control of bimanual movement enabled by recurrent neural networks. *Scientific Reports*, *14*(1), 1598. https://doi.org/10.1038/s41598-024-51617-3

KEEP THIS BLANK AND USE AS A TEMPLATE

Source Title	Stabilizing brain-computer interfaces through alignment of latent dynamics
Source citation (APA Format)	Karpowicz, B. M., Ali, Y. H., Wimalasena, L. N., Sedler, A. R., Keshtkaran, M. R., Bodkin, K., Ma, X., Rubin, D. B., Williams, Z. M., Cash, S. S., Hochberg, L. R., Miller, L. E., & Pandarinath, C. (2025, May 19). Stabilizing Brain-computer interfaces through alignment of Latent Dynamics. Nature News. https://www.nature.com/articles/s41467-025-59652-y#citeas
Original URL	https://www.nature.com/articles/s41467-025-59652-y#citeas
Source type	Journal Article

Keywords	Brain-computer interface
#Tags	#neural-decoding
Summary of key points + notes (include methodology)	This study analyzes how intracortical brain computer interfaces can minimize their instabilities wihtout knowledge of the subject's behaviors. Instabilities in the BCI can reduce accuracy on the model's predictions as the brain waves will shift creating new graphs that won't represent the ones the model was trained on. This study created a platform for nonlinear manifold alignment with dynamics to adjust the interface in real time. It focuses on an unsupervised training dataset in order to reduce the need to know the subject's behaviors and movemens. To do this it creates a day 0 decoding axis and compares all day k decoding axises to that. This encoding happens across days and can even be done offline, allowing the iBCI to be adjusted in real time across days or months. The iBCI with NoMAD was tested on monkeys and it was shown to do significantly better than other solutions. This study relates to my potential topic of understand how BCIs can affect mental health and possibly treat it by understanding how BCIs are adjusting to instabilities.
Research Question/Problem/ Need	How can instabilities in BCIs be reduced in order to increase accuracy of the model's predictions?
Important Figures	Half-life = 57.9 days → this is the half-life of the degradation of the BCI that this paper developed and is more than 8 times greater than some of the other current BCIs
VOCAB: (w/definition)	Manifold structures: mathematical structure to describe underlying pattern of data from neural decoding Neural population activity: data from several neurons (population) rather than a single one Intracortical Brain Computer Interfaces: implanted directly into cortex (inside the brain)
Cited references to follow up on	Ma, X. et al. Using adversarial networks to extend brain computer interface decoding accuracy over time.*eLife* **12**, e84296 (2023). → seems like another cool methodology to improve neural decoding and BCIs Sani, O. G., Abbaspourazad, H., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nat. Neurosci.* **24**, 140–149 (2021) → uses behavior to adapt model's prediction Deo, D. R. et al. Brain control of bimanual movement enabled by recurrent neural networks. *Sci. Rep.* **14**, 1598 (2024). → learn more

	about bimanual movement and how BCI and neural decoding helps
Follow up Questions	What instabilities come due to the physical design of the implant rather than the data?

Bulleted notes:

1. At first I had no idea if this paper was focusing on phsycially stabilizing the BCIs or just stabilizing the data
2. Then I found out it was digitally adjusting the data
3. I think this is a really interesting approach to account for the fact that there is so much noise in BCIs
4. Graphs are super interesting
    - Showing the different representations of how the different ways to stabilize
5. Method to stabilize:
    - Creates day 0 axis
    - Compares every axis for days after that to the initial axis
    - Encoding happens every day
    - Can even happen offline
6. This method allows iBCI to be adjust offline for days or months after being implanted into the patient's body
7. Tested on monkeys
8. Shown to do significantly better than other BCI models
9. I wonder how data stabilization can also be improved

Article #5 Notes: Impact of Environmental Pollutants on Gut Microbiome and Mental Health via the Gut–Brain Axis

Article notes should be on separate sheets

Citation: Singh, S., Sharma, P., Pal, N., Kumawat, M., Shubham, S., Sarma, D. K., Tiwari, R. R., Kumar, M., & Nagpal, R. (2022). Impact of Environmental Pollutants on Gut Microbiome and Mental Health via the Gut–Brain Axis. *Microorganisms*, *10*(7), 1457. https://doi.org/10.3390/microorganisms10071457

Follow Up References:

From gut dysbiosis to altered brain function and mental illness: Mechanisms and pathways | Molecular Psychiatry. (n.d.). Retrieved September 16, 2025, from https://www.nature.com/articles/mp201650

KEEP THIS BLANK AND USE AS A TEMPLATE

Source Title	Impact of Environmental Pollutants on Gut Microbiome and Mental Health via the Gut–Brain Axis
Source citation (APA Format)	gh, S., Sharma, P., Pal, N., Kumawat, M., Shubham, S., Sarma, D. K., Tiwari, R. R., Kumar, M., & Nagpal, R. (2022). Impact of Environmental Pollutants on Gut Microbiome and Mental Health via the Gut–Brain Axis. *Microorganisms*, *10*(7), 1457. https://doi.org/10.3390/microorganisms10071457
Original URL	https://www.mdpi.com/2076-2607/10/7/1457

Source type	Journal Article
Keywords	mental health; gut microbiota; gut–brain axis; gut dysbiosis; environmental pollutants
#Tags	#environment-for-health
Summary of key points + notes (include methodology)	This paper found that several environmental factors such as metals, air pollution, phthalates, bisphenol A, and more have a huge impact on the gut microbiome which has a resulting impact on mental health. Different factors such as leaky gut and the blood brain barrier are both positively and negatively affected by these various environmental factors. This paper's methodology involved doing a thorough review of the various studies in this area and composing their results and then analyzing them to draw conclusions. They also studied how different microbiota targeted interventions such as psychobiotics affected mental health.
Research Question/Problem/ Need	How do different environmental pollutants affect the microbiota-gut-brain-axis?
Important Figures	Mental health and drug addiction disorders have increased by 13% in the last decade → this was part of the introduction, and it shows the emphasis that understanding various factors that are related to mental health is important, one of them being the effect of environmental pollutants on the gut brain axis
VOCAB: (w/definition)	Gut dysbiosis: Imbalance in the gut microbiome that can affect health. Commensal gut microbiome: Microorganisms in the gut that normally live symbiotically without harming the host. Brain physiology: Study of how the brain's structures and systems function. Enteric nervous system: Network of neurons in the gut that regulates digestion independently of the brain. Tryptophan metabolism: Biological process that breaks down tryptophan into compounds like serotonin and kynurenine. Vagus nerve: Major nerve connecting the brain and gut, involved in parasympathetic control. Bidirectional communication: Two-way signaling between the gut and brain.

	Xenobiotics: Foreign chemical substances not naturally produced by the body. Neurotoxicity: Damage to the nervous system caused by chemicals or substances. Phthalates: Chemicals used to make plastics flexible; can disrupt endocrine function. Bisphenol A: Chemical in plastics that can act as an endocrine disruptor. Particulate matter: Tiny solid or liquid particles in the air that can harm health when inhaled. Bacteroidetes: Phylum of bacteria in the gut involved in breaking down complex molecules. Firmicutes: Phylum of gut bacteria involved in energy absorption and metabolism. Proteobacteria: Diverse bacterial phylum, some of which are pathogenic. Actinobacteria: Bacterial phylum including many beneficial gut microbes. Verrucomicrobia: Less common gut bacteria that can influence metabolism. Fusobacteria: Bacteria sometimes associated with infections and inflammation. Biomolecular networks: Interactions between molecules like proteins, genes, and metabolites in cells. HPA axis: Hypothalamus–pituitary–adrenal system controlling stress responses. Glucocorticoids: Hormones produced by the adrenal glands that regulate metabolism and stress. Monocarboxylate transporters: Proteins that move molecules like lactate and pyruvate across cell membranes.
Cited references to follow up	Rogers, G. B., Keating, D. J., Young, R. L., Wong, M.-L., Licinio, J., &

on	Wesselingh, S. (2016). From gut dysbiosis to altered brain function and mental illness: Mechanisms and pathways. Molecular Psychiatry, 21(6), 738–748. https://doi.org/10.1038/mp.2016.50 → Learn more about how the gut brain axis affects brain function – use for psychobiotics and using gut brain axis to improve mental health or neurological conditions
Follow up Questions	Is there a difference in the effect of pollutants that are breathed in vs ingested? How does duration of being affected change the results?

Bulleted Notes:

- Included several tables
    - Each with different environmental factors
    - Each table had various rows with different studies
    - Each study covered the same environmental factor
    - Some studies used mice
    - Others used drosophila
- Various pathways of microbiota gut brain axis
    - Immune
    - Endocrine (HPA axis)
    - Neural (vagus)
    - Metabolic
    - Tryptophan metabolism
    - Neuroactive metabolites (e.g. SCFAs)
- Dysbiosis: when the gut has increased permeability → leads to inflammatory signals and reduction of microbes that help benefit mental health
- Metals lead to psychiatric disordiers
    - Alter gut microbiome composition
    - Oxidative stress in gut and brain
- Phthalates (microplastics)
    - Problematic in early life (prenatal or neonatal)
- Bisphenol A
    - Alters gut microbiota composition – increases proteobacteria, decreases akkermansia (beneficial bacteria), decreases bifidobacteria
    - Leads to more gut permeatbility
- Air pollutants
    - Exposure causes increased mental health concerns
    - Can alter gut microbiota
    - May induce oxidative stress

Article #6 Notes: Health effects in people relocating between environments of differing ambient air pollution concentrations: A literature review

Article notes should be on separate sheets

Citation:

Edwards, L., Wilkinson, P., Rutter, G., & Milojevic, A. (2022). Health effects in people relocating between environments of differing ambient air pollution concentrations: A literature review. *Environmental Pollution*, *292*, 118314. https://doi.org/10.1016/j.envpol.2021.118314

Follow Up citations:

Vilcassim, M.J.R., Thurston, G.D., Chen, L.C., et al., Jun 11 2019. Exposure to air pollution is associated with adverse cardiopulmonary health effects in international travellers. J. Trav. Med. 26 (5), 11. https://doi.org/10.1093/jtm/taz032.

Lin, Y., Ramanathan, G., Zhu, Y., et al., 2019. Pro-oxidative and pro-inflammatory effects after traveling from Los Angeles to Beijing: a biomarker-based natural experiment. Circulation 20doi. https://doi.org/10.1161/CIRCULATIONAHA.119.042054.

KEEP THIS BLANK AND USE AS A TEMPLATE

Source Title	Health effects in people relocating between environments of differing ambient air pollution concentrations: A literature review
Source citation (APA Format)	Edwards, L., Wilkinson, P., Rutter, G., & Milojevic, A. (2022). Health effects in people relocating between environments of differing ambient air pollution concentrations: A literature review. Environmental Pollution, 292, 118314. https://doi.org/10.1016/j.envpol.2021.118314

Original URL	linkinghub.elsevier.com/retrieve/pii/S0269749121018960
Source type	Article
Keywords	Air pollution Particulate matter Migration Relocation Lung function Mortality
#Tags	#environment-for-health
Summary of key points + notes (include methodology)	In this paper, they do a literature review of the different studies that analzye the effects of relocation on health, specifically focusing on the change in air pollution levels. They had a total of 7 studies included and studied various particiulate matters. They looked at different health outcomes on lungs and the heart. They had certain eligibility criteria for the papers that would be included. These were that the studies had to focus on human populations, include certain particulate matters in their analysis, and they had to include health reports from before and after the relocation.
Research Question/Problem/ Need	What is the effect on health of a change in particulate matters on people who relocated?
Important Figures	(see below)

Author, Year	Overall Rating	1. Are the study groups free from baseline differences?	2. Was knowledge of the exposure groups adequately prevented during the study?	3. Were exposure assessment methods robust?	4. Were outcome assessment methods robust?	5. Were confounding and effect modification adequately addressed?	6. Were incomplete outcome data adequately addressed?	7. Are reports of the study free of suggestion of selective outcome reporting?	8. Was the study free of support from a company, study author, or other entity having a financial interest in any of the exposures studied?	9. Was the study apparently free of other problems that could put it at a risk of bias?
Avol, 2001	probably low	probably low	low	probably low	low	low	probably low	low	probably low	probably low
Awad, 2019	low	low	low	low	low	low	low	low	low	low
Gan, 2010	probably low	probably low	low	low	low	probably low	probably low	low	low	low
Kinney, 2000	probably high	low	probably low	probably low	probably low	probably high	probably low	low	low	probaly low
Krewski, 2003	probably low	probably low	low	probably low	low	low	low	low	low	low
Liang, 2019	probably high	probably high	low	probably low	low	probably high	probably low	probably low	probably low	probably low
Lin, 2019	probably low	low	low	low	low	probably low	probably low	low	probably low	probably low
Lleras-Muney, 2010	probably low	probably low	low	probably low	probably low	probably low	probably low	low	low	low
Madsen, 2010	probably low	probably low	low	low	low	probably low	low	low	low	low
Pereira, 2016	probably low	probably low	low	probably low	low	low	low	low	low	probably low
Renzetti, 2009	probably low	low	probably low	low	low	probably low	probably low	low	low	low
Sakai, 2004	probably high	low	probably high	low	low	probably high	low	low	low	probably low
Vilcassim, 2019	probably high	low	probably high	probably low	probably high	low	low	low	low	probably low
Wu, 2014	probably high	low	probably high	low	probably low	probably low	low	probably low	low	low
Wu, 2013	probably high	low	probably low	low	low	probably high	probably low	probably low	low	probably low

This figure is important because it shows what bias levels each study had and in what specific way they were found to be biased which is important to consider when analyzing the results that were provided in the literature review.

VOCAB: (w/definition)	FEV: forced expiratory volume – how quickly and forcefully a person can exhale air from their lungs Peak expiratory flow rate: maximum speed someone can exhale air after a deep breath – way to track lung health
Cited references to follow up on	Vilcassim, M.J.R., Thurston, G.D., Chen, L.C., et al., Jun 11 2019. Exposure to air pollution is associated with adverse cardiopulmonary health effects in

	international travellers. J. Trav. Med. 26 (5), 11. https://doi.org/10.1093/jtm/taz032. 11. I want to learn more about the specifics of how this study was conducted Lin, Y., Ramanathan, G., Zhu, Y., et al., 2019. Pro-oxidative and pro-inflammatory effects after traveling from Los Angeles to Beijing: a biomarker-based natural experiment. Circulation 20doi. https://doi.org/10.1161/CIRCULATIONAHA.119.042054. 12. This was another study that was included in the literature review that I want to further analyze
Follow up Questions	How do other things in the air affect heatlh (such as pollens or allergens)? What were some of the animal studies that were excluded from this literature review?

- It is well known that peole who move to a new location experience health effects due to the change in environment
- this lit review consists of 4 short term and 3 long term studies
- 1 in 3 deaths globally were due to exposure to fine particulate matter smaller than 2.5 nanometers
- Most common death due to air pollution are stroke, heart disease, lung cancer, chronic obstructive pulmonary disease
- Many studies have been done on the research of short and long term effects of air pollution on health but they focus on unchanging exposure
- Population, exposure, comparator, outcome, study design items used to develop study eligibility criteria
    - Must include human population not animals
    - Must include a health report from before and after the relocation
- Studied various health outcomes
    - Lung function metrics
    - Respiratory symptoms and hospitalization
    - Cardiovascular status metrics
    - Blood pressure
    - Pulse pressure
    - Heart rate
    - Heart rate variability
- There was a stronger trend for health change noted in children who moved at least 3 years prior to the follow up
- Studies also found that people who moved to urban areas (with significantly higher air pollution) from rural areas (that had signficantly lower air pollution)) had higher blood pressure and pulse pressure associated with that change

Article #7 Notes: Associations between Changes in Exposure to Air Pollutants due to Relocation and the Incidence of 14 Major Disease Categories and All-Cause Mortality: A Natural Experiment Study

Citation:

Chen, G., Qian, Z. (Min), Zhang, J., Wang, X., Zhang, Z., Cai, M., Arnold, L. D., Abresch, C., Wang, C., Liu, Y., Fan, Q., & Lin, H. (2024). Associations between Changes in Exposure to Air Pollutants due to Relocation and the Incidence of 14 Major Disease Categories and All-Cause Mortality: A Natural Experiment Study. *Environmental Health Perspectives*, *132*(9), 097012. https://doi.org/10.1289/EHP14367

Follow Up Citations:

Eslami B, Alipour S, Omranipour R, Naddafi K, Naghizadeh MM, Shamsipour M, et al. 2022. Air pollution exposure and mammographic breast density in Tehran, Iran: a cross-sectional study. Environ Health Prev Med 27:28, PMID: 35786683, https://doi.org/10.1265/ehpm.22-00027.

Source Title	Associations between Changes in Exposure to Air Pollutants due to Relocation and the Incidence of 14 Major Disease Categories and All-Cause Mortality: A Natural Experiment Study
Source citation (APA Format)	*Associations between Changes in Exposure to Air Pollutants due to Relocation and the Incidence of 14 Major Disease Categories and All-Cause Mortality: A Natural Experiment Study*. (n.d.). https://doi.org/10.1289/EHP14367

Original URL	https://ehp.niehs.nih.gov/doi/10.1289/EHP14367
Source type	Research Paper
Keywords	Air pollution, environment, health effects, relocation, statistical analysis
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	The goal of this paper was to analyze the impact of a change in air pollution when people relocate. They used the UK Biobank dataset to get information about people's locations, lifestyle, and health. They curated the dataset to only include people who relocated and lived in the new location for at least a year.
Research Question/Problem/ Need	How does a change in pollution levels due to relocation impact the risk of getting certain diseases or dying?
Important Figures	50,522 people – the number of people in the final dataset. This is important because it shows the significant size required to complete a data analysis of this size 93% white – this is is an important number because it shows significant bias toward people of one race which can skew the results
VOCAB: (w/definition)	Covariates – independent variables, factors that change independently Cox proportional hazards regression – a statistical method to analyze the possibility that something will happen over time relatively by looking at how it changes if one variable changes in a specific way
Cited references to follow up on	Awad YA, Di Q, Wang Y, Choirat C, Coull BA, Zanobetti A, et al. 2019. Change in PM2.5 exposure and mortality among Medicare recipients: combining a semirandomized approach and inverse probability weights in a low exposure population. Environ Epidemiol 3(4):e054, PMID: 31538135, https://doi.org/10.1097/EE9. 0000000000000054. I want to see more about their methodology, specifically how they collected data and how they analyzed that data. Eslami B, Alipour S, Omranipour R, Naddafi K, Naghizadeh MM, Shamsipour M, et al. 2022. Air pollution exposure and mammographic breast density in Tehran, Iran: a cross-sectional study. Environ Health Prev Med 27:28, PMID: 35786683, https://doi.org/10.1265/ehpm.22-00027. This paper was criticized for having potentially residual confounding. I want to compare the methodology of this paper with the paper I just read to see how the methodology and data collection varies so I can identify what might

	have caused the potential residual confounding and make sure to not do that in my project.
Follow up Questions	How did the bias present in this study affect the results? How might this study have looked different if it included people who moved to places outside of the UK?

Bulleted Notes:

- Lots of studies linking air pollution exposure to disease, limited to compare different exposures in the same people
- 50,522 participants, UK biobank, 2006-2010
- Measured exposure to particulate matter diameter <= 2.5 nanometers, <=10 nanometers, nitrogen oxide, nitrogen dioxide, sulfur dioxide, estimated levels for each participant based on their residential address and follow up
- Nine exposure groups classified by changes
- Incidence and mortality of 14 major diseases
    - Used links to hospital inpatient records and death registries
- Used cox proportional hazard models to estimate hazard ratios – the relative risk of getting disease over time based on changing one variable
- Final results presented with 95% confidence intervals
- Median follow up – 12.6 years
- 29,869 diagnosed with disease of interest
- 3,144 died
- Low to moderate relocation → increased risk of 14 diseases not death
- Each observation using hazard ratios has unique conclusions for the change in each of the particulate matter and chemical compound types tracked in the study
- 6.7 million deaths due to air pollution in 2019
- Research has shown that even in areas that are deemed to have good air quality, there are still risks to health for individuals → need to understand how reducing the air pollution will affect the people
- Not many focus on individual level changes such as residential moves
- Some other studies were observational studies in nature which means that it is hard to conclude causality due to other factors that could have changed and impacted the results as well
- Natural experiment design is becoming more popular → helps control issue of bias and time-invariant confounding
- Other studies only studied fine particulate matter <=2.5 nanometers
- Data from uk biobank – 500,000 people aged 40-69 in UK
    - Presents age bias because no data of people less than 40 when they started tracking
    - Older people are usually more immune to disease
- Address change info → administrative data liaison service of uk
- Used an air dispersion model from uk department for environment, food and rural affairs to estimate the ambient air pollution at a 1km by 1km precision
- Used survey to collect sociodemographic and behavioral covariates – focused on ones associated with the disease categories or air pollution
- Use various statistical analysis methods (townsend deprivation index, propensity score approach) to test sensitivy of different covariates and check their balance

Article #8 Notes: Geographic and demographic variations of inhalant allergen sensitization in Koreans and non-Koreans

Citations:

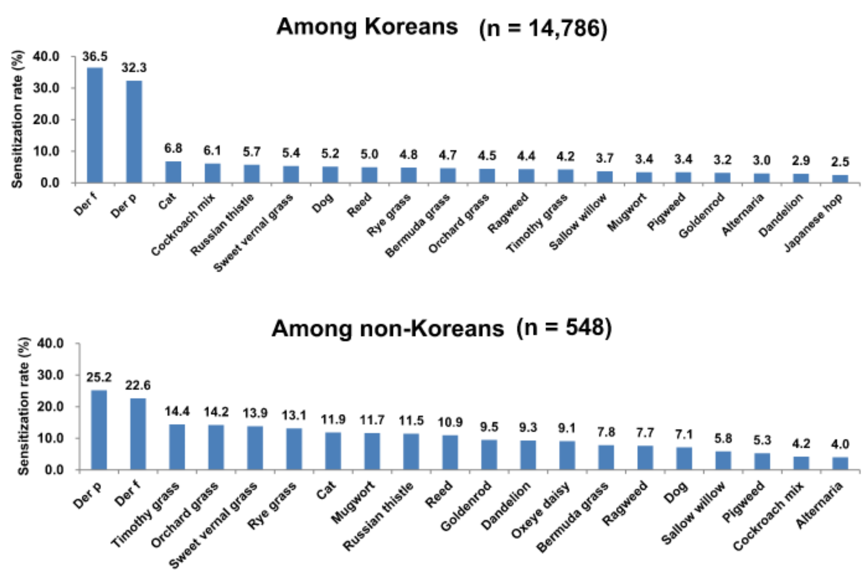
Park, S. C., Hwang, C. S., Chung, H. J., Purev, M., Al Sharhan, S. S., Cho, H.-J., Yoon, J.-H., & Kim, C.-H. (2019). Geographic and demographic variations of inhalant allergen sensitization in Koreans and non-Koreans. *Allergology International*, *68*(1), 68–76. https://doi.org/10.1016/j.alit.2018.07.005

Follow Up Citaitions:

Park, K. H., Lee, J., Lee, S. C., Son, Y. W., Sim, D. W., Lee, J.-H., & Park, J.-W. (2017). Comparison of the ImmunoCAP Assay and AdvanSure™ AlloScreen Advanced Multiplex Specific IgE Detection Assay. *Yonsei Medical Journal*, *58*(4), 786–792. https://doi.org/10.3349/ymj.2017.58.4.786

Park, H. J., Kim, E.-J., Yoon, D., Lee, J. K., Chang, W.-S., Lim, Y.-M., Park, J.-W., & Lee, J.-S. (2017). Prevalence of Self-reported Allergic Diseases and IgE Levels: A 2010 KNHANES Analysis. *Allergy, Asthma & Immunology Research*, *9*(4), 329–339. https://doi.org/10.4168/aair.2017.9.4.329

Source Title	Geographic and demographic variations of inhalant allergen sensitization in Koreans and non-Koreans
Source citation (APA Format)	Park, S. C., Hwang, C. S., Chung, H. J., Purev, M., Al Sharhan, S. S., Cho, H.-J., Yoon, J.-H., & Kim, C.-H. (2019). Geographic and demographic variations of inhalant allergen sensitization in Koreans and non-Koreans. *Allergology International*, *68*(1), 68–76. https://doi.org/10.1016/j.alit.2018.07.005
Original URL	https://www.sciencedirect.com/science/article/pii/S1323893018300893?vi

	a%3Dihub
Source type	Journal Article
Keywords	Allergen Epidemiology House dust mite Pollen Sensitization
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	This study aimed to understand how the differences in where people lived and their demographics affected their sensitivization rates to different allergies. This can give an insight as to hwo where people live affects their risk of getting certain allergen related respiratory diseases. They collected data using the advanSure allostation MAST assay to measure the sensitivization rates of 33 diferent inhalant allergens. They found distinct differences between Koreans and non-Koreans and there were also distinct differences in different regions within Korea, rural vs. urban, and different continents outside Korea. This indicates that the allergen sensitivatization is very closely linked to the area in which an individual is living.
Research Question/Problem/ Need	How does the sensitization rate of different inhalant allergens among different demographic and geographic factors of Koreans and non-Koreans?
Important Figures	 This figure shows the differences in the rate of sensitivization to different allergens for Koreans vs. non Koreans – important because it shows how kroeans are way more sensitivized to dar f and dar p than non Koreans but way less sensitivization to all other inhalant allergens
VOCAB: (w/definition)	Allergen sensitization: the process where the immune system learns to

	recognize a specific foreign substance, or allergen, through repeated exposure, and begins producing IgE antibodies against it (step before allergy symptoms show)
Cited references to follow up on	k, K. H., Lee, J., Lee, S. C., Son, Y. W., Sim, D. W., Lee, J.-H., & Park, J.-W. (2017). Comparison of the ImmunoCAP Assay and AdvanSure™ AlloScreen Advanced Multiplex Specific IgE Detection Assay. *Yonsei Medical Journal*, *58*(4), 786–792. https://doi.org/10.3349/ymj.2017.58.4.786 ■ Look at how to use this methodology k, H. J., Kim, E.-J., Yoon, D., Lee, J. K., Chang, W.-S., Lim, Y.-M., Park, J.-W., & Lee, J.-S. (2017). Prevalence of Self-reported Allergic Diseases and IgE Levels: A 2010 KNHANES Analysis. *Allergy, Asthma & Immunology Research*, *9*(4), 329–339. https://doi.org/10.4168/aair.2017.9.4.329 ■ More to see how people are affected by allergens
Follow up Questions	How does the allergen sensitization vary among migrants? How severe are the allergen symptoms and does this vary across the different axes studied?

Notes:

- Allergen sensitization is highly related to the risk of developing respiratory allergic diseases
- Pattern of inhalant allergen sensitization is affected by environmental factors and demographic factors
- Evaluated 15,344 individuals that were suspected to have respiratory allergic diseases (for example allergic rhinitis, asthma)
- 14,786 koreans (mean age: 35.8 years, range: 1-94 years)
- 548 non koreans (mean age: 33.1 years, range: 1-71 years)
- Allergen sensitization can be affected by residential environment, so classified based on area of long term residence or birth
- Non koreans were visitors, not non koreans who were living in korea or had been there for greater than 4 seasons
- Blood samples analyzed using advanSure allostation MAST assay that can detect 33 inhalant allergens from seven different groups
- Koreans have higher sensitization rates for dar f and dar p than non koreans
- Koreans have way lower sensitization rates for all other inhalant allergens
- For koreans there was variation between regions
    - In gangwon region, HDM positive rate was less than 30% which was much lower than average, but rates for pollen allergens were significantly higher
    - In jeju region, hdm positive rate was also less than 30%, high rate for japanese cedar was unique to this region
- Allergen sensitization rate higher in urban areas than rural areas

Article #9 Notes: Effects of changes in living environment on physical health: a prospective German cohort study of non-movers

Citations:

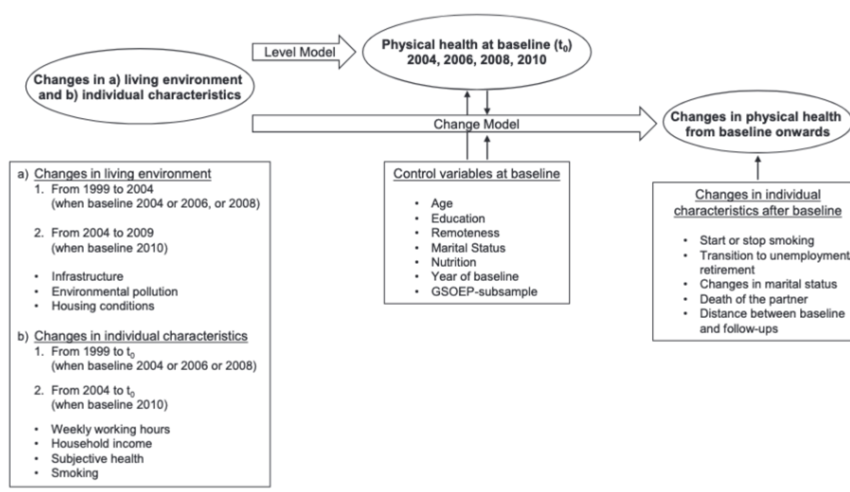
Aretz, B., Doblhammer, G., & Janssen, F. (2019). Effects of changes in living environment on physical health: A prospective German cohort study of non-movers. *European Journal of Public Health*, *29*(6), 1147–1153. https://doi.org/10.1093/eurpub/ckz044

Follow Up Citations:

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heterosked Asticity. *Econometrica (Pre-1986)*, *48*(4), 817.

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, *44*(4), 1049–1060. https://doi.org/10.2307/2531734

Source Title	Effects of changes in living environment on physical health: a prospective German cohort study of non-movers
Source citation (APA Format)	tz, B., Doblhammer, G., & Janssen, F. (2019). Effects of changes in living environment on physical health: A prospective German cohort study of non-movers. *European Journal of Public Health*, *29*(6), 1147–1153. https://doi.org/10.1093/eurpub/ckz044
Original URL	https://academic.oup.com/eurpub/article/29/6/1147/5384510
Source type	Journal Article
Keywords	Longitudinal data, effect of environment on health, living environment, changes in living environment, housing conditions, linear regression, prospective cohort study

#Tags	#environment-on-health
Summary of key points + notes (include methodology)	This study aimed to understand how changes in living conditions unrelated to actually relocating locations affected health for individuals. They collected data from the German Socio-Economic panel. They created a novel approach to strengthen the causal explanatory power of their findings. They did this by using four methodological strategies to reduce the confounding effects of selected migration or health selection into living environments. The strategies included using a strict time order, predicting changes with regard to time, including only non-movers so the effects of health didn't factor into decision and therefore create a bias, and controlling for time invariant vs time varying variables. They measured physical health through PCS, which is one of two main dimensions of the 12 item short form survey they had access to. This tool consisted of six self reported variables related to health. They also included predictors such as the living environment and individual characters as time invariant (at baseline) or time varying (up to baseline/baseline onwards). They used the identity link function and normally distributed outcome variables to control for multiple observations per person. They found that there were different amounts of changes in health based on the different baseline levels and based on the changes in their living conditions. Surprisingly, they found that men were more prone to experiencing worsened health due to worsened pollution in their living environment. Some limitations in their study included the fact that it only covered short term changes in living environments and that the living conditions were self reported and could present bias from the reporter's point of view.
Research Question/Problem/ Need	How does a change in living conditions unrelated to moving affect long term health?
Important Figures	

	This figure is important because it shows the novel methodology that they used in order to definitvely show a causal relationship
VOCAB: (w/definition)	Longitudinal studies: study that observes same group of people across extended period fo time in order to track changes and identify patterns
Cited references to follow up on	te, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heterosked Asticity. *Econometrica (Pre-1986)*, *48*(4), 817. ■ Good to understand the statistic model used in this paper er, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, *44*(4), 1049–1060. https://doi.org/10.2307/2531734 ■ Good to understand how to study longitudinal data
Follow up Questions	Were lifestyle factors included in this study? How can this study be further extended to other parts of the world?

Notes:

- Studied changes in living environment related to changes in physical health among non movers
- Collected data from german socio-economic panel
- Strengthen causal explanatory power of finding by using novel approach with four methodological strategies to reduce confounding effects by selected migration/health selection into living environments:
    - imposing a strict time order between living environment and physical health to exclude the possibility of reverse causation
    - predicting changes in health over time and not only in regard different health levels
    - including only non-movers, among whom health selection into living environments does not play a role
    - controlling for important time-invariant and time-varying individual characteristics.
- Physical health measured by PCS, one of two main dimensions of the 12-item short form survey invented by the RAND corporation
    - Psychometric tool consisting of six sel freported variables related to health
- They included predictors such as the living environment and individual characters as time invariant (at baseline) or time-varying (up to baseline/baseline onwards)
- Used identity link function and normally distributed outcome variable to control for multiple observations per person
- Found different amounts of changes in health based on different baseline levels
- Found that men were more prone to worsened pollution
- Limitations
    - Only covers short term changes in living environments
    - Living environment measures are from householders' subjective assessments → same-source bias

Article #10 Notes: Warming During Different Life Stages has Distinct Impacts on Host Resistance Ecology and Evolution

Citation:

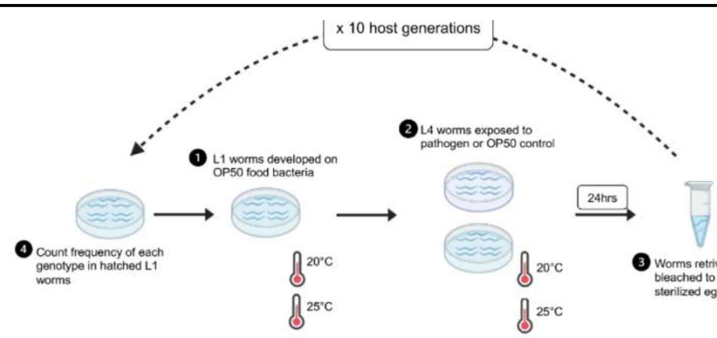
Li, J., Smith, C. A., Chen, J., Bates, K. A., & King, K. C. (2025). Warming During Different Life Stages has Distinct Impacts on Host Resistance Ecology and Evolution. *Ecology Letters*, *28*(2), e70087. https://doi.org/10.1111/ele.70087

Follow Up Citations:

Baugh, L. R., & Day, T. (2020). Nongenetic inheritance and multigenerational plasticity in the nematode C. elegans. *eLife*, *9*, e58498. https://doi.org/10.7554/eLife.58498

Kunze, C., Luijckx, P., Jackson, A. L., & Donohue, I. (2022). Alternate patterns of temperature variation bring about very different disease outcomes at different mean temperatures. *eLife*, *11*, e72861. https://doi.org/10.7554/eLife.72861

Source Title	Warming During Different Life Stages has Distinct Impacts on Host Resistance Ecology and Evolution
Source citation (APA Format)	J., Smith, C. A., Chen, J., Bates, K. A., & King, K. C. (2025). Warming During Different Life Stages has Distinct Impacts on Host Resistance Ecology and Evolution. *Ecology Letters*, *28*(2), e70087. https://doi.org/10.1111/ele.70087
Original URL	https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.70087
Source type	Journal Article

Keywords	C. elegans | host–pathogen interactions | infection | life stages | resistance evolution | warming
#Tags	#environment-on-health #model-organism-study
Summary of key points + notes (include methodology)	The study aimed to understand how warming at different points in a host's life would affect its susceptibility to pathogens. They set up evolutionary and ecological experiments. The ecological experiments were single generation and they looked at introducing the host to warming at different points in its life. They tracked mortality due to the pathogen and the relative fitness of the host. For the evolutionary study, they did multigenerational study and looked to see the frequency of the genotype that resisted pathogens across generations. They used statistic models such as the t-test and a generalized linear model to analyze and aggregate the data. They found that for the ecological one generation study, warming during the development stage increased relative fitness and resistance to the pathogen whereas warming in the adult stage increased mortality due to the pathogen. This indicates a strong causality that shows that warming in early development stages is much more beneficial compared to warming in later developmental stages. In the evolutionary study, they found that there was a cost to the resistance genotype. The resistance to the novel pathogen presence was most strongly favored when there was prolonged warming suggesting that global warming could be helpful to enhancing the animal's resistance to emerging infectious diseases.
Research Question/Problem/ Need	Does warming at different life stages of the host affect the ecological and evolution dynamics of resistance in *Caenorhabditis elegans* infected by a wild bacterial pathogen?
Important Figures	 This figure is important because it shows how they set up the experiment in order to study the evolutionary effects of the warming and pathogens on the worms.
VOCAB: (w/definition)	Fecundity: reproductive capability of organism Mechanistic model: explains how process works by uncovering underlying science

	Nematodes: diverse roundworms Assay: lab test to measure amount of substance
Cited references to follow up on	gh, L. R., & Day, T. (2020). Nongenetic inheritance and multigenerational plasticity in the nematode C. elegans. *eLife*, *9*, e58498. https://doi.org/10.7554/eLife.58498 This is interesting because it could give me insight for my project on how I could set up a multigenerational study for c elegans to see how relocation effects not only the immigrant but the future generations. ıze, C., Luijckx, P., Jackson, A. L., & Donohue, I. (2022). Alternate patterns of temperature variation bring about very different disease outcomes at different mean temperatures. *eLife*, *11*, e72861. https://doi.org/10.7554/eLife.72861 This could be helpful to show how different temperatures cause susceptibility to different diseases.
Follow up Questions	How does the health change for other health factors such as heart or lung health? Does this mimic human health (would a similar response in health effects happen to humans)?

Notes:

- Hosts can die more during infection if their immune defenses are weakened by higher temperatures
- Effects of warming across different host life stages on resistance remain unexplored
- As the pathogens change due to different temperatures, the cost and benefits of having genetics based resistance will change which could change how the organism evolves
- Used C elegans and bacterial pathogen Leucobacter musarum
- Compared fitness metrics (mortality and fecundity)
- Conducted evolution experiment → tracked frequency of resistant genotype in host populations across 10 generations
- Developed mechanistic model to disentangle impact of different environmental and host factors on evolutionary dynamics of resistance
- Manipulated temp: ambient vs. elevated, manipulated timing: periodic vs prolonged, host age: worm development vs. adult stage
- Analyses conducted in R 4.1.0
- Pathogen CFUs counts compared across host genotypes using t test
- Host mortality data fitted in generalized linear model with quasibinomial distribution
- Calculated relative fitness for evolution experiment → calculated as function of change in genotype frequencies using an equation
- For susceptible hosts, warming during host development reduced infection mortality, while during adult stage it caused higher mortality
- There was an evolutionary cost to resistant genotype shown as the fact that resistance was regularly lost across passages
- Presence of pathogen favored higher levels of resistance
- Pathogen exposure increased relative fitness of resistance under prolonged warming
- Resistance to novel pathogen in populations was strongly favored by pathogen presence when warming was prolonged suggesting that global warming could enhance animal resistance to emerging infectious diseases

Article #11 Notes: Modeling the zebrafish gut microbiome's resistance and sensitivity to climate change and parasite infection

Citation:

Sieler, M. J., Al-Samarrie, C. E., Kasschau, K. D., Kent, M. L., & Sharpton, T. J. (2025). *Modeling the zebrafish gut microbiome's resistance and sensitivity to climate change and parasite infection* (p. 2025.03.28.644597). bioRxiv. https://doi.org/10.1101/2025.03.28.644597

Follow Up Citations:

Tomanek, L. (2010). Variation in the heat shock response and its implication for predicting the effect of global climate change on species' biogeographical distribution ranges and metabolic costs. *Journal of Experimental Biology*, *213*(6), 971–979. https://doi.org/10.1242/jeb.038034

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120. https://doi.org/10.1128/AEM.01043-13

Source Title	Modeling the zebrafish gut microbiome's resistance and sensitivity to climate change and parasite infection
Source citation (APA Format)	er, M. J., Al-Samarrie, C. E., Kasschau, K. D., Kent, M. L., & Sharpton, T. J. (2025). *Modeling the zebrafish gut microbiome's resistance and sensitivity to climate change and parasite infection* (p. 2025.03.28.644597). bioRxiv. https://doi.org/10.1101/2025.03.28.644597
Original URL	https://www.biorxiv.org/content/10.1101/2025.03.28.644597v2

Source type	Journal Article
Keywords	Zebrafish; Gut microbiome, Development, Infection, Helminth, Temperature, Climate change, Pseudocapillaria tomentosa
#Tags	#environment-on-health #model-organism-study
Summary of key points + notes (include methodology)	The goal of this study was to understand the impact of change in water temperature on the gut microbiome diversity of the zebrafish. They also wanted to understand how this change would affect the parasite infection rate for zebrafish and what implications this would have for aquatic organisms. They set up adult zebrafish in water at 3 different temperatures and younger zebrafish in water in 3 different temperatures. They analyzed for parasite exposure and sample over 5 time points over 42 days to analyze the changes in the gut microbiome. They found that temperature altered the gut microbiome diversity and that as temperature increased, parasite infection rate decreased. They showed that water temperature and change in water temperature has a significant factor impacting the health of zebrafish which shows implications for how climate change could be good for the health of zebrafish in the future.
Research Question/Problem/ Need	How does a change in water temperature affect the gut microbiome health of zebrafish?
Important Figures	Fish at 28°C had highest mean infection burden – 4.86 worms per fish This figure is important because it shows how the result of parasite infection decreases as temperature increases, since the lowest temperature had the highest worm per fish rate
VOCAB: (w/definition)	Helminth exposure: contact with parasitic worms
Cited references to follow up on	anek, L. (2010). Variation in the heat shock response and its implication for predicting the effect of global climate change on species' biogeographical distribution ranges and metabolic costs. *Journal of Experimental Biology*, *213*(6), 971–979. https://doi.org/10.1242/jeb.038034 --- I want to follow up on this citation to learn more about the methodology they sued fto predict the effect of global climate change on species

	ich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120. https://doi.org/10.1128/AEM.01043-13 --- this could be an interesting use in my methodology since they are looking at how to analyze sequence data
Follow up Questions	How do bacterial diseases get changed (the impact on health) due to changes in environment temperature? How do the changes in microbiome due to temperature changes affect heart health? What about brain health via the gut brain axis?

Notes:

- Change in global water temperature due to climate change is expected to create health burden on aquatic organisms
- Gut microbiome can interact with parasites to influence infection outcomes → gut microbiome response to increasing temperatures may protect against increased infections
- Used zebrafish model organism to measure how resistant or resilient gut microbiome is against increased water temperature and helminth exposure
- Exposed adult zebrafish (already had grown up in normal temperatures) to various new temperature changes
- Analyzed fecal microbiome samples at 5 points across 42 days
- Parasite exposure and water temperature independently alter gut microbiome diversity
- Water temperature moderates association between parasite infection and gut microbiome
- Temperature altered the gut microbiome
- Increasing the water temperature reduced development of p. tomentosa worms in zebrafish
    - Due to thermal inhibition of parasite development or temperature mediated interactions between host microbiome and immune responses
- Gut microbiome composition varied over time
- Certain bacterial genera showed significant changes in response to the temperature changes

Article #12 Notes: Deep learning prediction models based on EHR trajectories: A systematic review

Citation:

Amirahmadi, A., Ohlsson, M., & Etminani, K. (2023). Deep learning prediction models based on EHR trajectories: A systematic review. *Journal of Biomedical Informatics*, *144*, 104430. https://doi.org/10.1016/j.jbi.2023.104430

Follow Up citations:

Men, L., Ilk, N., Tang, X., & Liu, Y. (2021). Multi-disease prediction using LSTM recurrent neural networks. *Expert Systems with Applications*, *177*, 114905. https://doi.org/10.1016/j.eswa.2021.114905

Zeng, X., Feng, Y., Moosavinasab, S., Lin, D., Lin, S., & Liu, C. (2019). Multilevel Self-Attention Model and its Use on Medical Risk Prediction. In *Biocomputing 2020* (pp. 115–126). WORLD SCIENTIFIC. https://doi.org/10.1142/9789811215636_0011

Source Title	Deep learning prediction models based on EHR trajectories: A systematic review
Source citation (APA Format)	irahmadi, A., Ohlsson, M., & Etminani, K. (2023). Deep learning prediction models based on EHR trajectories: A systematic review. *Journal of Biomedical Informatics*, *144*, 104430. https://doi.org/10.1016/j.jbi.2023.104430
Original URL	https://www.sciencedirect.com/science/article/pii/S153204642300151X
Source type	Literature Review Article
Keywords	Disease prediction, EHR trajectories, Systematic review, Electronic health records, Deep learning

#Tags

#environment-on-health

Summary of key points + notes (include methodology)

Recent breakthroughs in deep learning models have allowed them to have the capability to understand and model EHR data. This lead the authors to try and understand what work has already been done with deep learning models and how they have been analyzing EHR data. Some of the most common use cases for deep learning models analyzing EHR data were predicting the disease onset for the next visit and cardiovascular diseases in general. Some of the most common forms of modeling techniques were RNNs, Attention Mechanisms, and CNNs and GNNs. Some of the major challenges were the data dependencies within EHR data, a lack of a large amount of data, and low explainability and interpretability in the models.

Research Question/Problem/ Need

How can deep learning models be used to predict EHR trajectories?

Important Figures

Prediction task	Outcome	# of papers
Diseases	Multiple diseases	19
	Cardiovascular disease	17
	Diabetes	7
	Kidney diseases	7
	Chronic obstructive pulmonary disease	4
	Alzheimer's diseases	4
Other	Mortality prediction	6
	Readmission prediction	6

This figure is important because it shows the different outcomes analyzed by the model and it is important because it shows in general what other papers have focused on when creating their models. It is also interesting tht they didn't look into the input data or what type of people they were focusing on specifically, for example migrants versus people that lived in the one place for their whole life.

VOCAB: (w/definition)

EHR: electronic health records
Temporal: related to time

Cited references to follow up on

ı, L., Ilk, N., Tang, X., & Liu, Y. (2021). Multi-disease prediction using LSTM recurrent neural networks. *Expert Systems with Applications*, *177*, 114905. https://doi.org/10.1016/j.eswa.2021.114905

This one is important to follow up on because it would be very helpful for my methodology as RNNs could be a model structure that I pursue. It would also be helpful to see where they get their data.

	g, X., Feng, Y., Moosavinasab, S., Lin, D., Lin, S., & Liu, C. (2019). Multilevel Self-Attention Model and its Use on Medical Risk Prediction. In *Biocomputing 2020* (pp. 115–126). WORLD SCIENTIFIC. https://doi.org/10.1142/9789811215636_0011 This is another model structure that I could use for my project, self-attention model, and it would be useful to see where they get their data and how they develop the model to analyze and predict health outcomes.
Follow up Questions	Are there specific model types that work better with small amounts of data? Where did these papers get their data from?

Notes:

- EHRs are used more and more to predict future health risks by analyzing the patient trajectories over time
- Advancements in deep learning allow it to have the enhanced ability to model complex temporal dependencies within EHR data
- Reviewed papers in scopus, pubmed, ieee xplore, acm databases from January 2016 to April 2022
- Included studies that applied deep learning techniques to HER trajectories
- Extracted data on publication characteristics, objectives, and methodologies, and analyzed those
- A total of 63 studies were included
- Most commonly predicted – disease onset and cardiovascular disease
- Modeling techniques
    - RNNs for modeling long term dependencies in sequential data
    - Time aware attention mechanisms to focus on relevant time points in patient trajectories
    - Self aware mechanisms to facilitate the modeling of complex relationships within the data
    - CNNs for feature extraction from EHR sequences
    - GNNs for inter visit relationships
    - Attention scores to enhance model explainability
- Challenges
    - Data dependencies were hard to handle within EHR data
    - There was often insufficient data for training models
    - It was hard to have good explainability and interpretability for deep learning models
- Deep learning models have improved modeling of EHR trajectories
- Need for more publicly available EHR trajectory datasets
- Few models address all aspects of EHR trajectory data comprehensively

Patent #1 Notes: Personalized Health Tracker and Method for Destination Selection Based on Tracked Personalized Health Information

Citation:

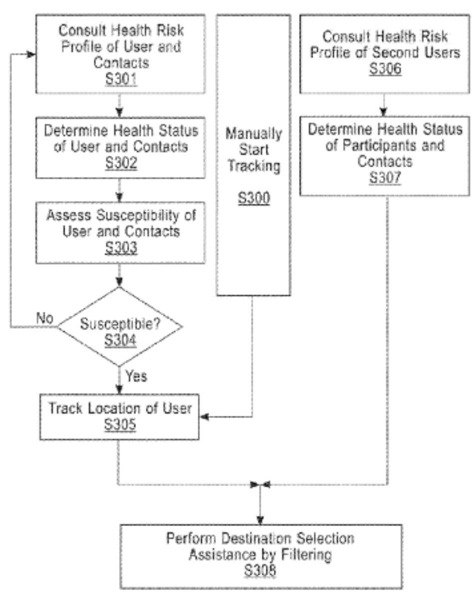
Balasubramanian, S., III, T. G. L., Malinowski, J. R., & Vasudevan, C. (2018). *Personalized health tracker and method for destination selection based on tracked personalized health information* (United States Patent No. US20180225421A1). https://patents.google.com/patent/US20180225421A1/en

Follow Up Citations:

Clapp, G. (2014). *Health related location awareness* (United States Patent No. US8758238B2). https://patents.google.com/patent/US8758238B2/en?oq=US20180225421A1

Moturu, S., & Madan, A. (2019). *Method and system for modeling behavior and heart disease state* (United States Patent No. US10265028B2). https://patents.google.com/patent/US10265028B2/en?oq=US20180225421A1

Source Title	Personalized Health Tracker and Method for Destination Selection Based on Tracked Personalized Health Information
Source citation (APA Format)	asubramanian, S., III, T. G. L., Malinowski, J. R., & Vasudevan, C. (2018). *Personalized health tracker and method for destination selection based on tracked personalized health information* (United States Patent No. US20180225421A1). https://patents.google.com/patent/US20180225421A1/en
Original URL	https://patents.google.com/patent/US20180225421A1/en

Source type	Patent
Keywords	Personalized health tracker, health risk information, geo-social data
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	This patent describes methods and systems to receive a user's destination request, retrieve health condition data, identify which communicable diseases pose risk to the user, and filter the destinations accordingly to reduce exposure to such risks. The system also provides navigation guidance to filtered destinations and alerts if necessary. The workflow is that the user would submit a destination type request, the system would search for destination candidates, obtain health information for the user, obtain health condition information for other users, identify potential health risks through rules, filter the destination candidates, give final destination and navigation guidance, and do alert and real time monitoring. The unique aspects of this patent are the integration of multiple data sources like wearable data, medical records, geo-social data, and contact inference. They also use social media to consider who the user is more likely to be in contact with.
Research Question/Problem/ Need	Existing navigation or destination recommendation systems typically only choose destinations or routes based on standard factors such as distance, traffic, speed, without considering health vulnerabilities which can pose a great risk to the users.
Important Figures	 This figure is important because it is a visual demonstration of the workflow of the proposed system in this patent.

VOCAB: (w/definition)	Communicable ailment: a disease that can be transmitted form one person to another Rules database: a repository of rules mapping which ailments are risky to which vulnerabilities
Cited references to follow up on	ɔp, G. (2014). *Health related location awareness* (United States Patent No. US8758238B2). https://patents.google.com/patent/US8758238B2/en?oq=US20180225421A1 ■ This interesting because it could give insight to how to relate location to health ɪuru, S., & Madan, A. (2019). *Method and system for modeling behavior and heart disease state* (United States Patent No. US10265028B2). https://patents.google.com/patent/US10265028B2/en?oq=US20180225421A1 ■ This could be helpful in modeling heart health in my project
Follow up Questions	Is this able to work for people that aren't on the system? How might widespread usage of this system this affect the traffic in different routes?

Notes:

- Navigation systems have evolved to take real time data for traffic and other factors into account, but not health risk
- There was a prior application for a patent that proposed avoiding people with communicable ailments but this patent aimed to refine the approach to not be too protective
- Health risk profile includes personal data, fitness data, medical data, family data, social data, geo-social data
- The filtering model uses health risk profile plus ruels from knowledge base to limit the destination choices
- Rule database creates decision tables that define relationships between susceptibilities and ailments
    - Must be manually constructed from medical knowledge
- Real time ambient health conditions by mapping people with communicable diseases and alerting the user to highlight risk zones
- They must retrieve data for people that are likely to be in contact with user
- Use social media to infer contacts
- Filter based on current location and people that are nearby
- Generate alerts rather than rejecting the entire route

Article #13 Notes: Linking the Gut Microbial Ecosystem with the Environment: Does Gut Health Depend on Where We Live?

Citation:

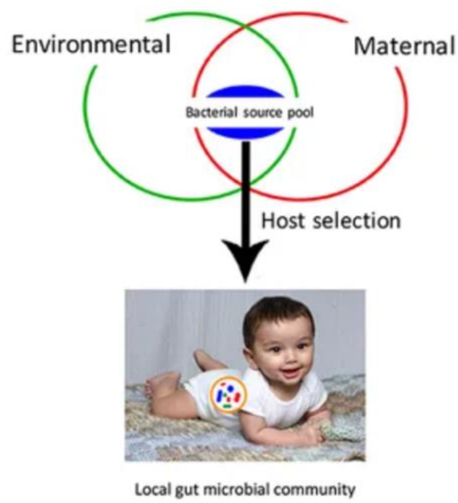
Tasnim, N., Abulizi, N., Pither, J., Hart, M. M., & Gibson, D. L. (2017). Linking the Gut Microbial Ecosystem with the Environment: Does Gut Health Depend on Where We Live? *Frontiers in Microbiology*, *8*, 1935-. https://doi.org/10.3389/fmicb.2017.01935

Follow Up Citations:

Haahtela, T., Holgate, S., Pawankar, R., Akdis, C. A., Benjaponpitak, S., Caraballo, L., Demain, J., Portnoy, J., & Hertzen, L. von. (2013). The biodiversity hypothesis and allergic disease: World allergy organization position statement. *World Allergy Organization Journal*, *6*. https://doi.org/10.1186/1939-4551-6-3

Wall, D. H., Nielsen, U. N., & Six, J. (2015). Soil biodiversity and human health. *Nature*, *528*(7580), 69–76. https://doi.org/10.1038/nature15744

Source Title	Linking the Gut Microbial Ecosystem with the Environment: Does Gut Health Depend on Where We Live?
Source citation (APA Format)	nim, N., Abulizi, N., Pither, J., Hart, M. M., & Gibson, D. L. (2017). Linking the Gut Microbial Ecosystem with the Environment: Does Gut Health Depend on Where We Live? *Frontiers in Microbiology*, *8*, 1935-. https://doi.org/10.3389/fmicb.2017.01935
Original URL	https://www.frontiersin.org/articles/10.3389/fmicb.2017.01935/abstract
Source type	Literature Review Article

Keywords	gut microbiome, immunity, environment, human health, immune tolerance, microbial colonization, biodiversity, microbe-rich environments
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	This paper aimed to understand the impact of environment on health, specifically on the gut microbiome and how that impacted health for individuals. They found that urbanization and modern lifestyles reduce the exposure to diverse microbes which alters the diversity of the human gut microbiome. This is linked to an increased risk of immune diseases. They reviewed peer reviewed studies that linked human gut microbiome composition to the environment and included studies that focused on different geographic regions and populations, but focused specifically on urban communities.
Research Question/Problem/ Need	How do environmental factors influence the composition of the gut microbiome and what are the health implications?
Important Figures	Figure 1  This figure is important because it shows the cycle of how the local gut microbial community forms and develops within a baby.
VOCAB: (w/definition)	Immunomodulation: process that alters the immune system's response Antibiotic resistance: the ability of microorganisms to resist effects of drugs
Cited references to follow up on	htela, T., Holgate, S., Pawankar, R., Akdis, C. A., Benjaponpitak, S., Caraballo, L., Demain, J., Portnoy, J., & Hertzen, L. von. (2013). The

	biodiversity hypothesis and allergic disease: World allergy organization position statement. *World Allergy Organization Journal*, *6*. https://doi.org/10.1186/1939-4551-6-3 This one could be helpful to learn more about allergies and the impact of environment on allergic responses, as that could be a focus of my project. l, D. H., Nielsen, U. N., & Six, J. (2015). Soil biodiversity and human health. *Nature*, *528*(7580), 69–76. https://doi.org/10.1038/nature15744 This could be interesting because it could show how a factor to consider in my project is soil biodiversity and what implications that has on human health.
Follow up Questions	How can the effects of the microbial community be predicted for each individual? How can the different effects of gut microbiome in different cultures be tracked to the gut brain axis?

Notes:

- Gut microbiota diversity decrease related to western diet and lifestyle
- Causes increase of several diseases and other health concerns
- Immune system develops in regard to the gut microbiome
- Evidence shows that gut microbiome system is directly transmitted from mother at birth
- Role of environment
- Proposed that reduced exposure to beneficial microbes due to urbanization leads to increased susceptibility to immune related diseases
- Proposed that improved sanitation and reduced microbial exposure contribute to immune system dysregulation
- Proposed that environmental biodiversity leads to decrease in gut microbiome diversity, potentially impacting health
- Factors influencing gut microbiome health:
    - Urbanization – reduced microbial exposure
    - Antibiotic use – alters microbial communities
    - Sanitation – reduce beneficial microbial exposure
    - Dietary changes – high processed foods impacts microbial composition
    - Soil biodiversity – could positively influence gut microbiome diversity if exposed to diverse soil microbes
- Showed that early life microbial exposure is critical for good gut microbiome health for healthy immune system

Article #14 Notes: Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank

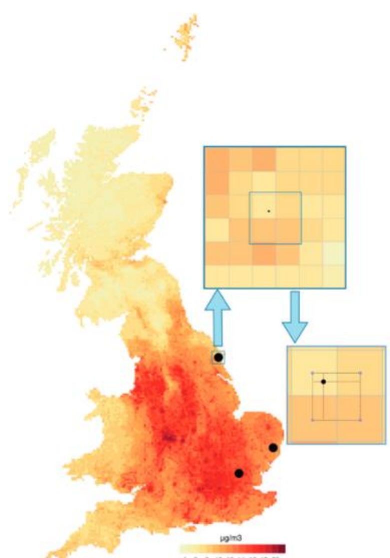
Citation:

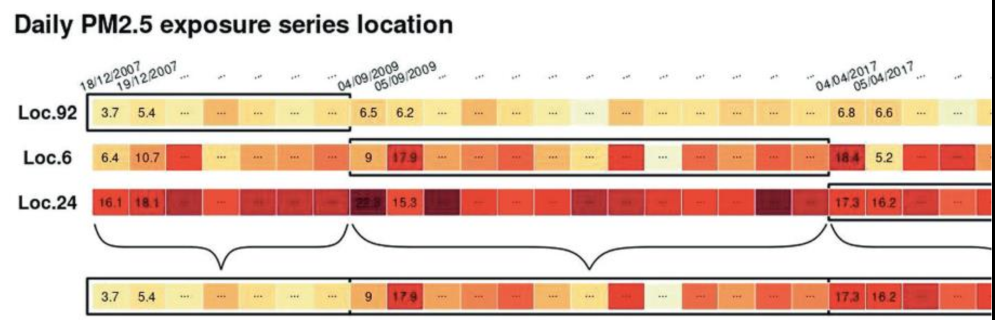
Vanoli, J., Mistry, M. N., De La Cruz Libardi, A., Masselot, P., Schneider, R., Ng, C. F. S., Madaniyazi, L., & Gasparrini, A. (2024). Reconstructing individual-level exposures in cohort analyses of environmental risks: An example with the UK Biobank. *Journal of Exposure Science & Environmental Epidemiology*, *34*(6), 1012–1017. https://doi.org/10.1038/s41370-023-00635-w

A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain. (n.d.). Retrieved December 16, 2025, from https://www.mdpi.com/2072-4292/12/22/3803

Schneider, R., Vicedo-Cabrera, A. M., Sera, F., Masselot, P., Stafoggia, M., de Hoogh, K., Kloog, I., Reis, S., Vieno, M., & Gasparrini, A. (2020). A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain. *Remote Sensing*, *12*(22), 3803. https://doi.org/10.3390/rs12223803

Source Title	Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank
Source citation (APA Format)	oli, J., Mistry, M. N., De La Cruz Libardi, A., Masselot, P., Schneider, R., Ng, C. F. S., Madaniyazi, L., & Gasparrini, A. (2024). Reconstructing individual-level exposures in cohort analyses of environmental risks: An example with the UK Biobank. *Journal of Exposure Science &*

	Environmental Epidemiology, *34*(6), 1012–1017. https://doi.org/10.1038/s41370-023-00635-w
Original URL	https://pmc.ncbi.nlm.nih.gov/articles/PMC11618064/
Source type	Research Paper
Keywords	Epidemiology; Exposure Modeling; Air pollution; Exposure linkage
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	The goal of this paper was to address the methodological challenge of assigining individual level environmental exposures to large cohorts, especially for people who relocated. The required data was individual residential histories and high resolution spatio temporal maps. Then the maps were aligned to extract the time series of address specific exposure values. Then the exposure was reconstructed, accounting for mobilitiy in residential history, and finally, a flexible summary of exposure was generated to tailor to specific research questions. They did a case study on particulate matter and linked each participant to an exposure level on pollution due to particulate matter in the atmosphere.
Research Question/Problem/ Need	Assigning individual-level exposures to large population-based cohorts poses methodological and practical problems
Important Figures	 This figure is important because it shows an example of the spatio temporal map that they used to create the exposure summaries.

	 This figure is important because it shows the time series of exposure to particulate matter at a certain place over a period of time.
VOCAB: (w/definition)	Case crossover: epidemiological study that is used to investigate triggers of sudden events by comparing exposure before event with same person's exposure during other control time periods Spatio temporal maps: a way to visualize data that has spatial and time compontents ICD-10 Code: alphanumerical medical codes to classify and record disease
Cited references to follow up on	*tellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain*. (n.d.). Retrieved December 16, 2025, from https://www.mdpi.com/2072-4292/12/22/3803 This reference is interesting to see the application of machine learning to reconstruct environmental exposures and to see if I can integrate this or something similar into my work. neider, R., Vicedo-Cabrera, A. M., Sera, F., Masselot, P., Stafoggia, M., de Hoogh, K., Kloog, I., Reis, S., Vieno, M., & Gasparrini, A. (2020). A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain. *Remote Sensing*, *12*(22), 3803. https://doi.org/10.3390/rs12223803 This is also interesting because it is another machine learning approach to environmental exposures that could be integrated into my approach.
Follow up Questions	Can this process be applied to multiple environmental exposures at once? Does this process take the ever changing climate of climate change into account?

Notes:

- Designed linkage framework to reconstruct environmental exposures for individual level epidemiological analyses
- Framework requires individual residential histories with related time periods and high resolution spatio-temporal maps of environmental exposures
- Research suggests health risks associated with a variety of environmental exposures such as pollutants, temperature, pollen, and other chemicals
- Exposures cause considerable health burdens but are usually associated with relatively low health risks at the individual level so estimating such associations requires large studies
- Large population based cohorts: European EPIC, UK Biobank, Japanese JECS
- Modern methodologies can combine mutli domain predictors in sophisticated analytical models to derive high resolution spatio temporal maps over large regions

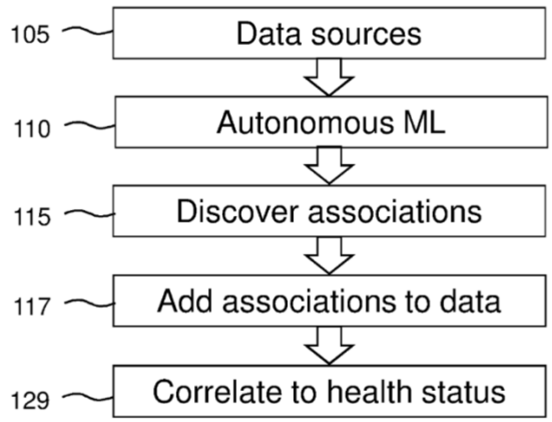
Patent #2 Notes:

Citation:

III, C. F. S., & Polk, J. B. (2019). *Methods for predicting or detecting disease* (United States Patent No. US20190108912A1). https://patents.google.com/patent/US20190108912A1/en

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Source Title	Methods for Predicting or Detecting Disease
Source citation (APA Format)	C. F. S., & Polk, J. B. (2019). *Methods for predicting or detecting disease* (United States Patent No. US20190108912A1). https://patents.google.com/patent/US20190108912A1/en
Original URL	https://patents.google.com/patent/US20190108912A1/en
Source type	Patent
Keywords	Autonomous ML, Health
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	This project presents a way to utilize machine learning to accurately predict patterns within health data in order to predict disease. Usually, people are only able to get diseases diagnosed after going to a doctor and they won't go all the time, so this is helpful to identify and diagnose diseases without having to see a doctor all the time. Machine Learning models are able to repeatedly see patterns that happen in data and then extrapolate them to new situations when it sees those patterns. The machine learning model can also make predictions that correlate to later onset of diseases which is helpful for early prevention of disease. There are many different types of machine learning models that could be used such as neural networks or random forest.
Research Question/Problem/	There is a need to identify patterns within clinical health data in order to

Need	predict disease.
Important Figures	 This figure shows a diagram of the methodology and is important to understand how this paper created their approach to health prediction.
VOCAB: (w/definition)	Electronic Medical Record: digital version of patient's paper chart Support Vector Machine: machine learning algorithm used for classification and regression tasks Prognosis: medical prediction of how disease or condition is expected to progress over time
Cited references to follow up on	n, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785 This references is interesting because it refers to XGBoost, a model I am using for my preliminary research, so it could be good to read this paper and learn more about how it is used. gedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). *Going Deeper with Convolutions* (No. arXiv:1409.4842). arXiv. https://doi.org/10.48550/arXiv.1409.4842

	This paper describes another type of machine learning model that I am interested in learning about since it was used for health predictions in this patent.
Follow up Questions	What types of models perform more accurately than other types of models? What kinds of data would be helpful for pattern prediction beyond medical history? Does a model that generally predicts disease work more accurately and efficiently than several models that each predict one disease?

Notes:

- Computer implemented methods and systems for discovering patterns within large sets of clinical data that are predictive of disease or specific health status
- Training phase and prediction phase
- Training
    - Accessing clinical data
    - Autonomous machine learning
    - Discovering associations
    - Correlating to health status
    - Refining data
- Prediction
    - Providing patient data
    - Pattern matching
    - Predicting health issues

Article #15 Notes: Algorithmic fairness in artificial intelligence for medicine and healthcare

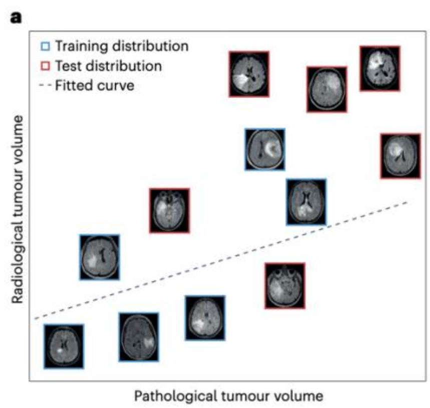
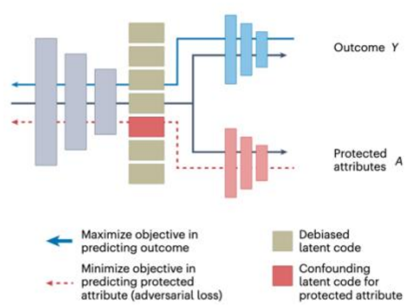
Citation:

Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., Sahai, S., & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering*, *7*(6), 719–742. https://doi.org/10.1038/s41551-023-01056-8

Howard, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O. I., Kather, J. N., Cipriani, N., Grossman, R. L., & Pearson, A. T. (2021). The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nature Communications*, *12*(1), 4423. https://doi.org/10.1038/s41467-021-24698-1

Kamishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012). Fairness-Aware Classifier with Prejudice Remover Regularizer. In P. A. Flach, T. De Bie, & N. Cristianini (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 35–50). Springer. https://doi.org/10.1007/978-3-642-33486-3_3

Source Title	Algorithmic fairness in artificial intelligence for medicine and healthcare
Source citation (APA Format)	n, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., Sahai, S., & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering*, *7*(6), 719–742. https://doi.org/10.1038/s41551-023-01056-8
Original URL	https://doi.org/10.1038/s41551-023-01056-8
Source type	Research Paper
Keywords	Machine Learning, fairness, AI for Health, health
#Tags	#environment-on-health

Summary of key points + notes (include methodology)	Machine learning, if not fair and equitable, can be detrimental to healthcare systems and cause more disparities across patient subpopulations. So, this paper goes through the biases that are present in some AI strategies and how to mitigate them. It identifies several biases such as ones present in data or the model structure. It also identifies several solutions such as changing the overall data that the model gets access to, what features the model gets access to, or making it more explainable.
Research Question/Problem/ Need	Is the use of machine learning in health fair, and if not, what are the causes and how can we mitigate the results so they are more fair?
Important Figures	 This figure shows an example of how the model performs when predicting tumors. This is important in order to understand how models perform in practice.  This figure shows how making one attribute protected could lead to more accurate results. This figure is important to understand one of the proposed solutions.
VOCAB: (w/definition)	Algorithmic bias: systematic and repeated errors in AI system that cause unfair outcomes True Positive rate: fairness metric that shows proportion of truly sick people identified as sick correctly (also known as recall or sensitivity) False positive rate: proportion of truly healthy people incorrectly identified as sick (if this is high, it can lead to unnecessary harmful treatments)

	Health disparities: difference in health status or healthcare access due to social, economic, or environmental disadvantages that arise in population groups
Cited references to follow up on	oward, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O. I., Kather, J. N., Cipriani, N., Grossman, R. L., & Pearson, A. T. (2021). The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nature Communications*, *12*(1), 4423. https://doi.org/10.1038/s41467-021-24698-1 I want to follow up on this ciation because it is important to understand this impact if I want to train my own deep learning model but make sure that it is fair. ishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012). Fairness-Aware Classifier with Prejudice Remover Regularizer. In P. A. Flach, T. De Bie, & N. Cristianini (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 35–50). Springer. https://doi.org/10.1007/978-3-642-33486-3_3 This paper could help me understand how to make a deep learning predictive model that is more fair, or how to add this fairness-aware classifier as part of my predictive pipeline in order to have more fair results.
Follow up Questions	What fairness metrics should be prioritized in these models? How can we distinguish between correlations between a demographic and health outcome that are not causational and the ones that are for more accurate predictions?

Notes:

- Current analysis of AI models that are deployed in healthcare situations show evidence of widespread inequality in how patients are diagnosed, treated, and even billed
- One real world example is a model that learned to use healthcare costs as a proxy for medical illness and lead to the prioritization of helathier white patients over sicker black patients
- Sources of bias:
    - Data that doesn’t represent sufficient diversity for minority or marginalized groups
    - Some features that are used by algorithms are less predictive for certain groups
    - Data that is sourced from a single unrepresentative healthcare system creates selection bias
    - If there are inconsistent labels or biased labels done by clinicians the model learns those patterns
    - Some bias occurs if the clinicians rely too heavily on the model’s outputs automatically
- Proposed solutions:
    - Teaching the model to value fair representation by making them invariant to protected attributes or unaware of features such as race (can this lead to lower accuracy because those features might lead to patterns that are actually insightful and important?)
    - Training model on more diverse data from multiple sites
    - Using explainable AI to make the model’s decision process more transparent in order to identify which features lead to the output

Article #16 Notes: Combinatorial prediction of therapeutic perturbations using causally inspired neural networks

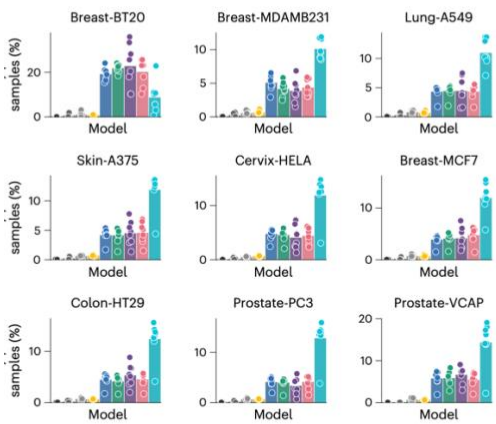
Citation:

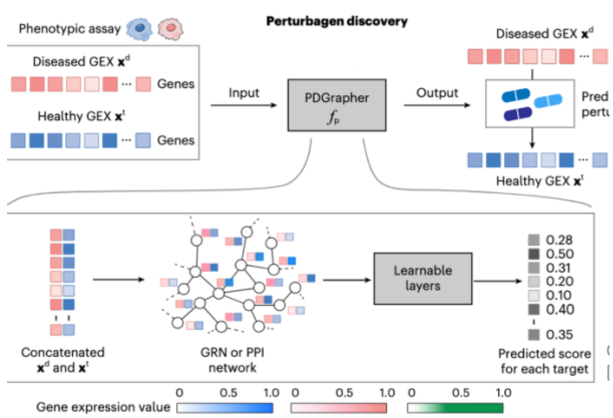
Gonzalez, G., Lin, X., Herath, I., Veselkov, K., Bronstein, M., & Zitnik, M. (2024). *Combinatorial prediction of therapeutic perturbations using causally-inspired neural networks*. Bioinformatics. https://doi.org/10.1101/2024.01.03.573985

Zhu, J., Wang, J., Wang, X., Gao, M., Guo, B., Gao, M., Liu, J., Yu, Y., Wang, L., Kong, W., An, Y., Liu, Z., Sun, X., Huang, Z., Zhou, H., Zhang, N., Zheng, R., & Xie, Z. (2021). Prediction of drug efficacy from transcriptional profiles with deep learning. *Nature Biotechnology*, *39*(11), 1444–1452. https://doi.org/10.1038/s41587-021-00946-z

Yuan, B., Shen, C., Luna, A., Korkut, A., Marks, D. S., Ingraham, J., & Sander, C. (2021). CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. *Cell Systems*, *12*(2), 128-140.e4. https://doi.org/10.1016/j.cels.2020.11.013

Source Title	Combinatorial prediction of therapeutic perturbations using causally inspired neural networks
Source citation (APA Format)	zalez, G., Lin, X., Herath, I., Veselkov, K., Bronstein, M., & Zitnik, M. (2024). *Combinatorial prediction of therapeutic perturbations using causally-inspired neural networks*. Bioinformatics. https://doi.org/10.1101/2024.01.03.573985
Original URL	https://doi.org/10.1101/2024.01.03.573985

<table>
<tr><td>Source type</td><td>Research Paper</td></tr>
<tr><td>Keywords</td><td>Graph Neural Network, Phenotype perturbations, disease countering, machine learning for health</td></tr>
<tr><td>#Tags</td><td></td></tr>
<tr><td>Summary of key points + notes (include methodology)</td><td>PDGrapher solves the inverse problem of finding the perturbagens that will create the desired change in the phenotypes. It does this by embedding the cell states into a biological network and then learns to map the desired phenotype changes to combinatorial interventions. It’s results are that it significantly outperforms other models in accuracy and efficiency.</td></tr>
<tr><td>Research Question/Problem/ Need</td><td>Currently, identifying disease countering phenotypes is a time-consuming and resource heavy process.</td></tr>
<tr><td>Important Figures</td><td>


This figure is important because it shows that the model is able to generalize to new situations which is necessary for accurate predictions when applied in real situations.


This figure is important because it shows the structure of the graph neural network.</td></tr>
</table>

VOCAB: (w/definition)	Perturbations: intentional changes applied to biological systems Phenotypes: observable characteristics or measurable biological states Graph neural network: neural network that learns from data structured as graphs Perturbagens: specific agents that induce perturbations in biological system
Cited references to follow up on	, J., Wang, J., Wang, X., Gao, M., Guo, B., Gao, M., Liu, J., Yu, Y., Wang, L., Kong, W., An, Y., Liu, Z., Sun, X., Huang, Z., Zhou, H., Zhang, N., Zheng, R., & Xie, Z. (2021). Prediction of drug efficacy from transcriptional profiles with deep learning. *Nature Biotechnology*, *39*(11), 1444–1452. https://doi.org/10.1038/s41587-021-00946-z This is interesting to follow up on because it is an application of deep learning for health so I can learn how they trained their model best and use a similar method. n, B., Shen, C., Luna, A., Korkut, A., Marks, D. S., Ingraham, J., & Sander, C. (2021). CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. *Cell Systems*, *12*(2), 128-140.e4. https://doi.org/10.1016/j.cels.2020.11.013 This is good to follow up on because it could show me how to make my model more interpretable so users are able to understand where the model results come from and trust the results more.
Follow up Questions	Can this be applied with genes instead of phenotypes? What if it took the person's individual biography into account as well?

Notes:

- Phenotype identification can help distinguish healthy from diseased cells
- This paper introduces PDGraph, a causally inspired neural network that is capable of reversing disease phenotypes
- Instead of predicting how the perturbagens will affect the phenotypes, this model takes the inverse problem and identifies which perturbagens will be needed to get the desired response
- Trains up to 25x faster than other models
- Model is trained on experimental perturbations and signatures

Article #17 Notes: The impact of site-specific digital histology signatures on deep learning model accuracy and bias

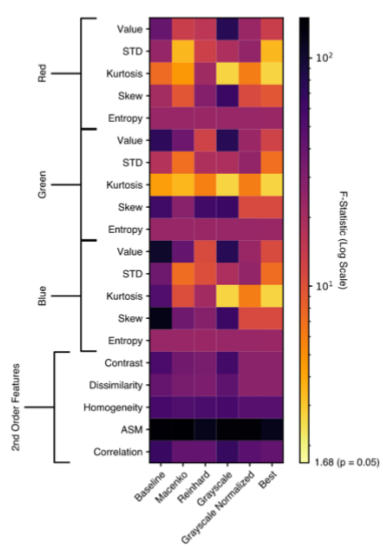
Citation:

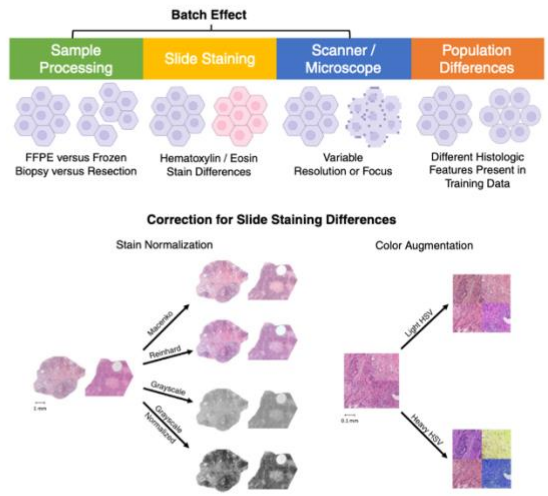
Howard, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O. I., Kather, J. N., Cipriani, N., Grossman, R. L., & Pearson, A. T. (2021). The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nature Communications*, *12*(1), 4423. https://doi.org/10.1038/s41467-021-24698-1

Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., Hulsbergen-van de Kaa, C., & Litjens, G. (2020). Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study. *The Lancet. Oncology*, *21*(2), 233–241. https://doi.org/10.1016/S1470-2045(19)30739-9

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier* (No. arXiv:1602.04938). arXiv. https://doi.org/10.48550/arXiv.1602.04938

Source Title	The impact of site-specific digital histology signatures on deep learning model accuracy and bias
Source citation (APA Format)	vard, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O. I., Kather, J. N., Cipriani, N., Grossman, R. L., & Pearson, A. T. (2021). The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nature Communications*, *12*(1), 4423. https://doi.org/10.1038/s41467-021-24698-1

Original URL	https://doi.org/10.1038/s41467-021-24698-1
Source type	Research Paper
Keywords	Histology, deep learning, ai for health
#Tags	
Summary of key points + notes (include methodology)	Public data, such as the Cancer Genome Atlas come from several different sources and are used to train CNNs, but they have their own site specific errors that come in batches. The model can learn to exploit these site specific errors and exploit them leading to decreased accuracy overall. The authors analyze over 3000 data points from TCGA and also treat site detection as its own task. They test model performance on this site detection task to identify how much the models are able to identify site specific signatures. Furthermore, the paper applies cross validation strategies holding out entire site data and then evaluating to see how the model results differ. The authors find that when a model is being trained on multi-site data, it is important to take site preservation into account and make sure that the model doesn't become over reliant on patterns from specific sites.
Research Question/Problem/ Need	How do site specific histology signatures affect deep learning model accuracy?
Important Figures	 This figure is important because it shows the variance in the histology of the first order and second order cancer images.

	 This figure is important because it shows the cause of different site specific histologies.
VOCAB: (w/definition)	Histology: common collection of image attributes Stain: intensity of image Site-specific: data collected from a specific medical center Data augmentation: rotating, changing images slightly to increase variance and reduce overfitting Cross-validation: holding out entire site data and then evaluating to ensure generalization
Cited references to follow up on	lten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., Hulsbergen-van de Kaa, C., & Litjens, G. (2020). Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study. *The Lancet. Oncology*, *21*(2), 233–241. https://doi.org/10.1016/S1470-2045(19)30739-9 This citation is interesting because it is an automated deep learning pipeline so I could apply this methodology into my work. eiro, M. T., Singh, S., & Guestrin, C. (2016). *“Why Should I Trust You?”: Explaining the Predictions of Any Classifier* (No. arXiv:1602.04938). arXiv. https://doi.org/10.48550/arXiv.1602.04938

	This reference is interesting because it talks more about how to make a classifier interpretable which could be very helpful for my project.
Follow up Questions	What training strategies reduce site specific reliance the most? How to separate biological variance from technological variance?

Notes:

- **Whole slide images of tissue to study cancer and diseases**
- **Deep learning used to predict cancer subtypes, predict mutations, predict survival outcomes, predict patient demographics**
- **Models are trained on large public datasets**
- **Data comes from many hospitals and labs called sites**
- **Sites have different staining protocols, scanner hardware, image resolution, slide preparation methods**
- **Differences introduce non biological variance into images**
- **Investigates whether deep learning models are unintentionally learning which site slide comes from**
- **This would mean that models do well at testing but would fail when applied to a new hospital**
- **Some models use shortcut learning: learn to recognize features that are unrelated to actual output such as site that the image was taken**
- **Uses cancer genome atlas dataset**
- **Creates task and measures whether a model can identify what site a slide was submitted from**
- **Applies image normalization and augmentation techniques and sees if site detectability decreases**
- **Test standard cross validation and preserved site cross validation**
- **Standard cross validation overestimates accuracy that shows up in preserved site cross validation (more similar to real life situations)**
- **This paper shows that models need to take into account site bias in order for maximum fairness**

Article #18 Notes: Healthcare predictive analytics using machine learning and deep learning techniques: a survey

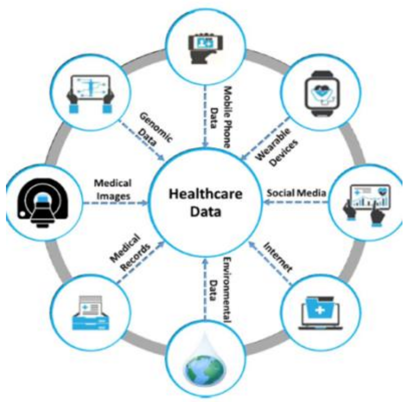
Citation:

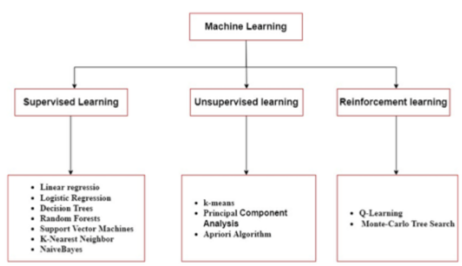
Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: A survey. *Journal of Electrical Systems and Information Technology*, *10*(1), 40. https://doi.org/10.1186/s43067-023-00108-y

Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in Using IoT with Machine Learning in Health Prediction System. *Forecasting*, *3*(1), 181–206. https://doi.org/10.3390/forecast3010012

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, *99*, 650–655. https://doi.org/10.1016/j.procir.2021.03.088

Source Title	Healthcare predictive analytics using machine learning and deep learning techniques: a survey
Source citation (APA Format)	awy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: A survey. *Journal of Electrical Systems and Information Technology*, *10*(1), 40. https://doi.org/10.1186/s43067-023-00108-y
Original URL	https://doi.org/10.1186/s43067-023-00108-y
Source type	Research Paper
Keywords	Healthcare prediction, Artificial intelligence (AI), Machine learning (ML), Deep learning (DL), Medical diagnosis
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	This paper give s a review of how machine learning and deep learning techniques are applied in healthcare predictive analysis. It outlines the

<table>
<tr><td></td><td>different types of medical healthcare data that is used for predictive modeling and shows that data quality and preprocessing are important to high quality models. Then it goes through the different types of machine learning and deep learning models and what each one is used for. Finally, it discusses the limitations and evaluation metrics used in AI research.</td></tr>
<tr><td>Research Question/Problem/ Need</td><td>How well are current machine learning and deep learning methods able to predict health and disease risk?</td></tr>
<tr><td>Important Figures</td><td>


This figure is important to understand the different sources that can contribute to healthcare data.


This figure is important to understand the different types of machine learning models that exist.</td></tr>
<tr><td>VOCAB: (w/definition)</td><td>Supervised learning: model is trained on labeled data
Linear regression: method used to predict continuous numerical values
Logistic regression: used for classification tasks
Naive Bayes: simple probabilistic classifier
Long short term memory: special kind of recurrent neural network capable of learning long term dependencies</td></tr>
</table>

Cited references to follow up on	Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in Using IoT with Machine Learning in Health Prediction System. *Forecasting*, *3*(1), 181–206. https://doi.org/10.3390/forecast3010012 I want to understand how IoT is used in the medical industry so this could be helpful. lemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, *99*, 650–655. https://doi.org/10.1016/j.procir.2021.03.088 Since I am planning on testing LSTMs in my project, it would be helpful to see how they have been used previously.
Follow up Questions	Can models be used to help improve the quality of medical data? How does using synthetic data affect the accuracy of the model?

Notes

- Healthcare systems generate lots of data every day
- Predictive analytics aim to anticipate future patient outcomes in order to support decision making for clinicians and improve efficiency and quality of care
- Machine learning allows computes to learn patterns from data and then make predictions without explicit algorithms telling them what to predict
- Types of healthcare data: HER, medical images, time series physiological data, genomic and omics data, wearable data
- Data preprocessing includes cleaning data, handling missing values, normalization, noise removal, feature engineering – all done before modeling
- Traditional ML relies heavily on handcrafted features whereas deep learning does not need as much manual feature engineering
- There are many tyeps of machine learning techniques
- Logistic regression is used for binary classification
- Decision tress are rule based and interpretable although prone to overfitting
- Support vector machines are effective in high dimensional spaces
- K-nearest neighbors is simple, but computationally expensive
- Naïve bayes is based in proability and fast but assumes each feature is independent
- There are many deep learning techniques in healthcare
- CNNs are used for image dta
- RNNs are used for sequential and time dependent data
- Long shorter term memory is a type of RNN used for solving the vanishing gradient problem and is popular for HER time series modeling
- Autoencoders learn compressed representations
- Applications include: disease diagnosis, disease prognosis, risk prediction, imaging analsyis, clicnical decision support, personalized medicine
- Evaluation metrics: accuracy, precision, recall, F1 score, AUROC
- Limitations: data quality issues, bias and fairness, interpretability, generalization, privacy and security

Article #19 Notes: Machine Learning Applications in Studying Mental Health Among Immigrants and Racial and Ethnic Minorities: A Systematic Review

Citation:

Park, K. K., Ahmed, A., & Al-Garadi, M. A. (2023). *Machine Learning Applications in Studying Mental Health Among Immigrants and Racial and Ethnic Minorities: A Systematic Review* (No. arXiv:2304.09233). arXiv. https://doi.org/10.48550/arXiv.2304.09233

Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Transactions on Computer-Human Interaction*, *27*(5), 1–53. https://doi.org/10.1145/3398069

Liu, Y., Qu, H.-Q., Mentch, F. D., Qu, J., Chang, X., Nguyen, K., Tian, L., Glessner, J., Sleiman, P. M. A., & Hakonarson, H. (2022). Application of deep learning algorithm on whole genome sequencing data uncovers structural variants associated with multiple mental disorders in African American patients. *Molecular Psychiatry*, *27*(3), 1469–1478. https://doi.org/10.1038/s41380-021-01418-1

Source Title	Machine Learning Applications in Studying Mental Health Among Immigrants and Racial and Ethnic Minorities: A Systematic Review
Source citation (APA Format)	, K. K., Ahmed, A., & Al-Garadi, M. A. (2023). *Machine Learning Applications in Studying Mental Health Among Immigrants and Racial and Ethnic Minorities: A Systematic Review* (No. arXiv:2304.09233). arXiv. https://doi.org/10.48550/arXiv.2304.09233
Original URL	https://doi.org/10.48550/arXiv.2304.09233

Source type	Research Paper
Keywords	Machine learning; mental health; minorities, disparities
#Tags	
Summary of key points + notes (include methodology)	Machine learning is being used increasingly in the medical field, especially for identifying and helping mental health. This paper aims to understand how machine learning has been used for identifying mental health in minority groups such as immigrants or racial and ethnic minorities. They searched up papers on the Google Scholar database and ultimately had 12 papers included in their survey. The survey showed that their has been use of machine learning for mental health applications for minority groups, but it is still under development and needs further work.
Research Question/Problem/ Need	How is machine learning currently applied to study mental health among immigrants and racial and ethnic minorities?

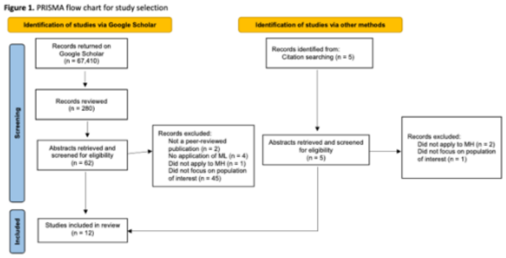
Important Figures

Table 2. Publication analysis

Characteristic	N	%	Reference
Year of publication			
2017	2	16.7%	32,33
2020	2	16.7%	42 43
2021	5	41.7%	36,37,39-41
2022	3	25.0%	34,35,38
Region			
Asia			
Turkey	1	8.3%	35
Jordan	1	8.3%	34
Europe			
United Kingdom	1	8.3%	37
Germany	2	16.7%	33,36
Switzerland	1	8.3%	43
US	6	50.0%	32,38-42
Population of focus			
Refugees	5	41.7%	33-37
Hispanics	3	25.0%	32,38,39
African Americans	2	16.7%	40,41
Korean immigrants	1	8.3%	42
European immigrants	1	8.3%	43

This figure is important because it shows a summary of all the different literatures analyzed in this survey.

Figure 1. PRISMA flow chart for study selection



This figure is important because it shows how studies were selected to be part of this literature review.

VOCAB: (w/definition)	Schizophrenia: mental disorder affecting how person thinks, feels, and behaves

	Sociodemographic data: combination of social and demographic characteristics to describe individual
Cited references to follow up on	hieme, A., Belgrave, D., & Doherty, G. (2020). Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Transactions on Computer-Human Interaction*, *27*(5), 1–53. https://doi.org/10.1145/3398069 This reference is interesting because it focuses on human computer interactions which could be helpful in making my project more accessible to human users. Y., Qu, H.-Q., Mentch, F. D., Qu, J., Chang, X., Nguyen, K., Tian, L., Glessner, J., Sleiman, P. M. A., & Hakonarson, H. (2022). Application of deep learning algorithm on whole genome sequencing data uncovers structural variants associated with multiple mental disorders in African American patients. *Molecular Psychiatry*, *27*(3), 1469–1478. https://doi.org/10.1038/s41380-021-01418-1 This reference is interesting because it uses machne learning and genes to identify mental health disorders.
Follow up Questions	How can similar machine learning techniques be applied for physical health rather than mental health? Do certain machine learning models perform better than others on predicting mental health outcomes?

Notes:

- Machine learning is being used increasingly for metnal health research
- ML is used to analyze more complex newer data types
- This survey systematically analyzes published literature to identify potential gaps in current use of ML to study mental health of immigrants, refugees, migratnts, and racial and ethnic minorites
- Quieried google scholar
- Filtered studies based on their requirements
    - Must be application of ML for MH
    - Must focus on minority for that respective country
- Found over 67 thousand articles
- Only included 12 articles
    - Each published within last 6 eyars
    - 50% studied populations in the US
    - Most used supervised learning
    - Some analyzed up to 16 models to find best model
    - Some did not discuss cross validation strategy
- Clinical application of Mahcine leanring for mental health specifically for minority groups is still under development

Article #20 Notes: Towards modeling evolving longitudinal health trajectories with a transformer based deep learning model

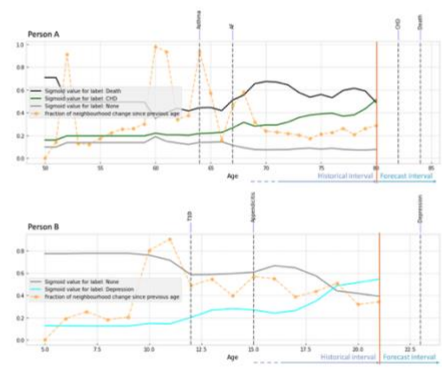
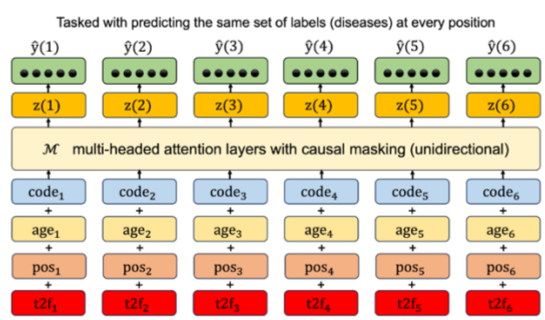
Citation:

Moen, H., Raj, V., Vabalas, A., Perola, M., Kaski, S., Ganna, A., & Marttinen, P. (2024). *Towards modeling evolving longitudinal health trajectories with a transformer-based deep learning model* (No. arXiv:2412.08873). arXiv. https://doi.org/10.48550/arXiv.2412.08873

Choi, E., Xu, Z., Li, Y., Dusenberry, M. W., Flores, G., Xue, Y., & Dai, A. M. (2020). *Learning the Graphical Structure of Electronic Health Records with Graph Convolutional Transformer* (No. arXiv:1906.04716). arXiv. https://doi.org/10.48550/arXiv.1906.04716

Kumar, Y., Ilin, A., Salo, H., Kulathinal, S., Leinonen, M. K., & Marttinen, P. (2024). Self-Supervised Forecasting in Electronic Health Records With Attention-Free Models. *IEEE Transactions on Artificial Intelligence*, *5*(8), 3926–3938. https://doi.org/10.1109/TAI.2024.3353164

Source Title	Towards modeling evolving longitudinal health trajectories with a transformer based deep learning model
Source citation (APA Format)	n, H., Raj, V., Vabalas, A., Perola, M., Kaski, S., Ganna, A., & Marttinen, P. (2024). *Towards modeling evolving longitudinal health trajectories with a transformer-based deep learning model* (No. arXiv:2412.08873). arXiv. https://doi.org/10.48550/arXiv.2412.08873
Original URL	https://doi.org/10.48550/arXiv.2412.08873
Source type	Research Paper
Keywords	Longitudinal health trajectories · Longitudinal data · Disease prediction ·

	Transformers
#Tags	#environment-on-health
Summary of key points + notes (include methodology)	This paper introduces Evolve, a transformed based deep learning model that is designed for handling evolving longitudinal health trajectories. This model will give continuous mutli time point predictiosn rather than one off disease predictions. It applies causal attention masking in order to ensrue that the temporal order remains the same (past affects future but not the other way around). It modifies the transformer training objective to give continuous predcitions. Evolve gives similar results in accuracy but gives much more insight into how the health changes continuously over time.
Research Question/Problem/ Need	What is the best way to continuously montiro people's health trajectories and enable interventions in ongoing health trajectories?
Important Figures	 This figure is important because it shows 2 example health progressions over time showing how health can drastically improve and worsen.  This figure shows a diagram of the model's structure which is helpful to understand how the model is structured and how the inputs become the outputs.
VOCAB: (w/definition)	Unidirectional transformer model: neural network architecture that processes input data in only one direction Clinical endpoints: specific measurable health outcomes Sigmoid values: numbers between 0 and 1 that result from passing data

	through sigmoid function
Cited references to follow up on	hoi, E., Xu, Z., Li, Y., Dusenberry, M. W., Flores, G., Xue, Y., & Dai, A. M. (2020). *Learning the Graphical Structure of Electronic Health Records with Graph Convolutional Transformer* (No. arXiv:1906.04716). arXiv. https://doi.org/10.48550/arXiv.1906.04716 This reference is interesting because it focuses on a transformer specifically for HER data which could be helpful for my project. nar, Y., Ilin, A., Salo, H., Kulathinal, S., Leinonen, M. K., & Marttinen, P. (2024). Self-Supervised Forecasting in Electronic Health Records With Attention-Free Models. *IEEE Transactions on Artificial Intelligence*, *5*(8), 3926–3938. https://doi.org/10.1109/TAI.2024.3353164 This reference is interesting because it could be another way to structure my model in my project to best handle the data.
Follow up Questions	How can the model be adapted to consider multimodal data or multimodal omics? How can the model take interventions such as medications or a sudden shift in location into account?

Notes:

- **Health trajectories are a sequence of health events**
- **Longitudinal health registers capture sequences for large populations**
- **Traditional models only make a single prediction for future window**
- **There is a need for models that can capture the evolving risk over time allowing for early detection, continuous monitoring, and interpretation of how the risk profile changes over time**
- **Transformers work well for this problem due to self attention and long range dependency monitoring**
- **Used nationwide longitudinal health data**
- **Features sucha s clinical codes, procedure codes, and drug purchases**
- **Data spans many years**
- **Model processes sequence of events in chronological order**
- **Model predicts disease onset at each time step**
- **Causal attention mask ensures model only uses past not future data**
- **Designed to predict multiple diseases at once**
- **Each data element includes medical code, time since previous event, age at event, and position in sequence**
- **Inputs are embedded into continuous vectors that capture medical semantics and temporal context**
- **Transformer produces latent embeddings that encode health state evolution**
- **Loss is computed for each disease at each time stamp**
- **Model learns what patterns indicate rising risk and how early events lead to later events**
- **Compares results to bidirectional transformers, logistic regression, gradient boosted trees**
- **Evolve matches baseline prediction performance for disease onset**
- **Some limitations are that it requires high quality longitudinal data, is computationally intensive, and not very interpretable for user**